Applications using infrared spectroscopy to detect and bridge the variability and heterogeneity before and after fruit processing : a case study on apple purees

Weijie Lan

THÈSE DE DOCTORAT D'AVIGNON UNIVERSITÉ

École Doctorale N° 536
Agrosciences & Sciences

Spécialité / Discipline de doctorat :
Chimie

INRAE UMR 408 Sécurité et Qualité des Produits d'Origine Végétale

Présentée par
Weijie Lan

Applications utilisant la spectroscopie infrarouge pour détecter et relier la variabilité et l'hétérogénéité avant et après transformation des pommes : cas des purées

Applications using infrared spectroscopy to detect and bridge the variability and heterogeneity before and after fruit processing: a case study on apple purees

---

Soutenue publiquement le 23/09/2021 devant le jury composé de :

Mme Nathalie Dupuy, Professeure-HDR, Université Aix-Marseille **Rapporteure**
M. Jean-Michel Roger, Directeur de Recherche-HDR, INRAE Montpellier **Rapporteur**
M. Vincent Baeten, Directeur Scientifique, CRA-Wallonie **Examinateur**
M. Pierre Picouet, Maitre de Conférences-HDR, ESA Angers, **Examinateur**
Mme Catherine Renard, Directrice de Recherche-HDR, INRAE Nantes **Directrice de thèse**
Mme Sylvie Bureau, Ingénieure de Recherche-HDR, INRAE Avignon **Co-directrice de thèse**
M. Benoit Jaillais, Ingénieur de Recherche, INRAE Nantes, **Encadrant**
M. Alexandre Leca, Chargé de Recherche, INRAE Avignon, **Encadrant**

“雄关漫道真如铁，而今迈步从头越”

However difficult it might seem, the challenge will be overcome.

毛泽东《忆秦娥 娄山关》

Acknowledgements

In the beginning, I would like to thank Chinese Scholarship Council (CSC) and Ms. Prof. Wen Qin from Sichuan Agricultural University. They supported me a doctoral grant and provided me a chance to live a unique experience in France.

I would like to express my sincere gratitude to Ms. Dr. Catherine M. G. C Renard, who provided me an opportunity to join the research unit UMR 408 “Securité et Qualité des Produits d’Origine Végétale” (SQPOV) at INRAE, and accommodated me into the ‘Interfaces’ project. Without her precious support it would not be possible to conduct this thesis. Thanks for all her experienced and foresighted advices in every task of my work.

I thank to my co-supervisor Ms. Dr. Sylvie Bureau for the continuous support of my Ph.D study and related research, and her patience, motivation, and immense knowledge. Her guidance helped me in all the time of my research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

My sincerely thanks also go to my co-supervisors Benoit Jaillais and Alexandre Leca, for introducing me into the world of chemometric analysis, food texture and rheology. Thank you for your excellent guidance throughout the thesis, allowing me to develop both my personal and scientific abilities.

Apart from my supervisors, I won’t forget to express the gratitude to Mr. Dr. Vincent Baeten, who provided me an opportunity to enrich my skills of Raman technique at Quality and Authentication of Products Unit, Walloon Agricultural Research Centre (CRA-W) in Belgium.

Besides my advisor, I would like to thank my thesis committee: Mr. Dr. Jean-Michel Roger, Ms. Prof. Nathalie Dupuy, and Dr. Pierre Picouet, for evaluating my work within the PhD jury.

I would also like to thank the members of my annual PhD committee: Ms. Dr. Layal Dahdouh, Ms. Dr. Carine Le Bourvellec, and Mr. Dr. Frederic Carlin for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

I thank all the members of ‘Interfaces’ project and Agropolis Fondation for establishing the framework, allowing me to study on a fascinating research question during my thesis.

Thanks to all our members of SQPOV, for their warm welcome and the friendly atmosphere. Particular thanks to Patrice Reling, Barbara Gouble, Line Touloumet, Marielle Boge, Caroline Garcia, Gisèle Riqueau and Romain Bott for their technical help.

I am also pleased to say thank you to my Chinese colleagues at INRAE: Dr. Songchao Chen, Dr. Jiantao Zhao, Dr. Shouyang Liu, Dr, Xiuliang Jin and Dr. Yuan Lin, who gave me several interesting and impressive discussions of my research work. Songchao, thank you for introducing me to the uses of R software to improve my works and your chemometric support throughout the thesis. I am looking forward to further cooperation with you in the near future.

I would always remember my friends at SQPOV, Carla, Alexandra and Xuwei, who always help me to solve several complex problems of scholar and social activities.

In the end, I am grateful to my parents, my girlfriend and all my family members. I consider myself nothing without them. They gave me enough moral support, encouragement and motivation to accomplish the personal goals.

Thank you, Merci & 谢谢!

Table of contents

Table of contents

List of figures

List of figures

List of tables

List of tables

List of abbreviations

List of abbreviations

Abbreviations	Common names
AIS	Alcohol Insoluble Solids
AI	Artificial Intelligence
ANPP	Association Nationale Pommes Poires
ANOVA	Analysis of Variance
ANN	Artificial Neural Network
ANR	Agence Nationale de la Recherche (English: French National Agency)
AS	Amplitude Sweep Tests
ASE	Accelerated Solvent Extractors
ATR-FTIR	Attenuated Total Reflection- Fourier-Transform Infrared
BR	Braeburn
BM	Mealy Braeburn
CA	Chemical Analysis
CIRAD	Centre de Coopération Internationale en Recherche Agronomique pour le Développement (English: Agricultural Research Centre for International Development)
CB	Combination of Vis-NIR and MIR
CRA-W	Walloon Agricultural Research Centre
CSC	Chinese Scholarship Council (CSC)
DA	Discriminant analysis
DW	Dry Weight
DMC	Dry Matter Contents
d(4:3)	Particle sizes averaged over volume
d(3:2)	Particle sizes averaged over surface area
ELM	Extreme learning machine
F1	The First Discriminant Factor
F2	The Second Discriminant Factor
FAO	Food and Agriculture Organization of the United Nations
F&V	Fruits and Vegetables
FAFCM	Fast Allied Fuzzy C-means Clustering Algorithm

NF	Fresh samples
FD	Freeze-dried samples
FDA	Factorial Discriminant Analyses
FT-IR	Fourier-Transform Infrared
FW	Fresh Weight
FS	Frequency Sweep Tests
G’	Storage Modulus
G”	Loss Modulus
GD	Golden Delicious
GS	Granny Smith
GA	Royal Gala
GO	Golden Smoothee
GC	Gas Chromatography
G0	Grinding speed at 300 rpm
G1	Grinding speed at 1000 rpm
G3	Grinding speed at 3000 rpm
HB	Hot Break
HPLC	High Pressure Liquid Chromatography
INRAE	Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (English: National Research Institute for Agriculture, Food and Environment)
LDA	linear discriminant analysis
LOO-PLS	Leave- one- out partial least square
MCR-ALS	multivariate curve resolution-alternative least square
MS	Mass Spectrometry
MALDI–TOF	Matrix-assisted Laser Desorption/Ionization Time-of-flight
MSI	Mass Spectrometry Imaging
M-DUV-cryo-imaging	Multispectral Deep-UV Autofluorescence Cryo-imaging
MIRS	Mid infrared spectroscopy
MRI	Magnetic Resonances Imaging (MRI)
MB-PCA	Multiblock Principal Component Analysis

MSC	Multiplicative Scatter Correction
MW-PLSDA	Moving Window Partial Least Squares Discriminant Analysis
NR	Non-refined
NIRS	Near infrared spectroscopy
NIR-HSI	Near infrared hyperspectral imaging
PC1	The first principal component
PC2	The second principal component
PLS-DA	Partial least squares Discriminant Analysis
PME	Pectin methylesterase
PG	Polygalacturonase
PSE	Pressurized Solvent Extractors
PCA	Principle Component Analysis
PAT	Process analytical technology
PLS	Partial Least Square
QDA	Quadratic Discriminant Analysis
Ra	Refined at 0.5 mm
ROIs	Region of Interests
R_c^2	Determination Coefficients of Calibration
R_v^2	Determination Coefficients of Validation
R_p^2	Determination Coefficients of Prediction
R_{cv}^2	Determination Coefficients of Cross-validation
RMSEP	Root Mean Square Error of Prediction
RMSEC	Root Mean Square Error of Calibration
RMSEV	Root Mean Square Error of Validation
RPD	Residual Predictive Deviation
RF	random forest regression
SNR	Signal-to-Noise Ratio
SQPOV	Securité et Qualité des Produits d'Origine Végétale
SO-PLS	Sequential and Orthogonalized partial least squares
S-cryo-XRF	Synchrotron cryo-X-tray fluorescence microscopy
SSC	Soluble solid contents
SVM	Support Vector Machine
SPA	Successive Projections Algorithm

SNV	Standard Normal Variate
SD	Standard Deviation
Th-	Non-thinning treatment
Th+	Thinning treatment
T0	Apples at harvest
T1	Apples stored at 4°C after one month
T3	Apples stored at 4°C after three month
T6	Apples stored at 4°C after six month
TSC	Total Sugars Contents
TA	Titratable Acidity
VIP	Variable Importance
VIS	Visible spectroscopy
VISSA-SR	Variable Iterative Space Shrinkage Approach with Stepwise Regression
VISNIR-HSI	Visible and near infrared hyperspectral imaging
WT	Wavelet Transform
XRI	X-ray imaging
η_{50}	Purees viscosity at a control share rate of 50 s^{-1}
η_{100}	Purees viscosity at a control share rate of 100 s^{-1}

I. Introduction

1. Background and challenges

Nowadays, the way consumers choose to buy and eat (Corollaro et al., 2013; Harker, Gunson, & Jaeger, 2003) has put much stress on agronomical (production system, genetic resources, climate change, etc.) (Davis, Downs, & Gephart, 2021; Piñeiro et al., 2020) and processing (environmentally friendly, clean label, etc.) systems (Knorr & Augustin, 2021; Knorr, Augustin, & Tiwari, 2020; Tiwari, Norton, & Holden, 2013). The food security for current consumers and future generations relies on to understand how to maintain food stability and guidance for producers and processors.

Whichever the agro-resources, especially fruit and vegetables (F&V), the varietal diversity (Doerflinger et al., 2015; Le Bourvellec et al., 2015) coupled with the origin and location, the agronomical practices (levels of fruit thinning and irrigation etc.) (Mills, Behboudian, & Clothier, 1996; Saei et al., 2011), the maturity stage before and after harvest (Nyasordzi et al., 2013) and the specific storage conditions of each fruit induce a large variability. The latter can be considered at two levels: among different lots (inter-variability) and between individual fruits (intra-variability). It is even more complex due to the heterogeneity inside each fruit defined as the gradients from stem to calyx part (Qiao et al., 2019), from periphery to center area (Rahman et al., 2017), and from top to bottom (Zhang, at al., 2018). Both, variability and heterogeneity introduce highly diverse appearances (color, shape and size, etc.), physical (firmness, crunchiness etc.), structural (cell size and shape, cell wall content and composition, etc.) and biochemical (sugars, acids and nutrients, etc.) characteristics of raw material, that enter the processing systems. The current processing systems are not always adapted and optimized to the raw material, but meet the required microbiological safety standards, involving high energy-cost unit operations (high temperature, pressure, vacuum, long-periods, etc.) (Masanet, 2008). Besides, the increasing concerns of natural products without artificial food additives (food colors, sweeteners etc.) force the food system revolution (Bearth, Cousin, & Siegrist, 2014). Therefore, it would be highly beneficial to develop innovative fruit processing systems, which can consider a

large diversity of raw materials, apply efficient processing conditions, and provide new solutions to develop natural fruit products.

In the case of F&V, according to a FAO study (FAO, 2015; Rezaei & Liu, 2017), over 20% of losses at harvest, during sorting and grading in European and American countries are due to the quality standards set by retailers. About 14-21% of losses in developing regions occur during processing techniques for safe products (**Fig. 1**). Fruit producers and manufacturers encounter difficulties to decide what and how to produce the expected and constant quality level of final products. Up to now, fruit industrial manufacturers deal with the variability and heterogeneity before and after food processing mainly based on their experiences or some quantitative quality traits such as consistency, SSC, color, pH, etc. However, they often do not have the good choices to know how to fractionate and process the raw materials and then formulate the food by often adding additives (Bearth et al., 2014). However, this rough and intensive-labor strategy causes losses of energy and waste of food resources, which are not environment-friendly and not sustainable.

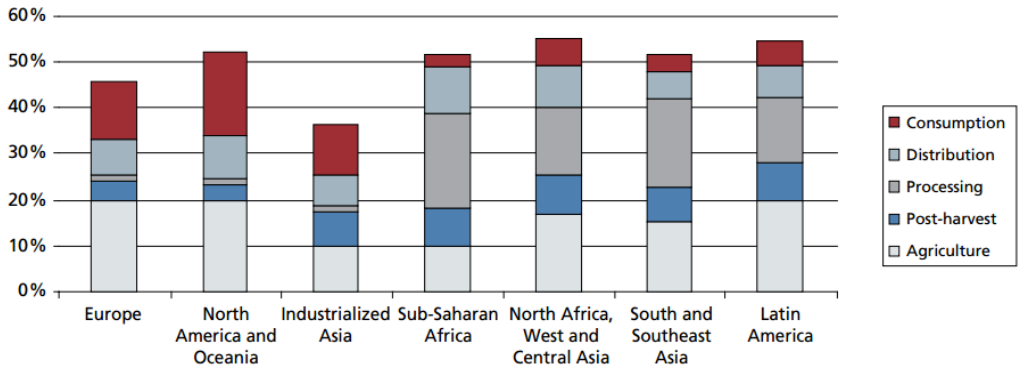

Fig. 1. Losses and wastes at different stages of the F&V supply chain in different countries.
(Figure adapted from FAO, 2015).

Apple accounts for the second largest market of fruit puree, after banana puree and followed by strawberry puree, with a global market value of about 2,000 million USD annually (Market Research Future, 2019). Apart from the specific apples dedicated to cider, nearly 20% of the French harvested apples are processed into puree (72%) and juice (17%) (FranceAgriMer, 2017). Apple puree is used as-is or as the basic ingredient

of jams, preserves or compotes (Defernez, Kemsley, & Wilson, 1995), which are popular among people of all ages, especially for babies and elders. Besides its economic importance and high consumption, apple puree is a processed product of primary interest, particularly suitable to introduce controlled variability coming from either raw material or processing conditions (Buergy, 2021b; Szczesniak & Kahn, 1971).

The quantification of quality traits is mandatory to understand and anticipate fruit selection and processing, and to do that, specific tools have been used over the past decades. They are intensively applied in fields or orchards, food processing industries and/or analysis laboratories (**Table 1**). Conventional chromatography such as high-performance liquid chromatography (HPLC) or gas chromatography (GC) coupled to mass spectrometry (MS) gives a precise determination of food quality. However, these equipments are almost always used at the laboratory scale, with limitations of cumbersome, time-consuming, laboriousness, complex sample preparation and analysis protocols using chemicals. Several handheld digital analyzers, such as digital refractometer to determine soluble solids content (SSC in °Brix), chromameter for color, pH-meter and penetrometer for texture, which are cost-effective, rapid, easy to perform both at field and industrial conditions. However, each of these techniques just gives specific information of quality, sometimes based on empirical calculations. Compared to them, visible (VIS), near (NIR) and mid (MIR) infrared techniques have several advantages of i) rapid and non-destructive (VIS and NIR) spectrum acquisition; ii) limited or no sample preparation, no chemical waste; iii) available for a wide range of samples (solids, powders, liquids, gels, pellets and pastes) and iv) suitable in field, industrial and laboratorial conditions. Moreover, several parameters are possibly evaluated from a single spectrum, with varying precision, but a calibration step is needed. Further, the development of new optical sensors and data treatment capacity, make possible to combine conventional imaging and spectroscopy to simultaneously explore spatial and spectral information, and then evaluate and predict the quality of fruit and processed products (Gowen, et al., 2007).

Till now, most of infrared applications have been dedicated to the detection of variability (variety, location, maturation, etc.) in raw or processed apples (**Table 13**, **Table 15**). However, there is limited knowledge on the use of such techniques to identify and analyze such variability on the same apples before and after processing. Indeed, a possible link between fresh and processed apples could allow the assessment of puree quality from a simple, rapid and non-destructive infrared scanning of raw apples.

Table 1. An overview of common and infrared spectroscopic techniques to detect food variability and heterogeneity at field, industrial and laboratorial scales.

Quality traits	Techniques	Available conditions	Advantages	Disadvantages
Size and shape (apples)	Electronic sorting and grading machine	• Industry	• Relatively cheap • Automatic	• Sorting based only on apple size
	Machine vision	• Industry	• More precise than electronic sorting • Integrated color and size information	• More expensive than electronic sorting • Mathematical data analyses
	Spectroscopic imaging (VIS-HSI, NIR-HSI, MRI, XRI)	• Lab	• High sensitivity • More precise and informative	• Expensive • Cumbersome (MRI, XRI)
Color (apples and purees)	Color charts	• Field/orchard • Industry • Lab	• Simple • Economic	• Subjective • Laborious
	Handheld colorimeters (CIE L*a*b*; CIE L*c*h)	• Field/orchard • Industry • Lab	• Rapid and simple • Non-destructive • Both liquid and solid maters	• Need standard calibration • Limited detection area
	Machine vision and VIS-spectroscopy and imaging	• Field/orchard • Industry • Lab	• Rapid and simple • Affordable costs • Non-destructive or destructive measure	• Mathematical data analyses • Impact of environment on spectral and imaging quality
Apple texture and firmness	Penetrometer	• Field/orchard • Industry • Lab	• Rapid and simple • Affordable costs	• Rough determination • Destructive • Laborious
	Texture analyzer	• Industry • Lab	• Multiple textural assessments (firmness, crunchiness, consistency etc.) • Precise detection	• Destructive • Cumbersome • High costs
	VIS and NIR Spectroscopy and imaging	• Field/orchard • Industry • Lab	• Rapid and simple • Non-destructive • Affordable costs	• Need modeling • Effect of environments and sample properties on spectral and imaging quality
	Magnetic resonance imaging (MRI)	• Lab	• Nondestructive • High contrast and resolution	• High costs • Cumbersome • Special skills for equipment • Use of radiations
	X-ray imaging	• Lab	• Nondestructive • Sensitive and precise • High contrast and resolution	• High costs • Cumbersome • Special skills for equipment • Use of radiations
Puree rheological properties	Bostwick consistometer	• Industry	• Rapid and simple • Low costs	• Rough detection • Limited precision and stability • Only for puree consistency
	Industrial viscometer	• Industry • Lab	• Potable and waterproof • For liquids	• Relatively precise detection • only for viscosity
	Rheometer	• Lab	• Multiple probes and cells • Informative rheological assessments (viscosity, viscoelasticity, consistency etc.) • Precise detection	• Complex operations • Long time testing • Cumbersome • High costs

Dry matter (apples and purees)	Ventilated thermal oven	• Industry • Lab	• Cheap • Simple	• Long-time • Need sample pre-treatment • Destructive
	Freeze dryer	• Lab	• Necessary to stabilize samples (polyphenols etc.)	• Expensive • Long-time and destructive • Need sample pre-treatment
	NIR spectroscopy (Hand-held)	• Field/orchard • Industry • Lab	• Non-destructive • Portable • Rapid and simple	• Relatively expensive, cheaper than freeze dryer • Need modeling
SSC (apples and purees)	Digital SSC refractometer	• Field/orchard • Industry • Lab	• Rapid and simple detections • Stable at different temperatures (5-70°C).	• Destructive • Very common and widespread
	VIS, NIR, and FT-IR	• Field/orchard • Industry • Lab	• Rapid and simple • Non-destructive (VIS, NIR) • On a wide range of sample statues • Continuous and numerous measurements	• Need modeling • High-cost of some FT-IR
Acidity (apples and purees)	Digital pH meter	• Field/orchard • Industry • Lab	• Rapid and simple detections • Low costs	• Destructive • Chemical standard calibration
	Automatic titrator	• Lab	• Automatic titration and calculation • Reducing labor work • High sensitivity	• Destructive • Cumbersome • Need sample preparation • Need chemical standard calibration
	NIR and FT-IR spectroscopy	• Field/orchard • Industry • Lab	• Rapid and simple • Non-destructive (NIR) or destructive • Sensitive detection by FT-IR	• Need modeling • High-cost of FT-IR
Biochemical composition (individual sugars and acids, polyphenols, polysaccharides etc.)	Enzymatic reactions coupled with spectrofluorometer	• Lab	• Accurate detection	• Chemical wastes • Complex sample preparation • Cumbersome
	Chromatographic techniques (HPLC and/or LC-MS coupled with MS)	• Lab	• High sensitivity • Stable detection	• Complex sample preparations • Chemical wastes • High costs • Cumbersome
	VIS and NIR spectroscopy	• Field/orchard • Industry • Lab	• Rapid and simple • Limited sample preparation • Continuous and numerous measurements	• Need modeling • Limited precisions • Only available for major components
	FT-IR and Raman spectroscopy	• Lab	• Rapid and simple • Sensitive detection	• Need modeling • Cumbersome • High costs

2. ‘Interfaces’ Project

This work was realized in an Agropolis Foundation Flagship project funded by ANR (the French National Research Agency) within the framework of the "Investissements d´avenir" program (ANR-10-LABX-001-01, Labex Agro, coordinated by Agropolis Fondation (https://www.agropolis-fondation.fr/The-foundation) (ID 1603-001) (**Fig. 2**). The project is titled “The interfaces between agricultural raw material and processing, a key point for bridging variability of raw materials and versatility of processing for innovative food systems” (“**Interfaces**”, https://www6.paca.inrae.fr/sqpov/Projets-Partenariats/Projets/Projets-nationaux-institutionnels/Interfaces-2017-2020), is coordinated by Dr. Véronique Broussolle (INRAE) and Dr. Dominique Pallet (CIRAD), and involves multiple French research units: UMR SQPOV, UMR QualiSud, UR PSH, and UMR MoiSA. The interfaces between production and processing domains are identified as key points to make food supply chain more durable and notably reduce losses and wastes by sorting raw materials for optimal use and adapting processes to variability of raw material. This project aims to better understand the variability and heterogeneity between fruit production and processing and their impact on microbiological and nutritional safety, organoleptic quality, and socio-economical sustainability. The questions of how fruits respond to processing, how they can be characterized with efficient tools, and how to manage them for sustainable and precise fruit processing are thus investigated.

Fig. 2. Sponsors, partners and research units involved in the “Interfaces” project.

This project is structured into 5 work packages (WPs) (**Fig. 3**), including the investigation of indicators and tools to better characterize the heterogeneity and variability of fresh and processed fruits (WP1), the interaction between raw material and processing (WP2), the fruit microbiota along the processing chain (WP3) and the relationship between different models of physiological and biochemical changes during fruit development and maturation in orchards, during postharvest storage and during processing (WP4). WP5 aims to better understand how knowledge of processability can impact relationships between the different actors of the food chain.

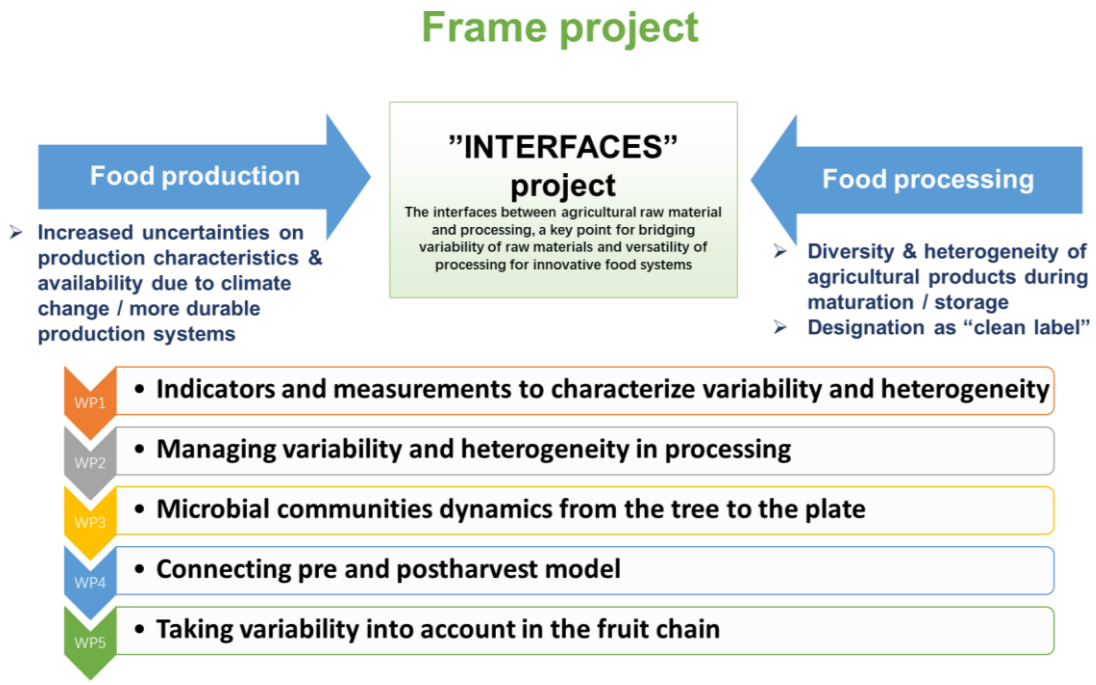

Fig. 3. The frame of "Interfaces' project.

The project is dedicated to three fruit – food systems: apple and apple puree, mango and mango dry slices, grape and grape wine. My PhD project belongs to the WP1 of the "Interfaces' project and performed in the UMR SQPOV. It is focused on apple (*Malus x domestica* Borkh.) and aims at investigating and determining relevant indicators and methods to evaluate the variability and heterogeneity of both, apples and processed purees. To reach this aim, the studied apples are issued from different varieties, agricultural practices, postharvest storage times and processing conditions.

3. Objectives of the thesis

The interfaces between production and processing domains are key points to make food supply chain more durable. In this context, the objective is to improve the quality of processed products and to reduce losses and wastes. The challenges are:

- How to face the variability and heterogeneity of agricultural raw materials, and how to manage and optimize their processing into food products to meet the consumers expectations or at least to reach constant and controlled product quality?

- How to make tools efficient enough and reliable to detect the variability and heterogeneity of raw materials, in order to predict the characteristics of processed products and to possibly adapt the processing conditions.

- And what strategy can facilitate and optimize the use of variable and heterogeneous raw materials to produce final processed products meeting consumers demands and habits?

This thesis presents a proof-of-concept of using infrared spectroscopy and chemometric -based methods to answer these challenges. In the future, the idea is to integrate them at key steps of the processing chain to better control the apple product quality.

4. Manuscript structure

After the '**Introduction**' above presenting briefly the context, I give further details in a **'literature review'** part, concerning:

i. The variability and heterogeneity of apples and purees due to production (variety, fruit thinning), postharvest storage (storage times) and processing conditions (temperature, grinding speed and refining levels) on the quality characteristics;

ii. Recent progress of infrared techniques to detect variability and heterogeneity of fresh and processed apples, and potential solutions to predict the properties of processed purees based on the spectral information of raw apples;

iii. Multivariate statistical and mathematical data analysis and advanced chemometric strategies to discriminate, link and explore spectra and reference data.

This review points out the current difficulties to determine apple heterogeneity and variability, identify the knowledge gaps to link them between apples and purees, and evaluate the use of infrared techniques to valorize and optimize processed products from the data of raw apples, leading to the '**Objective and strategy**' part.

The '**Results and discussion**' part is divided into three sections to answer our research questions:

i. How to identify the variability and heterogeneity of raw apples and processed purees using different spectroscopic and imaging techniques (**Papers I and II**) and taking into account the balance between data intensity and required sample preparation (**Paper III**);

ii. How is the variability of apples linked to the quality of processed purees and is it possible to predict the quality of processed purees using VIS-NIR, NIR and MIR spectral signals of raw apples before processing (**Papers IV, V and VI**);

iii. How to improve the puree formulation by infrared spectroscopy as an innovative solution to manage apple puree variability (**Paper VII**).

The final chapter is the '**Conclusions and Perspectives**' to synthetize the highlights of my works and present some promising research subjects that I have identified.

5. Publications and presentations

Five papers of my work were published:

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 (**Paper II**).

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 (**Paper III**).

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 (**Paper IV**).

Lan, W., M.G.C. Renard, 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 (**Paper VI**).

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 (**Paper VII**).

Two papers and two literature reviews are in preparation:

Lan, W., et al. NIR, MIR, Raman and Hyperspectral imaging techniques: Which is the best way to determine chemical, structural and rheological properties of apple purees? (Preparing for submission) (**Paper I**)

Lan, W., et al. Fruit variability impacts puree quality: assessment of individually processed apples using the visible and near infrared spectroscopy. (Preparing for submission) (**Paper V**).

Lan, W., et al. How to decipher the structural and chemical heterogeinity in fruits and vegetables: a review of advanced spectroscopic and imaging methods. (Preparing for submission)

Lan, W. et al. Contribution of near and mid infrared spectroscopy as a key point for detecting and bridging variability and heterogeneity of fresh and processed apples: a review. (Preparing for submission)

Three deliverables prepared for 'Interfaces' project:

Deliverable 1.1 – Report on 'Develop methods and identify relevant indicators to qualify the variability and heterogeneity of fresh apples and processed apple purees' (2018).

Deliverable 1.2 – Report on 'Recent advances of infrared spectroscopy for evaluating variability and heterogeneity of fresh apples and processed purees' (2019).

Deliverable 1.3 – 'Advanced infrared strategies to detect fruit variability and heterogeneity and bridge them between fresh and processed apples' (2020).

I have presented four oral conferences in national and international congresses:

Lan et al., Contribution of infrared spectroscopy to characterize the fresh and processed apples, International Plant Spectroscopy Conference (IPSC), International Plant Spectroscopy Conference (IPSC) in Berlin, Germany, March 24-28th, 2019.

Lan et al., Infrared spectroscopy: a potential tool to manage apple puree processing, 3rd Symposium on Fruit and Vegetable Processing conference, Avignon, France, November 24-25th, 2020.

Lan et al., Infrared-guided formulation: an innovative concept applied to apple puree, 21st HélioSPIR Meeting of Near Infrared Spectroscopy, Nantes, France, June 29th , 2021.

Lan et al., ATR-FTIR, an integrated tool to evaluate rheological, structural and biochemical variations during purees processing, Yong Rheologists Days, Giron, France, July 7-9th, 2021.

One abstract was submitted for a conference to the ICNIRS 2021:

Lan et al., Infrared guided processing: a potential strategy to predict processed purees properties from spectra of intact apples, 20th biennial meeting of the International Council for NIR Spectroscopy (ICNIRS), Beijing, China, October 18-21st, 2021. Submission.

An outcome named "les faits marquants" has been selected and recorded in 2021 for

the annual report of the department ‘TRANSFORM’ of INRAE.

II. Literature Review

1. Products and their source of variability and heterogeneity

1.1. Apple and puree

1.1.1. Apple definition

Cultivated apple (*Malus x domestica* Borkh.), a member of the Rosaceae family, is one of the most emblematic and widespread fruit crops in temperate region (Cornille et al., 2012; Juniper & Mabberley, 2006). Genetic studies revealed the evolutionary history of apple fruits (**Fig. 4**) firstly domesticated from *M. sieversii* in the Tian Shan Mountains for 4000–10,000 years, and dispersed from Central Asia to West Europe along the Silk Road, allowing hybridization and introgression of wild crabapples from Siberia (*M. baccata* (L.) Borkh.), Caucasus (*M. orientalis* Uglitz.), and Europe (*M. sylvestris* Mill.) (Cornille et al., 2012, 2014). After the interspecific hybridization, the scientific name *Malus* × *domestica* Borkh. of the modern cultivated apple is generally accepted and replaces the previous name of *Malus pumila* (Korban & Skirvin, 1984).

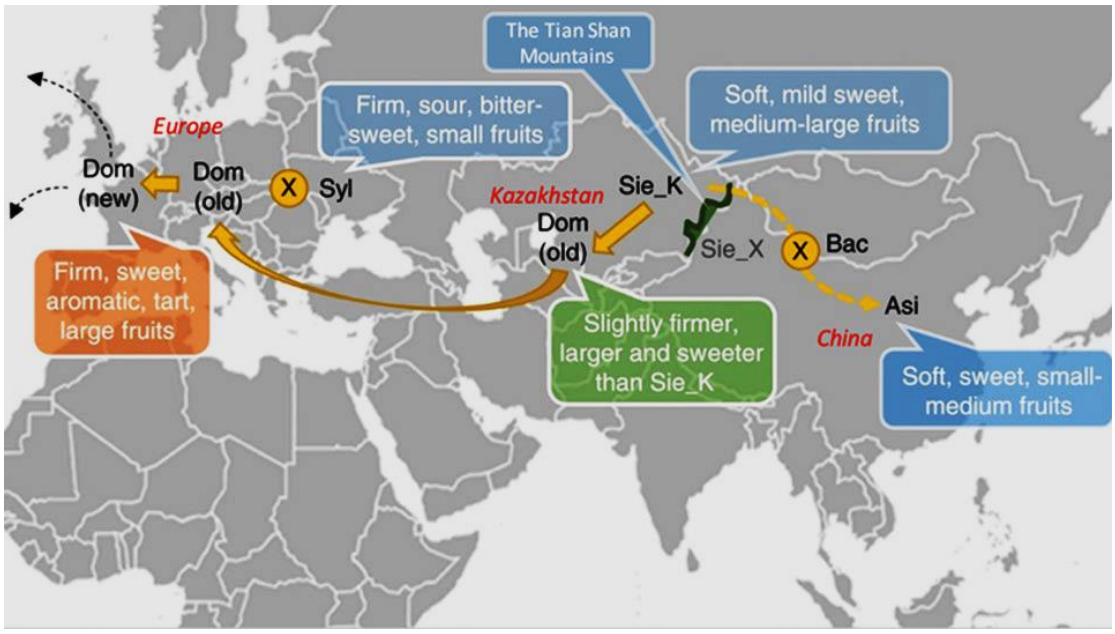

Fig. 4. Apple evolutionary map.
(Figure adapted from Duan et al., 2017)

1.1.2 Apple production and consumption

1.1.2.1 Apple production

Apple is one of the most widely cultivated fruits around the world, with a global production of 75.8 million tons in the 2019/2020 crop year (USDA, 2020), ranking it third in worldwide fruit production, following banana and watermelon. In the same year (**Fig. 5**), China was the world's largest producer of apples, with a production amount around 41 million metric tons. The European Union came in second place with about 11.48 million metric tons of apples (Statista, 2020). In Europe, France is the third apple producer with nearly 1.7 million metric tons in 2018, after Poland (around 4 million metric tons) and Italy (2.4 million metric tons) (FAO, 2018).

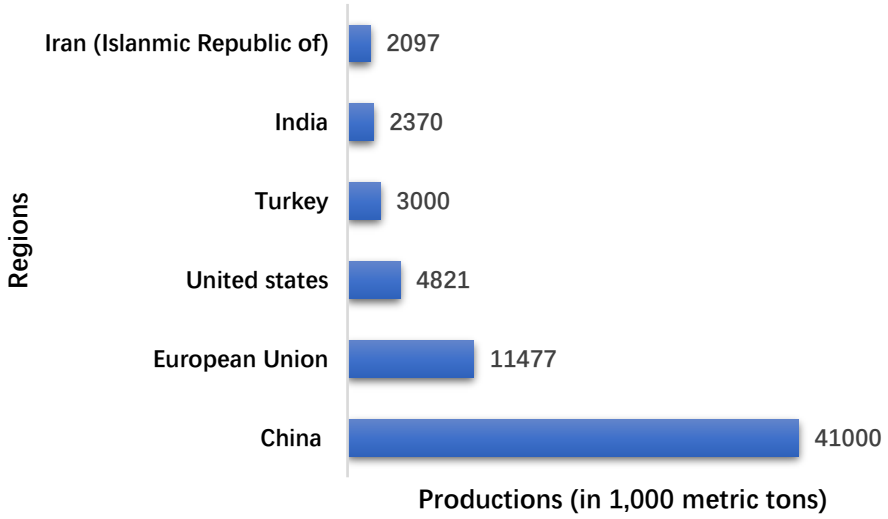

Fig. 5. Global leading countries of apple production in 2019-2020.

(Data adapted from Statista, 2020)

The success of apple fruit is undisputed, because today there are more than 20,000 varieties of apples of which 7,000 are regularly cultivated across the globe (Elzebroek, 2008). Particularly, there are mainly around 100 commercial apple cultivars, such as Fuji, Red Delicious, Golden Delicious, Gala, Granny Smith, Idared, Jonagold, Braeburn, Cripps Pink, Jonathan, Elstar and McIntosh (Bhushan, Kalia, Sharma, Singh, & Ahuja, 2008; Root & Barrett, 2005). In China, Fuji accounts for over 65% of the total apple production followed by Golden Delicious and Red Delicious (BEEDATA, 2020).

In France, Golden Delicious (29%), Gala (20%), Granny Smith (10%), Fuji (4%) and Braeburn (4%) are the most produced cultivars (ANPP, 2020). The leading regions in French apple production are Provence-Alpes-Côte d'Azur (25%), Midi-Pyrénées (16%) and Pays de la Loire (16%) where the climate conditions are especially favorable (FranceAgriMer, 2015).

Globally, around two thirds of total apples are marketed for fresh consumption, and the rest is mainly processed to juices, canned sauces, purees, dried and frozen products (Bhushan et al., 2008; Root et al., 2005). China is the largest consumer, absorbing over 48% of the global apple consumption in 2018 (BEEDATA, 2020). In China, around 90% of apples are dedicated for fresh consumption, whereas the rest is processed into apple products (BEEDATA, 2020). In France, apple is the most consumed fruit, with around 50% of the apple production designated for fresh consumption, whereas 19% are processed (ANPP, 2020).

1.1.2.1 Apple composition

Apple is one of the most interesting food in a healthy diet for its content in water (> 80%), sugars (fructose > glucose > sucrose), organic acids (0.2-0.8%, mainly malic acid), polyphenols, vitamins (mainly vitamin C), minerals (0.34% - 1.23%), dietary fibers (around 2% - 3%) and starch (Anses, 2020; Kiczorowska & Kiczorowski, 2005; Musacchi & Serra, 2018), varying according to cultivar, environment, ripening stage and post-harvest conditions (Downing, 2012) (**Table 2**). Besides, the soluble pectins and insoluble cellulose and hemicellulose in apple dietary fibers can give different flowing behavior, fermentation rate or binding potential of the final processed products (Gidley & Yakubov, 2019). Moreover, apple is one of the best resources of polyphenols for human health (Boyer & Liu, 2004a).

Table 2. The basic apple composition with skin

Compositions	**Amount** (per 100 g of fresh apple)	**Unit**	**Ranges**
water	85.6	g	82.4-87.5
energy	52	kcal	/
protein	0.26	g	0.17-0.57

total lipid (fat)	0.17	g	0.05-0.31
ash	0.19	g	0.07-0.48
carbohydrates	13.8	g	/
total fibers	2.4	g	1.4-3.5
sugars	10.4	g	8.77-12.0
glucose	2.43	g	1.45-3.66
sucrose	2.07	g	0.78-3.53
fructose	5.9	g	4.63-6.63
starch	0.05	g	/
Vitamin C (total ascorbic acid)	4.6	mg	4.0-5.5
Vitamin E	0.18	mg	0.08-0.38
Iron Fe	0.12	mg	0.09-0.18
Sodium, Na	1.0	mg	1.0-2.0
Potassium, K	107	mg	88-136
Calcium, Ca	6.0	mg	4.0-9.0

Notes: Recalculated based on (USDA, 2019) from 'Red delicious', 'Golden Delicious', 'Royal Gala', 'Granny Smith', and 'Fuji' apple varieties.

1.1.3 Apple quality determination

Comprehensive quality assessments of apples need to take into account evaluation of both, appearance and intrinsic physical and chemical characteristics (Pu, Feng, & Sun, 2015; Zhang et al, 2018). The external quality attributes of fruits, such as color, size, shape etc., are some of the most important sensory indexes. They significantly affect the marketing prices and the purchase behavior of consumers (Harker et al., 2003). Besides, the invisible and internal quality characteristics of fruits, such as firmness, crispness, soluble solids, total sugars, dry matter, organic acids, micronutrients etc., mainly impact their taste and nutritional values (Bondonno et al., 2017; Boyer et al., 2004b). The different techniques to determine apple quality at different scales, in industry and laboratory scales, are summarized (**Table 1**) and discussed in detail in the following parts:

1.1.3.1 Size

Apple size is a function of cell number, cell volume and cell density (Bain & Robertson, 1951; Coombe, 1976). Apple size is mainly related to: i) cell division occurring during the first period of fruit growth and ceases within three weeks after pollination; and ii) cell enlargement after division phrase until the final stages of fruit growth (Batjer & Westwood, 1958; Westwood, Batjer, & Billingsley, 1967).

Apple size plays a role in the consumer's purchase decisions. Large apples are preferred by consumers for fresh consumption (Iwanami, 2011; Kim, Lee, Kim, & Cho, 2009), but they cannot last as long in storage compared to smaller fruits (Bain et al., 1951; Kays, 1999). In industry, apple grading machines in sorting lines are commonly applied to classify apples in different size classes, which determine the marketing prices. Currently, automatic apple sorting systems using machine vision could be a potential solution to divide apples, rapidly and more accurately, based on their integrated information of color, size and weight properties (Sofu, Er, Kayacan, & Cetişli, 2016).

1.1.3.2 Color

A large diversity of apple colors exists, such as red, yellow, and green etc., but most are variegated or bicolor (red overcolor on yellow background). This should be an important parameter for variety and quality identification. Consumer color preference varies in the global market, according to countries and habits. Generally, bright red apples are usually catching the consumer's attention, whereas the dark red apples are less and less popular (Telias et al, 2011). However, the increase of red color, such as 'Red Delicious' which appears very early during fruit development, makes difficult to use this quality trait to determine the optimal date of harvest. The practicians need therefore to have new reliable quality indexes based on internal quality (Krishnaprakash et al., 1983). For some varieties, like 'Granny Smith', the red overcolor is a negative trait for the market and the blushed apples are downgraded (Hirst, Tustin, & Warrington, 1990). The different colors on apple fruit surface result from the combination of

pigments which are anthocyanins for red-violet, chlorophylls for green and carotenoids for yellow and orange (Lancaster, Grant, Lister, & Taylor, 1994).

The skin color is an important and fundamental parameter in apple industry to discriminate apple varieties. In sorting house, the grading machines sort apples based on the percentage of overcolor area on apple surface (Hamadziripi et al., 2014; Telias et al., 2011). Some classes are then defined, for example the 'Modi' with over 70% of red color, 'Demi Rouge ™' ranging from 40% to 70%, 'Pink Lady' with more than 40% of blush color, and 'Crispy Pink' with 10%-40% of blush color (Hirst et al., 1990).

Currently, there are mainly three kinds of methods to detect apple color:

- The color charts: a quite simple and economic method, but only provide the basic overview of color information based on the standard color references (Nickerson, 1957; Wilson, 1941). Besides, this method is very subjective and not suited for numerous determinations.
- The digital tri-stimulus colorimeter or Chroma meter (Konica Minolta Holdings, INC., Japan): it is the widely adopted instrument in fruit research and industry, because of easy to use, rapid, portable, and able to provide numerous and repetitive characterizations (J. Ahmed & Ramaswamy, 2006). It provides the color expression using the specific color space of CIE L*a*b*(1976), with three coordinates L* (lightness, black (0) and white 100)), a*(color between green (-60) and red (+60)), and b* (a color between blue (-60) and yellow (+60)). However, such equipment just evaluates the color in a limited 8 mm size area of the skin per measure (K. León, Mery, Pedreschi, & León, 2006).
- The digital imaging analysis: these techniques overcome the critical aspect of the limited measuring areas by colorimeters, and assess the color of whole apples while they pass through high-speed sorting grading machines (Zou & Zhao, 2015). Besides, both the visible spectroscopy and imaging techniques can be applied to evaluate the apple color variations in labs.

1.1.3.3 Texture and firmness

Apple textural properties refer to the firmness, crispness, mealiness and juiciness, which are usually evaluated with a combination of both, specific trained sensory panels and instrumental measurements (Brookfield et al., 2011). Fruit firmness is estimated by a puncture or compression test/measure of the force expressed in kilogram-force, pound-force, kg/cm^2 or Newton (N) (Lehman-Salada, 1996). This is linked to consumer's perception of apple texture (Harker et al., 2002). According to consumer's preferences, apple firmness below 45 N is too soft for 'Royal Gala' and 'Elstar', while 60 N is appropriate for 'Golden Delicious' and 'McIntosh' (Hoehn et al., 2003). Apple firmness varies according to variety, maturation stages, storage conditions, etc. This

parameter is usually used to determine the optimal harvest time (Harker, Maindonald, & Jackson, 1996; Kingston, 1992). During maturation, the advancing of the softening process and the decline of cell adhesion result in a decrease of apple firmness (DeEll, et al., 2001; Johnston, Hewett, & Hertog, 2002; Kingston, 1992).

There are several types of penetrometers to determine apple firmness. The most used are portable hand-penetrometers (Magness & Taylor, 1925) and electronic pressure testers (EPT) (Lehman-Salada, 1996). They can be used in orchards to quickly test apple firmness, and they are also available for scientific and industrial works. Besides, some advanced textural analyzers provide multiparameter and fine evaluation of apple texture, including crunchiness, mealiness, firmness, etc., but they are cumbersome and expensive (Liu, Cao, & Liu, 2019). However, these aforementioned methods usually suffer from drawbacks: i) destructive measurement on apple surface or peeled flesh for a metal probe penetration with a depth around 8 mm depending on the apple size (Johnston et al., 2002); ii) only available to estimate limited areas (mainly the outer region of apple flesh). However, the computerized penetrometer Mohr ® Digi-Test (Mohr and Associates, Richland, WA) can assess firmness throughout three regions in the apple from the outer to the middle region of the flesh (around 0.8–1.5 cm from the outer apple surface) (Mohr & Mohr, 2000); iii) the measured firmness is not always in agreement with the sensory analysis (Harker, et al., 2002).

More recently, the use of infrared spectroscopy and imaging techniques (mainly visible and near infrared) have been introduced to provide an integrated evaluation of apple texture, such as firmness (Park et al., 2003), roughness, crunchiness and mealiness (Mehinagic et al., 2003). They have the great advantage of non-destructive and rapid measurements which allow numerous fruit characterizations.

1.1.3.4 Dry matter content

Dry matter content (DMC) in apples presents the whole composition after removing water, mainly including sugars, starch, cell walls, organic acids and minerals. It is an important indicator to assess apple quality during both, development/ripening

and post-harvest storage (McGlone et al., 2003; Palmer et al., 2010). Apple are shown to be better accepted by consumers when DMC is higher (Palmer et al., 2010). Apple with a higher DMC (> 16%) at harvest might have a relatively slower starch degradation during post-harvest storage than those with lower DMC (< 13%) (Palmer et al., 2013). Besides, there is a positive correlation between DMC and firmness in apples at harvest and during post-harvest, independently of the fruit size (Palmer, 2014; Saei et al., 2011). The range of DMC is quite broad in apples and varies according to cultivar, rootstock, crop load, year, origin, storage conditions etc. (described in **Part 3**).

The conventional method to determine DMC in apples is time-consuming and destructive needing to grind apple in powder, weigh the samples before (for the fresh weight (FW)) and after (for the dry weight (DW)) drying in a ventilated oven at 60-65°C for 3 days until stable weight. Currently, DMC is also rapidly and non-destructively determined using NIR with for example the portable FELIX F-750 NIRS meter, with good predictions in apple (Kaur, Künnemeyer, & McGlone, 2017), sweet cherry (Escribano et al., 2017), mango (Anderson, Subedi, & Walsh, 2017) and kiwifruit (Shafie et al., 2015) etc., giving a possible integration of this measure into in-line sorting system.

1.1.3.5 Sugars

Determining the concentrations of total and individual sugars in the apple is crucial to understand the consumers' acceptance (Harker et al., 2002; Magwaza & Opara, 2015). Sugars in apples are mainly composed of fructose, glucose and sucrose. Their contents vary depending on apple cultivar, years, growing orchards and agricultural practices, ripening stages and storage conditions (Iwanami, 2011). They also increase during the post-harvest period, due to the starch hydrolysis (Visser et al., 1968). Generally, the individual sugars in apples are quantified by chemical methods, such as colorimetric assay, GC, HPLC etc., which are highly sensitive but limited to the laboratory work.

The most widely applied method used to measure total sugars in apples is the digital refractometer, which is rapid, simple and available at field, industry and laboratory

scales. It gives an overview of soluble solids content (SSC) in apples, presenting mainly the total soluble sugars (fructose, sucrose and glucose) (Bartolozzi et al., 1997; Kingston, 1992). SSC are expressed in degree Brix (°Brix), where 1 °Brix represents 1 g of sucrose in 100 g of aqueous solution at 17.5 °C (Bartolozzi et al., 1997; Kingston, 1992). The °Brix unit is applied in the case of fresh and processed apple products, such as purees, juices, jams etc. SSC has been shown to be the best indicator of apple sweetness perceived by sensory panelists, and is better than the results of chromatographic methods (Harker, Marsh, Young, Murray, Gunson, & Walker, 2002). In addition, consumers detect differences of apple sweetness with the SSC variations higher than 1°Brix (Harker, Marsh, et al., 2002).

In the last decades, the VIS-NIR and NIR spectroscopic techniques have become popular alternative methods to determine SSC in apples. Compared to digital refractometer, the major advantages are i) rapid and non-destructive; ii) simple and portable (F-750 FELEX USA, H-100F Sunforest Koera, etc.) with e.g. the apple SSC prediction from 5 to 25 °Brix with a standard error less than 0.5 °Brix (http://www.sunforest.kr/category_main.php?sm_idx=169), iii) possibility to be applied from field to on-line in industry.

1.1.3.6 Acidity

Apple acidity is a fundamental trait for determining cultivar and harvest dates, and provides information about flavor and taste due to a strong correlation between them (Hulme, Jones, & Wooltorton, 1963). The total organic acids contents in apples are mainly composed of malic acid with up to 90% of the whole, followed by citric, tartaric, lactic and oxalic acids (Kader, 2008). In Europe, the acid content in apples has been shown to strongly affect consumer preferences, with values ranging from 3.0 to 10.0 g/L (Bai et al., 2015; Iwanami, 2011). During apple ripening, the total acid content increases during the cell expansion phase, then decreases during ripening on trees and further decreases during post-harvest storage (Ali, Raza, Khan, & Hussain, 2004; Nybom, 1959).

Apple acidity is usually evaluated by measuring titratable acidity (TA) and pH. TA is measured by titration with a sodium hydroxide (NaOH) solution at a pH of 8.1. TA corresponds to the "potential acidity" meaning the total quantity of acids present both, in the acid form but also in the form dissociated and salified by mineral elements (citrate or malate of potassium, calcium or magnesium). It's generally expressed in g/L of the major malic acid or in %, being 1% TA equal to 10 g/L of malic acid (Musacchi & Serra, 2018). The hydrogen potential, noted pH, measures the chemical activity of hydronium ions in aqueous solution and represents the "real acidity", which in apple ranges is from 3.4 to 4.2 with an average around 3.7 (Eisele & Drake, 2005).

TA measurement is carried out using an automatic digital titrator (e.g. from the Mettler Toledo, Thermo Fisher companies etc.), and is designed to improve the efficiency of the measurement. There are currently pH meters that can be used in the field (mini pH meter) as well as in industry or in the laboratory. But these measurements are made on homogenized samples, and are therefore destructive.

In recent years, mid infrared spectroscopy has been reported as a sensitive tool for determining both TA, pH and malic acid in apples (Bureau, Ścibisz, Le Bourvellec, & Renard, 2012; Irudayaraj & Tewari, 2003) and apple juices (Kelly & Downey, 2005; Reid, et al., 2005). Current advances in mid-infrared sensors are geared towards micro-engineering and miniaturization, such as the Handheld MID IR Spectrometer (company ALLIED), but cost around 15,000 USD.

1.1.3.7 Cell wall

Cell walls and their constituent polysaccharides play an important role to forms a strong network, supporting the plasma membrane and preventing it from bursting under the turgor pressure contained osmotically in the cell (Cosgrove, 2005; Fricke, Jarvis, & Brett, 2000). Plant cell wall is primarily composed of cellulose (15–40% of cell wall dry weight), hemicelluloses (20–30%) and pectins (30– 50%) (Cosgrove & Jarvis, 2012; Fischer & Bennett, 1991; Jarvis, 2011; Renard & Thibault, 1993).

Traditional chemical methods are usually applied to extract cell wall materials and purify their polysaccharides. Renard (2005b) summarized and evaluated the existing

methods. The most common one is based on the cell wall extraction with ethanol and acetone to obtain the 'Alcohol Insoluble Solids' (AIS). This method is very simple but extracting AIS requires a large consumption of chemical solvents if starting from fresh samples (up to 1 L ethanol and 0.4 L acetone/ 1.0 - 1.5 g cell wall). Accelerated or pressurized solvent extractors (ASE, PSE) can allow multiplexing and thus faster and less solvent-consuming for cell wall preparation, but only from already freeze-dried (lyophilized) samples. Kurz et al., (2010) reported the potential of the NIR technique to assess the AIS contents of fruits.

After extraction of AIS, several specific cell wall polysaccharides, such as pectins, cellulose and hemicelluloses can be accurately determined by GC-MS or HPLC, but still suffer from the drawbacks of complex sample pre-treatments and several chemical wastes. MIR technique has been applied on the AIS samples from a large panel of F&V to rapidly evaluate their cell wall polysaccharides (Canteri et al., 2019).

1.1.3.8 Polyphenols

Apples are known to be a good source of polyphenols, which are very beneficial for health (Sun, Chu, Wu, & Liu, 2002). The main class of polyphenols in apples are flavan-3-ols, which include the monomeric (+)-catechin and (-)-epicatechin as well as their oligo-and polymers, the procyanidins (Vrhovsek, Rigo, Tonon, & Mattivi, 2004). Procyanidins account for more than 80% of the total polyphenol content (Le Bourvellec, Bouzerzour, Ginies, Regis, Plé, & Renard, 2011; Oszmiański, Wolniak, Wojdyło, & Wawer, 2008). Most apple polyphenols are concentrated on the fruit surface but are also present in the flesh (Le Bourvellec et al., 2011; Oszmiański et al., 2008). In red apples, in addition anthocyanins (1%), the glycosides of anthocyanidins are found.

Generally, freeze-dried apple samples were used to determine polyphenols by HPLC-DAD after thioacidolysis (Le Bourvellec et al., 2011). It can provide the individual polyphenols including procyanidins and monomeric flavanols, phenolic acids, dihydrochalcones and flavonols etc., but needs a long time sample preparation and HPLC analysis. Some previous studies reported the use of NIR and MIR techniques to determine the total and individual polyphenols in apples (Giovanelli et al., 2014; Pissard et al., 2013; Pissard et al., 2018; Bureau et al, 2012).

1.1.4 Puree definition

Fruit purees are defined as 'the fermentable but unfermented product obtained by suitable physical processes such as sieving, grinding, milling the edible part of whole or peeled fruit without removing the juice', based on the European Legislation (Directive 2012/127EU). Although plant-based purees with a large variety of fruits are available, all can be described as suspensions of soft particles in a viscous serum or gel (Colin-Henrion, Cuvelier, & Renard, 2007). Generally, the industrial production of apple purees consists typically in cooking at 93 - 98℃ for about 4 - 5 min, refining to remove seeds and skin pieces and then pasteurizing at 90℃ for around 20 min to obtain a shelf life of 6 months at room temperature (Oszmiański et al., 2008). Several additional ingredients can be added to the purees during processing, such as ascorbic acid, spices, sugars, honey or water etc. Sweetened apple puree should contain not less than 16.5° Brix and unsweetened apple puree not less than 9.0° Brix.

1.1.5 Puree production and consumption

Apple puree accounts for the second largest market of fruit puree, after banana puree and followed by strawberry puree, with a global market value of about 2,000 million USD annually (Future, 2019). Apart from cider production, nearly 20% of the French harvested apples are processed into purees (72%) and juices (17%) (FrenceAgriMer, 2017). Apple purees can be used as the basic ingredient of jams., preserves or compotes (Defernez et al., 1995). The products of apple purees are popular from all ages of people, giving apple purees (46%) and mixed purees with other fruits (54%) are steadily increasing over the last years (FranceAgriMer, 2017) (**Fig. 6**).

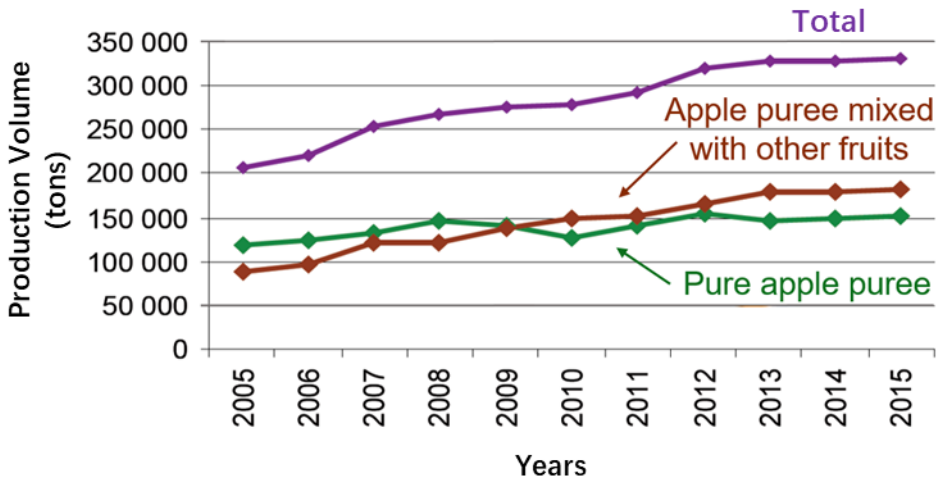

Fig. 6. French industrial production of apple purees during 2005-2015.

(Figure adapted from FranceAgriMer, 2017)

1.1.6 Puree composition, rheology and structure

1.1.6.1 Puree composition

The basic nutrients of an apple puree product are displayed in **Table 3**. Compared to the apple nutrients in **Table 2**, there is a clear loss of calories, protein, dietary fibers, carbohydrates, and total sugars after processing, due to heating and refining procedures. However, a higher concentration of Vitamin C was detected in the processed purees dataset, most likely due to its use as additive in the grinding step. Until now, there was no complete reference data regarding the biochemical compositions of both raw apple materials and corresponding processed purees.

Table 3. The basic nutrients of a commercial apple puree product.

Nutrient (per 100 g)	Amount
Calories	56.4 Kcal
water	85.1 g
Protein	1.13 g
Carbohydrates	11.7 g
Dietary fiber	1.7 g
Sugars	11.7 g
Total fat	< 0.3 g
Cholesterol	0.00 mg
fructose	7.24

galactose	< 0.1
glucose	2.79
sucrose	1.69
Beta-carotene	19,9 ug
Vitamin C	9.86 mg
Calcium	4.3 mg
Chloride	< 20 mg
Iron	0.10 mg
Sodium	< 5 mg

Notes: data adapted from USDA, https://ciqual.anses.fr/#/aliments/13187/puree-de-pommes-type-%22compote-sans-sucres-ajoutes%22.

1.1.2.1 Puree rheology

Rheological tests provide a meaningful insight on the structural organization of apple purees, and explain their sensory perception (Jasim Ahmed & Ramaswamy, 2007). Apple puree is a viscoelastic fluid, which can present both, viscous properties as liquid matter and elastic properties of solid matter (Buergy et al., 2021b). It is similar to most of plant-based purees, and behaves as a non-Newtonian fluid, more precisely shear thinning fluid, presenting a yield stress (Rao, Thomas, & Javalgi, 1992). Plant-based purees are usually described by a power law model and also by different models that include the yield stress as a fitting parameter, such as the Herschel–Bulkley and Casson models (Colin-Henrion, Cuvelier, & Renard, 2007; Rao, Thomas, & Javalgi, 1992). According to previous works, the rheological properties of apple purees are mainly attributed to the soluble solids in the serum phase, the insoluble solids (cell wall) and the particle features of insoluble solids (Espinosa-Muñoz et al. 2013). Generally, apple purees show dominant elastic properties with higher storage modulus (G') than loss modulus (G"), because of the elastic network of weak attractive or repulsive forces between cell wall particles (Espinosa-Muñoz et al., 2011). However, strong homogenization and intensive refining of apple purees can result in smaller and less structured particles in puree suspensions, consequently leading to dominant viscous behaviors (G''> G') (Kunzek, Opel, & Senge, 1997).

1.1.2.2 Puree structure

Structurally, apple puree can be divided into pulp and serum (Rao, Thomas, &

Javalgi, 1992) (**Fig. 7**). Puree pulp is a complex matrix with clusters of cells, individual cells or cell fragments of the apple parenchyma. The variation in particle size in pulp ranges from hundreds of µm to mm, and significantly affects the rheological properties and sensory perception of apple puree (Espinosa-Muñoz et al., 2013; Espinosa-Muñoz et al., 2011; Leverrier et al., 2016). Puree serum contains mainly water (85%), soluble sugars (12%), and some minor chemical compounds (polyphenols, pectins etc.), while pulp contains the same elements plus insoluble dietary fiber. (Ebermann & Elmadfa, 2008).

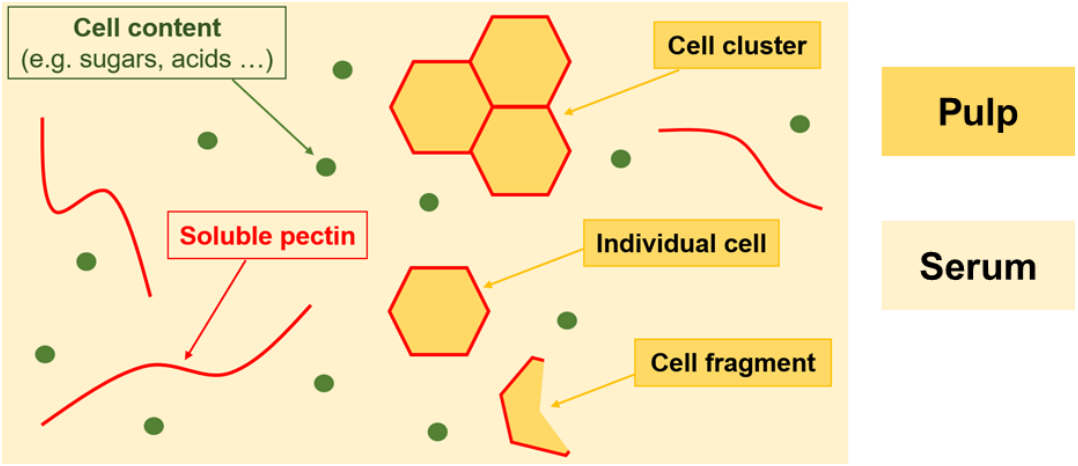

Fig. 7. A schematic representation of the composition of plant-based purees. (Figure adapted from Buergy, 2021).

1.1.7 Puree quality determination

The rheological and textural properties of apple purees, such as consistency, viscosity, viscoelasticity, particle size and volume etc. are determined using several specific equipments:

- Bostwick consistometer, the simple and cheap method to determine the consistency of apple purees (expressed as ‘Bostwick units’) is widely applied in the fruit puree industry. The puree consistency is evaluated by the flowing distance in cm forced by gravity along the equipment over a 30 s interval. However, the measurement is highly subjective and just provides a limited overview of puree yield stress and viscosity (Cullen, Duffy, & O'Donnell, 2001).

- Rheometer is a specific instrument to provide a precise determination of puree flow behavior and deformation properties by measuring the stress-strain relationship. It gives information of puree viscosity and viscoelastic properties corresponding to the moment when puree starts to flow (yield point), etc., at different temperatures or shear stresses/rates. The explanations of these rheological parameters have been described by Rao (2010), and the specific operations on apple purees by Buergy (2021b).

- Laser diffraction granulometry uses a laser beam which passes through the dispersed particles of purees coupled with the Mie or Fraunhofer theory of light scattering (Mastersizer, 2007), in order to accurately determine the particle size distributions of apple purees. Results are expressed as the diameter of every theoretical sphere of irregular particles and the distribution of puree particles.

Besides the puree particle size changes, cell wall contents (alcohol insoluble contents, AIS) have been linked to the rheological properties of apple purees, such as apparent viscosity, yield stress and elastic modulus etc. (Espinosa-Muñoz, Renard, Symoneaux, Biau, & Cuvelier, 2013). Extracting the cell wall requires times and a large consumption of chemical solvents to remove all soluble components (mainly sugars and acids), up to 1 L ethanol and 0.4 L acetone to prepare 1.0 - 1.5 g of dry cell wall from fresh matrices. Accelerated or pressurized solvent extractors (ASE, PSE) can allow multiplexing and thus a faster and less solvent-consuming cell wall preparation, but only from already freeze-dried matrices.

Therefore, in order to have an accurate and complete assessment of puree texture, these aforementioned methods must be adapted with complex sample preparation, time-consuming and expensive determination. A recent research highlighted the potential of determining the consistency of tomato juice using portable mid-infrared spectroscopy (Ayvaz et al., 2016). However, there was no other report on the determination of rheological and textural properties such as specific viscosity, viscoelastic properties and particle size variations.

Other common methods, especially chromatographic techniques (HPLC-MS, GC-MS, etc.) and colorimetric assay are also applied to characterize the biochemical composition (sugars, acids, polyphenols, etc.) in apples and purees (**Table 1**). Exploring the possibility of using infrared spectroscopy has been reported on mango (Labaky et al., 2021), tomato (Bureau, Page, Bogé, & Renard, 2015; Szuvandzsiev et al., 2014) and raspberry (Andrianjaka-Camps et al. 2015). However, there was no report on apple puree quality.

1.2 Heterogeneity

1.2.1 Definition of heterogeneity

The word ‘heterogeneity’ refers to any kind of variation of appearance, composition and/or properties in an individual fruit. Based on that, three spatial scales (macroscale, mesoscale and microscale) of fruit heterogeneity need to be considered:

- At the macroscopic scale, studies focus on the whole fruit. At this scale, individual fruit is considered as a whole, a continuum constituted of different connecting tissues (Ho et al., 2006; Mebatsion et al., 2008). At this scale, spatial studies of heterogeneity are mainly dedicated to the different parts of individual fruits, with gradients from stem to calyx part (Qiao et al., 2019) (**Fig. 8a**), along equatorial direction (Zhang et al., 2018) (**Fig. 8b**), and from periphery to center area (**Fig. 8c**) (Rahman et al., 2017) etc. At the macroscale, heterogeneity can be non-destructively viewed in an individual fruit (surface color, shape and size etc.) or analyzed using conventional and destructive methods on different parts of samples (Menesatti et al., 2009; Mo et al., 2017; Peiris et al., 1999; Pissard et al., 2012).

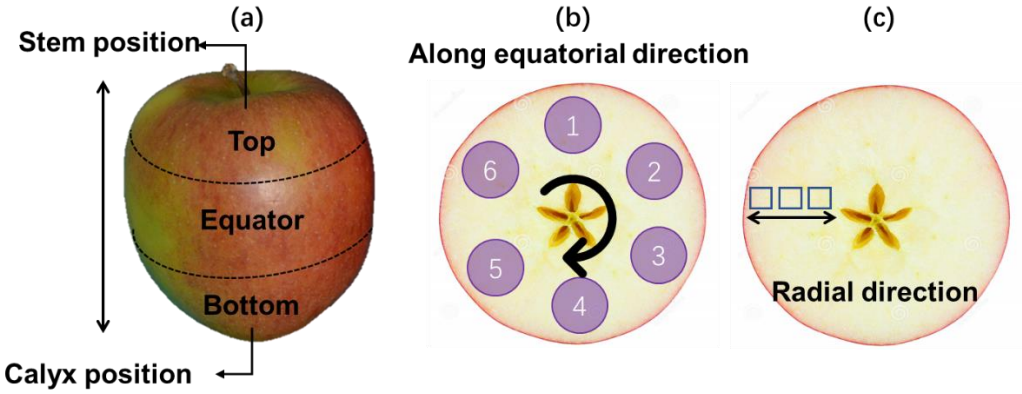

Fig. 8. The apple heterogeneity at macroscale.

- At the mesoscopic scale, the actual topology of the individual fruit tissue is considered, by visualizing the arrangement of the intercellular spaces, cell walls and individual cells as building blocks (Bondonno et al., 2017). At this scale, fruit tissues have the possible to show considerable heterogeneity between the different parts (**Fig. 9**), such as endocarp, exocarp and mesocarp tissues etc., which are tortuously or randomly connected with several scales in play (Mendoza et al., 2007; Sen, 2004). The biochemical and structural properties have been increasingly investigated in different fruit tissues in the past few years (Aregawi et al., 2013; Bassan et al., 2013; Qi & Shih, 2014; Stewart, 1996; Türker-Kaya & Huck, 2017). The characterization of heterogeneity of fruit tissues is destructive and usually uses both, the optical imaging and/or conventional chromatographical and/or mass spectrometric technologies after sample preparation.

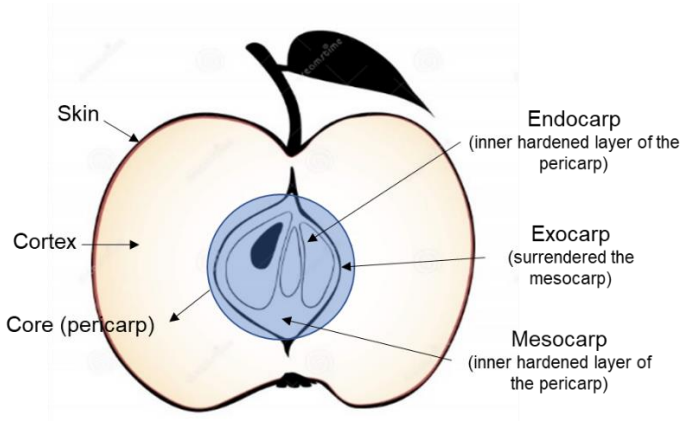

Fig. 9. The different apple tissues subdivided into endocarp, mesocarp, exocarp, cortex and skin regions.

(Figure adapted from Franceschi et al., 2012)

- At the microscopic scale, single cells are distinguished or separated from fruit tissues. The heterogeneity of cell walls and cell membranes can thus be addressed (**Fig. 10**). Specific functional and structural variations of cell walls (Barron et al, 2005; McCann et al., 1997; Szymańska-Chargot et al., 2016; Xiao et al., 2020) and cell membranes and their components were studied (Bassan et al., 2013; Chylińska, Szymańska-Chargot, & Zdunek, 2014; Guendel, Rolletschek, Wagner, Muszynska, & Borisjuk, 2018a; Murata et al., 2000; Pan, Pu, & Sun, 2017). Several advanced phenotypic analyses using imaging techniques allow to study the heterogeneity of single cells, even of single live cells (Evers et al., 2019).

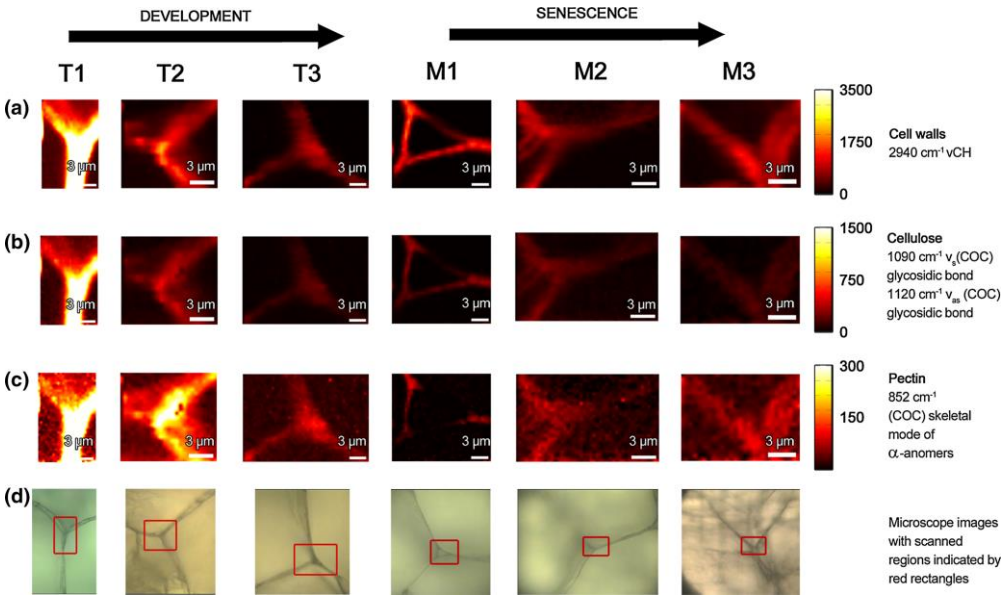

Fig. 10. Raman maps of the cell wall of apple parenchyma at different development stages T1, T2 and T3 and during 3 months storage M1, M2 and M3.

(Figure adapted from Szymańska-Chargot et al., 2016).

1.2.2 Apple heterogeneity

In the past few years, high heterogeneity has been identified for physical, biochemical and structural properties in individual apples at the macro, meso and micro levels (**Table 4**).

At the macroscopic scale, high heterogeneity of SSC (Fan et al., 2016; Ma et al.,

2018; Mo et al., 2017; Peiris et al., 1999), starch content (Menesatti et al., 2009; Peirs et al., 2003a), total polyphenols and vitamin C (Pissard, et al., 2013) was demonstrated based on the destructive chemical characterizations in the different pieces of apples. Briefly, some interesting results were concluded:

i) The heterogeneity of SSC in apples has been proven to truly exit from proximal to distal direction (**Fig. 8a**), along equatorial direction (**Fig. 8b**) and in radial direction from inside to outside (**Fig. 8c**).

ii) The heterogeneity of SSC, polyphenols and vitamin C varies with different trends among apple varieties (Pissard et al., 2013). For example, SSC gradually increases from the top to the bottom in 'Fuji' (Mo et al., 2017) and 'Red Delicious' (Peiris et al., 1999) apples, whereas it decreases in 'Hidala' apples (Pissard et al., 2012).

iii) The level of heterogeneity depends on the quality parameters with for example a significant variation of total polyphenols from the proximal to the distal direction, but not for vitamin C in the same apples (Pissard et al., 2013).

At the mesoscopic scale, the heterogeneity concerns several biochemical components (sucrose, sorbitol, malic acid, metal irons, polyphenols, etc.) and structural features (pore size and space, intercellular spaces etc.) in different tissues (**Table 4**). These studies provide specific insights to:

i) A significant heterogeneity is observed for malic acid, sucrose and flavanols in apple endocarp tissue (Zhang, Cha, & Yeung, 2007) and different levels for pore intensity in apple cortex tissue (Janssen et al., 2020a; Nugraha et al., 2019).

ii) The heterogeneity in apple of metal irons between cuticle and inner tissues (Vidot et al., 2020), or flavanols and hydroxycinnamic acids between cuticle and outer cortex tissues (Vidot et al., 2019).

At the microscopic scale, the heterogeneity concerns apple cell structure (cell size, cell volume, cell networks etc.), cell wall composition, water hydrogen bonding status (**Table 4**). For example, the pectin accumulation changes in apple during ripening between cell wall corners and along the cell wall (Szymańska-Chargot et al., 2016) (**Fig. 10**). The conventional chemical analyses do not work at this scale. Some imaging techniques with high sensitivity and resolution are good candidates, such as confocal Raman imaging, X-ray micro-CT imaging and MRI etc.

Table 4. Overview of the reported physical, structural and biochemical apple heterogeneity at the macro-, meso- or micro-scales.

Scales	Parameters	Methods	Variety	Conclusions	Ref.
Macro	SSC	CA, NIRS, VISNIR-HSI	Fuji	Higher SSC (1°Brix) in the calyx part than the stem part	(Fan et al., 2016; Mo et al., 2017)
	SSC	CA, NIRS, NIR-HSI	Red Delicious	Increase of SSC along the proximal-distal axis, and from inner towards outer surface	(Mo et al., 2017; Peiris et al., 1999)
	SSC	CA, NIR-HSI	Fuji	Higher SSC near peels than in the center	(T. Ma et al., 2018)
	SSC	CA, NIRS	Hidala, Pilot	Higher SSC near the stem part than the calyx one in 'Hidala'; homogeneous SSC in 'Pilot'	(Pissard et al., 2013)
	Starch	CA, NIR-HSI	Jonagold, Boskoop	Different levels in apples	(Peirs et al., 2003a)
	Starch	CA, NIR-HSI	Golden Delicious	Higher concentration in the outer cortex than in the core	(Menesatti et al., 2009)
	Total polyphenols	CA, NIRS	Hidala and Pilot	Significant higher polyphenol contents in the calyx part than in the stem and equator ones	(Pissard et al., 2013)
	Vitamin C	CA, NIRS	Hidala and Pilot	No significant difference in the vertical direction.	(Pissard et al., 2013)
Meso	Sucrose, sorbitol	MALID-TOF and MSI	Fuji	Sorbitol accumulation in the center; higher sucrose content in the cortex side than in the center	(Horikawa et al., 2019)
	Flavanols, dihydrochalcones	MALID-TOF and MSI	Golden Delicious	Different distribution of flavanols in endocarp, mesocarp and exocarp in apples	(Franceschi et al., 2012)
	Organic acids and flavonoids	MALID-TOF and MSI	/	Heterogeneity of malic acid, quinic acid, sucrose and flavonoids contents in apple endocarp region	(Zhang et al., 2007)
	Metal iron	S-cryo-XRF	Douce Moen, Guillevic	Variation of K, Ca, Fe, Mn from the cuticle to the inner tissue	(Vidot et al., 2020)
	Polyphenol compositions	M-DUV-cryo-imaging	Gala, Douce Moen, Guillevic	Differences in Flavanols, hydroxycinnamic acids between cuticle and outer cortex region	(Vidot et al., 2019)
	Pore size	XRI	Jonagold, Greenstar, Kanzi	Roughly spherical pores near the skin and elongated spores close to the core	(Mendoza et al., 2010)
	Intercellular spaces	XRI	Braeburn, Fuji,	Large differences of intercellular spaces in apple tissues	(Ting et al., 2013)

			Golden Delicious, Jazz		
	Pore structure and density	XRI	Braeburn	Both, low and high porous regions in the cortex tissue	(Janssen et al., 2020a; Nugraha et al., 2019)
	Pore space	XRI	Jonagold and Braeburn	Complex and variable pore space networks	(Mendoza et al., 2007)
	Moisture and void networks	XRI	Jonagold	Heterogeneous void networks in skin, cortex and vascular tissues and moisture around individual cells and in cell walls	(Verboven et al., 2008)
Micro	Cell size, volume	XRI	Kanzi	Large heterogeneity of cell size and volume in different flesh and mesocarp tissues	(Wang et al., 2017)
	Cell networks	XRI	Jonagold	Significant differences of cell networks in tissues	(Herremans et al., 2015)
	Cell wall polysaccharides	Raman imaging	Golden Delicious	Heterogeneity of pectin distribution in cell walls	(Szymańska-Chargot et al., 2016)
	Water hydrogen bonding status	Raman imaging	Fuji	The strongest hydrogen bonds in cell wall and the weakest in the intercellular regions	(Li et al., 2020)

1.2.3 Factors impacting apple heterogeneity

Some previous works have addressed the apple heterogeneity caused by various factors such as genetic (variety), microclimate with the sunlight exposure, ripening, storage and processing conditions, inducing macro and micro changes in morphology, physiology and biological components (**Table 5**). However, there is much less information regarding several important factors, such as the effect of:

i) Agricultural practice, such as fruit thinning and/or water stress on apple heterogeneity at macro to micro scales.

ii) Processing conditions (heating, grinding, refining etc.) on the different tissues of an individual apple fruit.

Table 5. An overview of the possible factors affecting apple heterogeneity

Factors	Impacts	References
Variety	Heterogeneity of flavanols and phenolic acids according to different apple tissues	(Vidot et al., 2019)
	Heterogeneity of SSC, polyphenols and Vitamin C in different apple varieties	(Pissard et al., 2012)
Microclimate	Higher redness and more anthocyanin in apple peel on the sunlight exposed side	(Matsuoka, 2019; Proctor, 1974)
Ripening	Different levels of starch concentrations in apples	(Peirs et al., 2003a)
	Change of pectin distribution in apple cell walls	(Szymańska-Chargot et al., 2016)
Storage	large effect on firmness of apple tissues	(Aregawi et al., 2013)
	Dispersion of pectin in cell walls	(Szymańska-Chargot et al., 2016)
	No impact on phenolic distribution	(Vidot et al., 2019)
Drying process	Migration of water between intracellular and intercellular spaces	(Khan et al., 2018)

1.2.4 Challenges to characterize apple heterogeneity

The heterogeneity met in each apple depending on the different parts, tissues and individual cells is difficult to assess and study due to the balance between the limitations of characterizing techniques and the complexity of sample preparation.

At the macroscopic scale, conventional chromatographic or mass spectrometric technologies face difficulties to have a holistic understanding of apple heterogeneity, because of i) long-time and intensive labor operations, ii) a large amount of targeted

components, iii) impossibility to use one method for analyzing multiple quality attributes.

At the mesoscopic and microscopic scales, several advanced imaging techniques have been successfully applied to study the heterogeneity of apple tissues and cell walls. However, most of them have drawbacks such as arduousness and high-cost, and some difficulties to provide comprehensive insights of the factors involved during fruit production or processing. Moreover, complex sample pre-treatments and initial calibration steps have to be performed to prevent the oxidization of the highly hydrated samples and to increase the signal-to-noise ratio (SNR) of the acquired images (the detailed discussion in **II-Part 6**).

These current problems prevent highly effective investigations of apple heterogeneity, which arise from several combined factors: genetic, agricultural practices, storage and processing. Therefore, developing a relatively simple, rapid, stable analyzing method for more numerous and synergistic studies of apple heterogeneity would be a great contribution in future projects.

1.3 Variability

1.3.1 Definition of apple variability

Apple variability is observable by differences in both, appearance and internal physical and chemical characteristics, which can be approached at two different levels:

- The ‘inter-variability’ between apple varieties: intensive researches have been performed to estimate the complex impact of genetics.
- The ‘intra-variability’ within a variety, between apple batches or individual apples: the quality of apple lots is affected by natural environment conditions (seasons and geographical locations), agricultural practices (production systems, harvesting dates) and post-harvest fruit management conditions (storage durations, temperature, humidity, atmosphere etc.). In addition, individual apples are also affected by micro-environmental conditions such as position in tree during growth and ripening changing for example the sunlight exposure and subsequently temperature. At this scale, each apple would need to be targeted and studied, which is difficult or even impossible due to a large number of conditions and a very small volume of apples for each.

The ‘inter- and intra-variability’ impact physical and biochemical properties of apples.

1.3.2 Factors impacting of apple variability

The possible factors impacting apple variability before harvesting can be grouped into three major sub-sources which are genetic materials (varieties, rootstocks), environmental conditions (light, temperature, humidity, soil, wind etc.) and agronomic practices (orchard design, row orientation, pruning, crop load/fruit thinning, pollination, irrigation etc.) (**Fig. 11**) (Musacchi et al., 2018).

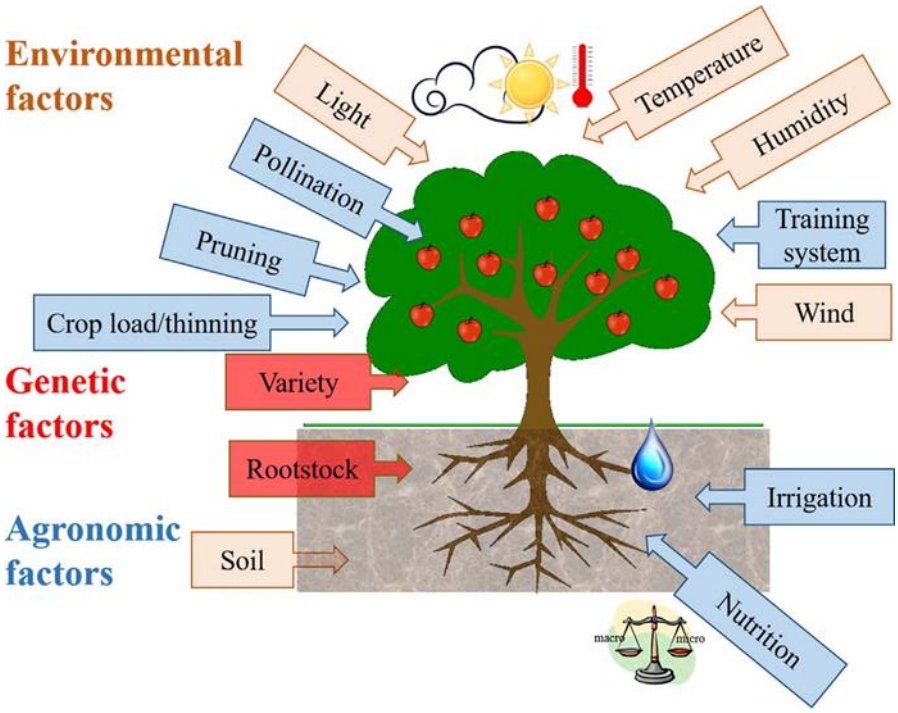

Fig. 11. Genetic, environmental, and agronomic sources affecting apple quality during pre-harvest.

(Figure adapted from Musacchi et al., 2018)

The genetic diversity mainly includes different apple varieties and rootstocks. Obviously, apple varieties give a large variation of external traits (fruit size, color etc.) (Kouassi et al., 2009), sensory traits (firmness, crispness, juiciness, flavor) (Cevik et al., 2010; King et al., 2000; Kouassi et al., 2009), and biochemical parameters (sugar, acidity, polyphenols etc.) (Hoehn et al., 2003; Volz & McGhie, 2011). Rootstocks control wood production in the tree, directing its energy into fruit production. Rootstock mechanisms have been reported to influence apple fruit size, firmness, color, SSC, carbohydrates and mineral concentration, ripening and respiration rates at harvest and during storage (Barritt et al., 1997; Kviklys et al., 2012; Musacchi et al., 2018; Tomala et al., 2008).

The complex environmental conditions affect apple growth and quality, especially light availability and temperature (Corelli-Grappadelli et al., 2004). Particularly, low temperatures during the first 40 days after full bloom can significantly reduce the apple fruit size (Stanley et al., 2000). The apple cell expansion rates at 20 °C can be 10 times

higher than at 6°C (Warrington et al., 1999). Apple grown at temperature between 12 and 16 °C have relatively higher SSC than in warmer conditions (Stanley et al., 2000). Light intensity is strongly correlated to apple quality, especially concerning the synthesis of secondary metabolites such as anthocyanins and carotenoids, as well as the degradation of chlorophyll (Matsuoka, 2019). Besides, a higher light exposition in apple planting systems provides higher fruit DMC (Wünsche et al., 1996).

In terms of agronomic practices, different orchard designs (location, altitude, planning density etc.) strongly affect apple quality due to a complex interaction of specific temperature, humidity, and light (Musacchi et al., 2018). Pollination is a key event to maintain or improve apple quality (Latimer, 1931, 1933, 1937). Agricultural practice such as apple thinning to remove excess fruits from trees, alters the crop load but improves fruit quality at harvest (Dennis, 2000; Yuan, 2007). Similarly, irrigation impacts apple quality and for example, deficit irrigation is a technique of applying less water to the tree than the evapo-transpiration demand at selected times during fruit growth (Musacchi et al., 2018). The deficit irrigated apples are generally smaller, but with higher SSC and lower acidity than standard fruits (Ebel, Proebsting, & Patterson, 1993), althrough this is not always an effective solution to improve fruit quality (Centofanti, Bañuelos, & Ayars, 2019).

1.3.2.1 Variety

There are more than ten thousand apple varieties listed in the European Apple Inventory (Watkins, 1985). This large number gives a wide range of inter-variability in fruit quality traits (Way et al., 1990). In America, just 16 varieties account for 90 % of the domestic apple production and particularly Golden Delicious, Granny Smith, Jonathan, McIntosh, Red Delicious, Rome Beauty, Stayman, and York make up 80% of the continent production (Wellness, 2021). In France, Golden Delicious, Granny Smith, Reine des Reinettes (Queen of the Pippin), Pink Lady and Royal Gala are the most favorite apple varieties (Statista, 2021). Apart from their strong appearance differences, a large diversity of sensory properties (texture and taste) and eating habits guide the choice for each apple varieties (in **Table 6**). The varieties of Golden Delicious, Granny

Smith, Gala, Fuji, Honeycrisp, Braeburn and Jonagold are widely applied for fresh consumption and/or sauce processing.

Table 6. The sensory properties and consuming advantages of the most popular apple varieties.

Varieties	Sensory properties	Consuming advantages
Golden Delicious	sweet, mild	snacking, salad, baking, beverage, pie, sauce
Granny Smith	tart, super crunchy	snacking, salad, baking, beverage, pie, sauce
Pink Lady	sweet-tart, crunchy	snacking, pie, sauce
Gala	very sweet	snacking, salad, baking, beverage, pie, sauce
Red Delicious	mildly sweet, crunchy	snacking, salad
Fuji	super sweet, crunchy	snacking, salad, baking, beverage, pie, sauce
Honeycrisp	distinctly sweet, crunchy	snacking, salad
Braeburn	sweet-tart, crunchy	snacking, salad, baking, sauce, pie
Jonagold	sweet-tart, crunchy	snacking, salad, baking, beverage, pie, sauce

(Data adapted from Washington State Apple Association https://bestapples.com/varieties-information/varieties/)

Two popular apple cultivars, ‘Golden Delicious’ and ‘Granny Smith’ have a relatively homogeneous color distribution of their skin, mainly yellow and green respectively. However, several previous studies demonstrated a large inter-variability of their biochemical and structural properties (T**able 7**). The cellular architecture, cell walls stiffness and crystallinity index of the cellulose fibers were significantly different between Golden Delicious and Granny Smith, explaining the inter-variability of their firmness (Rojas-Candelas et al., 2021). There were also significant differences of biochemical compositions, such as sugars, vitamin C, total phenolic contents etc., and antioxidant ability between these two varieties (Asale et al., 2021).

T**able 7**. A brief comparison of physicochemical, structural and biochemical properties of Golden Delicious and Granny Smith.

Properties	Golden Delicious	Granny Smith
firmness (N)	20.24 ± 1.23	24.61 ± 1.79
cellular density (cells/mm^2)	21.48 ± 0.34	13.5± 3.10
number of pores pore/mm^2)	7.30 ± 0.63	4.12 ± 0.38
Young's Modulus (MPa)	1.06 ± 1.20	1.76 ± 1.03
fiber diameters of cellulose (nm)	27.85 ± 6.90	31.85± 10.03
total sugars (g/100 g)	17.32 ± 0.40	15.54 ± 0.27
sucrose (g/100 g)	6.76 ± 0.16	6.14 ± 0.10
Vitamin C (mg/100g)	31.48 ± 2.18	14.97 ± 1.28
total phenolic contents (mg/g)	66.94 ± 1.62	71.88 ± 2.30
total flavonoid content (mg/g)	15.59 ± 0.23	21.78 ± 1.87
DPPH scavenging (ug/mL)	122.53 ± 3.48	93.25 ± 2.88

Note: Data adapted from Asale et al., 2021; Rojas-Candelas et al., 2021.

Apart from 'Golden Delicious' and 'Granny Smith', many varieties are bi-colored, with variable ranges of intensity and distribution of red skin color, especially for Gala, Braeburn and Fuji etc. (Trong et al. 2014; Iglesias, Echeverría, & Lopez, 2012; Iglesias, Echeverría, & Soria, 2008; Iglesias, Graell, Echeverría, & Vendrell, 1999). The intra-variability of red pigmentation is controlled by the relative concentration of anthocyanins in different apples of each variety (Matsuoka, 2019). A large diversity of red of apples skin might be also associated with a large variability of textural and structural properties (Defraeye et al., 2013; Janssen et al., 2020b; Van Beers et al., 2015).

1.3.2.2 Agricultural practice

Apple grows in two phases: an early exponential cell division phase that occurs between 1 and 4 weeks after full bloom, followed by a cell expansion phase during the rest part of the season (Bollard, 1970). Fruit thinning (hand thinning, mechanical thinning, chemical thinning etc.) at the apple early growing stages has been proven as an efficient method to significantly increase the cell numbers during the cell division phrase (**Fig. 12**) (Bergh, 1990; Milić et al., 2017; Wismer, Proctor, & Elfving, 1995).

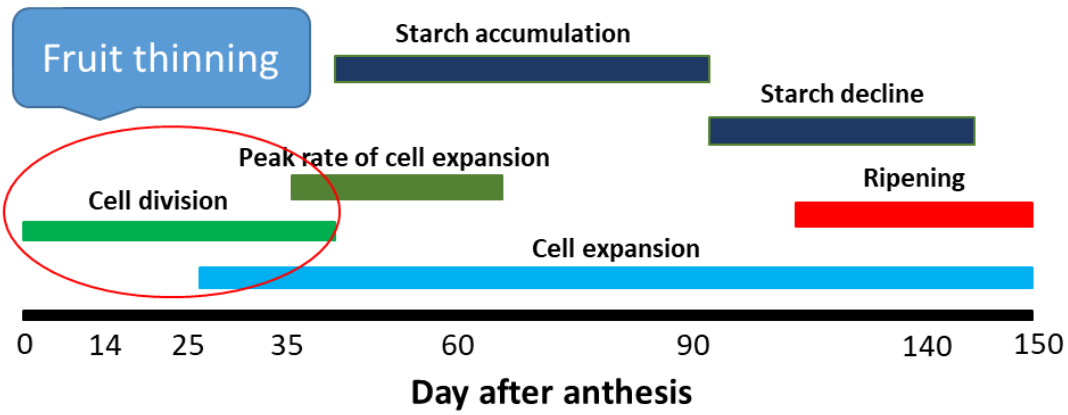

Fig. 12. Impact of fruit thinning at early stages during apple growth and maturation.

The size of apple fruit is mainly due to the variation in the number of cells and in a lesser extent to the average cell size. (Bain & Robertson, 1951). The averaged fruit size of Golden Delicious and Granny Smith is increased after fruit thinning (Bergh, 1992). Besides, adjusting the crop load by thinning practice can increase the SSC, DMC and TA and possibly decrease the firmness and Calcium in apples at harvest (**Table 8**) and during postharvest storage (Fadanelli et al., 2004; Saei et al., 2011). Although studies dealt with the effect of fruit thinning directly on the quality of crops, there is no information of its impact on i) the processed apple products, and ii) the heterogeneity of individual fruits (see the part of apple heterogeneity). Besides, there is a limited knowledge of using infrared techniques to investigate the possibility of evaluating the effect of thinning practices on harvested apples and even on their processed purees (see the part of infrared techniques).

Table 8. Effects of fruit thinning on apple physical, structural and biochemical properties.

Features	Fruit thinning / Non-fruit thinning	References
cell number (cells/fruit)	Boskoop: 8.6% ↗ Cox: 23.3% ↗ Golden Delicious: 3.0% ↗	(Link, 2000)
Cell size	Boskoop: 7.2% ↗ Cox: 4.4% ↗ Golden Delicious: 9.9% ↗	(Link, 2000)
Fruit size	Gala: 16% ↗ Golden Delicious:14% ↗ Braeburn:13% ↗	(Bergh, 1990, 1992; Mpelasoka et al., 2001; Solomakhin et al., 2010)
Firmness	Gala: 9.2-11.8% ↗ Golden Delicious: 2.7-12.5% ↘ Braeburn: 1.8% ↗	(Mpelasoka et al., 2001; Saei et al., 2011; Solomakhin et al., 2010)
Weight	Gala: 20% ↗ Elstar: 31.7% ↗ Braeburn: 8% ↗ Golden Delicious: 36.7% ↗	(Link, 2000; Meland, 2009; Mpelasoka et al., 2001; Solomakhin et al., 2010)
TA	Gala: 17.6% ↗ Golden Delicious: 3% ↗ Scifresh: 14.8% ↗	(Hehnen et al., 2012; Meland, 2009; Solomakhin et al., 2010);
SSC	Gala: 6.9% ↗ Golden Delicious: 5%↗ Elstar: 2.4% ↗ Braeburn: 2.5% ↗	(Meland, 2009; Mpelasoka et al., 2001; Solomakhin et al., 2010)
DMC	Gala: 15.7-29.5 ↗	(Saei et al., 2011)
Starch	Gala: 2.4-11.1% ↘ Golden Delicious: 25-36.6% ↗	(Solomakhin et al., 2010)
Polyphenols	Jonagold: 19% -178% ↗	(Stopar, et al., 2002)
Ca	Boskoop: 7.6% ↘ Cox: 31.2% ↘ Golden Delicious: 28.7% ↘	(Link, 2000)
K	Boskoop: 27.7% ↗ Cox: 25.7 ↗ Golden Delicious: 15.3 % ↗	(Link, 2000)

1.3.2.3 *Storage conditions*

The main postharvest factors that influence apple physical, structural and biochemical properties are temperature, storage time, relative humidity and storage atmosphere (O_2/CO_2 rates, ethylene concentration etc.) (DeEll et al., 2001). The most conventional storage temperature for apples has been reported to be between 0 °C and 3 °C, but it varies according to the variety (Johnston et al., 2002a; 2002b). The temperature strongly influences the postharvest life of apple fruit (**Table 9**). However, apples are often under non-optimal temperatures during grading, packing, distribution, loading and unloading, and at retail outlets during display (Musacchi et al., 2018).

Moreover, as apple is a living organism, long storage leads to loss of water, firmness, carbohydrates, organic acids etc. (Thompson, 2008). Intensive research have been dedicated to limit the loss of apple quality by adjusting and optimizing the storage environments, such as atmosphere and humidity (Mditshwa, Fawole, & Opara, 2018).

There are only few research studies on the effect of storage conditions applied on apples on the quality of their processed products. There was no report concerning how much variability caused by storage impacts both, apples and their corresponding purees, or if the processing reduces or enhances the variability of raw apples.

Table 9. Overview of factors affecting apple physical, structural, and biochemical properties during storage.

Factors	Properties	Conclusions	Ref.
Temperature	Firmness	Firmness decreases with rising temperature, depending on cultivars	(DeEll et al., 2001; Johnston et al., 2001, 2002a; 2002b)
	Mealiness	Higher temperature increasing the speed of apple mealiness	(Huang & Lu, 2010; Varela et al., 2005)
	SSC	Higher temperature increasing SSC in apples at the initial storage stages	(Prasanna et al., 2000)
	TA	Lower titratable acidity with higher storage temperature	(Kweon et al., 2013)
	polyphenols	Higher temperature leading to a decrease of total polyphenols and proanthocyanins	(Queiroz et al., 2011)
storage time	Firmness	Decrease of apple firmness during long storage	(Billy et al., 2008; Belie et al., 2000; Tu et al., 2000)
	Mealiness	Storage time promoting mealiness, depending on cultivars	(Billy et al., 2008; Huang et al., 2010; Varela et al., 2005)
	SSC	SSC increasing during short storage, but decreasing after long storage	(Jha et al., 2012; Tu et al., 2000; Veberic et al., 2010)
	DMC	DMC decreasing during apple storage	(Perring, 1989)
	Acidity	Acidity decreasing during storage	(Jha et al., 2012; Tu et al., 2000; Veberic et al., 2010)
	polyphenols	Limited effect on flavonoid or antioxidant activity during long cold storage	(Sluis et al., 2001; Veberic et al., 2010)
humidity	Firmness	Apple firmness decreasing more slowly at higher humidity	(Paull, 1999; Tu et al., 2000)
	Mealiness	Higher humidity promoting development of mealy texture	(Tu et al., 2000)
	Weight	Higher humidity reducing apple weight loss	(Lee, et al., 2019; Tu et al., 2000)
	SSC	No significant effect on SSC	(Tu et al., 2000)
	DMC	No significant effect on DMC	(Tu et al., 2000)
	Acidity	Limited effect on acidity, titratable acidity decreasing slower at low humidity	(Prange, et al., 2001; Tu et al., 2000)
controlled atmosphere	Firmness	Highly acceptable firmness than in normal atmosphere	(Konopacka & Plocharski, 2004; Siddiqui et al., 1996; Wang et al., 2020)
	Weight	Significant decay of weight loss during storage	(Wang et al., 2020)
	SSC	No significant effect on SSC	(Wang et al., 2020)
	Acidity	No significant effect on acidity	(Wang et al., 2020)
Ethylene	Firmness	Ethylene accelerating softening depending on cultivars	(Johnston et al., 2002)

1.3.3 Factors impacting puree variability

According to the previous reports, there are several resources contributing to a large diversity of chemical, textural and rheological properties of apple purees (**Table 10**), which can be mainly addressed to: 1) apple material, 2) thermal processing, and 3) mechanical treatment. The specific impacts of these four resources are described in the following parts.

1.3.3.1 Raw apple materials

A large diversity of raw apples, including different cultivars, harvesting (thinning practices, organic or conventional etc.) and postharvest storage conditions, have been reported to introduce strong chemical and textural variation on the processed purees.

Oszmianski et al. (2008) found that Idared and Shampion apples generated a significantly different concentration of total polyphenols and procyanidin B2 in their cooked purees. Calligaris et al. (2006) compared and explored the processing suitability of six Italian apple cultivars and finally obtained sensorial characteristics equal to the Golden Delicious purees. Besides, the texture of apple purees can be affected by cultivars, because of their different pulp contents, chemical composition and particle size (Le Bourvellec et al., 2011; Schijvens, Van Vliet, & Van Dijk, 1998). Buergy et al. (2021a) demonstrated the viscosity of apple puree strongly varied among Braeburn, Gala, Golden Delicious, Granny Smith apples from 562 to 1368 mPa s.

Even for the same apple cultivar, different agricultural practice during apple cultivation can result in a large diversity of puree quality. Rembialkowska et al. (2007) demonstrated that purees cooked with organic apples contained higher levels of flavonoids, polyphenols and vitamin C than conventional ones. Besides, the thinned Golden Delicious apples, with large particles, resulted in significantly higher puree viscosity than the non-thinned apples (Buergy et al., 2020).

Moreover, puree prepared from apples at unripe, ripe and overripe maturing stages had clearly different viscosity and yield stress (Schijvens et al., 1998). The post-harvest stored apples (storage periods at 4°C) can generate a wide range of different puree

textures, because of the decrease in particle size during prolonged apple post-harvest storage (Buergy et al., 2020).

1.3.3.2 Thermal processing

Thermal processing is a common step in apple puree preparation. The first heat treatment or 'break' plays a major role of inactivating polyphenol oxidases and pectolytic enzymes such as polygalacturonase [PG] or pectin methylesterase [PME], while softening the tissue. The conventional indirect heating treatments can be adapted with two strategies:

i. "Hot-break" products are sufficiently heated to inactivate enzymes with a heating temperature over 75°C;

ii. or 2) "cold-break" is performed at a temperature under 66 °C, allowing pectolytic enzyme action and leading to lower consistency and texture stability (Colin-Henrion et al., 2007).

Generally, thermal processing always results in the degradation of organic acids, tissue softening and textural loss on apples, whereas the sugars, cellulose and hemicellulose are stable and less influenced (Buergy, 2021a; Opatová et al., 1992).

Besides the indirect thermal process, apple puree can be produced by several techniques, such as microwave, ohmic heating and high-pressure processing. Microwave processing technique has the advantages of heating solids or liquids rapidly and uniformly, thus inactivating the enzymes more quickly and minimizing phenolic oxidation during fruit processing (Guo, Sun, Cheng, & Han, 2017). The microwave processed apple purees had higher phenolic compounds (like chlorogenic acid, polymeric procyanidin, phloretin-2'-glucoside and quercetin glycosides) than typical industrial processes (Oszmiański et al., 2008). Moreover, the viscosity of microwave heated apple purees remained stable during storage (Picouet et al., 2009). Microwave technique has been reported as a mini-processing strategy for an apple to purees (Picouet et al., 2009). Ohmic heating is a rapid and uniform thermal processing

technique based on the passage of an alternating current through a sample which responds by generating heat internally due to its inherent resistance (Kaur, Gul, & Singh, 2016). Rinaldi et al. (2021) demonstrated the ohmic cooked apple purees presented a lower reduction of ascorbic acid than conventional thermal processing. High-pressure processing is reported to be very effective in retaining quality traits in fruit products such as purees. The high-pressure processed apple puree showed the higher concentrations of ascorbic acid, but a significant decrease of viscosity due to β-eliminative degradation and the very fast acid hydrolysis of water-soluble pectins facilitated by the low pH (Diaz, Anthon, & Barrett, 2007; Krall & McFeeters, 1998; Rinaldi et al., 2021).

1.3.3.3 Mechanical treatments

Mechanical treatment plays a major role to disrupt the apple tissue structure into smaller cell fragments, mainly with grinding and refining.

Particularly, grinding treatment only modifies puree particle size and shape (an increase of grinding speed causes a decrease of particle size and apparent viscosity), but not the chemical compositions and cell wall contents (Espinosa-Muñoz et al., 2012; Espinosa et al., 2011).

The different levels of puree refining are known to be a determinant of particle size and of cell wall content (Colin-Henrion et al., 2007). After puree refining, a large part of cell wall in apple peels can be removed, thus resulting in a decrease of dry matter content (Colin-Henrion et al., 2007). However, there are few reports providing specific insight into chemical, structural and rheological properties of purees at different levels of refining (**Table 10**).

Table 10. Previous reports of the possible factors affecting the physical, biochemical, and textural properties of apple purees

Variety	Variable factors	Quality traits	Ref.
Unknown	storage temperature, periods	sugars and acids, sensory properties	(Opatová et al., 1992)
French market products	32 commercial purees	graininess, viscosity, moisture and grittiness	(Tarea et al., 2007)
Idared, Shampion	microwave and ascorbic acid addition	color, phenolics compositions and antioxidant activity	(Oszmiański et al., 2008)
Golden Delicious	grinding speed, time	rheological, structural and sensory properties	(Espinosa et al., 2011)
Idared and Fuji	storage time	phenolics compositions and antioxidant activity	(Loncaric et al., 2014)
Bramley	high hydrostatic pressure, temperature	polyphenolic stability and physico-chemical properties	(Keenan et al., 2011)
Idared	admixture to apricot jams	phloretin 2'-glucoside as marker compound by HPLC	(Dragovic-Uzelac et al., 2005)
Golden Delicious	the stages of processing	patulin content	(Janotová et al., 2011)
Granny Smith	microwave heating	vitamin C, viscosity, color, polyphenols, titratable acid	(Picouet et al., 2009)
Golden Delicious	particle content and size, serum viscosity	rheological properties	(Espinosa-Muñoz et al., 2013)
Unknown	purees and concentrates	polyphenol compounds	(Bengoechea et al., 1997)
French market products	purees and juices	total polyphenols and vitamin C	(Georgé et al., 2005)
Lobo, Boskoop, Cortland	organic and conventional apples, pasteurize	flavonoids, polyphenols, vitamin C	(Rembiałkowska et al.,2007)
Golden Delicious	pulp content, particle size, serum viscosity	texture and sensory perception	(Espinosa-Muñoz et al., 2012)
Braeburn, Gala, Golden Delicious, Granny Smith	mealiness and firmness of raw apples	puree particles and viscosity	(Buergy et al., 2021)
Golden Delicious	heating technology	vitamin C, color, furfural contents	(Pelacci et al., 2021)
Fuji	vacuum and oxygen-free	phenolic compounds, ascorbic acid, antioxidant activities, color, and enzyme activities	(Kim et al., 2021)
Golden Delicious	thermal, ohmic heating, high-pressure processing	color, viscosity, particle size distribution, total phenol content, ascorbic acid content and sensorial quality	(Rinaldi et al., 2021)
Golden Delicious, Granny Smith	fruit load, post-harvest storage	puree particles and viscosity	(Buergy et al.,2020)
17 varieties	apple flesh and purees	polysaccharides and polyphenols	(Le Bourvellec et al., 2011)
Idared, Rome	blanch temperature/time	rheological properties	(Godfrey Usiak et al., 1995)
R.I. Greening, Rome	apple firmness, finisher speed and size	particle size distributions	(Nogueira, et al., 1985)
Golden Delicious, Boskoop, Cox's Orange	apple ripeness, finisher size, cooking time	rheological properties	(Schijvens et al., 1998)

1.3.4 Challenges to manage processed food from raw materials

So far, some previous researches have investigated the possible relationship between raw and cooked apples by taking into account specific quality traits such as the polyphenolic compound and the degree of browning during apple juice processing (Song et al., 2007), the total antioxidant capacity during cider production (Khanizadeh et al., 2008), the texture of apples and purees (Buergy, Rolland-Sabaté, Leca, & Renard, 2021), and the cell wall polysaccharides before and after apple sauce processing (Le Bourvellec et al., 2011) etc.

These observed correlations of pre- and post-cooking physical, textural and biochemical properties could open the possibility of managing the quality of processed foods based on their highly correlated information from raw materials. Previous studies have reported the potential of using the NIR or MIR spectra of harvested apple and mango fruits to forecast their SSC, TA, DMC etc. during post-harvest storage, using the strong chemical internal relationships (Ignat et al., 2014; Nordey, Davrieux, & Léchaudel, 2019) (**Table 11**). For processed food, this strategy can be used to assess the texture of cooked meat and rice from the NIR spectra of raw materials (Meullenet, Jonville, Grezes, & Owens, 2004; Windham et al., 1997). However, this strategy failed at predicting butter quality from NIR spectra of milk (Lefébure et al., 2021). To our knowledge, there is no report on the use of spectroscopic information on raw fruit to predict their processed fruit products.

Table 11. Applications of infrared spectroscopy to forecast some quality traits of fruit and food.

Applications	Conclusions	Ref.
Apple ripening	Optimal harvesting time from initial apple ripening stages by MIRS	(Hazarika, Hebb, & Rizvi, 2018)
Apple storage	Apple internal composition during storage based on VIS-NIR spectra at harvest.	(Ignat et al., 2014)
Mango storage	Mango shelf-life and quality traits based on NIR spectra at harvest	(Nordey et al., 2019)
Form milk to butter	Butter characteristics from NIR spectra of milk	(Lefébure et al., 2021)

From raw to cooked meat	Texture of meat after cooking	(Meullenet et al., 2004)
From raw to cooked grain	Texture of cooked rice from raw grain by NIRS	(Windham et al., 1997)

2. Spectroscopic and imaging techniques

2.1. Spectroscopic techniques

For many years, several vibrational spectroscopy techniques have been considered as rapid, non-destructive, and inexpensive tools to estimate molecular structure and organic matter in food products (Abbas, Pissard, & Baeten, 2020; Nicolai et al., 2007). The development of instrumentation makes them applicable both in laboratory research (Dupuy, Duponchel, Huvenne, Sombret, & Legrand, 1996; Dupuy, Galtier, Ollivier, Vanloot, & Artaud, 2010; Galtier et al., 2011) and online or at-line food industrial production and process (De Beer et al., 2009; Picouet, Gou, Hyypiö, & Castellari, 2018; Porep, Kammerer, & Carle, 2015). In this section, we introduce the basic knowledge of visible (VIS), infrared (NIR and MIR) and Raman spectroscopic techniques, as well as their recent applications on the detection of variability of apple and their processed products.

The infrared spectral regions, with the NIR immediately following the visible (VIS) spectra, and then moving to the longer wavelength region of MIRS (**Fig. 13**). When a sample matrix is illuminated by the infrared light, molecular vibrations are measured by absorbances at each wavelength associated with different vibrations of specific chemical bonds. The absorption frequencies are usually presented as wavelengths (λ) with the unit of nm. The relationship between the attenuation of light through a substance and the properties of that substance can be described by the Beer-Lambert Law (Swinehart, 1962).

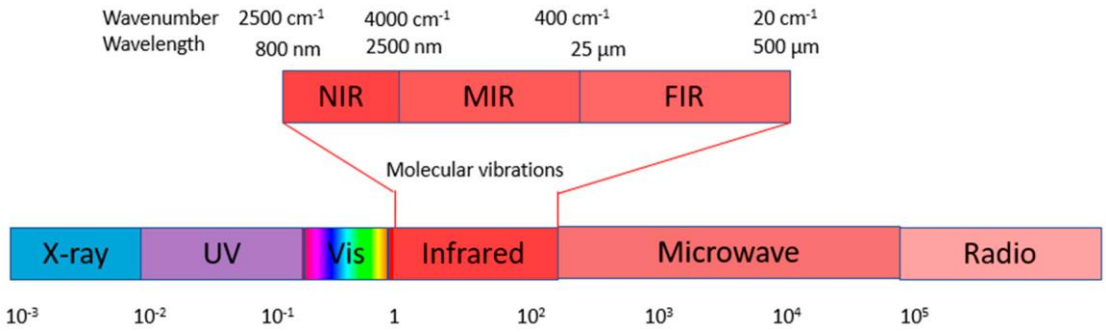

Fig. 13. Spectral range for near infrared (NIR) and mid-infrared (MIR) showing as wavelengths (nm) and wavenumbers (cm^{-1}).

(Figure adapted from Fox, 2020)

The interaction between infrared source and characterized samples can be analyzed through the transmittance, diffuse reflectance and transflectance modes:

- The transmission mode is mainly adapted to the transparent materials with constant sample thickness, placed in a quartz cuvette or a flow cell. In these conditions, the absorption depends solely on the concentration of the absorbing component.
- The diffuse reflection mode is suitable for the opaque liquids, solid and granular samples, in this case part of light is scattered. The diffuse reflectance measurements can be adapted both on apples and cooked purees.
- The transflection mode, with the combination of reflection and transmission, is suitable for characterizing turbid or clear liquids.

2.1.1. Visible (VIS) spectroscopy

2.1.1.1. Introduction

Visible spectroscopy presents the electronic transitions of molecules with the absorption of light from 400 to 780 nm. The absorption peaks are very stable under variable temperature conditions and often presented the typical full width half maximum of approximately 20 nm (Zude, Truppel, & Herold, 2002). VIS spectroscopy is one of the most economic and stable methods to assess color scale values correlated to fruit pigment levels. It can be acquired non-destructively and rapidly on fruit surface,

giving the estimations of chlorophylls, carotenes, xanthophylls, anthocyanins and other phenols in the pigments (De Jager & Roelofs, 1996; Walsh et al., 2020).

2.1.1.2 Application on apples and puree

Several specific VIS fingerprints have been pointed out corresponding to the variability of specific biochemical compositions in fruit pigments (**Table 12**). Particularly, the major absorbers at around 550 nm, 652 nm and 677 nm, associated with anthocyanins, chlorophyll *a* and *b* responsible for the changes in apple skin, are discriminative markers for apple ripening stages (Giusti & Wrolstad, 2001; Pourdarbani et al., 2020b; Zude et al., 2006). However, the chemical attributes of apple flesh are not always corelated to their skin color changes (Ignat et al., 2014). The VIS spectra of apples are limited to assess the soluble solids, titratable acidity, and firmness in fruits. Therefore, the combination of VIS and NIR wavelengths are generally considered to determine the quality of apple and their processed products (see summary in **part II 5.2**).

Table 12. The specific VIS spectra (400-780 nm) of biochemical compositions in fruit pigments.

Wavelength (nm)	Compositions	Ref.
420-503	Carotenes and xanthophylls	(Walsh et al., 2020)
475	beta-carotene	(Walsh et al., 2020)
~ 435, 350-500	xanthophylls, lutein, violaxanthin	(Walsh et al., 2020)
530-550	anthocyanin	(Iglesias & Alegre, 2009; Iglesias et al., 2012; Merzlyak et al., 2003; Toledo-Martín et al., 2016)
~ 650, ~ 680, ~ 720	chlorophyll	(Zude, Herold, & Geyer, 2000)

2.1.2 Near infrared (NIR) spectroscopy

2.1.1.1 Introduction

NIR spectroscopy, based on the electromagnetic radiation, covers the range from 780-2500 nm. NIR spectra are composed of the signals of almost all major structures

and functional groups of organic compounds existing in food (Osborne, 2006). In the assessment of intact fruits and their processed products, the broad NIR bands arising from overlapping absorption are mainly related to the overtones and combinations of vibrational C–H and O–H chemical bonds, primarily associated with water and storage reserves (the dominant macroconstituents of fruit) (Golic, Walsh, & Lawson, 2003; Kawano, 1994).

NIRS has several advantages such as rapid spectrum acquisition, limited sample preparation, absence of chemical waste and simultaneous multianalytes detection. Besides, NIR spectrometers have the potential to characterize a great variety of samples in various forms, including all kinds of solids, liquids, purees, powders, etc. For example, the **Fig. 14** displays the NIR detections on apple (**Fig. 14a**) and purees (**Fig. 14b**) in our work. However, when spectra are used to predict quality traits, they require an initial calibration step, which is time consuming. Indeed, for a set of samples, representative of the expected variability, both NIRS spectra and their corresponding reference data are required to established predictive models using multivariate statistical and mathematical data analyses. Several parameters can be evaluated from a single spectrum, with varying precision.

Fig. 14. An example of the measurements on **(a)** apple and **(b)** purees using the NIR spectrometer.

2.1.1.2 Applications on apples and purees

Since K.H Norris firstly applied NIR spectroscopy to measure moisture of grains in 1964 (Panford, 1987), it has been widely used for quality and safety inspection, classification and sorting of apple industry at-line, on-line or in-line (Huang, Yu, Xu, & Ying, 2008; Walsh et al., 2020; Walsh, McGlone, & Han, 2020; Wang et al., 2015; Xie et al., 2016). Biological characteristics of different apples, which are subject to the variety, harvesting season, geographical origin, maturity level, storage conditions and periods etc., can result in a large variability of NIR spectrum dataset and play an important role in the robustness of developed models (Zhang et al., 2018). Therefore, NIR technique coupled with advanced chemometrics has been applied on raw and processed apples, in order to 1) identify the pre-harvest, post-harvest and processing variability (**Table 13**); and 2) predict their quality attributes (**Table 14**).

Detection of apple and puree variability

For raw apples, both VIS-NIR and NIR techniques have a good ability to discriminate different apple varieties, harvesting from same or different countries (in **Table 13**). Generally, discrimination models based on the VIS-NIR spectra of apples have higher correct classification rates than NIR spectra, taking into account the information of color features. Eisenstecken et al (2019) demonstrated the possibility of NIR to classify apples from three different orchard elevation levels (225, 650, and 1000 m above sea level), with correct classification rates of 93.6% and 77.9% for 'Golden Delicious' grown 1000 and 225 m high, respectively. Schmutzler et al. (2014) perfectly discriminated the Golden Delicious apples harvested in Italy from over 20 varieties. Besides, different storage conditions (regular air or controlled atmosphere, low or warm temperature etc.) (Buccheri et al., 2019; Camps et al., 2007) and storing periods (Beghi et al., 2014) on apples can be successfully identified by NIR. However, there is no report concerning the use of NIR spectroscopy to investigate and discriminate the effects of fruit thinning on apples, as well as their processed product.

For apple products (**Table 15**), VIS-NIR spectroscopy has been successfully

applied to discriminate the adulterated apple juice with an accuracy of 91-100%, and to detect adulterating sucrose and fructose with a limit of 18.5% (León, Kelly, & Downey, 2005). Besides, Reid et al. identified apple juices under heated and non-heated conditions (discrimination accuracy of 77.2%), and the variability of four variety (discrimination accuracy of 82.4-100%) (Reid, Woodcock, O'Donnell, Kelly, & Downey, 2005). For fruit purees, a few studies have mainly aimed at detecting adulterations in mixed purees of different fruit species (Contal, León, & Downey, 2002; Downey & Kelly, 2004). However, we know no report concerning the uses of VIS-NIRS or NIRS techniques to detect the variability in apple purees, such as variety, fruit thinning, and storage periods, which have been highlighted to strongly affect the quality of apple puree.

Table 13. Applications of VIS and NIRS techniques to determine and discriminate the possible factors impacting apple variability.

Variable factors	Technique	Wavelength (nm)	Apples	Origin	Chemometric method	No. apples	Accuracy	Ref.
variety	VIS	380-700	Gala, Braeburn, Pink Lady	Australia, U.K., France, Germany, Italy	PLS-DA	132	94%	(Vincent et al., 2018)
	VIS-NIRS	400-2500	Gala, Elstar, Smoothee	France	FDA	450	> 95%	(Camps et al., 2007)
	VIS-NIRS	600-1200	Golden Delicious, Red Delicious	Italy	LDA	280	100%	(Beghi et al., 2014)
	VIS-NIRS	325-1075	Fuji, Red Delicious, Royal Gala	China	PCA, WT + ANN	90	100%	(He et al., 2005, 2007)
	VIS-NIRS	600-1000, 900-1700	5 varieties	France, Spain, Italy	PCA, LDA, QDA	500	85%, 98%	(Cortés et al., 2019)
	VIS-NIRS	400-1021	Fuji, Red star, Gala	China	SPA+ELM	300	> 96.7%	(Li et al., 2018)
	NIRS	1000-2500	Fuji, Huaniu, Gala, Huangjiao	China	LDA + FAFCM	200	100%	(Wu et al., 2015)
	NIRS	1000-2500	Fuji, New Jonagold, Red Start, Ralls Janet	-	MW-PLSDA	200	98.1%	(Luo et al., 2011)
	NIRS	1000-2500	Fuji, Huaniu, Gala, Huangjiao	China	Gath-Geva clustering	/	96.5%	(Wu et al., 2020)
	NIRS	1000-2500	9 varieties	Italy	PCA, QDA	/	> 86.3%	(Eisenstecken et al., 2019)
elevation levels	VIS-NIRS	380-1030	Fuji	China	VISSA-SR	276	> 97.1%	(Tian et al., 2020)
	FT-NIRS	1000-2500	Golden Delicious	Italy	PCA, QDA	/	> 87.5%	(Eisenstecken et al., 2019)
organic or non-organic	VIS	380-700	Gala, Braeburn, Pink Lady	Australia, U.K., France, Germany, Italy	PLS-DA	132	66%	(Vincent et al., 2018)
	FT-NIRS	900-1720	Gala	U.K.	PLS-DA	60	>96%	(Song et al., 2016)
maturation stages	VIS-NIRS	450-1000	Fuji	Iran	ANN	172	> 99.37	(Pourdarbani et al., 2020a, 2020b)
geographical origins	VIS-NIRS	590-1250	Fuji	China	KNN	600	> 92.3%	(Ma et al., 2020)
	FT-NIRS	1000-2500	Fuji	Japan, China	MW-PLSDA	200	98.6%	(Luo et al., 2011)
	NIRS	1000-2500	Golden Delicious	20 countries	PCA	235	100%	(Schmutzler et al., 2014)
	NIRS	1000-2500	Fuji	China	CARS+PLSDA	208	98.1%	(Li et al., 2018)
marketing grades	NIRS	1000-2500	New Jonagold	China	MW-PLSDA	200	96.0%	(Luo et al., 2011)

storage periods	NIRS	800-2700	Golden Delicious	Italy	LDA	280	> 93.7%	(Giovanelli et al., 2014)
	VIS-NIRS	600-1200	Golden Delicious, Red Delicious	Italy	LDA	280	100%	(Beghi et al., 2014)
stored in air or controlled atmosphere	NIRS	950-1650	Annurca	Italy	PLS-DA	240	93.3%	(Buccheri et al., 2019)
reddened treatments at post-harvest	NIRS	950-1650	Annurca	Italy	PLS-DA	240	96.6%	(Buccheri et al., 2019)
shelf life and cold storage	VIS-NIRS	400-2500	Gala, Elstar, Smoothee	France	FDA	450	> 75%	(Camps et al., 2007)

Prediction of apple and puree quality

Intensive investigations using NIRS have been reported regarding the measurement of apple internal attributes in the past decades (Nicolai et al., 2007). According to previous works (**Table 14**), NIRS has the possibility to evaluate SSC, DMC, TA, starch index, firmness, individual sugars, polypenols, and antioxidant capacity.

Table 14. Some applications of NIR technique to determine apple quality traits.

Apple quality	Ref.
SSC	(Lammertyn et al., 1998; McGlone, et al., 2002; Bosoon Park et al., 2003; Peirs et al., 2003b; Zou et al., 2007; Zude et al., 2006)
DMC	(Kaur et al., 2017; McGlone et al., 2003; Travers et al., 2014; Walsh et al., 2004; Zhang et al., 2019)
TA	(Ignat et al., 2014; Liu et al., 2006; McGlone et al., 2002; Peirs et al., 2002; Pissard et al., 2021)
starch index	(Ignat et al., 2014; Menesatti et al., 2009; Peirs et al., 2001; Peirs et al., 2003a)
firmness	(Fan et al., 2009; Lu, 2004; Lu et al., 2000; McGlone et al., 2002; Mendoza et al., 2012, 2014; Park, et al., 2003; Peng et al., 2005)
individual sugars	(Cho et al., 1998; Costa et al., 2003; Eisenstecken et al., 2015; Liu, et al., 2006)
polyphenols	(Pissard et al., 2018; Pissard et al., 2013; Schmutzler et al., 2014)
antioxidant capacity	(Schmutzler & Huck, 2016)

VIS-NIR and NIR spectroscopy have been applied for the quality analysis of apple based-products, for example, determining the sugars and acids in juices (Chang, Chen, & Tsai, 1998; Sinnaeve et al., 1997; Temma, Hanamatsu, & Shinoki, 2002) and predicting the total ester and volatile compounds in apple cider (Peng, Ge, Cui, & Zhao, 2016; Ye, Gao, Li, Yuan, & Yue, 2016; Ye, Yue, Yuan, & Li, 2014) (**Table 15**). However, so far, there has been no attempt to use such approaches for the quality assessments of apple purees.

Table 15. Applications of Vis-NIR and/or NIR techniques to determine the quality of apple products.

Apple products	Techniques	Wavelengths (nm)	Results	Ref.
apples juice	VIS-NIRS	400-2500	good predictions of total sugars, malic acid, sucrose, fructose and glucose with RPD >4.2	(Sinnaeve et al., 1997)
	VIS-NIRS	680-1235	Absorbance at 912 nm was an important wavelength to estimate SSC in apple juice	(Temma et al., 2002)
	VIS-NIRS	400-2500	detect adulterant apple juice with an accuracy of 91-100%	(León et al., 2005)
	VIS-NIRS	400-2500	discriminate the different heating (77.2%) and variety (82.4-100%)	(Reid et al., 2005)
	NIRS	1000-2500	Absorbance at 2270 nm was a dominant factor to estimate SSC in fruit juices.	(Chang et al., 1998)
	NIRS	1000-2500	detect bacterial contamination in juice	(Rodriguez-Saona et al., 2004)
	NIRS	800-2500	a RMSEP of 0.6 °Brix and R^2 of 0.99 for SSC in juice	(Lu et al., 2007)
	NIRS	900-1350	NIR spectra were sensitive to predict SSC, but not for color deterioration	(Zhu et al., 2011)
	NIRS	800-2500	Good prediction of SSC (R^2 = 0.88, RMSE = 0.28 °Brix,) and TA (R^2 = 0.76, RMSE = 0.24 g/L)	(Włodarska et al., 2018)
apple vinegar	NIRS	1280-2500	discriminate the varieties of fruit vinegars	(Liu et al., 2008)
apple cider	NIRS	850-2500	successfully and simultaneously predict SSC, pH, total acidity, and total ester content in wine	(Ye et al., 2014)
	NIRS	1340-1850; 800-1350	Good predictions of the alcohol strength and for the titratable acidity	(Peng et al., 2016)
	NIRS	800-2500	determine the volatile compounds in apple wines with RPD > 2.9	(Ye et al., 2016)
strawberry, raspberry and apple mixed purees	VIS-NIRS	400-2500	over 10% of apple puree can be detected in the mixed purees	(Contal et al., 2002)
mixed strawberry, apple, and raspberry purees	VIS-NIRS	400-2500	minimum detection levels of apple about 25% and 20% w/w for raspberry and strawberry, respectively.	(Downey et al., 2004)

2.1.3 MIR infrared (MIR) spectroscopy

2.1.3.2 Introduction

Mid-infrared spectroscopy is a powerful tool to elucidate sample structure and identify their chemical compounds, based on the functional groups absorb photons at characteristic frequencies of MIR radiation. The MIR region is between 4000 and 400 cm^{-1}. Typical bands observed by MIR spectroscopy correspond to the fundamental vibrations of the chemical bonds of molecules, which are described by the X-H stretching region (4000-2500 cm^{-1}), the triple-bond region (2500-2000 cm^{-1}), the double-bond region (2000-1500 cm^{-1}) and the fingerprint region (1500-600 cm^{-1}) (Türker-Kaya et al., 2017). MIR spectra have been comprehensively studied in F&V as fresh and processed materials (Bureau, Cozzolino, & Clark, 2019), freeze-dried powders (Bureau et al., 2012) and cell wall purified extracts (McCann et al. 1992; Canteri, Renard, Le Bourvellec, & Bureau, 2019; Kačuráková et al., 1999; Kyomugasho et al., 2015; Szymanska-Chargot et al., 2015). Generally, the fingerprint region and a part of the double-bond region contain complex molecular structure information of components constitutive of F&V. However, identifying or assigning the IR bands in these regions may be difficult because of the band overlapping.

Compared to the low structural selectivity in the broad bands of NIR spectra, more resolved fundamentals of MIR spectra allow to better elucidate the chemical and structural information of samples. However, the lower energy of MIR radiations and the strong water interactions in fruit suspensions prevent the sensitive evaluation of chemical compositions and structural properties (Abbas et al., 2020).

2.1.3.3 Applications on apples and purees

Detection of apple and puree variability

MIR spectroscopy provides rapid and accurate evaluations of raw and processed apple products' quality (**Table 16**). But it is a destructive characterizing method for raw apples, requiring sample pre-treatments to generate homogenates, freeze-dried powders or their extracts (alcohol insoluble solids, cell wall materials etc.). For apple purees,

existing studies were mainly aimed at detecting adulterations in mixed purees of different fruit species (Contal et al., 2002; Defernez et al., 1995; Kemsley, Holland, Defernez, & Wilson, 1996). Particularly, MIR technique combined with partial least squares discrimination analysis (PLS-DA) was able to detect admixture of apple in raspberry puree at the minimum level of 20% (Kemsley et al., 1996); similar detection limits of apple-strawberry mixed purees were obtained using VIS-NIR coupled with PCA and linear discriminate analysis (LDA) (Contal et al., 2002). To date, knowledge is still limited on the use of MIR technique to identify variability in apple puree from raw materials and cooking conditions is limited.

Table 16. Applications of MIR technique to analyze the quality of raw and processed apples.

Apple samples	Applications	Wavenumbers (cm^{-1})	Reference
parenchyma cell wall material	classification of cell wall	1800-900	(Belton et al., 1995)
puree	detect adulteration	1802-899	(Kemsley et al., 1996)
puree	detect adulteration	1802-899	(Holland, Kemsley, & Wilson, 1998)
juice	estimation of total sugar, glucose, fructose, sucrose	1250-900	(Rambla et al., 1998)
Juice	estimation of sucrose, glucose, fructose, citric acid	1250-900	(Tewari et al., 1999)
diluted juice and standard mixtures	estimation of malic acid, tartaric acid, citric acid	1400-1180	(Ayora-Cañada & Lendl, 2000)
juice	estimation of sucrose, glucose, fructose, sorbitol, citric acid, malic acid	1500-950	(Irudayaraj et al., 2003)
centrifuged juice	classification	1250-900	(Gestal et al., 2004)
juice	detect adulteration	1850-880	(Kelly et al., 2005)
juice	detect adulteration	1200-900	(Dhaulaniya et al., 2020)
Juice, phenol fraction	detect authentication	1800-750	(He et al., 2007)
pomace	perdition of sucrose, fructose, glucose, malic acid, total phenolic	4000-500	(Queji et al., 2010)
juice and jam	detect adulteration	4000-500	(Mohamed et al., 2011)
homogenate	estimation of DMC, SSC, TA, sucrose, glucose, fructose, citric acid, malic acid	1700-1010	(Bureau et al., 2012)
freeze-dried powder	estimation of flavan-3-ols, procyanidins, dihydrochalcones, hydrocinnamic acids	1568-1010	(Bureau et al., 2012)
cell wall materials	estimation of galacturonic acid, hemicellulose, cellulose	1500-800	(Szymanska-Chargot et al., 2015)
Alcohol Insoluble Solids	estimation of the degree of methylesterification of pectins	1800-1600	(Kyomugasho et al., 2015)
cell wall materials	estimation cell wall polysaccharides	1800-800	(Liu et al., 2021)

Prediction of apple and puree quality

MIR spectroscopy has been widely applied for the quality analysis of apple based-products, such as fresh homogenates (Bureau et al., 2012), juices (Kelly et al., 2005; León, Kelly, & Downey, 2005; Reid et al., 2005) and purees (Holland et al., 1998) (**Table 16**).

Particularly, direct MIR estimations on fresh apple homogenates performed well in predicting SSC, DMC, TA, some individual sugars and organic acids (Bureau et al., 2012). As infrared spectroscopy is extremely sensitive to changes of hydrogen bonding (Jackson & Mantsch, 1995), the main drawback of spectral measurements is the low sensitivity and limited specific signals of chemical compositions under strong water interactions in fresh apples. To overcome these limitations observed on highly hydrated apples, drying methods with as limited as possible alteration of composition and structure are needed (**Fig. 15**). But sample drying is an expensive and time-consuming operation. Specific MIR fingerprints of individual sugars and organic acids have been comprehensively summarized in a recent review (Bureau et al., 2019). However, these MIR signals come from the direct spectrum on fresh and processed fruits. No work compared the differences and limitations of MIR fingerprint regions on fresh and corresponding freeze-dried apples, as well as their processed purees.

Besides, MIR applications to assess fruit textural properties (mainly focus on cell wall compositions) are always performed on their cell wall materials (AIS) (**Fig. 15**) (Canteri, Renard, Le Bourvellec, & Bureau, 2019; Szymanska-Chargot, Chylinska, Kruk, & Zdunek, 2015). Although some cell wall modifications in plants (Femenia, García-Pascual, Simal, & Rosselló, 2003) and fruits (Cardoso et al., 2009) under heating and dehydration have been investigated by MIR, there is no report concerning MIR spectroscopy to detect cell wall change during fruit puree processing and to monitor rheological and mechanical properties.

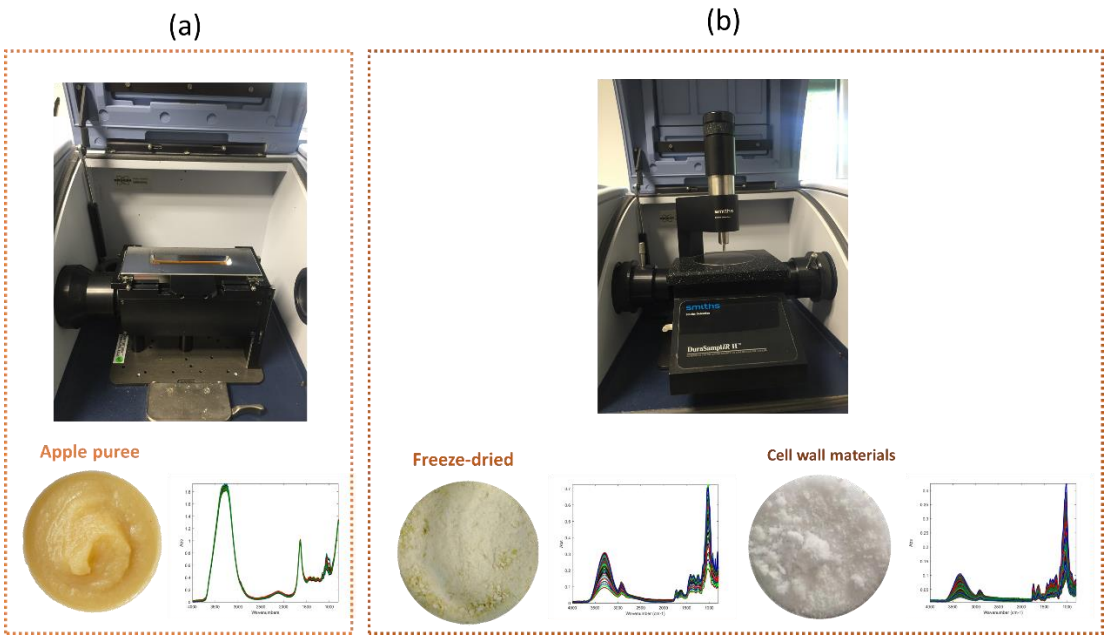

Fig. 15. MIR sensors to characterize **(a)** fresh, and **(b)** freeze-dried and cell wall samples.

2.1.4 Raman spectroscopy

2.1.4.2 Introduction

Raman is a light scattering technique, whereby a molecule scatters incident light from a high intensity laser light source. Generally, most of the scattered light is at the same wavelength (or color) as the laser source and does not provide useful information, which is called Rayleigh Scatter. However, a small amount of light (usually 0.0000001%) is scattered at different wavelengths (or colors), which depend on the chemical structure of the analyte. This is called Raman Scattering. The different peaks in a Raman spectrum show the intensity and wavelength position of the scattered light.

Raman technique is a powerful analytical tool in food quality and safety inspection, such as detection of mycotoxin and common harmful chemical residues in agri-food products (Huang, Cheng, & Lai, 2020; Jiang, Sun, Pu, & Wei, 2018; Wu, Pu, & Sun, 2021; Zhang, Pu, Huang, & Sun, 2021), microbiota evaluation between raw and processed food (He & Sun, 2015). Both Raman and MIR techniques have the potential to study the vibrations of molecular bonds, but these two spectroscopies are complementary in the fact that: i) the phenomenon observed in Raman is the elastic and

inelastic scattering of light (see detailed descriptions from Smith & Dent (2005)), while infrared spectroscopy relies on absorption of infrared light in different amounts and at distinct wavenumbers corresponding to different bond vibrations; ii) Raman spectra usually featured a series of much narrow and sharper spectral peaks than IR spectra. The intensity of these Raman peaks is linearly proportional to the concentration of the molecules, which is an advantage to quantify chemical compositions or identify structural changes in samples (Pelletier, 2003; Qin et al., 2014). iii) Raman is sensitive to homo-nuclear bonds (C-C, C=C and C≡C bonds etc.), whereas infrared spectroscopy is sensitive to hetero-nuclear functional group variations and polar bonds (C-O and C = O), especially OH stretching in water (Griffiths & Miseo, 2014; Pistorius, 1996).

2.1.4.3 Applications on apples and purees

Unlike IR spectroscopy, Raman spectroscopy has little applications for in-line quality evaluation in food industry. Most of the reported Raman works focused mainly on the detection of the pesticide residues on apple surfaces or their processed products (Xu, Gao, Han, & Zhao, 2017; Yang & Ying, 2011). It has been successfully applied to determine several nutrients, such as lycopene and β-carotene in tomato (Baranska, Schütze, & Schulz, 2006), the fructose in apple juice (Moreira, Buffon, de Sá, & Stradiotto, 2021). This technique has some limitations, especially the low ratio of signal to noise (SNR). If the sample contains a fluorescent compound, it becomes difficult to obtain Raman spectra. The disadvantage in this case is that a sample could be heated during measurement due to the high energy of laser light, which might alter or destroy it. It is then important to explore and optimize the balancing operation between the laser power and the time of measurement in order to obtain a spectrum with a high SNR without heating the sample.

To date, no detailed study has compared the differences and limitations of Raman and infrared spectroscopy (NIR and MIR) to determine the structural and rheological properties of fruit purees.

2.2 Spectroscopic imaging techniques

With the development of optical sensing and imaging techniques, it has become possible to combine conventional imaging and spectroscopy to simultaneous obtain spatial and spectral information within individual F&V (Gowen et al., 2007), for non-destructive defect detection (Lu & Lu, 2017) or the destructive quality inspection of individual fruits (Su & Sun, 2018). The purpose of this part is to give a detailed overview of spectroscopic imaging applications, mainly including VIS and NIR imaging, FT-IR imaging, Raman imaging, for detecting the heterogeneity of F&V. The advantages and disadvantages of these techniques are compared and discussed.

2.2.1 VISNIR and NIR imaging

2.2.1.2 Introduction

Since the term 'hyperspectral imaging' (HSI) was firstly mentioned for works on remote sensing (Goetz, Vane, Solomon, & Rock, 1985), it has emerged as an efficient scientific tool to inspect and assess the quality of F&V (Du et al., 2020; Hussain, Pu, & Sun, 2018; Lu, Saeys, Kim, Peng, & Lu, 2020; Pathmanaban, Gnanavel, & Anandan, 2019; Zhang et al., 2018). It opens the possibility to target, visualize and even qualify the chemical and physical heterogeneity within an individual F&V, using the informative spectrum acquired at each pixel of each image in visible (VIS) (380-780 nm) and/or near infrared (NIR) (780- 2500 nm) ranges (**Table 16**).

The VISNIR-HSI technique gives a relatively higher radiation penetration rate in samples with a lower cost of machines (McGlone, Clark, & Jordan, 2007; McGlone & Kawano, 1998), whereas the NIR-HSI technique is less influenced by light scattering (Boldrini, Kessler, Rebner, & Kessler, 2012) and is better correlated to chemical components (Baeten & Dardenne, 2005). A typical HSI system generally consists in five major modules (**Fig. 16**): i) a light source for the illumination; ii) a wavelength dispersion device for spectrograph; iii) a camera for two-dimensions geographic space, iv) a sample platform and v) a computer (Gowen et al., 2007; Li et al., 2017).

2.2.1.3 Applications of detecting heterogeneity

Several applications of VISNIR-HSI and NIR-HSI techniques to detect the heterogeneity of individual F&V (**Table 17**) have been reported. VISNIR-HSI (400-1100 nm) is most commonly used to detect the distribution of moisture content and SSC in individually measured apples (Mo et al., 2017), melon (Sugiyama, 1999), kiwifruits (Martinsen & Schaare, 1998) and carrots (Liu et al., 2016; Yang et al., 2020). However, NIR-HSI (900-2500 nm) also gives a good evaluation of quality parameters, such as SSC and starch content in apples (Ignat, et al., 2014; Ma et al., 2018). Moreover, the longer wavelengths (> 1100 nm) allow to enhance the sensitivity to detect SSC (Ma et al., 2018; Menesatti et al., 2009; Peirs et al., 2003b; Sun et al., 2019; Sun, Zhang, Liu, & Wang, 2017) and starch content (Menesatti et al., 2009; Peirs et al., 2003a). NIR-HSI is even able to evaluate other quality attributes, such as pH (Ma et al., 2021), and firmness (Sun et al., 2017) in individual F&V. Accordingly, both non-destructive and destructive HSI scanning provide in a rapid and simple way, an overview of the heterogeneity of properties and composition (mainly SSC, pH, DMC, firmness, starch) within tissues of individual F&V.

Table 17. An overview of VISNIR-HSI and NIR-HSI techniques to detect the heterogeneity of individual F&V.

Techniques	Wavelength (nm)	Species	Parameters	Sample preparation	Conclusions	References
NIR-HSI	913-2519	apple	SSC	slices	higher SSC near apple peels than the central parts	(Ma et al., 2018)
NIR-HSI	900-1700	apple	starch	slices	starch heterogeneity in different apple varieties during ripening	(Peirs et al., 2003a)
NIR-HSI	1000-1700	apple	starch	slices	higher concentration in the outer cortex than in the core	(Menesatti et al., 2009)
VISNIR-HSI	400-1000	apple	SSC	slices	large heterogeneity of SSC from the skin to the core	(Mo et al., 2017)
VISNIR-HSI	400-1000	melon	SSC	destructive	large heterogeneity of SSC between surface, inner cortex and core tissues.	(Sugiyama, 1999)
NIR-HSI	900-1700	melon	SSC, firmness	destructive	higher SSC close to the fleshy part than in the part near peel.	(Sun et al., 2017)
VISNIR-HSI	400-1000	kiwi	fructose, glucose, sucrose	slices	the heterogeneity of sugar contents between inner cortex and the core tissues.	(Hu et al., 2017)
VISNIR-HSI	650-1100	kiwi	SSC	slices	quick increase of SSC in the core than in the inner and outer pericarp.	(Martinsen et al., 1998)
NIR-HSI	1002-2300	kiwi	SSC, pH	non-destructive	large heterogeneity of SSC between stem and calyx parts.	(Ma et al., 2021)
NIR-HSI	1000-1650	kiwi	ripeness	destructive	heterogeneity of the core and outer pericarp during ripening	(Serranti et al., 2017)
VISNIR-HSI	380- 1023	kiwi	SSC, firmness	non-destructive	heterogeneous distribution of firmness and SSC	(Zhu et al , 2017)
VISNIR-HSI	450-950	blueberry	SSC, firmness	non-destructive	heterogeneity of SSC from outer to internal parts, and from stem to calyx parts.	(Qiao et al., 2019)
VISNIR-HSI	500-1000	blueberry	SSC, firmness	non-destructive	heterogeneity of SSC from stem and calyx parts	(Leiva-Valenzuela et a., 2013)
VISNIR-HSI	400-1000	pear	SSC	non-destructive	heterogeneity of SSC from the top to the bottom.	(Zhang et al., 2018)
VISNIR-HSI	380-1030 874-1734	jujube	SSC	non-destructive	heterogeneity of SSC between inner and outer tissues.	(Zhao et al., 2020)
NIR-HSI	900-1700	jujube	SSC	non-destructive	asymmetric and non-uniform distribution of SSC	(Sun et al., 2019)
VISNIR-HSI	400-1000 880-1720	mango	moisture	slices,	heterogeneity of moisture loss in mango slices	(Pu & Sun, 2015, 2017)
VISNIR-HSI	405-970	carrot	moisture	slices	heterogeneity of moisture distribution at boundary, middle and center positions	(Liu et al., 2016; Yang et al., 2020)
VISNIR-HSI	400-1000	tomato	moisture	non-destructive	slight change in the spatial distribution of moisture	(Mollazade et al., 2012)

NIR-HSI	1000-1550	tomato	moisture, SSC, pH	non-destructive	large heterogeneity of SSC and pH from peripheries to center areas	(Rahman et al., 2017)
VISNIR-HSI	400-1000	potato	moisture	slices	limited variation of moisture in slices	(Amjad et al., 2018; Xiao et al., 2020)
VISNIR-HSI	400-1000	potato	SSC	slices	significant decrease of SSC from center to epidermis areas	(Shao et al., 2020)

2.2.1.4 Challenges of detecting heterogeneity

To investigate the heterogeneity of individual F&V, two specific concerns have been addressed in detail:

- The light source: it plays an important role in reflectance HSI mode to acquire images and extract relevant information from them, such as reducing noise and shadow or enhancing contrast (Liu, Zeng, & Sun, 2015; Mollazade et al., 2012; Zhang et al., 2014). Moreover, the different brightness levels between the central parts and borders for spherical fruits significantly influence the detection accuracy (Zhao et al., 2020).
- The sample platform: two efficient sample platforms are usually used for a non-destructive global scanning of fruit surfaces and a destructive scanning of different shapes of fruit tissues (cylinder, cube, long strip, slice, etc.).
    - The non-destructive HSI scanning of sample surfaces gives an initial overview of internal sample heterogeneity, especially for materials with thin peels (Qiao et al., 2019; Rahman et al., 2017; Zhang et al., 2018). For example, a hyperspectral platform coupled with two motor-driven rollers is described to obtain a 360° HSI image of individual fruit and vegetable (**Fig. 16b)** (Ma et al., 2021; Sun et al., 2018). It allows also to have a better insight of fruit heterogeneity with a relatively more efficient and easier way to generate a 3D plot than the traditional 2D plot scanning on a fruit surface portion (Zhang et al., 2018; Zhu et al., 2017). However, the non-destructive HSI methods always suffer from several drawbacks: i) they are suitable for standard or approximately cylindrical fruits, but difficult for irregularly shaped fruits; ii) complex chemometrics are needed to normalize the influence of pixels in the center and edges of an image and iii) visualization of thick peel samples is impossible due to the limited radiation penetration.
    - The destructive HSI scanning of F&V gives more detailed insights of the heterogeneity among different tissues (**Fig. 16a**). A first look was then done of the starch distribution in apple slices using the first principal component scores of HSI imaging at specific wavenumbers (Peirs et al., 2003a). However, crucial concerns have to be taken into account: i) the limited stability of highly hydrated cut pieces of F&V due to their rapid

oxidization and ii) heavy workload to characterize all sample areas targeted for model calibration (Mo et al., 2017; Shao et al., 2020; Sun et al., 2017).

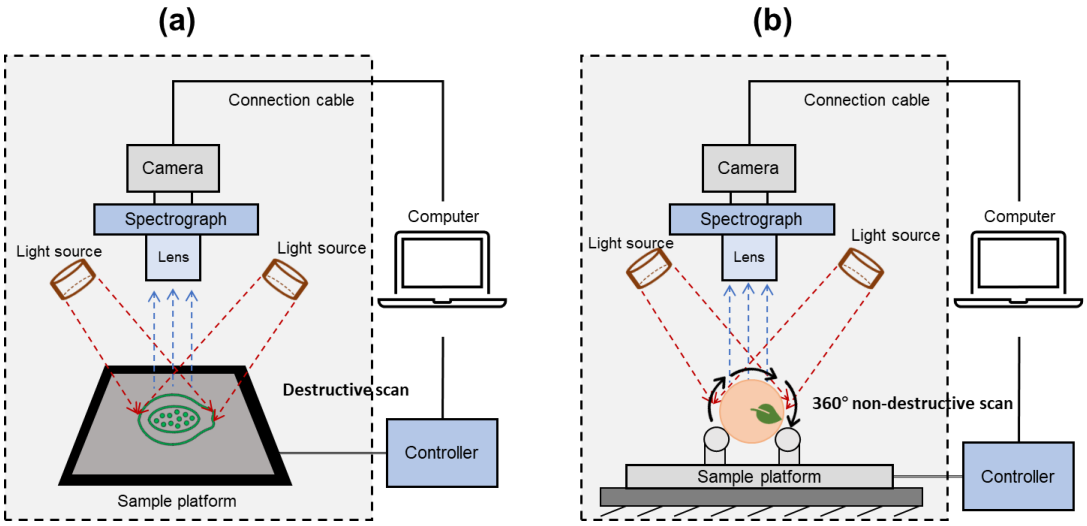

Fig. 16. The destructive **(a)** and non-destructive **(b)** HSI scanning methods to detect the heterogeneity of F&V.

2.2.2 FT-IR imaging

2.2.2.2 Introduction

FT-IR imaging is one way to create spatially resolved chemical images. Each pixel of these images consists of a whole MIR spectrum. A FT-IR imaging spectrometer is schematized (**Fig. 17a**) with the possibility to select among three main sampling modes which are transmission, transflection and attenuated total reflection (ATR) (**Fig. 17b**). The FT-IR imaging acquisition could be accomplished by raster-scanning with a point illumination or by using wide-field illumination and focal plane array or linear array detectors (Bhargava, 2012). These three kinds of acquisition operations are compared and further discussed in the **Part 6.3** of Raman imaging technique. For FT-IR images, the point scanning method is generally applied on F&V samples in order to get a high-resolution inspection of heterogeneity of tissues and more, between single cells (González-Cabrera, Domínguez-Vidal, & Ayora-Cañada, 2018). Particularly, the possibility to pretreat the experimental samples as a thin layer allows an accurate assessment under transmission and transflection modes (Kazarian & Chan, 2013).

Differently, ATR-FTIR mode is directed through an internal reflection element (IRE) with a high refractive index (diamond, zinc selenide, germanium or silicon etc.) and the samples must be in direct contact with IRE (Baker et al., 2014). The penetration depth of FT-IR generally reaches around 1 to 2 µm within the wavenumber regions 1800-900 cm^{-1} (Bassan et al., 2013). In most of the cases, it is necessary to prepare microtome samples with a thickness over 5 µm for ATR-FTIR imaging analysis, permitting the total absorption of the emitted IR radiation (Schulz, Krähmer, Naumann, & Gudi, 2014).

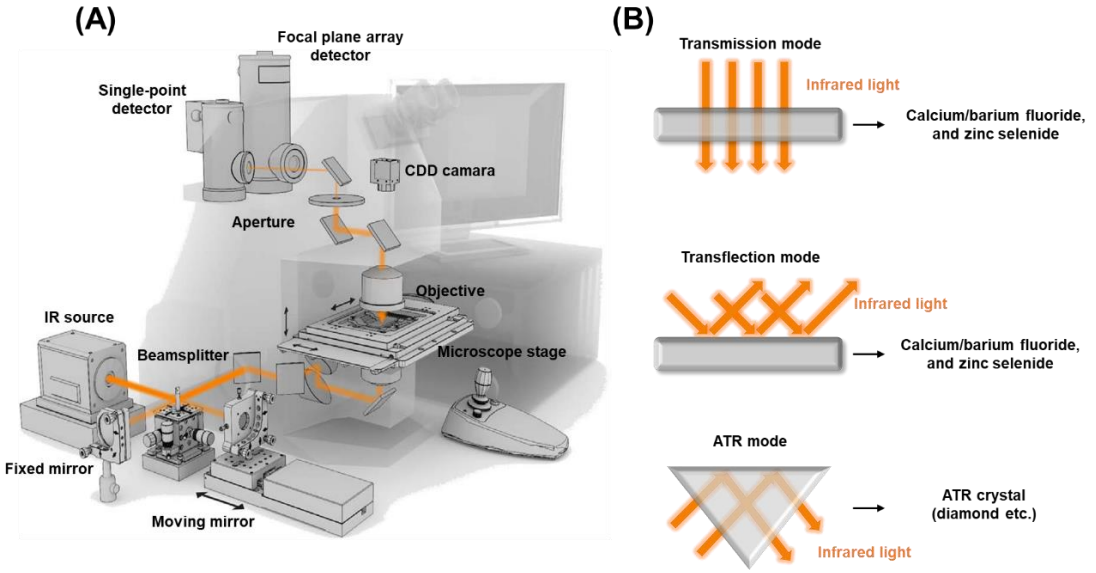

Fig. 17. (a) The schematic of modern FT-IR imaging spectrometer and **(b)** the three main sampling models for FT-IR imaging spectroscopy.

(Figures adapted from Baker et al., 2014).

2.2.2.3 Applications of detecting heterogeneity

The plant physiologists' community started developing FT-IR imaging in the 1990s, to investigate plant tissues (Stewart, 1996) and more specifically the chemical constituents of cell walls (McCann et al., 1997). Many reports have successfully described FT-IR imaging to gain further insight on the plant physiology related to heterogeneity of the structure and composition as it allows a high-resolution of single cell analysis (Chan et al., 2020; Vongsvivut et al., 2019). Some applications have been done to characterize seeds (Timilsena et al., 2019), cereals (Barron et al., 2005; Guendel et al., 2018b) and woods (Cuello et al., 2018). This technique provides important information regarding the major cell wall components such as lignin, cellulose and

various polysaccharides (Schulz et al., 2014). Coupled with PCA, it also allows to follow the biochemical changes during the ripening of olive fruits (Guinda, Rada, Delgado, Gutiérrez-Adánez, & Castellano, 2010) and particularly the solubilization and depolymerization of cell wall polysaccharides and the heterogeneity of polyphenols (carboxylic acids related to phenols) and sugars (galactose, arabinose and glucose etc.) in different tissues (González-Cabrera et al., 2018).

2.2.2.4 Challenges of detecting heterogeneity

FT-IR imaging technique shows limited applications directly on the surface of intact F&V, since the reflection intensity of the samples are relatively low. In order to obtain FT-IR images with a high-resolution, a long acquisition time is needed from several minutes to few days per sample. It is much slower than using VIS-NIR-HSI and NIR-HSI system, and not suitable for numerous analyses.

Further, fresh F&V tissues are almost always associated to a large amount of water, which strongly disturbs the informative infrared signals of interest, such as cellulose, pectins, polyphenols, individual sugars etc. Thus, a careful sample preparation using a cryogenic sectioning technique is necessary to obtain dehydrated or freeze-dried objects. Util now, this technique has mainly been dedicated to agricultural products with lower water contents, such as black peppers (Lafeuille, Frégière-Salomon, Michelet, & Henry, 2020) or walnut shell (Xiao et al., 2020). Some difficulties are specific to F&V such as a rapid oxidization of samples during drying and the fact they are soft tissues. Moreover, in such temperature conditions (usually around -18 °C), the vitreous ice formed within frozen sample tissues acts as the supporting medium, but the intercellular ice crystals can damage the samples (Duncan & Williams, 1983; Schulz et al., 2014). From our point of view, the investigations of F&V tissues or cells using FT-IR imaging technique would be more suitable for F&V peels, hypodermis and epidermis tissues, with harder tissues, a relatively lower moisture content and less sensitive to enzymatic oxidations (González-Cabrera et al., 2018).

2.2.3 Raman imaging

2.2.3.2 Introduction

Raman chemical imaging can combine both images and Raman spectra to detect the heterogeneity of biochemical and structural properties in F&V at the level of tissues. A Raman chemical imaging system consists of a light source, several optical lenses, a wavelength separation device (tunable filter for global imaging devices or diffraction grating for dispersive devices and interferometer for Fourier-Transform devices) and detectors (Qin, Chao, & Kim, 2010; Schulz, Baranska, & Baranski, 2005; Yaseen, Sun, & Cheng, 2017). In practice, specific comparative operations need to be considered for Raman imaging detections applied to F&V heterogeneity, especially for the two selections:

- The scanning methods: generally, there are three major methods to collect Raman images, including point-scan (PS), line-scan (LS) and area-scan (AS) (Qin et al., 2019; Yaseen et al., 2017) (**Fig. 18**):
    - The PS method (**Fig. 18a**) uses a specific point to acquire the Raman spectrum at each pixel to finally generate an image with similar resolution, step size and power density. However, it usually requires long acquisition durations from hundreds of milliseconds (at least a laser dwelling time of 100 ms for each pixel volume of ~1 μm^3 (Qi et al., 2014)) up to several seconds per point measurements depending on the type of samples and of Raman systems. Until now, PS is the most suitable and widely applied method to highlight the heterogeneity of minor components ($mg.kg^{-1}$) (lycopene, α- and β-carotene etc.) and the cell wall major ones (pectins, cellulose, lignins and various polysaccharides) in F&V tissues (**Table 18**).
    - The LS method (**Fig. 18b**) allows a simultaneous acquisition of a given line of spatial information instead of single points, and generates a Raman imaging hypercube after scanning sample entire surface. The LS method has been then considered as a potential high-throughput macro-scale Raman chemical imaging to map the carotenoids distribution in tomato (Qin, Chao, & Kim, 2011, 2012) and in carrot slices (Qin et al., 2017).

Indeed, it takes few minutes to scan over the entire surface of a sample, which is much faster than PS mapping (eg. 5.5 mins for LS *vs* 30 mins of PS according to Bonvenkamp et al. 2019), but it has the disadvantage of a limited power density (eg. 0.31 MW/cm^2 for PS *vs* 0.003 MW/cm^2 for LS according to Bonvenkamp et al. 2019) and therefore weak signals (Yaseen et al., 2017).

- AS method (**Fig. 18c**) exists to scan a relatively large area of sample surface by a defocused laser spot, but it is not a high-throughput strategy. It allows to create Raman images using specific wavenumber ranges.

- The laser wavelengths: the powerful monochromatic lasers at 448 nm, 532 nm, 785 nm and 1064 nm have been used as Raman excitation sources to detect the heterogeneity of F&V tissues (**Table 18**). The increase of excitation intensity or reduction of laser wavelength can enhance the Raman signals, but often introduces a strong fluorescence resulting in sample degradation or even burning. The lasers in the visible region, 488 nm and 532 nm, are not often used to visualize chlorophylls, carotenoids and anthocyanins in horticultural crops because of the strong fluorescence signals (Qin et al., 2019). However, they are used to evaluate the distribution and accumulation of cell wall polysaccharides (eg. cellulose and pectins) (Pan et al., 2017; Szymańska-Chargot et al., 2016) and the hydrogen bonding status of water in tissues (Li et al., 2020). The diode laser at 785 nm and Nd:YAG (neodymium-doped yttrium aluminum garnet) laser at 1064 nm minimize fluorescence interferences, and have been successfully applied to assess the distribution of carotenoids in carrot (Baranska, Baranski, Schulz, & Nothnagel, 2006; Qin et al., 2017), lycopene in tomato (Qin et al., 2011, 2012), as well as components (e.g. cellulose, lignins and pectins etc.) of F&V cell walls (Baranska & Schulz, 2005; Baranska et al., 2005).

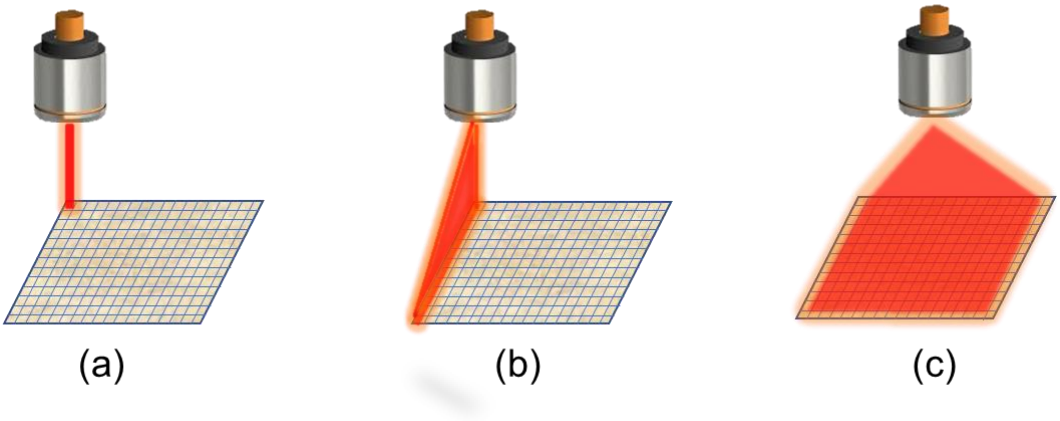

Fig. 18. Three different scanning methods of the Raman spectroscopic imaging acquisition: **(a)** point-scanning (PS); **(b)** line-scanning; **(c)** area-scanning (AS).

2.2.3.3 Applications of detecting heterogeneity

Raman imaging has been applied mainly for the study of carotenoids and cell-wall components in F&V (**Table 18**). Several applications aim to visualize the aggregation of lycopene in locular tissues and outer pericarp of individual tomato fruit (Ishigaki et al., 2017; Qin et al., 2011, 2012); the carotenoids in parenchymatic center, phloem, and peripheral parenchyma of carrots (Baranska & Schulz, 2005; Baranska et al., 2005; Gonzalvez et al., 2014) and some specific crystals of different carotenoids in carrot cells (Roman et al., 2015). Studies on cell wall concern the distribution of pectins in apple during maturation and storage (Szymańska-Chargot et al., 2016), the variations of cellulose and pectins in pear (Pan et al., 2017) and the specific accumulation of lignins in mature carrot cells (Roman, Dobrowolski, Baranska, & Baranski, 2011). Moreover, the water properties are evaluated in apple tissues by the study of hydrogen bond status (Li et al., 2020). These studies demonstrate the high potential of the macro or high-spatial-resolution Raman imaging for visualizing macro and micro chemical and structural composition, spatial distribution and morphological features in F&V at tissue and cellular levels.

2.2.3.4 Challenges of detecting heterogeneity

Compared to FT-IR imaging, Raman imaging technique requires little to no sample preparation and is not sensitive to water interferences, while the FT-IR method has

constraints on the thickness, uniformity and freeze-drying operations on samples. Raman spectroscopic information can be directly acquired on fresh F&V tissues with a thickness from tens or hundreds micrometers to few millimeters (**Table 18**). However, this technique is not suitable to give a rapid overview of heterogeneity in numerous fruits, because of the long acquisition and limited scanning area.

Table 18. Overview of Raman imaging applications to detect heterogeneity of individual F&V.

Species	Scan	Laser (nm)	Regions (cm^{-1})	Fresh samples	Applications	Conclusions	Ref.
apple	Point-scan	532	50 - 4000	slices, thickness (180 μm)	cell wall polysaccharides	a large heterogeneous distribution of pectins in cell walls during ripening, then evenly dispersed along cell wall during storage	(Szymańska-Chargot et al., 2016)
			2700 - 3800	slices, thickness (1 mm)	water hydrogen bonding status	the strongest and weakest hydrogen bonds mainly in cell wall or intercellular regions	(Li et al., 2020)
tomato	Point-scan	532	50 - 4000	slices, thickness (180 μm)	cell wall polysaccharides	high concentrations of pectic polysaccharides in cell wall corners in mature green parenchyma	(Chylińska et al., 2017)
			900 - 1700	tissues, height × area (2.21 cm × 3.38 cm^2)	lycopene	lycopene with different aggregates confirmations distributed inhomogeneously in tissues	(Ishigaki et al., 2017)
			150 - 3500	slices, thickness (~20 μm)	cellulose, pectins	middle lamella between primary walls enriched in pectins	(Chylińska et al., 2014)
		785	200 - 2500	slices (5, 10 mm thickness)	lycopene	a large difference of lycopene content in locular tissues and outer pericarp	(Qin et al., 2011, 2012)
carrot	Line-scan	785	100 - 2900	slices, thickness (2, 5, 8 mm)	carotenoids	non-uniform carotenoid concentration in tissues	(Qin et al., 2017)
	Point-scan	488, 532	0 - 3600	slices	carotenoids	various carotenoid molecules distributed homogeneously in root cells	(Roman et al., 2015)
		532	500 - 2500	slices, thickness (500 μm)	β-carotene	heterogeneous distribution of β-carotene predominantly in the secondary phloem tissue and periderm	(Gonzalvez et al., 2014)
			100 - 4000	slices, thickness (3 mm)	carotenoids	heterogeneous compositions of α-, β-carotene, lutein and lycopene	(Baranska, Baranski, et al., 2006)
		1064	100 - 4000	slices, thickness (3 mm)	polyacetylenes, carotenoids, starch, pectins	heterogeneous distribution of carotenoids in tissues: parenchymatic center of the root, phloem, and peripheral parenchyma	(Baranska & Schulz, 2005; Baranska et al., 2005)
			100 - 4000	slices	polyacetylenes, lignins, pectins, cellulose	specific tissue of accumulation of starch and cell wall components (lignins, pectins and cellulose)	(Roman et al., 2011)
mango, potato	Point-scan	532, 1064	690 - 990	slices, thickness (~400 μm)	β-carotene, starch	high degree of heterogeneity	(Brackmann et al., 2011)
pear	Point-scan	532	300 - 3100	slices, thickness (0.5 mm)	cellulose, pectins	different distributions of pectins and cellulose in cell walls	(Pan et al., 2017)

2.2.4 Other imaging techniques

Other imaging techniques, such as computer/machine vision, thermal imaging, X-ray imaging (XRI), thermal imaging, Magnetic resonance imaging (MRI) etc., have the capability to provide superior spatial information and have been proven to be efficient scientific tools to investigate fruit heterogeneity (Mebatsion et al., 2008; Pu, Feng, et al., 2015; Zhang et al., 2014).

Machine vision technology is based on a color camera and produce BW (black and write) or RGB (red, green, and blue) images. In the case of color machine vision, three filters imitate human vision (Hameed, Chai, & Rassau, 2018). This technique has been successfully applied for fruit classification, grading and localization in the tree (Gongal et al., 2015; Naik & Patel, 2017). However, the limited information in the visible area from machine vision cannot provide the estimation of internal chemical and physical properties in fruits.

Thermal imaging is a non-destructive technique of monitoring temperature based on the infrared radiation emitted by an object (Ali, Hashim, Aziz, & Lasekan, 2020). It can provide the key facts of heat distribution and basic dimension, as well as structural analysis on fruit surface. However, most chemical and physical heterogeneity in individual fruits can not be detected based only on the thermal imaging system.

MRI can map the density of proton molecules to obtain 2D and 3D imaging of individual fruits, with the advantages of high abundance and activeness (Clark et al., 1997; Pathmanaban et al., 2019). It has been used to detect the heterogeneity of water content and structure in apples (Defraeye et al., 2013; Herremans et al., 2014). However, the high-cost and long-time imaging acquisition limit its applications.

XRI can provide physical heterogeneity analysis in fruits based on the differences of X-ray attenuation in different kinds of tissues (Nugraha et al., 2019). Besides, the X-ray computed tomography images, taking from multiple angles, can generate 3D images of individual fruits using mathematical algorithm. This technique can be used to analyze the tissue density, pore structure, moisture and void networks in apples (Janssen et al., 2020a; Mendoza et al., 2007; Nugraha et al., 2019). In practice, the acquisition of high-

quality X-ray images needs a relatively long-time, and the reconstruction of the 3D X-ray images is difficult to achieve, as it requires powerful computational hardware and is time-consuming.

3. Chemometrics

The application of spectroscopic techniques has been greatly reinforced by the development of chemometrics methods and the calculation speed of computers in parallel. The chemometrics field is very large and some of the most common mathematical and statistical methods have been detailed in reviews (Mishra et al., 2021; Oliveri, Malegori, & Casale, 2020; Roger & Boulet, 2018). The applications of chemometrics tools to detect food quality can be mainly divided into two groups, including: i) classification modelling to discriminate the sample variability (**Part II, 7.2**), and ii) regression modelling to predict quantitative variables (**Part II, 7.3**).

3.1Spectral data pre-processing

Spectral data acquired on fruit samples with infrared spectrometer are associated with background information and noises. Before using chemometric tools, it is often necessary to apply a pre-treatment on spectral dataset. There are some common pre-processing methods, such as baseline correction, smoothing, derivatives, standard normal variate transformation (SNV) etc.

Baseline correction is an important pre-processing technique used to separate true spectroscopic signals from interferences or to remove background effects (Liland, Rukke, Olsen, & Isaksson, 2011). It has been reported to improve the VIS-NIR prediction of SSC in apples (Vincent et al., 2018; Xia, Fan, Li, Tian, Huang, & Chen, 2020), and the MIR evaluation of the sweetener in apple juices (Dhaulaniya, Balan, Sodhi, Kelly, Cannavan, & Singh, 2020).

Spectral smoothing, mainly includes moving smoothing and, Savitzky-Golay smoothing is one of the methods used to eliminate noises (Gorry, 1990). Optimizing the

smoothing window width of spectral matrix is crucial to increase the Signal-to-Noise Ratio (SNR). It has been widely applied to the NIR prediction of SSC in apples (Sun, Zhang, Pan, & Liu, 2009; Tian, Fan, Li, Xia, Huang, & Zhao, 2019) and the detection of harvesting optimal picking dates (Peirs et al., 2001). Moreover, it is often combined with several other pre-processing methods, such as SNV, multiplicative scatter correction (MSC) etc.

Regarding the elimination or minimization of unwanted systematic effects (the multiplicative interferences of scatter, particle size, and the change of light distance etc.) (Barnes, Dhanoa, & Lister, 1989), SNV as a typical mathematical method is useful to correct both, baseline shifts and global intensity variations (Barnes et al., 1989). The SNV pre-treated NIR (Dong & Guo, 2015; Qing, Ji, & Zude, 2007) and MIR spectra (Bureau et al., 2009; Labaky et al., 2021) give good a prediction accuracy of fruit quality traits.

Derivatives including first and second derivatives are used to remove background and increase spectral resolution. There are two algorithms, direct differentiation and Savitzky-Golay. The most used method is the Savitzky-Golay, where the data within a moving window are fitted by a polynomial of a given degree to generate a differential of a chosen degree (Gorry, 1990). The selection of the proper width of the moving window is very important in this function. Many calibration models using derivatives of NIR spectra, give good prediction of apple quality traits (Beghi, Giovenzana, Civelli, & Guidetti, 2016; Giovanelli et al., 2014; Lammertyn, Nicolai, Ooms, De Smedt, & De Baerdemaeker, 1998).

Recently, Mishra et al. (2020) demonstrated the benefits of using the multi-block approach for spectral data preprocessing, where several pre-treatments were combined using sequential and orthogonalized partial least squares (SO-PLS), thus leading to a boosting procedure. They purposed a chemometric strategy named ‘SPORT’, both allowing both the fusion of multiple preprocessings and the identification of the best preprocessing techniques and their combinations.

3.2Classification modelling

3.2.1 Principal component analysis (PCA)

PCA, which was proposed by K. Pearson (1901), is one of the most widely used basic tools for discriminating the samples based on their spectra and/or reference dataset. It is a data-reduction technique, which reduces the dimensions of a dataset into its principal components (PC) with minimal loss of information. The PCs are uncorrelated and the former retain most of the variation in all the original dataset (Jolliffe, 2002), indicating the directions where there is the most variance. Although some precision is lost, data analysis is facilitated since the number of variables is reduced and can be visualized graphically. This is an unsupervised method, which can be used to identify informative spectral features and reduce noise.

PCA depends on the assumption that a large variability (i.e., a high variance value) is synonymous with a high amount of information (Oliveri et al., 2020). For this reason, PCA algorithms search for the maximum variance direction, in the multidimensional space of the original dataset (**Fig. 19**). The maximum variance direction represents the first principal component (PC1). The second PC (PC2) keeps the maximum variance among all directions orthogonal to (= not correlated to) the PC1. These two new variables (PC1 and PC2) are thus not intercorrelated (Jolliffe & Cadima, 2016; Oliveri et al., 2020). All subsequent principal components are calculated iteratively and are by construction not correlated with one another.

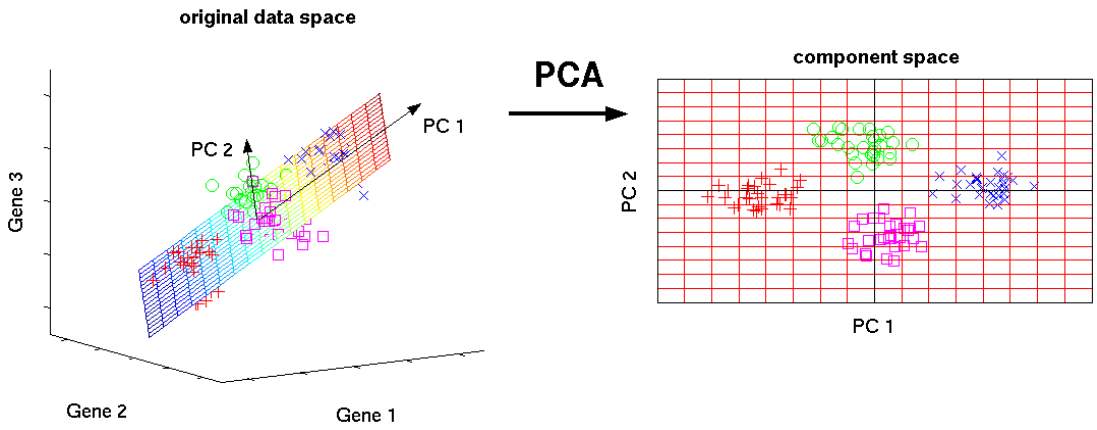

Fig. 19. A brief graph to explain PCA.

(Figure adapted from Ghosh, 2019)

VIS-NIR or NIR spectroscopic techniques coupled with PCA show a good ability to detect a large apple variability due to varieties (Cortés et al., 2019; Daniela Eisenstecken et al., 2019; He et al., 2005, 2007) and geographical origins (Daniela Eisenstecken et al., 2019; Schmutzler et al., 2014), as well as the concentrations of different mixed fruit purees (Contal et al., 2002).

3.2.2 Discriminant analysis (DA)

Discriminant analysis (DA) separates samples into classes, minimizing the variance within the class and maximizing the variance between classes, and finding the linear combination of the original variables (directions). It is a supervised method requiring the knowledge of group memberships for each sample. Thus, it is usually applied to the same sample types as PCA, where the latter technique can be used to reduce the number of variables in the data set and the resultant principal components (PCs) are then used in DA to define and predict classes (Mendlein, Szkudlarek, & Goodpaster, 2013).

DA is a form of supervised pattern recognition, such as linear discriminant analysis (LDA) or factorial discriminant analysis (FDA), well-known as chemometric approaches for solving classification problems in chemistry. Different from PCA establishing the directions of maximal variance, LDA and FDA aim to separate the

known classes by creating a new linear axis and projecting data on that axis (Raschka, 2014) (**Fig. 20**). All the variance between classes is calculated and defined as the distance between the mean of different classes to maximize the distance between classes. Besides, the variances of each class are also calculated and is defined as the distance between the mean and the sample of every class. Finally, the Fisher's criterion tries to find the lower-dimensional space that maximizes the variance between classes and minimizes the variance within classes.

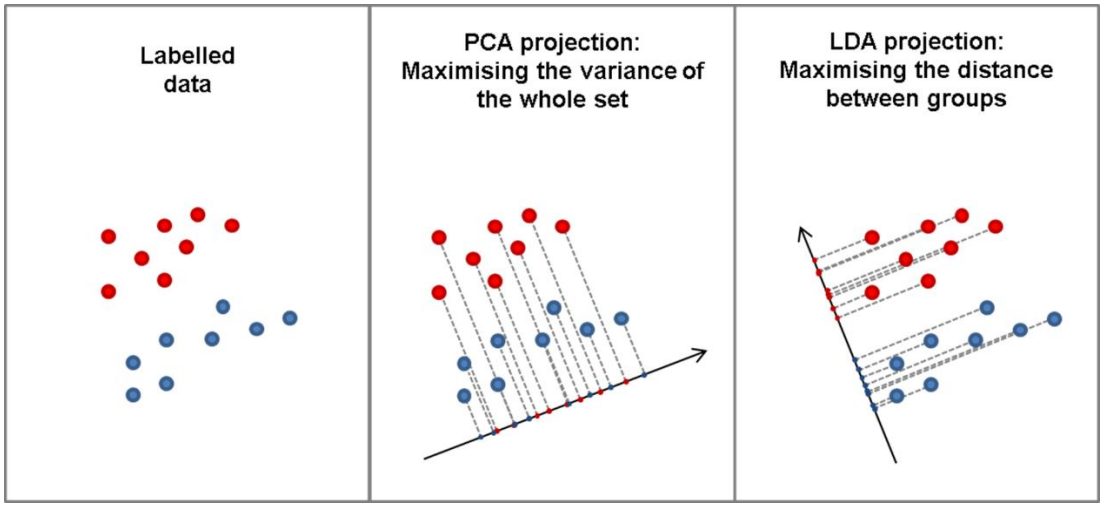

Fig. 20. The differences of PCA and LDA methods.

(Figure adapted from Pelliccia, 2018)

FDA has been applied on the NIR spectra of apples to successfully discriminate the different varieties and post-harvest storage conditions (Camps et al., 2007). Bureau et al. demonstrated the potential to apply FDA on MIR information to discriminate the variability of tomato purees (Bureau, Vilas Boas, Giovinazzo, Jaillais, & Page, 2020; Bureau, Vilas-Boas, Giovinazzo, & Page, 2019). Moreover, MIRS coupled with FDA is a useful tool to detect authenticate the food ingredient (Gerard Downey, 1998).

3.2.3 Multivariate curve resolution-alternating least squares (MCR-ALS)

MCR-ALS (multivariate curve resolution-alternative least square) is an effective multivariate self-modelling curve resolution method developed by Tauler et al. (1995; 2005). It is widely used to simultaneously elucidate the pure spectra of different species

present in processed products and their concentration profiles, such as edible oils from different vegetable sources (Le Dréau, Dupuy, Artaud, Ollivier, & Kister, 2009). The detailed mathematical description and some applications have been described in **Paper VII**.

3.3Regression modelling

3.3.1 Partial least square regression (PLS)

PLS is one of the most widely used multivariate linear regression methods to quantify the quality of F&V. In short, PLS regression combines principal component analysis and canonical correlation analysis. PLS models maximize the covariance between Y- matrix (references datasets of F&V) and X- matrix (spectral dataset of F&V) in a way that it has better prediction for Y- matrix by maximizing the variance of X-matrix.

The latent variables (LVs) of PLS models are directions in the space of the predictors. Particularly, the maximum covariance with the selected response variable is calculated as the first latent variable, subtracting from both the original predictors and the response. The second latent variable is orthogonal to the first one, being the direction of maximum covariance between the residuals of the predictors and the residuals of the response. This approach continues for the subsequent LVs.

The optimal PLS model is chosen according to the most appropriate number of latent variables, which are determined by a proper validation strategy, determination coefficients and prediction errors. The prediction ability of PLS models is usually presented by several parameters have been described (Abdi, 2003; Nicolaï et al., 2007), such as:

- determination coefficient of calibration (R_c^2) and/or validation (R_v^2) models, which determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, the coefficient of determination tells one how well the data fits the model.

- root mean square error of calibration ($RMSE_c$) and/or validation ($RMSE_v$) sets, presents the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells you the differences between predicted values by developed model and the observed values.
- Residual Predictive Deviation (RPD), which is defined as the standard deviation of observed values divided by the Root Mean Square Error (RMSE). It indicates the precision behavior of the prediction in comparison with the average composition of all the samples.

3.3.2 Machine learning regression

Machine learning (ML) methodologies aim at learning from training data to perform a task, including classification, regression, clustering, and dimensionality reduction models (Liakos et al., 2018). Usually, each feature of an individual example can be nominal (enumeration), binary (i.e., 0 or 1), ordinal (e.g., A+ or B −), or numeric (integer, real number, etc.). The ML regression models maybe linear (Ridge regression, Lasso regression etc.) and non-linear (random forest, support vector machine, decision tree, K Nearest Neighbors, Cubist etc.).

- RF is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time (Ho, 1995). For classification tasks, the output of the RF is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. This method frequently regards as "black-box" models, as it generates reasonable predictions across a wide range of data while requiring little configuration.
- SVM has been introduced for predicting numerical property values. SVM regression models can resolve nonlinear relationships in original feature spaces through dimensionality extension (Noble, 2006). However, it has black box characters, meaning that the predictions cannot be directly interpreted in chemical terms. Hence, it is generally difficult to rationalize model performance.

More specific explanations of different ML regression methods have been described in several outstanding publications (Jordan & Mitchell, 2015; Mitchell, 1997; Shwartz & David, 2014).

III. Objectives and strategy

1. Objectives

The interfaces between production and processing domains are key points to make food supply chain more sustainable. In this thesis, our objectives were to explore new solutions to manage the variability and heterogeneity during apple processing, to improve the quality of processed products and to reduce losses and wastes. Based on the literature study, the following challenges were identified:

- A large variability and heterogeneity of raw fruits and their processed products have been pointed out (see II. Literature review, parts 1.1.3). We need to know how to deal with variability and heterogeneity in agricultural raw materials, and how to manage and optimize their processing into food products to meet consumers expectations or at least achieve consistent and controlled product quality?

- Lots of different techniques and methods were applied to characterize the variability and heterogeneity of raw fruit materials (see II. Literature review, parts 1.2 & 1.3). How to develop sufficiently efficient and reliable approaches to detect the variability and heterogeneity of raw materials in order to predict the characteristics of processed products and possibly adapt the processing conditions?

- And which strategy can facilitate and optimize the use of variable and heterogeneous raw materials to produce processed end-products that meet consumers demands and habits?

This thesis presents a proof-of-concept for investigating the infrared spectroscopy and chemometrics to address these challenges with a process analytical technology (PAT) approach in apple processing.

2. Strategy

2.1. Experiments

A first step was to obtain apples representative of the different possible scales of variability and heterogeneity. Therefore, fruits were sourced to take into account both “inter-variability” i.e. to introduce pre- and post-harvest differences between apple

batches (using different varieties, thinning practice and storage) and “intra-variability” i.e. between apples from each batch and even within apples (**Fig. 21**). These fruits were subsequently processed into purees using different heating methods and conditions.

2.1.1. Apples

The choice to work on apples responds to specific facts: i) economical interest, widely cultivated around the world and particularly in Europe; ii) consumption as both, fresh and various processed products (juices, purees, compotes, jams, etc.); iii) a good source of nutrients (dietary fibers, polysaccharides, antioxidant compounds, polyphenols etc.).

Four apple varieties i.e. ‘Golden Delicious’ (GD), ‘Granny Smith’ (GS), ‘Royal Gala’ (GA) and ‘Braeburn’ (BR) were selected in our work. They are representative of the diversity of commercial apples, showing a high ‘inter’ and ‘intra’ variability in apple quality, and are also suitable for puree processing (see Literature Review part 1).

Two factors, fruit thinning in the orchard and storage time at 4°C (from 0 to 6 months) were used to introduce more variability in the apples. The previous literature review showed that these factors can generate a large variability in chemical (SSC, TA, DMC etc.) and structural (cell numbers, firmness, crunchiness cell etc.) properties of apples, but there is no report regarding the use of infrared techniques to detect this variability, either in apples or their corresponding purees.

2.1.2. Processing conditions

Several processing conditions, including different thermal and mechanical treatments, were studied to generate a large variability in puree quality. Particularly, temperature and time of cooking lead to variations in the chemical (mainly acids degradations) and textural (mainly tissue softening) properties of the purees, while grinding and refining play a major role in breaking up the apple tissue structure into smaller cell fragments. However, there is no report providing specific insight of using infrared techniques to identify these processing factors on fruit purees.

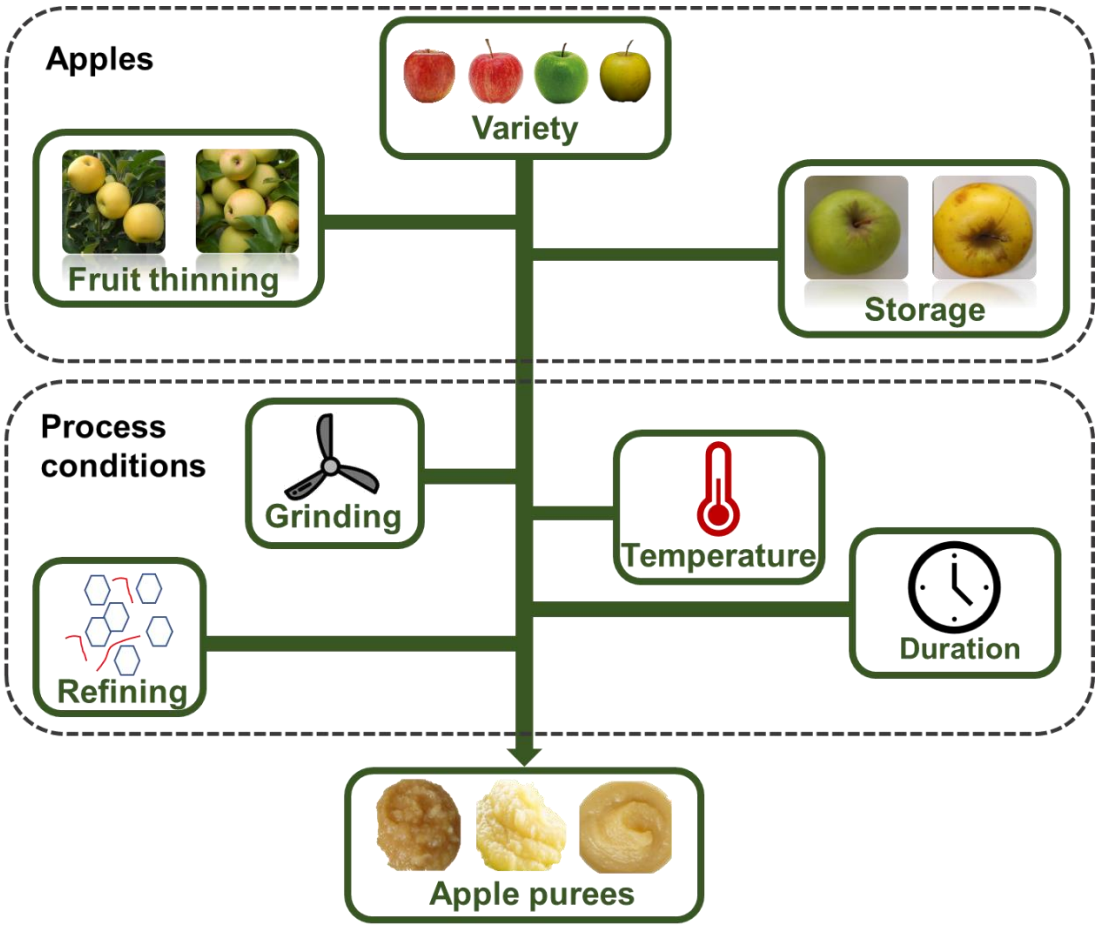

Fig. 21. Experimental factors applied to apples during pre-harvest and post-harvest periods and during processing into purees.

2.2 Experimental strategy

Our experimental strategy included trials during the past five years (**Table 19**).

Table 19. A summary of the experimental trials performed from 2016 to 2020.

Years	Experimental factors
2020 (Paper V)	• **Variety:** Golden Delicious, Granny Smith, Gala, Braeburn • **Processing**: individual apples cooked by microwave heating
2019 (Paper VI, VII)	• **Variety:** Golden Delicious, Granny Smith, Gala, Braeburn • **Processing**: 9 cooking conditions (3 temperatures x 3 grinding speeds) of Golden Delicious • **Puree formulation** by mixing 2 varieties among the 4 chosen ones
2018 (Paper I, III)	• **Variety:** Golden Delicious, Granny Smith, Gala, Braeburn • **Maturation**: thinning & non-thinning on Golden Delicious • **Storage**: with stress or without stress on Braeburn (for mealy texture) • **Processing:** 2 cooking conditions (3000 rpm at 70°C for 15 min or 400 rpm at 95°C for 17 min), 2 refining levels
2017 (Paper II, IV)	• **Variety:** Golden Delicious, Granny Smith • **Maturation**: thinning or non-thinning on Golden Delicious • **Storage**: 0, 1, 3, 6 months at 4°C • **Processing:** 1 cooking condition (90°C at 1500 rpm for 15 min), 2 refining levels
2016 (Paper IV)	• **Variety:** Golden Smoothee • **Storage**: 0, 1, 3, 6 months at 4°C • **Processing**: 1 cooking condition (1500 rpm at 90°C for 15 min), 2 refining levels

These experiments can be divided into three parts:

- **Part 1 focuses on the use of spectroscopic techniques to identify the variability and heterogeneity in raw apples and processed purees and to assess their quality properties.**
  - Comparison of NIR, MIR, Raman spectroscopies and hyperspectral imaging to identify a wide variability of purees from different apples and processing conditions (**Table 19)** and to predict their physical, chemical, and structural properties (**Paper I**).
  - Application of MIR technique on different fresh, freeze-dried (removing water interactions) and cell wall (removing soluble chemicals) materials of apples and purees to assess all puree quality properties and identify specific signals of apple processing variations (**Paper II**).
  - Development of NIR-HSI technique to reveal the heterogeneity in the contents of dry matter, total sugars, individual sugars, malic acid, and polyphenols by scanning individual apple slices of four varieties (**Table 19**), based on a simple and efficient calibration method (**Paper III**).
- **Part 2 intends to analyze the correlations between fresh and processed apples and to predict puree quality properties from spectral information of raw apples**
  - Evaluation of the correlations of quality traits and spectral signatures between batches of apples and their corresponding purees presenting a high variability (**Table 19**), in order to predict puree characteristics from the NIR spectral information of their corresponding raw and intact apples (**Paper IV**).
  - Evaluation of the correlation between apple and puree with the absolute definition of 'one apple to one puree' using microwave oven, to know the impact of 'intra-batch variability' before and after apple processing. The VIS-NIR technique was applied to each of the individual apples and their cooked purees to analyse spectral relationships before and after cooking, and to validate the possibility of predicting the quality traits of cooked purees using the VIS-NIR spectra of intact raw individual apples (**Paper V**).

- Evaluation of the correlation between apple homogenates and their processed purees using different cooking conditions (9 recipes) by MIR technique in order to explore the possibility of using MIR spectroscopy to predict quality properties of variable purees based on the prior information of raw apples and recipes (**Paper VI**).

For fruit processors, knowing predicted properties of final products after processing is not enough to produce the expected and constant final purees. To manage the food processing, the processing guidance (temperature, time, grinding speed…) must be provided and associated with the prediction models. In addition, this could enable the obtention of stable or even new food products using the spectral information coupled with advanced models from a large variability of raw materials. To date, there is no report on infrared spectroscopy to guide the fruit processing and monitor the quality of final products after several processing conditions.

- **Part 3 reports development of an innovative concept using infrared spectroscopy to guide puree formulation**

  - Application of VIS, NIR and MIR techniques on four single-variety purees and their formulated purees between two varieties to predict the texture and taste (viscosity, color, sugars and acids) of the formulated purees based on the spectra of single-variety purees using a multi-parameter optimization method.

IV. Results and discussion

Part 1. Using different spectroscopic and imaging techniques to detect variability and heterogeneity of apples and purees

The first chapter aims to answer the following questions:

- What is the most appropriate method to identify the puree variability and detect puree physical, chemical and structural properties by using different vibrational spectroscopic techniques?
- What is the most appropriate sampling method for spectroscopic analyses to evaluate efficiently the puree variability and quality?
- How to efficiently assess apple and puree variability and heterogeneity?

The results of Part 1 are presented as three papers (Papers I, II and III)

Developing highly efficient, economic and reliable process analytical approaches is a key point for apple puree quality control in industrial and scientific works. However, only few studies (see **II. Literature review Part 2.**) reported the use of spectroscopic techniques to evaluate apple puree quality. The comprehensive comparison of these techniques for chemical, structural and rheological determinations of food products remains limited. Therefore, identifying the most efficient spectroscopic method to assess quality of processed foods is a crucial point to prioritize further developments.

Vibrational spectroscopic (near-infrared, mid-infrared and Raman) and imaging techniques (near-infrared hyperspectral imaging), have been considered (**Paper I)** as candidates to integrate Process Analytical Technology (PAT) for the rapid qualification of apple purees. A comprehensive comparison of their capacity was done to:

i) Identify 72 different apple purees issued from 4 apple varieties with a thinning or non-thinning practice, different storage conditions and two different puree processing recipes;

ii) Predict the chemical (SSC, TA, DMC, individual sugars and malic acid), structural (particle sizes) and rheological properties (viscosity and viscoelasticity) of apple purees using PLS and machine learning (RF and SVM) regression models.

Paper I (Submitted)

Vibrational spectroscopy (NIR, MIR, Raman) and NIR hyperspectral imaging techniques: Which is the best way to determine chemical, structural and rheological properties of apple purees?

Weijie Lan, Vincent Baeten, Benoit Jaillais, Catherine M.G.C. Renard, Quentin Arnould, Songchao Chen, Alexandre Leca, Sylvie Bureau*.

Paper submitted to Journal of Food Engineering

1. Introduction

Apple puree is the basic ingredient of many fruit-based products, such as jams, preserves or compotes, yogurts and pie fillings for food industry (Defernez, Kemsley, & Wilson, 1995). Today, apple purees are predominantly analyzed by chromatography and specific rheometers to determine their biochemical (Keenan, Brunton, Butler, Wouters, & Gormley, 2011) and rheological properties (Buergy, Rolland-Sabaté, Leca, & Renard, 2020; Espinosa-Muñoz, Renard, Symoneaux, Biau, & Cuvelier, 2013). These methods provide accurate quantifications, but they are time-consuming, expensive and not suitable for fast and numerous characterizations.

Spectroscopic and imaging techniques have been considered to be some of the representative process analytical technologies (PAT) for the rapid qualification of agricultural commodities and processed food (Cullen, O'Donnell, & Fagan, 2014). In particular, near-infrared (NIR), mid-infrared (MIR), Raman and hyperspectral imaging (HSI) offer the advantages of a minimal sample preparation and a rapid data acquisition. However, because they are based on differences in the interaction between electromagnetic radiation and matters, it seems crucial to know: which is the most efficient spectroscopic method to assess quality of processed purees? Which are the best analytical methods based on a specific technique to provide an accurate determination of the chemical (soluble solids, titratable acidity, dry matter, individual sugars and malic acid), structural (particle sizes) and rheological properties (viscosity and viscoelasticity) of apple purees?

NIR technique has been widely applied for the safety inspection and quality assessment of apple fruits at the wavelength range from 780-2500 nm (Nicolai, et al., 2007; Pissard, et al., 2013). The broad bands of NIR contain the overlapping absorption bands corresponding mainly to overtones and combinations of vibrational mode C-H and O-H bonds of fruit components (Osborne, 2006). NIR technique can assess a diversity of quality traits in raw and processed apple products (**in Part II. 5.2**).

MIR spectroscopy on fresh and processed apples gives a good estimation of SSC, DMC, TA, malic acid and some individual sugars (**in Part II. 5.3**). Compared to the

low structural selectivity in the board bands of NIR spectra, more resolved fundamentals of MIR spectra allow to better elucidate the chemical and structural information of samples. However, the lower energy of MIR radiations and the strong water interactions in fruit suspensions prevent the sensitive evaluation of chemical compositions and structural properties.

Raman spectroscopy can provide a complementary interpretation of molecule vibration changes in polarizability, which is distinct from the vibration used in MIR by the changes in dipole moment (Pistorius, 1996). It has been successfully applied to determine several micronutrients, such as lycopene and β-carotene in tomato (Baranska, Schütze, & Schulz, 2006). For highly hydrated products, such as fresh and processed fruits, Raman presents two advantages in comparison with infrared: a weak scattering of the polar O-H group and more intense bands of homo-nuclear molecular bonds (C-C, C=C etc.). To date, no detailed study has compared the differences and limitations of Raman and infrared spectroscopy (NIR and MIR) to determine the structural and rheological properties of fruit purees.

Hyperspectral imaging (HSI) is an emerging platform technique that integrates imaging and spectroscopy to provide both spatial and spectral information (Baeten & Dardenne, 2005). Several applications of HSI were carried out on fresh fruits to estimate their external and internal quality (**in Part II. 6.1**). For fruit processed purees, no work has been done on HSI to detect their chemical and biochemical compositions, such as SSC, DMC, malic acid, individual sugars and cell wall composition. In addition, one of our interests is to investigate the possibility of using HSI to detect and even monitor the structural (particle sizes) and rheological (viscosity and viscoelasticity) properties of puree samples.

In this work, four different spectroscopic and imaging techniques, namely NIR, MIR, Raman and HSI, were applied on the same set of diverse (variety, thinning practice, fruit texture, processing) apple puree samples in order to: i) evaluate their potential to detect the puree variability; ii) compare their performance to predict chemical, structural and rheological characteristics of purees and then iii) identify

signals specific of the puree properties.

2. Materials and methods

2.1 Apple purees

2.1.1 Apples

Apples of four varieties: 'Golden Delicious' (GD), 'Granny Smith' (GS), 'Royal Gala' (GA) and 'Braeburn' (BR) were harvested at a commercial maturity in 2018 from the La Pugère experimental orchard (Chambre d'Agriculture des Bouches du Rhône) (Mallemort, Bouches du Rhône, France). GS, GA , BR and half of GD apples was grown under a standard chemical fruit thinning practice (Th+) with 50-100 fruits / tree. The other half of GD apples was non thinned (Th-) with 150-200 fruits/ tree.

After harvesting, five apple groups (GD Th+, GD Th-, GS, BR, GA) were stored at 4 ℃ in normal atmosphere to ensure starch regression (customised phytotron, Froid et Mesures, Beaucouzé, France). Half of the Braeburn apples (BM) were stored for 11 days at 23 ℃ and around 90% relative humidity. These two different stress treatments on Braeburn apples resulted a significant change ($p < 0.001$) of texture ($p < 0.001$), characterizing by the linear distances of puncture tests (13.96 ± 1.23 of BR crunchy apples and 11.69 ± 0.72 of BM mealy apples) based on the same method from **Paper IV**. The BM apples were clearly with a higher mealy texture than the BR apples. Totally, six apple groups (GD Th-, GD Th+, GS, GA, BR and BM) were used for puree processing (**Fig. 22**).

2.1.2 Purees processing

For all apple groups, three replicates of apples purees were processed from 3 kg of apples each. After sorting and washing, apples (3 kg) were cored, and sliced into 12 portions, then processed under vacuum by a multi-functional processing system (RoboQbo Qb8-3, Bentivoglio, Italy), following two different processing recipes: I) ground at 3000 rpm for 202 s during the increase of temperature and heated at 70 ℃ for 15 min, then pasteurized at 95 ℃ for 2 min; II) ground at 3000 rpm for 360 s during

the temperature increase step, followed by 400 rpm at 95 °C for 17 min. Afterwards, half of each processed puree was refined at 0.5 mm using a Robot Coupe C80 automatic refiner (Robot Coupe C80, Robot Coupe SNC, Vincennes, France) and the other was not refined. Finally, all processed apple purees were conditioned in hermetically sealed cans, then cooled at 23 °C before the measurements performed the day after. In total, 72 puree samples (6 apple groups × 2 processing recipes × 2 refining levels × 3 processing replicates) were obtained (**Fig. 22**).

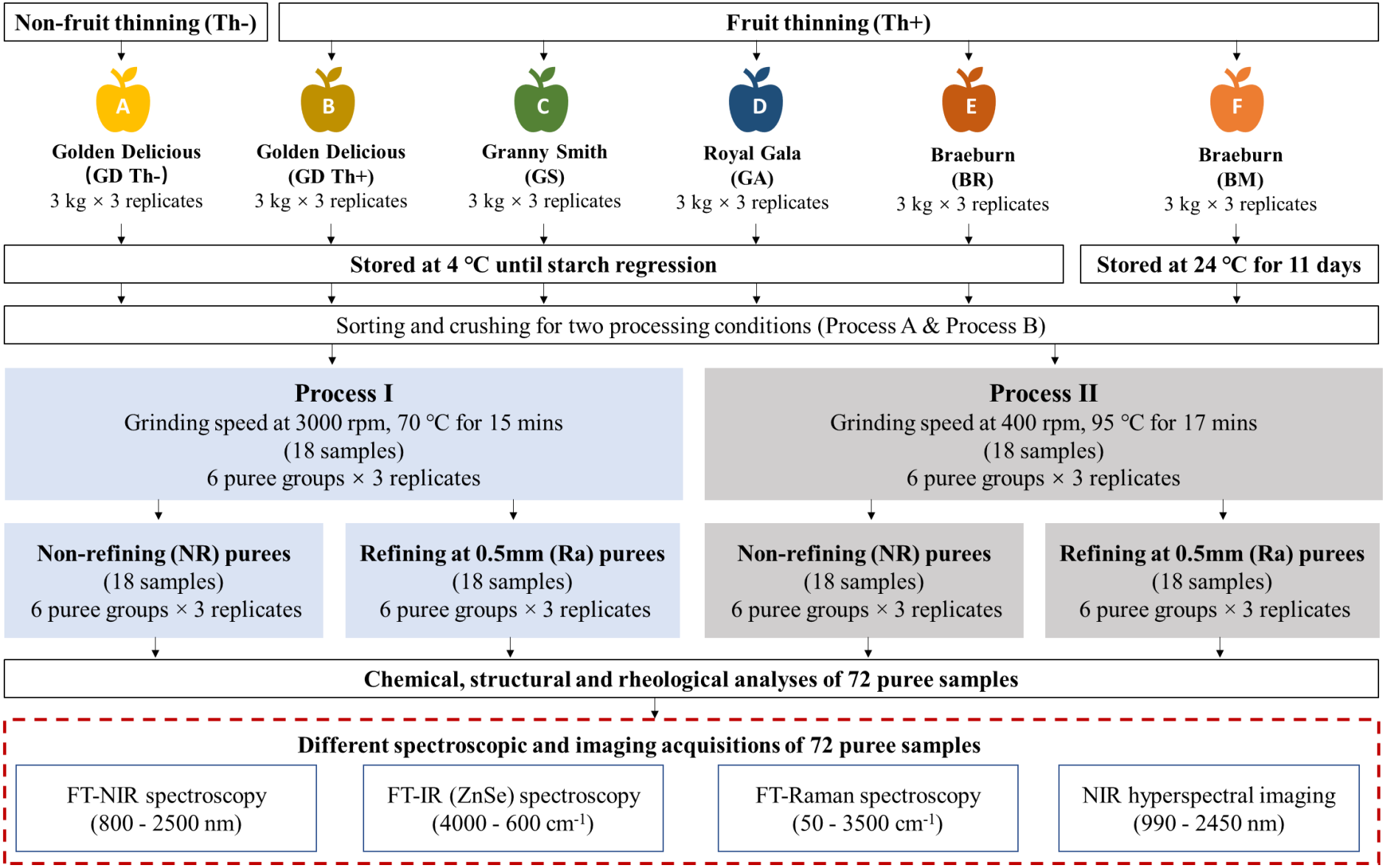

Fig. 22. Experimental scheme of apple puree processing, quality characterization and spectral acquisition**.**

2.2 Determination of quality traits

2.2.1 Rheological and structural analyses

The puree rheological measurements, consisting in rotational (flow curve) and oscillatory (amplitude sweep) tests, were carried out using a Physica MCR-301 controlled stress rheometer (Anton Paar, Graz, Austria)) equipped with a vane measuring system with a 3.46 mm gap (CC27/S cup and FL100/6W bob, Anton Paar), at 22.5 °C. The flow curves were performed after a pre-shearing period of 1 minute at 50 s^{-1} followed by 5 minutes at rest. The viscosity was then measured at a controlled shear rate range of [10; 250] s^{-1} on a logarithmic ramp, at a rate of 1 point every 15 seconds. The complete flow curves were fitted with a power law model, as described by Eq. (1).

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

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

Amplitude sweep tests were performed at an angular frequency of 10 $rad.s^{-1}$ in the deformation range of [0.01; 100] %, in order to determine the linear viscoelastic range of the purees and the yield stress, defined as the crossing point between the storage modulus (G') and the loss modulus (G'') curves. The damping factor $\tan\delta = G''/G'$ of purees was calculated.

The particle sizes were measured according to our **Paper II**. Puree samples were diluted in distilled water to separate particles and stained with calcofluor white at 0.1 g/L and highlighted with a 365 nm UV lamp . A high-resolution digital video camera (Baumer VCXU31C, Baumer SAS, Fillinges, France) with a macro lens (VSTech 0513, VS Technology Corporation, Tokyo, Japan.) was used to visualize the distribution and dispersion of puree particles. The particle sizes averaged over volume d(4:3) (de Brouckere mean) and over surface area d(3:2) (Sauter mean) were measured with a laser granulometer (Mastersizer 2000, Malvern Instruments, Malvern, UK).

2.2.2 Biochemical analyses

SSC was determined with a digital refractometer (PR-101 ATAGO, Norfolk, VA, USA) and expressed in °Brix at 23 °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). Individual sugars and malic acid were quantified using colorimetric enzymatic kits, according to the manufacturer's instructions (R-biopharm, Darmstadt, Germany). The content of glucose, fructose, sucrose and malic acid were expressed in g/kg of FW. These measurements were performed with a SAFAS flx-Xenius XM spectrofluorimeter (SAFAS, Monaco) at 570 nm for sugars and 450 nm for malic acid. DMC was estimated from the weight of freeze-dried samples upon reaching a constant weight (freeze-drying for 5 days). Cell wall materials (AIS) of purees were isolated using the alcohol insoluble solids method proposed by Renard (2005) and the cell wall contents (AIS contents) were expressed in FW.

2.3 Spectral and image data acquisition

2.3.1 NIR spectroscopy

NIR spectra were collected with a multi-purpose analyzer (MPA) spectrometer (Bruker Optics®, Wissembourg, France) at 23°C. Puree samples were transferred into 10 mL glass vials (5 cm height x 18 mm diameter) which were placed on the automated sample wheel of the spectrophotometer. Logarithmic transformed reflectance spectra (log (1/R)) were acquired with a spectral resolution of 8 cm^{-1} from 12500 to 4000 cm^{-1} (corresponding to wavelengths from 800 to 2500 nm). Each spectrum corresponded to the average of 32 scans. The spectral acquisition and instrument adjustments were controlled by OPUS software Version 5.0 (Bruker Optics®). A reference background measurement was automatically acquired before each data set acquisition using an internal Spectralon reference. Each puree sample was measured randomly three times on different aliquots. The mean of three replicate scans for each sample was used in subsequent chemometric analysis.

2.3.2 MIR spectroscopy

MIR spectra of purees were acquired at 23°C using a Tensor 27 FTIR spectrometer (Bruker Optics®, Wissembourg, France) equipped with a horizontal attenuated total reflectance (ATR) sampling accessory and a deuterated triglycine sulphate (DTGS) detector. The purees were placed at the surface of a zinc selenide (ATR-ZnSe) crystal with six internal reflections. Spectra with 32 scans each were collected from 4000 cm^{-1} to 800 cm^{-1} with a 4 cm^{-1} resolution and were corrected against the background spectrum of air. Three replications of spectral measurement were performed randomly on each puree, and the average for each sample was used for further analysis.

2.3.3 Raman spectroscopy

Raman spectra were acquired on a Confocal Raman Microscope Senterra II spectrometer obtained from (Bruker Optics, Ettlingen, Germany) with a 785 nm diode laser and a thermoelectrically cooled CCD detector, operating at -65 °C. For spectra collection, each puree sample was manually placed and compacted in 36 holes (those in the middle) of a 96 well aluminium plate (12 × 8) with an inner diameter of 6 mm each. After removing the water of purees by evaporation at the ambient temperature (~20 ℃), spectra were accumulated with a bleaching of 20 s, an integration time of 2 s and 7 coadditions using a 100 mW laser. Raman intensity were recorded from 50 to 3650 cm^{-1} with a spectral resolution of 4 cm^{-1} intervals. OPUS 7.8 Software (Bruker Optics, Ettlingen, Germany) was used for spectral data acquisition. Each sample was independently and randomly scanned six times. The main spectrum of 6 replicates for each puree sample was used for further analysis.

2.3.4 HSI acquisition

The hyperspectral images of apple purees were acquired on a pushbroom (a line-scanning type) near infrared hyperspectral imaging system (SPECIM, Oulu, Finland), which consisted of a SWIR camera (SWIR-CL-400-N25E, SPECIM) covering the spectral range of 900-2500 nm with a spectral resolution of about 12 nm, an OLES 56 camera lens (SPECIM), an illumination source (halogen lamps) and a translating scanner. Before measurements, the reflectance calibration was performed as the same as **Paper III**. All the image acquisition parameters (the exposure time of camera, the

scanning speed etc.) were controlled by the LUMO® software from SPECIM. Each puree sample was placed on a hole (with an inner diameter of 3 cm) of the standard white plate (nine holes totally). All images were acquired in the reflectance mode and the final image size for each kernel is 387 × 127 × 288, the two first values representing pixel dimensions in the x and y directions (field of view of 9.5 × 3.1 cm, with a spatial resolution of 245 μm) and the third value accounting for the number of spectral channels. As the beginning and ending wavelengths contained noise caused by the instrument itself, the 258 bands from 990 to 2450 nm were selected for further spectral analysis. The averaged HSI spectrum of each puree sample was calculated for further discrimination and regression analyses.

2.4 Statistical analyses of references data

After checking for normal distribution with a Shapiro-Wilk test (α=0.05), the reference data of processed purees were presented as mean values and the data dispersion within our experimental dataset expressed as standard deviation values (SD) (**Table 20**). Analysis of variance (ANOVA) was carried out to determine the significant differences due to the different apple varieties, process recipes and mechanical refining treatments (**Table 20**) using XLSTAT (version 2018.5.52037, Addinsoft SARL, Paris, France) data analysis toolbox. Principal component analysis (PCA) was carried out on all reference data of processed purees to evaluate their discriminant contributions using Matlab 7.5 software using the SAISIR package (Cordella & Bertrand, 2014).

2.5 Chemometric analysis

NIR, MIR, Raman and HSI spectra were pre-processed with Matlab 7.5 software using the SAISIR package (Cordella & Bertrand, 2014). The discriminant analysis and multivariate regression were performed with several packages of the R software (version 4.0.2) (R Core Team, 2019). After several pretests, the standard normal variate (SNV) with smoothing (a window size of 13 variables) transformed NIR data in 800-2500 nm; the SNV pre-processed MIR spectra in 1800-900 cm^{-1}; the SNV with smoothing (a window size of 13 variables) of Raman in 1800-800 cm^{-1} and the SNV

with 3 windows (a window size of 3 variables) smoothed HSI data in 990-2450 nm had the best performances to classify and assess the puree quality.

Partial least square (PLS), support vector machine (SVM) and random forest (RF) algorithms were used to discriminate purees (**Part 3.2**) and predict their quality traits (**Part 3.3**). The 10-fold full cross-validation was applied to all spectral datasets (NIR, MIR, Raman and HSI). For discrimination models (PLS-DA, SVM-DA and RF-DA), the discrimination accuracy (acc), the optimal numbers of latent variables (LVs) of PLS-DA models (**Table 21**) and the most attributed vibrational bands (**Table 22**) were used to describe the discriminating ability of the different spectroscopic techniques. For regression models (PLS-R, SVM-R and RF-R), the prediction performances were assessed by the determination coefficient of cross-validation (R_{cv}^2), the root mean square error of cross-validation ($RMSE_{cv}$) and the residual predictive deviation (RPD) value as described by Nicolai (Nicolai, et al., 2007). The most correlated spectral signals of the best developed models with RPD values over 2.0, indicating the possibility for approximate qualitative predictions (Nicolai, et al., 2007), were described in **Table 23** and **Table 24**.

3. Results and discussion

3.1 Characteristics of apple purees

After puree processing, the different apple varieties provided a large variability of chemical, textural and rheological properties (**Table 20**). According to PCA results taking into account all the characterized parameters (**Fig. 23**), 'Granny Smith' (GS) purees (C) were clearly discriminated from the other puree groups along the first principal component (PC1). The GS purees presented a significantly ($p < 0.001$) higher viscosity (K and n) and elasticity (yield stress, G' and G''), particle size d(4:3) and volume d (3:2), TA, malic acid and AIS content than the others (**Fig. 23b and Table 20**). Remarkable higher values ($p < 0.001$) for SSC and DMC allowed the separation of 'Golden Delicious' (A and B) and 'Royal Gala' purees (D) along the second principal component (**Fig. 23a, Fig. 23b**). Thinning practice (Th+) on GD apples (B) resulted in a less viscous purees than non-thinned GD purees (A) (**Table 20**), which is in line with

the results from **Paper IV**. For all non-refined (NR) purees, 'Royal Gala' had the lowest viscoelastic moduli (G' < 934.0 ± 35.4 Pa, G''< 194.3 ± 7.2 Pa), titratable acidity (TA < 3.8 ± 0.2 meq/kg) and cell wall contents (AIS < 128.4 ± 9.5 mg/g). However, the overlapping of the two kinds of 'Braeburn' purees (E and F) (**Fig. 23a**) revealed the difficulty to produce different purees after processing and refining of both, crunchy (the puncture linear distance of 13.96 ± 1.23) and mealy (the puncture linear distance of 11.69 ± 0.72) apples.

The two different processing recipes used here (Processes I and II) led to significant ($p < 0.01$) changes of puree rheological behaviors (*K, n*, G', G'', yield stress and tan δ) and particle distributions (d4:3 and d3:2), but not for SSC, DMC and AIS ($p > 0.05$) (**Table 20**). Particularly, purees processed at 95 °C and 400 rpm (Process II) had a soft solid-like behavior. They were more viscous (*K* and *n*) with higher G' and G'' and larger particles (d4:3 and d3:2) than the purees processed at 70 °C and 3000 rpm (Process I).

Moreover, as expected, the refining treatment generated a significant ($p < 0.01$) decrease of puree viscosity and elasticity (*K*, *n*, G', G'' and yield point), particle sizes (d4:3 and d3:2) and cell wall contents, but did not impact (p > 0.05) chemical attributes.

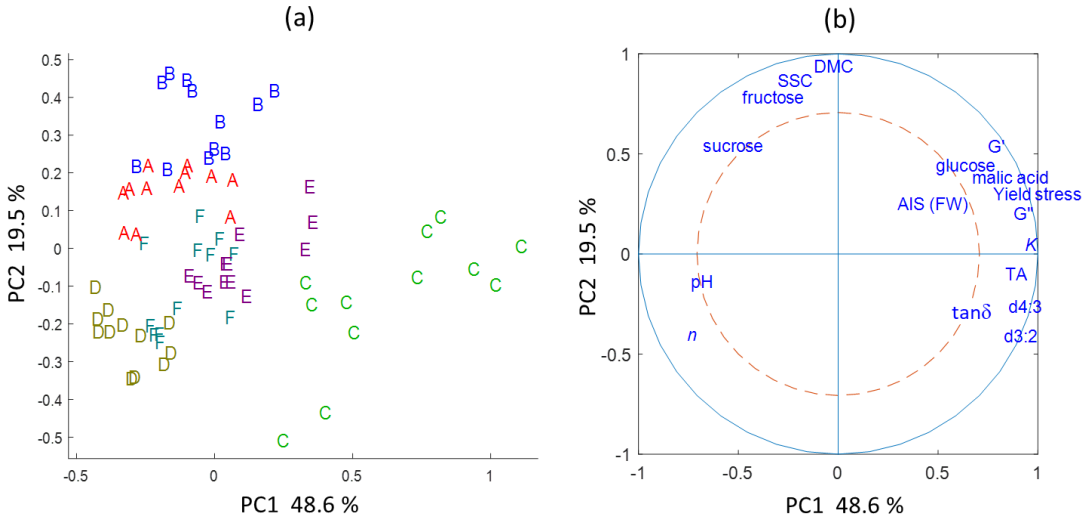

Fig. 23. Principal component analysis on chemical, structural and rheological parameters of six puree groups (A: GD Th-; B: GD Th+; C: GS; D: GA; E: BR; F:BM): **(a)** the scores plot of the two first components (PC1 and PC2); **(b)** the correlation plot of the PC1 and PC2.

Table 20. Chemical, structural, and rheological characteristics of studied apple purees.

Groups	Process	Refining	Viscosity		G'	G"	Yield stress	tan δ	d4:3	d3:2	SSC	DMC	pH	TA	malic acid	fructose	sucrose	glucose	AIS
			K	*n*	Pa	Pa		-	-	-	(°Brix)	(g/g)		(meq/kg)	(g/kg)	(g/kg)	(g/kg)	(g/kg)	mg/g
GD Th-	I	NR	15.4	0.24	1245.9	247.4	12.8	0.20	267.4	174.5	14.4	0.21	3.8	5.4	4.9	67.7	73.3	17.5	138.3
		Ra	14.6	0.24	1158.8	222.5	12.6	0.19	264.5	172.4	14.1	0.20	3.7	5.5	4.6	67.2	66.4	17.1	117.8
	II	NR	23.1	0.21	1601.3	341.6	20.1	0.21	397.0	230.7	14.4	0.21	3.8	4.9	4.0	60.5	55.2	17.0	141.8
		Ra	20.1	0.21	1351.1	273.9	17.4	0.20	384.0	226.2	14.1	0.20	3.8	5.0	4.6	70.6	65.3	17.5	115.5
GD Th+	I	NR	11.4	0.27	984.4	190.7	9.7	0.19	262.9	177.4	15.3	0.22	3.7	5.9	5.5	78.7	73.0	16.3	145.2
		Ra	10.4	0.27	922.1	172.3	9.4	0.19	256.5	174.4	14.8	0.22	3.7	6.2	5.7	81.9	72.3	16.2	119.4
	II	NR	20.0	0.22	1390.7	290.9	16.9	0.21	382.1	228.5	14.7	0.21	3.7	5.5	5.1	76.2	60.8	17.5	141.1
		Ra	17.8	0.22	1200.4	241.5	15.0	0.20	371.4	223.8	14.5	0.21	3.7	5.7	5.2	71.6	63.1	17.6	118.3
GS	I	NR	32.0	0.21	1835.5	385.9	25.4	0.21	598.5	314.2	12.3	0.19	3.4	10.7	7.2	44.1	29.7	20.5	182.5
		Ra	22.3	0.22	1131.7	227.5	15.6	0.20	545.1	287.4	11.7	0.19	3.4	10.6	6.7	41.6	28.9	19.2	147.8
	II	NR	44.8	0.20	1794.5	543.2	25.8	0.30	774.4	399.5	12.2	0.19	3.4	10.4	6.4	42.8	25.5	19.9	169.7
		Ra	23.8	0.22	944.1	280.0	14.4	0.30	488.2	256.4	12.2	0.18	3.4	10.4	4.7	25.9	13.9	13.4	145.5
GA	I	NR	7.3	0.33	720.2	137.8	7.6	0.19	383.1	226.6	12.6	0.19	4.0	3.8	4.1	60.7	72.3	13.5	128.4
		Ra	7.1	0.32	675.6	125.6	7.4	0.19	372.6	223.0	12.4	0.19	4.1	3.9	4.2	65.8	71.6	13.6	119.7
	II	NR	12.4	0.27	934.0	194.3	10.5	0.21	440.3	261.1	12.5	0.18	4.3	3.7	3.6	53.8	57.0	12.5	124.5
		Ra	11.3	0.27	810.1	160.6	9.7	0.20	431.2	256.8	12.2	0.18	4.3	3.7	3.5	47.8	50.8	11.7	122.7
BR	I	NR	11.2	0.28	1080.3	215.7	12.5	0.20	421.7	227.5	12.9	0.19	3.6	6.7	5.6	54.7	43.0	17.1	156.1
		Ra	9.8	0.29	987.8	192.6	11.6	0.20	412.5	223.8	12.7	0.19	3.6	7.8	5.7	61.4	39.5	18.2	132.8
	II	NR	22.6	0.23	1508.8	323.3	19.8	0.21	537.8	283.5	13.3	0.20	3.5	6.7	5.7	61.9	37.4	18.7	154.5
		Ra	15.9	0.24	1054.8	210.1	13.7	0.20	499.8	267.8	13.1	0.19	3.5	6.9	5.7	61.8	41.2	18.4	122.7
BM	I	NR	8.0	0.29	965.1	200.2	7.7	0.21	241.7	172.1	12.6	0.19	3.7	5.8	4.1	51.5	36.0	16.5	145.4
		Ra	8.0	0.28	957.8	195.8	8.0	0.20	240.1	170.7	12.4	0.19	3.7	5.7	4.4	55.6	38.0	17.9	125.7
	II	NR	13.9	0.24	1373.1	309.9	12.3	0.23	292.6	212.1	13.2	0.18	3.8	5.5	4.7	59.0	39.4	21.1	142.7
		Ra	13.7	0.23	1288.3	278.7	12.0	0.22	286.7	199.8	12.7	0.18	3.7	5.5	4.9	69.2	39.8	20.8	117.8
SD			*8.7*	*0.04*	*321.7*	*92.5*	*5.2*	*0.03*	*129.7*	*53.2*	*1.1*	*0.01*	*0.2*	*2.2*	*1.0*	*13.6*	*17.8*	*2.9*	*18.3*
F-value and significance		***Variety***	*192.0*	*120.5*	*50.9*	*73.3*	*74.9*	*1071.5*	*394.5*	*386.2*	*117.1*	*58.8*	*1285.8*	*215.0*	*43.4*	*154.9*	*218.1*	*30.4*	*2.8*
			***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	*
		Process	*110.1*	*98.6*	*13.3*	*52.1*	*35.4*	*1609.1*	*218.6*	*303.4*	*0.02*	*0.4*	*47.2*	*52.6*	*16.1*	*21.4*	*54.9*	*1.8*	*1.6*
			***	***	**	***	***	***	***	***	*ns*	*ns*	***	***	***	***	***	*ns*	*ns*

Refining	*70.7*	*4.3*	*66.8*	*77.2*	*54.3*	*107.3*	*82.5*	*114.1*	*5.8*	*1.2*	*5.7*	*1.7*	*2.4*	*2.3*	*2.5*	*2.2*	*120.9*
	***	*	***	***	***	***	***	***	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	***

Note: GD Th-: non-thinned Golden Delicious; GD Th+: thinned Golden Delicious; GS: Granny Smith; GA: Royal Gala; BR: crunchy Braeburn, stored at 4°C; BM: mealy Braeburn, stored at 24 °C. G', G'': storage and loss modulus, at an angular frequency of 10 rad/s; AIS: Alcohol insoluble solids. Data expressed in Fresh weight (FW) values correspond to the mean of 3 lots x 10 apples. Two processing strategies: Process I of 70 °C, 3000 rpm and Process II of 95 °C, 400 rpm. Processed purees with non-refining (NR) or refined at 0.5 mm. In grey, ANOVA results of puree variety, process and refining conditions. ns, *, **, ***: Non-significant or significant at $P < 0.05$, 0.01, 0.001 respectively.

3.2 Discrimination of variability of apple purees

The ability of the four different techniques coupled with PLS-DA, SVM-DA and RF-DA were compared to classify different factors: (a) varieties (48 purees from (Th+) GD, GA, GS and BR apples), (b) process recipes (72 samples of process I and II), (c) refining treatments (72 NR and Ra), (d) fruit thinning practices (24 GD purees from Th+ and Th-) and (e) fruit stress treatments (24 Braeburn purees with crunchy BR and mealy BM) (**Table 21**). The main vibrational bands observed in NIR, MIR, Raman and HSI datasets, which contributed to the best discrimination models are shown on purees for all factors (a-e) (**Table 22)**.

NIR technique coupled with PLS-DA models gave a correct discrimination of the four puree varieties (acc = 88.8%, 4 LVs), the two GD fruit thinning purees (acc = 86.7%, 2 LVs) and the two Braeburn purees (acc = 95.8%, 3 LVs). The specific NIR spectral regions at 818-850, 1849, 1880 and 2145-2155 nm mainly contributed to puree variety discrimination (**Table 22**). Particularly, the spectral area at 800-1000 nm, which is known as the absorption of apple carbohydrates and water variations (Giovanelli, Sinelli, Beghi, Guidetti, & Casiraghi, 2014; Zude, Herold, Roger, Bellon-Maurel, & Landahl, 2006), has been used for the apple variety classification (Bobelyn, et al., 2010). The absorption bands around 1880 nm are explained by the O-H combinations of water contents in apples (Camps, Guillermin, Mauget, & Bertrand, 2017). The broad band at 2100-2200 nm corresponds to the first combination band of C-H bond of sugars and acids, and has also been highlighted in **Paper IV**. Besides, the wavelengths around 1400 nm (1345, 1392 and 1379-1384 nm), related to the soluble solids variations in apple juices (Kaur, Künnemeyer, & McGlone, 2020), were one of the major contributors for the discriminations of apple thinning (Th+ and Th-) and stress treatments (crunchy BR and mealy BM). However, NIR technique was not able to well classify (acc < 55.6%) the processing recipe and refining, which nevertheless induced intensive structural and rheological variations of purees (**Table 20 and Table 22**).

MIR technique provided a better discrimination of all studied factors than NIR (**Table 21**). Particularly, three different discrimination models (PLS-DA, SVM-DA,

RF-DA) allowed to classify perfectly the four puree varieties. The specific spectral wavenumbers at 1723- 1718, 1107, 1061 and 1022 cm^{-1} (**Table 22**), attributed to the stretching bonds of C=O of malic acid, and the C-O and C-C of glucose, fructose and sucrose (Bureau, Cozzolino, & Clark, 2019), were consistent with the measured differences of purees coming from varieties (**Fig. 23 and Table 21**). Compared to NIR results, the satisfactory classifications by MIR of processing recipe (acc = 100 %) and refining (acc = 91.7%) were mainly based on the overlapped region between 1750 and 1650 cm^{-1} (1749 cm^{-1}, 1730-1715 cm^{-1} and 1640-1628 cm^{-1} in **Table 22**), related to the organic acids, soluble polysaccharides, pectins, phenolics and absorbed water in purees (described in detailed in **Paper II**). MIR was able to highlight the physicochemical modifications of apple purees generated by different processing strategies (heating temperature and grinding speed) and mechanical refining treatments. Besides the aforementioned spectral signals, the excellent PLS discriminations of apple thinning (acc = 100%) and stress treatments (acc = 100%) were linked to three specific wavenumbers at 1084, 1056 and 998 cm^{-1}, corresponding to the variations of glucose and sucrose in fruits (Bureau, et al., 2019).

For Raman spectroscopy, PLS-DA models developed over the range of 800-1800 cm^{-1} had a lower discrimination accuracy and more LVs to discriminate puree varieties (acc = 81.3%, 7 LVs), thinning practices (acc = 75.0%, 6 LVs) and stress treatments (acc = 70.8%, 6 LVs) than the models obtained with NIR and MIR techniques (**Table 22**). The main vibrational bands responsible for these discriminations were related to the variations of major sugars and acids in apple purees, which have been highlighted in honey products (Pompeu, et al., 2018) and soft drinks (Ilaslan, Boyaci, & Topcu, 2015). In particular, the C-C stretching and C-H deformation vibration of glucose at 840-842 cm^{-1} (Özbalci, Boyaci, Topcu, Kadılar, & Tamer, 2013); the stretching of C-O-C at 872 cm^{-1} and the deformation of C-OH of fructose at 872, 939, 944 and 1054 cm^{-1} (Cerchiaro, Sant'Ana, Temperini, & da Costa Ferreira, 2005; Mathlouthi & Luu, 1980; Özbalci, et al., 2013); the C-O and C-OH vibration of sucrose at 1126 cm^{-1} (Ilaslan, et al., 2015; Pierna, Abbas, Dardenne, & Baeten, 2011) and the C=O stretching of malic acid at 1734 cm^{-1} (Barańska, Kuduk-Jaworska, Szostak, & Romaniewska,

2003). Interestingly, Raman technique showed a good discrimination of puree processing conditions (acc = 82.3%, 8 LVs). Besides the aforementioned wavenumbers, the specific Raman bands at 845 cm^{-1} and 1433-1436 cm^{-1} were observed to discriminate puree processing changes, presenting the C-O-C and COO- antisymmetric stretching of pectins during the clarification of apple juice (Camerlingo, et al., 2007).

HSI technique coupled with PLS-DA showed a relatively higher discrimination accuracy of puree variety (acc = 100%, 7 LVs), processing recipes (acc = 86.1%, 10 LVs), fruit thinning (acc = 91.6%, 6 LVs) and stress treatments (acc = 100%, 4 LVs) than the conventional NIR spectroscopy, but using a higher number of latent variables (LVs). Besides the similar aforementioned wavenumber regions as in NIR around 1400, 1880 and 2100-2300 nm, the specific spectral areas at 1048-1088 nm and 1106-1145 nm were observed, corresponding to the SSC and DMC variations in fruits (Wang, Peng, Xie, Bao, & He, 2015). Comparing to NIR, PLS-DA on the averaged HSI puree spectra gave an impressive improvement of the discrimination of puree process recipes, from 51.4% to 86.1%. However, both NIR and HSI spectra had a limited ability to discriminate the different refining levels (< 58.3% correct identification). These results indicated these two techniques had the potential to detect puree variability (varieties, fruit thinning, process) involving significant differences in composition (**Table 20**), but not to estimate puree textural changes (refining) (**Table 20**).

Generally, PLS-DA models developed using NIR, MIR, Raman and HSI spectra of purees had the best performances to discriminate the variability of varieties (a), processes (b), fruit thinning (d) and stress treatments (e) (**Table 21**). However, specifically for refined purees, RF-DA gave a better discrimination of purees than SVM-DA and PLS-DA.

Table 21. Discrimination using 10-fold full cross-validation PLS-DA, SVM-DA and RF-DA models of apple purees according to **(a)** varieties, **(b)** processes, **(c)** refining levels, **(d)** fruit thinning practices of Golden Delicious apples, **(e)** stress treatments of Braeburn apples, using NIR, MIR, Raman and HSI data.

Spectral techniques	**NIR**			**MIR**			**Raman**			**HSI**		
Spectral ranges	800- 2500 nm			900- 1800 cm^{-1}			800- 1800 cm^{-1}			990-2450 nm		
Models	PLS-DA	SVM-DA	RF-DA	PLS-DA	SVM-DA	RF-DA	PLS-DA	SVM-DA	RF-DA	PLS-DA	SVM-DA	RF-DA
(a) Variety (GD/GS/BR/GA)												
No. of samples	48	48	48	48	48	48	48	48	48	48	48	48
Correct discrimination rate	88.8 %	81.25 %	84.6 %	100.0 %	100.0 %	100.0 %	81.3 %	50.0 %	60.4 %	100 %	72.9 %	72.9 %.
LVs	4	-	-	3	-	-	7	-	-	7		
(b) Process (I/ II)												
No. of samples	72	72	72	72	72	72	72	72	72	72	72	72
Correct discrimination rate	51.4 %	31.9 %	44.4 %	100 %	97.2 %	93.1 %	82.3 %	67.7 %	67.7 %	86.1 %	41.7 %	47.2 %
LVs	4	-	-	5	-	-	8	-	-	10		
(c) Refining levels (NR/ Ra)												
No. of samples	72	72	72	72	72	72	72	72	72	72	72	72
Correct discrimination rate	51.4 %	38.9 %	55.6 %	84.7 %	90.3 %	91.7 %	56.9 %	40.3 %	45.8 %	55.1 %	51.4 %	58.3 %
LVs	5	-	-	4	-	-	7	-	-	6		
(d) Fruit thinning (Th+/ Th-)												
No. of samples	24	24	24	24	24	24	24	24	24	24	24	24
Correct discrimination rate	86.7 %	53.3 %	82.5 %	100.0 %	100.0 %	100.0 %	75.0 %	16.7 %	45.8 %	91.6 %	79.2 %	87.5 %
LVs	3	-	-	3	-	-	6	-	-	6		
(e) stress treatments (BR/ BM)												
No. of samples	24	24	24	24	24	24	24	24	24	24	24	24
Correct discrimination rate	95.8 %	63.3 %	87.5 %	100.0 %	100.0 %	100.0 %	70.8 %	25.0 %	54.2 %	100.0 %	58.3 %	87.5 %
LVs	3	-	-	3	-	-	6	-	-	4		

Note: NIR spectra: SNV pre-treated and smoothed (13 windows) at 800-2500 nm; MIR spectra: the SNV pre-treated at 1800-900 cm^{-1}; Raman spectra: the SNV pre-treated and smoothed (13 windows) at 1800-800 cm^{-1}; HSI spectra: the SNV pre-treated and smoothed (3 windows) at 990-2450 nm. 'Variety': (four varieties of 'Golden Delicious', 'Braeburn', 'Granny Smith' and 'Royal Gala'); 'fruit thinning': different fruit thinning practices for Golden Delicious apples (50 - 100 fruits/ tree or 150-200 fruits/ tree); 'stress': two different textures of Braeburn apples (11 days at 24 ℃ or 2 months at 4 ℃); 'processing': two processing recipes (70 ℃ for 15 mins with 3000 rpm grinding or 95 ℃ for 17 mins with 400 rpm grinding); 'refining': two refining conditions after puree processing (refined at 0.5 mm or not refined).

Table 22. The main attributions for vibrational bands of the best overall discrimination models developed for puree samples.

Spectra	Spectral ranges	Factors	No. samples	Model	LVs	acc (%)	Key frequencies NIR (nm), MIR (cm^{-1}), Raman (cm^{-1}), HSI (nm)
NIR	800-2500 nm	variety	48	PLS-DA	5	88.8	818-850, 1849, 1880, 2145-2155
		process	72	PLS-DA	4	51.4	/
		refining	72	RF-DA	-	55.6	/
		fruit thinning	24	PLS-DA	2	86.7	904, 1392, 1864
		stress	24	PLS-DA	3	95.8	913, 1345, 1379-1384
MIR	1800- 900 cm^{-1}	variety	48	PLS-DA	4	100.0	1723-1718, 1107, 1061, 1022
		process	72	PLS-DA	5	100.0	1730-1715, 1640-1628, 1138, 1084, 1001-998
		refining	72	RF-DA	-	91.7	1749, 1636, 1061, 1018, 995
		fruit thinning	24	PLS-DA	3	100.0	1772, 1593, 1084, 1022, 998
		stress	24	PLS-DA	3	100.0	1658-1608, 1056, 1018, 1001
Raman	800-1800 cm^{-1}	variety	48	PLS-DA	7	81.3	842, 873, 1064, 1126, 1266, 1433, 1610
		process	72	PLS-DA	8	82..3	816-818, 845, 939, 972, 1362-1367, 1433-1436, 1734
		refining	72	PLS-DA	7	56.9	/
		fruit thinning	24	PLS-DA	6	75.0	842, 1054, 1077, 1427, 1608, 1675
		stress	24	PLS-DA	6	70.8	840, 904, 944, 1059-1063, 1334, 1734
HSI	990-2450 nm	variety	48	PLS-DA	7	100.0	1106-1145, 1259, 1338, 1406, 1869-1874, 1931-1964
		process	72	PLS-DA	10	86.1	1048-1088, 1191, 1242, 2117, 2274-2387, 2437
		refining	72	RF-DA	/	58.3	/
		fruit thinning	24	PLS-DA	6	91.6	1065-1088, 1338-1367, 2145, 2331-2342, 2376-2398, 2426
		stress	24	PLS-DA	4	100.0	1048, 1134, 1389, 1947, 2409

Note: acc: discrimination accuracy; PLS-DA: partial least square discrimination; RF-DA: random forest discrimination. NIR spectra: SNV pre-treated and smoothed (13 windows) at 800-2500 nm; MIR spectra: the SNV pre-treated at 1800-900 cm^{-1}; Raman spectra: the SNV pre-treated and smoothed (13 windows) at 1800-800 cm^{-1}; HSI spectra: the SNV pre-treated and smoothed (3 windows) at 990-2450 nm. 'Variety': four apple varieties of 'Golden Delicious', 'Braeburn', 'Granny Smith' and 'Royal Gala' ; 'fruit thinning': different fruit thinning practices for Golden Delicious apples (50 - 100 fruits/ tree or 150-200 fruits/ tree); 'stress': two stress treatments of Braeburn apples (11 days at 24 ℃ or 2 months at 4 ℃); 'processing': two processing recipes (70 ℃ for 15 mins with 3000 rpm grinding or 95 ℃ for 17 mins with 400 rpm grinding); 'refining': two refining conditions after puree processing (refined at 0.5 mm or not refined).

3.3 Prediction of apple puree quality traits

Prediction of puree rheological, structural and biochemical properties is compared according to the four techniques (NIR, MIR, Raman and HSI) (**Table 23, Table 24)**.

NIR showed a poor prediction (R_{cv}^2 < 0.52, RPD < 1.4) of puree rheological (*K*, *n*, G', G'', yield stress and tan δ) and structural parameters (d4:3 and d3:2). However it gave a good prediction of puree composition, such as DMC (R_{cv}^2 = 0.82, RPD = 2.3), SSC (R_{cv}^2 = 0.83, RPD = 2.5), TA (R_{cv}^2 = 0.83, RPD = 2.4) and pH (R_{cv}^2 = 0.85, RPD = 2.6). Particularly, the specific wavebands in the intervals 937-1050 nm, 1180-1210 nm and 1290-1330 nm highly contributed to the DMC and SSC models, corresponding to O-H and C-H vibrations of water and carbohydrates (Giovanelli, et al., 2014; Zude, et al., 2006). Besides the aforementioned absorbance regions, NIR wavenumbers between 2208 and 2254 nm, corresponding to the combination bands of C-H and O-H (Wang, et al., 2015), were also considered in the puree DMC prediction. The wavelengths located around 1600 nm (1534-1607 nm for TA models) and 1850 nm (1835-1873 nm for TA and pH models) were used to estimate puree acidity, already described to correspond to the C-O vibration of COOH and O-H combinations (Camps, et al., 2017; Wang, et al., 2015). The prediction of puree individual compounds was acceptable only for malic acid (R_{cv}^2 = 0.80, RPD = 2.1). Generally, NIR spectra coupled with PLS gave a better estimation of puree quality than SVM and RF regression.

MIR technique was potentially able to estimate the rheological parameters (*K*, *n*, G', G'' and tan δ) with acceptable R_{cv}^2 (> 0.81) and RPD (> 2.0) values (**Table 23**). Particularly, PLS and RF models obtained acceptable predictions of the consistency (*K*) (R_{cv}^2 > 0.81, RPD > 2.1) and flow (n) (R_{cv}^2 > 0.80, RPD > 2.0) parameters of the power-law viscosity model of apple purees. PLS models gave the best predictions (R_{cv}^2 > 0.82, RPD > 2.3) of the viscoelastic parameters G' and G'' of purees, but were less accurate for the yield stress (R_{cv}^2 = 0.77, RPD = 1.7). Impressively, MIRS coupled with PLS showed an excellent prediction of tan δ (R_{cv}^2 = 0.96, RPD = 5.1), corresponding to the integrative assessment of both elastic and viscous contributions of apple purees (Espinosa-Muñoz, et al., 2013). The spectral region at 1500-1750 cm^{-1} appeared to be

highly relevant to estimate puree viscosity and viscoelasticity. It corresponds to the C=O and C-O stretching of carboxylic acids at 1745-1740 cm^{-1} and the C=O vibration of pectic methyl ester at 1628-1634 cm^{-1} (Liu, Renard, Rolland-Sabaté, Bureau, & Le Bourvellec, 2020). Concerning the puree structural properties, RF model was the best to predict particle sizes over volume d(4:3) (R_{cv}^2 = 0.88, RPD = 2.9) and over surface area d(3:2) (R_{cv}^2 = 0.82, RPD = 2.2). For composition, acceptable to good PLS predictions were obtained for SSC, DMC, TA, pH, malic acid and sucrose, giving RPD from 2.2 to 3.9 (**Table 24**). The specific spectral signals related to the acids at 1736-1718 cm^{-1} and to the fructose and sucrose at 1065-1055 cm^{-1} and 1024-1016 cm^{-1} (Bureau, et al., 2019), were the major contributors of SSC and DMC models. The excellent predictions of TA and pH, with RPD values of 3.6 and 3.9 respectively, depended on the particularly strong absorptions between 1736-1715 cm^{-1}. However, a lower RPD (RPD = 2.2) and a higher LVs were obtained for malic acid than for TA. For individual sugars, an acceptable PLS prediction was obtained for fructose (R_{cv}^2 = 0.85, RPD = 2.6) based on its typical fingerprints at 1155 cm^{-1}, 1056 cm^{-1}, and 980 cm^{-1} (Bureau, et al., 2019), but neither for sucrose (R_{cv}^2 < 0.78, RPD <1.9) nor for glucose (R_{cv}^2 < 0.49, RPD <1.4).

Raman spectroscopy showed a limited ability to estimate the rheological and structural properties of apple purees with low R_{cv}^2 (< 0.48) and RPD (< 1.4) values (**Table 23**). These results were in line with the lower ability of the aforementioned Raman model to distinguish between non-refined and refined purees (acc = 56.9%) (**Part 3.2**). Moreover, none of the developed Raman models gave acceptable predictions of the global (SSC, DMC, TA and pH) and individual biochemical compositions (sugars, acids and cell wall contents) of apple purees. The best Raman model had a R_{cv}^2 of 0.71 and a RPD value of 1.8, indicating a possible application just to distinguish puree samples presenting a large variation of titratable acidity (TA).

The models based on HSI data could not predict rheological (*K*, *n* , G', G'', yield stress, tan δ) (R_{cv}^2 < 0.48, RPD < 1.4) and structural (d4:3 and d3:2) (R_{cv}^2 < 0.47, RPD < 1.4) properties. Acceptable PLS predictions were obtained for SSC (R_{cv}^2 = 0.86, RPD = 2.7), DMC (R_{cv}^2 = 0.84, RPD = 2.4), TA (R_{cv}^2 = 0.83, RPD = 2.4) and pH (R_{cv}^2 =

0.85, RPD = 2.6). Particularly, the most contributing wavelengths located at around 1180-1219 nm, 1282-1327 nm and 2179-2207 nm were the same as described with the NIR spectroscopy. However, none of the models could predict individual sugars (fructose, glucose and sucrose) ($R_{cv}^2 < 0.74$, RPD < 1.8) and AIS contents ($R_{cv}^2 < 0.42$, RPD < 1.3).

Table 23. Prediction of rheological and structural properties of apple purees using the full cross-validation PLS, SVM and RF regression based on their NIR, MIR, Raman and HSI spectra.

Parameter	Spectra	Ranges	SD	PLS-R				SVM-R			RF-R			Key frequencies
				R_{cv}^2	$RMSE_{CV}$	RPD	LVs	R_{cv}^2	$RMSE_{CV}$	RPD	R_{cv}^2	$RMSE_{CV}$	RPD	NIR (nm), MIR (cm^{-1}), Raman (cm^{-1}), HSI (nm)
Viscosity- *K*	NIR	6.6 - 46.8	8.7	0.41	6.8	1.3	6	0.31	8.2	1.1	0.32	7.7	1.1	/
	MIR			0.81	4.1	2.1	7	0.71	5.5	1.6	0.81	4.1	2.1	1712, 1682 - 1668, 1539, 1152, 1094, 1061, 998
	Raman			0.37	6.95	1.3	6	0.31	7.4	1.3	0.31	7.2	1.2	/
	HSI			0.54	6.1	1.4	10	0.27	7.5	1.2	0.36	6.6	1.3	/
Viscosity- *n*	NIR	0.19 - 0.34	0.04	0.52	0.03	1.4	6	0.30	0.04	1.0	0.35	0.03	1.2	/
	MIR			0.81	0.02	2.2	8	0.80	0.02	2.1	0.80	0.02	2.1	1745 - 1740, 1712 - 1710, 1539, 1140, 1081, 1065-1059,1036, 980
	Raman			0.48	0.03	1.4	8	0.36	0.03	1.4	0.44	0.03	1.3	/
	HSI			0.42	0.03	1.3	7	0.25	0.03	1.1	0.33	0.03	1.2	/
G' (Pa)	NIR	617 - 1962	322	0.32	270	1.2	6	0.11	320	1.0	0.27	282	1.1	/
	MIR			0.82	140	2.3	8	0.80	156	2.1	0.83	139	2.3	1745-1740, 1707, 1634, 1558 - 1537, 1140, 1078, 1063, 1036, 980
	Raman			0.10	326	1.0	6	0.11	303	1.0	0.25	276	1.2	/
	HSI			0.38	263	1.2	9	0.21	298	1.1	0.26	273	1.2	/
G'' (Pa)	NIR	114 - 593	92	0.36	77	1.2	6	0.21	92	1.0	0.26	84	1.1	/
	MIR			0.84	36	2.5	6	0.77	62	1.5	0.81	42	2.2	1745-1740, 1709, 1634-1628, 1558 - 1537, 1139, 1065, 1034, 980
	Raman			0.12	100	0.9	6	0.22	82	0.9	0.20	82	1.1	/
	HSI			0.41	74	1.3	10	0.16	84	1.1	0.19	85	1.1	/
yield stress	NIR	6.4 - 27.7	5.2	0.36	4.2	1.2	6	0.21	5.2	1.0	0.34	4.5	1.2	/
	MIR			0.77	3.0	1.7	7	0.73	2.8	1.8	0.67	3.0	1.8	/
	Raman			0.33	4.3	1.2	8	0.26	4.5	1.2	0.27	4.4	1.2	/
	HSI			0.47	4.0	1.3	11	0.27	4.4	1.2	0.36	4.3	1.2	/
tan δ	NIR	0.18 - 0.30	0.03	0.22	0.03	1.1	5	0.16	0.03	1.0	0.15	0.03	1.0	/
	MIR			0.96	0.01	5.1	7	0.95	0.01	3.7	0.96	0.01	4.5	1749, 1537, 1109 - 1105, 1040 - 1038, 1018 - 1016, 980
	Raman			0.44	0.02	1.3	5	0.45	0.02	1.3	0.42	0.02	1.3	/
	HSI			0.24	0.03	1.1	6	0.14	0.03	1.1	0.15	0.03	1.1	/
d4:3	NIR	239 - 777	130	0.47	95	1.4	6	0.21	130	1.0	0.32	106	1.2	/
	MIR			0.85	50	2.6	8	0.81	60	2.2	0.88	45	2.9	1745, 1626 - 1620, 1539- 1510, 1151, 1099 - 1092, 1061, 1001, 922
	Raman			0.47	93	1.4	6	0.17	117	1.4	0.19	117	1.1	/
	HSI			0.59	85	1.5	9	0.22	114	1.1	0.30	107	1.2	/
d3:2	NIR	170 - 402	53	0.42	41	1.3	6	0.22	53	1.0	0.29	47	1.1	/
	MIR			0.66	31	1.7	8	0.70	30	1.8	0.81	24	2.2	1745, 1699, 1626-1620, 1151, 1099 - 1092, 1061, 1001, 975, 922
	Raman			0.43	41	1.3	6	0.14	49	1.3	0.14	50	1.1	/
	HSI			0.50	40	1.3	9	0.26	46	1.2	0.29	44	1.2	/

Notes: Puree spectra and reference data from four varieties ('Golden Delicious', 'Braeburn', 'Granny Smith' and 'Royal Gala') with different fruit thinning practices for Golden Delicious apples (50 - 100 fruits/ tree or 150-200 fruits/ tree), stress treatments for Braeburn apples (11 days at 24 °C or 2 months at 4 °C), two processing

recipes (70 ℃ for 15 mins with 3000 rpm grinding or 95 ℃ for 17 mins with 400 rpm grinding) and two refining conditions (refined at 0.5 mm or not refined). NIR spectra: SNV pre-treated and smoothed (13 windows) at 800-2500 nm; MIR spectra: the SNV pre-treated at 1800-900 cm^{-1}; Raman spectra: the SNV pre-treated and smoothed (13 windows) at 1800-800 cm^{-1}; HSI spectra: the SNV pre-treated and smoothed (3 windows) at 990-2450 nm. All results corresponded to 10-fold full-crossed validation tests. R_{cv}^2: determination coefficient of the full-crossed validation test; $RMSE_{cv}$: root mean square error of full-cross validation test; RPD: the residual predictive deviation of full-crossed validation test, LVs: the optimal numbers of latent variables. PLS-R: partial least square regression; RF-R: random forest regression; SVM-R: support vector machine regression.

Table 24. Prediction of biochemical properties of apple purees using the full cross-validation PLS, SVM and RF regression based on their NIR, MIR, Raman and HSI spectra.

Parameter	Spectra	Ranges	SD	PLS-R				SVM-R			RF-R			Key frequencies
				R_{cv}^2	$RMSE_{CV}$	RPD	LVs	R_{cv}^2	$RMSE_{CV}$	RPD	R_{cv}^2	$RMSE_{CV}$	RPD	NIR (nm), MIR (cm^{-1}), Raman (cm^{-1}), HSI (nm)
DMC (g/g)	NIR	0.16 - 0.23	0.01	0.82	0.01	2.3	7	0.73	0.01	1.9	0.78	0.01	2.1	937, 946, 1139, 1180 - 1210, 1307 - 1330, 2208 - 2254
	MIR			0.85	0.01	2.7	5	0.76	0.01	1.8	0.78	0.01	1.9	1734 - 1718, 1655 - 1637, 1084, 1061, 1024 – 1016
	Raman			0.20	0.01	1.0	8	0.02	0.01	1.0	0.01	0.01	1.0	/
	HSI			0.84	0.01	2.4	7	0.70	0.01	1.6	0.79	0.01	2.1	1037-1065, 1145, 1180-1219,1305-1338, 2286, 2421
SSC (°Brix)	NIR	11.6 - 15.8	1.1	0.83	0.4	2.5	6	0.50	0.8	1.4	0.57	0.7	1.5	944 - 946, 992, 1180- 1210, 1239, 1290 - 1330
	MIR			0.88	0.4	2.9	3	0.78	0.5	2.2	0.82	0.4	2.4	1736 - 1718, 1065 – 1055, 1022 - 1016
	Raman			0.39	0.9	1.2	9	0.18	1.0	1.2	0.15	1.0	1.1	/
	HSI			0.86	0.4	2.7	8	0.66	0.7	1.5	0.76	0.6	1.9	1048-1071, 1140-1151, 1180-1219,1290-1338
TA (meq/kg)	NIR	3.5 - 11.1	2.2	0.83	0.9	2.4	7	0.43	1.7	1.3	0.72	1.2	1.8	1017, 1049, 1167, 1374, 1534 - 1607, 1835 - 1873
	MIR			0.92	0.6	3.6	5	0.92	0.6	3.6	0.91	0.6	3.4	1736 - 1718, 1605 - 1601, 1042 - 1030, 1001 - 995
	Raman			0.71	1.2	1.8	9	0.58	1.6	1.8	0.58	1.4	1.6	/
	HSI			0.83	0.9	2.4	7	0.66	1.3	1.7	0.76	1.1	2.0	1054-1071, 1085-1214, 1293-1316, 2179-2207
pH	NIR	3.4 - 4.3	0.2	0.85	0.09	2.6	7	0.71	0.13	1.8	0.73	0.13	1.9	912, 1018, 1178, 1280 - 1305, 1835 - 1875
	MIR			0.93	0.06	3.9	5	0.91	0.07	3.6	0.91	0.07	3.6	1718 - 1715, 1094, 1065, 1034, 998, 968
	Raman			0.59	0.2	1.5	9	0.43	0.2	1.5	0.37	0.2	1.3	/
	HSI			0.85	0.1	2.6	7	0.66	0.1	1.7	0.73	0.1	1.9	1054-1065, 1185-1280,1282-1327, 2179-2207
malic (g/kg)	NIR	3.0 - 7.5	1.0	0.80	0.5	2.1	8	0.61	0.7	1.4	0.66	0.7	1.5	912, 1018, 1178, 1365, 1384, 1843 - 1860, 1908
	MIR			0.81	0.5	2.2	6	0.79	0.6	1.6	0.78	0.6	1.8	1730 - 1715, 1095 - 1082, 1001 - 995, 968 - 962
	Raman			0.27	0.9	1.2	9	0.15	0.9	1.2	0.13	1.0	1.1	/
	HSI			0.80	0.5	2.0	7	0.65	0.7	1.5	0.70	0.6	1.7	1134, 1185-1280, 1338-1367, 1843-1860, 2196-2246
fructose (g/kg)	NIR	18.7 - 84.4	13.6	0.73	7.2	1.9	8	0.51	9.7	1.4	0.52	9.8	1.4	/
	MIR			0.85	5.2	2.6	6	0.79	7.2	1.9	0.84	5.4	2.5	1155, 1094, 1065, 1056, 1034, 980
	Raman			0.66	8.5	1.6	7	0.25	8.5	1.6	0.39	10.5	1.3	/
	HSI			0.74	7.1	1.9	7	0.43	9.6	1.4	0.57	9.2	1.5	/
sucrose (g/kg)	NIR	11.0 - 81.9	17.8	0.53	11.9	1.5	7	0.40	12.3	1.4	0.41	10	1.3	/
	MIR			0.78	9.4	1.9	8	0.76	8.9	1.7	0.75	9.8	1.8	/
	Raman			0.47	12.7	1.4	5	0.33	15.5	1.1	0.35	14.8	1.3	/

	HSI			0.61	11.1	1.6	7	0.15	16.3	1.1	0.27	15.1	1.2	/
glucose (g/kg)	NIR	10.0 - 22.5	2.9	0.35	2.3	1.2	4	0.31	2.6	1.1	0.39	2.2	1.3	/
	MIR			0.44	2.2	1.3	7	0.43	2.2	1.3	0.49	2.0	1.4	/
	Raman			0.11	2.9	1.0	8	0.03	2.8	1.0	0.09	2.8	1.0	/
	HSI			0.41	2.2	1.3	4	0.27	2.4	1.2	0.37	2.3	1.3	/
AIS (FW)	NIR	16.0 - 26.7	2.7	0.34	2.3	1.2	5	0.36	2.2	1.3	0.31	2.5	1.1	/
	MIR			0.42	2.2	1.2	10	0.57	1.8	1.5	0.51	1.9	1.4	/
	Raman			0.10	2.9	1.0	6	0.10	2.7	1.0	0.11	2.90	0.9	/
	HSI			0.35	2.3	1.2	6	0.21	2.5	1.1	0.30	2.4	1.1	/

Notes: Puree spectra and reference data from four varieties (‘Golden Delicious’, ‘Braeburn’, ‘Granny Smith’ and ‘Royal Gala’) with different fruit thinning practices for Golden Delicious apples (50 - 100 fruits/ tree or 150-200 fruits/ tree), stress treatments for Braeburn apples (11 days at 24 ℃ or 2 months at 4 ℃), two processing recipes (70 ℃ for 15 mins with 3000 rpm grinding or 95 ℃ for 17 mins with 400 rpm grinding) and two refining conditions (refined at 0.5 mm or not refined). NIR spectra: SNV pre-treated and smoothed (13 windows) at 800-2500 nm; MIR spectra: the SNV pre-treated at 1800-900 cm-1; Raman spectra: the SNV pre-treated and smoothed (13 windows) at 1800-800 cm-1; HSI spectra: the SNV pre-treated and smoothed (3 windows) at 990-2450 nm. All results corresponded to 10-fold full-crossed validation tests. R_{cv}^2: determination coefficient of the full-crossed validation test; $RMSE_{cv}$: root mean square error of full-cross validation test; RPD: the residual predictive deviation of full-crossed validation test, LVs: the optimal numbers of latent variables. PLS-R: partial least square regression; RF-R: random forest regression; SVM-R: support vector machine regression.

3.4 Comparison of NIR, MIR, Raman and HSI performances

NIR spectroscopy, the easiest to apply and the most affordable spectroscopic technique in this work, showed an acceptable ability (2.3 < RPD < 2.6) to predict puree major chemical composition, including SSC, DMC, TA and pH. Such good NIR predictions will probably contribute to the development of the rapid routine evaluation of the composition of fruit-based products. For individual components, a good estimation was only obtained for malic acid, depending on its positive correlation with TA (R^2 = 0.78) and pH (R^2 = 0.76). However, this technique can not provide acceptable estimations of puree textural changes, in line with our conclusions in **Paper IV**.

Compared to NIR, MIR technique had the potential to assess puree rheological properties, including both, viscosity and viscoelasticity. However, the predictions shown in **paper I** were less accurate (RPD > 2.0) than our works in **paper II** (RPD > 2.4), which concerned purees presenting a larger range of rheological behaviors. Interestingly, among puree viscoelastic parameters, tan δ was the best estimated by MIR (RPD = 5.1). Compared to machine learning models (SVM and RF), PLS regressions generally showed a better ability to predict puree rheological and biochemical properties. However, for the puree particle structure (sizes and volume), random forest regression provided better predictions. The informative wavenumber regions at 1500-1750 cm^{-1} and 900-1200 cm^{-1} should be considered for rheological and structural assessments of apple purees, which was in line with a previous work (Ayvaz, et al., 2016) and the results of **paper II**. For biochemical composition, MIR coupled with PLS regression provided the best prediction of puree global quality traits (SSC, DMC, TA and pH) with the possibility to evaluate some individual components (malic acid and sucrose). The lower prediction of malic acid than of TA was probably due to its relatively low concentration (3.0 - 7.5 g/kg of malic acid and 3.5 – 11.1 g/kg of TA) and limited variations (SD = 1.0 g/kg of malic acid, SD = 2.2 g/kg of TA). For puree individual sugars, the higher internal correlations between fructose and SSC (R^2 = 0. 78) than between sucrose and SSC (R^2 = 0. 51) probably explained the better prediction of fructose than of sucrose.

In this study, Raman technique showed a potential to discriminate different purees, according to variety and processing recipes (acc > 81.3%), but it was not able to predict puree rheological, structural and chemical parameters. However, Raman gives excellent biochemical predictions on homogeneous samples, such as commercial tomato purees (Baranska, et al., 2006) and honey products (Özbalci, et al., 2013; Pierna, et al., 2011), which contain high concentrations of the predicted components (lycopene in tomato and sugars in honey). It has also been used to detect the rheological changes of monotonous mixed food matrices (Nawrocka, Miś, & Szymańska-Chargot, 2016; Ngarize, Adams, & Howell, 2004). In this work, the unsatisfactory predictions using Raman spectroscopy could be due to i) the very weak spectral signals corresponding to the biochemical variations in apple purees (even after water evaporation before spectrum acquisition) and ii) the variable heterogeneity according to the puree refining and grinding, which make a barrier against an efficient light diffusion.

The models based on the averaged NIR-HSI spectra of apple purees provided a significant improvement of puree discrimination and a slight increase in quality prediction in comparison with the results issued from a measurement of a limited sample area (~2 cm^2) by NIR spectroscopy. The averaged NIR-HSI spectra, which contained a richer spectral information of puree heterogeneity than the local NIR spectra, could explain the better model performance (Cheng & Sun, 2017). However, both NIR spectroscopy and HSI technique had a limited ability to detect puree differences after refining and to predict their rheological and structural properties. Strangely, the PLS-DA models using the full number of HSI spectra of each puree had a relatively lower discriminating accuracy than their corresponding averaged spectra. Previous works noticed the heterogeinity of tested samples usually affected the NIR and HSI determination precisions (Prieto, Roehe, Lavín, Batten, & Andrés, 2009). The large heterogeneity, including irregular particle size and shape and their large water content on puree surface, could introduce a strong diffuse reflection and spectral noise during the HSI image acquisition. Although NIR-HSI on purees slightly improved the prediction of SSC and DMC than NIR results, the much larger volume of dataset and

the longer time needed for image pre-processing limited its use in comparison with NIR local measurements.

Further, the AIS, which contributes to the rheological properties of processed puree products, was not well evaluated in this study whatever the technique used directly on puree samples.

4. Conclusion

This study provided a first comprehensive assessment to choose the best techniques among NIR, MIR, Raman spectroscopy and HSI for evacuating apple puree variability and quality. MIR technique had the best performance to provide an accurate identification of puree properties due to apple variability (variety, fruit thinning and stress treatments) and processing conditions (heating, grinding and refining). It gave also a reliable evaluation of puree rheological and structural characteristics and composition (RPD values from 2.1 to 5.1). NIR and HSI techniques can be easier adapted to routine characterization of puree global parameters (soluble solids, titratable acidity and dry matter), but not of their textural changes. Raman spectroscopy offered an insufficient information to evaluate apple puree variability and quality. Clearly, Raman spectroscopy should not be prioritized in further studies on the determination of the properties of apple purees.

The current study also makes possible to consider future applications with NIR, NIR-HSI and MIR according to the industrial or research needs (speed of data acquisition and presentation of the sample). These techniques are very suitable for the development of Process Analytical Technology in order to trace samples and optimize conditions during processing.

Highlights of Paper I

This study answered our first question in Part 1:

What is the most appropriate method to identify the puree variability and detect puree physical, chemical and structural properties by using different vibrational spectroscopic techniques?

- MIR was the best tool to identify puree variability (> 91.7 % of correct classification) and to potentially predict average particle size (RPD = 2.9), viscosity (RPD ≥ 2.1) and viscoelastic properties (RPD > 2.3) of puree,
- NIR, MIR and HSI techniques had a good ability to estimate puree composition such as soluble solids (RPD > 2.5), titratable acidity (RPD > 2.4) and dry matter (RPD > 2.3).
- Raman spectroscopy could not provide sufficient assessment of puree quality.

Consequently, NIR, MIR and HSI should be prioritized as process analytical technologies to detect the variability of purees and assess their texture and taste.

Paper II (Published)

According to our **Paper I**, MIR (or FTIR) technique has been proven as the best method to identify puree variability and satisfactorily predict their quality. It also gave an acceptable assessment of rheological properties. However, it stayed limited to predict cell wall (AIS) contents, while they are major determinants of texture. That could be due to the low sensitivity (cell wall contents are typically below 20 g/kg) and limited specific signals for cell walls, probably due to some strong interactions with water in fresh purees and signal similarity to soluble sugars. To overcome this limitation, sample preparation was investigated as a key point for MIR (FT-IR) spectral analysis to improve analytical results. Specifically, freeze-drying to reduce the water interferences or extracting cell wall materials to remove major soluble chemicals were prioritized.

Therefore, the possibility of MIR technique to obtain sufficient information for chemical, textural and rheological properties based on fresh, freeze-dried (removing water interactions) and cell wall materials (removing soluble chemicals) of raw apples and their corresponding purees was studied in **paper II**. The MIR spectra of 36 apple sets and the corresponding 72 purees, issued from different varieties, agricultural practices, storage periods and processing conditions were treated in order to:

i) Evaluate how much sample preparation has improved the prediction of chemical, textural and rheological properties of purees (number of quality traits and the precision of prediction)
ii) Identify signals that are specific of the changes related to apple processing.

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

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?

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

[a] UMR SQPOV, INRAE, Avignon University, F-84000 Avignon, France
[b] Stat SC, INRAE, ONIRIS, F-44300 Nantes, France
[c] TRANSFORM, INRAE, F-44300 Nantes, France

1. Introduction

Sample preparation is a key point for quality of analytical data. Infrared spectroscopy (near or mid-infrared), because of its integrative nature, is one of the main candidates for a rapid qualification of agricultural commodities and processed food, especially in the view of process analytical technology (PAT). Advanced techniques based on infrared spectroscopy offer the advantages of a minimal sample preparation and a rapid data acquisition. However, this questions the balance between data intensity and required sample preparation hence man-power: are the data acquired on "raw" samples sufficient for process monitoring, quality control or process comprehension? A specific point is also that foods are frequently highly hydrated and not stable, so that appropriate steps must be taken to preserve samples for later quality control. As the time consumption and cost of sample preparation are generally barriers to a rapid and precise determination by spectroscopy, knowing the most efficient sample pretreatments could contribute to improve analytical results as well as to provide informative options at both, laboratory and industrial scales.

Different methods for the reference data acquisition such as HPLC, GC-MS or NMR (Bureau et al., 2013), types of spectroscopy or related hyperspectral images (NIR, MIR, Raman) (Baranska, Schütze, & Schulz, 2006) and modeling algorithms (Van Boekel, 2008) have been intensively compared on fruits. It seems also crucial to compare and determine the optimal sample form (fresh, freeze-dried or cell wall extracts) and the associated changes occurring during fruit processing, notably using

infrared spectroscopy which has the potential to be applied both, on-line and off-line.

Direct ATR-FTIR estimations on fruit fresh homogenates have obtained good results to predict soluble solids content, dry matter content, titratable acidity, some individual sugars and organic acids (Bureau, Ścibisz, Le Bourvellec, & Renard, 2012; Ayvaz et al., 2016). As infrared spectroscopy is extremely sensitive to changes of hydrogen bonding (Jackson & Mantsch, 1995), the main drawback of spectral measurements is the low sensitivity and limited specific signals of chemical compositions under strong water interactions in fresh fruit suspensions, such as citric acid in apples (Bureau, Ścibisz, Le Bourvellec, & Renard, 2012), lycopene and β-carotene in tomato (Baranska, Schütze, & Schulz, 2006). Moreover, classical measurements of rheological properties and particle size distribution of fruit products require costly rheometer, particle sizing equipment and experienced staffs. Therefore, one of the challenging works is to investigate the possibility of ATR-FTIR to estimate the specific rheological modifications (viscosity and viscoelastic parameters) and then to monitor textural changes (particle size and volume) for both, accurate determinations in scientific research or rapid and direct assessment in industrial processing.

Much more information can be extracted from dry food commodities, such as the structural changes of cereals (Georget & Belton, 2006), micronutrients in fruits (Lu et al., 2011) and even cell wall content variations (Canteri, Renard, Le Bourvellec, & Bureau, 2019). To overcome the limitations observed on highly hydrated products, such as fruits, drying methods with as limited as possible alteration of composition and structure are needed. Thus, freeze-drying prevents evolution of samples under the action of endogenous enzymes (notably oxidation and hydrolysis). It also carries out a concentration due to water elimination, so that specific components present in low concentrations can have significant spectral absorptions. But freeze-drying is expensive and time-consuming, needing at least 24-48 hours. It allowed to obtain similar predictions of chemical compositions than those in fresh samples (de Oliveira, de Castilhos, Renard, & Bureau, 2014; Oliveira-Folador et al., 2018). Few detailed studies compared the differences and limitations of ATR-FTIR fingerprint regions on fresh and corresponding freeze-dried plant leaves (Durak & Depciuch, 2020).

ATR-FTIR applications to assess fruit textural properties (mainly focus on cell wall compositions) are always performed on their cell wall materials (AIS) (Canteri, Renard, Le Bourvellec, & Bureau, 2019; Szymanska-Chargot, Chylinska, Kruk, & Zdunek, 2015). However, extracting the cell wall requires a large consumption of chemical solvents if starting from fresh samples (up to 1 L ethanol and 0.4 L acetone/ 1.0 - 1.5 g cell wall). The accelerated or pressurized solvent extractors (ASE, PSE) can allow multiplexing and thus a faster and less solvent-consuming cell wall preparation, but only from already freeze-dried samples. After removing all soluble components (mainly sugars and acids), specific signals related to pectins, cellulose and hemicelluloses have proven to be useful for the fast evaluation of cell wall polysaccharides during fruit growth and subsequent storage (Szymanska-Chargot, Chylinska, Kruk, & Zdunek, 2015). Although some cell wall modifications in plants (Femenia, García-Pascual, Simal, & Rosselló, 2003) and fruits (Cardoso et al., 2009) under heating and dehydration have been investigated by ATR-FTIR. However, for fruit processed purees, little work has been done on ATR-FTIR to detect their cell wall changes during processing and monitor rheological and mechanical properties (Ferreira, Barros, Coimbra, & Delgadillo, 2001).

In this study, ATR-FTIR spectroscopy was applied on the corresponding raw apples and processed purees. Spectra were acquired on different kinds of homogeneous samples such as fresh (NF for non-freeze-dried), freeze-dried (FD) and cell wall extracts (AIS for alcohol insoluble solids) in order to: i) evaluate how much sample preparation improved the prediction of chemical, textural and rheological characteristics of purees (number of quality traits and their precision) and ii) identify signals specific of the variations which occur during apple processing.

2. Materials and methods

2.1 Plant Material

Apples of two varieties: ‘Golden Delicious’ (GD) and ‘Granny Smith’ (GS) were harvested at commercial maturity in 2017 in an experimental orchard named La Pugère (Mallemort, Bouches-du-Rhône, France). Standard commercial fruit thinning practices

(Th+ to 50 to 100 fruits/tree) and no thinning (Th- to 150-200 fruits/tree) were compared during the ripening of 'Golden Delicious'. The three obtained apple groups (Th+ GD, Th- GD and GS) were stored in a cold chamber at 4°C and at around 90% of humidity during one, three and six months (respectively T1, T3 and T6), except the first batch (T0) were analyzed and processed the day after harvest without any storage time.

Each apple batch (T0, T1, T3 and T6) was divided into two subsets (**Fig. 24)**:

i) the first subset was dedicated to apples characterization: 3 replicates of 10 apples were selected and separated into two aggregate samples as described by Bureau (Bureau, Ścibisz, Le Bourvellec, & Renard, 2012). One sample corresponding to the NF sample was stored at -80°C and then homogenized at 11000 rpm with an Ultraturrax T-25 (IKA, Labortechnik, GmbH, Staufen, Germany) after 1.5 h of thawing at 22.5 °C for biochemical and spectral characterizations. The other sample corresponding to the freeze-dried (FD) was used to extract cell wall materials (AIS). Finally, 36 NF, FD and AIS samples (3 apple groups × 4 storage times × 3 biological replicates) of raw apple fruits were obtained.

ii) the second sub-set was dedicated to puree processing: 3 replicates of apples (4 kg each) were used to produce three puree lots. After sorting and washing, apples were cored and cut in 8 portions, then processed in a multi-functional processing system (Roboqbo, Qb8-3, Bentivoglio, Italy). Half of the each puree (2 kg) was refined with a 0.5 mm (Ra) sieve (Robot Coupe C80 automatic refiner, Robot Coupe SNC, Vincennes, France) whereas the other half was not refined (NR). Finally, fresh puree samples (NF) were conditioned in two hermetically sealing cans: one was cooled at room temperature (22.5 °C) before the next-day measurements of rheological, textual and some chemical (soluble solids and titratable acidity) properties, while the other was freeze-dried (FD) and stored at -20 °C for AIS extraction. Thus, in total 72 NF, FD and AIS samples of purees were prepared and characterized, corresponding to 3 apple groups × 4 storage times × 2 refining levels × 3 biological replicates.

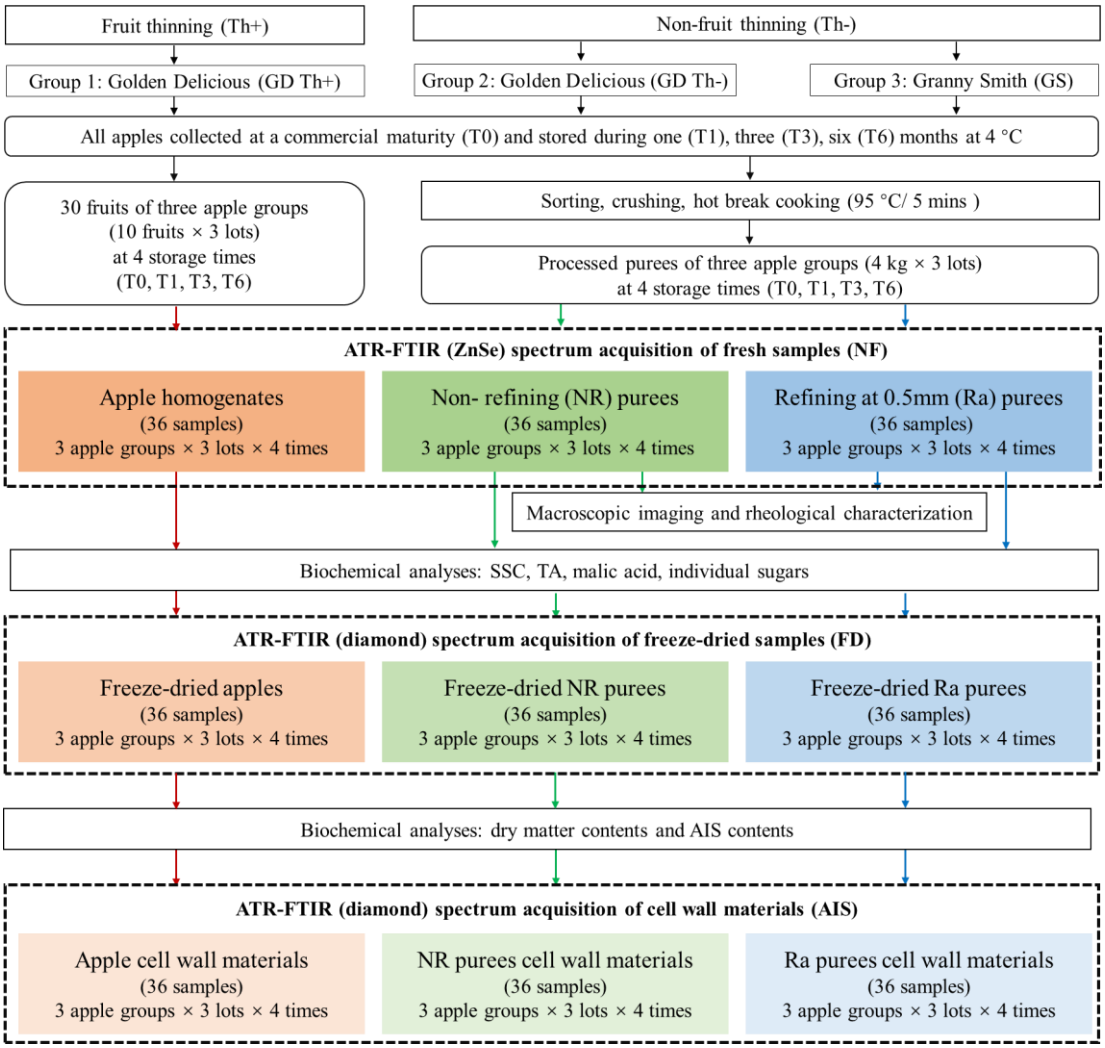

Fig. 24. Experimental scheme for apple and puree samples preparation, characterization using ATR-FTIR and reference analyses.

2.2 Biochemical Analyses

SSC, TA, DMC, individual sugars (glucose, fructose and sucrose) and malic acid were quantified based on the same method described in **Paper I**. Freeze-dried (FD) samples were acquired based on a freeze-dryer of 5 days upon reaching a constant weight. Cell wall materials (AIS) were isolated using the method proposed by Renard (Renard, 2005). and the cell wall contents (AIS contents) were expressed in both, fresh weight (FW) and dry matter weight (DW). Three biological replicates were characterized for each biochemical trait and each sample.

2.3 Rheological Analyses

The puree rheological measurements consisted in one rotational (flow curve) and

two oscillatory (amplitude and frequency sweeps) tests, carried out using a Physica MCR-301 controlled stress rheometer (Anton Paar, Graz, Austria) at 22.5 °C. 50 mL of each puree sample was placed in a C-CC27 with an inner radius of 14.46 mm measuring cup (Anton Paar, Graz, Austria). All tests were performed by a six blade vane geometry FL 100/6W with a radius of 11 mm (Anton Paar, Graz, Austria). The flow curves were performed after a pre-shearing period of 1 minute at 50/s followed by 5 minutes at rest. The viscosity was then measured at a controlled shear rate range of [10; 250]/s on a logarithmic ramp, at a rate of 1 point every 15 seconds. The values of the viscosity at 50/s and 100/s (η_{50} and η_{100} respectively) were kept as indicators of the sensorial puree texture (Engelen & de Wijk, 2012; Espinosa-Muñoz et al. 2012) during consumption. Amplitude Sweep (AS) tests were performed at an angular frequency of 10 rad./s in the deformation range of [0.01; 100] %, in order to determine the linear viscoelastic range of the purees and the yield stress, defined as the crossing point between the storage modulus (AS-G') and the loss modulus (AS-G'') curves. Frequency Sweep (FS) measurements were operated within the linear viscoelastic region as determined by the AS test (0.05%) in the angular frequency range of [0.1; 100] rad./s. For means of comparison the storage and loss moduli (FS-G' and FS-G'') were taken at 1 rad./s to evaluate the viscoelastic properties of the studied purees. Puree samples were diluted in distilled water to separate particles and stained with calcofluor white at 0.1 g/L and highlighted with a 365 nm UV lamp (Soukup, 2014). A high-resolution digital video camera (Baumer VCXU31C, Baumer SAS, France) with a macro lens (VSTech 0513, VS Technology Corporation, Japan.) was used to visualize the distribution and dispersion of puree particles. The particle sizes averaged over volume d(4:3) (de Brouckere mean) and over surface area d(3:2) (Sauter mean) were measured with a laser granulometer (Rawle, 2003) (Mastersizer 2000, Malvern Instruments, Malvern, UK).

2.4 ATR-FTIR spectrum acquisition

ATR-FTIR spectra were collected at room temperature using a Tensor 27 FTIR spectrometer (Bruker Optics, Wissembourg, France) equipped with a horizontal attenuated total reflectance (ATR) sampling accessory and a deuterated triglycine

sulphate (DTGS) detector. Three replications of spectral measurement were performed on all raw and processed apples for fresh (NF for non freeze-dried), freeze-dried (FD) and cell wall (AIS) samples. The spectra of all samples were acquired in random order. The instrument adjustment and spectral acquisition were controlled by OPUS software Version 5.0 (Bruker Optics®). The spectra of raw and processed apples were acquired using two different crystals. A big zinc selenide (ATR-ZnSe) crystal with dimensions of 6 cm x 1 cm and six internal reflections was used for fresh samples (apple homogenates and purees) containing water. For the freeze-dried and cell wall samples, a small crystal was used characterized by a single-reflectance horizontal ATR-Diamond Cell (Golden Gate Bruker Optics) equipped with a press tip flap system to press sample on the crystal always in the same way. Spectra (32 scans for ATR-ZnSe and 16 scans for ATR-Diamond) were collected from 4000 cm^{-1} to 650 cm^{-1} and were corrected against the background spectrum of air.

2.5 Statistical Analyses and Chemometrics

After ensuring normal distribution with a Shapiro-Wilk test (α=0.05), the reference data were presented as mean values and the data dispersion within our experimental dataset expressed as standard deviation values (SD). Analysis of variance (ANOVA) was carried out to determine the significant differences due to the controlled factors (thinning, storage and puree mechanical refining) on both apples (**Table 25**) and purees of each variety (**Table 26 and Table 27**) using XLSTAT (version 2018.5.52037, Addtionsoft SARL, Paris, France) data analysis toolbox.

Spectral pre-processing and multivariate data analysis were performed with Matlab 7.5 (Mathworks Inc. Natick, MA) software using the SAISIR package (Bertrand & Cordella, 2008). The absorption band between 2400-2300 cm^{-1}, due to carbon dioxide, was discarded prior to the calculation. All FT-IR data were pre-processed with baseline correction, standard normal variate (SNV) and a derivative transform calculation Savitzky–Golay method, gap size = 11, 21, 31) of first or second order. After pretests of these pre-processing treatments applied on several different spectral regions, the best results of prediction and discrimination were obtained on the range 1800-900 cm^{-1}, which has been already highlighted (Bureau et al., 2009). Particularly, Principal

Component Analysis (PCA) and Factorial Discriminant Analysis (FDA) were applied on SNV pre-treated spectra (in **Part 3.1** and **Part 3.2**). The specificity and sensitivity values of FDA discriminations were calculated by the already reported method of Nargis (Nargis et al., 2019), in order to better evaluate sample differentation. For PLS (Partial least square) modelling (in **Part 3.3**), the baseline correction coupled with SNV pre-processing had the best performances to correct multiplicative interferences and variations in baseline shift, and reached the best prediction results.

Leave-one-out PLS models were developed using spectra of fresh (NF), freeze-dried (FD) and AIS of puree samples, for which the three spectral matrices (NF, FD and AIS) corresponded to the same reference dataset. A total number of 72 averaged spectra for each puree form (NF, FD and AIS) corresponding to 3 apple groups (GS, GD Th+ and GD Th-) × 4 storage times × 2 puree refining modalities × 3 biological replicates was used as modelling dataset. PLS model performance was assessed using the determination coefficient of cross-validation (R_{cv}^2), the root-mean-square error of cross-validation (RMSECv), the number of latent variables (LVs), the ratio of the standard deviation values (RPD) and the linkable spectral regions (**Table 28** and **Table 29**).

3. Results and discussion

3.1 Spectral characterization of NF (non-freeze-dried) apple purees

PCA and FDA applied on the spectra of NF puree samples successfully allowed to detect puree differences coming from the raw apple variabilities (variety, fruit thinning and storage period) **(Fig. 25)**. They also highlighted the modifications of puree structure by the mechanical refining over several months of apple storage **(Fig. 26)**.

In **Fig. 25**, the first principal component (PC1) clearly discriminated the two varieties (‘Golden Delicious’ and ‘Granny Smith’) and thinning practices for Golden delicious (Th- and Th+), in relation with the fructose variation followed at 1061 cm^{-1} (Bureau, Cozzolino, & Clark, 2019). Moreover, the peak at 1022 cm^{-1}, reported as a peak specific to sucrose in apple juices (Leopold, Leopold, Diehl, & Socaciu, 2011),

appeared to be the main contributor of the second principal component (PC2), which distinguished the storage times. Along the PC2 axis, the discrimination of storage durations from T0 at the top to T6 at the bottom was in relation with the decrease of sucrose (1022 cm^{-1}) and the increase of fructose (1065 cm^{-1}) in purees, in accordance with the reference chemical dataset **(Table 26)**. Consequently, factors such as variety, thinning practice and storage duration affecting raw apple characteristics induced changes in the corresponding purees after processing. ATR-FTIR applied directly on processed purees could then be useful for traceability of these effects impacting raw fruits based on the specific C-C and C-O-C bonds of carbohydrates, such as 1022 cm^{-1}, 1061 cm^{-1}, 1065 cm^{-1}.

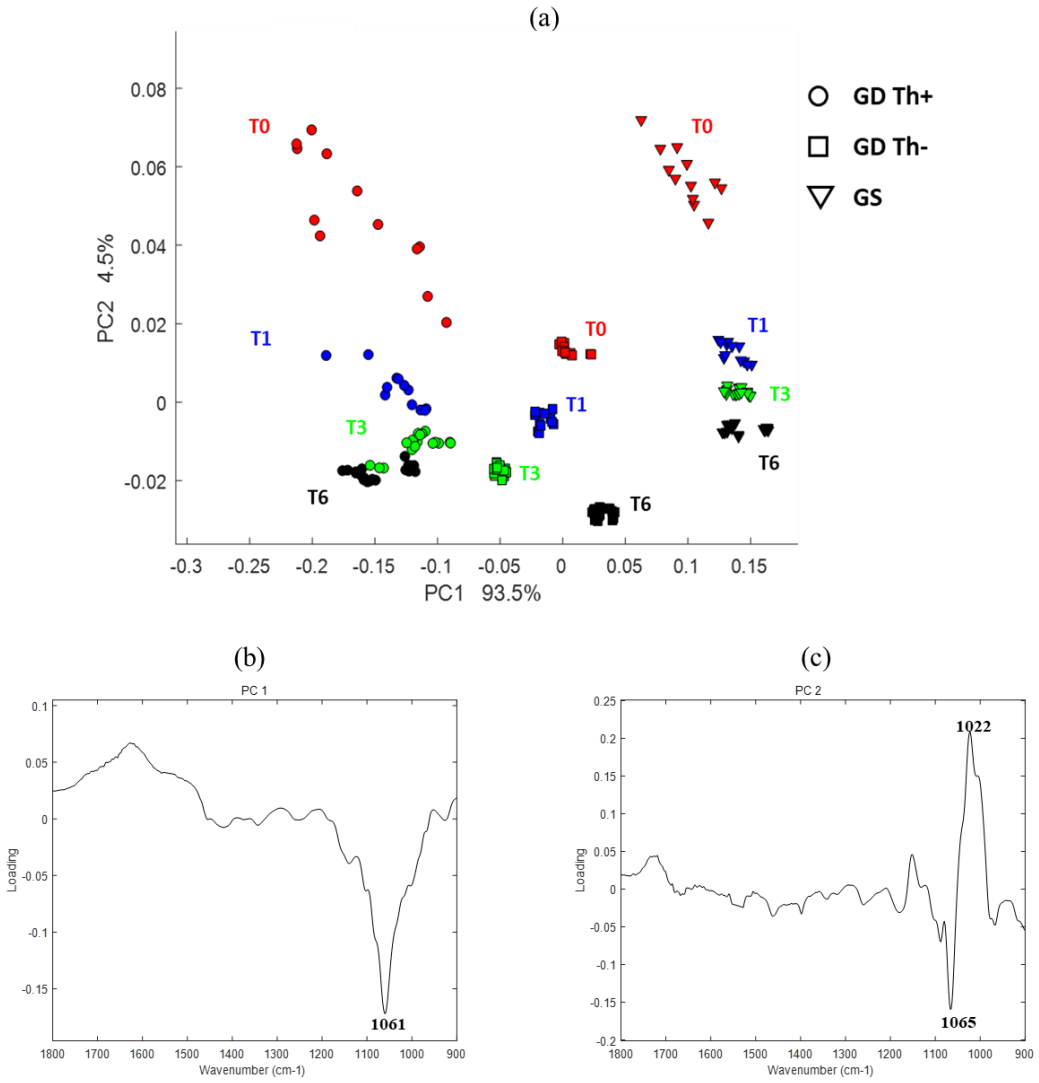

Fig. 25. PCA on the SNV pre-treated ATR-FTIR spectra (900-1800 cm-1)of purees (NF

samples) prepared with normal thinned ‘Granny Smith’ apples (GS marked with △), thinned (Th+) ‘Golden Delicious’ apples (GD Th+ marked with ○) and non-thinned ‘Golden Delicious’ apples (GD Th- marked with □) stored in cold storage room (4°C) during 0, 1, 3 and 6 months (T0, T1, T3 and T6): **(a)** the scores plot of the two first components (PC1 and PC2); **(b)** the loading plot of PC1; **(c)** the loading plot of PC2.

According to the reference data **(Table 26)** and their PCA results **(Fig. 27)** the mechanical refining resulted a clear reduction of cell wall contents (AIS in DW and FW), viscosity (η_{50} and η_{100}), viscoelasticity (yield stress, G’ and G’’ in both oscillatory tests), particle size (d(4:3) and d(3:2)) in T0 purees prepared with apples at harvest (T0). However, gradually over apple storage, less differences were detected between the non-refined (NR) and refined (Ra) purees. The non-refined (NR) 'Golden Delicious' and ‘Granny Smith’purees were characterized by large apple particles and only few small separated cells at the beginning of cold storage (T0) **(Fig. 26)**. The refining treatments mainly led to lower particle size by removing the big puree particles **(Fig. 26a)**. However, at the end of storage (T6), both non-refined (NR) and refined (Ra) purees were mostly composed of single cells and no clear difference was observed between them **(Fig. 26d)**. This similar structure of NR and Ra purees at T6 could be due to an increase in cell separability linked to a decrease of the intermolecular bonding between cell wall polymers and a notable increase of pectin solubility during apple storage (Varela, Salvador, & Fiszman, 2007).

FDA performed on the spectra of all NR and Ra purees (NF samples) at each apple storage time gave highly consistent observations with the reference data and macroscopic images showed above **(Fig. 26)**. According to the third factorial components (F3) (F1 and F2 for variety and thinning discriminations, **Fig. 28**), the two puree refining levels were well separated at T0, then appeared progressively overlapped at T3 and T6 **(Fig. 26)**. Especially along the F3 axis, at T0, intensive spectral variations were related to the decrease of soluble organic acids (1718 cm^{-1} and 1709 cm^{-1}), soluble polysaccharides, pectins and absorbed water (1740 cm^{-1}, 1695 cm^{-1}, 1682 cm^{-1}, 1668 cm^{-1}, 1655 cm^{-1} and 1468 cm^{-1}) between the two refining conditions **(Fig. 29)**. Although

the peaks of carbohydrates at 1019 cm^{-1} and 1049 cm^{-1} (glucose/fructose) and 1155 cm^{-1} (the glycosidic linkage) are known to successfully monitor the consistency of tomato juice (Ayvaz et al., 2016), the region between 1750 and 1450 cm^{-1} highly contributed to the discrimination of apple purees according to their particle size and their rheological behavior after mechanical refining treatments. These differences between tomato and apple might be due to the nature of the datasets and in particular the impact of post-harvest storage on chemical compositions (sugars and acids) and textural properties (pectins degradations) as confounding factors in this apple processing experiment.

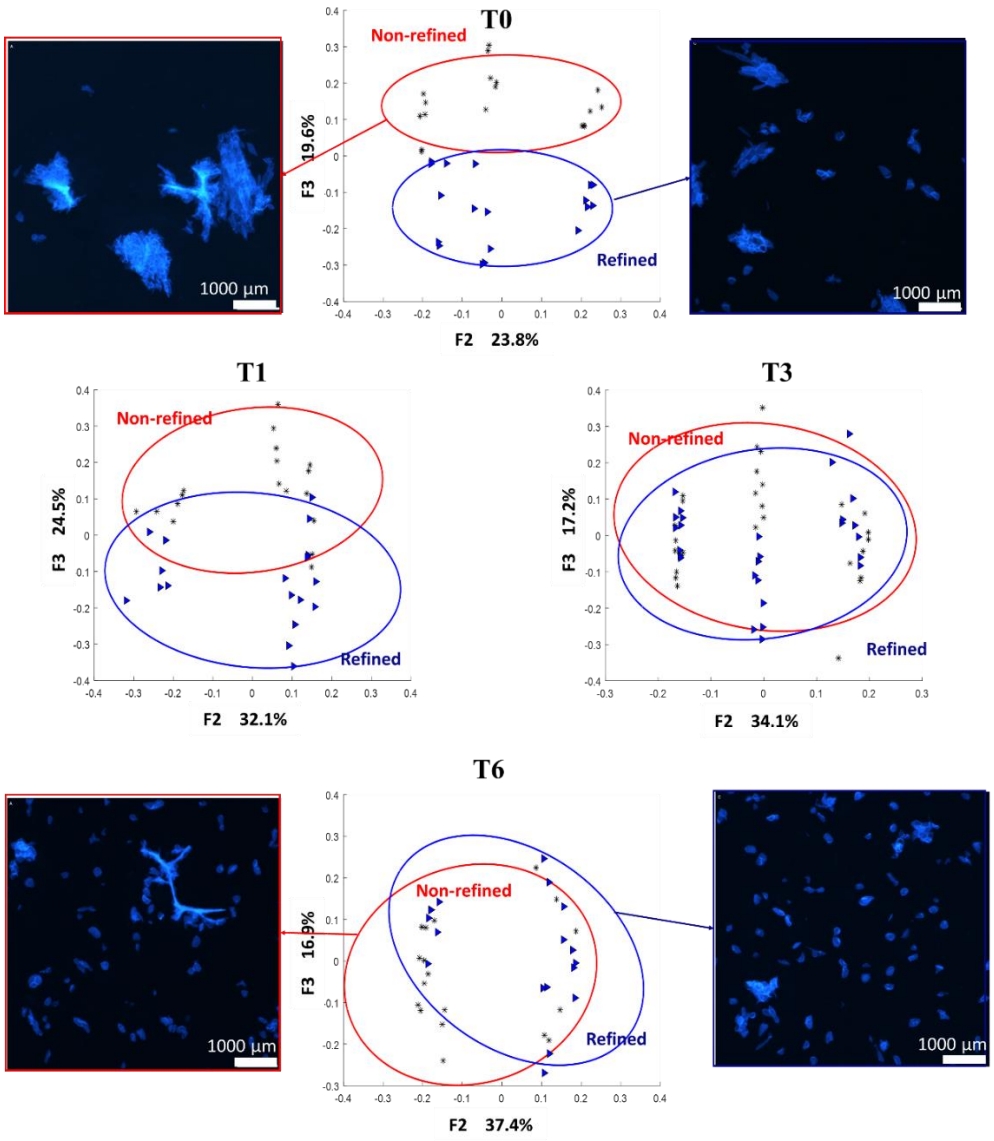

Fig. 26. FDA on the SNV pre-treated ATR-FTIR spectra (900-1800 cm^{-1}) of non-refined (* with 95% confidence ellipse circles) and refined (△ with 95% confidence ellipse

circles) ‘Golden Delicious’ and ‘Granny Smith’ purees at harvest (T0), after one-month (T1), three months (T3) and six months (T6) of storage at 4°C. Macroscopic laser scanning images of puree particle distributions at harvest (T0) and after six-month storage (T6).

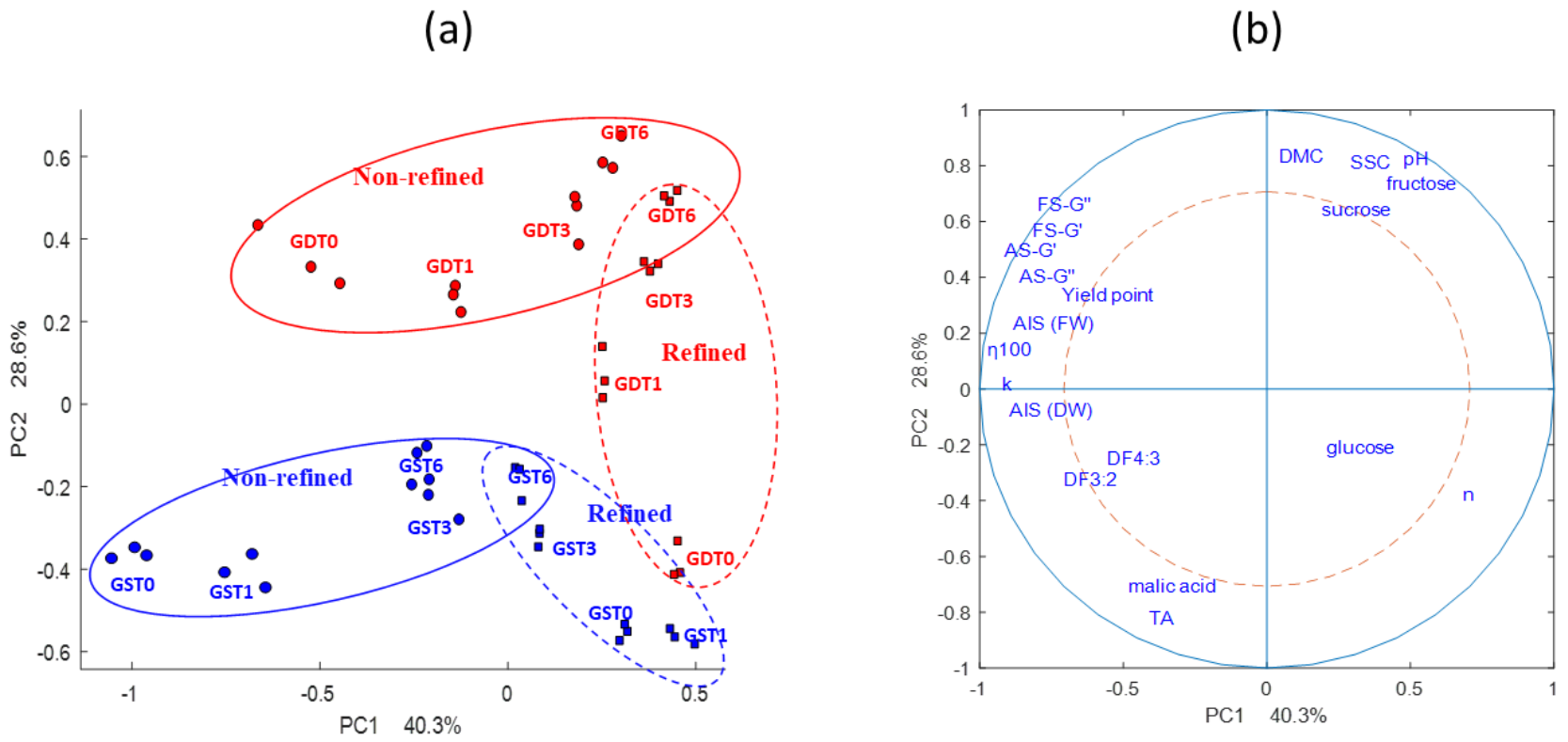

Fig. 27. **(a)** PCA on the reference data of purees prepared with apples (Golden Delicious and Granny Smith) stored in a cold storage room until six months (T0, T1, T3 and T6) and submitted to two refining levels ('Non-refined' and 'Refined') after cooking; **(b)** the first two PCs loading plot of references data.

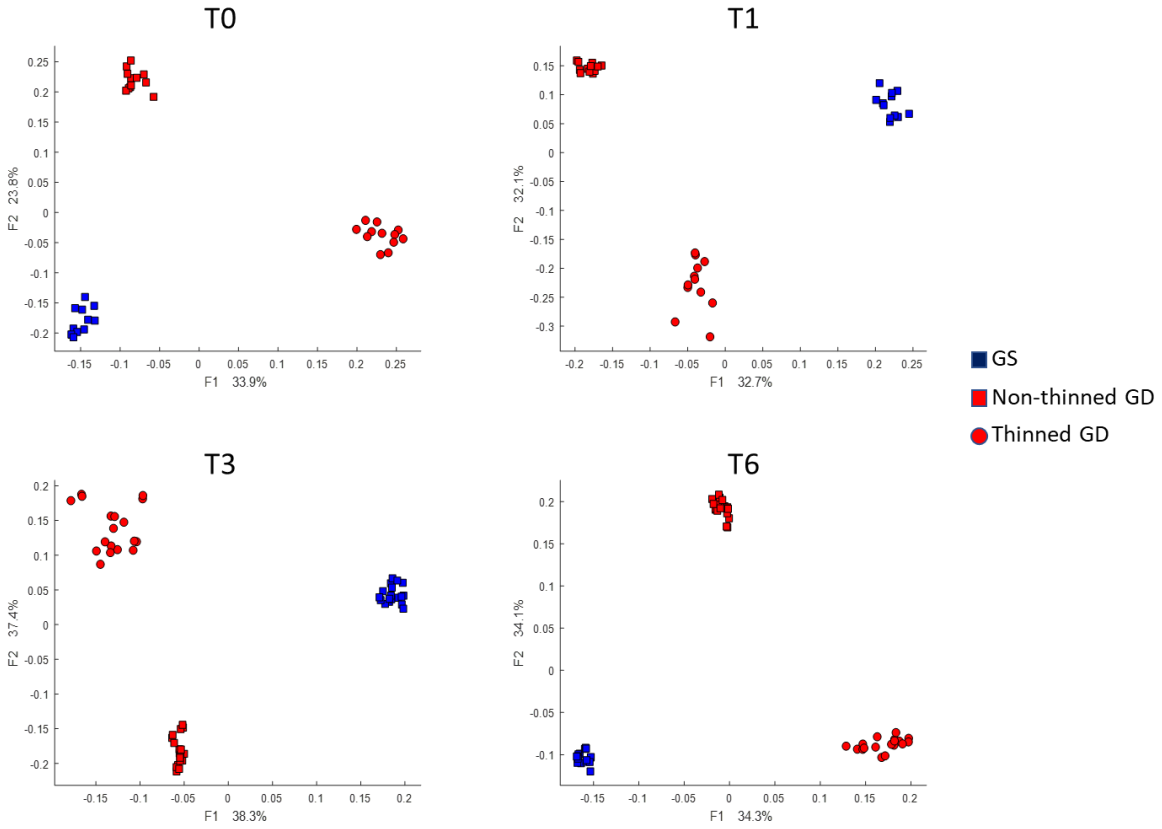

Fig. 28. Map of the first two discriminant factors (F1 and F2) of FDA results on the SNV pre-treated ATR-FTIR spectra (900-1800 cm^{-1}) of non-thinned and thinned 'Golden Delicious' (GD) and 'Granny Smith' (GS) purees at harvest (T0), after one-month (T1), three months (T3) and six months (T6) of storage at 4°C.

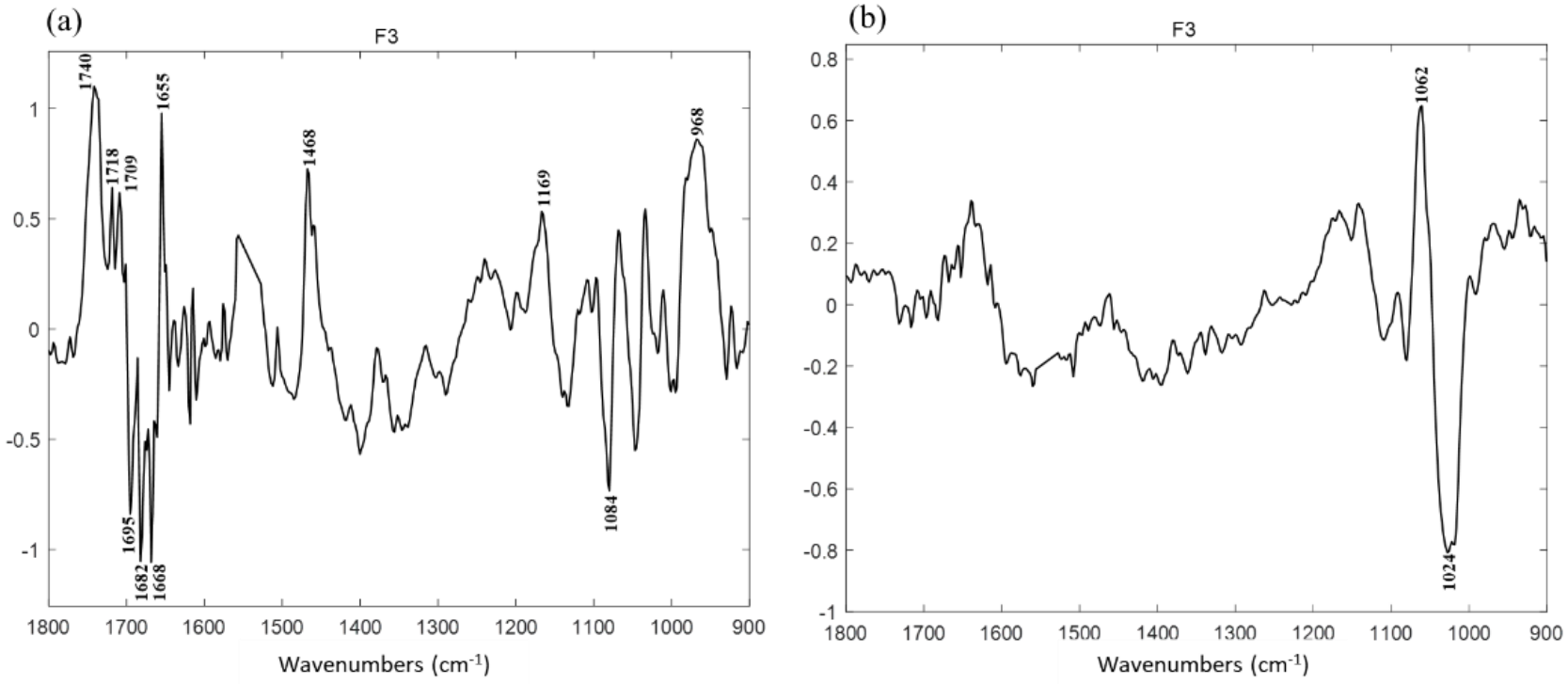

Fig. 29. The third factorial loadings (F3) of the FDA discriminating the non-refined (NR) and Refined (Ra) 'Golden Delicious' and 'Granny Smith' purees between 1800 and 900 cm^{-1} at: (a) harvest (T0); (b) after six month cold storage (T6).

3.2 Spectral evaluation of the link between fresh and processed apples

FDA results showed a good ability to discriminate puree processing changes (**Fig. 30**) and variety differences **(Fig. 31**), according to the first two discriminant factors (F1 and F2). Whatever the sample preparation (NF, FD and AIS), a clear separation was observed between raw apples (homogenates) and processed purees **(Fig. 30 a, c, e)**. The changes occurring during processing between raw (homogenates) and processed (purees) products were illustrated on the first factorial axis (F1) for the NF samples (with 97.2% specificity and 98.6% sensitivity) and AIS materials (100% specificity and sensitivity), and on the second factorial axis (F2) for FD samples (100% specificity and sensitivity).

Combining the main discriminant coefficients of the FDA models separating raw and processed materials (F2 for NF and FD samples, F1 for AIS samples) **(Fig. 30 b, d, f)** and using the absorption band assignments described in literature, allowed to identify phenomena occurring during apple processing. In both NF and FD samples, highly consistent variations of spectral intensity were commonly found between 1800 and 1500 cm^{-1}, this region giving overlapped information related to pectins, proteins, phenolics and absorbed water (Kačuráková et al., 1999), detailed in the following section:

- The increase of the bands at 1750 cm^{-1} in NF **(Fig. 30b)**, 1788 cm^{-1} and 949 cm^{-1} in FD **(Fig. 30d)** were specific of C=O, C-O and C-C stretching vibrations of carboxylic acids and polysaccharides (Canteri, Renard, Le Bourvellec, & Bureau, 2019; Kyomugasho et al., 2015). These observations were in accordance with the increase of soluble fiber fractions and total polysaccharide contents after apple cooking (Colin-Henrion, Mehinagic, Renard, Richomme, & Jourjon, 2009).

- The bands at 1610-1620 cm^{-1} (1614 cm^{-1} in NF; 1618 cm^{-1} in FD) have been reported to correspond to the vibration of C=O from protein or pectic acid ester (Abidi, Cabrales, & Haigler, 2014). These peaks were consistent with the aforementioned pectic absorption peaks (1750 cm^{-1} and 1788 cm^{-1}), in accordance with the increase of pectin content in purees. In the same way, this absorbance displays the same variations

in a simplified experiment of apple cell wall (mainly soluble pectins) submitted to similar puree processing conditions (100°C for 20 min at pH 3.0) (Liu, 2019). In addition, the negligible concentration of proteins in fresh and processed apples (0.17-0.57 g/100 g FW) limited the hypothesis concerning the protein change during apple processing (U.S. Department of Agriculture, Agricultural Research Service , 2019).

- the strong decrease of bands near 1630 cm^{-1} and 1560 cm^{-1} could be attributed to the degradation of phenolic compounds during processing. These bands have been already identified to quantify the polyphenol contents in freeze-dried apples (Bureau, Ścibisz, Le Bourvellec, & Renard, 2012).

- the specific bands of soluble acids (1712 cm^{-1} in NF, 1718 cm^{-1} in FD) (Clark, 2016) and of sugars (fructose at 1084 cm^{-1} and 1061 cm^{-1}; sucrose at 1113 cm^{-1}) (Bureau, Cozzolino, & Clark, 2019), which have been validated with standard chemicals in ATR-FTIR, could partially contribute to the dynamics of puree changes. These spectral variations relating the decreases of acid contents and increases of fructose at 1712 cm^{-1} were also in line with the results of chemical measurements **(Table 25 and Table 26).**

In cell wall materials (AIS), two negative peaks at 1100 cm^{-1} and 984 cm^{-1} **(Fig. 30f)**, could be attributed to the solubilization of the cell wall pectins after thermal processing (Coimbra, Barros, Barros, Rutledge, & Delgadillo, 1998; Kacurakova, Capek, Sasinkova, Wellner, & Ebringerova, 2000), consistent with the acid hydrolysis and β-elimination of pectins depolymerization while apple processing (Le Bourvellec et al., 2011). Conversely, two positive peaks at 1595 cm^{-1} and 1030 cm^{-1} could be linked to the increase of lignin (Garside & Wyeth, 2003) and cellulose contents (Fasoli, et al., 2016; Schulz & Baranska, 2007) in cell wall materials. A possible explanation is the depolymerization of cell wall polysaccharides (mainly pectins) during maturation resulting in a relative enrichment of lignin and cellulose in comparison with pectins after apple processing.

ATR-FTIR detected the processing changes from raw apples to purees by scanning fresh, freeze-dried and cell wall samples. Particularly, spectra of fresh and freeze-dried samples, i.e. with or without water, provided highly consistent information on internal soluble matters (sugars, acids, pectins and phenolics). Concerning the cell wall

depolymerization (mainly pectin solubilization and galactose loss), these changes could be detected only by scanning the cell wall materials (AIS), thus highlighting the solubilization of pectins diffusing from pulp to serum (Burgy et al., 2018; Sila et al., 2009).

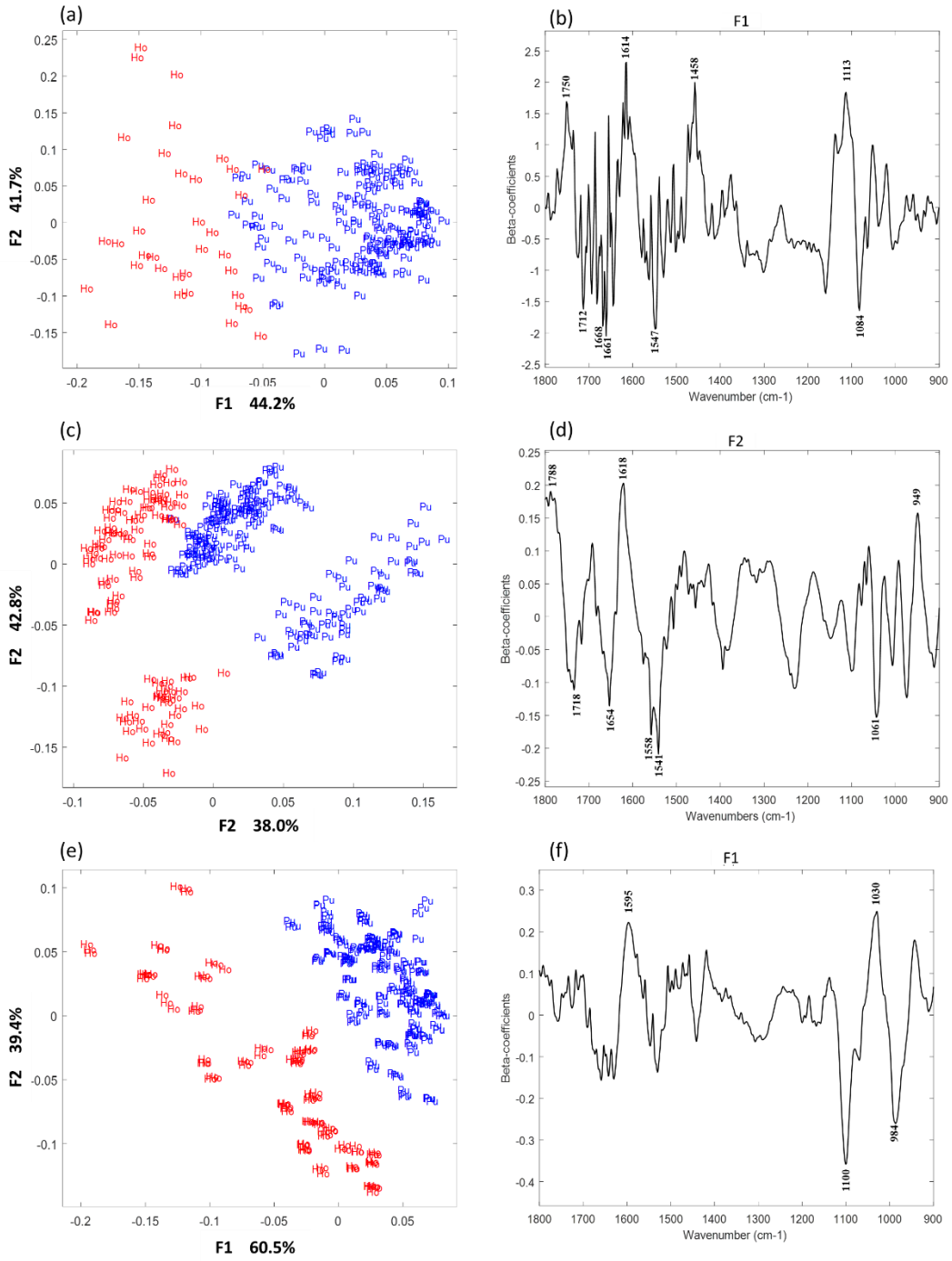

Fig. 30. Maps of Factorial Discriminant Analysis (FDA) performed on the SNV-pre-treated ATR-FTIR spectra (900-1800 cm-1) of all fresh apple homogenates (named 'Ho') and the corresponding processed purees (named 'Pu') with: **(a)** fresh samples

('NF'), **(c)** freeze-dried samples ('FD'), **(e)** cell wall samples ('AIS'); **(b)** the second factorial score ('F2') of fresh samples, **(d)** the second factorial score ('F2') of freeze-dried samples ('FD'); **(f)** the first factorial score ('F1') of cell wall samples.

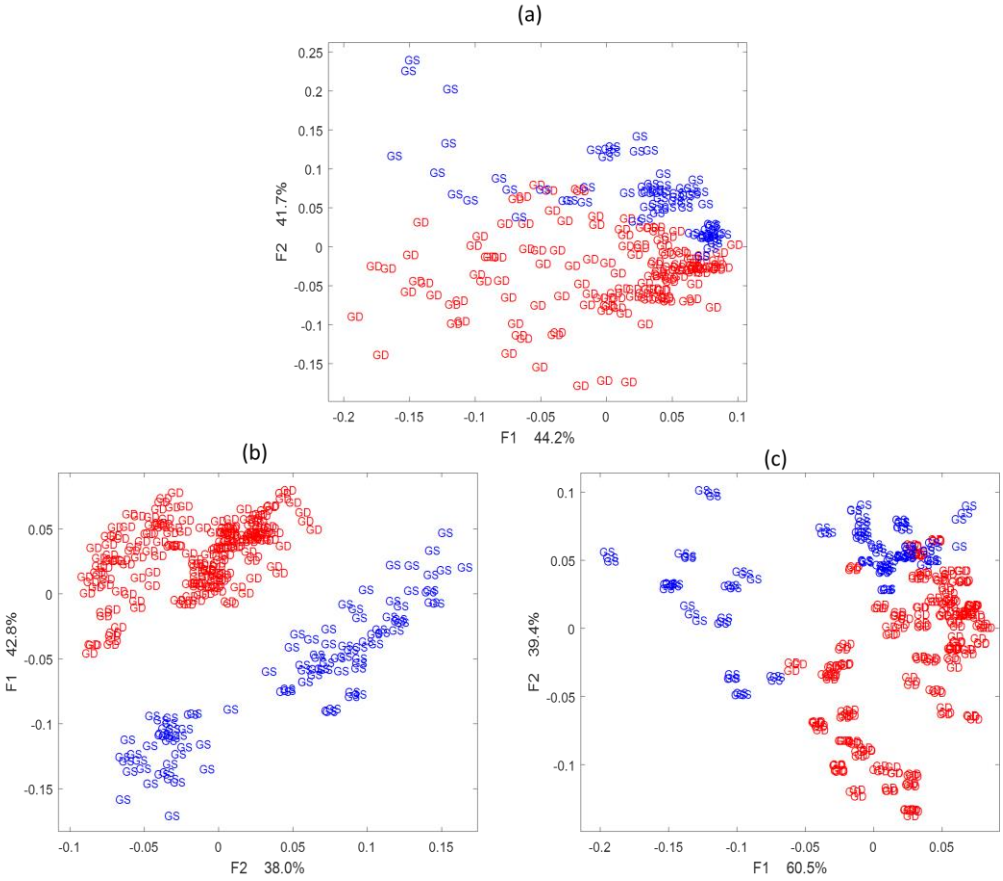

Fig. 31. Maps of Factorial Discriminant Analysis (FDA) performed on the SNV-pre-treated ATR-FTIR spectra (900-1800 cm^{-1}) of all homogenates and purees of Golden Delicious ('GD') and Granny Smith ('GS') with: **(a)** fresh samples ('NF'), **(b)** freeze-dried samples ('FD'), **(c)** cell wall samples ('AIS').

3.3 Prediction of quality traits: comparison according to sample forms

Acceptable to good predictions of SSC, TA. DMC, fructose and malic acid could be obtained on fresh (NF) and/or freeze-dried (FD) purees by ATR-FTIR, giving RPD from 3.1 to 5.2 (NF) and from 3.6 to 7.6 (FD) **(Table 28)**.

The prediction of global fruit quality traits, such as SSC and DMC, depended on two major spectral peaks, respectively, related to the sugars in NF (1061 cm^{-1}) (Bureau,

Cozzolino, & Clark, 2019) and to the acids in FD (1724 cm^{-1}) (Clark, 2016). In purees, the prediction accuracy of these two quality traits was similar in NF and FD samples with a R_{cv}^2 higher than 0.94 for SSC and higher than 0.89 for DMC. A good correlation between SSC and DMC in purees (R^2=0.78) and the similar related spectral signals used in models (mainly 1724 cm^{-1} and 1061 cm^{-1}) explained the good prediction of both SSC and DMC in NF and FD samples. For the third global quality trait, TA, its prediction was excellent with RPD higher than 6 in NF and FD samples. A particularly strong absorption at 1718 cm^{-1} was used in the TA models in both NF and FD samples.

Concerning the main individual sugars and acids (sucrose, fructose and malic acid), ATR-FTIR on FD samples provided more accurate prediction results (R_{cv}^2>0.87 and RPD>3.2) than on NF samples (R_{cv}^2>0.79 and RPD>2.3). For fructose and sucrose, the regression coefficients of the models showed numerous characteristic peaks in the region 1150-900 cm^{-1} in FD samples. But, despite the similar typical peaks, specific peaks such as 1034 cm^{-1} for sucrose and 1084 cm^{-1} for fructose were detected and used in their respective models. The lower RPD and the higher RMSECv in NF than in FD samples were due to the presence of water leading to a lower concentration of components and then a lower sensitivity to their variations. Moreover, to obtain the best prediction of sugars in fresh samples, the spectral region 1700-1550 cm^{-1} specific to soluble substances, was useful. In fresh samples, the linear models for TA, SSC, DMC and malic acid prediction depended foremost on the sugar absorption (fructose and sucrose), because of their relatively higher total concentrations (99.4-228.9 g/kg FW) than those of acids (TA: 25-109.1 meq H^+/kg FW). After freeze-drying, the specific spectral area (1725-1710 cm^{-1}) corresponding to acidity (Clark, 2016) became the main area of PLS models, due to their larger variations during storage than those of individual sugars.

Another quality trait of interest is the AIS contents, which contributes to the rheological properties of the processed apple purees products (Espinosa-Muñoz et al. 2012). The prediction of AIS contents is acceptable with RPD of 3.3 on FD purees, when expressed in dry matter (DW). Its prediction was not possible directly on NF purees. The significant signals at 985 cm^{-1} corresponding to CH stretching of cellulose

(Fahey, Nieuwoudt, & Harris, 2017) and at 1147 cm^{-1} for C-O-C vibration of glycosidic bound between uronic acids (Coimbra, Barros, Barros, Rutledge, & Delgadillo, 1998) were in line with the previous PLS models built to predict AIS yield in freeze-dried fruit and vegetables (Canteri, Renard, Le Bourvellec, & Bureau, 2019).

Briefly, ATR-FTIR technique worked well to evaluate global quality traits of interest in apple purees: SSC, TA and DMC. The prediction of cell wall contents (AIS) was possibility only on freeze-dried apple purees. Concerning the detailed composition including the individual components, the prediction was possible directly on fresh puree for malic acid whereas the prediction of the main individual sugars (fructose and sucrose) required the puree freeze-drying. The prediction of glucose was not acceptable in apple purees whatever the tested conditions.

Surprisingly, prediction was acceptable (R_{cv}^2 >0.87, RPD >3.1) for rheological parameters such as puree viscosity (η_{50} or η_{100}) and visco-elasticity (G', G'' in both amplitude and frequency sweep tests and yield stress) on FD samples with less than 10 LVs and was better than on NF and AIS samples **(Table 29)**. The single shear rate value at 50/s (η_{50}) has been described to be the best correlated with the in-mouth texture perception of fluid foods (Chen & Engelen, 2012). For the two parameters measured at η_{50} and η_{100}, predictions were better in FD samples than in NF and AIS samples. Particularly, two main spectral areas (1718 cm^{-1} and 1620-1595 cm^{-1}) in NF and FD samples appeared to be highly relevant to predict the puree viscosity. Differently, in AIS samples, the two major peaks (1018 cm^{-1} and 1110 cm^{-1}) linked to the viscosity prediction have been conventionally attributed to the pectin changes in fruit cell walls (Coimbra, Barros, Barros, Rutledge, & Delgadillo, 1998). For the specific viscoelastic parameters of purees (AS-G', AS-G'' and yield stress) by amplitude sweep tests, their prediction by ATR-FTIR was excellent in FD samples with RPD values higher than 3.4. The yield stress, corresponding to the moment when the puree starts to flow at the macroscopic level, could be predicted directly on NF purees with the better RPD and RMSECv than on FD samples. From frequency sweep tests (FS), the gel-like behaviors (FS-G' > FS-G'') of all purees could be well estimated in FD samples (R_{cv}^2> 0.90), even with a large variation of FS-G' and FS-G'' **(Table 27)**. Surprisingly, fresh NF samples

were the suitable sample type to evaluate the particle size, both d(4:3) and d(3:2), with a good performance of the PLS models (RPD>3.0).

Although acceptable results of PLS regression were obtained on the three sample types for the prediction of puree rheological properties (viscosity and viscoelasticity) and particle information (sizes and volume), it is worth signaling the differences of their fingerprint peaks: i) for fresh NF samples, the major region between 1750 and 1500 cm^{-1} was attributed to the absorbed water and complex soluble substances (pectins, polyphenols and proteins); ii) for cell wall AIS extracts, the typical peaks (1018 cm^{-1}, 1083 cm^{-1}) were mainly related to their pectic and phenolic variations; iii) for freeze-dried FD samples, the specific peaks, 1500-1750 cm^{-1} and 1200-900 cm^{-1}, combining with those observed separately in NF and AIS samples were used. The limited spectral sensitivity for the fresh suspensions (NF) and the restricted variations for the cell wall extracts (AIS) resulted in a less accurate prediction of the rheological behaviors than for freeze-dried FD samples. These results demonstrated the possibility of ATR-FTIR technique to accurately estimate viscosity, elasticity and the particle distributions directly on freeze-dried purees (FD). However, ATR-FTIR on fresh purees (FD) had a good ability to directly evaluate the particle size and properties (RPD>3.0), and also can probably to be used to evaluate the rheological behaviors (viscosity and viscoelasticity) according the results of RPD values over 2.5 (Nicolai et al., 2007).

4. Conclusion

As far as we know, this is the first report concerning the assessment of quality variations in fruit products during processing depending on ATR-FTIR spectral information of the same samples but characterized as fresh, freeze-dried and cell wall extracts. Direct spectral measurements on fresh samples could provide a reliable assessment of texture and major composition characteristics of purees. Thus, ATR-FTIR technique can be adapted to routine analysis in fruit industries, a simple method, using few steps for manufacturers. Long-time freeze-drying preparations still keep the stability and consistency of the ATR-FTIR signals in comparison with those of fresh samples, and provided more detailed assessments of rheological properties and cell wall

contents. ATR-FTIR on cell wall materials was the only way to identify the variations of cell wall compositions, but not enough to overview the changes during fruit processing.

Briefly, ATR-FTIR associated with suitable sample pre-treatments in fruit processing could offer sufficient information for the industrial and research demands. Balancing the pre-treated methods to stabilize samples and knowing the potential ability of infrared spectroscopy are both crucial for rapid and accurate analyses in fruit processing. Based on our results, future works could be extended to a wide span of complex processing strategies (drying, juicing, fermentation etc.) and/or operational units.

Table 25. Reference data of fresh apples including different varieties, agronomic conditions and times of a cold storage.

Variety	Year	Storage periods	Fruit thinning	AIS content mg/g FW	AIS content mg/g DW	DMC g/g FW	SSC °Brix	TA meq/kg FW	glucose g/kg FW	fructose g/kg FW	sucrose g/kg FW	malic acid g/kg FW
Granny Smith	2017	T0	Th -	38.49	256.57	0.151	11.10	97.83	21.35	43.56	34.45	6.04
		T1	Th -	26.73	182.19	0.149	11.80	103.43	21.59	26.11	40.51	6.09
		T3	Th -	24.30	169.68	0.145	11.57	85.97	23.15	29.19	36.41	4.60
		T6	Th -	23.02	161.21	0.143	10.77	60.90	21.64	50.41	30.06	3.47
SD				*6.5*	*39.8*	*0.004*	*0.5*	*17.3*	*1.48*	*11.3*	*6.1*	*1.2*
F-value and significance		Storage periods		*136.52* ***	*142.52* ***	*3.47* *ns*	*4.40* *	*115.27* ***	*0.91* *ns*	*3.16* *ns*	*1.88* *ns*	*63.30* ***
Golden Delicious	2017	T0	Th -	26.82	154.66	0.174	13.53	64.93	24.08	67.47	59.19	4.38
		T1	Th -	23.18	136.38	0.174	14.43	55.90	21.72	65.57	65.26	3.41
		T3	Th -	24.37	137.98	0.175	14.53	50.40	20.97	58.67	59.29	2.38
		T6	Th -	19.94	121.41	0.164	12.80	26.50	22.04	76.36	46.01	2.58
		T0	Th +	49.41	218.02	0.225	17.37	84.00	20.93	96.27	86.28	6.50
		T1	Th +	32.24	153.48	0.209	15.87	68.40	17.07	67.82	80.51	5.03
		T3	Th +	28.14	145.54	0.194	16.40	61.23	15.66	65.62	82.76	3.49
		T6	Th +	27.25	133.39	0.204	16.93	34.50	23.73	89.27	57.17	2.27
SD				*8.7*	*28.7*	*0.021*	*1.6*	*17.9*	*3.1*	*13.4*	*14.2*	*1.4*
F-value and significance		Storage periods		*117.99* ***	*100.28* ***	*21.19* ***	*4.41* *ns*	*315.68* ***	*14.07* ***	*19.20* ***	*45.97* ***	*227.70* ***
		Fruit thinning		*332.73* ***	*97.66* ***	*535.29* ***	*432.74* ***	*146.14* ***	*22.40* ***	*28.10* ***	*159.05* ***	*160.52* ***

Data expressed in Fresh weight (FW) or Dry matter weight (DW); values correspond to the mean of 3x10 apples. Four storage periods at 4°C: from harvest (T0), one (T1), three (T3) and six months (T6). Two conditions of fruit load: non-thinning with 100% number of apples (Th-) and thinning with 50% number of apples (Th+) per tree. Two refining conditions of purees with 0.5 mm (Ra) and non-refining (NR). In grey, one way- ANOVA results of Granny Smith apples and two-way ANOVA results of Golden Delicious apples. ns, *, **, ***: Non significant or significant at $P < 0.05$, 0.01, 0.001 respectively.

Table 26. Chemical and biochemical data of processed purees including different varieties, agronomic conditions and times of a cold storage.

Variety	Year	Storage periods	Fruit thinning	Refining	AIS content mg/g FW	AIS content mg/g DW	DMC g/g FW	SSC °Brix	TA meq/kg FW	glucose g/kg FW	fructose g/kg FW	sucrose g/kg FW	malic acid g/kg FW
Granny Smith	2017	T0	Th -	NR	45.2	267.6	0.169	10.5	101.5	20.3	50.7	58.0	8.5
		T1	Th -	NR	31.5	185.3	0.170	11.5	98.1	18.5	51.0	60.6	7.5
		T3	Th -	NR	24.2	147.1	0.164	11.3	86.5	23.0	47.3	46.6	6.1
		T6	Th -	NR	29.3	177.4	0.165	10.6	63.6	21.4	43.0	35.7	5.2
		T0	Th -	Ra	20.5	126.1	0.164	10.4	107.8	19.8	54.7	55.1	7.9
		T1	Th -	Ra	17.6	108.0	0.163	11.5	100.7	19.4	49.8	61.9	7.4
		T3	Th -	Ra	23.2	142.9	0.162	11.6	88.0	22.9	51.2	51.9	6.3
		T6	Th -	Ra	24.7	152.2	0.162	11.2	64.3	22.6	53.4	34.6	5.2
SD					*47.7*	*8.4*	*0.005*	*0.51*	*16.2*	*2*	*4.1*	*11*	*1.2*
F-value and significance			storage times		*20.4* ***	*27.6* ***	*0.6* *ns*	*28.5* ***	*619.5* ***	*13.4* ***	*3.4* *	*34.1* ***	*163.6* ***
			refining levels		*146.8* ***	*183.2* ***	*3.8* *ns*	*5.2* *	*14.2* **	*0.7* *ns*	*16.5* ***	*0.1* *ns*	*1.1* *ns*
Golden Delicious	2017	T0	Th -	NR	31.6	164.5	0.192	13.4	58.1	18.9	50.5	66.7	4.5
		T1	Th -	NR	27.6	147.2	0.188	15.0	54.4	15.4	49.4	59.1	2.8
		T3	Th -	NR	27.3	140.0	0.195	14.1	46.7	18.6	84.1	84.8	3.6
		T6	Th -	NR	27.6	145.7	0.189	13.8	26.8	23.0	85.1	77.3	2.7
		T0	Th -	Ra	18.3	126.6	0.197	12.6	59.0	24.5	45.2	43.5	3.2
		T1	Th -	Ra	20.8	113.8	0.183	14.5	48.6	17.4	57.4	67.3	3.3
		T3	Th -	Ra	22.3	116.0	0.192	14.1	48.5	19.1	79.0	88.5	3.6
		T6	Th -	Ra	21.3	118.7	0.179	13.8	27.0	22.1	82.9	65.3	2.7
		T0	Th +	NR	41.1	193.6	0.206	15.5	70.9	23.5	85.3	64.4	5.5
		T1	Th +	NR	31.9	150.8	0.213	17.6	69.3	16.8	80.3	115.9	5.6
		T3	Th +	NR	31.6	143.3	0.221	16.9	59.9	13.8	88.0	102.5	4.9
		T6	Th +	NR	34.8	150.3	0.231	17.5	34.7	23.8	95.7	44.0	3.6
		T0	Th +	Ra	24.8	117.3	0.208	16.5	70.6	14.8	82.3	65.3	5.8
		T1	Th +	Ra	25.4	121.8	0.208	16.5	71.4	17.8	81.3	118.6	5.9
		T3	Th +	Ra	25.9	118.1	0.219	16.8	60.1	14.4	88.5	108.6	5.2
		T6	Th +	Ra	29.5	124.0	0.238	17.4	36.0	24.5	97.0	57.3	3.7
SD					*18.5*	*5*	*0.019*	*1.7*	*15*	*3.8*	*16.6*	*24.1*	*1.2*
F-value and significance			storage times		*1.9* *ns*	*5.7* **	*21.4* ***	*20.0* ***	*415.9* ***	*28.9* ***	*57.6* ***	*64.4* ***	*28.5* ***

	refining levels	*140.9* ***	*170.1* ***	*8.6* **	*3.0* *ns*	*0.0* *ns*	*0.0* *ns*	*1.5* *ns*	*1.9* *ns*	*0.1* *ns*
	fruit thinning	*68.3* ***	*0.7* *ns*	*432.5* ***	*331.0* ***	*308.3* ***	*3.9* *ns*	*41.7* ***	*19.3* ***	*191.4* ***

Data expressed in Fresh weight (FW) or Dry matter weight (DW), values correspond to the mean of 3 puree lots (4kg of each). Four storage periods at 4°C: from harvest (T0), one (T1), three (T3) and six months (T6); two conditions of fruit load: non-thinning with 100% number of apples (Th-) and thinning with 50% number of apples (Th+) per tree; ns, *, **, ***: In grey, standard derivation (SD) of two puree varieties, two way- ANOVA results of Granny Smith purees and three-way ANOVA results of Golden Delicious purees. ns, *, **, ***: Non-significant or significant at $P < 0.05$, 0.01, 0.001 respectively. Cell wall content: ‘AY’; dry matter contents: ‘DMC’, Soluble solid contents: ‘SSC’; titratable acidity: ‘TA’.

Table 27. Rheological and textural of processed purees from different varieties, agronomic conditions and times of a cold storage.

Variety	Year	Storage periods	Fruit thinning	Refining	CSR (η50) Pa.s	CSR (η100) Pa.s	AS-G' (Pa)	AS-G" (Pa)	Yield point (Pa)	FS-G' (Pa)	FS-G" (Pa)	d 4:3 -	d 3:2 -
Granny Smith	2017	T0	Th -	NR	1.52	0.91	2980.6	786.1	31.0	2211.4	509.6	779.7	402.1
		T1	Th -	NR	1.29	0.84	2626.6	683.6	30.1	2021.2	467.2	814.2	416.7
		T3	Th -	NR	1.17	0.72	2160.5	540.1	14.2	1585.0	361.5	412.2	272.6
		T6	Th -	NR	1.05	0.66	2356.0	583.1	15.0	1873.2	446.6	347.8	246.3
		T0	Th -	Ra	0.53	0.35	156.0	62.0	10.5	58.7	15.3	593.3	301.7
		T1	Th -	Ra	0.34	0.26	108.6	35.7	5.6	102.7	25.7	604.7	294.9
		T3	Th -	Ra	0.79	0.48	1341.9	324.0	9.2	1036.7	213.2	359.7	252.4
		T6	Th -	Ra	0.79	0.48	1729.1	394.7	11.4	1431.5	309.9	309.3	231.6
SD					*0.38*	*0.23*	*1039.4*	*264.3*	*9.2*	*795.7*	*185.3*	*190.7*	*67.4*
F-value and significance			storage times		*24.54* ***	*14.01* ***	*26.9* ***	*16.5* ***	*28.2* ***	*67.3* ***	*46.8* ***	*955.7* ***	*579.9* ***
			refining levels		*1242.57* ***	*1746.73* ***	*941.7* ***	*1161.7* ***	*290.7* ***	*1556.0* ***	*1257.1* ***	*360.1* ***	*658.3* ***
Golden Delicious	2017	T0	Th -	NR	1.28	0.77	3127.8	626.7	47.5	2754.5	388.2	909.9	251.5
		T1	Th -	NR	1.13	0.70	1960.2	466.7	21.9	1631.3	310.8	694.0	351.9
		T3	Th -	NR	0.87	0.55	1849.0	453.0	13.9	1360.7	311.7	339.8	205.9
		T6	Th -	NR	0.92	0.50	1816.0	427.0	14.0	1506.3	330.5	316.1	223.6
		T0	Th -	Ra	0.08	0.32	10.5	5.0	0.7	3.1	1.1	709.9	151.5
		T1	Th -	Ra	0.71	0.42	1241.5	274.6	13.5	938.5	189.9	539.9	272.5
		T3	Th -	Ra	0.66	0.40	1376.6	308.1	10.9	1059.9	218.9	299.5	191.6
		T6	Th -	Ra	0.75	0.39	1608.0	361.4	14.5	1192.1	256.7	284.4	210.0
		T0	Th +	NR	1.75	0.97	3375.1	816.4	52.1	2759.3	475.9	831.6	231.6
		T1	Th +	NR	1.54	0.94	2783.7	639.5	25.2	2102.3	432.9	489.0	261.8
		T3	Th +	NR	1.25	0.70	2517.6	609.0	22.3	1794.1	401.2	405.1	228.3
		T6	Th +	NR	1.60	0.88	3168.2	751.7	33.9	2372.8	522.7	393.5	255.1
		T0	Th +	Ra	0.09	0.44	10.5	5.0	0.7	3.1	1.1	681.6	171.6
		T1	Th +	Ra	0.99	0.51	1852.7	406.8	17.7	1433.8	278.1	416.9	236.1
		T3	Th +	Ra	0.90	0.51	1978.3	445.8	17.9	1382.6	280.9	339.1	209.7
		T6	Th +	Ra	1.15	0.71	2497.0	532.2	28.8	1997.0	394.9	353.8	239.1
SD					*0.47*	*0.21*	*981.0*	*226.0*	*14.2*	*796.4*	*145.7*	*198.3*	*46.3*
F-value and significance			storage times		*13.33* **	*12.25* ***	*4.84* **	*5.48* **	*4.48* *	*11.17* ***	*7.71* ***	*142.98* ***	*18.53* ***

	refining levels	*20.04* ***	*44.40* ***	*11.26* ***	*15.83* ***	*5.74* *	*6.65* *	*17.19* ***	*1.70* *ns*	*0.15* *ns*
	fruit thinning	*60.10* ***	*93.17* ***	*41.93* ***	*57.71* ***	*24.00* ***	*39.40* ***	*68.09* ***	*28.92* ***	*24.43* ***

Data expressed in Fresh weight (FW) or Dry matter weight (DW), values correspond to the mean of 3 puree lots (4kg of each). Four storage periods at 4°C: from harvest (T0), one (T1), three (T3) and six months (T6); two conditions of fruit load: non-thinning with 100% number of apples (Th-) and thinning with 50% number of apples (Th+) per tree; ns, *, **, ***: In grey, standard derivation (SD) of two puree varieties, two way- ANOVA results of Granny Smith purees and three-way ANOVA results of Golden Delicious purees. ns, *, **, ***: Non-significant or significant at $P < 0.05$, 0.01, 0.001 respectively.

Table 28. Prediction of apple processed purees composition using the leave-one-out PLS regression based on the fresh ('NF') and freeze-dried ('FD') ATR-FTIR spectra and reference data.

				Leave-one-out PLS (n=72)				
Parameter	Sample	Range	SD	R_{cv}^2	RMSECv	LVs	RPD	Linkable regions (cm^{-1})
SSC (°Brix)	NF	10.3-18.6	2.4	0.94	0.6	4	4.1	1055-1065, 1028-1030, 1558-1562, 1649-1653
	FD			0.95	0.5	3	4.9	1058-1065, 1724-1735, 998-1001
Sucrose (g/kg FW)	NF	32.2-123.1	24.2	0.79	10.5	8	2.3	1084-1095, 1030-1034, 1574- 1583, 1225-1229, 916-920, 998-1102
	FD			0.87	7.8	7	3.2	998-1001, 1080-1084, 1030-1034, 1124-1137, 998-1102
Glucose (g/kg FW)	NF	13.5-25.7	3.4	0.65	2.0	9	1.7	1720-1715, 1656-1645, 1539-1562, 1886-1753, 1163, 1067, 1015
	FD			0.70	1.8	6	1.9	1028-1034, 1578-1570, 1010-1015, 1420- 1397, 1079, 985-998
Fructose (g/kg FW)	NF	40.0-99.9	18.9	0.88	6.0	8	3.1	1635-1655, 1078-1086, 1028-1034, 987-998, 1137-1142
	FD			0.90	5.3	6	3.6	1082-1090, 1030-1034, 987-989, 926-928, 1061-1665, 1035-1046
TA (meq/kg FW)	NF	25.0-109.1	22.8	0.97	3.8	4	6.0	985-998, 1084-1095, 1715-1730, 1695-1701
	FD			0.98	3.0	3	7.6	1716-1724, 987-989, 962-968
Malic acid (g/kg FW)	NF	2.35-8.97	1.63	0.91	0.5	4	3.3	1082-1095, 995-1001, 1715-1730, 1539
	FD			0.94	0.4	5	4.3	1716-1733, 1541-1558, 1695-1705, 1022-1024
DMC (g/g FW)	NF	0.16-0.24	0.03	0.89	0.01	6	3.1	1055-1068, 1443-1430, 1113-1135, 965-978, 1741-1730
	FD			0.92	0.01	5	3.6	1710-1728, 1541-1558, 1514-1507
AIS content (mg/g DW)	NF	100.4-271.7	33.3	0.75	16.9	10	1.9	1665-1685, 1701-1718, 1113-1128, 962-968, 1548-1560, 1605-1620
	FD			0.88	10.1	7	3.3	1142-1150, 985-995, 1058-1065, 1058, 995-1005, 1650-1665
AIS content (mg/g FW)	NF	16.5-48.9	6.1	0.76	3.5	9	2.0	1655-1685, 1605-1620, 1665-1685, 1700-1722, 965-985, 1094-1105
	FD			0.83	2.3	8	2.7	1055-1065, 985-995, 1030-1035, 1142-1150, 1165-1193, 1096-1101

Puree spectra and reference data from two varieties ('Granny Smith', 'Golden Delicious') with different thinning conditions, a cold storage (during 0, 1, 3 and 6 months) and two puree refining conditions. Spectral area: 1800-900 cm^{-1} and spectrum pre-processing: baseline-correction and SNV. They are same in Table 29.

Table 29. Prediction of apple processed purees rheological parameters and textural properties using the leave-one-out PLS regression based on the fresh (NF), freeze-dried (FD) and cell wall (AIS) ATR-FTIR spectra and reference data.

				Samples (n=72)				
Parameter	Sample	Range	SD	R_{cv}^2	$RMSEC_V$	LVs	RPD	Linkable regions (cm^{-1})
	NF			0.84	0.18	8	2.5	1620-1635, 1662-1670, 1718-1726, 1110-1122, 1080-1109, 1450-1456
η50	FD	0.69-1.94	0.44	0.88	0.14	9	3.1	940-952, 1060-1065, 1455-1471, 925-935, 1078-1084, 1145-1150, 1718-1726
	AIS			0.86	0.16	8	2.8	1018-1023, 1110-1115, 1160-1168, 1057-1083, 925- 935, 1618-1625
	NF			0.83	0.09	8	2.5	1610-1620, 1718-1726, 1560-1584, 1080-1110, 1450-1456
η100	FD	0.25-1.06	0.21	0.89	0.06	9	3.4	940-952, 1060-1065, 1150-1161, 1455-1471, 1020-1038, 983-995,
	AIS			0.84	0.08	9	2.6	1018-1023, 1092-1110, 924- 935, 1057-1083, 1610-1625, 946-958
	NF			0.82	425	10	2.4	1645-1665, 1047-1055, 1082-1088, 1450-1456, 1530-1547, 925-932,
AS-G' (Pa)	FD	6-3612	1001	0.88	297	9	3.4	1020-1036, 1618-1635, 1060-1065, 1455-1471, 1084-1090, 983-995
	AIS			0.85	332	9	3.0	1610-1625, 1078-1113, 1018-1023, 924- 935, 1039-1043, 1193-1216
	NF			0.83	98	9	2.5	1530-1547, 1456-1464, 1645-1665, 1080-1088, 1610-1618, 925-932
AS-G'' (Pa)	FD	2-860	234	0.89	69	10	3.4	1015-1030, 1060-1068, 930-944, 1084-1090, 1465-1482, 1624-1643
	AIS			0.86	72	9	3.1	1018-1023, 1078-1110, 1560-1584, 1610-1625, 924-935, 1193-1216
	NF			0.86	4.4	9	2.9	1082-1088, 1530-1547, 1686-1699, 1030-1043, 1610-1618, 1090-1111,
yield stress	FD	0.6-57.6	12.9	0.87	4.2	9	3.1	984-992, 1463-1470, 1048-1054, 935-944, 1142-1151, 1465-1482, 1090-1104
	AIS			0.82	4.9	9	2.6	1039-1056, 1018-1023, 1078-1110, 946-958, 924- 935, 1610-1625
	NF			0.84	303.5	8	2.6	1645-1665, 1530-1549, 1456-1464, 1610-1620, 1058-1063
FS-G' (Pa)	FD	0.3-3105.6	798.2	0.90	217.6	10	3.3	946-955, 1015-1030, 1455-1471, 1090-1104, 1060-1068, 1612-1620
	AIS			0.84	292.4	8	2.5	1018-1023, 1610-1625, 1092- 1110, 912-930, 1039-1056
	NF			0.82	63.3	10	2.5	1645-1665, 1456-1464, 1530-1549, 1685-1695, 1058-1063, 1610-1618,
FS-G'' (Pa)	FD	0.3-511.1	158.7	0.91	48.1	8	3.3	937-949, 1060-1068, 1455-1471, 1011-1028, 1455-1462, 1092-1104
	AIS			0.87	56.1	10	2.9	1018-1023, 1570-1584, 1528-1542, 1092-1110, 1610-1625, 912-924
	NF			0.90	59	9	3.3	1701-1710, 1655-1668, 1034-1038, 1718-1726, 986-995, 1534-1541,1145-1152
d (4:3)	FD	277-920	195	0.93	53	9	3.5	934-949, 1464-1482, 1540-1558, 1050-1056, 915-920, 1740-1765
	AIS			0.87	65	8	3.0	1045-1083, 1502-1516, 1059-1067, 956-980, 1605-1615
	NF			0.86	21	10	3.0	1146-1158, 1034-1038, 1405-1412, 1082-1119, 1560-1597, 986-995, 1730-1742
d (3:2)	FD	132-422	64	0.85	23	10	2.8	1027-1039, 1056-1065, 1110-1124, 915-939, 1008-1015, 1625-1648
	AIS			0.81	26	9	2.3	974-995, 1018-1023, 1235-1256, 1045-1083, 1727-1735, 1605-1615

Highlights of Paper II

This study answered our second question in Part 1:

What is the most appropriate sampling method for spectroscopic analyses to evaluate efficiently the puree variability and quality?

- Fresh and freeze-dried samples presented similar MIR fingerprint spectral variations due to processing.
- MIR on fresh purees predicted well particle size and volume (RPD > 3.0) affecting texture.
- MIR on freeze-dried purees improved assessment of chemical composition (RPD > 3.2).
- MIR on freeze-dried purees could assess viscosity and viscoelasticity (RPD> 3.1).
- MIR on cell wall extract could highlight their evolution during processing.

Paper III (Published)

A large heterogeneity of apples has been pointed out in relation with genetic, ripening, agricultural practice, environment and climate change (see **II. Literature review, Part 2**). This could be a limitation for the use of NIR spectroscopy to predict fruit quality. Thus, using efficient and rapid methods to know more about the distribution of components in apples could help researchers, field growers or industrial manufacturers to determine where and how many infrared measurements are needed to optimize the quality prediction and fruit sorting.

A rapid and effective method, using NIR-HIS, was tested to explore the internal biochemical heterogeneity in single apple fruits of four varieties (**Paper III**). First, NIR-HSI images were acquired on the 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 only 141 representative cylinders from the total dataset (1056 samples), in which the DMC, total sugars content (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 intended to describe the distribution of these components in individual apples.

Contents lists available at ScienceDirect

Postharvest Biology and Technology

journal homepage: www.elsevier.com/locate/postharvbio

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

1. Introduction

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

Consequently, the main objective of this work was to provide a simple and efficient method to reduce the intensive reference measurements (contents of dry matter, total sugars, individual sugars, acids and polyphenols) in order to develop a HSI modelling

calibration and to evaluate the apple variability and heterogeneity.

2. Material and methods

2.1 Apple fruit

The experiment was conducted on four different apple varieties: 'Golden Delicious' (GD), 'Granny Smith' (GS), 'Braeburn' (BR) and 'Royal Gala' (GA). In 2018, all apples were harvested in the experimental orchard at La Pugère (Bouches du Rhône, France). 'Braeburn', 'Granny Smith' and 'Royal Gala' apples were grown under a commercial fruit thinning (Th+, 50-100 fruits/ tree). 'Golden Delicious' apples were grown under two thinning conditions, the commercial fruit thinning (Th+, 50-100 fruits/ tree) and without thinning (Th-, 150-200 fruits/ tree). After the commercial harvesting ('Royal Gala' on August 28th, Golden Delicious on September 19th, 'Granny Smith' on September 20th, and 'Braeburn' on October 3rd), all apples 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 fruits × 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 fruits × 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. 32)**.

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 fruits × 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. 32**).

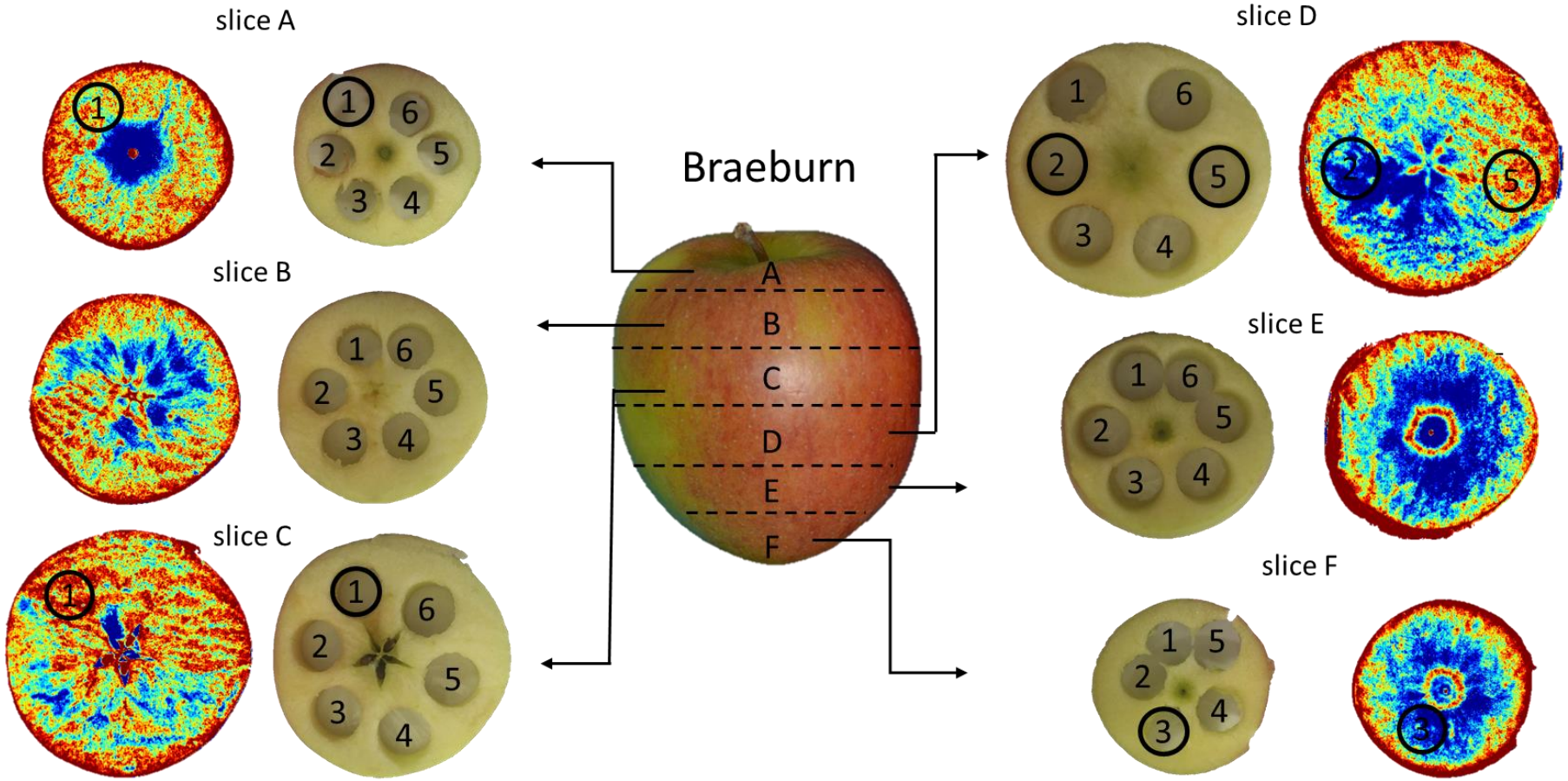

Fig. 32. The photographs of Braeburn apple slices and the first principal component (PC1) score plot of all near-infrared hyperspectral pixels (990-2450 nm) for each slice (A, B, C, D, E, F). The selected ROIs were labelled with black circles.

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 ***X*** (5 apple groups × 6 fruits × 6 slices × 10,000 rows by 258 columns). After pre-tests, matrix ***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 ***X*** to check the major components causing variability in the apples (**Fig. 33**). 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. 32**).

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 (**Fig. 33**). 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).

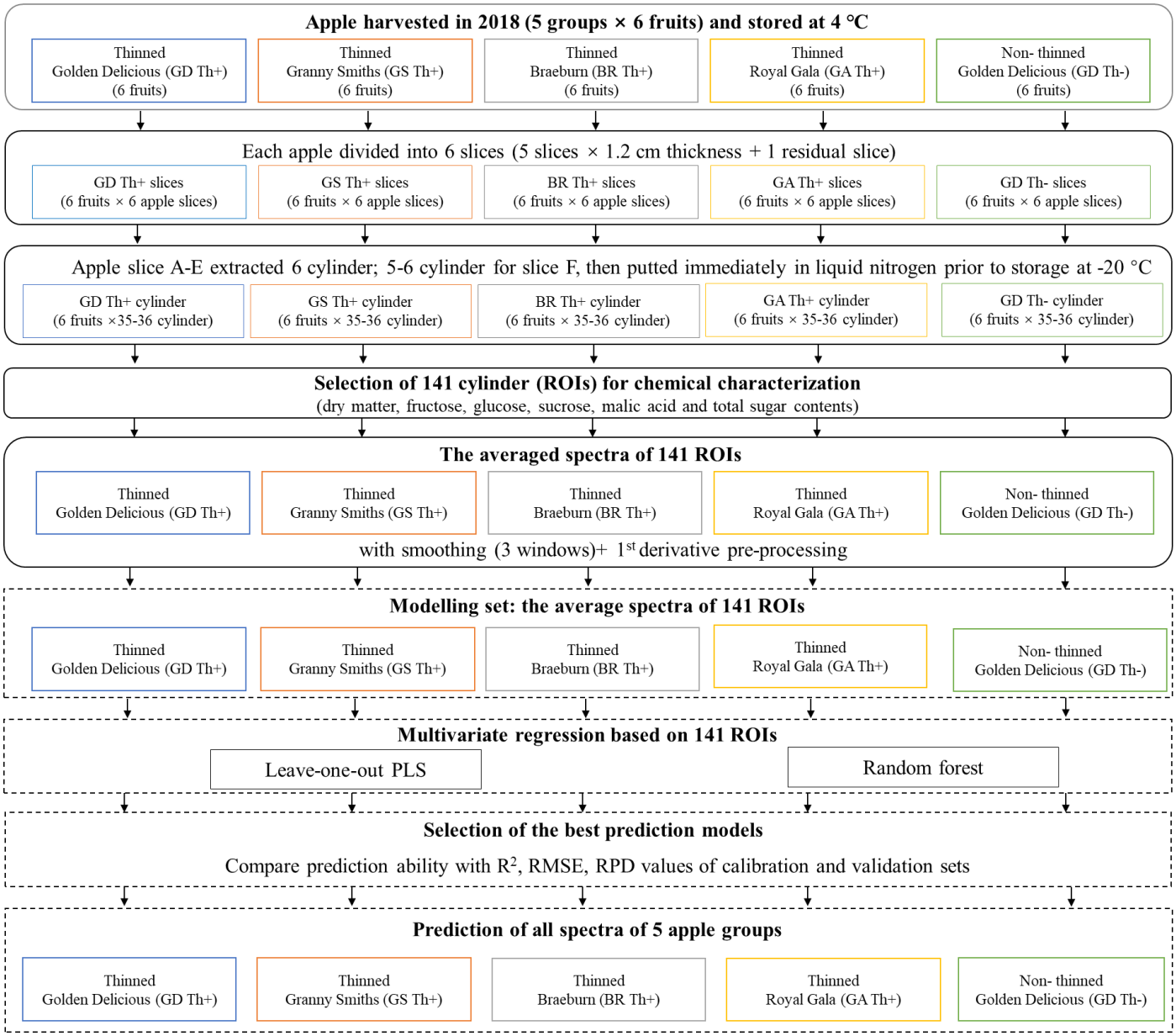

Fig. 33. Chemometric strategies for apple internal quality modelling and prediction.

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 ***X***) was able to discriminate the variability and heterogeneity of apple fruits between the top (slice A) and the bottom (slice F) (**Fig. 34a**). The first two principal components represented 68.0 % of the total variability, with the first component (PC1) of 43.8 % 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 (**Fig. 34b and c**): 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 fruits (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 (43.8 % 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. 32**). 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. 32**). Besides, the ROIs with clear color differences inside the middle slices (area No. 2 and No. 5 of slice D in **Fig. 32**) were also selected. 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.

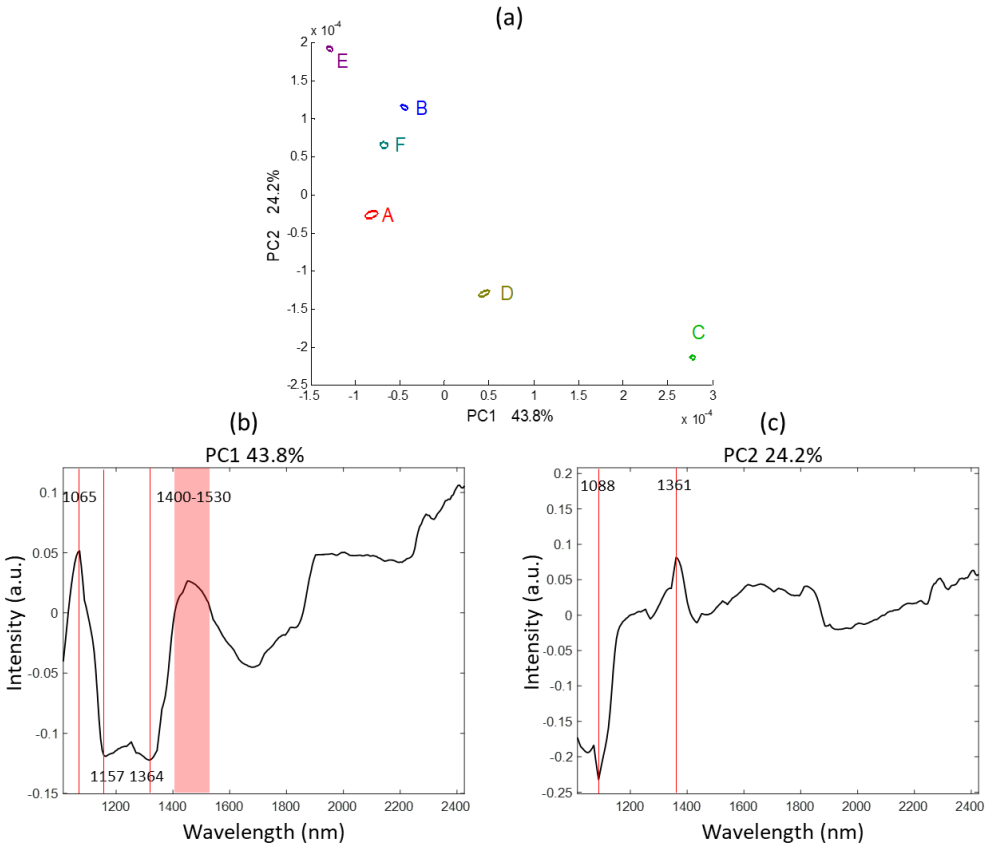

Fig. 34. PCA on the randomly selected HSI spectra acquired on all apple slices (1,0000 spectra of each apple slice, and totally 36 slices of 6 fruits) in the range from 990 to 2450 nm of all groups and the loading plots of six different apple slices from the top to the bottom (named ‘A’, ‘B’, ‘C’, ‘D’, ‘E’, ‘F’) on the first (PC1) and the second (PC2) principal components.

3.2 Chemical characteristics of ROIs

The boxplot of chemical reference data (**Fig. 35**) of the 141 selected ROIs showed a large variation of contents of dry matter, total sugars, malic acid and polyphenols in the different apple varieties.

Royal Gala apples had the most intensive variations of DMC among the five apple groups (**Fig. 35a**). Conversely, the lowest variations of DMC and of TSC (**Fig. 35b**) 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 varieties (**Fig. 35c**). Moreover, the contents of polyphenols varied a lot in each apple variety (**Fig. 35f**). 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 fruits (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 results observed during the 2017 harvested season (**Paper IV**). Interestingly, the tree thinning treatment, by increasing the individual apple growth potential, led to a lower variability of malic acid (**Fig. 35f**) and sucrose (**Fig. 35d**) 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.

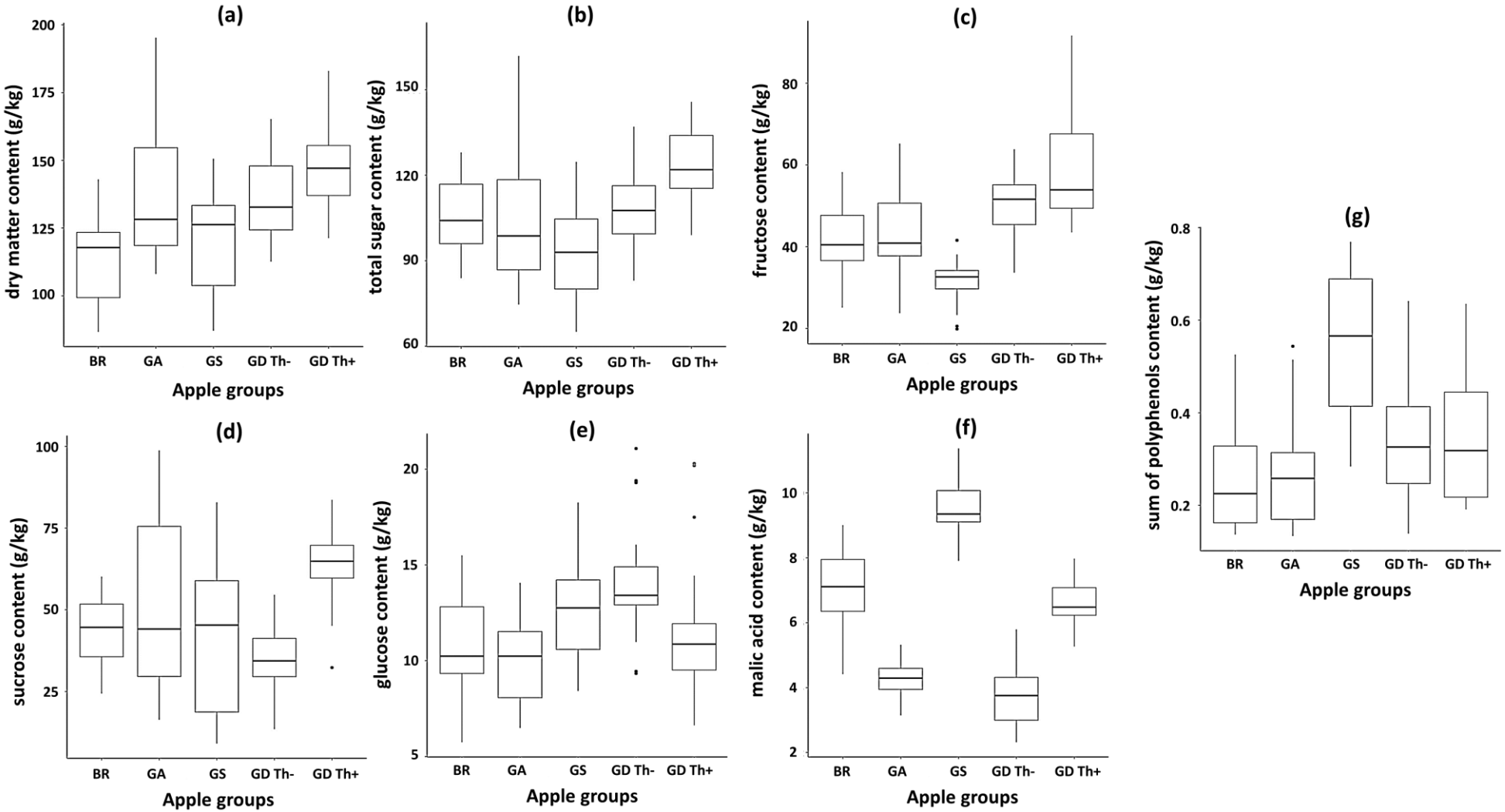

Fig. 35. The boxplots of: **(a)** dry matter, **(b)** total sugars, **(c)** fructose, **(d)** sucrose, **(e)** glucose, **(f)** malic acid, **(g)** sum of polyphenols of 'Braeburn' (BR); 'Granny Smith' (GS); 'Royal Gala' (GA); thinned 'Golden Delicious' (GD Th+) and non-thinned 'Golden Delicious' (GD Th-) apples.

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 30**). According to Nicolai et al. (2007), a RPD over 2 indicates the possibility of a coarse qualitative prediction of the internal attributes of fruits. The linear models (PLS) were much better than the random forest (RF) (**Table 30**), as described by Sun et al. (2017) to predict soluble solids content in melon fruits. 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. 36a, b)**. Moreover, the beta-coefficients showed strong positive or negative bands (**Fig. 36c, d**) 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 fruits (Giovanelli et al., 2014; Peirs et al., 2003c). 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.

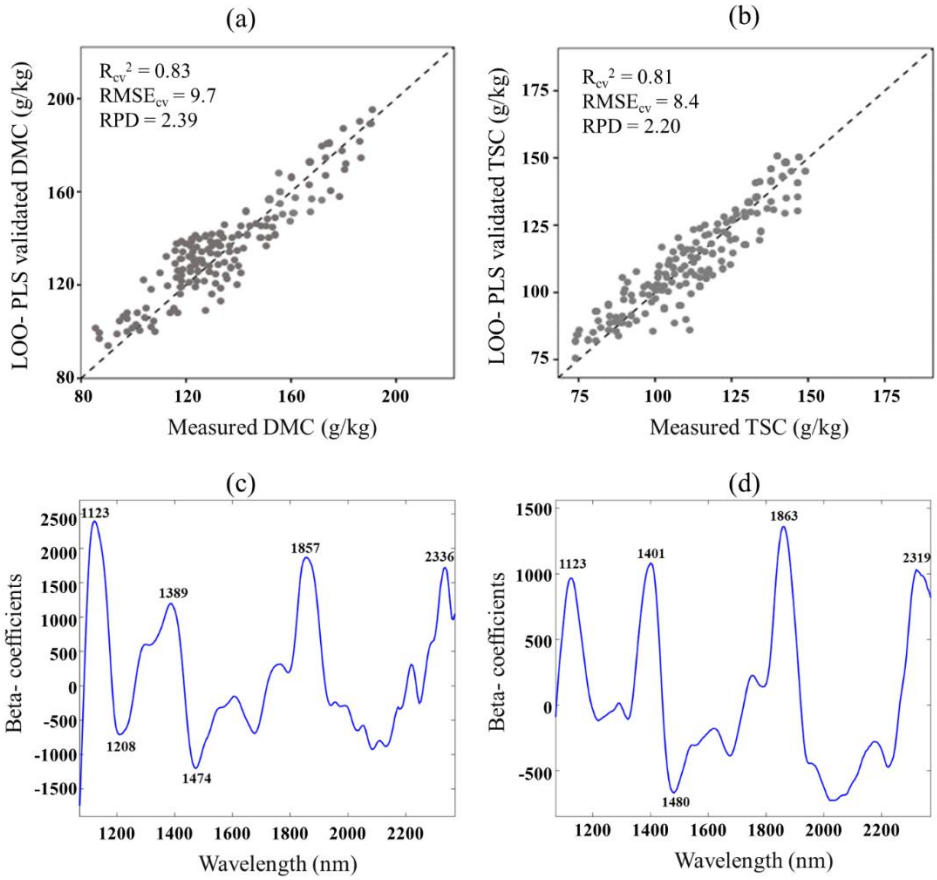

Fig. 36. Comparison of the measured and the full-cross validated **(a)** dry matter content (DMC) and **(b)** total sugars content (TSC) of the 141 ROI samples; and the most contributing wavelengths for **(c)** DMC and **(d)** TSC prediction, using the leave-one-out PLS regression on the ROI averaged spectra.

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 30**). 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~~s~~, such as dry matter and total sugars, but not of individual compounds (fructose, glucose, sucrose and malic acid) or microcomponents (sum of polyphenols).

Table 30. Leave- one- out partial least square (LOO-PLS) and random forest (RF) results of apple internal quality traits using the averaged 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

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 DMC (**Fig. 37**) and TSC (**Fig. 38**) 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 fruits 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., 2003a; 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 'variety', which is the major factor influencing the fruit heterogeneity, and the possible reason to explain the aforementioned disagreement. 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 varieties. 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 varieties = 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 varieties = 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 varieties = 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 fruits. NIRS information from top to bottom of apple surfaces therefore needs to be considered for all apple varieties. Apple heterogeneity along the direction from top to bottom presented the major internal chemical variations for all studied varieties.

Consequently, 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 (**Fig. 38**) and DMC (**Fig. 37**) in apple slices were observed in most apple slices of each variety (at least 4 over 6 fruits). These results provided an important opportunity to advance our knowledge on the quality measurement: where and how many specific positions need to be measured on apple surfaces with NIRS, in order to develop accurate and robust prediction models.

The previous HSI models mainly detected the soluble solids content and firmness changes in single fruit (Mo et al., 2017; Sun et al., 2017), because of the quick and easy reference data quantification of all targeted samples using digital refractometers and hardness detectors. Compared to these studies, our work provided an efficient solution for the HSI modelling calibration step, depending on the reference data measured on 141 representative samples instead of the 1056 prepared samples. Importantly, this method offered a new sight on contents of total sugars (sum of the fructose, glucose and sucrose) and dry matter in apples, with a limited number of complicated (individual sugars measured by spectrometry using enzymatic kits) and time-consuming (at least

24 hours for freeze-drying) analyses for HSI modelling. In future, such a rapid and efficient approach for HSI modelling calibration would be helpful to detect the variations of apple internal quality parameters according to different environmental conditions (crop load, irrigation, light penetration and elevations of regions etc.) and growing stages, and then contribute to an improvement of apple quality and production. The objective at the end could be to have a better knowledge of the apple homogeneity in order to manage them better for fresh market and processing taking into account the sustainability of practices.

4. Conclusion

In this study, the power of chemometric methods was harnessed in a two-steps 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~~s~~. 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 ~~It~~ opens the possibility to more systematically evaluate the fruit variability and heterogeneity in future projects.

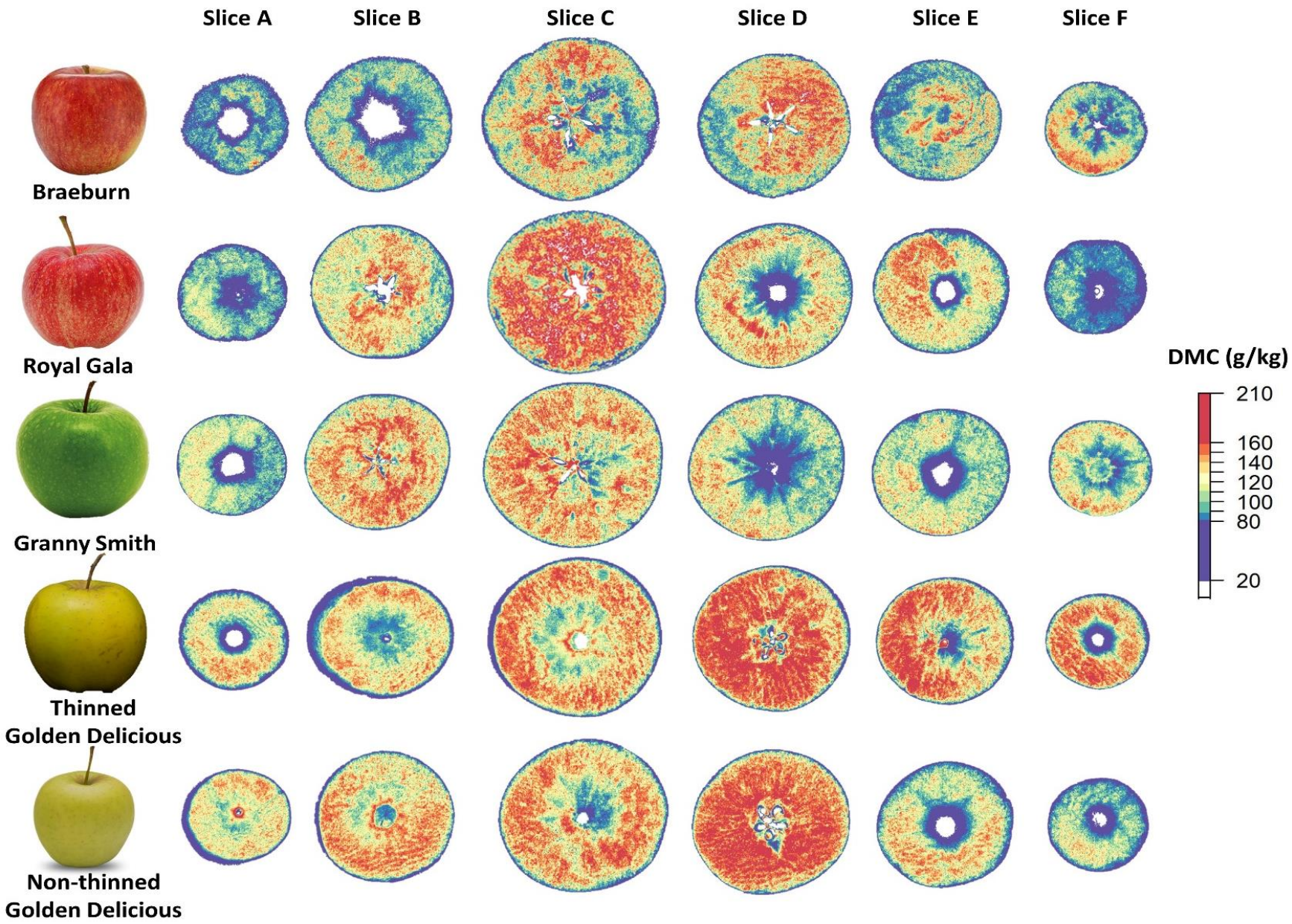

Fig. 37. The distribution of dry matter content (DMC) in apple slices predicted by the LOO- PLS models developed based on the ROI averaged spectra.

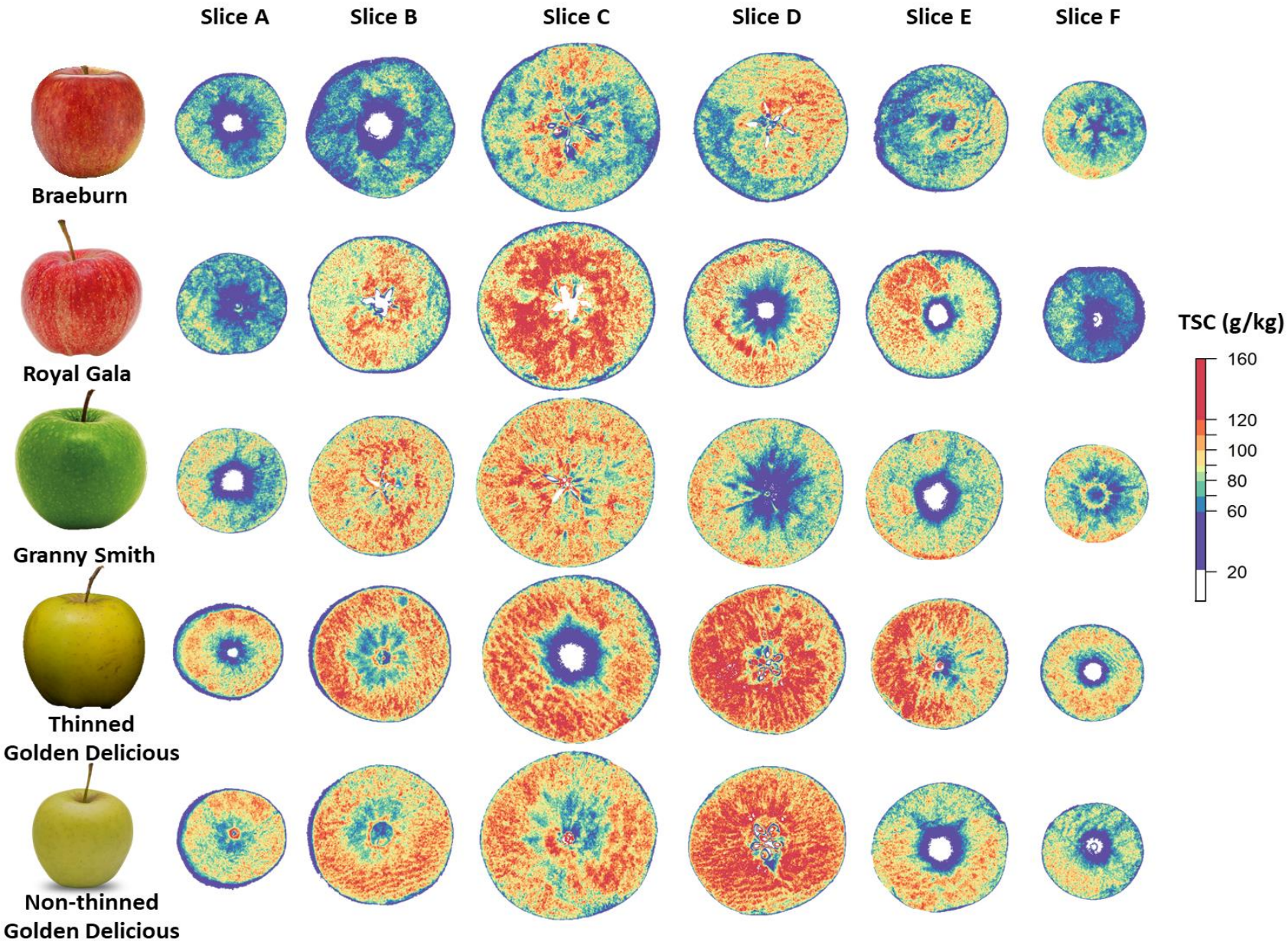

Fig. 38. The distribution of total sugars content (TSC) in apple slices predicted by the LOO- PLS models developed based on the ROI averaged spectra.

Highlights of Paper III

This study introduced a simple and efficient NIR-HSI approach to assess the apple heterogeneity. It reduced the need of numerous chemical characterizations to describe the distribution of quality traits within and between fruits. NIR-HSI could thus contribute to a better management of the fruit quality measurement.

Based on this method:

- PCA on the NIR-HSI of apple slices allowed to select the regions of interest (ROIs), i.e. representative parts for reference analyses, in apples selecting the more heterogeneous part regions (ROIs) in apples.
- A large heterogeneity of DMC, TSC, contents of fructose, glucose, sucrose, malic acid and polyphenols were highlighted on ROIs by spectrophotometry and chromatography.
- PLS models based on the selected ROIs from the total dataset successfully mapped the distribution of DMC (R^2 = 0.83, RPD = 2.39) and TSC (R^2 = 0.81, RPD =2.20) in each apple slice. The method did not allow to quantify the other quality traits: fructose, glucose, sucrose, malic acid and polyphenols.

Part 2. Analyzing the correlations between fresh and processed apples and predicting puree quality using spectral information of apples

From the results of **Part 1**, NIR and MIR techniques can assess a large variability of apple fruits or cooked purees, while some of the apple characteristics were directly linked to their processed products (see **Literature review** part). The challenge works in **Part 2** were to assess the possibility of predicting the properties of processed fruit products based on the raw fruit material spectral information, and so, to provide practical and suitable strategies to estimate the quality potential of fruits, to monitor their processing and to control the quality of fruit products.

To develop reliable and relevant models to predict cooked purees from spectra of raw apples, the following four questions needed to be answered:

- What are the internal correlations of physical, biochemical and textural properties between apples and purees, taking into account a large variability and heterogeneity?
- What are the spectral correlations between raw and processed apples?
- To what extent does the fruit variability impact the puree processed quality?
- Is it possible to predict processed purees' quality traits from the spectral information of corresponding raw apples?

The results of Part 2 are presented as three papers (Papers IV, V, and VI)

Paper IV (Published)

Up to now, in apple industry, manufacturers use their own experience and knowhow to make blend of apples in order to obtain the same purees over time. In **Paper IV**, an original work was performed on totally 960 apples of three varieties ('Golden Smoothee', non-thinned and thinned 'Golden Delicious', 'Granny Smith') and their 96 corresponding purees processed at harvest and after one, three and six months of cold storage. The potential of NIRS was studied to:

i) Evaluate the quality variations of both, apples and their corresponding purees and the correlations between them.

ii) Focus on the prediction of puree characteristics from the spectral information of raw apples.

As far as we know, this is the first report investigating the potential of NIR technique to predict the texture and taste of purees from only spectral data of raw apples.

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

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

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

[a] UMR408 Sécurité et Qualité des Produits d'Origine Végétale, INRAE, Avignon University, F-84000 Avignon, France
[b] INRAE-ONIRIS Unité Statistiques, Sensométrie, Chimiométrie (Stat SC), Nantes F-44322, France
[c] CEPIA Division INRAE, Nantes 44000, France

1. Introduction

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

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

quality assessment of apple purees have been mainly focused on the changes in polyphenol contents and total antioxidant activity (Loncaric, Dugalic, Mihaljevic, Jakobek, & Pilizota, 2014; Sukhonthara, Kaewka, & Theerakulkait, 2016), color (Oszmiański, Wolniak, Wojdyło, & Wawer, 2008), ascorbic acid (Picouet, Landl, Abadias, Castellari, & Viñas, 2009), organic acids (Bengoechea *et al.*, 1997), sugars (Keenan, Brunton, Butler, Wouters, & Gormley, 2011), polysaccharides (Le Bourvellec, Bouzerzour, Ginies, Regis, Plé, & Renard, 2011), rheological properties (Espinosa-Muñoz, Renard, Symoneaux, Biau, & Cuvelier, 2013; Espinosa-Muñoz, Symoneaux, Renard, Biau, & Cuvelier, 2012) and sensory appreciation (Espinosa-Muñoz, To, Symoneaux, Renard, Biau, & Cuvelier, 2011). However, almost all of these quality parameters have been measured through specific laboratory analyses, such as chromatography, which are time-consuming, expensive and not suitable for fast and numerous characterizations. Consequently, the development of rapid, accurate and reliable methods is required to control the quality of the raw apples and processed purees, and meet the ever-increasing demands for consistent and high quality fruit products.

Near infrared spectroscopy (NIRS) has been increasingly used for the safety inspection and quality assessment of agricultural products (Nicolai *et al.*, 2007). It has several advantages such as rapid spectrum acquisition, limited preparation requirements and no chemical waste, but it requires an initial calibration step, which is time consuming. Indeed, on a set of samples, representative of the expected variability, both NIRS spectra and their corresponding reference data are needed to establish predictive models using multivariate statistical and mathematical data analyses. Several parameters can thus be evaluated from a single spectrum, with varying precision. Intensive investigations using NIRS have been reported regarding the measurement of apple internal attributes in the past decades (Nicolai *et al.*, 2007). Satisfactory evaluation results are reported for soluble solids contents (Peirs, Tirry, Verlinden, Darius, & Nicolaï, 2003b), dry matter (McGlone, Jordan, Seelye, & Clark, 2003), titratable acidity (Liu & Ying, 2005), starch index (Menesatti *et al.*, 2009), chlorophyll content (Zude, Truppel, & Herold, 2002), firmness (Zude *et al.*, 2006), individual

sugars (Liu, Ying, Yu, & Fu, 2006) and antioxidant capacity (Schmutzler & Huck, 2016). Further, NIRS spectra are shown to classify apples according to varieties (Luo *et al.*, 2011), geographical origins (Bobelyn *et al.*, 2010) and postharvest storage periods (Camps, Guillermin, Mauget, & Bertrand, 2007; Giovanelli, Sinelli, Beghi, Guidetti, & Casiraghi, 2014).

Thus, NIRS can assess a diversity of quality traits in raw apples and processed products, while some of the puree characteristics are directly linked to those of the raw fruit. We therefore can suppose that NIRS spectra of the raw fruits could be used to predict the properties of the processed products, here purees, at least given a constant process operation. However, as far as we know, there is no literature related to the feasibility of using NIRS to evaluate the changes of apple puree properties and to trace back to their corresponding raw apple quality. The challenge here was to assess the possibility of predicting the properties of processed fruit products based on the raw fruit material spectral information, and so, to provide practical and suitable strategies to estimate the quality potential of fruits, to monitor their processing, and to control the quality of fruit products.

The specific objectives of our current work were to assess the potential of NIRS to 1) detect different factors such as variety, fruit thinning, storage period and mechanical puree refining on apples and/or their corresponding purees, 2) evaluate the quality traits of interest in both apples and the corresponding purees such as textural/rheological properties, soluble solids content, titratable acidity, dry matter content, insoluble solids content 3) and then establish the links between fruit materials before and after processing.

2. Materials and methods

2.1 Fruit materials

2.1.1 Apples

The experiment was conducted on three apple varieties: ‘Golden Smoothee’, ‘Golden Delicious’, and ‘Granny Smith’ during two subsequent harvesting seasons,

2016 and 2017, that are summarized in **Fig. 39**. In 2016, 240 ‘Golden Smoothee’ apples were harvested from the experimental orchard of INRA (Drôme, France). In 2017, 480 ‘Golden Delicious’ and 240 ‘Granny Smith’ apples were obtained from the experimental orchard at La Pugère (Bouches du Rhône, France).

Two fruit thinning levels were also compared during the ripening of ‘Golden Delicious’ apples in 2017. For this, trees were thinned 40 days after flowering (on June 2nd), and the treatment named Th+ corresponded to 50 to 100 fruits/tree (a standard commercial fruit load) and Th- to 150-200 fruits/tree (highly loaded trees). The ‘Golden Smoothee’ in 2016 and ‘Granny Smith’ in 2017 grew under regular fruit thinning (Th-). After harvesting (Golden Smoothee on September 14th, 2016, ‘Golden Delicious’ on September 11th, 2017 and ‘Granny Smith’ on September 25th, 2017), apples were kept in a cold storage chamber at 4°C and at around 90% of humidity during one, three and six months (respectively T1, T3 and T6), except the first group (T0) for which apples were analyzed and processed the day after harvest. These storage durations were chosen in order to increase the fruit variability linked to firmness and biochemical changes, such as demethylation and depolymerization of pectins (Billy, Mehinagic, Royer, Renard, Arvisenet, Prost, et al., 2008). Each set (T0, T1, T3 and T6) was divided into two sub-groups. The first one was dedicated for fresh apple characterization and the second one for processing. For characterization, three replicates of 10 apples, representative of the total apple set, were analyzed after one night temperature equilibration at 23°C.

First, nondestructive measurements (NIR, color, ethylene releasing rate, fruit weight), and then texture tests (puncture mean load and puncture linear distance) were performed on each apple. After that, each apple was cored and divided as described by Bureau *et al.* (Bureau, Scibisz, Le Bourvellec, & Renard, 2012) in order to create, for each replicate of ten apples, three batches of 40 pieces representative of each apple. The pieces were immediately put in liquid nitrogen to avoid any oxidization. Finally, one batch was stored at -20 °C and then was subsequently freeze-dried to evaluate the alcohol insoluble solids (AIS) contents. The other two batches were stored at -80°C for biochemical characterization measurement.

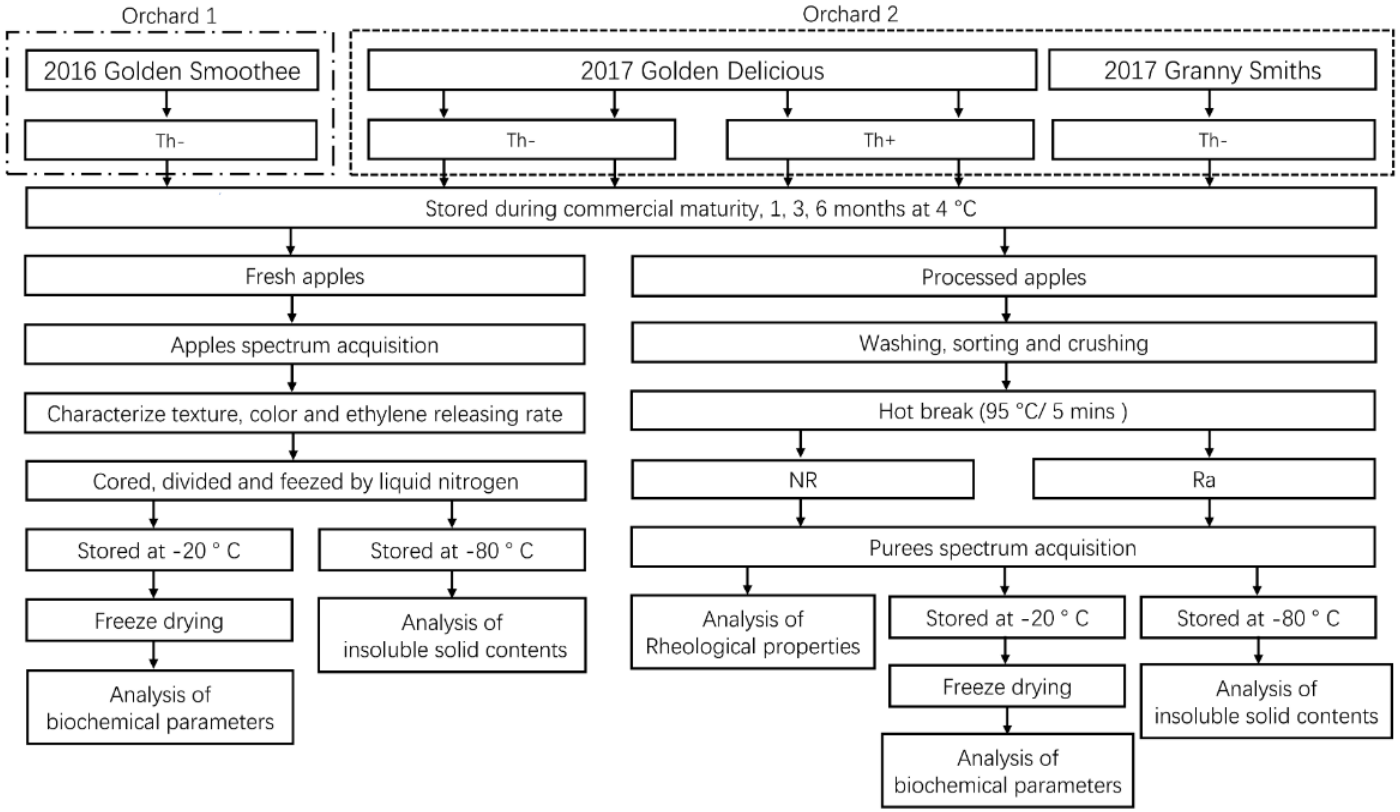

Fig. 39. Experimental scheme of sample preparation, spectrum acquisition and quality parameter characterization.

2.1.2 Purees

For each raw apple condition, three replicates of apple puree were produced with 4 kg of apples each. After sorting and washing, apples were cored and cut in 8 portions, then processed in a multi-functional processing system (Roboqbo, Qb8-3, Bentivoglio, Italy) following a Hot Break recipe: cooked at 95°C for 5 min at a 1500 rpm grinding speed, then cooled down to 65°C while maintaining the grinding speed. Half of the batch was refined at 0.5 mm using a Robot Coupe C80 automatic refiner (Robot Coupe SNC, Vincennes, France) in order to study two levels of granularity: non-refined (NR) and 0.5 mm refined (Ra). Finally, processed purees were conditioned in hermetically sealed cans, then placed and cooled at 23 °C before measurements, which took place the next day.

2.2 Determination of quality traits

2.2.1 Color

The skin color (un-blushed and blushed sides) was determined using a CR-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). The puree color was measured three times through a dedicated glass cuvette using the same method and equipment.

2.2.2 Ethylene production

Each group of ten apples was put in a hermetic jar at 23°C for 1 hour, and ethylene production was analyzed by taking 500 µL of the headspace and injecting it in gas chromatography (Agilent, California, United States) equipped with a porapak Q column and a FID detector and expressed in nmol kg^{-1} h^{-1}.

2.2.3 Fruit texture and puree rheology

Fresh apple texture was evaluated by a puncture test using a multipurpose texture analyzer (TAPlus, Lloyd Instruments, Farenham, UK). The puncture tests were operated with a punch probe (diameter 1.2 mm), which could penetrate up to a depth of

17 mm into each peeled section of apple. Firmness was then evaluated as the mean load value calculated by the division of penetration energy by the height of testing. Crunchiness of apple flesh (Gregson & Lee, 2002) was estimated as the linear distance values from the area under the force-distance curve in the range of 10 mm at the load plateau, consisting in summing the lengths between consecutive points.

The puree rheological measurements were carried out using a Physica MCR-301 controlled stress rheometer (Anton Paar, Graz, Austria) at 22.5 °C. The flow curves were performed after a pre-shearing period of 1 minute at 50 s^{-1} followed by 5 minutes at rest. The viscosity was measured at a rate of 1 point every 15 seconds, at a controlled shear rate range of [10; 250] s^{-1} on a logarithmic ramp. The value of the viscosity at 100 s^{-1} η_{100} was kept as an indicator of the puree viscosity. As often used to model fruit purees (Colin-Henrion, Cuvelier, & Renard, 2007), the complete flow curves were fitted with a power-law viscosity model as described by Eq. 1.

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

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

2.2.4 Biochemical characterization of apples and purees

Biological characterization of apples and purees has been described in **Papers I and II.**

2.3 FT-NIR spectrum acquisition

The spectral data of apples and purees were both acquired with a multi-purpose analyzer spectrometer (Bruker Optics®, Wissembourg, France), which provides diffuse reflectance measurements with a spectral resolution of 2 nm from 800 to 2500 nm. For each spectrum, 32 scans were recorded and averaged. The spectral acquisition and instrument adjustments were controlled by OPUS software Version 5.0 (Bruker Optics®). The apples were placed on an automated 30-positions sample wheel, each position corresponding to a measured area of 18 mm diameter. For each intact apple, two spectra were collected (on the blushed and un-blushed sides) though the 18 mm

diameter areas at 23°C. Puree were transferred into 10 mL glass vials (5 cm height x 18 mm diameter) which were placed on the automated sample wheel of the spectrophotometer. Each puree sample was measured three times on different aliquots. A reference background measurement was automatically activated before each data set acquisition using an internal Spectralon reference.

2.4 Statistical analyses and chemometrics

After ensuring the normal distribution of dataset, the results were presented as mean values and the data dispersion within our experimental dataset expressed as standard deviation values (SD). Analysis of variance (ANOVA) was carried out to determine the significant differences due to the tested factors on both apples and purees using XLSTAT (version 2018.5.52037, Addionsoft SARL, Paris, France) data analysis toolbox. The pairwise comparison between means was performed using Tukey's test at the 95% level of certainty ($p < 0.05$(*), 0.01 (**) and 0.001 (***)). For apples, a one-way ANOVA was applied to access the effect of storage period on 'Golden Smoothee' in 2016 and 'Granny Smith' in 2017; a two-way ANOVA concerned the effects of storage period and fruit thinning on 'Golden Delicious' in 2017. For purees, a two-way ANOVA accessed the effects of storage periods and refining treatments on 'Golden Smoothee' and Granny Smith, and a three- way ANOVA for 'Golden Delicious' in terms of fruit thinning, storage periods and puree refining levels. Pearson's determination coefficients (R^2) were calculated in order to study the significance of the relationship between apples and purees and then output as an heat map using R software (version 3.5.2) (R Core Team, 2018) and the additional package named 'ComplexHeatmap' (Gu, Eils, & Schlesner, 2016).

Spectral pre-processing and multivariate data analysis were performed with Matlab 7.5 (Mathworks Inc. Natick, MA) software using the SAISIR package (Bertrand & Cordella, 2008). All the NIR data were pre-processed with standard normal variate (SNV) and a derivative transform calculation (Savitzky–Golay method, gap size = 11, 21, 31, 41) of first or second order. Each of the preprocessing methods was tested in the discrimination models. As SNV pre-processing had the best performances to correct

multiplicative interferences and variations in baseline shift, the results shown are those obtained with the SNV pretreatment. Principal Component Analysis (PCA) and Factor Discriminant Analysis (FDA) were carried out on spectral data to evaluate the possibility to discriminate samples according to the tested factors (varieties, thinning and storage). The specificity and sensitivity values of FDA discriminations, which help for a better evaluation of the rate of sample differentiation, were calculated by the already reported method of Nargis (Nargis *et al.*, 2019). The Partial least-square (PLS) regression method was used to develop predictive models of the quality traits of interest in apples and purees. The whole spectral dataset included 840 spectra of apples and 240 spectra of purees. The dataset was randomly split, two third of dataset (560 spectra of apples and 160 spectra of purees) were used for calibration and one third of dataset (280 spectra of apples and 80 spectra of purees) for validation. The procedure was repeated 10 times in order to obtain the suitable dimensions of the PLS models. The latter performance was described by the root mean square error of calibration (RMSEC), the root mean square error of validation (RMSEV), the number of latent variables (LVs), the determination coefficient (R^2) between the predicted and measured parameters and the RPD (Residual Predictive Deviation) value as described by Nicolai (Nicolai *et al.*, 2007).

3. Results and discussion

3.1 Apple and puree characteristics measured by classical methods

3.1.1 Fresh apples

In this experiment, three varieties, two agricultural conditions and a cold storage for up to 6 months provided an interesting apple fruit variability (**Fig. 40a, b**). Clear discriminations were shown between the different storage periods along the first principal component, and apple varieties and thinning levels along the second principal component (**Fig. 40a**). 'Granny Smith' was clearly differentiated from the two 'Golden' varieties. Remarkably, the two non-thinned 'Golden' samples were close to each other (blue in 2016 and red in 2017), in spite of different growing seasons and locations,

while the thinned samples were clearly differentiated. The most discriminant quality traits were: mean load, linear distance, AIS content (FW and DW), TA, malic acid content, ethylene production rate and color changes (L*, a* and b*) on the first principal component, and SSC, DMC and sucrose content on the second principal component (**Fig. 40b**). The totality of the acquired data is presented in **Table 31**.

During cold storage, mean load, linear distance and AIS content (FW and DW) decreased remarkably ($p < 0.001$) in all apples, indicating an intensive reduction of apple firmness, crunchiness and cell wall material contents (Johnston, Hewett, & Hertog, 2002). Good correlations were observed between AIS and mean load (R^2 =0.78 in 'Golden Smoothee', R^2=0.75 in 'Granny Smith', R^2 =0.82 in 'Golden Delicious'). The acidity in all apples decreased significantly with storage ($p < 0.001$) at a large range from 103.4 to 26.5 meq/kg FW for TA and 6.5 to 2.3 g/kg FW for malic acid. The ethylene production rate increased and then decreased during storage with significant changes ($p < 0.001$). All color parameters (L*, a* and b*) increased clearly for all apples, linked to a degreening and a yellowing during the long-term storage. The changes of all individual sugar contents were significant in 'Golden Smoothee' ($p < 0.05$) and 'Golden Delicious' ($p < 0.001$), but not in 'Granny Smith' ($p > 0.05$). For 'Golden Delicious', SSC and DMC from thinned trees (Th +) appeared to be significantly ($p < 0.001$) higher than from non-thinned trees (Th-).

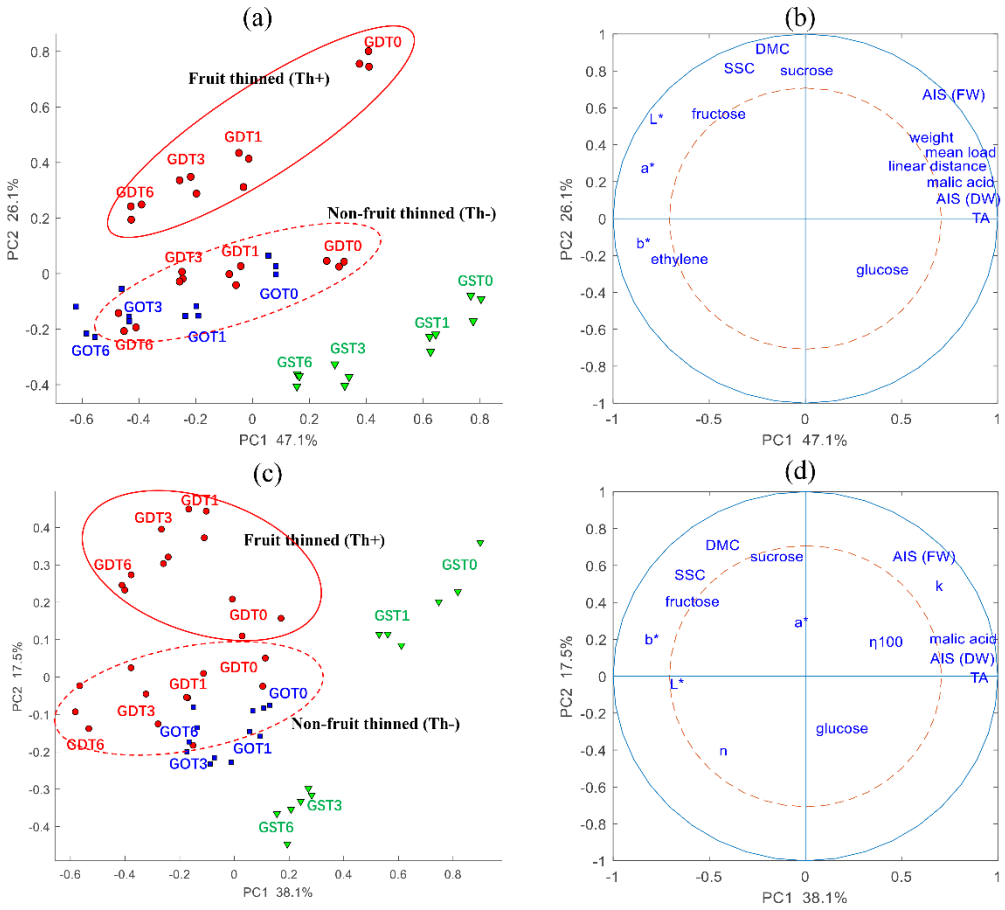

Fig. 40. PCA results of physical, physiological and biochemical compositions for: apples **(a)** and processed NR (no refined) purees **(c)** in three varieties (GO: 'Golden Smoothee', GD: 'Golden Delicious' and GS: Granny Smith) growing under two different thinning conditions (Th+ marked with red solid circle and Th- with red dotted circle) and stored at 4°C from harvest (T0), 1 (T1), 3 (T3) and 6 (T6) months. Correlation plot of the first principal components for apples **(b)** and NR purees **(d)**.

Table 31. Reference data of fresh apples with different varieties and growing conditions at harvest and during cold storage.

Variety	Year	Storage periods	Fruit thinning	weight	mea	linear		colour		Ethylen	AIS	AIS	DMC	SSC	TA	glucos	fructos	sucrose	malic
				g	(N)	(-)	L*	a*	b*	nmol/h. kg	mg/g FW	mg/g DW	g/g FW	°Brix	meq/kg FW	g/kg FW	g/kg FW	g/kg FW	g/kg FW
Golden Smoothee	2016	T0	Th -	144.2	2.71	12.33	70.	-15.6	44.0	2621	19.5	102.5	0.189	15.5	79.5	22.4	81.1	54.4	5.7
		T1	Th -	164.3	1.82	11.30	72.	-11.8	49.0	9223	15.3	86.8	0.173	13.0	74.6	18.3	76.4	49.0	5.3
		T3	Th -	126.0	1.75	10.84	75.	-6.1	54.5	7508	15.5	91.3	0.180	14.2	56.0	19.9	77.6	48.8	3.9
		T6	Th -	121.8	1.48	10.81	75.	-0.8	60.0	6933	14.5	88.5	0.166	13.8	40.4	19.2	85.0	40.4	2.6
SD				*17.8*	*0.48*	*0.65*	*2.4*	*5.9*	*6.2*	*2596*	*2.0*	*11.03*	*0.013*	*1.2*	*16.3*	*1.9*	*4.0*	*6.3*	*1.3*
F-value and significance		Storage periods		*23.0* ***	*224.* ***	*27.9* ***	*78.* **	*334.4* ***	*121* ***	*60.2* ***	*84.3* ***	*7.3* *	*2.2* *ns*	*6.1* *	*325* ***	*5.7* *	*8.6* **	*5.6* *	*82.0* ***
Granny Smith	2017	T0	Th -	196.6	3.09	17.23	63.	-18.8	41.1	2	38.5	256.6	0.151	11.1	97.8	21.4	43.6	34.5	6.0
		T1	Th -	188.9	2.68	15.78	62.	-17.7	40.3	714	26.7	182.2	0.149	11.8	103.4	21.6	26.1	40.5	6.1
		T3	Th -	189.5	1.84	11.40	66.	-16.9	44.1	2147	24.3	169.7	0.145	11.6	86.0	23.2	29.2	36.4	4.6
		T6	Th -	182.0	1.85	11.50	66.	-13.3	45.2	1082	23.0	161.2	0.143	10.8	60.9	21.6	50.4	30.1	3.5
SD				*7.0*	*0.58*	*2.70*	*1.9*	*2.2*	*2.2*	*810*	*6.5*	*39.8*	*0.004*	*0.5*	*17.3*	*1.48*	*11.3*	*6.1*	*1.2*
F-value and significance		Storage periods		*4.1* *ns*	*62.1* ***	*383.1* ***	*27.* **	*87.4* ***	*92.1* ***	*889.1* ***	*136.5* ***	*142.5* ***	*3.5* *ns*	*4.4* *	*115.3* ***	*0.9* *ns*	*3.2* *ns*	*1.9* *ns*	*63.4* ***
Golden Delicious	2017	T0	Th -	176.2	2.71	16.47	70.	-15.7	40.8	322	26.8	154.7	0.174	13.5	64.9	24.1	74.9	59.2	4.4
		T1	Th -	178.8	1.76	11.43	72.	-14.1	43.2	5072	23.2	136.4	0.174	14.4	55.9	21.7	65.6	65.3	3.4
		T3	Th -	189.9	1.50	10.76	75.	-8.5	48.9	6367	24.4	138.0	0.175	14.5	50.4	21.0	58.7	59.3	2.4
		T6	Th -	154.2	1.20	10.49	76.	-3.6	53.3	4393	19.9	121.4	0.164	12.8	26.5	22.0	73.0	46.0	2.6
		T0	Th +	205.4	3.31	17.59	75.	-10.3	43.6	12	49.4	218.0	0.225	17.4	84.0	20.9	96.3	86.3	6.5
		T1	Th +	208.2	2.03	11.84	75.	-8.6	43.7	5268	32.2	153.5	0.209	15.9	68.4	17.1	67.8	80.5	5.0
		T3	Th +	195.5	1.69	11.23	77.	-4.4	47.7	4074	28.1	145.5	0.194	16.4	61.2	15.7	65.6	82.8	3.5
		T6	Th +	192.8	1.56	10.80	76.	1.1	53.0	3718	27.3	133.4	0.204	16.9	34.5	23.7	89.3	57.2	2.3
SD				*17.3*	*0.67*	*2.75*	*2.3*	*5.4*	*4.5*	*2126.0*	*8.7*	*28.7*	*0.021*	*1.6*	*17.9*	*3.1*	*13.4*	*14.2*	*1.4*
F-value and significance		Storage periods		*20.2* ***	*788.* ***	*92.0* ***	*18.* **	*1103.* ***	*434.* ***	*92.4* ***	*118.0* ***	*100.3* ***	*21.2* ***	*4.4* *Ns*	*315.7* ***	*14.1* ***	*19.2* ***	*46.0* ***	*227.7* ***
		Fruit thinning		*148.3* ***	*186.* ***	*3.4* *ns*	*39.* **	*963.5* ***	*3.5* *ns*	*90.9* ***	*332.7* ***	*97.7* ***	*535.3* ***	*432.* ***	*146.1* ***	*22.4* ***	*28.1* ***	*159.1* ***	*160.5* ***

Data expressed in Fresh weight (FW) or Dry matter weight (DW), values correspond to the mean of 3x10 apples. Four storage periods at 4°C: from harvest(T0), one (T1), three (T3) and six months (T6); two conditions of fruit load: non-thinning with 100% number of apples (Th-) and thinning with 50% number of apples (Th+) per tree; ns, *, **, ***: In grey, one way- ANOVA results of Golden Smoothee, Granny Smith apples and two-way ANOVA results of Golden Delicious apples. ns, *, **, ***: Non-significant or significant at $P < 0.05$, 0.01, 0.001 respectively.

Table 32. Reference data of purees during a cold storage.

Variety	Year	Storage periods	Fruit thinning	Refining	η_{100}	Viscosity		colour			AIS	AIS	DMC	SSC	TA	glucos	fructos	sucrose	malic
					Pa.s	*K*	n	L*	a*	b*	mg/g FW	mg/g DW	g/g FW	°Brix	meq/kg FW	g/kg FW	g/kg FW	g/kg FW	g/kg FW
Golden Smoothee	2016	T0	Th -	NR	0.93	21.6	0.39	44.8	-5.0	14.9	20.4	114.9	0.177	14.9	66.4	23.4	71.8	56.8	5.2
		T1	Th -	NR	0.61	31.3	0.21	45.4	-4.4	15.3	13.8	82.9	0.166	14.1	65.7	18.9	68.7	55.5	4.9
		T3	Th -	NR	0.60	21.7	0.25	46.6	-4.3	16.2	14.2	86.7	0.177	13.8	54.4	17.9	73.3	46.3	4.3
		T6	Th -	NR	0.59	31.4	0.21	48.7	-4.3	19.6	15.0	96.1	0.159	13.4	43.0	19.0	80.8	39.4	3.8
		T0	Th -	Ra	0.67	39.1	0.20	45.2	-4.4	11.2	14.4	83.0	0.173	14.5	66.9	22.9	70.7	58.8	5.1
		T1	Th -	Ra	0.49	27.4	0.20	45.8	-3.8	11.9	9.8	61.7	0.163	14.2	66.9	20.0	70.4	49.2	5.0
		T3	Th -	Ra	0.49	27.2	0.20	47.5	-4.3	15.3	9.5	60.1	0.210	14.0	55.3	18.1	73.7	42.7	4.2
		T6	Th -	Ra	0.48	26.9	0.21	50.4	-4.6	18.4	12.2	81.1	0.158	13.6	37.8	18.6	78.5	38.2	3.6
SD					*0.21*	*5.4*	*0.06*	*1.9*	*0.4*	*2.8*	*31.1*	*6.0*	*0.019*	*0.55*	*11.3*	*2.2*	*5.01*	*8.2*	*0.6*
F-value and significance			storage periods		*51.2* ***	*10.3* ***	*72.2* ***	*258.2* ***	*5.6* **	*66.6* ***	*47.6* ***	*42.1* ***	*10.5* ***	*12.6* ***	*128.6* ***	*47.8* ***	*8.7* ***	*33.8* ***	*32.0* ***
			refining levels		*70.0* ***	*17.9* ***	*159.0* ***	*50.1* ***	*4.4* *ns*	*48.9* ***	*58.6* ***	*55.4* ***	*1.66* *ns*	*0.1* *ns*	*0.4* *ns*	*0.1* *ns*	*0.04* *ns*	*2.4* *ns*	*0.4* *ns*
Granny Smith	2017	T0	Th -	NR	0.91	54.6	0.10	44.6	-4.4	13.4	45.2	267.6	0.169	10.5	101.5	20.3	50.7	58.0	8.5
		T1	Th -	NR	0.84	49.9	0.09	46.4	-3.8	9.45	31.5	185.3	0.170	11.4	98.1	18.5	51.0	60.6	7.5
		T3	Th -	NR	0.72	22.2	0.25	46.2	-4.4	11.2	24.3	147.1	0.164	11.2	86.5	23.0	47.3	46.6	6.1
		T6	Th -	NR	0.66	18.8	0.27	46.5	-4.2	12.2	29.3	177.4	0.165	10.6	63.6	21.4	43.1	35.8	5.3
		T0	Th -	Ra	0.35	6.9	0.37	47.2	-4.5	13.6	20.5	126.1	0.164	10.3	107.8	19.8	54.7	55.1	7.9
		T1	Th -	Ra	0.26	3.2	0.47	48.0	-4.0	11.6	17.6	108.0	0.163	11.5	100.7	19.4	49.8	61.9	7.4
		T3	Th -	Ra	0.48	15.4	0.25	46.1	-4.4	11.4	23.2	142.9	0.162	11.6	88.0	22.9	51.2	51.9	6.4
		T6	Th -	Ra	0.48	16.5	0.23	46.8	-4.4	13.0	24.7	152.2	0.162	11.2	64.3	22.6	53.4	34.6	5.2
SD					*0.23*	*18.2*	*0.12*	*1.0*	*0.3*	*1.5*	*47.7*	*8.4*	*0.005*	*0.51*	*16.2*	*2.0*	*4.1*	*11.0*	*1.2*
F-value and significance			storage periods		*14.0* ***	*29.8* ***	*7.4* **	*8.6* **	*11.0* ***	*21.8* ***	*20.4* ***	*27.6* ***	*0.6* *ns*	*28.5* ***	*619.5* ***	*13.4* ***	*3.4* *	*34.1* ***	*163.6* ***
			refining levels		*1746.7* ***	*507.5* ***	*469.5* ***	*31.2* ***	*3.2* *ns*	*8.6* **	*146.8* ***	*183.16* ***	*3.8* *ns*	*5.2* *	*14.2* **	*0.7* *ns*	*16.5* ***	*0.1* *ns*	*1.1* *ns*

Table 32 (continued)

	T0	Th -	NR	0.77	40.8	0.12	46.7	-4.5	15.0	31.6	164.5	0.192	13.4	58.1	18.9	50.5	66.7	4.5
	T1	Th -	NR	0.70	22.7	0.23	47.0	-4.4	17.9	27.6	147.2	0.188	15.0	54.4	15.4	49.4	59.1	2.8
	T3	Th -	NR	0.55	17.0	0.25	47.9	-4.6	17.8	27.4	140.0	0.195	14.1	46.7	18.6	84.1	84.8	3.6
	T6	Th -	NR	0.50	17.0	0.24	52.9	-4.0	22.2	27.6	145.7	0.189	13.8	26.8	23.0	85.1	77.4	2.7
	T0	Th -	Ra	0.32	10.3	0.18	44.3	-4.0	9.0	18.3	126.6	0.197	12.6	59.0	24.5	45.2	43.5	3.2
	T1	Th -	Ra	0.42	13.5	0.25	46.6	-4.2	12.2	20.8	113.8	0.183	14.5	48.6	17.4	57.4	67.3	3.3
	T3	Th -	Ra	0.40	13.8	0.24	47.8	-4.6	15.5	22.3	116.0	0.192	14.1	48.5	19.1	79.0	88.5	3.6
	T6	Th -	Ra	0.39	14.9	0.22	53.3	-4.9	20.1	21.3	118.7	0.179	13.8	27.0	22.1	83.0	65.3	2.7
Golden Delicious 2017																		
	T0	Th +	NR	0.97	22.1	0.25	46.2	-4.5	15.8	41.1	193.6	0.206	15.5	70.9	23.5	85.4	64.4	5.5
	T1	Th +	NR	0.94	33.9	0.21	46.9	-4.0	17.1	31.9	150.9	0.213	17.6	69.3	16.8	80.3	115.9	5.6
	T3	Th +	NR	0.70	24.1	0.24	46.9	-4.0	16.5	31.6	143.3	0.221	16.9	59.9	13.8	88.0	102.5	4.9
	T6	Th +	NR	0.88	33.8	0.20	48.7	-4.2	22.7	34.8	150.3	0.231	17.5	34.7	23.8	95.7	44.0	3.6
	T0	Th +	Ra	0.44	13.0	0.25	45.2	-5.1	14.3	24.8	117.3	0.208	16.5	70.6	14.8	82.3	65.3	5.9
	T1	Th +	Ra	0.51	19.7	0.22	46.5	-4.2	15.2	25.4	121.8	0.208	16.5	71.4	17.8	81.3	118.6	5.9
	T3	Th +	Ra	0.51	18.8	0.22	46.3	-4.0	15.8	25.9	118.1	0.219	16.8	60.1	14.4	88.5	108.6	5.2
	T6	Th +	Ra	0.71	26.0	0.20	48.7	-4.6	22.0	29.5	124.0	0.238	17.4	36.0	24.5	97.0	57.3	3.7
SD				*0.21*	*8.5*	*0.03*	*2.4*	*0.4*	*3.6*	*18.5*	*5.0*	*0.019*	*1.7*	*15.0*	*3.8*	*16.6*	*24.1*	*1.2*
F-value and significance		storage periods		*8.3* ***	*3.8* **	*25.4* ***	*145.5* ***	*3.0* *	*196.0* ***	*1.9* *ns*	*5.7* **	*21.4* ***	*20.0* ***	*415.9* ***	*28.9* ***	*57.6* ***	*64.4* ***	*28.6* ***
		refining levels		*344.0* ***	*91.9* ***	*4.2* *	*8.7* **	*4.8* *	*109.1* ***	*140.9* ***	*170.1* ***	*8.6* **	*3.0* *ns*	*0.01* *ns*	*0.03* *ns*	*1.5* *ns*	*1.9* *ns*	*0.1* *ns*
		fruit thinning		*163.9* ***	*23.4* ***	*6.4* *	*52.6* ***	*1.0* *ns*	*23.7* ***	*68.3* ***	*72.7* ***	*432.5* ***	*331.0* ***	*308.3* ***	*3.9* *ns*	*41.7* ***	*19.3* ***	*191.4* ***

Data expressed in Fresh weight (FW) or Dry matter weight (DW); values correspond to the mean of 3x10 apples. Four storage periods at 4°C: from harvest (T0), one (T1), three (T3) and six months (T6). Two conditions of fruit load: non-thinning with 100% number of apples (Th-) and thinning with 50% number of apples (Th+) per tree. Two refining conditions of purees with 0.5 mm (Ra) and non-refining (NR). In grey, two way- ANOVA results of Golden Smoothee, Granny Smith purees and three-way ANOVA results of Golden Delicious purees. ns, *, **, ***: Non significant or significant at $P < 0.05$, 0.01, 0.001 respectively.

3.1.2 Apple purees

The fresh apple variability described above affected the characteristics of the corresponding non-refined (NR) purees cooked using the same recipe (**Fig. 40c, d**). The NR purees were discriminated according to the apple variety, fruit thinning and storage periods (**Fig. 40c**). The first principal component was positively correlated to TA, content of malic acid, AIS (DW and FW) and rheological parameters (η_{100}, K), and negatively linked with colors (L*, b*) and fructose content. The storage periods could be well-classified with this component. The second principal component was highly related to DMC, sucrose content, SSC and AIS content (FW) allowing the separation of varieties and fruit thinning conditions.

In all NR purees, clear decreases ($p < 0.001$) of TA, malic acid and AIS (in DW) were observed during storage, which were highly consistent with their changes in raw apples. At the same time, the rheological properties (η_{100} and K) decreased, with statistically significant differences ($p < 0.01$) in all NR purees, but not in Ra purees (**Table 32**). A good correlation was found between AIS expressed in fresh weight (FW) and the values of η_{100} in 'Golden Smoothee' (R^2=0.77), 'Granny Smith' (R^2=0.73) and non-fruit thinned (Th-) 'Golden Delicious' (R^2=0.84), meaning there was a good relationship between cell wall content and viscosity in NR purees. Visually perceptible differences of color with an increase of L* and b* (with $\Delta E > 2$) were detectable only after 6 months of storage for 'Golden Delicious' and 'Golden Smoothee' (Hunter, & Harold, 1987). Fructose content, as the major individual sugar, increased significantly ($p < 0.001$) in 'Golden Smoothee' and 'Golden Delicious' purees during storage, but not for 'Granny Smith', again in good agreement with the behavior observed in the raw fruits.

The changes of SSC, sucrose content and DMC in purees were still the major discriminative contributors for apple varieties and fruit thinning conditions during puree processing. Obvious differentiations were observed for SSC and sucrose content between 'Granny Smith' purees and the other purees, in accordance with their changes in raw apples. In 'Golden Delicious', tree thinning (Th+) led to a significant increase

($p < 0.001$) of DMC both in apples and their corresponding processed purees at each storage period (**Table 31** and **Table 32**), in accordance with the fact that the thinned apples, in addition to being larger, also accumulates more cell materials per volume unit (Palmer, Harker, Tustin, & Johnston, 2010). Additionally, the apples of the tree thinning (Th+) gave purees more viscous with significant higher values of η_{100} and K ($p < 0.001$) than the non-thinning condition (Th-). The tree thinning treatments, by affecting individual apple growth potential, affected physical properties of raw apples and processed purees, including their viscosity. Small fruits from non-thinned trees (Th-) resulted in less viscous purees than large fruits from thinned trees.

Concerning the refined (Ra) purees, an expected clear reduction was obtained for both AIS content (in DW) and viscosity (η_{100}) after refining (**Table 32**). That could be due to the loss of insoluble fibers in the removed particle fraction (Colin-Henrion, Mehinagic, Renard, Richomme, & Jourjon, 2009) leading to a loss of puree viscosity (Espinosa-Muñoz, To, Symoneaux, Renard, Biau, & Cuvelier, 2011; Leverrier, Almeida, & Cuvelier, 2016).

3.1.3 Relationship between the fresh apples and puree characteristics

In order to study the link between physical and chemical parameters of raw apples and their processed purees, the coefficients of determination (R^2) were calculated with the dataset including all three varieties under two thinning conditions (Th+ and Th-) and two refining levels and are displayed as heat maps for R^2 values from red ($R^2 > 0.8$) to blue ($R^2 < 0.2$) (**Fig. 41**). A clear similarity was observed between the two maps (**Fig. 41a, b**) with the same blue and red areas. Between all apples and their processed purees, high R^2 values (0.92 in NR and 0.91 in Ra) were obtained for TA. Acceptable correlations were also found for SSC (0.79 in NR and 0.81 in Ra), DMC (0.72 in NR and 0.73 in Ra) and malic acid content (0.65 in NR and 0.61 in Ra). For the AIS (DW) contents, good correlations (R^2) were obtained for each variety between raw apples and NR purees (not for Ra purees): 0.76 in 'Golden Smoothee', 0.83 in 'Granny Smith' and 0.77 in 'Golden Delicious', but lower when using all NR purees ($R^2 = 0.65$). Moreover, acceptable correlations ($R^2 > 0.71$) were obtained between texture characteristics (mean

load and linear distances) of all apples and rheological parameters (η_{100} and *K*) of their corresponding purees, whether non-refined (NR) (**Fig. 41a**) or refined (Ra) (**Fig. 41b**). The rheological variations in processed purees under the effects of different genotypes, storage periods and refining treatments were consistent with the textural changes in apples. However, no significant correlations ($R^2 < 0.44$) were found for individual sugars (glucose, sucrose and fructose) between all apples and their purees (NR and Ra), probably because: i) these concentrations changed less during long-term storage than TA and malic acid content, and ii) water content varied during thermal processing and refining treatments.

The good correlations of acidity (TA and malic acid content), SSC, DMC and physical properties (textual and rheological parameters) between apples and their purees were in line with PCA results (**Fig. 40b, d**). The chemical and physical variations in purees were shown to be potentially linked to the raw apple properties.

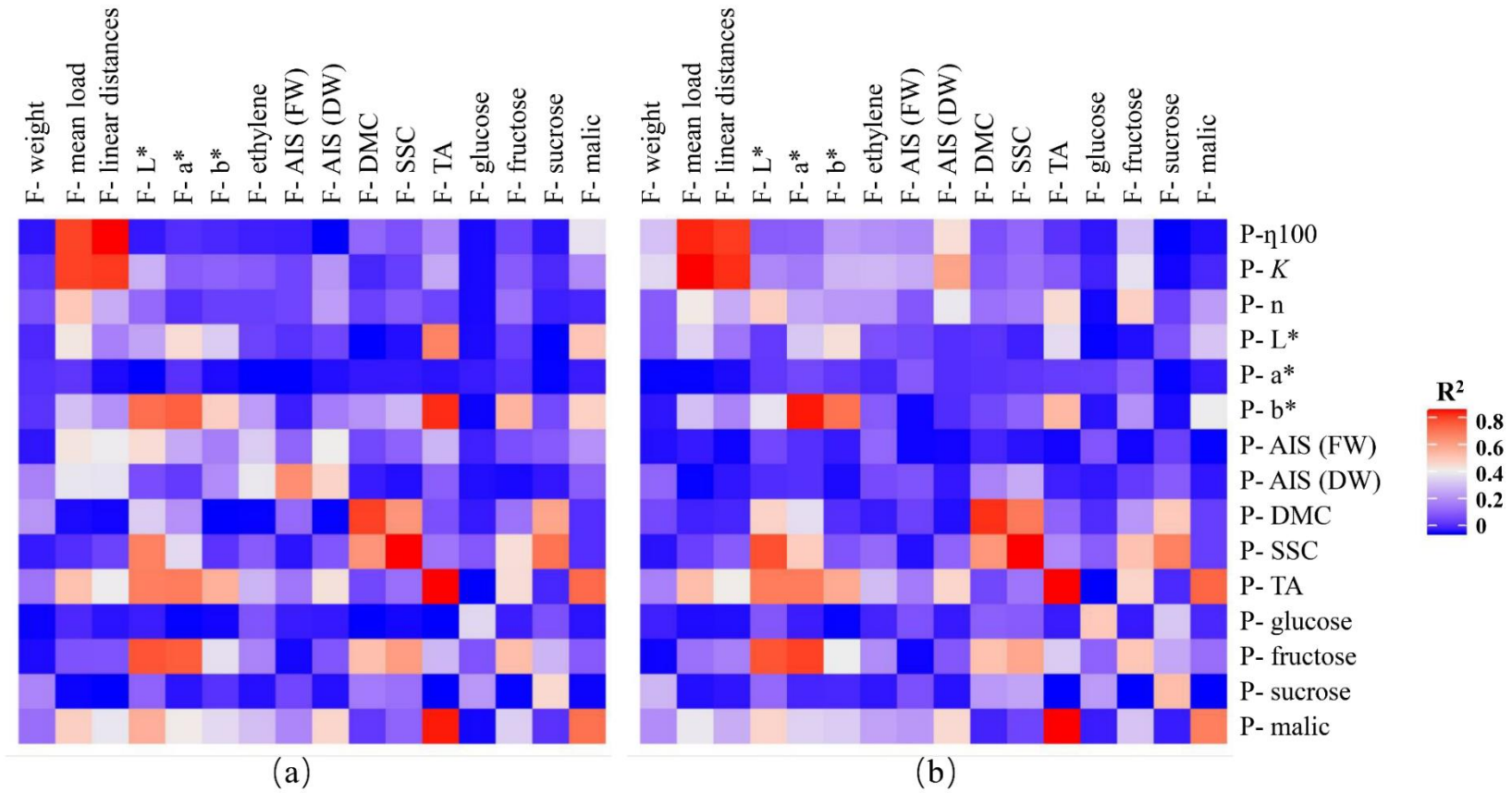

Fig. 41. Determination coefficient (R^2) of all physical and biochemical parameters between the 'Golden Smoothee', 'Granny Smith' and 'Golden Delicious' apples (titled with "F") and their processed purees (titled with "P"): **(a)** raw apples and non-refined (NR) purees; **(b)** raw apples and refined (Ra) purees.

3.2 Apple and puree characteristics measured by NIRS

ANOVA was performed on the SNV pre-treated NIR spectra of apples and processed purees (**Fig. 42**), in order to point out the wavelengths that varied during processing. According to the F-values, the variability was clearly higher for the spectra of apples (**Fig. 42a, b**) than those of purees (**Fig. 42c, d**), as could be expected given that each puree was prepared from 4 kg of fruit. For raw apples, the effect of variety (F-values of 800) was higher than the effect of storage (F-value of 120) (**Fig. 42a, b**). However, after processing into purees, the opposite conclusion was obtained: the effect of storage (F-values of 110) was almost three times higher than the effect of variety (F-value of 40) (**Fig. 42c, 42d**). Combine with their averaged spectral results (not shown), when apples were processed into puree, the peaks at 1930 nm in apples (**Fig. 42a**) were not variable in purees (**Fig. 42c**), demonstrating that the water contents had very limited variations in the purees.

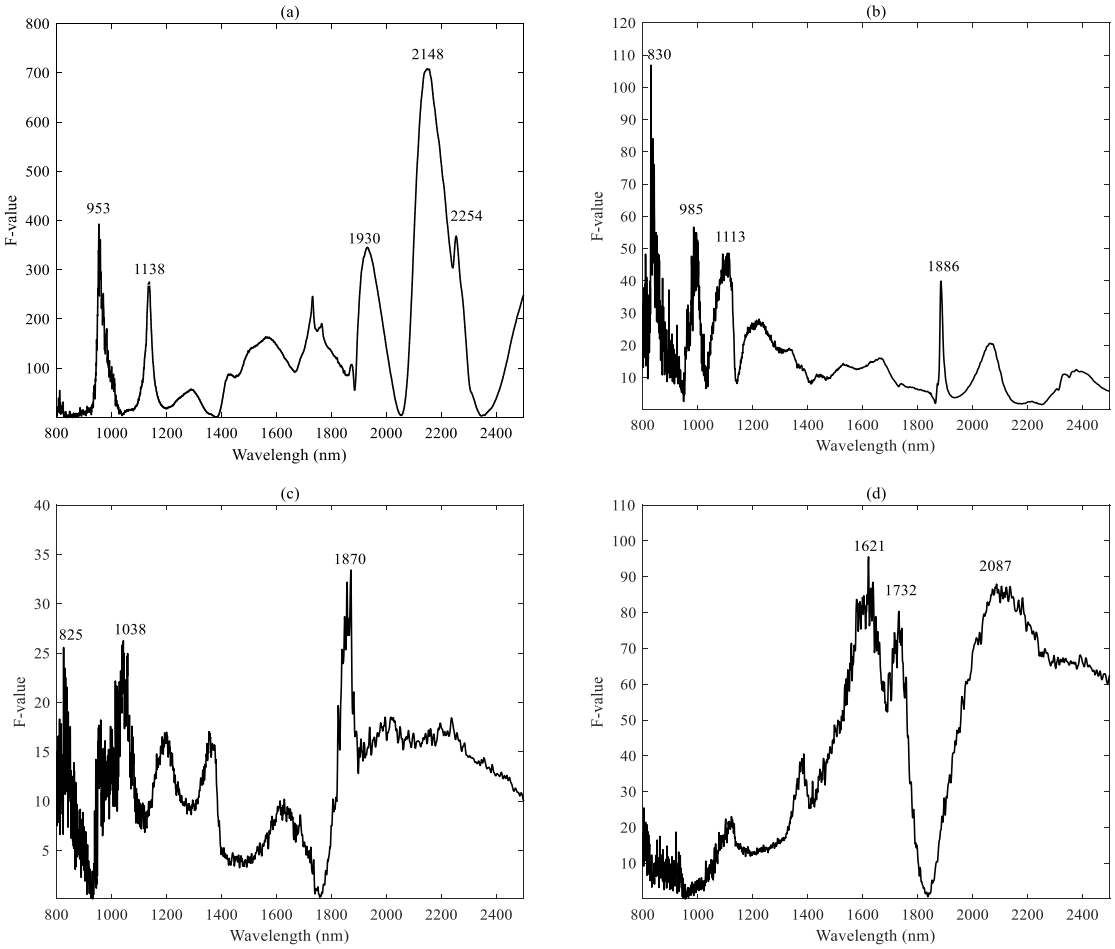

Fig. 42. ANOVA results of the SNV pre-treated NIR spectra between 800 and 2500 nm: **(a)** effect of variety on all apples spectra; **(b)** effect of storage period on all apples spectra; **(c)** effect of variety on all purees spectra; **(d)** effect of storage period on all purees spectra.

3.2.1 Discrimination of fresh apples

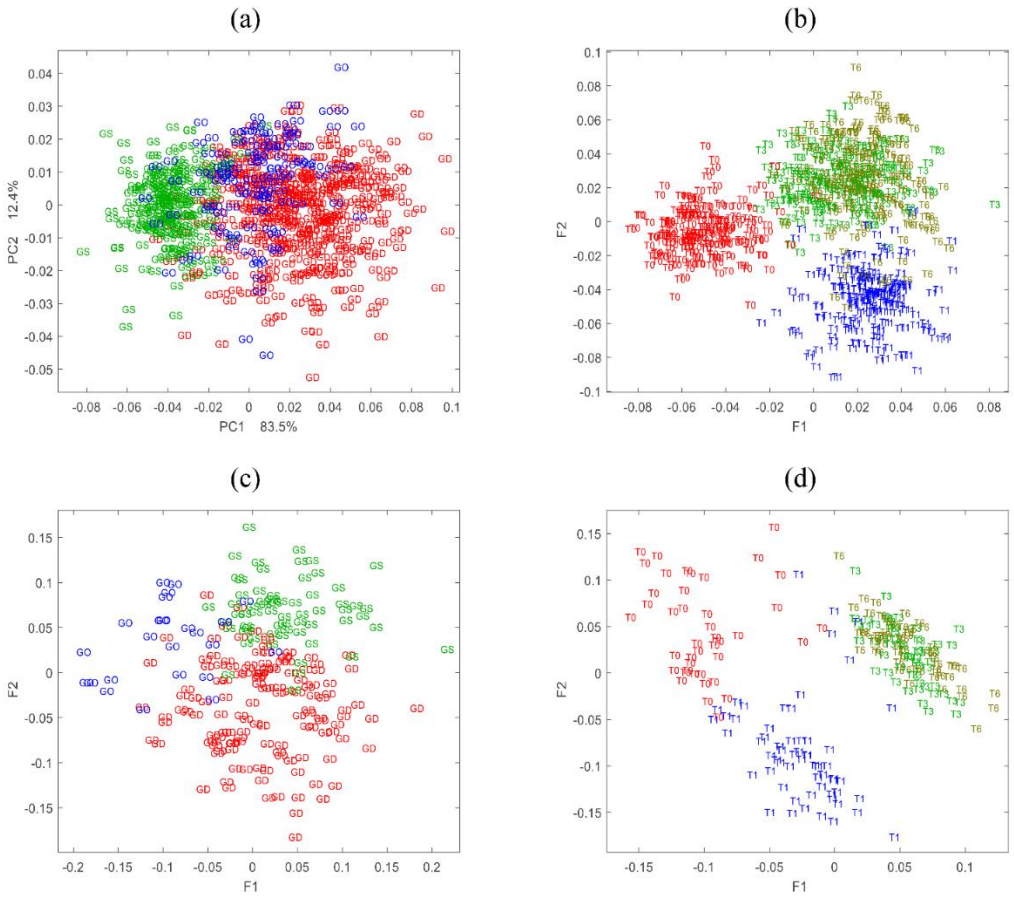

Fig. 43. Factorial maps of the SNV pre-treated NIR spectra of all apples between 1700 and 2350 nm or all purees between 800 and 2500 nm: **(a)** Principal Component Analysis showing apples varieties; **(b)** the Factorial Discriminant Analysis of storage periods of apples; **(c)** Factorial Discriminant Analysis of varieties of all purees; **(d)** Factorial Discriminant Analysis of storage periods of all purees. 'Golden Smoothee' (GO), 'Golden Delicious' (GD), 'Granny Smith' (GS); storage duration at 4°C from harvest (T0), 1 (T1), 3 (T3) and 6 (T6) months.

The wavelength range with the most variability (between 1700 and 2350 nm), identified by the ANOVA (**Fig. 42a**), was chosen to discriminate the effects of variety and storage. This range was used to perform PCA and FDA (**Fig. 43a, b**).

The first PCA displayed the discrimination of apples according to the variety (**Fig. 43a**). The first PC-score (PC1) discriminated 'Granny Smith' (GS) on the left and 'Golden Delicious' (GD) and 'Golden Smoothee' (GO) on the right, and accounted for 83.5% of the total variability. As observed for the reference data, 'Golden Smoothee'

spectra were overlapped with those of ‘Golden delicious’. The wavelengths at around 1880 nm, 1930 nm and 2100-2300 nm were the main contributors of the PC1 (not shown).The two bands at 1880 nm and 1930 nm are explained by the O-H combinations, which have been reported to characterize the water content in apples (Camps, Guillermin, Mauget, & Bertrand, 2007). The broad band at 2100-2300 nm corresponds to the first combination band of C-H bond of sugars or organic acids, already used to determine the concentration of individual sugars in apple juices (León, Kelly, & Downey, 2005; Liu, Ying, Yu, & Fu, 2006). These fingerprint wavelengths are consistent with the discrimination of apple varieties harvested in France (Camps, Guillermin, Mauget, & Bertrand, 2007).

In a second step, the different storage periods of all apples could be separated (100% of discrimination sensitivity and specificity between T0 and T1 apples, and 98.5% for sensitivity and 99.5% for specificity between T1 and T3 apples) by FDA (**Fig. 43b**). It was observed that T3 and T6 apples were overlapped (**Fig. 43b**), in line with their changes of mean load and linear distance (PCA could not well-classified storage periods). However, this result was inconsistent with previous report regarding well classification of storage periods of ‘Golden Delicious’ at 2°C by FDA (Giovanelli, Sinelli, Beghi, Guidetti, & Casiraghi, 2014). In our experiment, the strong variability and heterogeneity from different apple varieties and fruit-thinning treatments could provide more variations of water contents and carbohydrates, and thus introduced difficulties to well classify the storage stages after 3 months (T3). The use of different storage temperatures might also be involved, with a faster evolution at 4°C than at 2°C. The relevant wavelengths were mainly located in the ranges from 1700-1900 nm and 2250 nm (not shown).

The wavelengths around 1880 nm, 1930 nm, and 2100-2300 nm could be applied to the discrimination of apple varieties, while those at 1700-1900 nm and 2250 nm could be used for the classification of apple storage periods.

3.2.2 Discrimination of apple purees

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

ranges: 800-1050 nm, 1550-1730 nm, 1870 nm and 2100-2200 nm (**Fig. 43c, d**). Thus, the whole wavelength range from 800 to 2500 nm was used for FDA on the spectral dataset of all purees (not well-classified with PCA). The first two factors of the FDA allowed the discrimination of the three varieties, 'Golden Smoothee' (GO), 'Granny Smith' (GS) and 'Golden Delicious' (GD), with the discrimination specificity and sensitivity values of 86.8% and 84.6% in GD and GO apples; 84.0% and 82.2% in GD and GS apples; 88.5% and 91.9% in GD and GS apples (**Fig. 43c**). The F1 and F2 coefficients were both highly correlated with the area between 800 and 1000 nm (not shown), which is known as the absorption of apple carbohydrates and water (Giovanelli, Sinelli, Beghi, Guidetti, & Casiraghi, 2014; Zude, Herold, Roger, Bellon-Maurel, & Landahl, 2006), and already used for apple variety classification (Bobelyn *et al.*, 2010). Purees could be classified according to the storage periods with a distinct group for T0 and T1 (91.7% of sensitivity and 95.0% of specificity), but a mixed group for T3 and T6 (**Fig. 43d**). Besides the aforementioned absorbance region between 800 and 1000 nm, the wavelengths around 1400 nm and between 2100 and 2300 nm were also major contributors for discrimination of storage durations. These regions have been shown to be related to water loss and SSC variations, and could be regarded as the fingerprint wavelengths of apple storage periods (Camps, Guillermin, Mauget, & Bertrand, 2007; Giovanelli, Sinelli, Beghi, Guidetti, & Casiraghi, 2014; León, Kelly, & Downey, 2005). Other interesting results were obtained with the FDA applied on 'Granny Smith' purees taking into account the refining levels and the storage durations (not shown). According to the F1 and F2 axes, the two refining levels (Na and Ra) were separated both at T0 and T1, but not at T3 and T6. This result is highly consistent with the rheological changes of refined and non-refined purees of 'Granny Smith' (**Table 32**). This refining treatment led to stronger losses of viscosity and cell wall (AIS) content before the first month of apple storage (T1), compared to purees prepared after three months (T3 and T6).

3.3 Prediction of quality traits by NIRS

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

PLS to predict the physical and biochemical parameters of: (1) all apples from their NIR spectra (**Table 33**); (2) all processed purees (NR and Ra) from their NIR spectra (**Table 34**); (3) and all purees from the spectral information of apples (**Table 35** and **Table 36**).

3.3.1 Prediction of quality traits by NIRS on fresh apples

The prediction models were developed based on the 840 NIR spectra of apples combining 3 varieties, 2 years, 2 fruit thinning practices and 4 storage periods (**Table 33**). Selected results ($R^2 > 0.8$ obtained for the validation and RPD values > 2) were further discussed.

The prediction results of AIS content (R^2=0.85 expressed in dry weight and R^2=0.83 in fresh weight) during cold storage stood out in **Table 33**. The prediction of AIS content has already been studied but using the destructive mid-infrared technique on freeze-dried apple powder (Canteri, Renard, Le Bourvellec, & Bureau, 2019), and using NIRS to predict AIS content on 'Golden Delicious' apples during seven months storage at 2 ± 0.5°C with a R^2 of 0.96 (Lovász, Merész, & Salgó, 1994). The lower R^2 value in our study was probably in relation with the fact that three varieties and fruit thinning conditions were introduced in the same models. For the crunchiness (linear distance), the prediction result was acceptable (R^2=0.82, RPD=2.34), in accordance with previous results obtained on 'Golden Delicious', 'Braeburn' and 'Fuji' apples during a 7 months cold storage (R^2=0.84), but using the crunchiness data from sensory evaluation and the averaged NIR spectra of each group (Mehinagic *et al.*, 2003). Good predictions were obtained with DMC (R^2=0.87, RPD= 2.53) and SSC (R^2=0.81, RPD= 2.21), which were in accordance with previous studies (Giovanelli, Sinelli, Beghi, Guidetti, & Casiraghi, 2014; McGlone, Jordan, Seelye, & Clark, 2003). Moreover, acceptable correlation coefficients were obtained for TA (R^2=0.80, RPD= 2.09) and for the main organic malic acid (R^2=0.78, RPD=2.03). For the other individual sugar compounds, results were acceptable for fructose content (R^2=0.81, RPD=1.93) and sucrose content (R^2=0.81, RPD=2.14).

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

crunchiness (linear distance), SSC, TA, DMC, content of individual sugars (fructose and sucrose) and AIS. The benefit of AIS content prediction by NIRS was evident because the classical method of extraction and analysis needs a long time and lots of chemical solvents. Our models were also robust, given the large fruit variability used, with factors such as varieties, thinning practices and cold storage periods.

3.3.2 Prediction of quality traits of all purees

For purees, good predictions were observed for global parameters such as DMC (R^2=0.85, RPD=2.42) and SSC (R^2=0.92, RPD=3.12) (**Table 34**). The higher R^2 regarding SSC in purees than in apples, could possibly due a better homogeneity of the puree samples after processing. Additionally, acceptable results were also be observed for TA (R^2=0.80, RPD=2.22). For individual compounds, results were acceptable only for fructose (R^2=0.83, RPD=2.51). For the physical properties, rheological parameters (η 100, *K* and n) and color (L*, a* and b*), only poor results were obtained in all purees ($R^2 < 0.51$). However, in 'Granny Smith' purees (not shown), good correlations were obtained between NIRS and η100 (R^2=0.94, RPD=6.53), and *K* (R^2=0.93, RPD=3.52), and n (R^2=0.87, RPD=2.94). It seemed NIRS provided the possibility to access the evolution of rheological parameters in 'Granny Smith' purees from different apple storage times, but this relationship was not robust if a large variability of genotypes and agricultural practices were involved. Moreover, the results were surprising for AIS which was well-predicted in the corresponding intact apples (**Table 33**), but not in all processed purees ($R^2 < 0.69$, RPD < 1.59). The acceptable prediction of AIS content in apples probably depended on the good correlation between AIS and textural changes (firmness and crunchiness). z

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

In this part, PLS models were developed by combining spectral data acquired on fresh apples and reference data acquired on purees with two approaches: a) use the 48 averaged apple spectra (means of spectra of 2 faces x 10 apples by set) and the 48 reference data of their corresponding NR or Ra purees (3 replicates x 4 storage periods

x 4 puree groups); b) use the 480 averaged spectra of individual apples (means of faces a and b only) and their 48 reference data of corresponding NR or Ra purees. In this case, the same values of puree characteristics were linked to the 10 apples of the same set. These two methods (a and b) obtained similar prediction results and only results of the method b taking into account the apple spectra variability are shown, for both the NR purees (**Table 35**) and Ra purees (**Table 36**).

In NR and Ra purees, good predictions were obtained for rheological parameters (η_{100}, *K* and n). Especially for η_{100}, impressive R^2 and RPD values were observed for NR purees (R^2=0.88, RPD=2.31) and Ra purees (R^2=0.82, RPD=2.44). Good results were also obtained for AIS content (expressed in FW and DW) in NR purees (R^2=0.81, RPD=2.23) and Ra purees (R^2=0.84, RPD=2.48). As the AIS content is one of the main contributors of puree viscosity (Leverrier, Almeida, & Cuvelier, 2016), these concomitant results between AIS content and rheological parameters could probably be related to their good correlations in purees. In all studied purees, good correlations were obtained between their AIS and viscosity behaviors (η_{100}) (R^2 =0.75), but not for the AIS and SSC values (R^2=0.32). For coloration, acceptable prediction results of b* value were obtained both in NR purees (R^2=0.81, RPD=2.19) and Ra purees (R^2=0.79, RPD=2.12). Moreover, considering the DMC, SSC and TA, the PLS regression models had a good ability to estimate each characteristic for all purees on the basis of acceptable R^2 and RPD values (R^2 > 0.80, RPD > 2.11). However, the NIR technique cannot be used to estimate satisfactorily the content of individual sugars (fructose, sucrose, glucose) and of malic acid of purees depending on the spectral information of raw apples.

What stands out in these results was the better predictions of some quality traits of puree from fresh apple spectra (**Table 35** and **Table 36**) than from the puree spectra directly (**Table 34**). It was the case for rheological parameters, R^2 =0.82 from apples and R^2 < 0.44 from purees, possibly owing to the acceptable links (R^2 > 0.71) between apple texture (mean load and linear distance) and puree rheological properties (**Fig. 41**). In addition, better PLS results to predict AIS content from fresh apple spectra (R^2=0.81 in FW and 0.84 in DW) were obtained than from purees spectra (R^2 < 0.69) (**Table 34**),

probably due to good relationships between puree viscosity and AIS content mentioned above. Besides, the prediction of DMC, SSC and TA were still acceptable in all cases. Therefore, NIR technique showed a potential to directly predict the viscosity properties, b*, AIS content, SSC, DMC and TA of processed purees using their corresponding apple spectral information directly.

Compared with the PLS models used to predict the characteristics of purees based on their own spectra **(Table 33** and **Table 34)**, more LVs and lower prediction accuracy (RPD values) have generally found when models were built using the spectra of the intact apples to predict the characteristics of the processed purees. This fact has been also observed when other raw materials, e.g. meat (Meullenet, Jonville, Grezes, & Owens, 2004) or whole grain (Windham et al., 1997) were used to predict quality traits of the final cooked food. Such indirect prediction is a challenge as the spectra were not acquired on the material for which prediction was done, and because the chemical and textural traits from the material on which the spectra were acquired (the raw apples) are modified by processing. However, such predictions, albeit only semi-quantitative, are relevant for industrial use. Indeed, these developed models provide a promising solution to evaluate puree viscosity, a primary quality trait of this product, and cell wall contents (AIS), only based on spectra of fresh apples. A remarkable fact was that these predictions could not be done using the NIR spectra of the purees themselves (**Table 34**).

Table 33. NIR spectra and PLS method for prediction of internal quality traits of apples from three varieties (‘Granny Smith’, ‘Golden Delicious’ and ‘Golden Smoothee’), two thinning conditions and two harvest seasons, at 4 cold storage durations (0, 1, 3 and 6 months).

			Calibration n = 560		Validation n= 280				
Parameter	range	SD	r^2	RMSEC	R^2	RMSEV	LVs	RPD	spectral range(nm)
weight (g)	65.6-295.2	43.5	0.42	31.5	0.40	32.0	9	1.36	900- 2500
mean load (N)	0.97-4.81	0.68	0.72	0.36	0.70	0.37	11	1.84	900- 2500
linear distance (-)	10.44-19.49	2.51	0.84	1.01	0.82	1.07	11	2.34	900- 2500
L*	56.8-81.4	5.0	0.77	2.7	0.74	2.8	7	1.80	900- 2500
a*	-21.7-1.5	6.5	0.74	3.5	0.73	3.6	9	1.81	900- 2500
b*	28.9-64.8	5.5	0.62	3.4	0.61	3.5	9	1.60	900- 2500
Ethylene (nmol/h.kg)	0-11674	2639	0.74	1339	0.71	1403	9	1.88	900- 2500
AIS（mg/g DW）	69.2-266.7	38.38	0.85	16.1	0.85	17.7	12	2.16	900- 2500
AIS（mg/g FW）	14.1-51.1	8.02	0.86	3.1	0.83	3.6	12	2.24	900- 2500
DMC (g/g FW)	0.14-0.23	0.02	0.87	0.01	0.87	0.01	11	2.53	900- 2500
SSC (°Brix)	10.4-17.5	2.00	0.83	0.9	0.81	0.9	9	2.21	900- 2500
TA (meq/kg FW)	21.6-105.6	20.75	0.82	9.1	0.80	9.9	10	2.09	900- 2500
glucose (g/kg FW)	15.2-30.9	2.98	0.47	1.9	0.41	1.9	9	1.54	900- 2500
fructose (g/kg FW)	24.6-98.8	17.75	0.82	8.6	0.81	9.2	9	1.93	900- 2500
sucrose (g/kg FW)	28.7-91.1	16.33	0.84	7.2	0.81	7.6	10	2.14	900- 2500
malic (g/kg FW)	1.7-6.8	1.38	0.78	0.7	0.78	0.7	10	2.03	900- 2500

Table 34. NIR spectra and PLS method for prediction of quality traits of apple purees prepared from three varieties ('Granny Smith', 'Golden Delicious' and 'Golden Smoothee'), two thinning conditions and two harvest seasons, at 4 cold storage durations (0, 1, 3 and 6 months).

			Calibration n= 160		Validation n= 80				
Parameter	range	SD	r^2	RMSEC	R^2	RMSEV	LVs	RPD	spectral range(nm)
CSR (η_{100})	0.25-1.45	0.23	0.61	0.16	0.39	0.18	4	1.33	900- 2500
viscosity-*K*	2.8-59.5	11.78	0.59	7.2	0.39	9.7	6	1.22	900- 2500
viscosity-n value	0.06-0.49	0.16	0.57	0.05	0.44	0.10	6	1.59	900- 2500
L*	43.9-53.8	2.0	0.56	1.3	0.51	2.4	5	0.88	900- 2500
a*	-(5.2-3.4)	0.4	0.25	0.3	0.20	0.4	4	1.03	900- 2500
b*	7.8-23.0	3.6	0.55	2.5	0.51	2.7	5	1.33	900- 2500
AIS（mg/g DW）	90.9-271.7	35.2	0.76	15.2	0.66	32.6	7	1.08	900- 2500
AIS（mg/g FW）	14.8-48.9	6.2	0.71	3.5	0.69	3.9	7	1.59	900- 2500
DMC (g/g FW)	0.15-0.24	0.02	0.86	0.01	0.85	0.01	8	2.42	900- 2500
SSC (°Brix)	10.3-18.6	2.1	0.95	0.5	0.92	0.7	8	3.12	900- 2500
TA (meq/kg FW)	23.3-109.1	20.2	0.84	8.4	0.80	9.1	7	2.22	900- 2500
glucose (g/kg FW)	12.0-26.4	3.4	0.32	2.7	0.28	2.8	4	1.23	900- 2500
fructose (g/kg FW)	40.0-100.0	19.5	0.89	6.0	0.83	7.8	8	2.51	900- 2500
sucrose (g/kg FW)	24.7-123.1	24.3	0.63	14.4	0.60	16.6	6	1.46	900- 2500
malic (g/kg FW)	1.8-9.0	1.5	0.75	0.8	0.64	0.8	7	1.99	900- 2500

Table 35. Prediction of puree quality traits from spectral data of fresh apples: PLS results using NIR spectra of fresh apples from three varieties (‘Granny Smith’, ‘Golden Delicious’ and ‘Golden Smoothee’), two thinning conditions and two harvest seasons, at 4 cold storage durations (0, 1, 3 and 6 months) for prediction of quality traits of non-refined (NR) purees.

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

Table 36. Prediction of puree quality traits from spectral data of fresh apples: PLS (Partial Least Squares) results using NIR spectra of fresh apples from three varieties ('Granny Smith', 'Golden Delicious' and 'Golden Smoothee'), two thinning conditions and two harvest seasons, at 4 cold storage durations (0, 1, 3 and 6 months) for prediction of quality traits of refined (Ra) purees.

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

4. Conclusion

As far as we know, this was the first report concerning the assessment of quality variation of apple purees depending on NIR spectral information of the corresponding raw apples. Up to now, in apple industry, manufacturers use their experience and knowhow to make blend of apples in order to obtain always the same puree. From our results, NIR had the potential to predict internal quality of apples but also that of their processed products: a reliable assessment of texture and taste of the purees could be obtained based only on spectral data of fresh apples. This opens the possibility to sort or select apples according to the expected purees. By systematically scanning all apples, this could provide some objective data to predict the final product characteristics and thus reduce waste of materials along the processing chain. Further work will be needed to investigate the interaction of the processing conditions (temperature, time, oxygen and so on) with raw apples under various growing and storage conditions, so as to provide guidance for adapted processing procedures to reach stable final puree qualities.

Highlights of Paper IV

This study provided the answers of the first and the last questions in Part 2:

- Strong correlations were shown between apple texture (firmness and crunchiness) and puree viscosity ($R^2 > 0.79$), TA ($R^2 > 0.91$), and also for SSC ($R^2 > 0.79$) and DMC ($R^2 > 0.72$) between apples and purees.
- NIRS on apples and purees allowed their good classification at over 82% and 88% according to varieties and storage times, respectively.
- NIR coupled with PLS models showed a good ability to estimate puree characteristics from spectra acquired on corresponding apples such as viscosity ($R^2 > 0.82$), cell wall content ($R^2 > 0.81$) and also dry matter ($R^2 > 0.83$), soluble solids content ($R^2 > 0.80$) and titratable acidity ($R^2 > 0.80$).

  These new results open the possibility of using NIR technique to give a reliable assessment of texture and taste of the final products based on the evaluation of non-destructive raw materials.

Paper V (Preparing for submission)

Strong correlations of physical and biochemical properties have been identified between raw and processed apples (**Paper IV**). It also opens the possibility to predict puree composition (SSC, TA etc.) and viscosity from the NIR spectra on corresponding raw apples. However, these relationships between fresh and processed apples were obtained on a set of fruits needing at least 2.5 kg (approximately 10-15 apples). This means around 15 apples were processed in a single puree, ignoring the 'intra-batch variability' brought by each individual apple. Knowing the 'intra-variability' between raw fruits and cooked purees can help field growers and industrial manufacturers to sort fruits and produce sustainable final products.

Therefore, an experiment was designed to link individual apples to their purees. Totally 120 individual apples of 4 varieties ('Golden Delicious', 'Granny Smith', 'Braeburn' and 'Royal Gala') and their corresponding processed purees were used (**Paper V**). Apples were individually processed to puree using a microwave process with the strict definition of 'one apple to one puree', and the VIS-NIR spectra were acquired both, on each couple of individual apple and cooked puree to:

i) Know how much the inter- (between varieties) and/or intra-variability (between individual fruits) of raw apples impacts cooked purees;

ii) Determine the VIS and NIR spectral variations and relationships between raw and processed apples considering the experimental design of 'one apple to one puree';

iii) Further validate the potential of predicting the quality traits of the final cooked purees using the VIS-NIR spectra of intact raw apples.

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*

(This paper is under preparation for submission)

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 industrial conditions to process apples 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' between apple varieties (Buergy, Rolland-Sabaté, Leca, & Renard, 2021; **Paper IV**). 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 stability 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.

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 apples 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, the 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, has been applied for detecting the different fruit species in mixed purees (Contal, León, & Downey, 2002). From our previous work, strong correlations of chemical and textural properties have been pointed out between raw apples and their corresponding purees (**Papers I and II**). Based on that, these works 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 and the reference data of corresponding processed purees (**Paper I**). 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 (**Paper III**; Pissard, Baeten, Romnée, Dupont, Mouteau, & Lateur, 2012) and a large variability of puree characteristics (different varieties) have been clearly highlighted (**Papers I and IV**). 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. 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’? Besides, 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 spectra were acquired on 120 individual apples of 4 varieties and their corresponding individual processed purees to: i) study the variability of both, the individual apples and corresponding purees; ii) explore the spectral

correlations and variations before and after each apple processing; and iii) predict 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. 44**). All apples were stored for four months at 4°C before processing. During three successive weeks, 10 apples of each variety were processed and so were put at room temperature (22.5 °C) one day before. In total 120 individual apples (4 varieties × 10 apples × 3 weeks) were measured with the non-destructive techniques (color, VIS-NIR spectra). Afterwards, they were processed in 120 purees (one apple processed in one puree) and characterized for their biochemical composition and rheological properties.

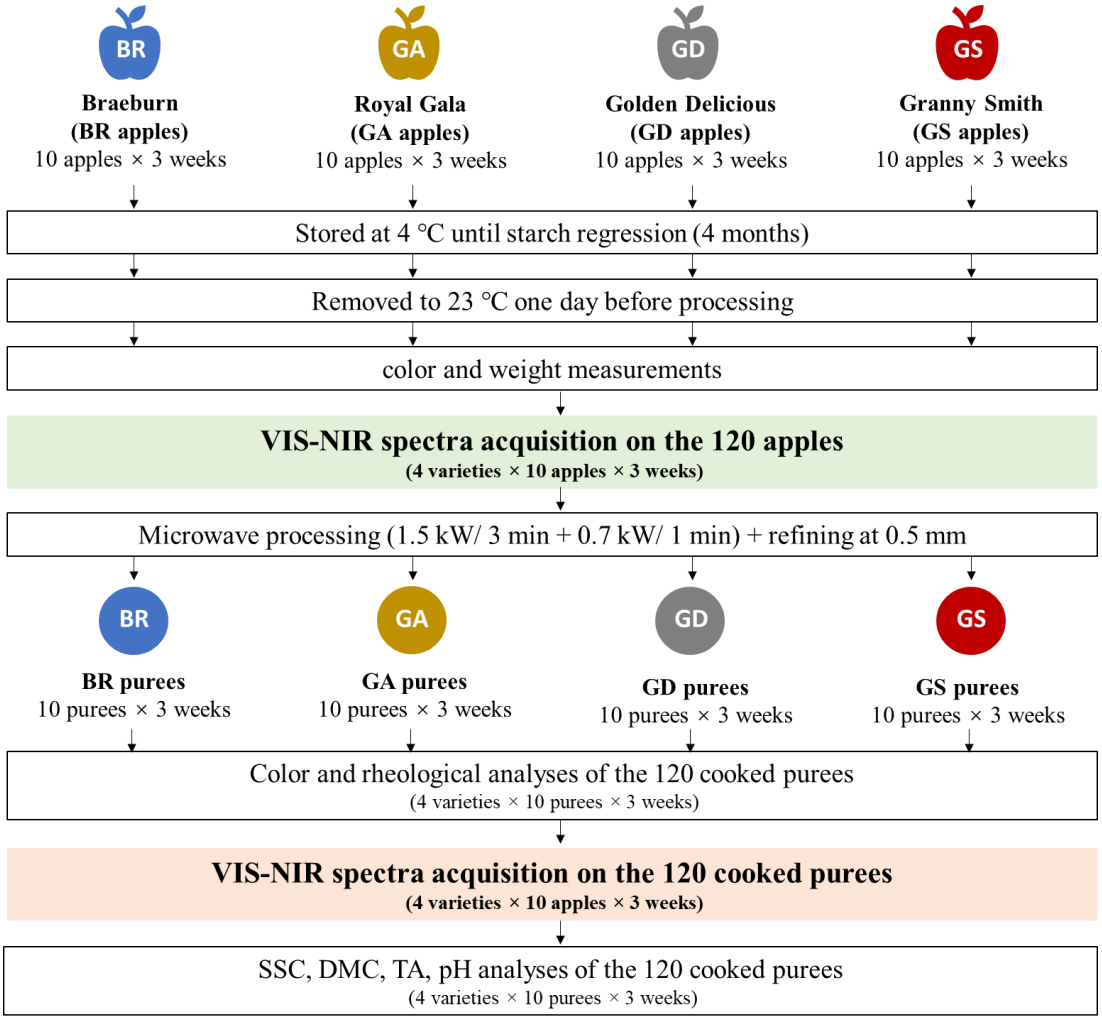

Fig. 44. Experimental design of apple processing, spectral acquisition, and quality characterization.

2.2. Non-destructive 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 500-800 nm (VIS) and 800-2500 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 cooking.

2.3 Individual apple processing

Individual and intact apples (about 160 g each) were sealed in a domestic preserving container 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 processed puree was conditioned in a hermetically sealed can, and then placed at 23 °C during one day prior to further analyses. Totally 120 purees (4 varieties × 10 purees × 3 weeks) were obtained from a large intra-variability of four apple varieties.

2.4 Determination of quality traits of individual purees

2.4.1 Physical characteristics

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

The puree rheological measurements, and in particular flow curves were carried out using a Physica MCR-301 controlled stress rheometer (Anton Paar, Graz, Austria) and a 6-vane geometry (FL100/6W) with a gap of 3.46 mm, at 22.5°C. The flow curves were performed after a pre-shearing period of 1 min at a shear rate of 50 s^{-1}, followed by 3 min at rest. The viscosity was then measured at a controlled shear rate range of [10; 250] s^{-1} on a logarithmic ramp. The values of viscosity at 50 s^{-1} and 100 s^{-1} (η_{50} and η_{100} respectively) were took as the final indicators of puree viscosity, which are

representative of the mouth sensory characteristics during consumption (Chen & Engelen, 2012).

2.4.2 Biochemical analyses

The biochemical analyses of all purees have been described in **Papers I and II**.

2.5 Spectrum acquisition on individual purees

VIS-NIR spectral data of processed purees were acquired using the same conditions than 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 spectrophotometers. 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. 45**) 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. Besides, multiblock principal component analysis (MB-PCA) was carried out on the VIS-NIR spectral matrices of apples and purees to evaluate the impact of fruit variability on processed products (Abdi, Williams, & Valentin, 2013). 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 developed 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 their 120 corresponding puree spectra. The dataset was randomly split as follow: two third of dataset (80 spectrum of apples and their related 80 spectra of cooked purees) were used for calibration and one third of dataset (40 spectrum of apples and their related 40 spectra of cooked purees) for validation. The procedure was repeated 10 times and the developed model performances were 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 Nicolai (Nicolai et al., 2007).

3. Results and discussion

3.1 Comparison of purees cooked by cooker-cutter and microwave ststem

In this part, the quality variations of different 'Golden Delicious' purees cooked from cooker-cutter processing system (Roboqbo, Qb8-3, Bentivoglio, Italy) on apple batches (at least 2.5 kg per batch) (**Paper IV**) or the microwave process on individual apples (**Paper V**) were displayed in **Table 37.** Both two processing strategies resulted

in similar ranges levels of puree chemical and physical quality traits. The microwave process can thus be considered as an efficient method to produce apple purees, presenting similar quality levels of the conventional cooked purees. Particularly, the microwave cooked purees from individual apples generated a more intensive chemical and physical variations (SD values) than the conventional purees cooked from apple batches. This result indicated the large intra-variability of apples is still present after microwave cooking, whereas averaging resulted in loss of intra-variability when apple batches were homogenized in one puree in cooker-cutter processing system.

Table 37. A comparison of puree quality from thinned Golden Delicious apples cooked at harvest in 2017 using the cooker-cutter robot (Roboqbo, Qb8-3, Bentivoglio, Italy) (**Paper IV**) and in 2020 using the microwave oven (**Paper V**).

Golden Delicious	Cooker-cutter processed purees (**Paper IV**)		Microwave processed purees (**Paper V**)	
	Ranges	SD	Ranges	SD
L*	45.89 - 46.48	0.26	46.11-51.33	1.33
a*	-(4.71-4.24)	0.21	-(6.09-5.07)	0.32
b*	15.39-16.22	0.36	17.61-23.71	1.96
η_{100}	0.79-1.06	0.13	0.64-1.15	0.13
DMC (g / g FW)	0.20-0.21	0.01	0.14-0.20	0.02
SSC (°Brix)	15.2-15.9	0.3	11.8-17.4	1.73
TA (meq/kg FW)	6.73-7.35	0.28	2.89-5.29	0.73
pH	3.60-3.63	0.01	3.82-4.44	0.14

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

In this study, both the inter-variability of apple varieties 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. 45** and **Fig. 46**).

3.2.1 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. 46**). Both, for apples and purees, a significant ($p < 0.0001$) higher redness and a 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 (b* values) 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. 45**) 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 (a* values) ($R^2 = 0.70$) and yellowness (b* values) ($R^2 = 0.58$) between apples and purees.

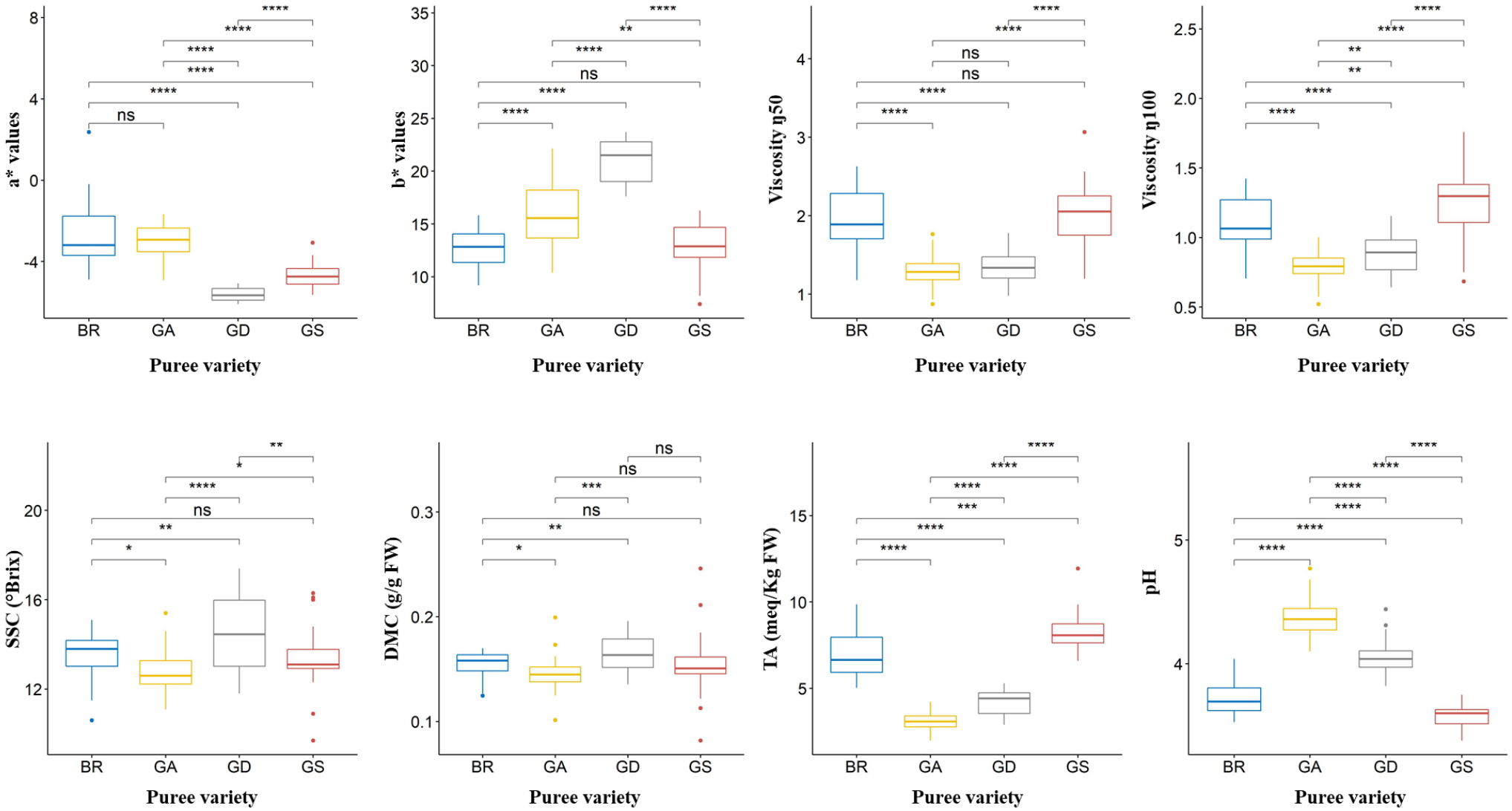

Fig. 45. The boxplots and the T-test results of color, rheological and biochemical properties of four apple puree varieties. (The significances were displayed as 'ns' (p values > 0.05), '*' (p values ≤ 0.05), '**' (p values ≤ 0.01), '***' (p values ≤ 0.001) and '****' (p values ≤ 0.0001)).

3.2.2 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 et al., 2021). Moreover, the viscosity at the shear rate of 50 s^{-1} (η_{50}), which is commonly used to describe the in-month texture perceptions of fluid food (Chen et al., 2012), was similar ($p > 0.05$) in GD and GA purees. This result was different from our work in **Paper VII** giving a higher viscosity of GD than of GA purees. 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). 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. 45**). 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^{-1} and η_{50} = 1.57 ± 0.18 Pa s^{-1}). However, this was not the case for two GA apple sets with different redness (a* = 27.2 ± 5.0 and a* = 14.2 ± 2.3) resulted a similar puree viscosity of η_{50} = 1.26 ± 0.18 Pa s^{-1} and 1.39 ± 0.30 Pa s^{-1}, respectively **(Fig. 46)**.

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 apple will not allow anticipating the viscosity of cooked purees.

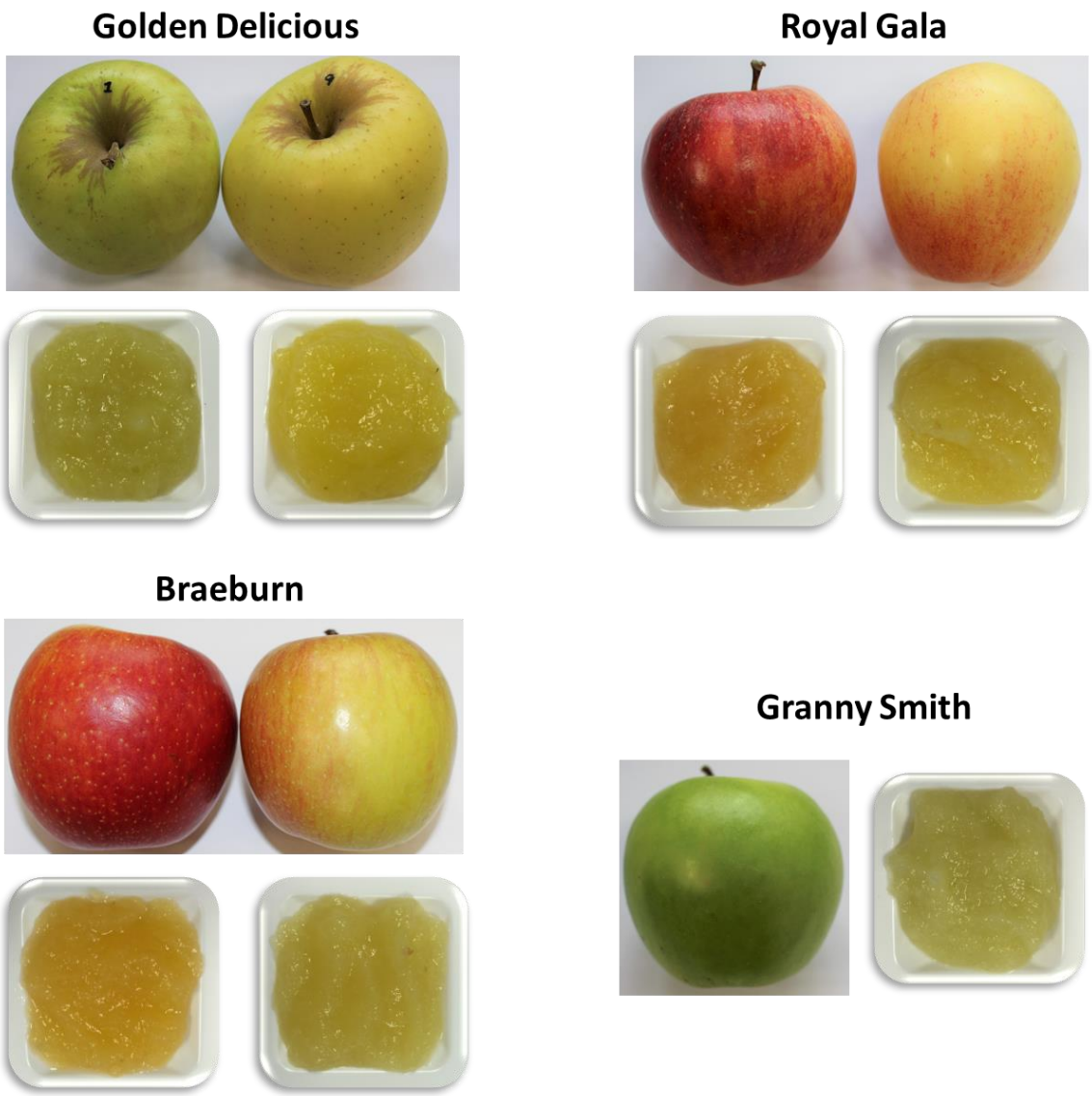

Fig. 46. The pictures of individual apples and the corresponding microwave cooked purees.

3.2.3 Biochemical compositions of purees

Significant inter-variability ($p < 0.05$) was observed for SSC between the four puree varieties, except for BR and GS purees ($p > 0.05$) (**Fig. 45**). 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 significant different ($p < 0.05$) in BR, GA and GD purees, but not for GS. This result can be explained by the large intra-variability of

DMC in GS purees (**Fig. 45**). 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 the ranking was reversed 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 introduced 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.

3.3 Spectral analysis of apples and purees

3.3.1 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 variability could be observed at around 500-700 nm, 1140 nm, 1386-1392 nm, 1880 nm, 1930-2197 nm and 2250-2450 nm (**Fig. 47a, b, c)**. Generally, the spectral variability was clearly higher in apples than in their corresponding purees, probably because the disappearing of intra-variability while 4 kg of apples were homogenized in one puree (**Paper IV**). Besides the possible effects of process, this difference between apples and purees perhaps also due to the sample structure (solid fruit and liquid purees), impacting the diffuse reflectance.

A Multi Blocks PCA (MB-PCA) was performed on the SNV pretreated VIS-NIR spectra to evaluate the distribution of apples and related purees, according to the variability brought by varieties and apples (**Fig. 48**). The four varieties could be discriminated based on the first two components (PC1 and PC2) for apples and purees with an explanation rate of 65.55% (**Fig. 48a**). Moreover, these two PCs can be explained by a clear higher contribution of apples than their corresponding purees, which indicates a reduction of the variability after processing (**Fig. 48b**).

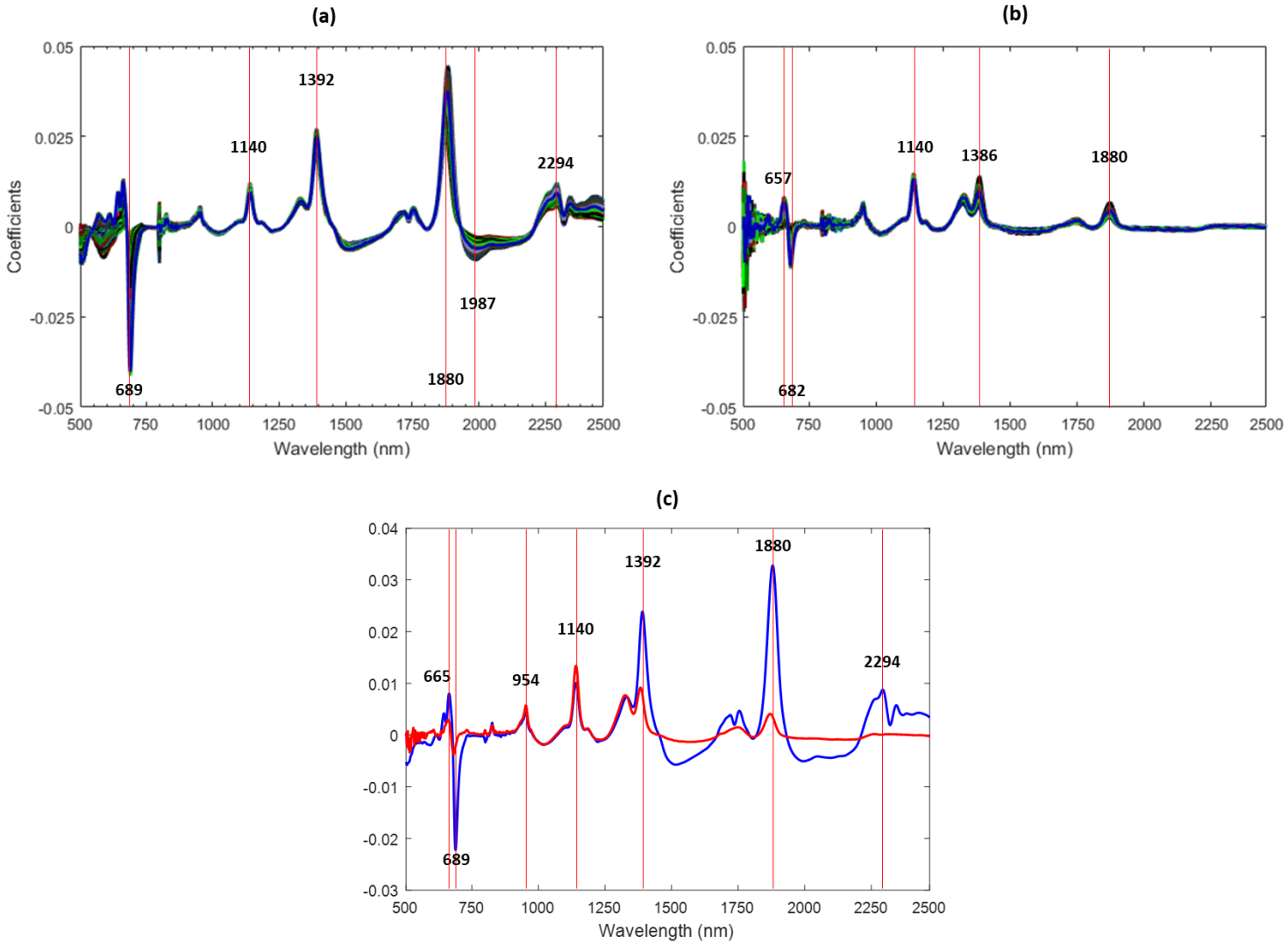

Fig. 47. The pre-processed (smoothing with 13 windows + SNV+ 1st derivation with 11 windows) VIS-NIR spectra of **(a)** individual apples and **(b)** their related cooked purees, and **(c)** the averaged pre-processed spectra of all apples (blue line) and cooked purees (red line).

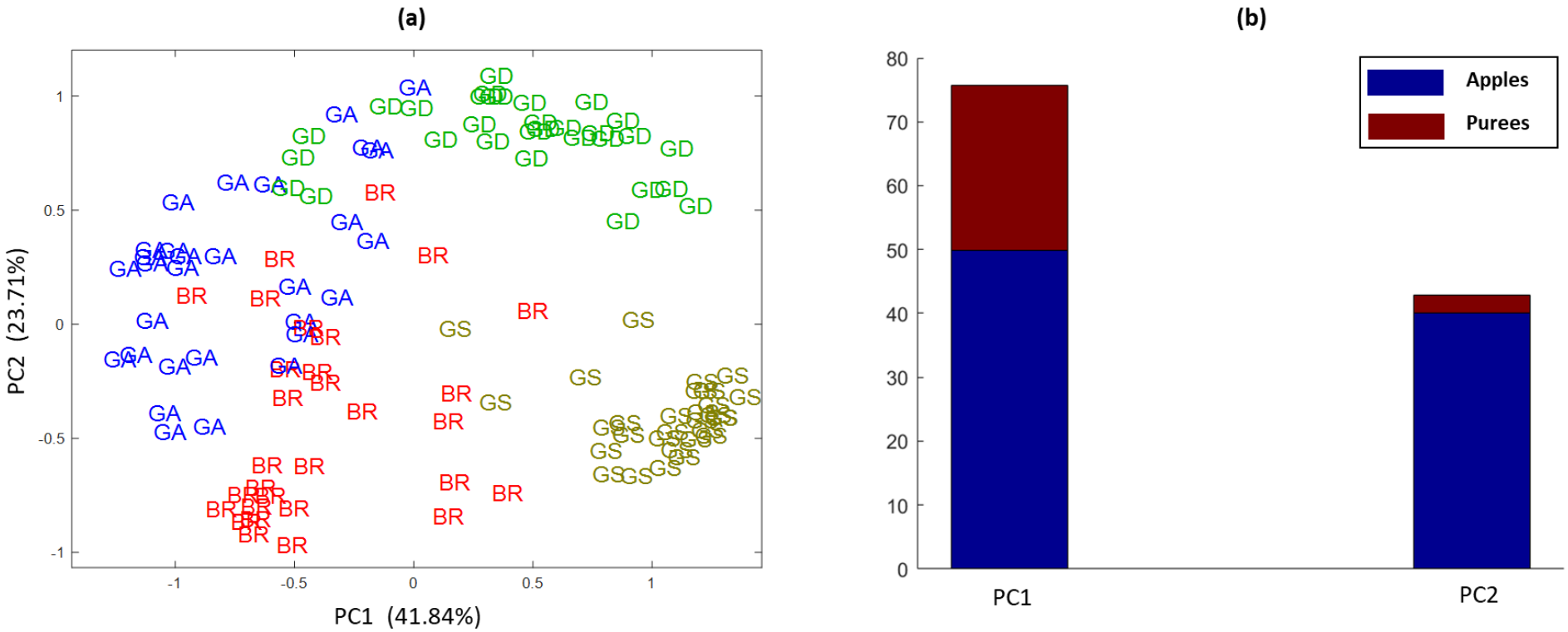

Fig. 48. Multiblock principal component analysis (MB-PCA) on two VIS-NIR spectral metrics of all apples and their related cooked purees: (a) the discrimination plot of four varieties based on the first two principal components (PC1 and PC2); (b) the contributions of the blocks of variables in apple or puree spectral matrices for PC1 and PC2.

3.3.2 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. 49a**) and NIR (800-2500 nm) spectra (**Fig. 49b**) in order to point out the highly correlated wavelengths between apples and their processed purees. The correlations were much higher in Visible than in NIR ranges. Particularly, in visible (**Fig. 49a**), a clear positive correlation was obtained from 665 nm to 685 nm, that confirming a strong color relationship 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 were positively correlated to the corresponding spectral areas at the same wavelengths of the purees. These positive correlations may be due to the SSC in apples and purees, which was stable during processing (**Paper IV**). Reversely, the wavelengths between 1125-1400 nm in apples were negatively correlated to the spectral regions 1850-2150 nm and 2250-2450 nm in cooked purees. One possible reason was the loss of water contents compared to the limited increases of sugars concentrations during apple thermal processing.

As mentioned in **Paper IV**, a strong relationship of physical and biochemical properties between fresh and processed apples has been observed and allow to predict the puree properties from the spectra information of apples. The observed spectral correlations in this study supported these previous results, in the current dataset taking into account both the large inter-variability with varieties and their intra-variability with apple individually processed.

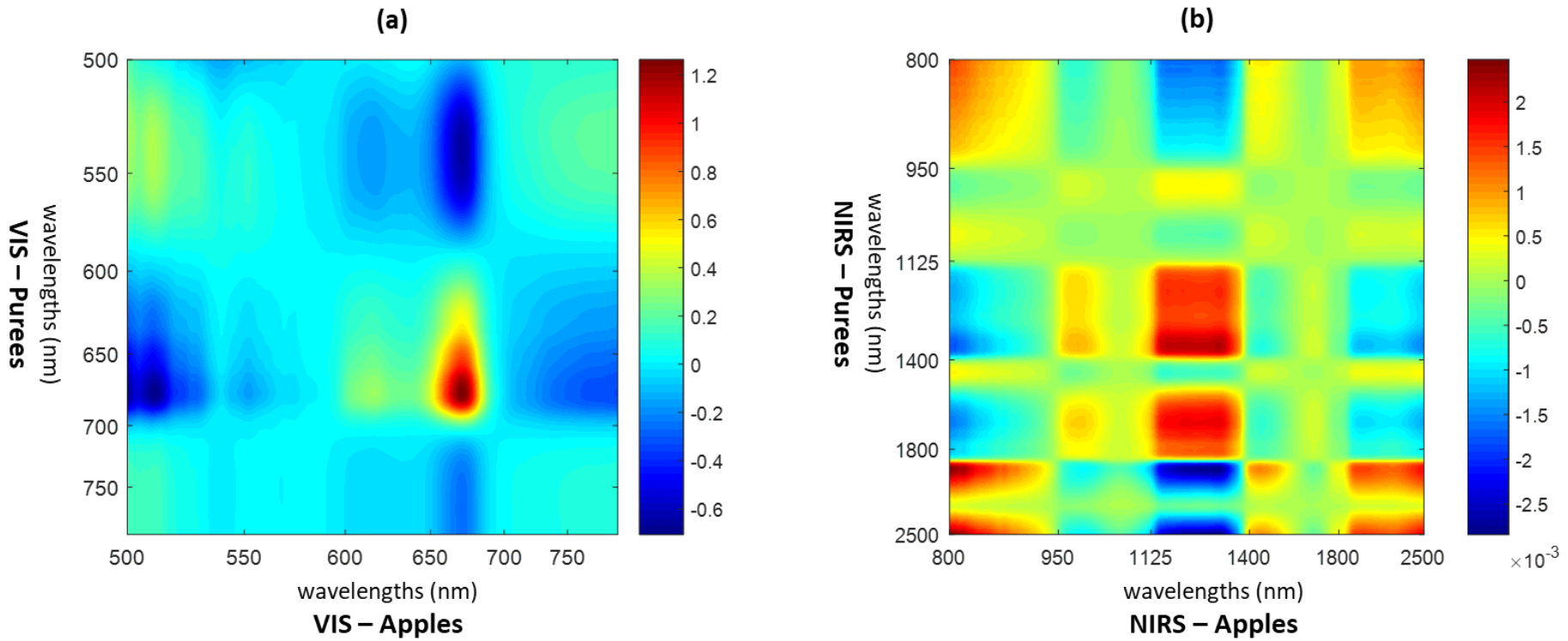

Fig. 49. The 2D-COS (two-dimensional correlation spectroscopy) plot between the spectra of all individual apples and their related cooked purees in the **(a)** visual (500-780 nm) and **(b)** near infrared (800-2500 nm) ranges.

3.4 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 38**), or on the corresponding individual raw apples (in **Table 39**).

3.4.1 Prediction using the 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 obtained to predict puree viscosity at the shear 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) techniques ($R_v^2 = 0.39$, RPD = 1.3) (**Paper IV**). However, they were much lower than the acceptable VIS-NIR predictions obtained on a previous experiment consisting in studying the preparation of apple puree mixtures with different proportions of variety ($R_v2 > 0.73$, RPD> 1.9) (**Paper VII**) presenting over half as much variability (the SD of η_{50} = 0.12 Pa. s) than it was here (the SD of η_{50} = 0.36 Pa. s). Thus, the VIS-NIR or NIR techniques cannot provide acceptable estimations of the puree viscosity, considering both the 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 in the visible range. These same wavelengths are also identified in apple purees (**Paper VII**), 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 b* values. A relatively large puree variability for yellowness (b*) (SD = 4.1) compared to our previous study (SD = 1.7) significantly improved the VIS-NIR

prediction results, with the RPD from 1.5 to 2.2 (**Paper IV**). When variability enough large, a possible prediction of color parameters by VIS-NIR.

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 38**). 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 a good correlation between SSC and DMC (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 (**Paper IV**). 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) (**Paper IV**). 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 (**Paper IV**), 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 predictions 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 (cold storage periods of raw fruits) to be acceptable by VIS-NIR. However, for acidity, VIS-NIR models developed on the variability of different apples and varieties was enough to give an excellent estimation of TA and pH.

3.4.2 Prediction using the 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 39**). Particularly, PLS models provided the better predictions of η_{50}, η_{100}, b*, SSC and TA than SVM and RF results. Compare to the PLS models developed from puree spectra (**Table 33**), more spectral latent variables (LVs from 5-11) were required for their good predictions using the VIS-NIR spectral of apples.

What stands out in these results was the much better PLS predictions of puree rheological parameters (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 spectra regions strongly correlated between apple and puree (**Fig. 49**). This result is in keeping 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 (**Paper IV**). 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 prevents 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 on raw apples gave the acceptable predictions of the viscosity of their corresponding cooked purees.

For puree color parameters, VIS-NIR spectra on the 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 the L* values ($R_v^2 < 0.59$, RPD < 1.5). The major wavelengths contributing to these models were highly consistent with the developed models directly using the puree spectra, such as 514-514 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 cautions because they concern only microwave cooked apples and not the conventional thermic processing (in the lab)

in relation probably with a rapid inactivation of enzymes by microwaves.

For puree biochemical parameters, PLS regression models had a good ability to estimate SSC, TA and pH of all purees based on the 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 from corresponding apples at 950 nm, 1150 nm, 1400 nm and 1880 nm, corresponding to the sugars and water variations (**Part 3.2)**. Impressively, the TA and pH of cooked purees can be 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 acidity (TA) using the apple NIR spectra and giving RPD around 2.1-2.3 (**Paper IV**). Indeed, some specific peaks in the visible region at 524 nm and 672 nm contributed to the good predictions of TA in purees. It can thus be suggested that taking into account the visible range of apples can provide a better prediction of puree acidity than just using the NIR range. Concerning DMC, the bad predictions might be explained by the lower variations here compared to our previous work (**Paper IV**), 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.

Table 38. 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 (meq/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 39. 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 (meq/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.

4. Conclusion

This study was designed based on the absolutely definition of ‘one apple to one puree’, which gave a first insight of the impacts of fruit inter- or intra-variability during processing, from the spectroscopic point of view. Importantly, the intra-variability in fruits can introduce the 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 of each one, could improve the prediction accuracy of regression models. Further, the strong correlations while apple processing obtained both from conventional characterizations and spectral analyses provided further evidences for such the indirect predictions of puree colors, viscosity and global biochemical parameters (SSC, TA and pH) from the non-destructive spectral information on raw apples.

Highlights of Paper V

This study gave a first insight into the relative impacts of fruit inter-variability (between varieties) and intra-variability (between individual fruits) on the quality of processed purees. And it provided further evidences that VIS-NIR and NIR techniques should be useful tools for industry in so far as they can give a reliable assessment of texture and taste of final puree products, based on non-destructive evaluations of the raw material.

- Both the inter-variability of apple varieties and the intra-variability of single apples introduced intensive changes of appearance, chemical and textural properties of the cooked purees.
- The intra-variability of cooked purees was different according to apple varieties.
- 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, related to chlorophyll and carbohydrate contents.
- VIS-NIRS allowed to predict puree color (a* and b*, RPD $\geqq$ 2.1), viscosity (RPD $\geqq$ 2.3), SSC (RPD = 2.1), TA (RPD = 2.8) and pH (RPD = 2.5) based on the spectra of raw individual apples.

Paper VI (Published)

From **Papers IV** and **V**, the quality of final processed purees was predicted by the VIS-NIR or NIR spectral information acquired on raw apples using PLS regressions. However, the main drawback of this strategy is the need, for modelling, to systematically acquire the corresponding spectra on both, raw and processed materials with a large number of conditions representative of the variability, giving often only rough assessments. In addition, the internal correlations of quality traits during puree processing were only confirmed under one of the most commonly used processing conditions.

An experiment was then conducted on 'Golden Delicious' apples harvested in 2017 and 2019, with two thinning practices and four storage times (at harvest, 1, 3, 6 months of cold storage), then cooked under three heating temperatures (70°C, 83°C and 95°C for 30 min) and three grinding speeds (300 rpm, 1000 rpm and 3000 rpm) (**Paper VI**). The aim of this work was to find the MIR spectral relationships between raw apples and processed purees at different cooking conditions, and explore the possibility of using MIR spectroscopy to predict puree quality obtained in different cooking conditions based on the prior information of raw apples and processing recipes. Direct standardization method was applied to find the spectral relationships between fresh and processed apples in different cooking conditions.

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Mid-infrared technique to forecast cooked puree properties from raw apples: A potential strategy towards sustainability and precision processing

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

[a] INRAE, Avignon University, UMR408 Sécurité et Qualité des Produits d'Origine Végétale, F-84000 Avignon, France
[b] INRAE, TRANSFORM Division, F-44000 Nantes, France
[c] INRAE, ONIRIS, Unité Statistiques, Sensométrie, Chimiométrie (StatSC), F-44322 Nantes, France
[d] INRAE, Unité InforSol, F-45075 Orléans, France

1. Introduction

Apple puree is one of the major industrially processed fruit products (over 0.3 million tons consumed per year in France and the world market value about 2000 million USD annually) (FranceAgriMer, 2017; Market Research Future, 2019), and can be used as the basic ingredient of jams, preserves or compotes (Defernez, Kemsley, & Wilson, 1995). The quality of apple purees is systematically influenced by both raw material characteristics (**Papers I, II and IV**, Buergy, Rolland-Sabaté, Leca, & Renard, 2020; Rembiałkowska, Hallmann, & Rusaczonek, 2007) and cooking strategies (heating, grinding speed and refining levels etc.) (Espinosa, To, Symoneaux, Renard, Biau, & Cuvelier, 2011; Oszmiański, Wolniak, Wojdyło, & Wawer, 2008; Picouet, Landl, Abadias, Castellari, & Viñas, 2009). In practical apple processing, industrial manufactures have to face the variability and heterogeneity of raw apples, optimize their processing strategies to maintain the sustainable and expected quality level of final puree products. Thus, developing rapid, efficient and integrated methods is needed to guide suitable fruit processing procedures, even to design innovative foods by using the raw material variability, and to reduce fruit wastes all along the processing chain.

Mid infrared spectroscopy (MIRS) is one of the main candidates for both the quantification and qualification of agricultural commodities and processed food (Bureau, Cozzolino, & Clark, 2019; Downey, 1998). Although MIRS presents a relatively lower ability for quantification than chromatography or mass spectrometry, it has the advantage of a rapid data acquisition and can provide informative fundamental

vibrations of molecular bonds (Karoui, Mazerolles, & Dufour, 2003). It does require a minimal sample preparation as the measurement must be done on homogeneous samples as liquid, puree or powder due to the very low penetration of radiation into the samples. Direct MIR characterizations of raw and processed fruits have shown considerable aptitudes to evaluate soluble solids content (SSC), dry matter content (DMC), titratable acidity (TA), some individual sugars, organic acids, rheological (viscosity and viscoelasticity) and structural (particle average size and volume) properties (Ayvaz, Sierra-Cadavid, Aykas, Mulqueeney, Sullivan, & Rodriguez-Saona, 2016; **Paper II**). These parameters are related to the taste (SSC, DMC, TA, malic acid), the texture (viscosity, viscoelasticity, particles and cell wall contents) and the basic nutrients (fructose, sucrose and glucose) impacting in a large amount puree quality (Bureau, Ścibisz, Le Bourvellec, & Renard, 2012; Espinosa-Muñoz, Renard, Symoneaux, Biau, & Cuvelier, 2013; Fügel, Carle, & Schieber, 2005).

Currently, our interest is to determine the possibility of using this technique to anticipate the characteristics of processed materials from the data acquired on homogenized raw fruit. According to our previous studies, strong correlations of spectral, chemical and textural properties between raw apples and the corresponding purees have been pointed out (**Papers IV and V**). Based on that, the quality of final processed purees could be predicted by the infrared spectral information acquired on raw apples using partial least square (PLS) regression (**Papers IV and V**). The main drawback of this strategy is the need, for modelling, to systematically acquire the corresponding spectra on both raw and processed materials with a large number of conditions representative of the variability, giving often only rough assessments. In addition, the internal correlations of quality traits during puree processing were only confirmed under one of the most commonly used processing conditions (**Papers IV and V**).

Direct standardization (DS) is a simple and efficient chemometric tool for the calibration transfer between spectral measurements or between two different sets of conditions, such as the spectral calibration from the off-line to on-line spectra of olive fruits (Salguero-Chaparro, Palagos, Peña-Rodríguez, & Roger, 2013). As far as we

know, this method has never been considered to bridge the spectral variations along the fruit processing chain. Our interest of this method is thus to find the spectra relationships of all processed purees and the corresponding spectral information acquired on homogenized apples, and to calculate the reconstructed processed puree spectra according to their relative spectral information acquired on apples by DS, taking into account both the variability of raw materials and of commonly used processing conditions. If so, the predictive models of puree quality traits (biochemical and physical) using their reconstructed spectra dataset open the possibility to i) predict the properties of processed apples based on the prior information of raw materials; ii) provide sustainable and precise processing strategies to estimate the quality potential of final products, and iii) to compare *in silico* the prediction results of different processing approaches to better control the quality of fruit products.

Accordingly, this work aimed to assess the potential of MIRS to: i) detect the variability of the cooked apple purees according to the pre- and post-harvest conditions (fruit thinning and storage periods) and the main processing conditions (heating temperature and grinding speed); ii) calculate reconstructed spectra of purees taking into account the variability of raw fruits and processing conditions; and iii) characterize and anticipate the biochemical (SSC, DMC, TA, individual sugars and malic acid), rheological (viscosity and viscoelasticity) and textural (particle size and volume) properties of the processed purees.

2. Materials and methods

2.1 Apples and purees

2.1.1 Apples

The experiment was conducted on the variety ‘Golden Delicious’ in 2017 and 2019. All apples were harvested at commercial maturity from La Pugère experimental orchard (Mallemort, Bouches du Rhône, France) (**Fig. 45)**.

- In 2017, half of the ‘Golden Delicious’ apples were subjected to a commercial chemical fruit thinning (Th+) with standard fruit load (50-100 fruits/tree), the other

half was not thinned (Th-), resulting in a high fruit load (150-200 fruits/tree). After harvesting, apples were processed into purees the day after harvest (T0), and after one (T1), three (T3) and six months (T6) of cold storage at 4°C.

- In 2019, the commercially thinned 'Golden Delicious' apples (Th+) were stored for up to one month (T1) at 4 °C until starch regression, then processed into purees under different processing conditions.

2.1.2 Puree processing

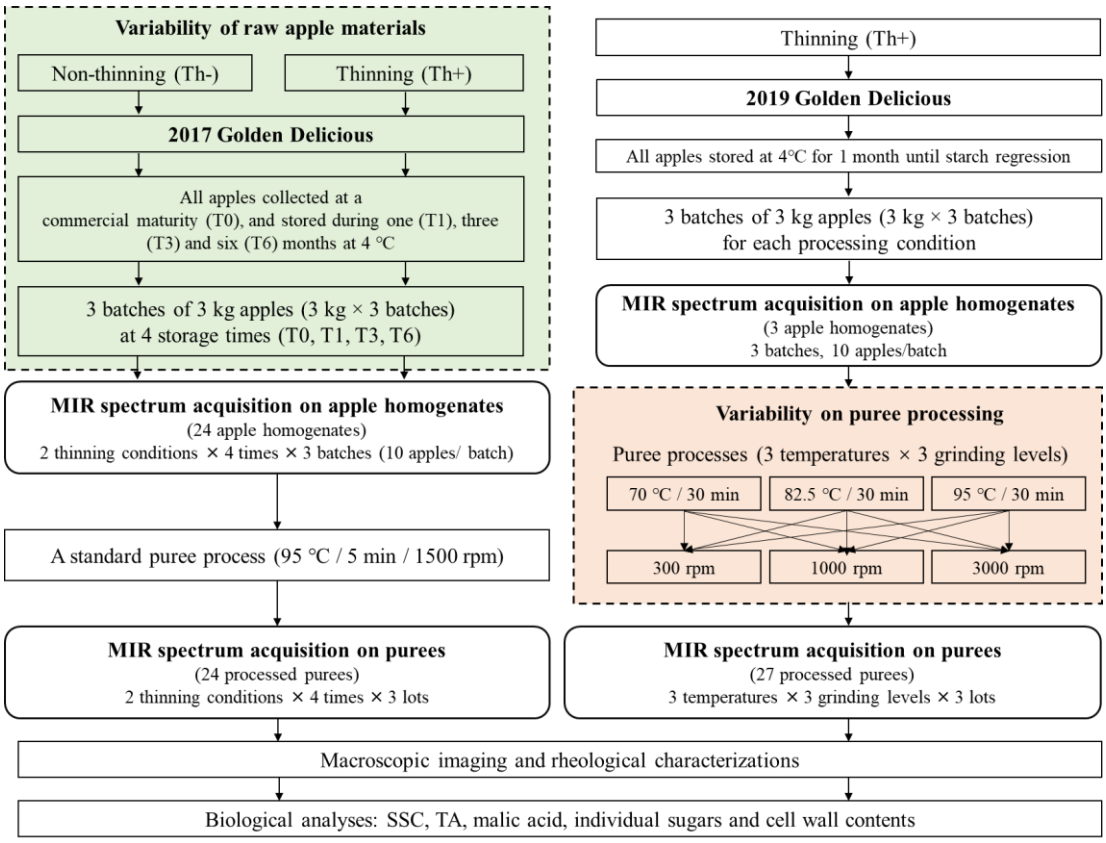

Fig. 50. Experimental scheme for apple production, puree preparation and the sample characterization by infrared spectroscopy and reference measurements.

Before processing, and for each condition, around 2 kg apples were homogenized at 11000 rpm with an Ultraturrax T-25 (IKA, Labortechnik, GmbH, Staufen, Germany) as raw apple homogenates. Three batches of apples (3 kg each) were used to produce three puree lots for each condition. After sorting and washing, Golden Delicious apples were cored and cut in 8 portions, then processed in a multi-functional processing system (Roboqbo, Qb8-3, Bentivoglio, Italy) by different conditions (**Fig. 50**):

- In 2017, all apple groups (2 thinning practices × 4 storage periods) were cooked with a standard Hot Break recipe of 95°C for 5 min at a 1500 rpm grinding speed, then cooled down to 65°C while maintaining the grinding speed. Finally, 24 different cooked purees (2 thinning practices × 4 storage periods × 3 lots) were prepared.
- In 2019, each apple group was processed with three different heating temperatures of 70°C, 83°C and 95°C for 30 min, while ground at three speed levels of 300 rpm, 1000 rpm and 3000 rpm, respectively. Totally, 27 different cooked purees (3 heating temperatures × 3 grinding speeds × 3 lots) were prepared.

Finally, all cooked purees were conditioned in two hermetically sealing plastic bags: one was cooled at room temperature (22.5 °C) before the next-day measurements of rheological, structural and some biochemical (SSC, TA, fructose, sucrose, glucose and malic acid) properties. And the other one was freeze-dried (FD) and stored at -20 °C for the determination of the content of cell wall, which are known to be a major contributor of rheological properties of apple purees (Espinosa-Muñoz, Renard, Symoneaux, Biau, & Cuvelier, 2013).

2.2 Determination of puree quality traits

2.2.1 Rheological and structural characterizations on cooked purees

The cooked puree rheological measurements have been described in detailed in **Paper II**.

2.2.2 Biochemical analyses on cooked purees

Biochemical analyses of cooked purees were described in **Papers I and II**.

2.3 Spectrum acquisition on raw apple homogenates and cooked purees

The MIR spectra acquisition has been described in **Papers I and II**.

The whole spectral dataset of MIR is described in **Fig. 51**. It included i) 81 spectra of raw apple homogenates, of which 72 spectra acquired in 2017 (3 apple batches × 2 fruit thinning conditions × 4 storage times × 3 spectral replicates) and 9 spectra acquired in 2019 (3 apple batches × 3 spectral replicates); and ii) 153 spectra of cooked apple

purees, containing 72 spectra acquired in 2017 (2 fruit thinning conditions × 4 storage times × 3 processing lots × 3 spectral replicates) and 81 spectra acquired in 2019 (3 heating temperatures × 3 grinding levels × 3 processing lots × 3 spectral replicates).

2.4 Statistical analyses of reference data

The reference data of cooked purees processed in 2017 and 2019 are presented as the mean values and the data dispersion within our experimental dataset expressed as standard deviation values (SD). After the Shapiro-Wilk tests, the references data of processed purees affected by fruit thinning and storage times were normal distributed (α=0.05), but not for the dataset of heating temperature and grinding effects during puree processing. Thus, analysis of variance (ANOVA) was carried out to determine the significant differences of cooked purees due to fruit thinning and storage times applied on raw apples (**Table 40**) using XLSTAT (version 2018.5.52037, Addinsoft SARL, Paris, France) data analysis toolbox. The normal distribution hypothesis was not verified for the dataset of heating temperature and grinding effects during puree processing Therefore Kruskal-Wallis tests were performed to evaluate the effects of heating temperature and grinding levels during puree processing (**Table 41**).

2.5 Spectra transferred by direct standardization (DS)

In this study, DS was used to find the relationship between the spectra matrices of all cooked purees ($\boldsymbol{P}$) and the corresponding spectra of raw apple homogenates ($\boldsymbol{F}$), taking into account the effects of raw material variability and processing conditions. The DS transfer works were performed in R software (version 4.0.2) (R Core Team, 2019) following a previous report (Ji, Viscarra Rossel, & Shi, 2015):

$$\boldsymbol{P} = \boldsymbol{FB} + \boldsymbol{E} \qquad (1)$$

where $\boldsymbol{B}$ is the transfer matrix ($\lambda \times \lambda$) presenting the variations in both $\boldsymbol{F}$ and $\boldsymbol{P}$, $\boldsymbol{E}$ is the residual matrix used to correct the baseline difference. $\boldsymbol{F}$, $\boldsymbol{P}$ and $\boldsymbol{E}$ matrices have the same size $n \times \lambda$, where n presents the numbers of transfer spectra and λ is the number of wavenumbers between 1800 and 900 cm^{-1}.

First, to compute the transfer $\boldsymbol{B}$ and error $\boldsymbol{E}$ matrices, the whole MIR spectral dataset ($\boldsymbol{P}$ and $\boldsymbol{F}$) was divided into: the calibration matrices, presenting the first two batches of raw apple homogenates ($\boldsymbol{Fc}$) and the first two lots of cooked purees ($\boldsymbol{Pc}$),

and the validation matrices with the third batch of raw apple homogenates ($\boldsymbol{Fv}$) and the third lot of cooked purees ($\boldsymbol{Pv}$) (**Fig. 51**).

In a second step, DS was performed separately on the calibration matrices of raw apple homogenates ($\boldsymbol{Fc}$) and cooked purees ($\boldsymbol{Pc}$) in 2017 ($\boldsymbol{Fc}_{2017}$ and $\boldsymbol{Pc}_{2017}$) and 2019 ($\boldsymbol{Fc}_{2019}$ and $\boldsymbol{Pc}_{2019}$) (**Fig. 51**, **Fig. 52**):

- the calibration matrices of apples ($\boldsymbol{Fc}_{2017}$) and purees ($\boldsymbol{Pc}_{2017}$) were processed to obtain the $\boldsymbol{B_0}$ and $\boldsymbol{E_0}$ related to the effects of raw materials on the processed purees as follows:

$$\boldsymbol{Pc}_{2017} = \boldsymbol{Fc}_{2017}\boldsymbol{B_0} + \boldsymbol{E}_0 \qquad (2)$$

Both ($\boldsymbol{Fc}_{2017}$) and ($\boldsymbol{Pc}_{2017}$) have the same size $n \times \lambda$, where n = 48; (2 thinning practices × 4 storage periods × 2 apple batches/ puree lots × 3 spectral replicates.

- the calibration matrices of apples ($\boldsymbol{Fc}_{2019}$) and purees ($\boldsymbol{Pc}_{2019}$) were performed for each puree processing condition, as follows:

$$\boldsymbol{Pc}^{(i)}_{\boldsymbol{2019}} = \boldsymbol{Fc}_{2019}\boldsymbol{B}_i + \boldsymbol{E}_i \qquad (3)$$

where i from 1 to 9, corresponding to 9 different processing conditions (3 heating temperatures × 3 grinding speeds). To each spectral replicate of $\boldsymbol{Fc}_{2019}$ corresponds nine spectra according to each processing condition ($\boldsymbol{Pc}_{2019}$). All the spectra of $\boldsymbol{Pc}_{2019}$ matrix corresponding to the same processing conditions were gathered in a specific matrix $\boldsymbol{Pc}^{(i)}_{\boldsymbol{2019}}$. The size of this matrix (one for each processing condition) is equal to that of raw apple homogenates $\boldsymbol{Fc}_{2019}$ (n = 6 spectra (2 apple batches × 3 spectral replicates) × λ).

Thirdly, once all the transfer $\boldsymbol{B}$ ($\boldsymbol{B}_0$ and $\boldsymbol{B}_i$) and error $\boldsymbol{E}$ ($\boldsymbol{E}_0$ and $\boldsymbol{E}_i$) matrices were computed, they were used to calculate the cooked puree reconstructed calibration and validation spectra matrices, as follows (**Fig. 51**):

$$\boldsymbol{Tc}_{2017} = \boldsymbol{Fc}_{2017}\boldsymbol{B}_0 + \boldsymbol{E}_0 \qquad (4)$$

$$\boldsymbol{Tv}_{2017} = \boldsymbol{Fv}_{2017}\boldsymbol{B}_0 + \boldsymbol{E}_0 \qquad (5)$$

$$\boldsymbol{Tc}_{2019} = \boldsymbol{Fc}_{2019}\boldsymbol{B}_i + \boldsymbol{E}_i \qquad (6)$$

$$\boldsymbol{Tv}_{2019} = \boldsymbol{Fv}_{2019}\boldsymbol{B}_i + \boldsymbol{E}_i \qquad (7)$$

Finally, the reconstructed calibration and validation spectral matrices of cooked

puree, $\boldsymbol{Tc}$ ($\boldsymbol{Tc}_{2017}$ + $\boldsymbol{Tc}_{2017}$) and $\boldsymbol{Tv}$ ($\boldsymbol{Tv}_{2017}$ + $\boldsymbol{Tv}_{2019}$) of the two years (2017 and 2019) were obtained with the same sizes of the real spectral matrices of cooked puree, $\boldsymbol{Pc}$ and $\boldsymbol{Pv}$, for the further multivariate regressions.

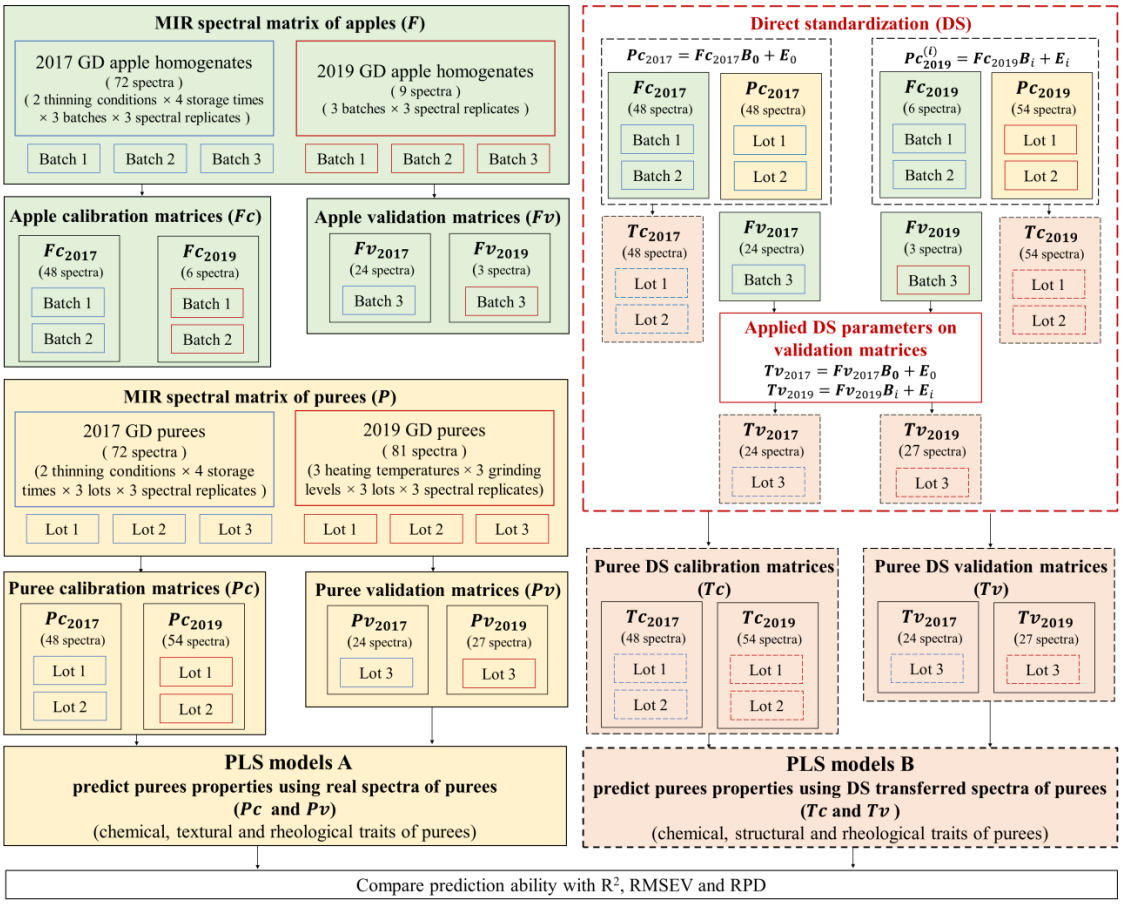

Fig. 51. Overview of MIR spectra pre-processing, direct standardization (SD) and multivariate regression.

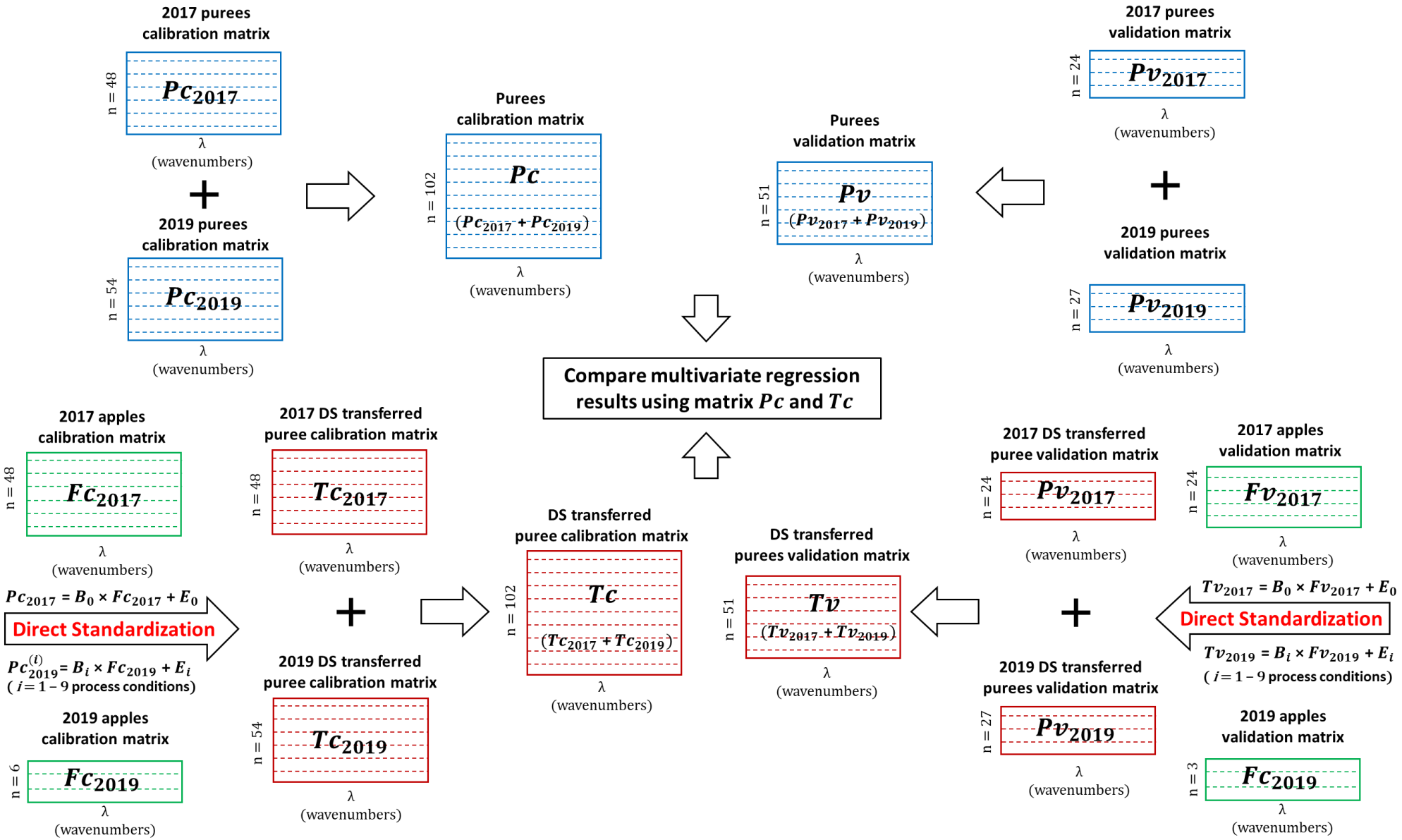

Fig. 52. Overview of the applied methodology to exploit reconstructed MIR spectra of purees and multivariate regression.

2.6 Multivariate regression

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). After pretests of several pre-processing treatments (baseline correction, standard normal variate (SNV) and a derivative transform calculation using Savitzky–Golay method (window size = 11, 21, 31) applied on several different spectral regions, the best results of prediction and discrimination were obtained on the range 1800-900 cm^{-1}, which has been already highlighted (**Paper II**), using SNV pre-processing. Principal Component Analysis (PCA) and Factorial Discriminant Analysis (FDA) were applied on SNV pre-treated spectra of cooked purees to detect their differences related to the variability of both, raw apples and processing conditions. The specificity and sensitivity values of FDA discriminations were calculated by the already reported method of Nargis et al. (2019).

PLS models were developed using the SNV pre-processed puree spectra (1800-900 cm^{-1}) of the calibration set $\boldsymbol{Pc}$ and the DS transferred spectra of purees ($\boldsymbol{Tc}$), corresponding to the same reference dataset. The two calibration matrices of cooked purees included a total of 102 spectra (48 spectra in 2017: 2 thinning practices × 4 storage periods × 2 lots × 3 spectral replicates, and 54 spectra in 2019: 3 heating temperatures × 3 grinding speeds × 2 lots × 3 spectral replicates). Then, the developed PLS models were applied on their corresponding validation spectra sets of $\boldsymbol{Pv}$ and $\boldsymbol{Tv}$, with a total of 51 spectra in 2017: 2 thinning practices × 4 storage periods × 1 lot × 3 spectral replicates and in 2019: 3 heating temperatures × 3 grinding speeds × 1 lot × 3 spectral replicates. PLS model performance was assessed using the determination coefficient of calibration (R_c^2) and validation (R_v^2), the root-mean-square error of validation (RMSEV), the number of latent variables for calibration (LVs), the residual predictive deviation of validation set (RPD), as described by Nicolai et al. (2007). The linkable spectral regions of the acceptable PLS models presenting RPD values higher than 2.5 (Nicolai, Beullens, Bobelyn, Peirs, Saeys, Theron, et al., 2007) were displayed based on their β- coefficients (**Table 43** and **Table 44**).

3. Results and discussion

3.1 Variability of cooked purees based on their MIR Spectra

3.1.1 Variability induced by the raw materials

According to ANOVA (F-values), fruit thinning applied on apples during their growth in orchard resulted in a significant variation ($p < 0.001$) of viscosity (η_{50} and η_{100}), viscoelasticity (G', G'', yield stress), particle sizes (d4:3 and d3:2) and compositions (DMC, SSC, TA, pH, malic acid, sucrose, fructose and AIS) of the cooked purees. Particularly, the impact of thinning on the viscosity, DMC and SSC of purees was higher than the effect of post-harvest storage at 4°C (**Table 40**). Purees processed from thinned apples (Th+) had higher viscosity values (η_{50} and η_{100}) and larger particle sizes (d4:3) than those from the non-thinned apples (Th-), observed after three or six months of cold storage (T3 and T6) (Buergy, Rolland-Sabaté, Leca, & Renard, 2020). Moreover, an intensive decrease of average particle sizes (d4:3) was observed in the purees cooked with the apples stored one month at 4°C (T1) for both 'thinning' (Th+) and 'non-thinning' (Th-) treatments. PCA applied on the spectra of cooked purees in 2017 showed a good ability to detect the effects of treatments applied on raw apples (fruit thinning and storage periods) (**Fig. 53a, b**). The effect of thinning on the first principal component (PC1 90.1%) was much higher than that of storage on the second principal component (PC2 6.9%), which was in line with our previous results (**Paper II**). In addition, the increase of the band at 1022 cm^{-1} and the decrease of the bands at 1061-1065 cm^{-1}, attributed to sucrose and fructose respectively (Bureau, Cozzolino, & Clark, 2019), were the major contributors of the observed discriminations on the two PCs (**Fig. 53c, d**).

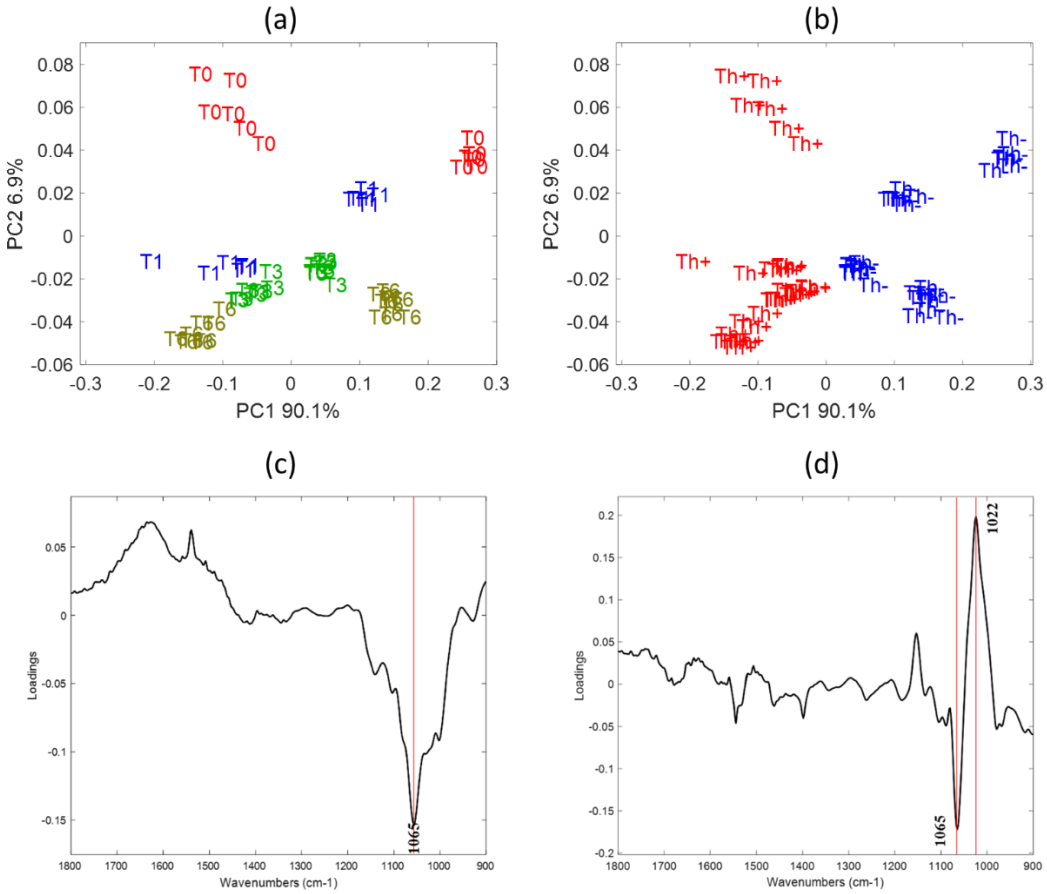

Fig. 53. PCA on the SNV pre-treated MIR spectra (900-1800 cm^{-1}) of purees cooked with thinned (Th+) and non-thinned (Th-) 'Golden Delicious' apples stored at 4°C during 0, 1, 3 and 6 months (T0, T1, T3 and T6): **(a)** the scores plot of the first two components (PC1 and PC2) related to fruit thinning; **(b)** the scores plot of the first two components (PC1 and PC2) related to storage periods; **(c)** the loading plot of PC1; **(d)** the loading plot of PC2.

3.1.2 Variability induced by processing conditions

The different grinding speeds affected significantly ($p < 0.05$) the viscosity (η_{50} and η_{100}), viscoelasticity (G' and G'') and particle size (d4:3 and d3:2) of the cooked purees (**Table 41**). Particularly, the increase of grinding speed significantly ($p < 0.001$) decreased viscosity (η_{50} and η_{100}), viscoelasticity (yield stress, G' and G'') and particle sizes (d4:3) which was observed at each tested temperature. From macroscopic images of purees (data not shown), the larger particles disappeared with increasing grinding speeds, which was enough to cause a decrease in the puree viscosity (Espinosa-Muñoz,

Renard, Symoneaux, Biau, & Cuvelier, 2013). Inversely, the increasing heating temperatures induced no significant ($p > 0.05$) changes of puree viscosity (η_{50} and η_{100}) and viscoelasticity (yield stress, G' and G''). The highest heating temperature (95°C) resulted in a significant ($p < 0.05$) increase of DMC and SSC and a decrease of TA and malic acid. Consequently, the changes of grinding speed during the puree processing significantly modified the structural properties and viscoelastic behaviors of purees, whereas heating temperature affected strongly the biochemical composition of purees.

FDA performed on the cooked puree spectra in 2019 successfully classified the processing changes induced at the different heating temperatures (**Fig. 54a**) and grinding speeds (**Fig. 54b**). The samples cooked at 95 °C were well-separated from the other two conditions (4 factors, 100% of sensitivity and specificity in **Table 42a**), according to the first factorial component (F1) (**Fig. 54a**). The specific bands at 1745 cm^{-1} and 1539 cm^{-1} were attributed to the increase of soluble pectins, probably in relationship with their solubilization in puree serum from apple cell walls, enhanced with the increasing heating temperature (Liu, Renard, Rolland-Sabaté, Bureau, & Le Bourvellec, 2020). Moreover, the negative peaks at 1057 cm^{-1} and 998 cm^{-1} could be due to the hydrolysis of sucrose during thermal processing, thus resulting in the increase of fructose (1022 cm^{-1}) and glucose (1107 cm^{-1}) contents.

The three different grinding levels could be discriminated according to the first two factorial components (F1 and F2) (**Fig. 54b**), especially for the highest grinding speed at 3000 rpm ('G3' in **Fig. 54b**) (4 factors, over 85.19% of specificity and sensitivity in **Table 42b**). The intensive negative spectral peaks at 1558, 1539-1541, 1508 and 1458 cm^{-1} along both discriminant factors (F1 and F2 in **Fig. 54d and Fig. 54e**) were all located in the region between 1450 cm^{-1} and 1600 cm^{-1}, which has been already attributed to the changes of particle size and rheological behavior after apple puree mechanical refining in a previous experiment (**Papers I and II**). These peaks indicated the decrease of particle size (d4:3 and d3:2) and viscosity of purees with increasing grinding speeds, which was in line with our reference measurements (**Table 41**) and Espinosa et al. (2011).

Briefly, MIR technique could detect several variability sources such as thinning

practices during fruit cultivation, cold storage and processing conditions (temperature and grinding) in the cooked purees. In addition, the spectral region 1450-1750 cm^{-1} was validated as being a reliable analytical signal of processing linked to the textural and rheological changes in purees.

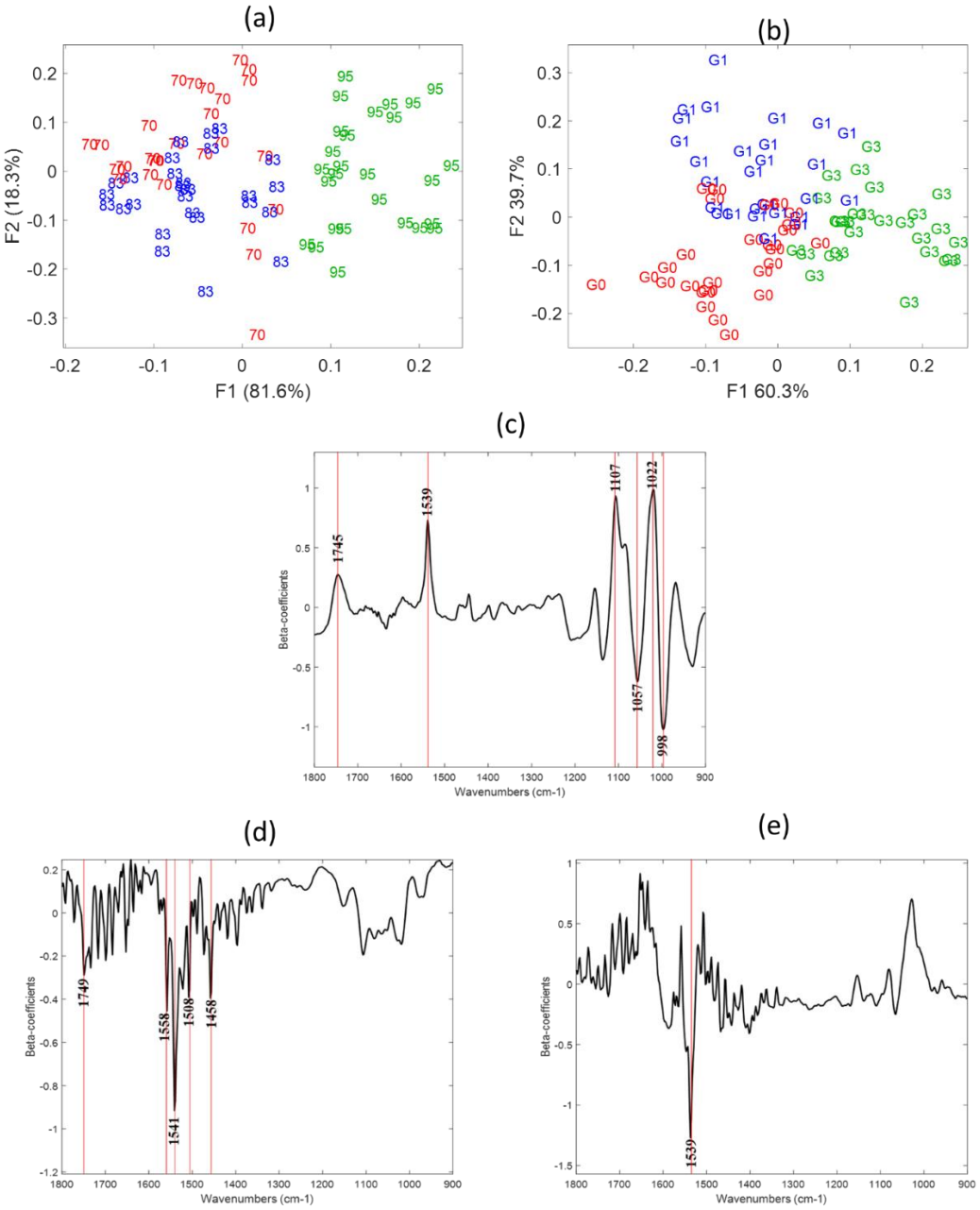

Fig. 54. Maps of FDA performed on the SNV pre-treated MIR spectra (900-1800 cm^{-1}) of purees cooked with: **(a)** three different temperatures (70 °C, 83 °C and 95 °C) and **(b)** three grinding speeds (G0 at 300 rpm, G1 at 1000 rpm and G3 at 3000 rpm); **(c)** the first factorial score ('F1') of heating temperature discrimination; **(d)** the first factorial

score ('F1') of grinding discrimination; **(e)** the second factorial score ('F2') of grinding discrimination.

3.2 Prediction of quality traits of cooked purees by MIRS

For all developed PLS models, as expected the decreases of determination coefficients between the calibration set (R_c^2) and the validation set (R_v^2) were observed in **Table 43**. According to RPD values over 2.5 (Nicolai, Beullens, Bobelyn, Peirs, Saeys, Theron, et al., 2007), prediction was acceptable to good (RPD from 2.6 to 3.3) for viscosity (η_{50} and η_{100}), average particle sizes (d4:3), SSC, TA, pH values and malic acid content in cooked purees by MIRS, taking into account a large variability of raw apple materials and processing conditions.

Apparent puree viscosity at a shear rate value of 50 s^{-1} (η_{50}), which has been described to be highly correlated with the in-mouth texture perception of fluid food (Chen & Engelen, 2012), could be predicted by MIRS with a R_v^2 of 0.87 and a RPD of 3.2. MIR prediction of apparent puree viscosity at a single shear rate value of 100 s^{-1} (η_{100}) (R_v^2 = 0.85, RPD= 3.0) observed here were much better than its prediction by NIRS (RPD = 1.3) (**Paper IV**), These results evidenced the possibility of MIRS to estimate puree viscosity. For the two apparent puree viscosity values measured at η_{50} and η_{100}, the main wavenumber regions at 1718-1730 cm^{-1} and 1618-1678 cm^{-1} were still observed in **paper II**. This could validate the application of MIRS to predict puree viscosity by taking into account not only the raw fruit variability but also the complex effects of processing conditions. However, the predictions (R_v^2 < 0.81, RPD < 2.1) of the viscoelastic parameters of purees (G', G'' and yield stress) were not precise enough to estimate the viscoelastic behaviors and the moment when puree starts to flow. Indeed, heating and grinding affected puree viscoelasticity (SD values of 2362 Pa for G', 595 Pa for G'', 27.1 Pa for yield stress) and resulted in more than twice higher variations of these parameters than those induced by thinning and cold storage on raw materials (SD values of 1001 Pa for G', 234 Pa for G'' and 12.9 Pa for yield stress). These new prediction accuracies of the viscoelastic parameters of purees (G', G'' and yield stress) were not as good as our previous ones by MIRS (**Paper II**), but could be more robust

to be considered for future applications. MIR coupled with the linear regression (PLS) showed a good performance (R_v^2= 0.87, RPD= 3.1) to evaluate the volume average particle size of purees (d4:3), but not the surface average particle size (d3:2). Particularly, the most informative wavenumbers to evaluate puree particle size at 1701-1713 cm^{-1}, 1655-1668 cm^{-1} and 1537-1541 cm^{-1} have been already observed previously to predict puree viscosity, to discriminate the purees prepared with different grinding speeds (mentioned in **part 3.1**) and with different refining levels (**Paper II**). Such good prediction of puree average particle size (d4:3) could not come from internal correlations with puree composition such as SSC, DMC or AIS contents because of their poor correlation ($R^2 < 0.48$), but probably some specific signals needing to be identified and confirmed. Moreover, we confirmed here the impossibility to predict the cell wall content directly in puree by MIRS, without any preparation such as freeze-drying (**Paper II**)

A good prediction of global puree quality traits, SSC (RPD= 3.1) and DMC (RPD= 2.9), was obtained with 5 LVs (**Table 43**). The variation of SSC and DMC in purees were highly correlated ($R^2 = 0.76$), which explained the good estimations of these two parameters. The respective fingerprint wavenumbers of SSC and DMC prediction were similar and detected at 996-1001 cm^{-1}, 1048-1057 cm^{-1} and 1109-1112 cm^{-1}, corresponding to the variations of sucrose, fructose and glucose in purees (Bureau, Cozzolino, & Clark, 2019). Moreover, the prediction of TA was excellent (RPD = 3.2), with a limited RMSEV of 7.6 mmol H^+/kg FW. The typical wavenumber region at 1709-1720 cm^{-1} in TA prediction has already been attributed to the C=O vibration of carboxylic acids (Bureau, Quilot-Turion, Signoret, Renaud, Maucourt, Bancel, et al., 2013; Clark, 2016). Depending on the good correlations between the different contributors of apple acidity ($R^2 = 0.81$ between TA and malic acid, $R^2 = 0.76$ between TA and pH), MIRS provided an acceptable prediction of pH ($R_v^2 = 0.83$ and RPD = 2.5) and malic acid content (R_v^2 = 0.85 and RPD= 2.7). Despite the similar typical fingerprints observed in the β-coefficients of PLS models of malic acid and pH, the relative lower RPD values and R_v^2 of pH compared to malic acid were probably due to the low pH variation. Concerning the main individual sugars, acceptable prediction was

obtained only for fructose (RPD = 2.6), but neither for sucrose (RPD= 1.3) nor for glucose (RPD = 1.5), which was in line with our previous results (**Paper II**). The lower concentration of glucose (10.4-25.4 g/kg FW) than the other individual sugars (34.9-98.7 g/kg FW of fructose, 39.1-118.5 g/kg FW of sucrose) led to its worse prediction by MIR results. A higher internal biochemical correlation between the major compounds (SSC, TA) and fructose (R^2 = 0.79 for SSC and fructose, R^2 = 0.76 for TA and fructose) in apple purees might explain the better prediction of fructose than the one of sucrose (R^2 = 0.58 for SSC and sucrose, R^2 = 0.44 for sucrose and TA).

Briefly, MIR technique can provide a simultaneous and robust estimation of biochemical compositions (dry matter, soluble solids, titratable acidity, pH, malic acid and fructose), rheological behaviors (viscosity at η_{50} and η_{100}) and particle size (d4:3) of apple purees, taking into account the large variability along the apple puree production chain (agricultural practices, post-harvest storage and processing conditions).

3.3 Reconstructed spectra for prediction of puree quality traits

In this part, MIR prediction models were developed using the reconstructed spectra of the calibration set of cooked purees ($\boldsymbol{Tc}$), only done for the well-predicted parameters mentioned in **part 3.2**, which were η_{50} and η_{100}, d4:3, SSC, DMC, TA, pH, malic acid and fructose. Then, these models were applied on the validation reconstructed spectra of the cooked purees ($\boldsymbol{Tv}$).

Overall, based on the PLS regression applied on the puree reconstructed spectra, acceptable predictions were obtained (RPD > 2.5) for rheological (η_{50}, η_{100}), structural (d4:3) and global biochemical (SSC, DMC, TA) parameters (**Table 44**). In contrast, predictions appeared not acceptable for malic acid (RPD = 2.3), fructose (RPD = 1.7) and pH (RPD = 2.1). Compared to the previous prediction from the real puree spectra (**Table 43**), lower R_v^2 and higher LVs have been generally obtained for all parameters giving some lower prediction performance (**Table 44**). Particularly, the use of the reconstructed spectra showed a good ability to predict puree viscosity parameters (η_{50} and η_{100}) with R_v^2 > 0.82, RPD > 2.5 and prediction errors (RMSEV) of 0.21 Pa.s and

0.10 Pa.s for η_{50} and η_{100}, respectively. These results were close to those from the real spectra of purees (**Table 43**). The fingerprint wavenumbers used in the PLS models were similar for both reconstructed and real spectra, mainly 1718-1734 cm^{-1}, 1616-1336 cm^{-1} and 1547-1553 cm^{-1} as described in **Part 3.2**. Although a relative lower R_v^2 and RPD (R_v^2 = 0.84 and RPD = 2.6) was obtained for particle size (d4:3) compared to the results on the real puree spectra (**Table 43**), the consistent fingerprints were highly related to the puree texture such as 1701-1715 cm^{-1}, 1537-1541 cm^{-1} and 1101-1107 cm^{-1}. These prediction performances revealed for the first time the possibility to evaluate the variation of average particle sizes in the cooked purees based on the MIR information of the corresponding raw apple homogenates. Considering the other global quality parameters, acceptable predictions were obtained for SSC (R_v^2 = 0.85 and RPD = 2.8) and DMC (R_v^2 = 0.84 and RPD = 2.6) contents. The specific wavenumbers in the ranges 997-1001 cm^{-1} and 1048-1057 cm^{-1} for sucrose and in the ranges 1009-1112 cm^{-1} for fructose mainly contributed to the PLS models for both reconstructed and real spectra. These ranges have been already mentioned to be linked to these sugars (Bureau, Cozzolino, & Clark, 2019), which are the main ones in apples. For acidity, the reconstructed spectra gave an excellent prediction of TA (R_v^2 = 0.86 and RPD = 2.9), using the spectral regions between 1709-1720 cm^{-1} of the typical C=O absorption (Clark, 2016). Consequently, PLS applied on the reconstructed MIR spectra calculated from the spectra of raw apple homogenates showed the possibility to directly predict the viscosity, averaged particle sizes, SSC, DMC and TA of cooked purees.

Several initial attempts have been tested to monitor the quality of cooked food from infrared information of their raw materials, with the objectives to predict the texture of cooked poultry pectoralis major muscles (Meullenet, Jonville, Grezes, & Owens, 2004), of cooked rice (Windham, Lyon, Champagne, Barton, Webb, McClung, et al., 1997) and of apple purees (**Papers IV, V**). In these studies, the spectra matrix of the raw materials and the reference data of the corresponding processed materials were used to calibrate models. The predictions thus obtained are mainly due to the strong internal correlations of quality traits between materials before and after processing, which could provide semi-quantitative prediction accuracy for practical uses. However, the internal

correlations of quality traits during fruit processing still remain unreliable when using a large variability of raw materials and various industrial processing systems (**Papers IV, V**). Further, such a direct modelling method requires a necessary step of acquisition of the infrared information on raw material batches for each processing condition, in order to obtain the same matrix sizes of spectra and reference data for calibration.

Here, a potential strategy has been firstly proposed to build reconstructed MIR spectra of processed purees from the spectra of raw apple homogenates using a spectral transfer method. The high consistency of the specific fingerprints used in the PLS models obtained for both the real spectra and the reconstructed spectra, confirmed our choice for this modelling strategy. Compared to the direct modelling method, a great advantage of using spectral transfer strategy is that the calibration dataset only needs the infrared information and reference data of a few processed purees and just a limited number of spectra of corresponding raw apples. For example, in our dataset of 2019, the reconstructed spectra of 27 different processed purees could be transferred from only 3 corresponding spectra of the same apple batches.

After a simple scanning of raw apple homogenates by MIRS, the models revealed the possibility to i) predict the quality of apple purees, such as viscosity, SSC and TA using a standard processing recipe (95 °C for 5 min and grinding at 1500 rpm), even though a large variability of raw apples was used (different fruit thinning and cold storage periods); and ii) to monitor and anticipate the organoleptic properties of cooked purees under different processing strategies, which is relevant for the processors and market. For example, a higher viscosity and acidity in-mouth feeling (predicted η_{50} = 1.42 ± 0.09 Pa.s, predicted TA = 65.8 ± 3.5 meq/kg FW) were predicted with the recipe at 83 °C for 30 min and grinding speed of 1000 rpm than with the recipe at 95 °C for 30 min and grinding speed of 3000 rpm (predicted η_{50} = 0.98 ± 0.14 Pa.s, predicted TA = 56.4 ± 4.5 mmol H^+/kg FW).

4. Conclusion

As far as we know, this is the first study that shows the ability of MIRS to estimate the quality of processed fruit products taking into account a large variability coming

from agricultural practices, post-harvest storage and processing conditions along the whole processing chain. MIR technique provided reliable assessment of viscosity, averaged particle sizes and major compositions (SSC, DMC, TA and malic acid) of apple purees.

Further, a simple spectroscopic transfer method (direct standardization) was applied for the first time to develop the reconstructed spectra of purees from their corresponding spectra of raw apple homogenates. MIRS coupled with PLS regression obtained acceptable predictions of TA, DMC, SSC, viscosity (η 50 and η 100) and averaged particle sizes of the final puree based on their reconstructed spectra. With a simple scanning of raw apple homogenates, MIR technique opens the possibility to i) predict the quality of final purees under a standard processing procedure, which is beneficial for fruit processing sustainability; and even ii) to monitor the texture and tastes of purees under different processing conditions for a better management.

Table 40. Biochemical, structural, and rheological data of apple purees and ANOVA results.

Fruit thinning	Storage periods	η_{50} Pa.s	η_{100} Pa.s	G' Pa	G'' Pa	Yield stress Pa	d 4:3 -	d 3:2 -	DMC g/g FW	SSC °Brix	TA mmol H+/kg FW	glucose g/kg FW	fructose g/kg FW	sucrose g/kg FW	malic acid g/kg FW	pH	AIS mg/g FW	AIS mg/g DW
Th-	T0	1.28	0.77	3127.8	626.7	47.5	909.9	251.5	0.19	13.4	58.1	18.9	50.5	66.7	4.5	3.7	164.5	31.6
	T1	1.13	0.70	1960.2	466.7	21.9	694	351.9	0.19	15.0	54.4	15.4	49.4	59.1	2.8	3.8	147.2	27.6
	T3	0.87	0.55	1849	453	13.9	339.8	205.9	0.20	14.1	46.7	18.6	84.1	84.8	3.6	4.0	140.0	27.3
	T6	0.92	0.50	1816	427	14	316.1	223.6	0.19	13.8	26.8	23.0	85.1	77.3	2.7	4.4	145.7	27.6
Th+	T0	1.75	0.97	3375.1	816.4	52.1	831.6	231.6	0.21	15.5	70.9	23.5	85.3	64.4	5.5	3.6	163.3	33.9
	T1	1.54	0.94	2783.7	639.5	25.2	489	261.8	0.21	17.6	69.3	16.8	80.3	115.9	5.6	3.8	150.9	31.9
	T3	1.25	0.70	2517.6	609	22.3	405.1	228.3	0.22	16.9	59.9	13.8	88.0	102.5	4.9	3.8	143.3	31.6
	T6	1.60	0.88	3168.2	751.7	33.9	393.5	255.1	0.23	17.5	34.7	23.8	95.7	44.0	3.6	4.3	150.3	34.8
Storage time	significance	***	***	***	***	***	***	***	**	**	***	***	***	***	***	***	*ns*	*
	F-values	*20.4*	*13.6*	*24.7*	*15.0*	*72.8*	*216.1*	*41.5*	*6.6*	*8.3*	*279.4*	*74.7*	*76.2*	*38.5*	*13.8*	*436.3*	*1.8*	*4.3*
Fruit thinning	significance	***	***	***	***	***	***	*	***	***	***	*ns*	***	**	***	***	**	**
	F-values	*138.2*	*61.5*	*67.3*	*91.5*	*29.6*	*47.2*	*5.5*	*157.7*	*115.7*	*176.9*	*1.3*	*187.8*	*13.8*	*57.9*	*48.8*	*15.9*	*10.3*

Data expressed in fresh weight (FW) or dry weight (DW); values correspond to the mean of 3 puree replications (3 kg per replication). Raw apples were stored at 4°C: from harvest (T0) and during one (T1), three (T3) and six months (T6). Two conditions of fruit load during cultivation: non-thinning with 100% number of apples (Th-) and thinning with 50% number of apples (Th+) per tree. In grey, two way- ANOVA results obtained for Golden Delicious purees. ns, *, **, ***: Non significant or significant at $P < 0.05$, 0.01, 0.001 respectively.

Table 41. Biochemical, textural, and rheological data of apple purees and results of Kruskal-Wallis non-parametric test.

Temperatures	Grinding speeds	η_{50}	η_{100}	G'	G''	Yield stress	d 4:3	d 3:2	DMC	SSC	TA	glucose	fructose	sucrose	malic acid	pH	AIS	AIS
°C	rpm	Pa.s	Pa.s	Pa	Pa	Pa	-	-	g/g FW	°Brix	mmol H+/kg FW	g/kg FW	g/kg FW	g/kg FW	g/kg FW		mg/g FW	mg/g DW
70	300	1.27	0.89	9629.8	2389.5	97.7	583.8	103.6	0.17	13.8	63.6	16.1	67.9	70.9	6.3	3.9	26.8	158.0
	1000	1.42	0.91	3295.4	768.8	38.8	633.7	274.8	0.16	13.8	67.8	18.0	66.2	76.6	6.4	3.9	28.0	171.4
	3000	0.64	0.40	1111.3	215.6	10.2	353.5	207.6	0.17	14.4	69.1	17.3	65.5	86.0	6.6	3.9	28.1	165.1
83	300	1.23	0.93	8437.5	2078.9	97.9	553.0	212.8	0.18	14.9	59.9	14.6	67.7	70.6	5.8	3.8	29.5	166.7
	1000	1.38	0.87	3036.9	764.4	35.6	647.8	324.8	0.16	13.4	69.8	17.4	71.9	75.2	5.2	3.9	25.6	159.6
	3000	0.91	0.54	1312.2	259.9	11.8	297.4	192.4	0.17	14.8	70.3	16.6	66.8	76.1	5.7	3.9	28.2	162.6
95	300	1.93	1.16	3708.2	1101.3	30.0	492.9	262.1	0.17	14.9	60.2	17.0	72.6	66.9	4.8	3.9	27.9	160.8
	1000	1.44	0.84	1955.4	522.2	16.6	332.9	209.8	0.17	14.9	64.1	17.8	71.3	64.5	5.1	3.9	26.2	157.3
	3000	1.07	0.63	1399.4	362.2	14.0	206.4	153.2	0.17	14.8	62.0	18.9	69.2	65.8	5.4	3.8	26.2	154.1
Temperature	significances	*ns*	*ns*	*ns*	*ns*	*ns*	*	*ns*	*	*	*	*	*	**	**	*ns*	*ns*	*ns*
	F-values	*4.1*	*1.3*	*0.9*	*0.4*	*1.5*	*7.0*	*1.2*	*6.3*	*6.8*	*8.4*	*6.1*	*8.9*	*9.1*	*10.2*	*2.1*	*2.3*	*4.2*
Grinding speeds	significances	**	***	***	***	***	**	**	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
	F-values	*10.9*	*17.2*	*21.6*	*22.6*	*19.4*	*13.7*	*9.2*	*1.1*	*2.8*	*1.4*	*0.4*	*0.1*	*1.4*	*0.2*	*1.3*	*0.4*	*0.2*

Data expressed in fresh weight (FW) or dry weight (DW); values correspond to the mean of 3 puree replications (3 kg per replication). Processing conditions variations were: three heating temperatures at 70°C, 83°C and 95°C for 30 min, and three grinding speeds at 300, 1000 and 3000 rpm at each temperature. In grey, Kruskal-Wallis results obtained on Golden Delicious purees. ns, *, **, ***: Non significant or significant at $P < 0.05$, 0.01, 0.001 respectively.

Table 42. The results of sensitivity (in blue cells) and specificity (in yellow cells) from: (a) the FDA discrimination (4 factors) of three different heating temperatures; and (b) the FDA discrimination (4 factors) of three different grinding speeds.

(**a**)

Temperatures (℃)	70	83	95
70	/	76.67%	100%
83	83.33%	/	100%
95	100%	100%	/

(**b**)

Grinding speeds (rpm)	300	1000	3000
300	/	75.00%	85.19%
1000	76.92%	/	87.10%
3000	85.19%	100%	/

Table 43. Prediction of biochemical, structural and rheological properties of apple purees using PLS regression based on their MIR spectra between 900-1800 cm^{-1}.

Parameter	Range	SD	R_c^2	R_v^2	RMSEV	RPD	LVs	Linkable regions (cm^{-1})
η 50	0.57-2.28	0.39	0.91	0.87	0.19	3.2	7	1780-1788, 1718-1734, 1659-1678, 1636-1616, 1549-1558, 1157-1163
η 100	0.26-1.22	0.27	0.89	0.85	0.07	3.0	8	1780-1788, 1718-1734, 1659-1678, 1636-1616, 1549-1558, 1157-1163
AS-G' (Pa)	1069-11154	2362	0.85	0.75	854	2.0	12	/
AS-G'' (Pa)	210-2707	595	0.87	0.73	298	1.9	12	/
yield stress	8.7-131	27.1	0.86	0.81	12.8	2.1	11	/
d 4:3	196-920	197	0.90	0.87	65	3.1	8	1780-1788, 1701-1713, 1649-1653, 1537-1541, 1101-1107, 1032-1043 1011-1026
d 3:2	44-360	60.1	0.76	0.62	41.7	1.4	11	/
AIS (DM)	133.1-193.6	12.4	0.79	0.50	7.6	1.4	10	/
AIS (FW)	21.8-14.1	3.4	0.80	0.50	2.2	1.5	6	/
DMC (g/g FW)	0.15-0.23	0.02	0.87	0.84	0.01	2.9	5	997-1001,1051-1057, 1101-1109
SSC (°Brix)	12.6-18.6	1.9	0.91	0.86	0.6	3.1	5	997-1001,1051-1057, 1101-1109
TA (mmol H^+/kg FW)	5.1-73.5	25.5	0.89	0.86	7.6	3.2	6	1713-1709, 1105-1109, 1016-1018, 1074-1072, 1038-1042
pH	3.6-4.4	0.2	0.89	0.83	0.1	2.6	7	1713-1709, 1105-1109, 1016-1018, 1074-1072, 1038-1042
malic (g/kg FW)	2.4-7.0	1.2	0.88	0.85	0.4	2.7	6	1721-1709, 1105-1109, 1016-1018, 1074-1072
fructose (g/kg FW)	34.9-98.7	16.0	0.89	0.84	6.1	2.6	9	1709-1713, 1259-1265, 1105-1109, 1074-1080,1038-1042, 1016-1020, 970-974
sucrose (g/kg FW)	39.1-118.5	18.5	0.79	0.64	14.0	1.3	9	/
glucose (g/kg FW)	10.4-25.4	3.6	0.74	0.68	2.4	1.5	7	/

Note: Puree spectra and references data from 'Golden Delicious' apples, including variability of two different thinning conditions, cold storage (during 0, 1, 3 and 6 months), three heating temperatures (70, 83 and 95 °C) and three grinding levels (300, 1000, 3000 rpm). All results based on the SNV pre-treated MIR spectra at 900-1800 cm^{-1}. R_c^2: determination coefficient of the calibration set; R_v^2: determination coefficient of the validation set; RPD: the residual predictive deviation of validation set; the linkable regions based on the β-coefficients of PLS models with the RPD values higher than 2.5; "/" presented the unacceptable results with the RPD values lower than 2.5.

Table 44. Prediction of biochemical, structural and rheological properties of apple purees using PLS regression based on their reconstructed MIR spectra of raw apple homogenates between 900-1800 cm^{-1}.

Parameter	Range	SD	R_c^2	R_v^2	RMSEV	RPD	LVs	Linkable regions (cm^{-1})
η 50	0.57-2.28	0.39	0.85	0.82	0.21	2.5	8	1720-1734, 1636-1614, 1556-1560, 1547-1533, 1506-1510, 1448-1470, 1157-1169
η 100	0.26-1.22	0.27	0.86	0.83	0.10	2.6	9	1720-1734, 1661-1675, 1636-1616, 1549-1558, 1507-1512,1445-1468, 1157-1163
d 4:3	196-920	197	0.89	0.84	76	2.6	9	1740-1745, 1701-1715, 1645-1659, 1583-1587, 1537--1541, 1508-1510, 1452-1470, 1100-1112
DMC (g/g FW)	0.15-0.23	0.02	0.87	0.84	0.01	2.6	6	1161-1165, 1101-1107, 1084-1090, 1051-1063, 989-1001
SSC (°Brix)	12.6-18.6	1.9	0.89	0.85	0.7	2.8	5	1101-1112, 1084-1090, 1051-1069, 997-1001
TA (mmol H^+/kg FW)	5.1-73.5	25.5	0.88	0.86	8.9	2.9	7	1715-1710, 1107-1113, 1082-1086, 1059-1063, 1038-1042, 1001-993
pH	3.6-4.4	0.2	0.84	0.79	0.1	2.1	8	1715-1709, 1105-1110, 1016-1018, 1074-1072, 1038-1042
malic (g/kg FW)	2.4-7.0	1.2	0.87	0.82	0.6	2.3	9	1713-1709, 1105-1109, 1080-1088, 1058-1064, 1016-1018, 1001-998
fructose (g/kg FW)	34.9-98.7	15.0	0.82	0.72	8.9	1.7	11	/

Note: Puree spectra and references data from 'Golden Delicious' apples, including variability of two different thinning conditions, cold storage (during 0, 1, 3 and 6 months), three heating temperatures (70, 83 and 95 °C) and three grinding levels (300, 1000, 3000 rpm). All results based on the SNV pre-treated MIR spectra at 900-1800cm^{-1}. R_c^2: determination coefficient of the calibration set; R_v^2: determination coefficient of the validation set; RPD: the residual predictive deviation of validation set; the linkable regions based on the β-coefficients of PLS models with the RPD values higher than 2.5; "/" presented the unacceptable results with the RPD values lower than 2.5.

Highlights of Paper VI

First, this study provided the answers of the last question in Part 2:

- The MIR spectra of cooked purees processed under different cooking conditions were calculated from the spectra of corresponding homogenized raw apples by direct standardization.
- PLS models using the spectra of homogenized raw apples predicted the titratable acidity (RPD = 2.9), soluble solid content (RPD = 2.8), particle averaged size (RPD = 2.6) and viscosity (RPD ≥ 2.5) of cooked purees.

MIR technique predicted the quality of cooked purees obtained from different cooking conditions and thus could improve the product sustainability or even help the development of new products in apple industry.

Part 3. Management of apple puree variability: puree formulation guided by infrared spectroscopy

Part 2 described the possibility to predict quality of cooked purees from the VIR-NIR and MIR spectral information of raw apples. These works are just the initial step in food processing management. For fruit processors, knowing predicted properties of final products after processing is not sufficient to produce the expected and constant final purees. The processing guidance (temperature, time, grinding speed…) should be provided and associated with the prediction models. Moreover, creating stable or even new food products from the mixture of a chosen variability of already processed materials, using the spectra information coupled with advanced models, should be also relevant. Until now, there is no report on infrared spectroscopy to guide the fruit processing and monitor the quality of final products after formulation.

Part 3 described an innovative concept concerning the feasibility of using infrared spectroscopy to:

i) Trace the composed varieties and compositions of formulated purees

ii) Develop an innovative chemometric method using infrared data to control the quality of final apple purees issued from the mix of single-variety purees using spectra reconstruction.

Four single-variety purees were prepared and mixed two by two with 9 different proportions (by weight) from 5% to 95%. The concentration profiles from multivariate curve resolution-alternative least squares (MCR-ALS) were tested to reconstruct spectra of formulated purees, thus giving the possibility to optimize admixtures of purees from their composed single-variety purees.

The results of Part 3 are presented as a paper (Paper VII)

Paper VII (Published)

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Visible, near- and mid-infrared spectroscopy coupled with an innovative chemometric strategy to control apple puree quality

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

[a] UMR408 Sécurité et Qualité des Produits D'Origine Végétale, INRAE, Avignon University, F-84000, Avignon, France
[b] Unité InfoSol, INRAE, Orléans, 45075, France
[c] TRANSFORM Division, INRAE, 44000, Nantes, France
[d] INRAE-ONIRIS Unité Statistiques, Sensométrie, Chimiométrie (Stat SC), F-44322, Nantes, France

1. Introduction

Apple puree is an ideal source of healthy constituents such as polyphenols and fibers (Le Bourvellec et al., 2011) and can be used as an intermediate for smoothies, fruit sauce, pie fillings and fruit-based baby food (Opatová, Voldřich, Dobiáš, & Čurda, 1992). Puree quality characteristics vary with fruit genetics (Rembiałkowska, Hallmann, & Rusaczonek, 2007), storage (Loncaric, Dugalic, Mihaljevic, Jakobek, & Pilizota, 2014), cooking parameters (Picouet, Landl, Abadias, Castellari, & Viñas, 2009), grinding intensity (Espinosa et al., 2011) and refining (**Papers I, II and IV**). In order to reach an apple puree with anticipated and constant taste and texture, a mixture of proportions of different apple varieties is generally done, presenting also the most economic and efficient strategies for manufacturers (O'sullivan, 2016). Thus, developing rapid and reliable approaches to determine the puree formulation, including fruit varieties and the proportions of each one, could be highly beneficial for fruit processed products and traceability control.

Infrared spectroscopy (visible-near and mid infrared), known as a rapid, relatively cheap, easy-to-use, non-destructive and automatable technique, has been applied for detecting adulterations in mixed purees of different fruit species (Contal, León, & Downey, 2002; Defernez, Kemsley, & Wilson, 1995; Kemsley, Holland, Defernez, & Wilson, 1996). Particularly, the MIR technique combined with partial least squares

discrimination analysis (PLS-DA) detects the presence of apple starting at 20% in apple-raspberry mixed purees (Kemsley, Holland, Defernez, & Wilson, 1996). Similar detectable limits are obtained using Vis-NIRS coupled with a principal component analysis (PCA) and a linear discriminant analysis (LDA) in apple-strawberry mixed purees (Contal, León, & Downey, 2002). However, so far, there has been no attempt to use such approaches for more advanced works on purees of apples only, but resulting from various proportions of different varieties.

Further, for fruit processors, the ever-increasing variability of raw fruits may mean that their empirical knowhow may not be sufficient to produce expected and constant final purees. The challenge is therefore to provide specific guidance for formulation of final purees based on information of individual batches of single variety puree. Multivariate curve resolution-alternative least square (MCR-ALS) has been widely used to simultaneously elucidate the pure spectra of different species present in processed products and their concentration profiles (de Juan & Tauler, 2006), such as edible oils from different vegetable sources (Le Dréau, Dupuy, Artaud, Ollivier, & Kister, 2009) and fruit juices with various organic acids (Silva, Lourenço, & de Araujo, 2018). The interest of this approach is to reconstruct the spectra of final processed products (in our case, formulated purees) according to the relative spectra of individual components (here single variety purees) by MCR-ALS. If so, the predictive models of processed puree quality traits (physical and chemical) using the reconstructed spectra dataset could open the possibility to provide a multicriteria optimization of puree formulation based on the prior information of single variety purees.

Partial least squares (PLS), a typical linear algorithm, has been used to successfully determine the global quality parameters of apple purees using NIRS information, such as titratable acidity, dry matter and soluble solids (**Papers IV, V**). However, the overlapping of absorption bands linked to non-linear rheological variations gave poor prediction of puree's texture by PLS regression. Machine learning approaches, such as random forest (RF) and Cubist, have been specially constructed to address large and complex nonlinear systems. Indeed, RF algorithm allows a better detection of adulteration in formulated oils than PLS (de Santana, Borges Neto, & Poppi, 2019).

Cubist regression working as decision tree models, gives a higher prediction accuracy than RF and PLS regression in palm-based cooking oil (Goh et al., 2019).

Accordingly, Vis-NIRS, MIRS and the combination of both (CB) infrared spectra coupled with machine learning (RF and Cubist) and PLS regressions were applied in our work on apples to: i) assess the possibility to detect the proportions of specific variety purees in formulated purees and evaluate the limits of the detection; ii) build models to evaluate the quality parameters of formulated purees obtained from different proportions of single-variety purees; and then iii) use information of single-variety purees to reconstruct spectra of formulated purees by MCR-ALS and investigate the possibility to develop regression models to guide formulation (admixing of purees from different varieties) for specific quality of final purees.

2. Material and methods

2.1 Apple purees

2.1.1 Purees processing

Apples of four varieties: 'Golden Delicious'(GD), 'Granny Smith'(GS), 'Braeburn'(BR) and 'Royal Gala'(GA) were harvested at a commercial maturity from La Pugère experimental orchard (Mallemort, Bouches du Rhône, France) in 2019, and stored for up to 2 months at 4 °C and around 90% relative humidity to ensure starch regression. After sorting and washing, on three consecutive weeks, a batch of each apple variety (2 kg) was processed into purees in a multi-functional processing system (Roboqbo, Qb8-3, Bentivoglio, Italy) following a Hot Break recipe: cooked at 95°C for 5 min at a 1500 rpm grinding speed, then cooled down to 65°C while maintaining the grinding speed. Finally, processed purees were conditioned in two hermetically sealed cans: one was cooled in a cold room (4°C) before formulation, while the other was stored at -20°C for biochemical measurement of individual sugars (fructose, sucrose and glucose) and malic acid.

2.1.2 Puree formulations

After processing the single-variety purees, a total of 6 experimental groups (named

A to E) were prepared, each, with two apple varieties (**Fig. 55**). Each group (A-F) included 9 samples with different formulated proportions of weight, was divided into two subsets: the first including 6 proportions (10%-90%, 25%-75%, 50%-50%, 75%-25%, 90%-10%, 95%-5%) for the modeling set, while the second with 3 proportions (80%-20%, 33%-67%, 14%-86%) for the external prediction set. Finally, spectral measurements (Vis-NIR and MIR), chemical (soluble solids, titratable acidity, pH, dry matter) and physical (color and rheological tests) characterizations were performed on each sample (single and formulated purees).

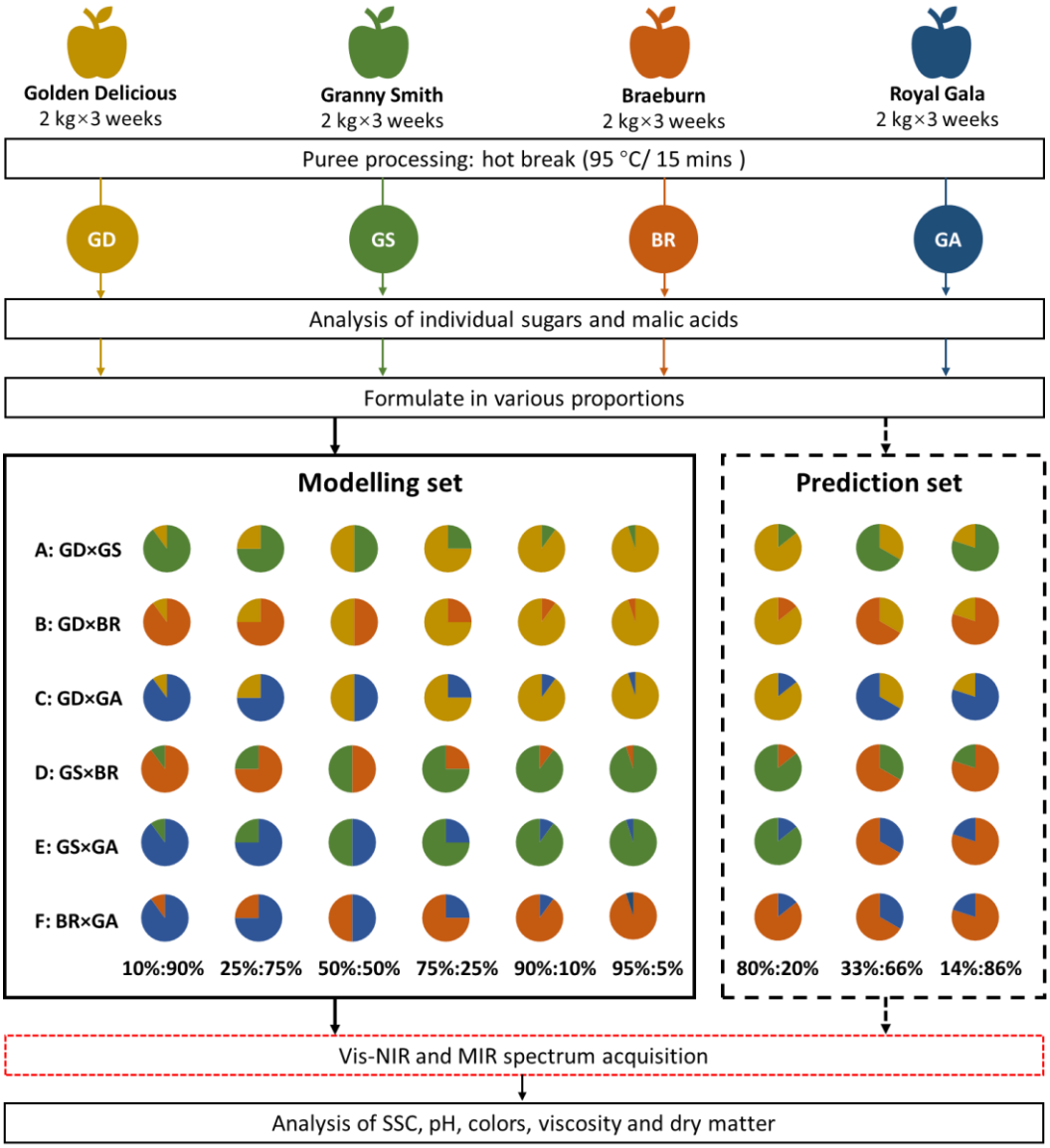

Fig. 55. Experimental scheme of purees reformation, quality characterizations and spectral acquisition.

2.2 Determination of quality traits

2.2.1 Physical characterizations

The puree color was determined three times through a dedicated glass cuvette using a CR-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). The puree rheological measurements, as flow curves, were carried out using a Physica MCR-301 controlled stress rheometer (Anton Paar, Graz, Austria) and a 6-vane geometry (FL100/6W) with a gap of 3.46 mm, at 22.5°C. The flow curves were performed after a pre-shearing period of 1 minute at a shear rate of 50 s^{-1}, followed by 5 minutes at rest. The viscosity was then measured at a controlled shear rate range of [10; 250] s^{-1} on a logarithmic ramp. The values of viscosity at 50 s^{-1} and 100 s^{-1} (η_{50} and η_{100} respectively) were kept as final indicators of the puree texture linked to sensory characteristics during consumption (Chen & Engelen, 2012).

2.2.2 Biochemical analyses

The biochemical analyses of single-varietal purees have been described in **Papers I and II**. The individual sugars (fructose, glucose, sucrose) and malic acid contents of formulated puree samples were calculated based on the measured values of processed single variety purees.

2.3 Spectrum acquisition

The Vis-NIR spectral data of purees was acquired with a multi-purpose analyzer spectrometer (Bruker Optics®, Wissembourg, France) at 23°C, which provides diffuse reflectance measurements with a spectral resolution of 8 cm^{-1} from 25000 to 4000 cm^{-1} (wavelength from 400 to 2500 nm). For each spectrum, 32 scans were recorded and averaged. The spectral acquisition and instrument adjustments were controlled by OPUS software Version 5.0 (Bruker Optics®). Puree were transferred into 10 mL glass vials (5 cm height x 18 mm diameter) which were placed on the automated sample wheel of the spectrophotometer. Each puree sample was measured three times on different aliquots. A reference background measurement was automatically activated before each data set acquisition using an internal Spectralon reference.

The MIR spectra of purees was collected at 23°C using a Tensor 27 FTIR

spectrometer (Bruker Optics®, Wissembourg, France) equipped with a horizontal attenuated total reflectance (ATR) sampling accessory and a deuterated triglycine sulphate (DTGS) detector. Three replications of spectral measurement were performed on different aliquots. The purees were placed at the surface of a zinc selenide (ATR-ZnSe) crystal with six internal reflections. Spectra with 32 scans for ATR-ZnSe were collected from 4000 cm^{-1} to 650 cm^{-1} with a 4 cm^{-1} resolution and were corrected against the background spectrum of air.

The whole spectral dataset of Vis-NIR or MIR included 36 spectra (3 replicates × 3 processing weeks × 4 varieties) of single-variety purees, 324 spectra of formulated purees spectra for the modelling set (3 replicates × 3 processing weeks × 6 formulated puree groups × 6 proportions) and 162 spectra for the external prediction set (3 replicates × 3 processing weeks × 6 formulated puree groups × 3 proportions) described in 2.2.1 and **Fig. 55**.

2.4 Statistical analyses of reference data

After ensuring normal distribution with a Shapiro-Wilk test (α=0.05), the reference data of processed purees were presented as mean values and the data dispersion within our experimental dataset expressed as standard deviation values (SD). Analysis of variance (ANOVA) was carried out to determine the significant differences due to the different single apple varieties (**Table 45**) or formulated puree groups (**Table 46**) using XLSTAT (version 2018.5.52037, Addinsoft SARL, Paris, France) data analysis toolbox. And the pairwise comparison between means was performed using Tukey's test. Principal component analysis (PCA) was carried out on all reference data of single-variety purees or of formulated purees to evaluate their discriminant contributions using Matlab 7.5 (Mathworks Inc. Natick, MA, USA) software.

2.5 MCR-ALS and spectra reconstruction

MCR-ALS (multivariate curve resolution-alternative least square) is an effective multivariate self-modelling curve resolution method developed by Tauler (de Juan & Tauler, 2006). The relative contributions given by MCR-ALS were obtained for both,

the Vis-NIR (400-2500 nm) and MIR (900-1800 cm^{-1}) spectral information, using the formulated purees and their corresponding single-variety purees (**Fig. 56**). For the formulated samples, one matrix D ($n\times \lambda$) was made up with the number of samples (n) and the intensity at each wavenumbers or wavelengths (λ). The S^T matrix ($s\times \lambda$) is the spectroscopic matrix describing the 'pure' infrared spectra (λ) of all single-variety purees (s). The D matrix can be mathematically decomposed into the individual contributions related to the spectral information of 'pure' purees in matrix S^T according to Eq. (1) and is interactively transformed using an alternative least square (ALS) procedure as Eq (2).

$$D = CS^T + E \qquad (1)$$

$$C = R(S^T)^+ \qquad (2)$$

Matrix C ($n\times q$) is the concentration matrix describing the contribution of every single-variety purees (q) in reconstructed purees (n). E is the error matrix that provides the data variation not explained by their contributions. The matrix $(S^T)^+$ is the pseudo-inverse matrix of S^T. A general constraint used in curve resolution method is the non-negativity on the concentration profiles.

Once the concentration profiles (matrix C) for each single-variety spectrum, including Golden Delicious (C_{GD}), Granny Smith (C_{GS}), Braeburn (C_{BR}) and Royal Gala (C_{GA}), were obtained, they were used to reconstruct a new spectroscopic matrix R ($n \times k$) for monitoring all formulated purees. Each row R_i. (i=1,…n) was made up of a reconstructed spectrum. And each column R.j (j=1,…k) gave the reconstructed spectral intensity at a wavenumber of MIRS or a wavelength of Vis-NIRS based on the corresponding pure puree spectra of Golden Delicious (λ_{GD}), Granny Smith (λ_{GS}), Braeburn (λ_{BR}) and Royal Gala (λ_{GA}), following Eq (3).

$$R = C_{GD}\lambda_{GD} + C_{GS}\lambda_{GS} + C_{BR}\lambda_{BR} + C_{GA}\lambda_{GA} \qquad (3)$$

2.6 Spectral multivariate regression

Spectral pre-processing and multivariate regression were performed with several packages ('prospectr' (Stevens & Ramirez-Lopez, 2013), 'pls' (Mevik, Wehrens, & Liland, 2019), 'Cubist' (Kuhn, Weston, Keefer, Coulter, & Quinlan, 2014) and 'caret'

(Kuhn, 2015)) of the R software (version 2.6.2) (R Core Team, 2019). As demonstrated in previous works (Bureau et al., 2013; Ncama, Opara, Tesfay, Fawole, & Magwaza, 2017), the wavelengths from 400 to 2500 nm of Vis-NIR and the wavenumbers from 900 to 1800 cm^{-1} in MIR were selected (**Fig. 57**). For all spectral datasets, standard normal variate (SNV), resampling (intervals= 5, 10, 15), and derivative transform calculation (Savitzky–Golay method, gap size = 11, 21, 31, 41) of first or second order were compared before multivariate regression. SNV pre-processing applied on the Vis-NIR and MIR data had the best performances to predict puree quality and was then systematically used.

The partial least square (PLS), Cubist and RF regression models were developed to i) detect the proportions of each apple varieties in puree samples (**Table 47**) and predict the quality characteristics of formulated purees based on ii) the acquired Vis-NIR, MIR and their combined infrared spectra (CB) (**Table 48**) or iii) the reconstructed Vis-NIR, MIR and CB spectra (**Table 50**). All aforementioned spectral matrices (Vis-NIRS, MIRS and CB) corresponded to the same reference dataset. The set of all modelling spectra (324 spectra) was randomly split, with two-thirds of the dataset (216 spectra) used for calibration and a third (108 spectra) for internal validation. Then, calibrated models were further validated with the external prediction set (162 spectra). The procedure was repeated 10 times in order to obtain the dispersion of values giving an idea of the model stability and robustness. The developed models performance was then described by the 10-times averaged values of the determination coefficients of internal validation (R_v^2) and external prediction (R_p^2), root mean square error of prediction (RMSEP), RPD (Residual Predictive Deviation) value as described by Nicolai (Nicolai et al., 2007). During model training, the variable importance (VIP) for each puree characteristics were computed using the 'varImp' function by 'caret' package in R software (Kuhn, 2015), which could be applied both on PLS and machine learning regressions (Parmley, Higgins, Ganapathysubramanian, Sarkar, & Singh, 2019).

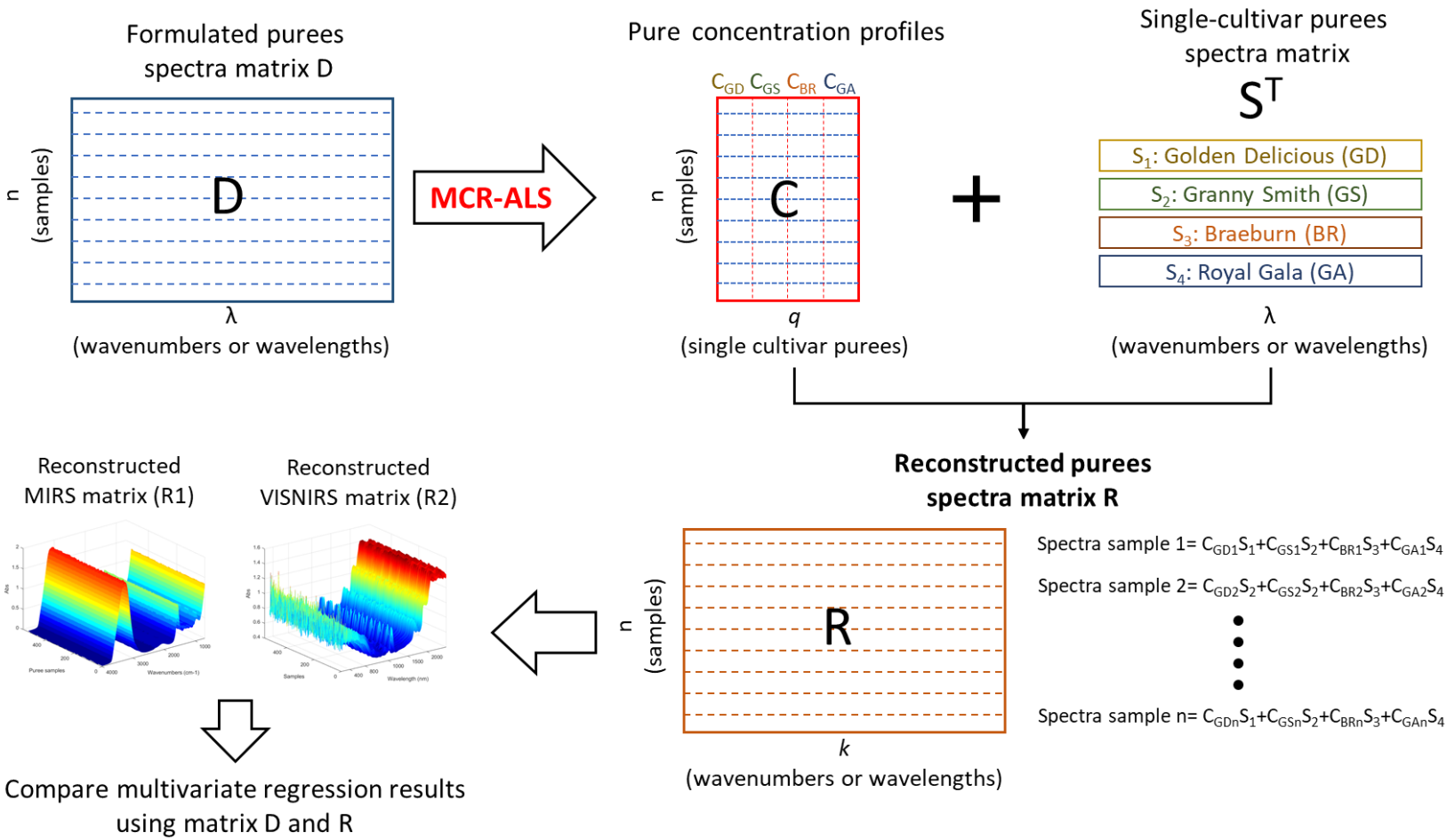

Fig. 56. Process of VIS-NIRS and MIRS data by multivariate curve resolution- alternative least square (MCR-ALS) and spectral reconstruction of reformulates puree samples.

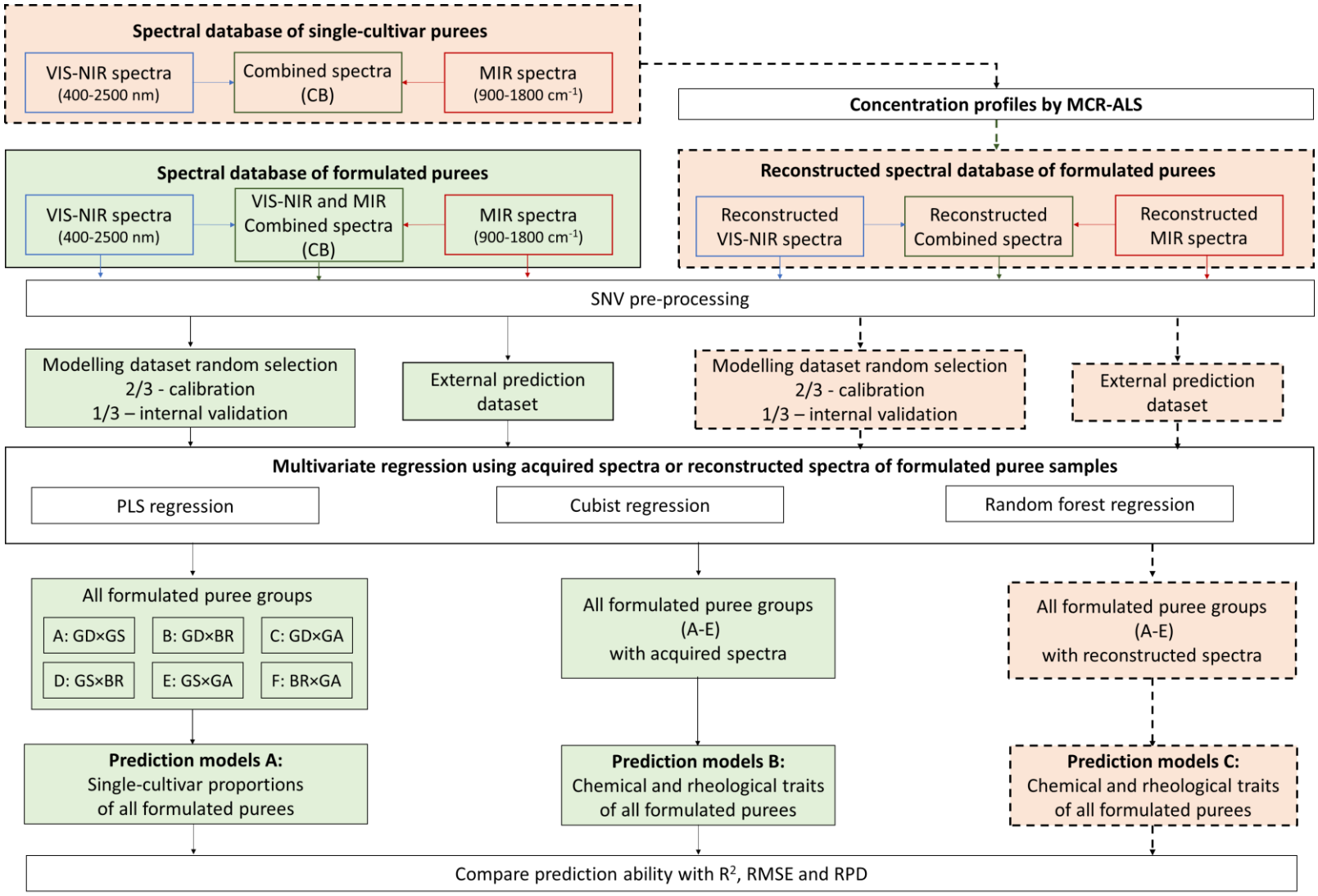

Fig. 57. Overview of the applied methodology of VIS-NIR and MIR spectra pre-processing and multivariate regression.

3. Results and discussion

3.1 Characteristics of single-variety purees and formulated purees

After puree processing, the four different varieties provided a large variability of appearance, in particular color and texture **(Fig. 58)**. According to PCA results taking into account their rheological and biochemical characteristics **(Fig. 59)**, 'Royal Gala' (GA) purees were clearly discriminated from the other purees along the first principal component (PC1), with significantly (p <0.001) lower TA, pH, glucose, malic acid and viscosity (η_{50} and η_{100}) (**Table 45**). Particularly, the values of viscosity at a shear rate of 50 s^{-1} (η_{50}), which is commonly used to describe the in-mouth texture perception of fluid foods (Chen & Engelen, 2012), were much more lower in GA purees (547 ± 13 $Pa.s^{-1}$) than in 'Golden Delicious' (GD) (839 ± 53 $Pa.s^{-1}$) and 'Granny Smith' (GS) (904 ± 31 $Pa.s^{-1}$) purees (**Table 45**). As expected, the viscosity and global quality (SSC and TA) of the formulated purees were affected when prepared with GA purees (**Fig. 60**). For example, the formulated GA-GD (group C) or GA-GS purees (group E) provided a high range of viscosity **(Fig. 60c and d)** and chemical composition (**Fig. 60e and f**), but with a limited variation of color (a* and b* values) **(Fig. 60a and b)**.

Fig. 58. Pictures of apples and processed purees.

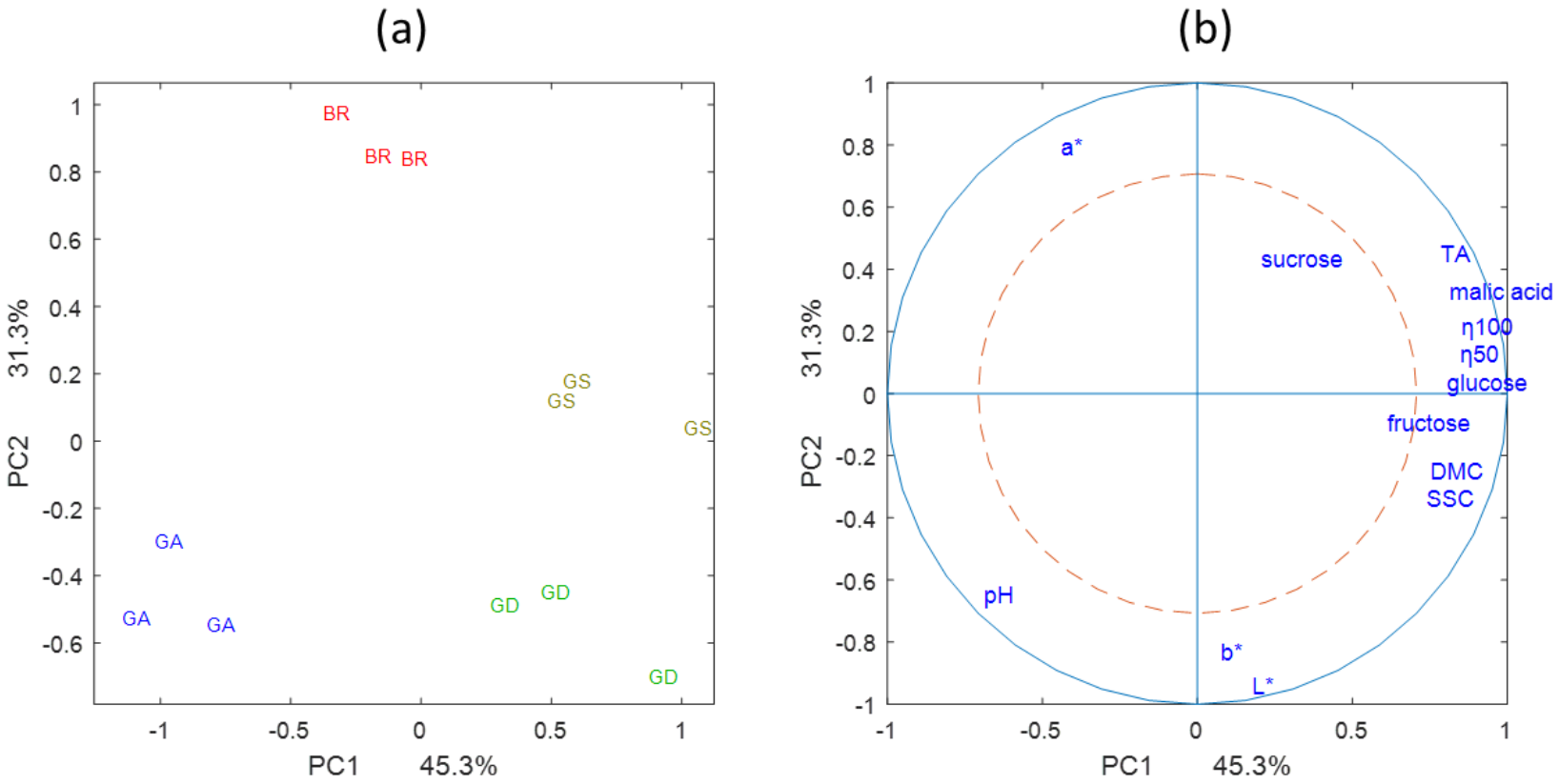

Fig. 59. PCA results of four kinds of apple purees during three processing periods.

Remarkable changes (p <0.001) of color parameters (L*, a* and b*) allowed the separation of 'Braeburn' (BR) purees and the others along the second principal component (**Fig. 59 and Table 46**). Particularly the redness (a* values) of formulated puree groups (**Fig. 60a**), the admixture of BR (groups B, D and F) introduced more intensive variations (from -4.33 to 2.35) than the others (groups A, C and E, from -4.77 to -1.52). The limited variations of yellowness (b* values) in formulated GD-GA purees resulted in differences below the visual detection threshold **(Fig. 60b)**. Consequently, different strategies of puree formulation, especially the mixtures with 'Royal Gala' or with 'Braeburn' purees, could provide variability in taste, texture and color.

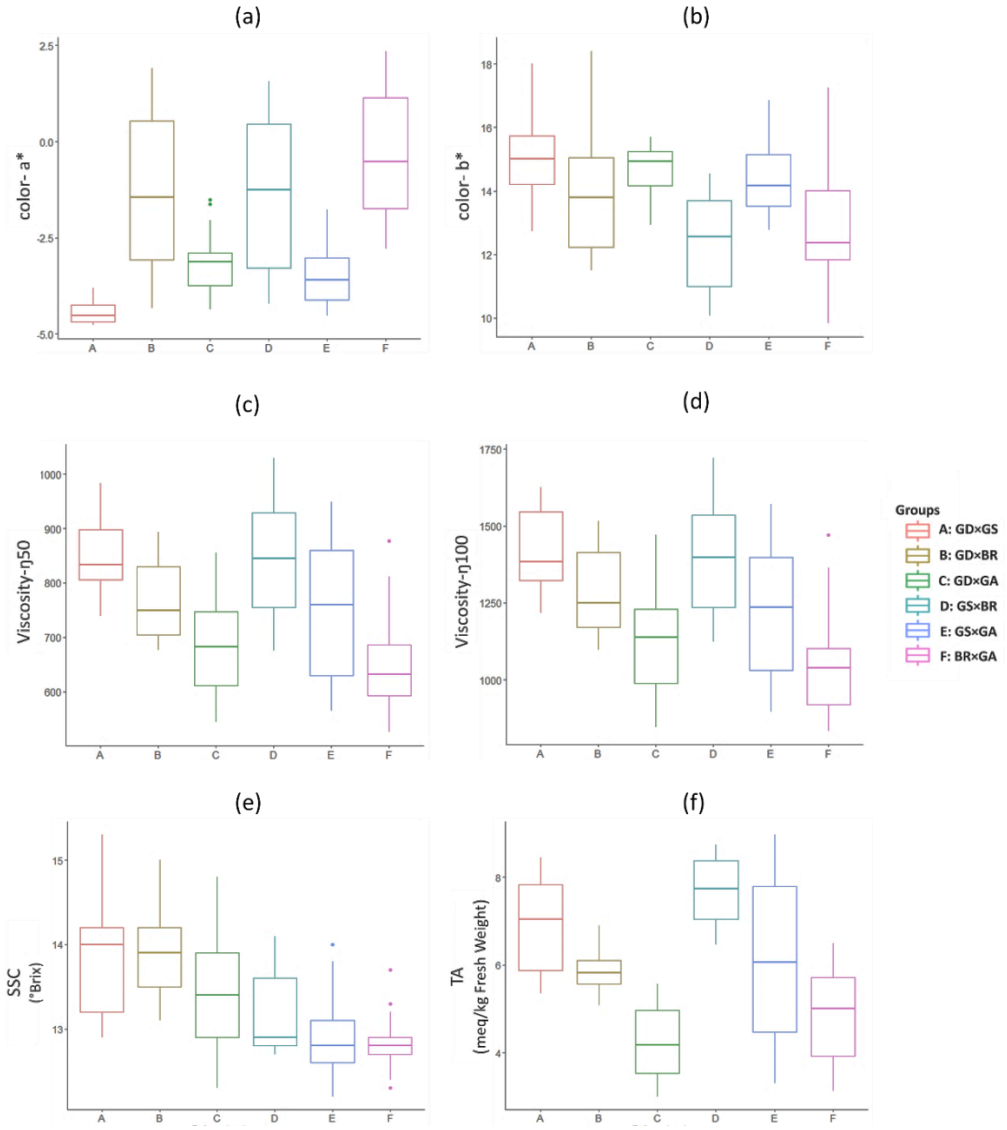

Fig. 60. Boxplot of colors (a* and b*), rheological parameters (ŋ50 and ŋ100), soluble solids (SSC) and titratable acidity (TA) of different formulated puree groups.

Table 45. Mean values with the characteristics of single-variety purees differed significantly using Tukey's test.

Variety	Viscosity η50	Viscosity η100	L*	a*	b*	SSC (°Brix)	DMC (g/g FW)	pH	TA (meq/kg FW)	malic acid (g/kg FW)	fructose (g/kg FW)	sucrose (g/kg FW)	glucose (g/kg FW)
GD	838.6± 69.6 a	1388.5± 138.2 ab	47.0± 0.3 a	-4.1± 0.3 c	15.8± 0.5 a	14.5± 0.4 a	0.167± 0.003 a	3.9± 0.1 b	53.9± 0.8 c	6.0± 0.1 b	72.8± 6.2 a	46.7± 6.6 a	18.7± 0.8 b
GS	904.2± 18.8 a	1501.2± 18.9 a	45.1± 0.8 b	-4.3± 0.9 c	14.4± 1.4 a	13.2± 0.3 b	0.152± 0.003 b	3.6± 0.2 c	89.1± 1.3 a	8.1± 0.5 a	58.6± 14.9 ab	37.4± 4.3 b	26.8± 1.3 a
BR	736.8± 61.1 b	1229.1± 106.2 b	42.2± 0.4 c	1.8± 0.3 a	10.2± 1.2 b	13.1± 0.5 b	0.151± 0.005 b	3.7± 0.2 bc	62.7± 1.5 b	5.9± 0.1 b	59.4± 6.6 ab	50.0± 3.7 a	17.2± 1.5 b
GA	547.1± 38.0 c	860.9± 59.5 c	45.6± 0.5 b	-2.5± 0.2 b	15.0± 0.8 a	12.4± 0.6 b	0.143± 0.008 b	4.3± 0.2 a	29.3± 1.3 d	3.5± 0.7 c	49.3± 4.0 b	36.0± 1.3 b	14.6± 1.3 c

Note: Data are expressed as puree fresh weight (FW) ± standard deviation. Puree varieties: Golden Delicious ('GD'); Granny Smith ('GS'), Braeburn ('BR') and Royal Gala ('GA').

Table 46. Mean values with the characteristics of formulated puree groups differed significantly using Tukey's test.

Groups	Viscosity η50	Viscosity η100	L*	a*	b*	SSC (°Brix)	DMC (g/g FW)	pH	TA (meq/kg FW)	malic acid (g/kg FW)	fructose (g/kg FW)	sucrose (g/kg FW)	glucose (g/kg FW)
A: GD×GS	854.1± 70.3 a	1412.8± 124.5 a	46.3± 1.0 a	-4.5± 0.2 d	15.0± 1.2 a	13.9± 0.7 a	0.160± 0.007 a	3.7± 0.1 e	70.0± 10.5 b	7.0± 0.8 a	66.1± 6.4 a	42.3± 3.5 b	22.6± 2.8 a
B: GD×BR	767.2 ± 67.4 b	1278.8± 128.7 b	45.2± 1.6 b	-1.3± 1.9 b	13.9± 1.9 b	13.9± 0.5 a	0.160± 0.006 a	3.7± 0.1 d	58.7± 4.3 c	5.9± 0.1 b	66.4± 5.5 a	48.2± 3.1 a	18.0± 0.9 c
C: GD×GA	684.7± 90.5 c	1127.0± 172.3 cd	46.1± 0.9 a	-3.2± 0.7 c	14.7± 0.7 a	13.4± 0.7 b	0.156± 0.008 b	4.1± 0.2 a	42.7± 7.9 e	4.8± 0.8 c	61.6± 11.7 b	41.6± 4.7 b	16.8± 1.6 cd
D: GS×BR	853.8± 99.2 a	1408.5± 173.0 a	43.6± 1.0 c	-1.4± 2.0 b	12.4± 1.4 c	13.2± 0.5 b	0.154± 0.007 d	3.6± 0.1 e	76.6± 7.5 a	7.2± 0.8 a	59.0± 2.9 b	43.4± 5.1 b	22.3± 3.1 ab
E: GS×GA	743.2± 115.2 b	1209.1± 197.0 bc	45.1± 0.5 b	-3.5± 0.8 c	14.4± 1.1 ab	12.9± 0.5 c	0.150± 0.007 c	3.8± 0.2 c	60.5± 17.8 c	5.9± 1.5 b	54.2± 8.7 c	36.7± 2.0 c	21.0± 4.0 b
F: BR×GA	651.7± 86.2 c	1061.4± 161.1 d	44.1± 1.2 c	-0.3± 1.5 a	12.8± 1.7 c	12.8± 0.3 c	0.148± 0.003 c	4.0± 0.2 b	48.7± 10.2 d	4.7± 0.8 c	54.6± 7.2 c	43.3± 5.6 b	16.0± 1.1 d

Note: Data are expressed as puree fresh weight (FW) ± standard deviation. Puree varieties: Golden Delicious ('GD'); Granny Smith ('GS'), Braeburn ('BR') and Royal Gala ('GA').

3.2 Characteristics of formulated purees: determination of composed single-variety puree proportions

In this part, the ability of SNV pre-processed Vis-NIR, MIR and CB coupled with PLS, Cubist and RF regressions was compared to estimate the proportions of single-variety in all formulated purees (**Table 47**).

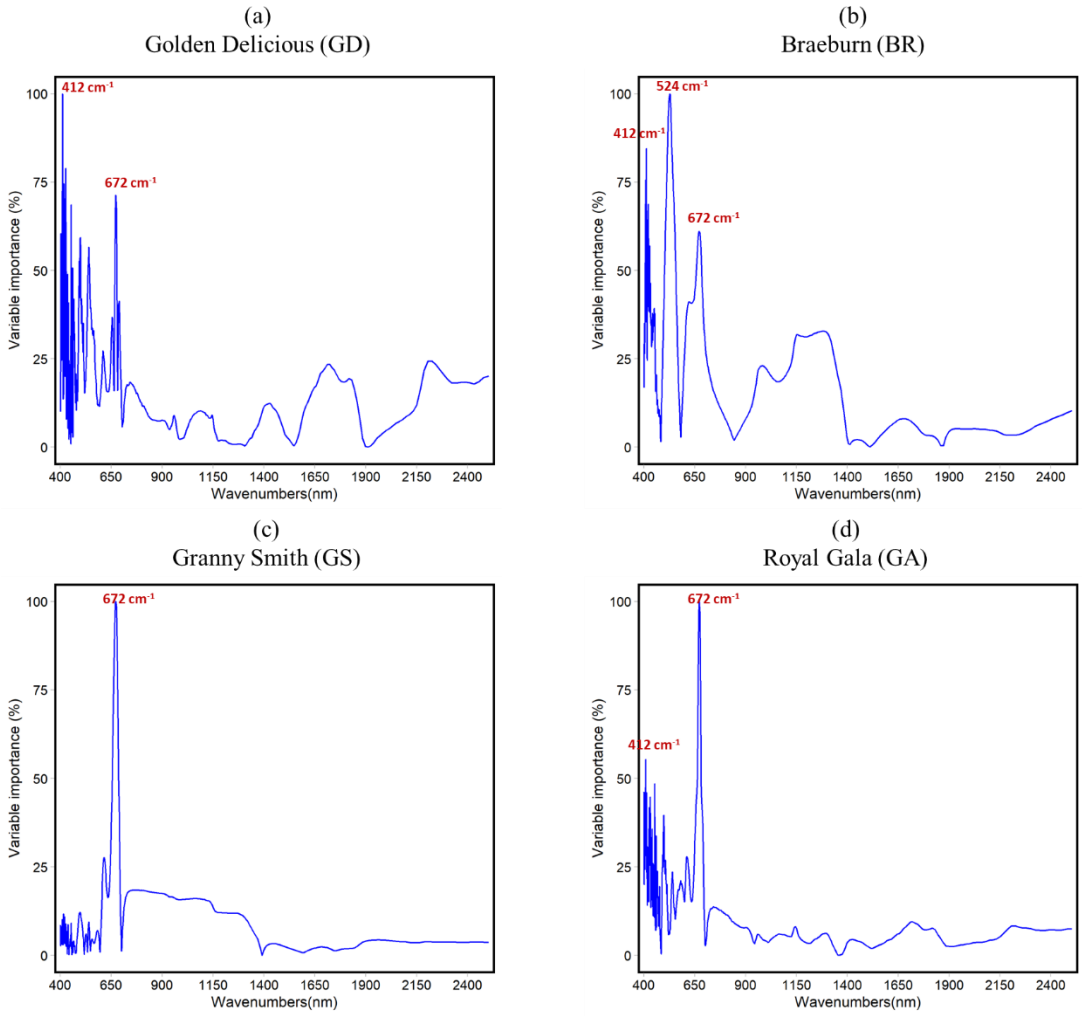

Fig. 61. PLS variable importance (VIP) of VIS-NIRS (400-2500 nm) prediction models for formulated puree proportions of each apple variety.

Both, Vis-NIR and MIR techniques were potentially able to estimate the proportions of single-variety puree in the formulated purees with good models presenting robust determination coefficients for both internal validation (R_v^2) and external validation (R_v^2), acceptable RMSEP (<10%) and RPD values at least higher than 2.5 (Nicolai et al., 2007). For Vis-NIR technique, two regression methods, PLS

and RF, showed an acceptable ability to estimate proportions of GS ($R_P^2 > 0.92$, RPD > 3.4, RMSEP < 9.2%) and of BR ($R_P^2 > 0.95$, RPD > 4.2, RMSEP < 7.9%) varieties in all formulated purees, based on the VIP wavelengths at 412 nm, 524 nm and 672 nm (**Fig. 61b and c**). The predictive errors obtained here for the mixture of two varieties of the same species, apple, were lower than those obtained earlier for the mixture of two species, namely apple/raspberry (11.3%) (Contal, León, & Downey, 2002). The poor Vis-NIRS prediction results for GD (RMSEP > 17.4%, RPD < 1.7) and GA (RMSEP > 16.2%, RPD < 2.1) were probably due to their similar color (**Fig. 58**). As the VIP wavelengths of Vis-NIR models were mainly dominated in the visible spectral region (412-672 nm), the color variations were not enough to be used for prediction of proportions in formulated purees, especially in the group C (GD-GA) (**Fig. 60a & b**).

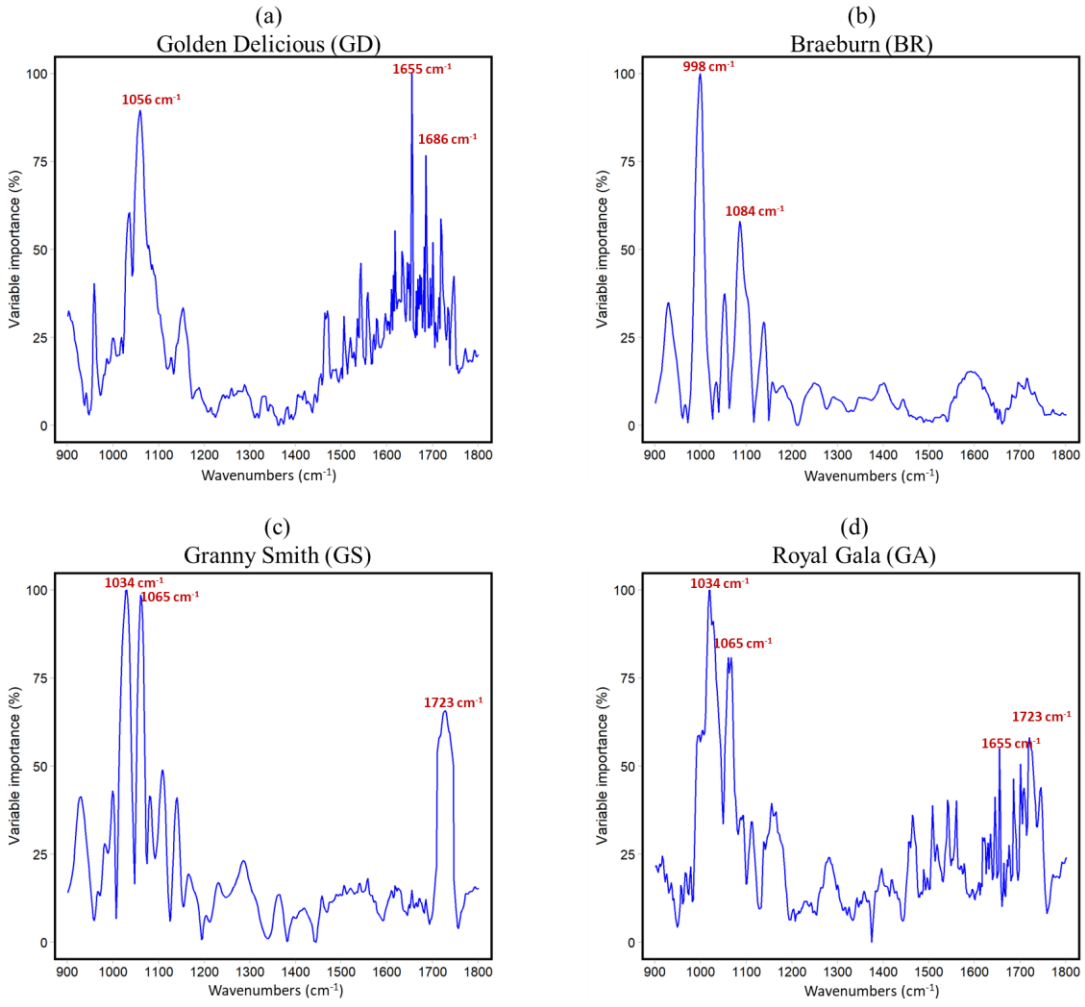

Fig. 62. PLS variable importance (VIP) of MIRS (900-1800 cm^{-1}) prediction models for formulated puree proportions of each apple variety.

MIR provided a better prediction of the proportions of single-variety purees in the formulated purees than Vis-NIR. Moreover, the regression method affected the prediction results of MIR. PLS gave better prediction results (RMSEP<8.1%, RPD> 3.6) than Cubist (RMSEP<15.1%, RPD> 2.3) and RF (RMSEP<10.6%, RPD> 2.7). Particularly, MIRS combined with PLS reached the lowest determination error (RMSEP=2.7%, RPD=11.4) for GS compared with other varieties (GD, GA, BR). The highest VIP values (**Fig. 62c**) at 1723 cm-1, 1065 cm^{-1} and 1034 cm^{-1} attributed respectively to malic acid, fructose and glucose (Bureau, Cozzolino, & Clark, 2019; Clark, 2016), were consistent with the existence of marked differences in chemical composition (SD and significance) between purees containing GS (**Table 46**). The excellent PLS predictions obtained for BR (RMSEP=4.3%, RPD=7.7) were based on the VIP wavenumbers at 998 cm^{-1} and 1084 cm^{-1} related to sucrose and fructose (Bureau, Cozzolino, & Clark, 2019) (**Fig. 62b**). Besides the aforementioned spectral signal, the satisfactory assessments of GD and GA proportions (RMSEP<8.1%, RPD>3.6) were linked to the MIRS region between 1750 and 1650 cm^{-1} related to organic acids, pectins, proteins, phenolics and absorbed water (**Fig. 62a & d**) (Abidi, Cabrales, & Haigler, 2014; Canteri, Renard, Le Bourvellec, & Bureau, 2019; Kačuráková et al., 1999).

The CB spectra, including Vis-NIR and MIR regions, coupled with PLS (RPD>2.8, RMSEP<11.5%) and RF (RPD>3.0, RMSEP<9.5%) provided a satisfactory assessment of the proportions of single-variety purees (**Table 47**). However, the results on CB were not as good as for MIR only.

Consequently, to predict proportions of single-variety purees, Vis-NIR was suitable for the formulated samples presenting large diversity in the color range, with the use of Braeburn and Granny Smith apples for example, and under vacuum processing conditions providing a good puree color preservation. MIRS coupled with PLS was evidenced as a powerful tool to provide excellent estimations of puree proportions, mainly based on differing concentrations of individual sugars and acid. Combining Vis-NIR and MIR did not improve prediction.

Table 47. Prediction of the proportions (%) of single-variety purees in all formulated purees based on VIS-NIR (400- 2500 nm), MIR (900- 1800 cm^{-1}) and their combined spectra (CB; VIS-NIR-MIR). Comparison of three regression models (PLS, Cubist and Random forest).

Single-variety	Spectra	PLSR				Cubist				Random forest			
		R_v^2	R_p^2	RMSEP	RPD	R_v^2	R_p^2	RMSEP	RPD	R_v^2	R_p^2	RMSEP	RPD
GD	Vis-NIR	0.66	0.60	19.0	1.5	0.82	0.58	19.4	1.5	0.88	0.64	17.4	1.7
	MIR	0.94	0.92	8.1	3.6	0.95	0.86	11.3	2.6	0.94	0.87	10.6	2.7
	CB	0.91	0.88	10.3	2.8	0.93	0.82	12.7	2.3	0.96	0.90	9.5	3.0
BR	Vis-NIR	0.97	0.95	7.5	4.4	0.98	0.93	8.7	3.8	0.97	0.95	7.9	4.2
	MIR	0.99	0.98	4.3	7.7	0.98	0.97	5.2	6.3	0.98	0.95	7.6	4.3
	CB	0.99	0.98	5.0	6.6	1.00	0.97	5.6	5.9	0.98	0.97	6.1	5.4
GS	Vis-NIR	0.93	0.92	9.2	3.4	0.97	0.89	10.5	3.0	0.97	0.94	8.2	3.8
	MIR	0.99	0.99	2.7	11.4	0.99	0.93	8.1	3.8	0.98	0.97	5.3	5.8
	CB	0.99	0.98	4.3	7.3	0.99	0.98	4.9	6.4	0.98	0.97	5.8	5.4
GA	Vis-NIR	0.79	0.65	16.2	2.1	0.67	0.68	20.2	1.7	0.75	0.73	18.5	1.9
	MIR	0.96	0.94	7.4	4.7	0.90	0.82	15.1	2.3	0.91	0.90	10.3	3.4
	CB	0.89	0.83	11.5	3.0	0.88	0.79	16.2	2.2	0.94	0.92	9.4	3.7

Notes: single-variety purees of Golden Delicious named 'GD', Braeburn named 'BR', Granny Smith named 'GS', Royal Gala named 'GA'. All results corresponded to the averaged values of 10 replicates. R_v^2: determination coefficient of the validation test (internal); R_p^2: determination coefficient of the prediction test (external); RMSEP: root mean square error of prediction test (external) expressed as the puree proportions (%); RPD: the residual predictive deviation of prediction test. (external).

3.3 Characteristics of formulated purees: prediction of quality traits

As previously, the different spectral areas, Vis-NIR, MIR or CB, of all formulated purees coupled with the different regression methods, PLS, Cubist and RF, were compared for their ability to predict color, rheological and biochemical characteristics of formulated purees (**Table 48**). MIR spectra coupled with PLS obtained the best predictions in comparison with Vis-NIR and CB, except for color. Indeed concerning the color parameters, a good prediction of a* values was obtained for all spectral areas with a RPD decreasing order Vis-NIR (RPD>4.0), CB (RPD>3.6) and MIR (RPD>3.3) for both PLS and machine learning regressions (Cubist and Random forest). Particularly, the best prediction of a* values was obtained on CB with PLS models (R_P^2=0.96, RPD=5.0), slightly better than in Vis-NIR (R_P^2=0.95, RPD=4.7).

MIR spectra coupled with PLS gave the best prediction (R_P^2>0.90, RPD>4.1) of the rheological parameters (η_{50} and η_{100}) (**Table 48**). The identified VIP wavenumbers were 1026, 1065, 1113 and 1720 cm^{-1} (**Fig. 58**). These dominant carbohydrate bands centered at 1000-1200 cm^{-1}, associated with C-OH and C-O-C vibration of glucose and fructose (Bureau, Cozzolino, & Clark, 2019), have also been identified to predict viscosity of tomato purees (Ayvaz et al., 2016). And an acceptable estimation of DMC was observed for all developed MIR models (RMSEP< 0.003, RPD>2.7).

For biochemical parameters, MIR coupled with PLS allowed a very good prediction of SSC (RMSEP=0.1, RPD=5.1) in accordance with previous results of apple and tomato purees (Ayvaz et al., 2016; **Papers I, II**). In apples, SSC is strongly correlated to the presence of sugars, namely fructose, sucrose and glucose. The two main sugars, fructose and sucrose, were satisfactorily predicted with PLS (RPD>3.0) and the non-linear regressions, Cubist and RF (RPD>2.9). However, MIR could not predict the glucose content (RPD<2.4) (**Table 48**).

Considering the different expressions of acidity such as pH, TA and malic acid content, MIR coupled with PLS provided their excellent prediction with R_P^2>0.92 and RPD>4.0. It can be noticed that Vis-NIRS gave also acceptable prediction of TA and

malic acid (R_P^2>0.87, RPD>2.9), better than our previous results in NIRS on apple purees (Lan, Jaillais, Leca, Renard, & Bureau, 2020).

In comparison with Vis-NIRS and MIRS, the slight improvements of using the combined spectra (CB) concerned only the prediction of a* values (**Table 48**). Combining Vis-NIRS and MIRS spectra offered little improvement or even degraded the results in comparison with MIRS alone for analyzing puree viscosity and chemical variations (**Table 48**). These conclusions were in accordance with previous works on forage (Reeves, 1997) and beers (Iñón, Garrigues, & Guardia, 2006). They can be explained by i) the limited ability to balance the important variables after combination of two spectral domains with different resolutions (**Fig. 63**); and ii) the involvement of non-relevant or unimportant spectral regions which disturbed the calibration modelling by producing more noise.

In summary, MIRS coupled with PLS had promising ability to well estimate viscosity, a* color parameter, DMC, SSC, pH, TA, malic acid, sucrose and glucose of formulated purees, but not for fructose. Acceptable assessments of a*, TA, malic acid and glucose were obtained with the Vis-NIR region, in which sensors could be easily adapted for fruit processing.

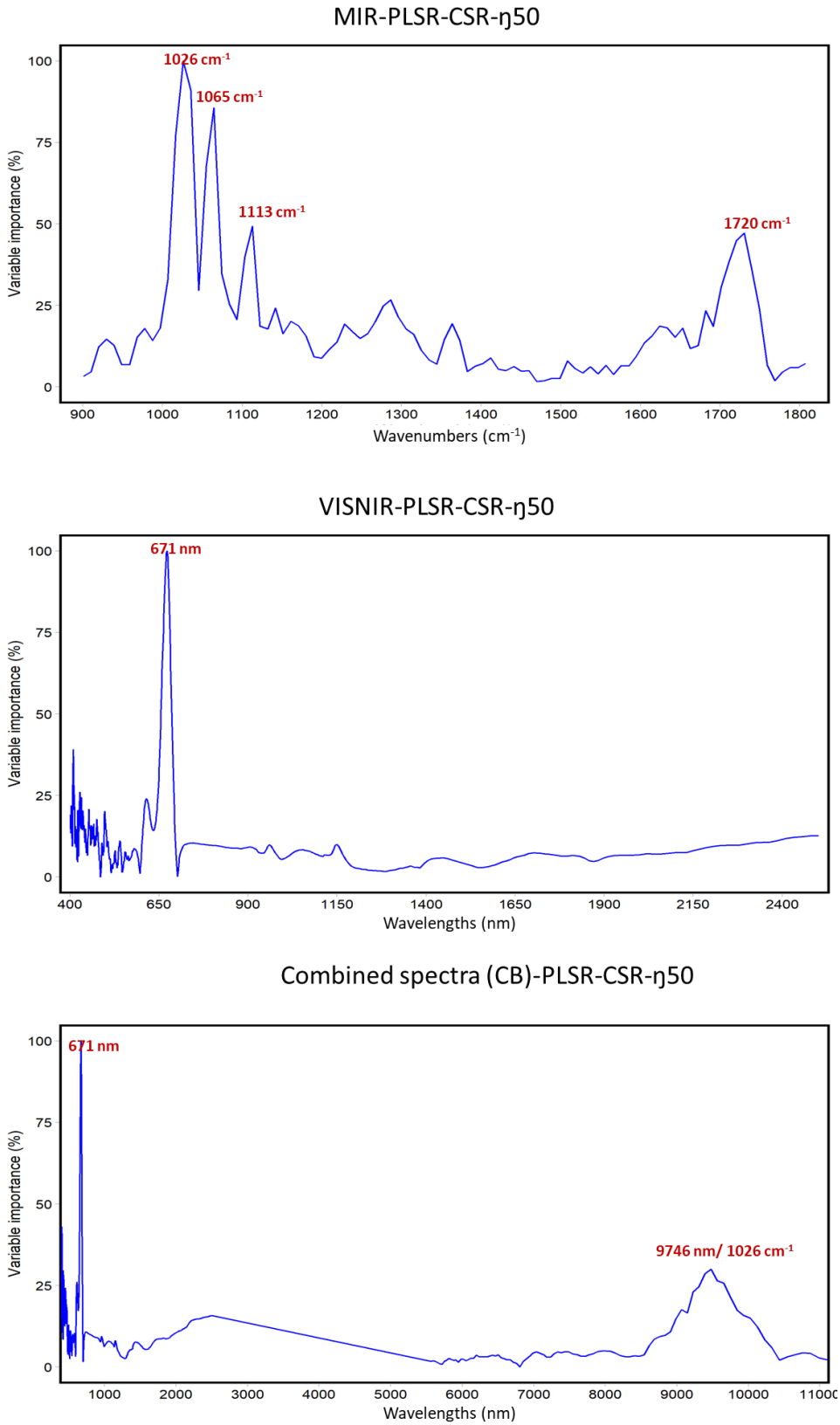

Fig. 63. The variable importance (VIP) of the VIS-NIRS, MIRS and CB of PLS prediction models for puree viscosity at 50 s^{-1} in control shear rate test (CSR).

Table 48. Prediction of chemical and rheological parameters of all formulated purees using Vis-NIR (400-2500 nm), MIR (900-1800 cm^{-1}) or their combined spectra (CB) and regression methods, PLS, Cubist or Random forest.

Parameter	Spectra	Range	SD	PLSR				Cubist				Random forest			
				R_v^2	R_p^2	RMSEP	RPD	R_v^2	R_p^2	RMSEP	RPD	R_v^2	R_p^2	RMSEP	RPD
L*	Vis-NIR	41.6-48.6	1.5	0.81	0.70	0.8	1.6	0.87	0.63	0.9	1.4	0.88	0.75	0.6	1.9
	MIR			0.88	0.80	0.6	2.0	0.96	0.83	0.6	2.2	0.94	0.80	0.6	2.2
	CB			0.89	0.79	0.6	2.1	0.95	0.79	0.6	1.9	0.94	0.83	0.5	2.4
a*	Vis-NIR	(-4.8)-2.4	2	0.97	0.96	0.4	4.7	0.98	0.94	0.5	4.0	0.96	0.94	0.5	4.1
	MIR			0.96	0.94	0.5	4.0	0.98	0.92	0.5	3.6	0.97	0.91	0.6	3.3
	CB			0.98	0.96	0.4	5.0	0.99	0.93	0.5	3.6	0.98	0.94	0.5	4.1
b*	Vis-NIR	9.6-18.4	1.7	0.62	0.55	1.2	1.5	0.76	0.46	1.5	1.3	0.72	0.53	1.3	1.4
	MIR			0.67	0.56	1.2	1.5	0.86	0.48	1.4	1.3	0.84	0.62	1.1	1.6
	CB			0.67	0.53	1.3	1.5	0.88	0.46	1.4	1.3	0.81	0.57	1.2	1.5
Viscosity η_{50}	Vis-NIR	526-1029	119	0.79	0.81	54.6	2.2	0.85	0.85	49.8	2.4	0.82	0.78	57.8	2.1
	MIR			0.94	0.90	29.8	4.1	0.95	0.89	39.4	3.1	0.9	0.87	43.6	2.8
	CB			0.91	0.87	43.5	2.8	0.93	0.88	43.2	2.8	0.91	0.89	42.8	2.8
Viscosity η_{100}	Vis-NIR	834-1721	210	0.73	0.74	108.0	2.0	0.87	0.79	98.9	2.2	0.82	0.75	109.3	1.9
	MIR			0.94	0.91	52.0	4.1	0.96	0.86	81.2	2.6	0.90	0.88	74.4	2.9
	CB			0.88	0.87	79.6	2.7	0.91	0.87	76.5	2.8	0.91	0.88	77.3	2.8
DMC (g/g FW)	Vis-NIR	0.14-0.17	0.009	0.85	0.79	0.004	2.1	0.81	0.75	0.004	1.9	0.79	0.77	0.004	2.0
	MIR			0.93	0.89	0.003	3.1	0.91	0.88	0.003	2.7	0.93	0.90	0.003	3.0
	CB			0.85	0.83	0.003	2.5	0.96	0.83	0.003	2.5	0.93	0.87	0.003	2.8
SSC (°Brix)	Vis-NIR	12.1-15.3	0.7	0.61	0.53	0.5	1.5	0.79	0.56	0.5	1.3	0.78	0.62	0.5	1.5
	MIR			0.96	0.95	0.1	5.1	0.96	0.93	0.2	3.9	0.94	0.94	0.2	4.1
	CB			0.89	0.94	0.2	4.0	0.95	0.92	0.2	3.4	0.95	0.96	0.1	4.4
fructose (g/kg FW)	Vis-NIR	40.2-80.3	9.1	0.37	0.38	7.3	1.2	0.52	0.25	8.4	1.1	0.70	0.50	6.3	1.4
	MIR			0.82	0.78	3.7	2.4	0.93	0.81	4.0	2.2	0.92	0.70	4.8	1.8
	CB			0.67	0.56	5.8	1.5	0.83	0.74	4.4	2.0	0.91	0.76	4.4	2.0
sucrose (g/kg FW)	Vis-NIR			0.54	0.49	3.9	1.4	0.69	0.52	4.3	1.3	0.76	0.46	4.0	1.4

	MIR	33.2-57.3	5.5	0.89	0.89	1.8	3.0	0.88	0.89	1.8	2.9	0.92	0.92	1.6	3.3
	CB			0.60	0.67	3.1	1.7	0.92	0.87	2.0	2.6	0.87	0.78	2.5	2.1
	Vis-NIR			0.92	0.93	1.0	3.6	0.96	0.87	1.3	2.6	0.91	0.89	1.2	2.9
glucose (g/kg FW)	MIR	13.2-28.3	3.7	0.98	0.98	0.5	6.7	0.99	0.97	0.6	5.6	0.97	0.94	0.9	4.1
	CB			0.95	0.93	1.0	3.7	0.98	0.96	0.7	4.9	0.98	0.95	0.8	4.4
	Vis-NIR			0.84	0.83	0.1	2.4	0.94	0.76	0.1	2.0	0.82	0.76	0.1	2.0
pH	MIR	3.39-4.47	0.23	0.94	0.92	0.1	4.0	0.89	0.85	0.1	2.5	0.94	0.92	0.1	3.4
	CB			0.83	0.86	0.1	2.7	0.96	0.67	0.1	1.7	0.9	0.85	0.1	2.4
	Vis-NIR			0.93	0.87	5.0	2.9	0.95	0.90	5.1	3.1	0.96	0.89	5.2	3.0
TA (meq/kg FW)	MIR	28.0-94.8	16.2	0.99	0.96	3.5	4.3	0.99	0.94	3.9	3.9	0.96	0.91	4.7	3.3
	CB			0.95	0.91	4.9	3.1	0.98	0.95	3.8	4.0	0.96	0.9	4.8	3.2
	Vis-NIR			0.90	0.88	0.5	2.9	0.91	0.85	0.5	2.6	0.94	0.87	0.5	2.8
malic acid (g/kg FW)	MIR	3.0-8.8	1.3	0.97	0.97	0.2	5.9	0.95	0.92	0.4	3.7	0.94	0.94	0.3	4.2
	CB			0.92	0.92	0.4	3.4	0.91	0.84	0.5	2.5	0.96	0.93	0.4	3.7

Notes: all results corresponded to the averaged values of 10 replicates. R_v^2: determination coefficient of the validation test (internal); R_p^2: determination coefficient of the prediction test (external); RMSEP: root mean square error of prediction test (external); RPD: the residual predictive deviation of prediction test. (external).

3.4 Characteristics of formulated purees: prediction of quality traits based on the reconstructed spectra

In order to compute the concentration profiles of relevant single-variety puree compositions, MCR-ALS was applied on the Vis-NIR and MIR spectra of all formulated purees and of the four single-variety purees, using two approaches: the 54 averaged formulated puree spectra and the 4 averaged single-variety puree spectra of a) each week or b) over the three weeks. These two methods (a and b) obtained similar concentrations, indicating their robustness over different processing weeks. Results are only shown for method b) taking into account different processing periods (**Table 50).** Based on that, in total 486 spectra of formulated purees were reconstructed based on their corresponding 36 single-variety spectra (4 varieties x 3 replications x 3 weeks).

Accurate predictions of the concentrations were obtained with MIRS. These predictions were highly related to the proportions of the single-variety purees (**Table 49**). However, the results were not acceptable with Vis-NIRS (**Table 49**). The limited ability of Vis-NIRS was due to the high similarity in color between GA and GD and so a poor prediction of the proportions GA/GD in formulated purees (**Table 48**). The concentration profiles of MIRS in each group (A-E) appeared to follow a non-linear relationship along the variation of puree proportions.

Prediction models were then developed using these reconstructed MIR spectra and the reference data characterized on the formulated purees (**Table 50**). Overall, reconstructed MIR spectra with PLS regression better predicted the puree characteristics than Cubist and RF regressions. What stands out in these results was the highly accurate PLS predictions (R_p^2>0.85, RPD>4.0) of rheological parameters (η_{50} and η_{100}) from reconstructed spectra (**Table 50**), which were close to those obtained from the real spectra of formulated purees (R_p^2>0.90, RPD>4.1) (**Table 48**). Particularly, similar MIRS fingerprint wavenumbers were obtained in reconstructed spectra and directly on formulated purees described above, mainly 1720, 1113, 1065 and 1026 cm^{-1} related to acid and sugars (Bureau, Cozzolino, & Clark, 2019). The prediction of DMC was acceptable (RPD>2.5) as mentioned above with real spectra in **Table 48**. For color,

a good prediction of a* value was obtained with both, PLS (R_p^2=0.92, PRD=3.5) and machine learning methods (R_p^2>0.89, PRD>3.2) but not for L* and b*. For SSC, although the slight lower R_p^2 and RPD values than the best results obtained directly on MIR spectra (RMSEP=0.13, RPD=5.1) (**Table 48**), the PLS and Cubist models had an acceptable ability to estimate it for all formulated purees (RMESP<0.20, RPD>4.1) (**Table 50**). Considering the global acidity parameters, acceptable PLS predictions (R_p^2>0.88, PRD>3.2) were obtained for pH and TA, with a lower performance than directly on real spectra (R_p^2>0.92, PRD>4.0 in **Table 48**). For individual sugars and acids, PLS models showed an excellent prediction of glucose and malic acids (R_p^2>0.94, RPD>4.3), and an acceptable prediction of sucrose (R_p^2=0.86, RPD=2.8) but not for fructose (RPD<2.5). The specific wavenumbers at 1034 cm^{-1} for glucose, 1723 cm^{-1} for malic acid and 998 cm^{-1} for sucrose, mainly contributed to the PLS models both from reconstructed spectra and directly on puree spectra. The decrease of prediction accuracy was possibly owing to the non-negativity of the concentration profiles which could constrain the spectral reconstruction (Le Dréau, Dupuy, Artaud, Ollivier, & Kister, 2009). Briefly, MIR spectra coupled with the concentration profiles of MCR-ALS showed a potential way to directly estimate the viscosity, a* color parameter, SSC, TA, malic acid, pH, fructose and glucose for formulated purees depending only on the spectral information of the single-variety purees.

Compared to the previous prediction models obtained on the real spectra of formulated purees (**Table 48**), highly consistent specific fingerprints and acceptable prediction results (**Table 50)** provided a justifiable explanation to use the spectra reconstruction of formulated purees from spectra of single-variety purees. MCR-ALS has been used in other ways to identify precisely the chemical species or track their evolutions (Garrido, Rius, & Larrechi, 2008; de Juan & Tauler, 2006). Here, it was firstly used with the concentration profiles to reconstruct spectra of processed products based on the spectra of raw materials.

Table 49. The VIS-NIR (400-2500 nm) and MIR (900-1800 cm^{-1}) spectral concentration profiles of each apple variety in formulated puree obtained from MCR-ALS.

Groups	Proportions	MIRS (900-1800 cm^{-1})				VIS-NIRS (400-2500 nm)			
		BR	GA	GD	GS	BR	GA	GD	GS
A: GD×GS	95%:5%	0.012	0.000	0.925	0.057	0.000	0.162	0.635	0.193
	90%:10%	0.010	0.000	0.911	0.073	0.000	0.275	0.381	0.333
	75%:25%	0.015	0.049	0.790	0.143	0.000	0.000	0.833	0.169
	50%:50%	0.015	0.000	0.566	0.415	0.000	0.000	0.595	0.406
	25%:75%	0.018	0.000	0.328	0.649	0.000	0.000	0.151	0.857
	10%:90%	0.075	0.000	0.100	0.809	0.000	0.000	0.254	0.731
	80%:20%	0.008	0.000	0.772	0.211	0.012	0.231	0.358	0.398
	33%:66%	0.079	0.003	0.337	0.571	0.033	0.061	0.043	0.868
	14%:86%	0.090	0.120	0.179	0.603	0.005	0.010	0.000	0.979
B: GD:BR	95%:5%	0.050	0.004	0.942	0.004	0.034	0.243	0.564	0.148
	90%:10%	0.081	0.010	0.897	0.000	0.000	0.250	0.621	0.133
	75%:25%	0.235	0.017	0.738	0.000	0.146	0.017	0.812	0.034
	50%:50%	0.495	0.004	0.492	0.000	0.252	0.000	0.764	0.000
	25%:75%	0.711	0.003	0.279	0.000	0.657	0.159	0.073	0.117
	10%:90%	0.847	0.013	0.135	0.000	0.866	0.079	0.000	0.058
	80%:20%	0.141	0.024	0.826	0.000	0.144	0.000	0.865	0.000
	33%:66%	0.515	0.000	0.425	0.058	0.599	0.164	0.103	0.133
	14%:86%	0.827	0.000	0.171	0.002	0.673	0.000	0.328	0.000
C: GD:GA	95%:5%	0.000	0.062	0.933	0.000	0.000	0.078	0.815	0.101
	90%:10%	0.048	0.047	0.897	0.005	0.000	0.000	0.995	0.000
	75%:25%	0.028	0.239	0.728	0.000	0.000	0.407	0.533	0.061
	50%:50%	0.006	0.474	0.500	0.016	0.019	0.584	0.272	0.113
	25%:75%	0.000	0.732	0.263	0.000	0.000	0.601	0.416	0.000
	10%:90%	0.000	0.907	0.087	0.000	0.000	0.521	0.483	0.000
	80%:20%	0.019	0.165	0.807	0.005	0.000	0.528	0.281	0.183
	33%:66%	0.003	0.642	0.352	0.001	0.000	0.367	0.632	0.000
	14%:86%	0.000	0.836	0.142	0.021	0.000	0.549	0.461	0.000
D: GS:BR	95%:5%	0.220	0.090	0.002	0.677	0.000	0.000	0.068	0.927
	90%:10%	0.283	0.259	0.000	0.444	0.001	0.000	0.000	0.994
	75%:25%	0.403	0.163	0.013	0.410	0.205	0.000	0.011	0.768
	50%:50%	0.636	0.138	0.000	0.217	0.417	0.000	0.092	0.492
	25%:75%	0.782	0.067	0.037	0.106	0.631	0.000	0.064	0.312
	10%:90%	0.951	0.005	0.000	0.040	0.853	0.000	0.000	0.144
	80%:20%	0.341	0.162	0.000	0.488	0.143	0.000	0.031	0.822
	33%:66%	0.738	0.082	0.000	0.171	0.543	0.000	0.201	0.249
	14%:86%	0.846	0.000	0.000	0.144	0.668	0.000	0.352	0.000
E: GS:GA	95%:5%	0.000	0.265	0.000	0.725	0.000	0.000	0.142	0.849
	90%:10%	0.006	0.219	0.000	0.787	0.000	0.000	0.178	0.822
	75%:25%	0.030	0.419	0.000	0.537	0.091	0.071	0.058	0.773
	50%:50%	0.009	0.655	0.000	0.334	0.093	0.000	0.560	0.340
	25%:75%	0.003	0.820	0.000	0.169	0.024	0.436	0.348	0.195
	10%:90%	0.047	0.702	0.238	0.000	0.000	0.068	0.943	0.000

	80%:20%	0.050	0.365	0.000	0.574	0.000	0.000	0.286	0.721
	33%:66%	0.000	0.799	0.000	0.193	0.000	0.453	0.114	0.431
	14%:86%	0.000	0.991	0.000	0.000	0.000	0.428	0.530	0.046
	95%:5%	0.785	0.209	0.000	0.000	0.706	0.000	0.314	0.000
	90%:10%	0.849	0.149	0.000	0.000	0.737	0.131	0.076	0.058
	75%:25%	0.599	0.398	0.000	0.000	0.513	0.000	0.496	0.000
	50%:50%	0.406	0.593	0.000	0.000	0.292	0.620	0.000	0.097
F: BR:GA	25%:75%	0.138	0.857	0.000	0.000	0.000	0.541	0.471	0.000
	10%:90%	0.051	0.946	0.000	0.000	0.000	0.533	0.475	0.000
	80%:20%	0.732	0.263	0.000	0.000	0.649	0.194	0.000	0.162
	33%:66%	0.254	0.745	0.000	0.000	0.176	0.644	0.033	0.136
	14%:86%	0.067	0.932	0.000	0.000	0.804	0.000	0.183	0.014

Puree varieties: Golden Delicious (‘GD’); Granny Smith (‘GS’), Braeburn (‘BR’) and Royal Gala (‘GA’).

Table 50. Prediction results of chemical and rheological parameters of all formulated purees from the reconstructed MIR spectra computed by the concentrations of MCR-ALS and the spectra of single-variety purees.

Parameter	Range	SD	PLSR				Cubist				Random forest			
			R_v^2	R_p^2	RMSEP	RPD	R_v^2	R_p^2	RMSEP	RPD	R_v^2	R_p^2	RMSEP	RPD
L*	41.6-48.6	1.5	0.91	0.86	0.5	2.4	0.9	0.83	0.6	1.9	0.86	0.78	0.6	2.1
a*	(-4.8)-2.4	2	0.92	0.92	0.5	3.5	0.94	0.89	0.6	3.2	0.93	0.91	0.6	3.4
b*	9.6-18.4	1.7	0.62	0.59	1.2	1.6	0.56	0.48	1.2	1.5	0.58	0.54	1.2	1.5
Viscosity ŋ50	526-1029	119	0.93	0.86	32.3	4.0	0.86	0.82	45.6	3.1	0.86	0.79	47.4	2.8
Viscosity ŋ100	834-1721	210	0.94	0.85	55.5	4.0	0.86	0.83	81	2.8	0.85	0.78	85.3	2.7
DMC (g/g FW)	0.14-0.17	0.009	0.87	0.85	0.003	2.7	0.85	0.84	0.003	2.6	0.89	0.82	0.004	2.5
SSC (°Brix)	12.1-15.3	0.7	0.95	0.9	0.2	4.1	0.9	0.85	0.2	4.1	0.79	0.73	0.3	2.3
fructose (g/kg FW)	40.2-80.3	9.1	0.84	0.79	4.0	2.1	0.88	0.82	3.7	2.5	0.83	0.8	3.7	2.3
sucrose (g/kg FW)	33.2-57.3	5.5	0.88	0.86	2.0	2.8	0.87	0.85	2.1	2.7	0.88	0.83	2.1	2.7
glucose (g/kg FW)	13.2-28.3	3.7	0.94	0.94	0.9	4.3	0.97	0.9	1.1	3.2	0.93	0.94	0.9	3.7
pH	3.39-4.47	0.23	0.89	0.88	0.1	3.2	0.89	0.83	0.1	2.8	0.86	0.79	0.1	2.7
TA (meq/kg FW)	28.0-94.8	16.2	0.92	0.91	4.4	3.4	0.91	0.88	5.9	2.7	0.92	0.92	4.4	3.4
malic (g/kg FW)	3.0-8.8	1.3	0.95	0.93	0.3	4.7	0.94	0.87	0.4	3.9	0.95	0.95	0.3	4.3

Notes: all results corresponded to the averaged values of 10 replicates. R_v^2: determination coefficient of the validation test (internal); R_p^2: determination coefficient of the prediction test (external); RMSEP: root mean square error of prediction test (external); RPD: the residual predictive deviation of prediction test (external).

4. Conclusion

This was the first detailed work to show the ability of infrared spectroscopy coupled to suitable chemometric methods as a powerful tool to trace different composed varieties and estimate their corresponding compositions in apple purees. Moreover, an innovative chemometric method based on MCR-ALS was developed to reach simultaneous targets in terms of composition (in % of different variety) and physico-chemical properties (rheology, SSC, TA, DMC) of final puree products. As far as we know, this was the first report concerning the control of the final fruit product quality variations depending on the spectral information of the initial purees using a spectral reconstruction approach.

Vis-NIR on formulated purees could detect the composed single varieties purees with large color differences, such as 'Granny Smith' (R_P^2> 0.92, RPD> 3.4, RMSEP< 9.2%) and 'Braeburn' (R_P^2> 0.95, RPD> 4.2, RMSEP< 7.9%), but not for 'Golden Delicious' and 'Royal Gala'. MIR had the potential to trace the composed apple varieties with the excellent evaluations of 'Granny Smith' and 'Braeburn proportions' (RMSEP<4.3%, RPD> 7.7) and the satisfactory assessments of 'Golden Delicious' and 'Royal Gala' proportions (RMSEP<8.1%, RPD>3.6). And MIR could also predict the internal quality (SSC, TA, DMC, viscosity, pH, fructose, malic acid) of formulated purees coupled with PLS and machine learning regressions.

Innovatively, MIR technique opens the possibility to control and guidance the final puree characteristics by simply scanning the single-variety apple purees, in order to maintain the product quality or to drive the development of new products in apple industry. For instance, after acquiring MIR spectra of the four single-variety purees, our developed PLS models might be used in industry: i) to formulate purees with defined SSC and viscosity (e.g. 15.0 ± 0.3 °Brix and 1500 ± 100 $Pa.s^{-1}$, which might be reached with the formulate solutions as 75% GD-25% GS, 80% GD-20% BR and 90% GD-10% GA purees); or ii) to compare *in silico* the results of different puree formulation strategies, such as 33.3% GD and 66.6% GS purees (low redness, high acidity and viscosity) versus another strategy of 80% BR and 20% GA purees (more redness, low

acidity, low viscosity), depending on the market.

Further challenging works will be to investigate the possibility to reconstruct spectra of final processed purees based on spectra acquired directly on raw apples to provide non-destructive information guidance.

Highlights of Paper VII

This study provided the answers of the last question in Part 3:

How to manage apple puree variability?

- MIRS coupled with PLS estimated proportions of each apple variety in puree mixtures (RMSEP < 8.1%, RPD > 3.6)
- The concentration profiles from MCR-ALS made possible to reconstruct spectra of purees formulated by blending different single-variety purees.
- MIRS allowed to predict the final puree quality, such as redness (a values) (RPD = 3.5), viscosity (RPD > 4.0), SSC (RPD = 4.1), DMC (RPD = 2.7) and malic acid (RPD = 4.7), based only on the spectral data of initial single-variety purees.

On the one hand, Mid-infrared spectroscopy has the potential to trace – and thus authenticate the varieties in formulated apple purees. On the other hand, the spectral data of single-variety purees can also be used to control puree formulation, with a multiparameter optimization of texture and taste (viscosity, color, sugars and acids) of final blended apple purees.

V. Conclusions and perspectives

1. Conclusions

This thesis aimed to highlight the variability and heterogeneity of both apples and purees using vibrational spectroscopic methods (NIR, MIR, Raman and NIR-HSI) and advanced chemometrics (PCA, FDA, PLS, MCR-ALS and machine learning). The experimental trials were designed to obtain variability on apples and on corresponding purees by modulating several factors acting during production in the field (varieties, agricultural practices with fruit thinning), during post-harvest storage (4°C during 0, 1, 3 and 6 months) and processing (heating temperatures, grinding speeds and refining levels). The apple cooking was performed at a laboratory scale with a cutter-cooker robot to deeply study the link between before and after processing. The spectral information and reference data were acquired in parallel on raw apples and on corresponding purees, in order to explore new solutions to manage variability and heterogeneity during apple processing. Consequently, our research works were focused on specific questions, which were:

- **How to identify the variability and heterogeneity of raw apples and processed purees using different spectroscopic and imaging techniques?**

NIR-HSI spectrophotometry and chromatography allowed to highlight a large heterogeneity of DMC, TSC, contents of individual sugars, malic acid, and polyphenols in individual apples of four varieties (**Paper III**). This variability was observed in all directions inside each apple, from proximal to distal, from inside to outside and along equator. Compared to literature, our study provided an efficient solution for the HSI modelling calibration step, using the reference data measured on 141 representative samples instead of the 1056 prepared samples. Importantly, this method offered a new sight on TSC and DMC in apples, with a limited number of complex (individual sugars measured by spectrophotometry using enzymatic kits) and time-consuming (at least 24 hours for freeze-drying) analyses for HSI modelling. Based on our predicted results, the TSC and DMC in Braeburn and Royal Gala apples varied more intensively than in Granny Smith and Golden Delicious. However, this NIR-HSI method was not suitable to describe the distribution of the other studied variables in apples (fructose, glucose,

sucrose, malic acids and polyphenols).

Most previous studies (**Table 9**) investigating the relation between apple fruit and purees either investigated the variability factors (variety, agronomic system, storage duration etc.) one by one or separately apples and purees. In most of our experiments (**Papers I, II, IV, V, VI**), a wide range of apple variability was considered, combining the different factors, and a total traceability was ensured between apples and purees. The introduced variability on raw fruits resulted in significant differences of color (L*, a* and b*), texture (viscosity and viscoelasticity), TA, contents of SSC, fructose, sucrose, malic acid and AIS in the cooked purees (**Papers I, II & IV, Table 28**).

NIR had the potential to discriminate different raw apples and their corresponding cooked purees according to variety and storage duration, with a well-classified rate over 82% (**Paper IV**). Besides, it had a good ability to predict the global parameters (SSC, DMC, TA) both on raw apples and processed purees ($R^2 > 0.82$, RPD > 2.3) (**Papers I and IV**). However, it did not provide an acceptable estimation of puree viscosity as it was limited to discriminate their textural changes from different refining treatments. Combining the VIS and NIR spectroscopy allowed to predict the color parameters (L*, a* and b*), when the puree variability was large enough (**Paper IV, Table 29**).

VIS-NIR, NIR, MIR, Raman and HSI techniques were compared on purees (**Paper I**). MIR spectroscopy was the best tool to assess puree variability (variety, thinning practice and storage) and processing conditions (heating, grinding and refining), with a discrimination accuracy over 90.3%. MIRS was able to evaluate the puree's biochemical properties (SSC, TA, DMC, contents of fructose, sucrose and malic acid) with a RPD from 2.2 to 6.0 (**Papers I & VI**). It gave also the possibility to estimate puree rheological (viscosity and viscoelastic moduli) and textural (particle size and volume) properties, which were not predictable by VIS-NIR and NIR techniques (**Papers I & II**). MIRS also allowed determining cell wall (AIS) content and visualizing cell wall depolymerization (mainly pectin solubilization and galactose loss) but only on freeze-dried purees and on isolated cell wall materials, respectively. Consequently, MIR technique has been proved to be a valid method, using few steps to provide simultaneous assessments of biochemical, rheology and textural properties of apple

purees. It associated with suitable sample pre-treatments could offer valuable and sufficient information on puree quality for industrial and laboratorial demands.

- **How is the variability of apples linked to the quality of processed purees? And, is it possible to predict the quality of processed purees from VIS-NIR, NIR and MIR spectral signals of raw apples before processing?**

Two original experiments were conducted to explore how the intra- and the inter-variability of apple batches affected the properties of cooked purees (**Papers IV and V**):

Different purees were cooked, using a cutter-cooker robot, from a set of apples with a large inter-variability coming from different varieties, thinning practices and storage durations (**Paper IV**). A PCA performed on reference data of raw and processed apples evidenced highly consistent individual clusters, indicating that factors impacting raw fruits are retrieved in their corresponding purees (**Fig. 40**). From the reference data, good linear correlations were firstly determined for some characteristics between raw apples and cooked purees, for apple texture (firmness and crunchiness) and puree viscosity ($R^2 > 0.79$), TA ($R^2 > 0.91$), and for SSC ($R^2 > 0.79$) and DMC ($R^2 > 0.72$) (**Paper IV, Fig. 41**). However, these relationships were obtained on fruit batches needing at least 2.5 kg (approximately 10-15 apples sized for the use of our cutter-cooker robot), ignoring the 'intra-variability' brought by each individual apple.

An experiment was then designed to meet the absolute definition of 'one apple to one puree' using a microwave processing, which gave a first insight of the impact of intra-variability between individual apples on the quality of processed purees (**Paper V**). The variability coming from varieties, observed earlier, was confirmed here. However, a striking fact was that some quality traits were found to be much more variable between apples of a given variety. For example, puree color was much more variable for Braeburn and Royal Gala, while SSC was particularly variable for Golden Delicious and viscosity for Granny Smith. Conversely, some characteristics appeared to be quite stable and reliable within a given batch of one variety, like acidity for Royal Gala or colour for Golden delicious (**Fig. 45**). The skin color properties (L*, a* and b*)

of single apples were not linked to the viscosity of their cooked purees. A limitation of this study was that it was not possible to perform destructive quality measurements of individual apples as intact apples were directly cooked. Briefly, we confirmed that both, inter- and intra- variability of raw apples impacted the quality of cooked purees.

Regarding the relationships of reference quality traits between raw and processed apples, the VIS, NIR and MIR spectral information were supposed to be also linked, which was explored in **papers IV, V & VI**.

Two-dimensional correlation spectroscopy (2D-COS) was used to investigate VIS and NIR spectral correlations between each individual apple and its related puree (**Paper V**). Strong spectral correlations were observed in the VIS and NIR regions, particularly at 650-680 nm and 1125-1400 nm (**Fig. 49**). VIS-NIR gave a reliable assessment of cooked puree viscosity ($R^2 > 0.81$), SSC ($R^2 = 0.78$) and TA ($R^2 = 0.87$) based on their spectral information of individual raw apples (**Paper V, Table 34**). Similarly, satisfactory predictions of puree viscosity ($R^2 > 0.82$), SSC ($R^2 > 0.80$), TA ($R^2 > 0.80$) and even cell wall contents ($R^2 > 0.81$) were also obtained using the averaged NIR spectra of apple sets before cutter-cooker robot cooking (**Paper IV, Tables 31 & 32**). Therefore, VIS-NIR and NIR techniques should be useful tools for industry to assess the taste and texture of cooked purees based on the non-destructive VIS-NIR or NIR scanning of individual raw apples (**Fig. 64**). However, the main drawback of these two experiments was the need, for modelling, to systematically acquire the corresponding spectra on both, raw and processed materials with a large number of conditions, representative of the variability, giving often only rough assessments. In addition, our developed VIS-NIR or NIR models only provided some prediction results for only one standard cooking condition due to the involved labour.

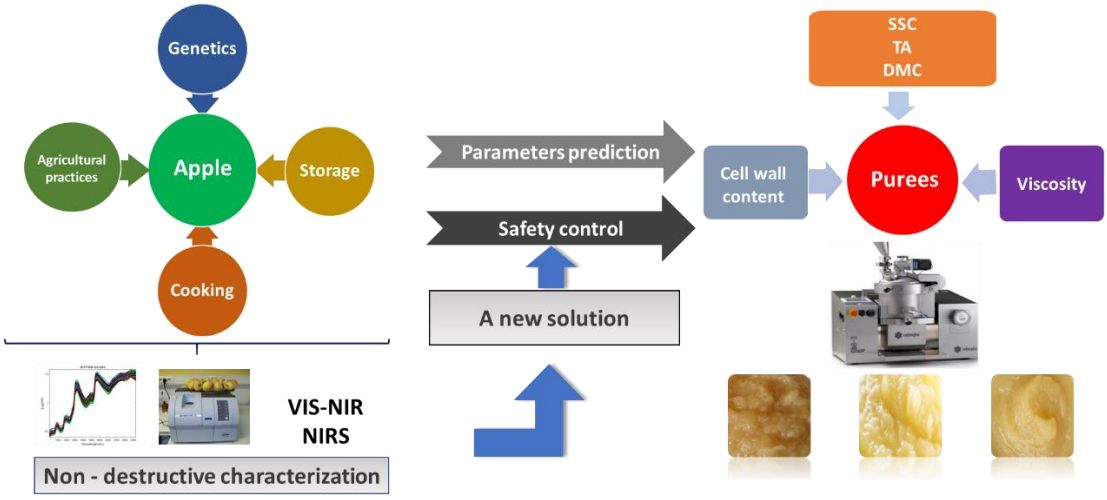

Fig. 64. A potential strategy to predict processed purees properties from VIS-NIR or NIR spectra of raw apples

In order to improve the prediction accuracy and explore the possibility to evaluate puree quality from multiple cooking conditions, an original experiment was designed in which different cooking conditions were compared (heating temperatures, grinding levels), and in which the direct standardization method was applied to establish the MIR spectral relationships between raw and processed apples (**Paper VI**). This strategy was proposed to build reconstructed MIR spectra of processed purees from the spectra of raw apple homogenates using a spectral transfer method. In this way, MIRS coupled with PLS models using these reconstructed puree spectra allowed to predict TA (R^2 > 0.86), SSC (R^2 > 0.85), DMC (R^2 > 0.84), particle averaged size (R^2 > 0.84) and viscosity (R^2 > 0.82) of cooked purees (**Paper VI**). After a simple scanning of raw apple homogenates by MIRS, our models allowed prediction of the quality traits of apple purees using nine different processing conditions (3 heating temperatures x 3 grinding levels). This approach is relevant for the processors and market to monitor and anticipate the organoleptic properties of cooked purees under different processing conditions (**Fig. 65**). However, scaling up tests have to be performed from the laboratory scale to reach the processors, especially as the processing units are not cooker-cutter robots.

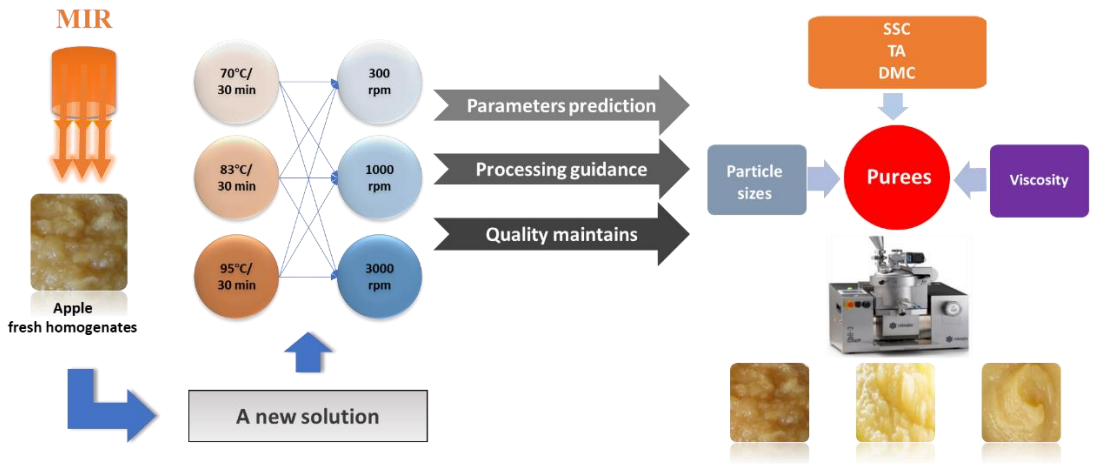

Fig. 65. A potential strategy to predict and maintain processed purees properties at different cooking conditions from MIR spectra of raw apple homogenates.

How to improve the puree formulation by infrared spectroscopy as an innovative solution to manage apple puree variability?

Papers IV, V & VI were mainly dedicated to explore solutions to manage variability from raw apples to their processed purees using infrared techniques. In **paper VII**, a focus was done on the development of a simple method to provide guidance for puree formulation after initial processing, in order to manage the final quality of puree products. To do that, four single-variety purees were mixed 2 by 2 following a systematic series of proportions (between 5% and 95%) and varieties to obtain 54 mixed purees.

An innovative chemometric method based on the concentration profiles from MCR-ALS was tested to reconstruct spectra of formulated purees, thus making the possibility to optimize physico-chemical properties of final puree products. As far as we know, this was the first report concerning the control of the quality of final fruit products depending on the spectral information of the initial purees using a spectral reconstruction approach.

MIR has the potential to guide puree formulation: a multi-parameter optimization was obtained for texture and taste (mainly viscosity, SSC, TA, malic acid, with RPD values > 4.0) of final apple purees using only the spectral data of single-variety purees.

After acquiring the MIR spectra on single-variety purees, this new method could optimize admixtures of purees and thus control the product quality, or even help the development of new products in apple industry (**Fig. 66**). However, VIS-NIR technique was not powerful enough to succeed in this strategy.

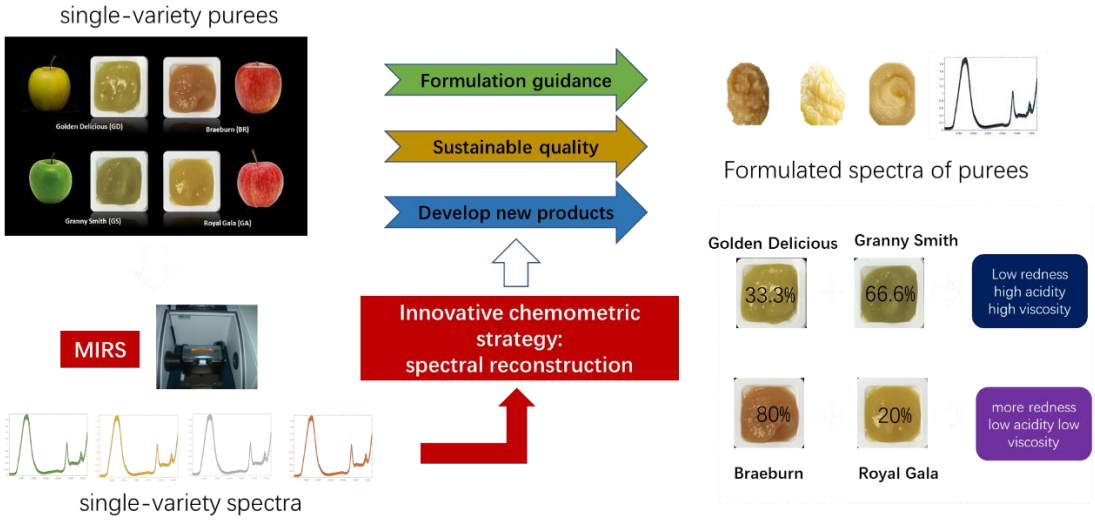

Fig. 66. Puree formulation guided by MIRS: a potential strategy to manage apple variability during processing.

Accordingly, this thesis provided several potential solutions for apple puree industry, in order to handle the ever-increasing variability and heterogeneity coming from climate change, different systems of production and processing technology. It is a first step towards a more sustainable and precise fruit processing by enlarging knowledge of how the spectra of raw materials can be used to manage the quality of final processed products. Our innovative results open the possibility to know which infrared range can be used to integrated analytical techniques in apple puree processing (**Fig. 67**) with for example:

i) At apple harvest, NIR-HSI technique can be used to rapidly characterize apple heterogeneity and map the sugars and dry matter content in apples, thus help researchers, field growers or industrial manufacturers to determine where and how many infrared measurements are needed to optimize the quality prediction and fruit sorting.

ii) After apple harvesting, VIS-NIR, NIR or MIR spectroscopy applied on raw apples provided an objective determination of quality properties of purees according to different processing conditions, and to know how to use apples in the best way and thus reduce wastes along the processing chain

iii) After puree processing, MIR technique applied on single-variety purees allows to guide puree formulation to reach stable standard processed products or innovate towards personalized purees.

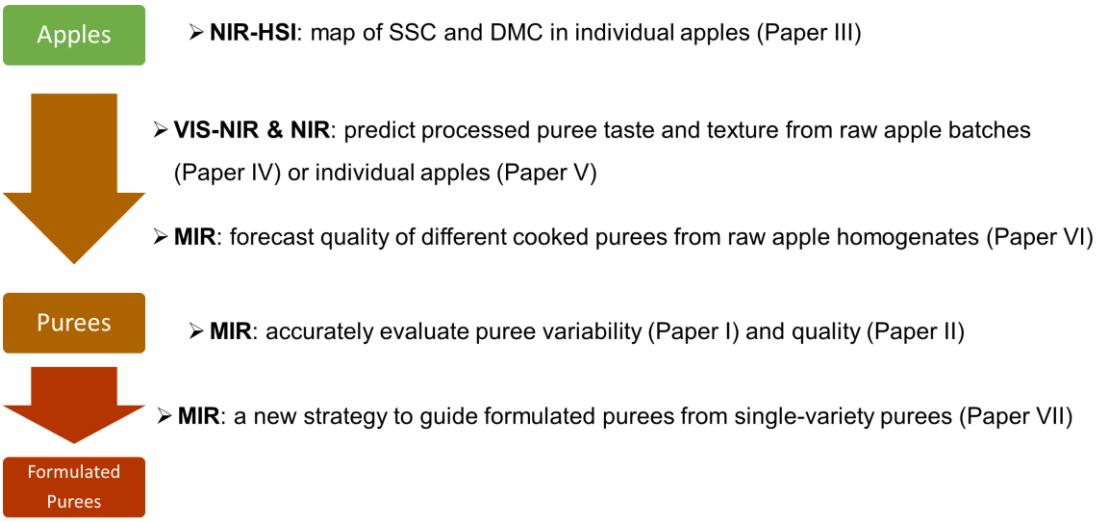

Fig. 67. A summary of our innovative solutions to manage variability and heterogeneity along apple puree processing.

2. Perspectives

Based on our results, there are some questions and ideas that need to be explored in our future researches, concerning the applications of several advanced chemometrics or the use of other methods to improve the detection of variability and heterogeneity

and the understanding of mechanisms occurring during apple processing.

- **Improvement of data treatment methods:**

  - Use 'SPORT' spectral pre-processing method to improve modelling accuracy. In our works, several basic pre-processing chemometrics including SNV, baseline correction, smoothing and derivatives, have been applied on spectra before modelling. It would be interesting to apply a new chemometric strategy named 'SPORT' (Mishra et al., 2020) allowing the fusion of multiple preprocessing methods and identifying the best preprocessing techniques and their combinations.
  - Investigate the fusion of NIR, MIR and Raman data to improve the evaluation of apple puree quality in a PAT strategy. A further work could be to investigate the fusion of these NIR, MIR and Raman data and evaluate the effect on prediction ability, such as low-level fusion with the direct combination of NIR, MIR and Raman spectra; middle-level fusion with selected bands of the NIRS, MIRS and Raman spectra according to evaluate parameters; high-level fusion directly conducted by outer-product analysis (OPA) and Grangere Ramanathan averaging (GRA).

- **Exploration of possibilities of infrared methods to better know F&V change during processing:**

  - Predict the diffusion of cell wall pectic content from pulp to serum using MIR technique on purees.

    A specific MIR region at 1595-1640 cm^{-1} was identified in fresh and freeze-dried samples before and after apple processing to contribute to the prediction of puree viscosity and viscoelasticity (**Paper II**). It is probably related to the solubilization of pectins, diffusing from puree's pulp to serum. In the 'Interfaces' project, pectin content of the raw apples and their corresponding cooked puree pulp have been characterized, according to the work of Alexandra Bürgy (2021). To go further, it should be interesting to evaluate not only the possibility to investigate the cell wall (AIS) content but also the cell wall composition and more specifically the pectin content impacting the texture of puree.

- Explore the VIS or VISNIR imaging of fruit surface to describe the internal quality distribution, and use microscopic MIR imaging to investigate the interactions of cell wall and polyphenols in different apple tissues.

  NIR-HSI technique successfully mapped the TSC and DMC in individual apples, with a destructive scanning of slices (**Paper III**). From a functional anatomy point of view, some interesting spatial relationships were observed between the internal chemical distribution (the highest values) and the fruit skin color (red color with the highest anthocyanin content) using our developed models (**Fig. 68**). If a good correlation between skin color and internal DMC and TSC was further identified, it would open the possibility to assess the chemical heterogeneity in apples based only on the VIS or VIS-NIR imaging acquired on apple surface, nondestructively. Further, a strong heterogeneity of total polyphenol contents was characterized in apples. This is one of the future interests in collaboration with INRAE BIA (Unité Biopolymères Interactions Assemblages) at Nantes to investigate the interaction of cell wall and polyphenols using MIR micro-spectroscopic imaging technique.

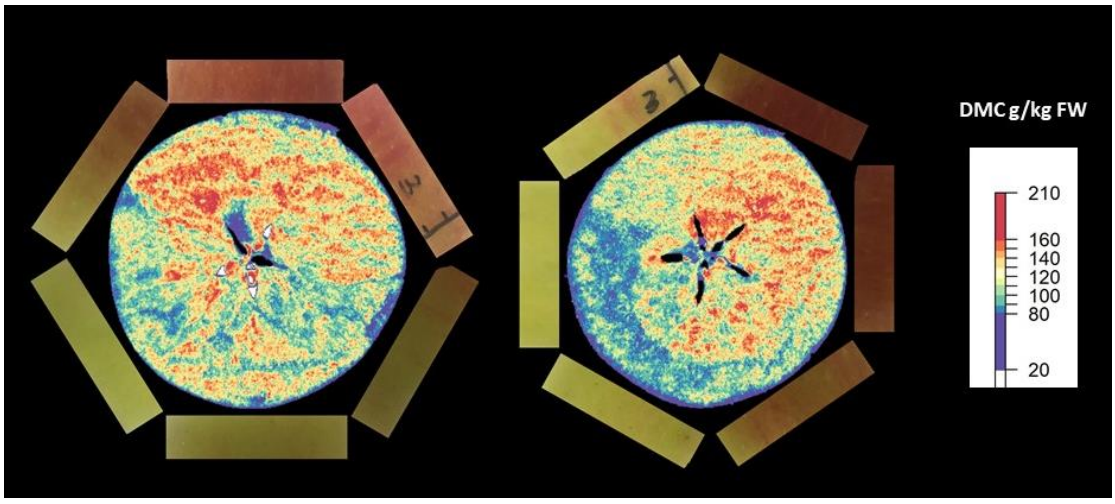

Fig. 68. A possible relationship between external skin color and internal DMC and TSC in Braeburn apples.

- **Integrate explicit PAT by using sensors in situ**

In our study, we have applied two solutions to bridge raw apples and cooked purees. The first one was the direct modelling using the reference data of final processed purees

and the infrared information of corresponding raw apples (**Papers IV and V**). The second one is to reconstructed puree spectra from their corresponding raw apples (**Paper VI**) or composed single-variety purees based on the spectral relationships before and after processing (**Paper VII**). One of our further interests is to explore the new solutions to guide apple processing in real time during the process. For example, a current project 'TransQuaPil' is on-going at UMR SQPOV concerning the use of a VIS-NIR sensor to acquire spectral information continuously during processing. Combining my results to this project might lead to better follow and control quality changes of fruit materials along the entire processing chain.

Infrared spectroscopy (near or mid-infrared) is one of the main candidates for a rapid qualification of agricultural commodities and processed food, especially in the view of PAT. Most of the previous infrared applications have been reported for the real-time evaluation of food quality and safety traits. However, knowing how to produce high-quality food using an efficient and precise strategy is even more crucial for industrial manufacturers, especially with continuous or increasing variability and heterogeneity along current food processing chain.

This thesis is a first step to explore the potential solutions to manage puree processing from the infrared information on apple materials. It is a proof-of-concept of what can be done and which infrared range can be used to integrate PAT in apple processing. With the development of advanced chemometrics and integrated infrared sensors, our results might open a new research topic named '**Smart Food Processing**', which can consider a large diversity of raw materials, apply efficient, sustainable and precise processing conditions, and provide new solutions to develop natural fruit products. With the scaling up and adapting our methods to large scale processing units, it could help food processors to reach several meaningful objectives based only on the rapid and simple spectral information on raw food materials, such as

- Rapid sorting or selecting raw fruit materials for suitable uses, including the fresh consumption or the consumption of various processed products (puree, juice or dried etc.) according to the market demand;
- Automatic choosing and adjusting of processing conditions (heating time,

temperature, pressure, grinding speed, etc.) to produce expected food products according to raw materials presenting variable properties and compositions. A recent Chinese national research program (Zou et al., 2019) has successfully made a first step to control the production of vinegar, based on infrared information acquired at several key stages during industrial fermentation. Till now, it has brought the incremental benefits over 40 million euros for 15 Chinese vinegar companies during 2016-2018.

- Developing and creating new products with the combination of infrared datasets and Artificial Intelligence (AI) powered system. Nowadays, some food and beverage companies, such as Coca Cola, use AI-powered tracking systems to extrapolate the potential response of their upcoming products. For example, Gastrograph AI (https://www.gastrograph.com/) is one such platform which provides historical consumer data to various companies. These data could be linked to the spectral dataset of 'Smart Food Processing' system, in order to further help food producers in understanding their consumer's preferences and provide specific processing guidance to reach the expected products for their target audiences.

VI. References

A

Abbas, O., Pissard, A., & Baeten, V. (2020). 3 - Near-infrared, mid-infrared, and Raman spectroscopy. In Y. Pico (Ed.), *Chemical Analysis of Food (Second Edition)* (pp. 77-134): Academic Press.

Abdi, H. (2003). Partial least square regression (PLS regression). *Encyclopedia for research methods for the social sciences, 6*(4), 792-795.

Abdi, H., Williams, L. J., & Valentin, D. (2013). Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *WIREs Computational Statistics, 5*(2), 149-179.

Abidi, N., Cabrales, L., & Haigler, C. H. (2014). Changes in the cell wall and cellulose content of developing cotton fibers investigated by FTIR spectroscopy. *Carbohydrate Polymers, 100*, 9-16.

Ahmed, J., & Ramaswamy, H. S. (2006). High pressure processing of fruits and vegetables. *Stewart Postharvest Review, 1*(8), 1-10.

Ahmed, J., & Ramaswamy, H. S. (2007). Dynamic rheology and thermal transitions in meat-based strained baby foods. *Journal of Food Engineering, 78*(4), 1274-1284.

Ali, M. A., Raza, H., Khan, M. A., & Hussain, M. (2004). Effect of different periods of ambient storage on chemical composition of apple fruit. *Int J Ag Biol, 6*(2), 568-571.

Amjad, W., Crichton, S. O. J., Munir, A., Hensel, O., & Sturm, B. (2018). Hyperspectral imaging for the determination of potato slice moisture content and chromaticity during the convective hot air drying process. *Biosystems Engineering, 166*, 170-183.

Anderson, N. T., Subedi, P. P., & Walsh, K. B. (2017). Manipulation of mango fruit dry matter content to improve eating quality. *Scientia Horticulturae, 226*, 316-321.

Andrianjaka-Camps, Z. N., Baumgartner, D., Camps, C., Guyer, E., Arrigoni, E., & Carlen, C. (2015). Prediction of raspberries puree quality traits by Fourier transform infrared spectroscopy. *LWT-Food Science and Technology, 63*(2), 1056-1062.

ANPP. (2020). Association Nationale Pommes Poires.

Aregawi, W. A., Defraeye, T., Verboven, P., Herremans, E., De Roeck, G., & Nicolai, B. M. (2013). Modeling of Coupled Water Transport and Large Deformation During Dehydration of Apple Tissue. *Food and Bioprocess Technology, 6*(8), 1963-1978.

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.

Asale, Y., Dessalegn, E., Assefa, D., & Abdisa, M. (2021). Phytochemicals and antioxidant activity of different apple cultivars grown in South Ethiopia: case of the wolayta zone. *International Journal of Food Properties, 24*(1), 354-363.

Ayora-Cañada, M. a. J., & Lendl, B. (2000). Sheath-flow Fourier transform infrared spectrometry for the simultaneous determination of citric, malic and tartaric acids in soft drinks. *Analytica Chimica Acta, 417*(1), 41-50.

Ayvaz, H., Sierra-Cadavid, A., Aykas, D. P., Mulqueeney, B., Sullivan, S., & Rodriguez-Saona, L. E. (2016). Monitoring multicomponent quality traits in tomato juice using portable mid-infrared (MIR) spectroscopy and multivariate analysis. *Food Control, 66*, 79-86.

B

Baeten, V., & Dardenne, P. (2005). Applications of Near-infrared Imaging for Monitoring Agricultural Food and Feed Products. In *Spectrochemical Analysis Using Infrared Multichannel Detectors* (pp. 283-301).

Bai, Y., Dougherty, L., Cheng, L., Zhong, G. Y., & Xu, K. (2015). Uncovering co-expression gene network modules regulating fruit acidity in diverse apples. *BMC Genomics, 16*(1).

Bain, J. M., & Robertson, R. (1951). The physiology of growth in apple fruits I. Cell size, cell number, and fruit development. *Australian Journal of Biological Sciences, 4*(2), 75-91.

Bain, J. M., & Robertson, R. N. (1951). The physiology of growth in apple fruits. I. Cell size, cell number, and fruit development. *Australian journal of scientific research. Ser. B: Biological sciences, 4*(2), 75-107.

Baker, M. J., Trevisan, J., Bassan, P., Bhargava, R., Butler, H. J., Dorling, K. M., . . . Martin, F. L. (2014). Using Fourier transform IR spectroscopy to analyze biological materials. *Nature Protocols, 9*(8), 1771-1791.

Barańska, H., Kuduk-Jaworska, J., Szostak, R., & Romaniewska, A. (2003). Vibrational spectra of racemic and enantiomeric malic acids. *Journal of Raman Spectroscopy, 34*(1), 68-76.

Baranska, M., Baranski, R., Schulz, H., & Nothnagel, T. (2006). Tissue-specific accumulation of carotenoids in carrot roots. *Planta, 224*(5), 1028-1037.

Baranska, M., & Schulz, H. (2005). Spatial tissue distribution of polyacetylenes in carrot root. *Analyst, 130*(6), 855-859.

Baranska, M., Schulz, H., Baranski, R., Nothnagel, T., & Christensen, L. P. (2005). In Situ Simultaneous Analysis of Polyacetylenes, Carotenoids and Polysaccharides in Carrot Roots. *Journal of Agricultural and Food Chemistry, 53*(17), 6565-6571.

Baranska, M., Schütze, W., & Schulz, H. (2006). Determination of lycopene and β-carotene content in tomato fruits and related products: comparison of FT-Raman, ATR-IR, and NIR spectroscopy. *Analytical Chemistry, 78*(24), 8456-8461.

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy, 43*(5), 772-777.

Barritt, B. H., Konishi, B. S., Drake, S. R., & Rom, C. R. (1997). Influence of sunlight level and rootstock on apple fruit quality. *Acta Horticulturae* (Vol. 451, pp. 569-577).

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.

Barron, C., Parker, M. L., Mills, E. N. C., Rouau, X., & Wilson, R. H. (2005). FTIR imaging of wheat endosperm cell walls in situ reveals compositional and architectural heterogeneity related to grain hardness. *Planta, 220*(5), 667-677.

Bartolozzi, F., Bertazza, G., Bassi, D., & Cristoferi, G. (1997). Simultaneous determination of soluble sugars and organic acids as their trimethylsilyl derivatives in apricot fruits by gas-liquid chromatography. *Journal of Chromatography A, 758*(1), 99-107.

Bassan, P., Sachdeva, A., Lee, J., & Gardner, P. (2013). Substrate contributions in micro-ATR of thin samples: implications for analysis of cells, tissue and biological fluids. *Analyst, 138*(14), 4139-4146.

Batjer, L. P., & Westwood, M. N. (1958). Size of Elberta and JH Hale peaches during the thinning period as related to size at harvest. *Proceedings of the American Society for Horticultural Science, 72*, 102-105.

Bearth, A., Cousin, M.-E., & Siegrist, M. (2014). The consumer's perception of artificial food additives: Influences on acceptance, risk and benefit perceptions. *Food Quality and Preference, 38*, 14-23.

BEEDATA. (2020). 2020 China Apple Industry Analysis Report.

Beghi, R., Giovanelli, G., Malegori, C., Giovenzana, V., & Guidetti, R. (2014). Testing of a VIS-NIR system for the monitoring of long-term apple storage. *Food and Bioprocess Technology, 7*(7), 2134-2143.

Beghi, R., Giovenzana, V., Civelli, R., & Guidetti, R. (2016). Influence of packaging in the analysis of fresh-cut Valerianella locusta L. and Golden Delicious apple slices by visible-near infrared and near infrared spectroscopy. *Journal of Food Engineering, 171*, 145-152.

Belton, P. S., Kemsley, E. K., McCann, M. C., Ttofis, S., Wilson, R. H., & Delgadillo, I. (1995). The identification of vegetable matter using Fourier Transform Infrared Spectroscopy. *Food Chemistry, 54*(4), 437-441.

Bengoechea, M. L., Sancho, A. I., Bartolomé, B., Estrella, I., Gómez-Cordovés, C., & Hernández, M. T. (1997). Phenolic Composition of Industrially Manufactured Purées and Concentrates from Peach and Apple Fruits. *Journal of Agricultural and Food Chemistry, 45*(10), 4071-4075.

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

Bergh, O. (1990). Effect of time of hand-thinning on apple fruit size. *South African Journal of Plant and Soil, 7*(1), 1-10.

Bergh, O. (1992). Cumulative effect of time of hand-thinning on fruit size of Golden Delicious and Granny Smith apples. *South African Journal of Plant and Soil, 9*(2), 64-67.

Bertrand, D., & Cordella, C. (2008). SAISIR package. Free toolbox for chemometrics in the Matlab, Octave or Scilab environments. In). https://www.chimiometrie.fr/saisir_webpage.html.

Bhargava, R. (2012). Infrared Spectroscopic Imaging: The Next Generation. *Applied Spectroscopy, 66*(10), 1091-1120.

Bhushan, S., Kalia, K., Sharma, M., Singh, B., & Ahuja, P. S. (2008). Processing of apple pomace for bioactive molecules. *Critical reviews in biotechnology, 28*(4), 285-296.

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

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

Boldrini, B., Kessler, W., Rebner, K., & Kessler, R. W. (2012). Hyperspectral Imaging: A Review of Best Practice, Performance and Pitfalls for in-line and on-line Applications. *Journal of Near Infrared Spectroscopy, 20*(5), 483-508.

Bollard, E. G. (1970). The physiology and nutrition of developing fruits. *The Biochemistry of Fruits and Their Products, 1*, 387-425.

Bondonno, N. P., Bondonno, C. P., Ward, N. C., Hodgson, J. M., & Croft, K. D. (2017). The cardiovascular health benefits of apples: Whole fruit vs. isolated compounds. *Trends in Food Science & Technology, 69*, 243-256.

Boyer, J., & Liu, R. H. (2004a). Apple phytochemicals and their health benefits. *Nutrition Journal, 3*(1), 1-15.

Brackmann, C., Bengtsson, A., Alminger, M. L., Svanberg, U., & Enejder, A. (2011). Visualization of β-

carotene and starch granules in plant cells using CARS and SHG microscopy. *Journal of Raman Spectroscopy, 42*(4), 586-592.

Brookfield, P. L., Nicoll, S., Gunson, F. A., Harker, F. R., & Wohlers, M. (2011). Sensory evaluation by small postharvest teams and the relationship with instrumental measurements of apple texture. *Postharvest Biology and Technology, 59*(2), 179-186.

Buccheri, M., Grassi, M., Lovati, F., Petriccione, M., Rega, P., Scalzo, R. L., & Cattaneo, T. M. P. (2019). Near infrared spectroscopy in the supply chain monitoring of Annurca apple. *Journal of Near Infrared Spectroscopy, 27*(1), 86-92.

Buergy, A. (2021). Modulation of the texture and tissue fragmentation of fruits during thermal treatments by the methods of cultivation and maturation : impact on the texture of the purees. *Ph.D Thesis*, Avignon University, Avignon, France.

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.

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.

Burgy, A., Rolland-Sabaté, A., Leca, A., Le Bourvellec, C., Renard, C. M. G. C. (2018). Molecular size of soluble pectins in apple fruits is not affected by heat processing into puree. International Conference on Agrophysics Soil, Plant & Climate, Dublin, POL.

Bureau, S., Arbex de Castro Vilas Boas, A., Giovinazzo, R., Jaillais, B., & Page, D. (2020). Toward the implementation of mid-infrared spectroscopy along the processing chain to improve quality of the tomato based products. *LWT- Food Science and Technology, 130*, 109518.

Bureau, S., Cozzolino, D., & Clark, C. J. (2019). Contributions of Fourier-transform mid infrared (FT-MIR) spectroscopy to the study of fruit and vegetables: A review. *Postharvest Biology and Technology, 148*, 1-14.

Bureau, S., Page, D., Bogé, M., & Renard, C. M. G. (2015). Mid-infrared spectroscopy as a rapid and convenient tool for the characterisation of tomato purees. *Acta Horticulturae* (Vol. 1081, pp. 317-322).

Bureau, S., Quilot-Turion, B., Signoret, V., Renaud, C., Maucourt, M., Bancel, D., & Renard, C. M. G. C. (2013). Determination of the Composition in Sugars and Organic Acids in Peach Using Mid Infrared Spectroscopy: Comparison of Prediction Results According to Data Sets and Different Reference Methods. *Analytical Chemistry, 85*(23), 11312-11318.

Bureau, S., Ruiz, D., Reich, M., Gouble, B., Bertrand, D., Audergon, J. M., & Renard, C. (2009). Application of ATR-FTIR for a rapid and simultaneous determination of sugars and organic acids in apricot fruit. *Food Chemistry, 115*(3), 1133-1140.

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. *Journal of Agricultural and Food Chemistry, 60*(14), 3551-3563.

Bureau, S., Vilas-Boas, A., Giovinazzo, R., & Page, D. (2019). Quality improvement of the tomato-based products using infrared tool along the production chain. (1233 ED., pp. 169-174): International Society for Horticultural Science (ISHS), Leuven, Belgium.

C

Calligaris, S., Manzocco, L., Kravina, G., Sovrano, S., Anese, M., & Nicoli, M. (2006). Processing suitability of different apple cultivars from Friuli-Venezia Giulia. *Industrie Alimentari (Italy)*.

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.

Camerlingo, C., Zenone, F., Delfino, I., Diano, N., Mita, D. G., & Lepore, M. (2007). Investigation on clarified fruit juice composition by using visible light micro-Raman spectroscopy. *Sensors, 7*(10), 2049-2061.

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

Cardoso, S. M., Mafra, I., Reis, A., Barros, A. S., Nunes, C., Georget, D. M. R., Smith, A. C., Saraiva, J., Waldron, K. W., & Coimbra, M. A. (2009). Traditional and industrial oven-dry processing of olive fruits: influence on textural properties, cell wall polysaccharide composition, and enzymatic activity. *European Food Research and Technology, 229*(3), 415-425.

Centofanti, T., Bañuelos, G. S., & Ayars, J. E. (2019). Fruit nutritional quality under deficit irrigation: the case of table grapes in California. *Journal of the Science of Food and Agriculture, 99*(5), 2215-2225.

Cerchiaro, G., Sant'Ana, A. C., Temperini, M. L. A., & da Costa Ferreira, A. M. (2005). Investigations of different carbohydrate anomers in copper(II) complexes with d-glucose, d-fructose, and d-galactose by Raman and EPR spectroscopy. *Carbohydrate Research, 340*(15), 2352-2359.

Cevik, V., Ryder, C. D., Popovich, A., Manning, K., King, G. J., & Seymour, G. B. (2010). A FRUITFULL-like gene is associated with genetic variation for fruit flesh firmness in apple (Malus domestica Borkh.). *Tree Genetics & Genomes, 6*(2), 271-279.

Chan, K. L. A., Altharawi, A., Fale, P., Song, C. L., Kazarian, S. G., Cinque, G., Sockalingum, G. D. (2020). Transmission Fourier Transform Infrared Spectroscopic Imaging, Mapping, and Synchrotron Scanning Microscopy with Zinc Sulfide Hemispheres on Living Mammalian Cells at Sub-Cellular Resolution. *Applied Spectroscopy, 74*(5), 544-552.

Chang, W. H., Chen, S., & Tsai, C. C. (1998). Development of a universal algorithm for use of NIR in estimation of soluble solids in fruit juices. *Transactions of the American Society of Agricultural Engineers, 41*(6), 1739-1745.

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

Cheng, J., & Sun, D. (2017). Partial Least Squares Regression (PLSR) Applied to NIR and HSI Spectral Data Modeling to Predict Chemical Properties of Fish Muscle. *Food Engineering Reviews, 9*(1), 36-49.

Cho, R. K., Sohn, M. R., & Kwon, Y. K. (1998). New observation of nondestructive evaluation for sweetness in apple fruit using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy, 6*(1-4), A75-A78.

Chylińska, M., Szymańska-Chargot, M., Deryło, K., Tchórzewska, D., & Zdunek, A. (2017). Changing of biochemical parameters and cell wall polysaccharides distribution during physiological development of tomato fruit. *Plant Physiology and Biochemistry, 119*, 328-337.

Chylińska, M., Szymańska-Chargot, M., & Zdunek, A. (2014). Imaging of polysaccharides in the tomato cell wall with Raman microspectroscopy. *Plant Methods, 10*(1), 14.

Clark, C. J. (2016). Fast determination by Fourier-transform infrared spectroscopy of sugar–acid composition of citrus juices for determination of industry maturity standards. *New Zealand*

Journal of Crop and Horticultural Science, 44(1), 69-82.

Clark, C. J., Hockings, P. D., Joyce, D. C., & Mazucco, R. A. (1997). Application of magnetic resonance imaging to pre- and post-harvest studies of fruits and vegetables. *Postharvest Biology and Technology, 11*(1), 1-21.

Coimbra, M. A., Barros, A., Barros, M., Rutledge, D. N., & Delgadillo, I. (1998). Multivariate analysis of uronic acid and neutral sugars in whole pectic samples by FT-IR spectroscopy. *Carbohydrate Polymers, 37*(3), 241-248.

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

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

Contal, L., León, V., & Downey, G. (2002). Detection and Quantification of Apple Adulteration in Strawberry and Raspberry Purées Using Visible and near Infrared Spectroscopy. *Journal of Near Infrared Spectroscopy, 10*(4), 289-299.

Coombe, B. (1976). The development of fleshy fruits. *Annual Review of Plant Physiology, 27*(1), 207-228.

Cordella, C. B. Y., & Bertrand, D. (2014). SAISIR: A new general chemometric toolbox. *TRAC Trends in Analytical Chemistry, 54*, 75-82.

Corelli-Grappadelli, L., & Lakso, A. N. (2004). Fruit development in deciduous tree crops as affected by physiological factors and environmental Conditions. *Acta Horticulturae* (Vol. 636, pp. 425-441).

Cornille, A., Giraud, T., Smulders, M. J., Roldan-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends Genet, 30*(2), 57-65.

Cornille, A., Gladieux, P., Smulders, M. J., Roldan-Ruiz, I., Laurens, F., Le Cam, B., Feugey, L. (2012). New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet, 8*(5), e1002703.

Corollaro, M. L., Endrizzi, I., Bertolini, A., Aprea, E., Demattè, M. L., Costa, F., . . . Gasperi, F. (2013). Sensory profiling of apple: Methodological aspects, cultivar characterisation and postharvest changes. *Postharvest Biology and Technology, 77*, 111-120.

Cortés, V., Cubero, S., Blasco, J., Aleixos, N., & Talens, P. (2019). In-line Application of Visible and Near-Infrared Diffuse Reflectance Spectroscopy to Identify Apple Varieties. *Food and Bioprocess Technology, 12*(6), 1021-1030.

Costa, G., Noferini, M., Fiori, G., Orlandi, A., & Miserocchi, O. (2003). Non-destructive technique to assess internal fruit quality. *Acta Horticulturae* (Vol. 604, pp. 571-576).

Cosgrove, D. J., & Jarvis, M. C. (2012). Comparative structure and biomechanics of plant primary and secondary cell walls. Frontiers in Plant Science, 3(204).

Cuello, C., Marchand, P., Laurans, F., Grand-Perret, C., Laine-Prade, V., Pilate, G., & Dejardin, A. (2018). ATR-FTIR imaging: phenotyping at the cell wall level in poplar wood. *Paris single cell day 2018*.

Cullen, P. J., O'Donnell, C. P., & Fagan, C. C. (2014). Benefits and challenges of adopting PAT for the food industry. In *Process analytical technology for the food industry* (pp. 1-5): Springer.

Cullen, P. J., Duffy, A. P., & O'Donnell, C. P. (2001). On-line rheological characterization of pizza sauce using tube viscometry. *Journal of Food Process Engineering, 24*(3), 145-159.

D

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.

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.

Davis, K. F., Downs, S., & Gephart, J. A. (2021). Towards food supply chain resilience to environmental shocks. *Nature Food, 2*(1), 54-65.

De Beer, T. R. M., Vercruysse, P., Burggraeve, A., Quinten, T., Ouyang, J., Zhang, X., Baeyens, W. R. G. (2009). In-line and real-time process monitoring of a freeze drying process using Raman and NIR spectroscopy as complementary process analytical technology (PAT) tools. *Journal of Pharmaceutical Sciences, 98*(9), 3430-3446.

De Belie, N., Schotte, S., Coucke, P., & De Baerdemaeker, J. (2000). Development of an automated monitoring device to quantify changes in firmness of apples during storage. *Postharvest Biology and Technology, 18*(1), 1-8.

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.

De Juan, A., & Tauler, R. (2006). Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications. *Critical Reviews in Analytical Chemistry, 36*(3-4), 163-176.

De Jager, A., & Roelofs, F. P. M. M. (1996). Prediction of optimum harvest date of Jonagold. *Determination and Prediction of Optimum Harvest Date of Apples and Pears*, 21-31.

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 Chemistry. 143*, 223-230.

De Oliveira, G. A., de Castilhos, F., Renard, C. M. G. C., & Bureau, S. (2014). Comparison of NIR and MIR spectroscopic methods for determination of individual sugars, organic acids and carotenoids in passion fruit. *Food Research International, 60*, 154-162.

De Santana, F. B., Borges Neto, W., & Poppi, R. J. (2019). Random forest as one-class classifier and infrared spectroscopy for food adulteration detection. *Food Chemistry, 293*, 323-332.

DeEll, J. R., Khanizadeh, S., Saad, F., & Ferree, D. C. (2001). Factors affecting apple fruit firmness - A review. *Journal of the American Pomological Society, 55*(1), 8-27.

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

Defraeye, T., Lehmann, V., Gross, D., Holat, C., Herremans, E., Verboven, P., Nicolai, B. M. (2013). Application of MRI for tissue characterisation of 'Braeburn' apple. *Postharvest Biology and Technology, 75*, 96-105.

Dennis Jr, F. G. (2000). The history of fruit thinning. *Plant Growth Regulation, 31*(1-2), 1-16.

Dhaulaniya, A. S., Balan, B., Sodhi, K. K., Kelly, S., Cannavan, A., & Singh, D. K. (2020). Qualitative and quantitative evaluation of corn syrup as a potential added sweetener in apple fruit juices using mid-infrared spectroscopy assisted chemometric modeling. *LWT- Food Sciecnce and Technology, 131*, 109749.

Dhaulaniya, A. S., Balan, B., Yadav, A., Jamwal, R., Kelly, S., Cannavan, A., & Singh, D. K. (2020). Development of an FTIR based chemometric model for the qualitative and quantitative

evaluation of cane sugar as an added sugar adulterant in apple fruit juices. *Food Additives and Contaminants - Part A Chemistry, Analysis, Control, Exposure and Risk Assessment, 37*(4), 539-551.

Diaz, J. V., Anthon, G. E., & Barrett, D. M. (2007). Nonenzymatic degradation of citrus pectin and pectate during prolonged heating: effects of pH, temperature, and degree of methyl esterification. *Journal of Agricultural and Food Chemistry, 55*(13), 5131-5136.

Do Trong, N. N., Erkinbaev, C., Tsuta, M., De Baerdemaeker, J., Nicolai, B., & Saeys, W. (2014). Spatially resolved diffuse reflectance in the visible and near-infrared wavelength range for non-destructive quality assessment of 'Braeburn' apples. *Postharvest Biology and Technology, 91*, 39-48.

Doerflinger, F. C., Miller, W. B., Nock, J. F., & Watkins, C. B. (2015). Relationships between starch pattern indices and starch concentrations in four apple cultivars. *Postharvest Biology and Technology, 110*, 86-95.

Dong, J., & Guo, W. (2015). Nondestructive Determination of Apple Internal Qualities Using Near-Infrared Hyperspectral Reflectance Imaging. *Food Analytical Methods, 8*(10), 2635-2646.

Downey, G. (1998). Food and food ingredient authentication by mid-infrared spectroscopy and chemometrics. *TrAC Trends in Analytical Chemistry, 17*(7), 418-424.

Downey, G., & Kelly, J. D. (2004). Detection and Quantification of Apple Adulteration in Diluted and Sulfited Strawberry and Raspberry Purées Using Visible and Near-Infrared Spectroscopy. *Journal of Agricultural and Food Chemistry, 52*(2), 204-209.

Dragovic-Uzelac, V., Pospišil, J., Levaj, B., & Delonga, K. (2005). The study of phenolic profiles of raw apricots and apples and their purees by HPLC for the evaluation of apricot nectars and jams authenticity. *Food Chemistry, 91*(2), 373-383.

Du, Z., Zeng, X., Li, X., Ding, X., Cao, J., & Jiang, W. (2020). Recent advances in imaging techniques for bruise detection in fruits and vegetables. *Trends in Food Science and Technology, 99*, 133-141.

Duncan, W. D., & Williams, G. P. (1983). Infrared synchrotron radiation from electron storage rings. *Applied Optics, 22*(18), 2914-2923.

Dupuy, N., Duponchel, L., Huvenne, J. P., Sombret, B., & Legrand, P. (1996). Classification of edible fats and oils by principal component analysis of Fourier transform infrared spectra. *Food Chemistry, 57*(2), 245-251.

Dupuy, N., Galtier, O., Ollivier, D., Vanloot, P., & Artaud, J. (2010). Comparison between NIR, MIR, concatenated NIR and MIR analysis and hierarchical PLS model. Application to virgin olive oil analysis. *Analytica Chimica Acta, 666*(1-2), 23-31.

Durak, T., & Depciuch, J. (2020). Effect of plant sample preparation and measuring methods on ATR-FTIR spectra results. *Environmental and Experimental Botany, 169*, 103915.

E

Ebel, R. C., Proebsting, E. L., & Patterson, M. E. (1993). Regulated deficit irrigation may alter apple maturity, quality, and storage life. *HortScience, 28*(2), 141-143.

Ebermann, R., & Elmadfa, I. (2008). *Lehrbuch Lebensmittelchemie und Ernährung*: Springer-Verlag.

Eisele, T. A., & Drake, S. R. (2005). The partial compositional characteristics of apple juice from 175 apple varieties. *Journal of Food Composition and Analysis, 18*(2), 213-221.

Eisenstecken, D., Panarese, A., Robatscher, P., Huck, C. W., Zanella, A., & Oberhuber, M. (2015). A near

infrared spectroscopy (NIRS) and chemometric approach to improve apple fruit quality management: A case study on the cultivars "cripps pink" and "braeburn". *Molecules, 20*(8), 13603-13619.

Eisenstecken, D., Stürz, B., Robatscher, P., Lozano, L., Zanella, A., & Oberhuber, M. (2019). The potential of near infrared spectroscopy (NIRS) to trace apple origin: Study on different cultivars and orchard elevations. *Postharvest Biology and Technology, 147*, 123-131.

Elzebroek, A. T. G. (2008). *Guide to cultivated plants*: CABI.

Engelen, L., & de Wijk, R. A. (2012). Oral Processing and Texture Perception. In J. Chen & L. Engelen (Eds.), *Food Oral Processing*, (pp. 157-176): Wiley-Blackwell.

Escribano, S., Biasi, W. V., Lerud, R., Slaughter, D. C., & Mitcham, E. J. (2017). Non-destructive prediction of soluble solids and dry matter content using NIR spectroscopy and its relationship with sensory quality in sweet cherries. *Postharvest Biology and Technology, 128*, 112-120.

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

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

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

Eurostat. (2017). Where are our Fruit and Veg produced?

Evers, T. M. J., Hochane, M., Tans, S. J., Heeren, R. M. A., Semrau, S., Nemes, P., & Mashaghi, A. (2019). Deciphering Metabolic Heterogeneity by Single-Cell Analysis. *Analytical Chemistry, 91*(21), 13314-13323.

F

Fadanelli, L., Comai, M., Dorigoni, A., Mattivi, F., & Boschetti, A. (2004). Influence of crop load and production site on quality of Golden Delicious apples during storage. *V International Postharvest Symposium 682* (pp. 749-756).

Fahey, L. M., Nieuwoudt, M. K., & Harris, P. J. (2017). Predicting the cell-wall compositions of Pinus radiata (radiata pine) wood using ATR and transmission FTIR spectroscopies. *Cellulose, 24*(12), 5275-5293.

Fan, G., Zha, J., Du, R., & Gao, L. (2009). Determination of soluble solids and firmness of apples by Vis/NIR transmittance. *Journal of Food Engineering, 93*(4), 416-420.

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.

FAO. (2015). Global initiative on food loss and waste reduction. *Key facts on food loss and waste you should know*, 1-2.

Fasoli, M., Dell'Anna, R., Dal Santo, S., Balestrini, R., Sanson, A., Pezzotti, M., Monti, F., & Zenoni, S. (2016). Pectins, Hemicelluloses and Celluloses Show Specific Dynamics in the Internal and External Surfaces of Grape Berry Skin During Ripening. *Plant and Cell Physiology, 57*(6), 1332-1349.

Femenia, A., García-Pascual, P., Simal, S., & Rosselló, C. (2003). Effects of heat treatment and dehydration on bioactive polysaccharide acemannan and cell wall polymers from Aloe barbadensis Miller. *Carbohydrate Polymers, 51*(4), 397-405.

Ferreira, D., Barros, A., Coimbra, M. A., & Delgadillo, I. (2001). Use of FT-IR spectroscopy to follow the effect of processing in cell wall polysaccharide extracts of a sun-dried pear. *Carbohydrate Polymers, 45*(2), 175-182.

Fischer, R. L., & Bennett, A. B. (1991). Role of cell wall hydrolases in fruit ripening.Annual Review of Plant Physiology and Plant Molecular Biology, 42(1), 675-703.

Franceschi, P., Dong, Y., Strupat, K., Vrhovsek, U., & Mattivi, F. (2012). Combining intensity correlation analysis and MALDI imaging to study the distribution of flavonols and dihydrochalcones in Golden Delicious apples. *Journal of Experimental Botany, 63*(3), 1123-1133.

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

Fox, G. (2020). The brewing industry and the opportunities for real-time quality analysis using infrared spectroscopy. *Applied Sciences,* 10(2), 616.

Fügel, R., Carle, R., & Schieber, A. (2005). Quality and authenticity control of fruit purées, fruit preparations and jams—a review. *Trends in Food Science & Technology, 16*(10), 433-441.

G

Galtier, O., Abbas, O., Le Dréau, Y., Rebufa, C., Kister, J., Artaud, J., & Dupuy, N. (2011). Comparison of PLS1-DA, PLS2-DA and SIMCA for classification by origin of crude petroleum oils by MIR and virgin olive oils by NIR for different spectral regions. *Vibrational Spectroscopy, 55*(1), 132-140.

Garrido, M., Rius, F. X., & Larrechi, M. S. (2008). Multivariate curve resolution–alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes. *Analytical and Bioanalytical Chemistry, 390*(8), 2059-2066.

Garside, P., & Wyeth, P. (2003). Identification of cellulosic fibres by FTIR spectroscopy-thread and single fibre analysis by attenuated total reflectance. *Studies in conservation, 48*(4), 269-275.

Georgé, S., Brat, P., Alter, P., & Amiot, M. J. (2005). Rapid Determination of Polyphenols and Vitamin C in Plant-Derived Products. *Journal of Agricultural and Food Chemistry, 53*(5), 1370-1373.

Georget, D. M. R., & Belton, P. S. (2006). Effects of Temperature and Water Content on the Secondary Structure of Wheat Gluten Studied by FTIR Spectroscopy. *Biomacromolecules, 7*(2), 469-475.

Gestal, M., Gomez-Carracedo, M. P., Andrade, J. M., Dorado, J., Fernandez, E., Prada, D., & Pazos, A. (2004). Classification of apple beverages using artificial neural networks with previous variable selection. *Analytica Chimica Acta, 524*(1-2), 225-234.

Ghosh S. (2019). Principal Component Analysis in R – Walk Through. https://dimensionless.in/principal-component-analysis-in-r/

Gidley, M. J., & Yakubov, G. E. (2019). Functional categorisation of dietary fibre in foods: Beyond 'soluble' vs 'insoluble'. *Trends in Food Science & Technology, 86*, 563-568.

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

Giusti, M. M., & Wrolstad, R. E. (2001). Characterization and Measurement of Anthocyanins by UV-Visible Spectroscopy. *Current Protocols in Food Analytical Chemistry, 00*(1), F1.2.1-F1.2.13.

Goetz, A. F. H., Vane, G., Solomon, J. E., & Rock, B. N. (1985). Imaging Spectrometry for Earth Remote

Sensing. *Science, 228*(4704), 1147-1153.

Goh, K. M., Maulidiani, M., Rudiyanto, R., Wong, Y. H., Ang, M. Y., Yew, W. M., Abas, F., Lai, O. M., Wang, Y., & Tan, C. P. (2019). Rapid assessment of total MCPD esters in palm-based cooking oil using ATR-FTIR application and chemometric analysis. *Talanta, 198*, 215-223.

Golic, M., Walsh, K., & Lawson, P. (2003). Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature. *Applied Spectroscopy, 57*(2), 139-145.

Gongal, A., Amatya, S., Karkee, M., Zhang, Q., & Lewis, K. (2015). Sensors and systems for fruit detection and localization: A review. *Computers and Electronics in Agriculture, 116*, 8-19.

González-Cabrera, M., Domínguez-Vidal, A., & Ayora-Cañada, M. J. (2018). Hyperspectral FTIR imaging of olive fruit for understanding ripening processes. *Postharvest Biology and Technology, 145*, 74-82.

Gonzalvez, A. G., Martin, D., Slowing, K., & Gonzalez Ureña, A. (2014). Insights into the β-carotene distribution in carrot roots. *Food Structure, 2*(1), 61-65.

Gorry, P. A. (1990). General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method. *Analytical Chemistry, 62*(6), 570-573.

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*(12), 590-598.

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

Griffiths, P. R., & Miseo, E. V. (2014). Infrared and Raman Instrumentation for Mapping and Imaging. In *Infrared and Raman Spectroscopic Imaging* (pp. 1-56).

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

Guendel, A., Rolletschek, H., Wagner, S., Muszynska, A., & Borisjuk, L. (2018a). Micro Imaging Displays the Sucrose Landscape within and along Its Allocation Pathways. *Plant Physiology, 178*(4), 1448-1460.

Guinda, Á., Rada, M., Delgado, T., Gutiérrez-Adánez, P., & Castellano, J. M. (2010). Pentacyclic Triterpenoids from Olive Fruit and Leaf. *Journal of Agricultural and Food Chemistry, 58*(17), 9685-9691.

Guo, Q., Sun, D., Cheng, J., & Han, Z. (2017). Microwave processing techniques and their recent applications in the food industry. *Trends in Food Science & Technology, 67*, 236-247.

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.

H

Hamadziripi, E. T., Theron, K. I., Muller, M., & Steyn, W. J. (2014). Apple compositional and peel color differences resulting from canopy microclimate affect consumer preference for eating quality and appearance. *HortScience, 49*(3), 384-392.

Hameed, K., Chai, D., & Rassau, A. (2018). A comprehensive review of fruit and vegetable classification techniques. *Image and Vision Computing, 80*, 24-44.

Hampson, C. R., Sanford, K., & Cline, J. (2002). Preferences of Canadian consumers for apple fruit size.

Canadian Journal of Plant Science, 82(1), 165-167.

Harker, F. R., Gunson, F. A., & Jaeger, S. R. (2003). The case for fruit quality: an interpretive review of consumer attitudes, and preferences for apples. *Postharvest Biology and Technology, 28*(3), 333-347.

Harker, F. R., Maindonald, J., Murray, S. H., Gunson, F. A., Hallett, I. C., & Walker, S. B. (2002). Sensory interpretation of instrumental measurements 1: texture of apple fruit. *Postharvest Biology and Technology, 24*(3), 225-239.

Harker, F. R., Maindonald, J. H., & Jackson, P. J. (1996). Penetrometer measurement of apple and kiwifruit firmness: Operator and instrument differences. *Journal of the American Society for Horticultural Science, 121*(5), 927-936.

Harker, F. R., Marsh, K. B., Young, H., Murray, S. H., Gunson, F. A., & Walker, S. B. (2002). Sensory interpretation of instrumental measurements 2: Sweet and acid taste of apple fruit. *Postharvest Biology and Technology, 24*(3), 241-250.

Hazarika, S., Hebb, R. L., & Rizvi, S. S. H. (2018). Prediction of Ripening Stage of Cameo Apple Using Fourier-Transform Infrared Spectroscopy. *International Journal of Fruit Science, 18*(2), 188-198.

He, H., & Sun, D. (2015). Microbial evaluation of raw and processed food products by Visible/Infrared, Raman and Fluorescence spectroscopy. *Trends in Food Science & Technology, 46*(2, Part A), 199-210.

He, J., Rodriguez-Saona, L. E., & Giusti, M. M. (2007). Midinfrared spectroscopy for juice authentication - Rapid differentiation of commercial juices. *Journal of Agricultural and Food Chemistry, 55*(11), 4443-4452.

He, Y., Li, X., & Shao, Y. (2005). Quantitative analysis of the varieties of apple using near infrared spectroscopy by principal component analysis and BP model. *Australasian Joint Conference on Artificial Intelligence* (pp. 1053-1056): Springer.

He, Y., Li, X., & Shao, Y. (2007). Fast Discrimination of Apple Varieties Using Vis/NIR Spectroscopy. *International Journal of Food Properties, 10*(1), 9-18.

Hehnen, D., Hanrahan, I., Lewis, K., McFerson, J., & Blanke, M. (2012). Mechanical flower thinning improves fruit quality of apples and promotes consistent bearing. *Scientia Horticulturae, 134*, 241-244.

Herremans, E., Melado-Herreros, A., Defraeye, T., Verlinden, B., Hertog, M., Verboven, P., . . . Nicolaï, B. M. (2014). Comparison of X-ray CT and MRI of watercore disorder of different apple cultivars. *Postharvest Biology and Technology, 87*, 42-50.

Herremans, E., Verboven, P., Verlinden, B. E., Cantre, D., Abera, M., Wevers, M., & Nicolaï, B. M. (2015). Automatic analysis of the 3-D microstructure of fruit parenchyma tissue using X-ray micro-CT explains differences in aeration. *BMC Plant Biology, 15*(1), 264.

Hirst, P. M., Tustin, D. S., & Warrington, I. J. (1990). Fruit colour responses of 'granny smith' apple to variable light environments. *New Zealand Journal of Crop and Horticultural Science, 18*(4), 205-214.

Ho, Q. T., Verlinden, B. E., Verboven, P., Vandewalle, S., & Nicolaï, B. M. (2006). A permeation–diffusion–reaction model of gas transport in cellular tissue of plant materials. *Journal of Experimental Botany, 57*(15), 4215-4224.

Ho, T. K. (1995, August). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). IEEE.

Hoehn, E., Gasser, F., Guggenbühl, B., & Künsch, U. (2003). Efficacy of instrumental measurements for determination of minimum requirements of firmness, soluble solids, and acidity of several apple varieties in comparison to consumer expectations. *Postharvest Biology and Technology, 27*(1), 27-37.

Holland, J. K., Kemsley, E. K., & Wilson, R. H. (1998). Use of Fourier transform infrared spectroscopy and partial least squares regression for the detection of adulteration of strawberry purees. *Journal of the Science of Food and Agriculture, 76*(2), 263-269.

Horikawa, K., Hirama, T., Shimura, H., Jitsuyama, Y., & Suzuki, T. (2019). Visualization of soluble carbohydrate distribution in apple fruit flesh utilizing MALDI–TOF MS imaging. *Plant Science, 278*, 107-112.

Hu, W., Sun, D., & Blasco, J. (2017). Rapid monitoring 1-MCP-induced modulation of sugars accumulation in ripening 'Hayward' kiwifruit by Vis/NIR hyperspectral imaging. *Postharvest Biology and Technology, 125*, 168-180.

Huang, C., Cheng, C., & Lai, Y. (2020). Paper-based flexible surface enhanced Raman scattering platforms and their applications to food safety. *Trends in Food Science & Technology, 100*, 349-358.

Huang, H., Yu, H., Xu, H., & Ying, Y. (2008). Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *Journal of Food Engineering, 87*(3), 303-313.

Huang, M., & Lu, R. (2010). Apple mealiness detection using hyperspectral scattering technique. *Postharvest Biology and Technology, 58*(3), 168-175.

Hulme, A. C., Jones, J. D., & Wooltorton, L. S. C. (1963). The respiration climacteric in apple fruits. *Proceedings of the Royal Society B: Biological Sciences, 158*, 514-535.

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

Hussain, A., Pu, H., & Sun, D. (2018). Innovative nondestructive imaging techniques for ripening and maturity of fruits – A review of recent applications. *Trends in Food Science and Technology, 72*, 144-152.

I

Iglesias, I., & Alegre, S. (2009). The effects of reflective film on fruit color, quality, canopy light distribution, and profitability of 'Mondial Gala' apples. *HortTechnology, 19*(3), 488-498.

Iglesias, I., Echeverría, G., & Lopez, M. L. (2012). Fruit color development, anthocyanin content, standard quality, volatile compound emissions and consumer acceptability of several 'Fuji' apple strains. *Scientia Horticulturae, 137*, 138-147.

Iglesias, I., Echeverría, G., & Soria, Y. (2008). Differences in fruit colour development, anthocyanin content, fruit quality and consumer acceptability of eight 'Gala' apple strains. *Scientia Horticulturae, 119*(1), 32-40.

Iglesias, I., Graell, J., Echeverría, G., & Vendrell, M. (1999). Differences in fruit color development, anthocyanin content, yield and quality of seven 'Delicious' apple strains. *Fruit Varieties Journal, 53*(3), 133-145.

Ignat, T., Lurie, S., Nyasordzi, J., Ostrovsky, V., Egozi, H., Hoffman, A., Schmilovitch, Z. (2014). Forecast of Apple Internal Quality Indices at Harvest and During Storage by VIS-NIR Spectroscopy. *Food and Bioprocess Technology, 7*(10), 2951-2961.

Ilaslan, K., Boyaci, I. H., & Topcu, A. (2015). Rapid analysis of glucose, fructose and sucrose contents of commercial soft drinks using Raman spectroscopy. *Food Control, 48*, 56-61.

Iñón, F. A., Garrigues, S., & de la Guardia, M. (2006). Combination of mid- and near-infrared spectroscopy for the determination of the quality properties of beers. *Analytica Chimica Acta, 571*(2), 167-174.

Irudayaraj, J., & Tewari, J. (2003). Simultaneous monitoring of organic acids and sugars in fresh and processed apple juice by Fourier transform infrared-attenuated total reflection spectroscopy. *Applied Spectroscopy, 57*(12), 1599-1604.

Ishigaki, M., Meksiarun, P., Kitahama, Y., Zhang, L., Hashimoto, H., Genkawa, T., & Ozaki, Y. (2017). Unveiling the Aggregation of Lycopene in Vitro and in Vivo: UV–Vis, Resonance Raman, and Raman Imaging Studies. *The Journal of Physical Chemistry B, 121*(34), 8046-8057.

Iwanami, H. (2011). Breeding for Fruit Quality in Apple. In *Breeding for Fruit Quality* (pp. 173-200): John Wiley and Sons.

J

Jackson, M., & Mantsch, H. H. (1995). The use and misuse of FTIR spectroscopy in the determination of protein structure. *Critical Reviews in Biochemistry and Molecular Biology, 30*(2), 95-120.

Janotová, L., Čížková, H., Pivoňka, J., & Voldřich, M. (2011). Effect of processing of apple puree on patulin content. *Food Control, 22*(6), 977-981.

Janssen, S., Verboven, P., Nugraha, B., Wang, Z., Boone, M., Josipovic, I., & Nicolaï, B. M. (2020a). 3D pore structure analysis of intact 'Braeburn' apples using X-ray micro-CT. *Postharvest Biology and Technology, 159*, 111014.

Jarvis, M. C. (2011). Plant cell walls: Supramolecular assemblies. *Food Hydrocolloids*, 25(2), 257-262.

Jha, S. N., Rai, D. R., & Shrama, R. (2012). Physico-chemical quality parameters and overall quality index of apple during storage. *Journal of Food Science and Technology, 49*(5), 594-600.

Ji, W., Viscarra Rossel, R., & Shi, Z. (2015). Accounting for the effects of water and the environment on proximally sensed vis–NIR soil spectra and their calibrations. *European Journal of Soil Science, 66*(3), 555-565.

Jiang, Y., Sun, D.-W., Pu, H., & Wei, Q. (2018). Surface enhanced Raman spectroscopy (SERS): A novel reliable technique for rapid detection of common harmful chemical residues. *Trends in Food Science & Technology, 75*, 10-22.

Johnston, J. W., Hewett, E. W., Banks, N. H., Harker, F. R., & Hertog, M. L. A. T. M. (2001). Physical change in apple texture with fruit temperature: effects of cultivar and time in storage. *Postharvest Biology and Technology, 23*(1), 13-21.

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

Johnston, J. W., Hewett, E. W., Hertog, M. L. A. T. M., & Harker, F. R. (2002). Temperature and ethylene affect induction of rapid softening in 'Granny Smith' and 'Pacific Rose™' apple cultivars. *Postharvest Biology and Technology, 25*(3), 257-264.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(2065), 20150202.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260.

Juniper, B. E., & Mabberley, D. J. (2006). *The story of the apple*: Timber Press (OR).

K

Kačuráková, M., Wellner, N., Ebringerová, A., Hromádková, Z., Wilson, R. H., & Belton, P. S. (1999). Characterisation of xylan-type polysaccharides and associated cell wall components by FT-IR and FT-Raman spectroscopies. *Food Hydrocolloids, 13*(1), 35-41.

Kacurakova, M., Capek, P., Sasinkova, V., Wellner, N., & Ebringerova, A. (2000). FT-IR study of plant cell wall model compounds: pectic polysaccharides and hemicelluloses. *Carbohydrate Polymers, 43*(2), 195-203.

Kader, A. A. (2008). Flavor quality of fruits and vegetables. *Journal of the Science of Food and Agriculture, 88*(11), 1863-1868.

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.

Karoui, R., Mazerolles, G., & Dufour, É. (2003). Spectroscopic techniques coupled with chemometric tools for structure and texture determinations in dairy products. *International Dairy Journal, 13*(8), 607-620.

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

Kaur, H., Künnemeyer, R., & McGlone, A. (2020). Investigating aquaphotomics for temperature-independent prediction of soluble solids content of pure apple juice. *Journal of Near Infrared Spectroscopy, 28*(2), 103-112.

Kaur, H., Künnemeyer, R., & McGlone, A. (2017). Comparison of hand-held near infrared spectrophotometers for fruit dry matter assessment. *Journal of Near Infrared Spectroscopy, 25*(4), 267-277.

Kaur, R., Gul, K., & Singh, A. K. (2016). Nutritional impact of ohmic heating on fruits and vegetables—A review. *Cogent Food & Agriculture, 2*(1), 1159000.

Kawano, S. (1994). Non-destructive NIR quality evaluation of fruits and vegetables in Japan. *NIR News, 5*(6), 10-12.

Kays, S. J. (1999). Preharvest factors affecting appearance. *Postharvest Biology and Technology, 15*(3), 233-247.

Kazarian, S. G., & Chan, K. L. A. (2013). ATR-FTIR spectroscopic imaging: recent advances and applications to biological systems. *Analyst, 138*(7), 1940-1951.

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

Kelly, J. F. D., & Downey, G. (2005). Detection of sugar adulterants in apple juice using Fourier transform infrared spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry, 53*(9), 3281-3286.

Kemsley, E. K., Holland, J. K., Defernez, M., & Wilson, R. H. (1996). Detection of adulteration of raspberry purees using infrared spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry, 44*(12), 3864-3870.

Khan, M. I. H., Nagy, S. A., & Karim, M. A. (2018). Transport of cellular water during drying: An understanding of cell rupturing mechanism in apple tissue. *Food Research International, 105*, 772-781.

Khanizadeh, S., Tsao, R., Rekika, D., Yang, R., Charles, M. T., & Vasantha Rupasinghe, H. P. (2008). Polyphenol composition and total antioxidant capacity of selected apple genotypes for

processing. *Journal of Food Composition and Analysis, 21*(5), 396-401.

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.

Kim, A. N., Lee, K. Y., Rahman, M. S., Kim, H. J., Kerr, W. L., & Choi, S. G. (2021). Thermal treatment of apple puree under oxygen-free condition: Effect on phenolic compounds, ascorbic acid, antioxidant activities, color, and enzyme activities. *Food Bioscience, 39*.

Kim, K. B., Lee, S., Kim, M. S., & Cho, B. K. (2009). Determination of apple firmness by nondestructive ultrasonic measurement. *Postharvest Biology and Technology, 52*(1), 44-48.

King, G. J., Maliepaard, C., Lynn, J. R., Alston, F. H., Durel, C. E., Evans, K. M., . . . Verhaegh, J. (2000). Quantitative genetic analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (Malus pumila Mill.). *Theoretical and Applied Genetics, 100*(7), 1074-1084.

Kingston, C. M. (1992). Maturity indices for apple and pear. *Horticultural Reviews, 13*, 407-432.

Knorr, D., & Augustin, M. A. (2021). Food processing needs, advantages and misconceptions. *Trends in Food Science & Technology, 108*, 103-110.

Knorr, D., Augustin, M. A., & Tiwari, B. (2020). Advancing the role of food processing for improved integration in sustainable food chains. *Frontiers in nutrition, 7*.

Konopacka, D., & Plocharski, W. J. (2004). Effect of storage conditions on the relationship between apple firmness and texture acceptability. *Postharvest Biology and Technology, 32*(2), 205-211.

Korban, S., & Skirvin, R. (1984). Nomenclature of the cultivated apple. *HortScience, 19*(2), 177-180.

Kouassi, A. B., Durel, C.-E., Costa, F., Tartarini, S., van de Weg, E., Evans, K., . . . Laurens, F. (2009). Estimation of genetic parameters and prediction of breeding values for apple fruit-quality traits using pedigreed plant material in Europe. *Tree Genetics & Genomes, 5*(4), 659-672.

Krall, S. M., & McFeeters, R. F. (1998). Pectin hydrolysis: effect of temperature, degree of methylation, pH, and calcium on hydrolysis rates. *Journal of Agricultural and Food Chemistry, 46*(4), 1311-1315.

Krishnaprakash, M. S., Aravindaprasad, B., Krishnaprasad, C. A., Narasimham, P., Ananthakrishna, S. M., Dhanaraj, S., & Govindarajan, V. S. (1983). Effect of apple position on the tree on maturity and quality. *Journal of Horticultural Science, 58*(1), 31-36.

Kunzek, H., Opel, H., & Senge, B. (1997). Rheological examination of material with cellular structure II. Creep and oscillation measurements of apple material with cellular structure. *Zeitschrift für Lebensmitteluntersuchung und-Forschung A, 205*(3), 193-203.

Kuhn, M. (2015). Caret: classification and regression training. R package version 6.0-85. https://CRAN.R-project.org/package=caret

Kuhn, M., Weston, S., Keefer, C., Coulter, N., & Quinlan, R. (2014). Cubist: Rule- And Instance-Based Regression Modeling. R package version 0.2.3. https://CRAN.R-project.org/package=Cubist.

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *J Stat Softw, 36*(11), 1-13.

Kurz, C., Leitenberger, M., Carle, R., & Schieber, A. (2010). Evaluation of fruit authenticity and determination of the fruit content of fruit products using FT-NIR spectroscopy of cell wall components. *Food Chemistry*, 119(2), 806-812.

Kviklys, D., Kvikliene, N., Bite, A., Lepsis, J., Univer, T., Univer, N., . . . Buskiene, L. (2012). Baltic fruit rootstock studies: Evaluation of 12 apple rootstocks in North-East Europe. *Horticultural*

Science, 39(1), 1-7.

Kweon, H., Kang, I., Kim, M., Lee, J., Moon, Y., Choi, C., Watkins, C. B. (2013). Fruit maturity, controlled atmosphere delays and storage temperature affect fruit quality and incidence of storage disorders of 'Fuji' apples. *Scientia Horticulturae, 157*, 60-64.

Kyomugasho, C., Christiaens, S., Shpigelman, A., Van Loey, A. M., & Hendrickx, M. E. (2015). FT-IR spectroscopy, a reliable method for routine analysis of the degree of methylesterification of pectin in different fruit-and vegetable-based matrices. *Food Chemistry, 176*, 82-90.

L

Labaky, P., Dahdouh, L., Ricci, J., Wisniewski, C., Pallet, D., Louka, N., & Grosmaire, L. (2021). Impact of ripening on the physical properties of mango purees and application of simultaneous rheometry and in situ FTIR spectroscopy for rapid identification of biochemical and rheological changes. *Journal of Food Engineering, 300*, 110507.

Lafeuille, J. L., Frégière-Salomon, A., Michelet, A., & Henry, K. L. (2020). A Rapid Non-Targeted Method for Detecting the Adulteration of Black Pepper with a Broad Range of Endogenous and Exogenous Material at Economically Motivating Levels Using Micro-ATR-FT-MIR Imaging. *Journal of Agricultural and Food Chemistry, 68*(1), 390-401.

Lammertyn, J., Nicolaï, B., Ooms, K., De Smedt, V., & De Baerdemaeker, J. (1998). Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR-spectroscopy. *Transactions of the American Society of Agricultural Engineers, 41*(4), 1089-1094.

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 Biology and Technology, 18*, 121-132.

Lancaster, J. E., Grant, J. E., Lister, C. E., & Taylor, M. C. (1994). Skin color in apples - Influence of copigmentation and plastid pigments on shade and darkness of red color in five genotypes. *Journal of the American Society for Horticultural Science, 119*(1), 63-69.

Latimer, L. P. (1931). Further observations on factors affecting fruit setting of the McIntosh apple in New Hampshire. *Proceedings of the American Society for Horticultural Science, 28*, 87-92.

Latimer, L. P. (1933). Pollination and fruit setting in the apple. *New Hampshire Agricultural Experiment Station*, 274.

Latimer, L. P. (1937). Self- and cross-pollination in the McIntosh apple and some of it's hybrids. *Proceedings of the American Society for Horticultural Science, 34*, 19-21.

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

Le Bourvellec, C., Bureau, S., Renard, C., Plenet, D., Gautier, H., Touloumet, L., Simon, S. (2015). Cultivar and Year Rather than Agricultural Practices Affect Primary and Secondary Metabolites in Apple Fruit. *Plos One, 10*(11).

Le Dréau, Y., Dupuy, N., Artaud, J., Ollivier, D., & Kister, J. (2009). Infrared study of aging of edible oils by oxidative spectroscopic index and MCR-ALS chemometric method. *Talanta, 77*(5), 1748-1756.

Lee, J., Mattheis, J. P., & Rudell, D. R. (2019). High storage humidity affects fruit quality attributes and incidence of fruit cracking in cold-stored 'Royal Gala'apples. *HortScience, 54*(1), 149-154.

Lefébure, É., Troch, T., Noutfia, Y., Colinet, F., Gérard, A., Dehareng, F., Sindic, M. (2021). Is it possible

to predict milk processing into butter using infrared spectroscopy? *Biotechnologie, Agronomie, Société et Environnement, 25*(1), 21-31.

Lehman-Salada, L. (1996). Instrument and operator effects on apple firmness readings. *HortScience, 31*(6), 994-997.

Leiva-Valenzuela, G. A., Lu, R., & Aguilera, J. M. (2013). Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging. *Journal of Food Engineering, 115*(1), 91-98.

Leopold, L. F., Leopold, N., Diehl, H. A., & Socaciu, C. (2011). Quantification of carbohydrates in fruit juices using FTIR spectroscopy and multivariate analysis. *Journal of Spectroscopy, 26*(2), 93-104.

León, K., Mery, D., Pedreschi, F., & León, J. (2006). Color measurement in L*a*b* units from RGB digital images. *Food Research International, 39*(10), 1084-1091.

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

Leverrier, C., Almeida, G., Espinosa-Mu noz, L., & Cuvelier, G. (2016). Influence of Particle Size and Concentration on Rheological Behaviour of Reconstituted Apple Purees. *Food Biophysics, 11*(3), 235-247.

Li, C., Li, L., Wu, Y., Lu, M., Yang, Y., & Li, L. (2018). Apple variety identification using near-infrared spectroscopy. *Journal of Spectroscopy, 2018*.

Li, D., Zhu, Z., & Sun, D.-W. (2020). Visualization of the in situ distribution of contents and hydrogen bonding states of cellular level water in apple tissues by confocal Raman microscopy. *Analyst, 145*(3), 897-907.

Li, X., Huang, J., Xiong, Y., Zhou, J., Tan, X., & Zhang, B. (2018). Determination of soluble solid content in multi-origin 'Fuji' apples by using FT-NIR spectroscopy and an origin discriminant strategy. *Computers and Electronics in Agriculture, 155*, 23-31.

Li, X., Li, R., Wang, M., Liu, Y., Zhang, B., & Zhou, J. (2017). Hyperspectral imaging and their applications in the nondestructive quality assessment of fruits and vegetables. In *Hyperspectral imaging in agriculture, food and environment*: IntechOpen.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *Sensors (Switzerland), 18*(8).

Liaw, A., Wiener, M., 2002. Classification and regression by Random Forest. R news 2, 18-22.

Liland, K. H., Rukke, E.-O., Olsen, E. F., & Isaksson, T. (2011). Customized baseline correction. *Chemometrics and Intelligent Laboratory Systems, 109*(1), 51-56.

Link, H. (2000). Significance of flower and fruit thinning on fruit quality. *Plant Growth Regulation, 31*(1), 17-26.

Liu, C., Liu, W., Lu, X., Chen, W., Yang, J., & Zheng, L. (2016). Potential of multispectral imaging for real-time determination of colour change and moisture distribution in carrot slices during hot air dehydration. *Food Chemistry, 195*, 110-116.

Liu, D., Zeng, X.-A., & Sun, D.-W. (2015). Recent Developments and Applications of Hyperspectral Imaging for Quality Evaluation of Agricultural Products: A Review. *Critical Reviews in Food Science and Nutrition, 55*(12), 1744-1757.

Liu, F., He, Y., & Wang, L. (2008). Determination of effective wavelengths for discrimination of fruit vinegars using near infrared spectroscopy and multivariate analysis. *Analytica Chimica Acta, 615*(1), 10-17.

Liu, X., Renard, C. M. G. C., Rolland-Sabaté, A., Bureau, S., & Le Bourvellec, C. (2020). Modification of apple, beet and kiwifruit cell walls by boiling in acid conditions: Common and specific responses. *Food Hydrocolloids*, 106266.

Liu, X., Renard, C. M. G. C., Rolland-Sabaté, A., Bureau, S., Le Bourvellec, C. (2019). Modification of cell walls of apple, red beet and kiwifruit by heating in acid conditions: common and specific responses. 15th Cell wall meeting, Cambridge, UK.

Liu, X., Renard, C. M. G. C., Bureau, S., & Le Bourvellec, C. (2021). Revisiting the contribution of ATR-FTIR spectroscopy to characterize plant cell wall polysaccharides. *Carbohydrate Polymers, 262*.

Liu, Y., Cao, M., & Liu, G. (2019). 17 - Texture analyzers for food quality evaluation. In J. Zhong & X. Wang (Eds.), *Evaluation Technologies for Food Quality* (pp. 441-463): Woodhead Publishing.

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

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

Liu, Y., Ying, Y., & Jiang, H. (2006). Rapid determination of maturity in apple using outlier detection and calibration model optimization. *Transactions of the ASABE, 49*(1), 91-95.

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

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

Lu, H., Ying, Y., Fu, X., Yu, H., Liu, Y., & Tian, H. (2007). Estimation of soluble solids content of apple fresh juice by FTNIR spectroscopy. *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis, 27*(3), 494-498.

Lu, R. (2004). Multispectral imaging for predicting firmness and soluble solids content of apple fruit. *Postharvest Biology and Technology, 31*(2), 147-157.

Lu, R., Guyer, D. E., & Beaudry, R. M. (2000). Determination of firmness and sugar content of apples using near-infrared diffuse reflectance. *Journal of Texture Studies, 31*(6), 615-630.

Lu, X., Wang, J., Al-Qadiri, H. M., Ross, C. F., Powers, J. R., Tang, J., & Rasco, B. A. (2011). Determination of total phenolic content and antioxidant capacity of onion (Allium cepa) and shallot (Allium oschaninii) using infrared spectroscopy. *Food Chemistry, 129*(2), 637-644.

Lu, Y., Saeys, W., Kim, M., Peng, Y., & Lu, R. (2020). Hyperspectral imaging technology for quality and safety evaluation of horticultural products: A review and celebration of the past 20-year progress. *Postharvest Biology and Technology, 170*.

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

M

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.

Ma, T., Xia, Y., Inagaki, T., & Tsuchikawa, S. (2021). Non-destructive and fast method of mapping the distribution of the soluble solids content and pH in kiwifruit using object rotation near-infrared hyperspectral imaging approach. *Postharvest Biology and Technology, 174*, 111440.

Ma, Y., Guo, J., Guo, Z., Huang, H., Shi, Y., & Zhou, J. (2020). Origin Tracing of Red Fuji Apple Based on Near Infrared Transmission Spectrum and Various Dimension Reduction Methods. *Modern Food Science and Technology, 36*(6), 303-309.

Magness, J. R., & Taylor, G. F. (1925).

Magwaza, L. S., & Opara, U. L. (2015). Analytical methods for determination of sugars and sweetness of horticultural products-A review. *Scientia Horticulturae, 184*, 179-192.

Mathlouthi, M., & Luu, D. V. (1980). Laser-raman spectra of d-fructose in aqueous solution. *Carbohydrate Research, 78*(2), 225-233.

Market Research Future (2019). Fruit Puree Market information: by fruit type (apple puree, banana puree, plum puree, strawberry puree, and others), by application (baby food, bakery, beverages, and others), by category (conventional and organic) and by Region - Global Forecast till 2023. Accessed March 2019, from https://www.marketresearchfuture.com/reports/fruit-puree-market-5281.

Martinsen, P., & Schaare, P. (1998). Measuring soluble solids distribution in kiwifruit using near-infrared imaging spectroscopy. *Postharvest Biology and Technology, 14*(3), 271-281.

Masanet, E. (2008). Energy efficiency improvement and cost saving opportunities for the fruit and vegetable processing industry. An Energy Star Guide for Energy and Plant Managers.

Mastersizer. (2007). *User manual. Malvern Instruments*.

Matsuoka, K. (2019). Anthocyanins in apple fruit and their regulation for health benefits. *Flavonoid–a Coloring Model for Cheering Up Life*.

McCann, M. C., Chen, L., Roberts, K., Kemsley, E. K., Sene, C., Carpita, N. C., Wilson, R. H. (1997). Infrared microspectroscopy: Sampling heterogeneity in plant cell wall composition and architecture. *Physiologia Plantarum, 100*(3), 729-738.

McCann, M. C., Hammouri, M., Wilson, R., Belton, P., & Roberts, K. (1992). Fourier transform infrared microspectroscopy is a new way to look at plant cell walls. Plant Physiology, 100(4), 1940-1947.

McGlone, V. A., Clark, C. J., & Jordan, R. B. (2007). Comparing density and VNIR methods for predicting quality parameters of yellow-fleshed kiwifruit (Actinidia chinensis). *Postharvest Biology and Technology, 46*(1), 1-9.

McGlone, V. A., Jordan, R. B., & Martinsen, P. J. (2002). Vis/NIR estimation at harvest of pre- and post-storage quality indices for 'Royal Gala' apple. *Postharvest Biology and Technology, 25*(2), 135-144.

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

McGlone, V. A., & Kawano, S. (1998). Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biology and Technology, 13*(2), 131-141.

Mditshwa, A., Fawole, O. A., & Opara, U. L. (2018). Recent developments on dynamic controlled atmosphere storage of apples—A review. *Food Packaging and Shelf Life, 16*, 59-68.

Mebatsion, H. K., Verboven, P., Ho, Q. T., Verlinden, B. E., & Nicolaï, B. M. (2008). Modelling fruit (micro)structures, why and how? *Trends in Food Science & Technology, 19*(2), 59-66.

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

Meland, M. (2009). Effects of different crop loads and thinning times on yield, fruit quality, and return bloom in Malus × domestica Borkh. 'Elstar'. *The Journal of Horticultural Science and Biotechnology, 84*(6), 117-121.

Mendlein, A., Szkudlarek, C., & Goodpaster, J. V. (2013). Chemometrics. In J. A. Siegel, P. J. Saukko & M. M. Houck (Eds.), *Encyclopedia of Forensic Sciences (Second Edition)* (pp. 646-651). Waltham: Academic Press.

Mevik, B.-H., Wehrens, R., & Liland, K. H. (2019). pls: Partial least squares and principal component regression. R package version 2.7-2. https://CRAN.R-project.org/package=pls.

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*, 67-81.

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 Biology and Technology. 62*, 149-160.

Mendoza, F., Lu, R., & Cen, H. (2012). Comparison and fusion of four nondestructive sensors for predicting apple fruit firmness and soluble solids content. *Postharvest Biology and Technology, 73*, 89-98.

Mendoza, F., Lu, R., & Cen, H. (2014). Grading of apples based on firmness and soluble solids content using Vis/SWNIR spectroscopy and spectral scattering techniques. *Journal of Food Engineering, 125*(1), 59-68.

Mendoza, F., Verboven, P., Ho, Q. T., Kerckhofs, G., Wevers, M., & Nicolaï, B. (2010). Multifractal properties of pore-size distribution in apple tissue using X-ray imaging. *Journal of Food Engineering, 99*(2), 206-215.

Mendoza, F., Verboven, P., Mebatsion, H. K., Kerckhofs, G., Wevers, M., & Nicolaï, B. (2007). Three-dimensional pore space quantification of apple tissue using X-ray computed microtomography. *Planta, 226*(3), 559-570.

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

Merzlyak, M. N., Solovchenko, A. E., & Gitelson, A. A. (2003). Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. *Postharvest Biology and Technology, 27*(2), 197-211.

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

Milić, B., Tarlanović, J., Keserović, Z., Zorić, L., Blagojević, B., & Magazin, N. (2017). The Growth of Apple Central Fruits as Affected by Thinning with NAA, BA and Naphthenic Acids. *Erwerbs-Obstbau, 59*(3), 185-193.

Mills, T., Behboudian, M., & Clothier, B. (1996). Preharvest and storage quality of 'Braeburn'apple fruit grown under water deficit conditions. *New Zealand journal of crop and horticultural science, 24*(2), 159-166.

Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends in Analytical Chemistry, 132*, 116045.

Mishra, P., Roger, J. M., Jouan-Rimbaud-Bouveresse, D., Biancolillo, A., Marini, F., Nordon, A., & Rutledge, D. N. (2021). Recent trends in multi-block data analysis in chemometrics for multi-source data integration. *TrAC - Trends in Analytical Chemistry, 137*.

Mitchell, T. M. (1997). Machine learning.

Mo, C., Kim, M. S., Kim, G., Lim, J., Delwiche, S. R., Chao, K., . . . Cho, B.-K. (2017). Spatial assessment of soluble solid contents on apple slices using hyperspectral imaging. *Biosystems Engineering, 159*, 10-21.

Mohamed, G. F., Shaheen, M. S., Khalil, S. K., Hussein, A., Kamil, M. M. J. N., & Science. (2011). Application of FT-IR spectroscopy for rapid and simultaneous quality determination of some fruit products. *9*(11), 21-31.

Mohd Ali, M., Hashim, N., Aziz, S. A., & Lasekan, O. (2020). Emerging non-destructive thermal imaging technique coupled with chemometrics on quality and safety inspection in food and agriculture. *Trends in Food Science & Technology, 105*, 176-185.

Mohr, C. L., & Mohr, B. C. (2000). US6643599-B6643591.

Mollazade, K., Omid, M., Tab, F. A., Mohtasebi, S. S., & Zude, M. (2012, July). Spatial mapping of moisture content in tomato fruits using hyperspectral imaging and artificial neural networks. In *Proceedings of the CIGR-AgEng2012: IV International workshop on Computer Image Analysis in Agriculture, Valencia, Spain* (pp. 8-12).

Moreira, L. F. P. P., Buffon, E., de Sá, A. C., & Stradiotto, N. R. (2021). Fructose determination in fruit juices using an electrosynthesized molecularly imprinted polymer on reduced graphene oxide modified electrode. *Food Chemistry, 352*.

Mpelasoka, B., Behboudian, M., & Mills, T. (2001). Water relations, photosynthesis, growth, yield and fruit size of 'Braeburn'apple: Responses to deficit irrigation and to crop load. *The Journal of Horticultural Science and Biotechnology, 76*(2), 150-156.

Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J. B., . . . Fujiyoshi, Y. (2000). Structural determinants of water permeation through aquaporin-1. *Nature, 407*(6804), 599-605.

Musacchi, S., & Serra, S. (2018). Apple fruit quality: Overview on pre-harvest factors. *Scientia Horticulturae, 234*, 409-430.

N

Naik, S., & Patel, B. (2017). Machine vision based fruit classification and grading-a review. *International Journal of Computer Applications, 170*(9), 22-34.

Nargis, H., Nawaz, H., Ditta, A., Mahmood, T., Majeed, M., Rashid, N., Muddassar, M., Bhatti, H., Saleem, M., & Jilani, K. (2019). Raman spectroscopy of blood plasma samples from breast cancer patients at different stages. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 222*, 117210.

Nawrocka, A., Miś, A., & Szymańska-Chargot, M. (2016). Characteristics of Relationships Between Structure of Gluten Proteins and Dough Rheology – Influence of Dietary Fibres Studied by FT-Raman Spectroscopy. *Food Biophysics, 11*(1), 81-90.

Ncama, K., Opara, U. L., Tesfay, S. Z., Fawole, O. A., & Magwaza, L. S. (2017). Application of Vis/NIR spectroscopy for predicting sweetness and flavour parameters of 'Valencia' orange (Citrus

sinensis) and ‘Star Ruby’ grapefruit (Citrus x paradisi Macfad). *Journal of Food Engineering, 193*, 86-94.

Ngarize, S., Adams, A., & Howell, N. K. (2004). Studies on egg albumen and whey protein interactions by FT-Raman spectroscopy and rheology. *Food Hydrocolloids, 18*(1), 49-59.

Nickerson, D. (1957). Horticultural Colour Chart names with Munsell key. *Journal of the Optical Society of America, 47*, 619-621.

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.

Noble, W. S. (2006). What is a support vector machine?. *Nature biotechnology*, 24(12), 1565-1567.

Noda, I. (1993). Generalized two-dimensional correlation method applicable to infrared, Raman, and other types of spectroscopy. *Applied Spectroscopy, 47*(9), 1329-1336.

Nordey, T., Davrieux, F., & Léchaudel, M. (2019). Predictions of fruit shelf life and quality after ripening: Are quality traits measured at harvest reliable indicators? *Postharvest Biology and Technology, 153*, 52-60.

Nogueira, J. N., McLellan, M. R., & Anantheswaran, R. C. (1985). Effect of fruit firmness and processing parameters on the particle size distribution in applesauce of two cultivars. *Journal of food science*, 50(3), 744-746.

Nugraha, B., Verboven, P., Janssen, S., Wang, Z., & Nicolaï, B. M. (2019). Non-destructive porosity mapping of fruit and vegetables using X-ray CT. *Postharvest Biology and Technology, 150*, 80-88.

Nyasordzi, J., Friedman, H., Schmilovitch, Z., Ignat, T., Weksler, A., Rot, I., & Lurie, S. (2013). Utilizing the IAD index to determine internal quality attributes of apples at harvest and after storage. *Postharvest Biology and Technology, 77*, 80-86.

Nybom, N. (1959). On the inheritance of acidity in cultivated apples. *Hereditas, 45*(2-3), 332-350.

O

Oliveri, P., Malegori, C., & Casale, M. (2020). 2 - Chemometrics: multivariate analysis of chemical data. In Y. Pico (Ed.), *Chemical Analysis of Food (Second Edition)* (pp. 33-76): Academic Press.

Oliveira-Folador, G., de Oliveira Bicudo, M., de Andrade, E. F., Renard, C. M. G. C., Bureau, S., & de Castilhos, F. (2018). Quality traits prediction of the passion fruit pulp using NIR and MIR spectroscopy. *LWT - Food Science and Technology, 95*, 172-178.

Opatová, H., Voldřich, M., Dobiáš, J., & Čurda, D. (1992). Quality changes during the storage of apple puree. *Food / Nahrung, 36*(2), 129-134.

Osborne, B. G. (2006). Near-infrared spectroscopy in food analysis. *Encyclopedia of analytical chemistry: applications, theory and instrumentation*.

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.

O'Sullivan, M. (2016). *A handbook for sensory and consumer-driven new product development: Innovative technologies for the food and beverage industry* (pp 156-159): Woodhead Publishing.

Özbalci, B., Boyaci, İ. H., Topcu, A., Kadılar, C., & Tamer, U. (2013). Rapid analysis of sugars in honey by processing Raman spectrum using chemometric methods and artificial neural networks. *Food Chemistry, 136*(3), 1444-1452.

P

Palmer, J. (2014). The future role of crop physiologists, a personal view. In K. Theron (Ed.), *Acta Horticulturae* (Vol. 1058, pp. 209-220): International Society for Horticultural Science.

Palmer, J., Diack, R., Johnston, J., & Boldingh, H. (2013). Manipulation of fruit dry matter accumulation and fruit size in 'Scifresh' apple through alteration of the carbon supply, and its relationship with apoplastic sugar composition. *Journal of Horticultural Science and Biotechnology, 88*(4), 483-489.

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

Parmley, K. A., Higgins, R. H., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2019). Machine Learning Approach for Prescriptive Plant Breeding. *Scientific Reports, 9*(1), 17132.

Pan, T., Pu, H., & Sun, D. (2017). Insights into the changes in chemical compositions of the cell wall of pear fruit infected by Alternaria alternata with confocal Raman microspectroscopy. *Postharvest Biology and Technology, 132*, 119-129.

Panford, J. (1987). Application of near-infrared reflectance spectroscopy in North America.

Park, B., Abbott, J. A., Lee, K. J., Choi, C. H., & Choi, K. H. (2003). Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of Delicious and Gala apples. *Transactions of the American Society of Agricultural Engineers, 46*(6), 1721-1731.

Pathmanaban, P., Gnanavel, B. K., & Anandan, S. S. (2019). Recent application of imaging techniques for fruit quality assessment. *Trends in Food Science & Technology, 94*, 32-42.

Paull, R. (1999). Effect of temperature and relative humidity on fresh commodity quality. *Postharvest Biology and Technology, 15*(3), 263-277.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2*(11), 559-572.

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*(1), 114-118.

Peirs, A., Lammertyn, J., Ooms, K., & Nicola, B. M. (2001). Prediction of the optimal picking date of different apple cultivars by means of VIS/NIR-spectroscopy. *Postharvest Biology and Technology, 21*(2), 189-199.

Peirs, A., Scheerlinck, N., De Baerdemaeker, J., & Nicolaï, B. M. (2003a). Starch Index Determination of Apple Fruit by Means of a Hyperspectral near Infrared Reflectance Imaging System. *Journal of Near Infrared Spectroscopy, 11*(5), 379-389.

Peirs, A., Scheerlinck, N., Touchant, K., & Nicolaï, B. M. (2002). Comparison of Fourier transform and dispersive near-infrared reflectance spectroscopy for apple quality measurements. *Biosystems engineering, 81*(3), 305-311.

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

Peirs, A., Scheerlinck, N., Nicolaï, B.M., (2003c). Temperature compensation for near infrared reflectance measurement of apple fruit soluble solids contents. *Postharvest Biology and Technology, 30*, 233-248.

Pelacci, M., Malavasi, M., Cattani, L., Gozzi, M., Tedeschi, F., Vignali, G., Gervais, S. (2021). Impact of indirect and ohmic heating sterilization processes on quality parameters of apple puree: Application in a real industrial line. *Journal of Physics: Conference Series* (1 ed., Vol. 1868).

Pelletier, M. J. (2003). Quantitative Analysis Using Raman Spectrometry. *Applied Spectroscopy, 57*(1), 20A-42A.

Pelliccia D. (2018). Classification of NIR spectra by Linear Discriminant Analysis in Python. https://nirpyresearch.com/classification-nir-spectra-linear-discriminant-analysis-python/

Peng, B., Ge, N., Cui, L., & Zhao, H. (2016). Monitoring of alcohol strength and titratable acidity of apple wine during fermentation using near-infrared spectroscopy. *LWT - Food Science and Technology, 66*, 86-92.

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 Chemistry. 293*, 358-367.

Peng, Y., Lu, R., 2008. Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. *Postharvest Biology and Technology, 48*, 52-62.

Peng, Y., & Lu, R. (2005). Modeling multispectral scattering profiles for prediction of apple fruit firmness. *Transactions of the American Society of Agricultural Engineers, 48*(1), 235-242.

Perring, M. A. (1989). Changes in dry matter concentration gradients in stored apples. *Journal of the Science of Food and Agriculture, 46*(4), 439-449.

Philippe, G., Gaillard, C., Petit, J., Geneix, N., Dalgalarrondo, M., Bres, C., . . . Bakan, B. (2016). Ester Cross-Link Profiling of the Cutin Polymer of Wild-Type and Cutin Synthase Tomato Mutants Highlights Different Mechanisms of Polymerization. *Plant Physiology, 170*(2), 807-820.

Picouet, P. A., Gou, P., Hyypiö, R., & Castellari, M. (2018). Implementation of NIR technology for at-line rapid detection of sunflower oil adulterated with mineral oil. *Journal of Food Engineering, 230*, 18-27.

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.

Piñeiro, V., Arias, J., Dürr, J., Elverdin, P., Ibáñez, A. M., Kinengyere, A., Torero, M. (2020). A scoping review on incentives for adoption of sustainable agricultural practices and their outcomes. *Nature Sustainability, 3*(10), 809-820.

Pissard, A., Baeten, V., Dardenne, P., Dupont, P., & Lateur, M. (2018). Use of NIR spectroscopy on fresh apples to determine the phenolic compounds and dry matter content in peel and flesh. *Biotechnology, Agronomy and Society and Environment, 22*(1), 3-12.

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. *Biotechnology, Agronomy and Society and Environment*.

Pissard, A., Fernandez Pierna, J. A., Baeten, V., Sinnaeve, G., Lognay, G., Mouteau, A., Lateur, M. (2013). Non-destructive measurement of vitamin C, total polyphenol and sugar content in apples using near-infrared spectroscopy. *Journal of the Science of Food and Agriculture, 93*(2), 238-244.

Pissard, A., Marques, E. J. N., Dardenne, P., Lateur, M., Pasquini, C., Pimentel, M. F., Baeten, V. (2021). Evaluation of a handheld ultra-compact NIR spectrometer for rapid and non-destructive determination of apple fruit quality. *Postharvest Biology and Technology, 172*.

Pistorius, A. M. A. (1996). *Biochemical applications of FT-IR spectroscopy*: [Sl: sn].

Porep, J. U., Kammerer, D. R., & Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science and Technology, 46*(2), 211-230.

Pompeu, D. R., Larondelle, Y., Rogez, H., Abbas, O., Pierna, J. A. F., & Baeten, V. (2018).

Characterization and discrimination of phenolic compounds using Fourier transform Raman spectroscopy and chemometric tools. *Biotechnology, Agronomy and Society and Environment.*

Pourdarbani, R., Sabzi, S., Kalantari, D., Karimzadeh, R., Ilbeygi, E., & Arribas, J. I. (2020a). Automatic non-destructive video estimation of maturation levels in Fuji apple (Malus Malus pumila) fruit in orchard based on colour (Vis) and spectral (NIR) data. *Biosystems engineering, 195*, 136-151.

Pourdarbani, R., Sabzi, S., Kalantari, D., Paliwal, J., Benmouna, B., García-Mateos, G., & Molina-Martínez, J. M. (2020b). Estimation of different ripening stages of Fuji apples using image processing and spectroscopy based on the majority voting method. *Computers and Electronics in Agriculture, 176*.

Prange, R. K., Delong, J. M., & Harrison, P. A. (2001). Storage humidity and post-storage handling temperature affect bruising and other apple quality characteristics. *Acta Horticulturae* (Vol. 553, pp. 717-720): International Society for Horticultural Science.

Proctor, J. T. A. (1974). Color stimulation in attached apples with supplymentary light. *Canadian Journal of Plant Science, 54*(3), 499-503.

Prieto, N., Roehe, R., Lavín, P., Batten, G., & Andrés, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review. *Meat Science, 83*(2), 175-186.

Pu, Y., Feng, Y., & Sun, D. (2015). Recent Progress of Hyperspectral Imaging on Quality and Safety Inspection of Fruits and Vegetables: A Review. *Comprehensive Reviews in Food Science and Food Safety, 14*(2), 176-188.

Pu, Y., & Sun, D. (2015). Vis–NIR hyperspectral imaging in visualizing moisture distribution of mango slices during microwave-vacuum drying. *Food Chemistry, 188*, 271-278.

Pu, Y., & Sun, D. (2017). Combined hot-air and microwave-vacuum drying for improving drying uniformity of mango slices based on hyperspectral imaging visualisation of moisture content distribution. *Biosystems Engineering, 156*, 108-119.

Q

Qi, J., & Shih, W. (2014). Performance of line-scan Raman microscopy for high-throughput chemical imaging of cell population. *Applied Optics, 53*(13), 2881-2885.

Qiao, S., Tian, Y., Gu, W., He, K., Yao, P., Song, S., Zhang, F. (2019). Research on simultaneous detection of SSC and FI of blueberry based on hyperspectral imaging combined MS-SPA. *Engineering in Agriculture, Environment and Food, 12*(4), 540-547.

Qin, J., Chao, K., Cho, B. K., Peng, Y., & Kim, M. S. (2014). High-throughput Raman chemical imaging for rapid evaluation of food safety and quality. *Transactions of the ASABE, 57*(6), 1783-1792.

Qin, J., Chao, K., & Kim, M. S. (2011). Investigation of Raman chemical imaging for detection of lycopene changes in tomatoes during postharvest ripening. *Journal of Food Engineering, 107*(3), 277-288.

Qin, J., Chao, K., & Kim, M. S. (2012). Nondestructive evaluation of internal maturity of tomatoes using spatially offset Raman spectroscopy. *Postharvest Biology and Technology, 71*, 21-31.

Qin, J., Chao, K., & S. Kim, M. (2010). Raman Chemical Imaging System for Food Safety and Quality Inspection. *Transactions of the ASABE, 53*(6), 1873-1882.

Qin, J., Kim, M. S., Chao, K., Dhakal, S., Cho, B.-K., Lohumi, S., Huang, M. (2019). Advances in Raman spectroscopy and imaging techniques for quality and safety inspection of horticultural products. *Postharvest Biology and Technology, 149*, 101-117.

Qin, J., Kim, M. S., Chao, K., Schmidt, W. F., Cho, B.-K., & Delwiche, S. R. (2017). Line-scan Raman imaging and spectroscopy platform for surface and subsurface evaluation of food safety and quality. *Journal of Food Engineering, 198*, 17-27.

Qing, Z., Ji, B., & Zude, M. (2007). Wavelength selection for predicting physicochemical properties of apple fruit based on near-infrared spectroscopy. *Journal of Food Quality, 30*(4), 511-526.

Queiroz, C., Lopes, M. L. M., Fialho, E., & Valente-Mesquita, V. L. (2011). Changes in bioactive compounds and antioxidant capacity of fresh-cut cashew apple. *Food Research International, 44*(5), 1459-1462.

Queji, M. D., Wosiacki, G., Cordeiro, G. A., Peralta-Zamora, P. G., & Nagata, N. (2010). Determination of simple sugars, malic acid and total phenolic compounds in apple pomace by infrared spectroscopy and PLSR. *International Journal of Food Science & Technology, 45*(3), 602-609.

R

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

R Core Team, R. C. (2019). R: A Language and Environment for Statistical Computing. 1(1358), 34.

Rawle, A. (2003). Basic of principles of particle-size analysis. *Surface coatings international. Part A, Coatings journal, 86*(2), 58-65.

Rahman, A., Kandpal, L. M., Lohumi, S., Kim, M. S., Lee, H., Mo, C., & Cho, B.-K. (2017). Nondestructive Estimation of Moisture Content, pH and Soluble Solid Contents in Intact Tomatoes Using Hyperspectral Imaging. *Applied Sciences, 7*(1), 109.

Rambla, F. J., Garrigues, S., Ferrer, N., & De La Guardia, M. (1998). Simple partial least squares-attenuated total reflectance Fourier transform infrared spectrometric method for the determination of sugars in fruit juices and soft drinks using aqueous standards. *Analyst, 123*(2), 277-281.

Rao, M. A. (2010). *Rheology of fluid and semisolid foods: principles and applications*: Springer Science & Business Media.

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.

Reeves, J. B. (1997). Concatenation of Near- and Mid-Infrared Spectra To Improve Calibrations for Determining Forage Composition. Journal of Agricultural and *Food Chemistry, 45*(5), 1711-1714.

Reid, L. M., Woodcock, T., O'Donnell, C. P., Kelly, J. D., & Downey, G. (2005). Differentiation of apple juice samples on the basis of heat treatment and variety using chemometric analysis of MIR and NIR data. *Food Research International, 38*(10), 1109-1115.

Rembiałkowska, Ewa; Hallmann, Ewelina and Rusaczonek, Anna (2007) Influence of Processing on Bioactive Substances Content and Antioxidant Properties of Apple Purée from Organic and Conventional Production in Poland. Poster at: 3rd QLIF Congress: Improving Sustainability in Organic and Low Input Food Production Systems, University of Hohenheim, Germany, May 31-June 3 2005.

Renard, C. M. G. C., & Thibault, J. F. (1993). Structure and properties of apple and sugar-beet pectins extracted by chelating-agents. *Carbohydrate Research, 244*(1), 99-114.

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

Rezaei, M., & Liu, B. (2017). Food loss and waste in the food supply chain. *International Nut and Dried Fruit Council: Reus, Spain*, 26-27.

Rinaldi, M., Langialonga, P., Dhenge, R., Aldini, A., & Chiavaro, E. (2021). Quality traits of apple puree treated with conventional, ohmic heating and high-pressure processing. *European Food Research and Technology*.

Rodriguez-Saona, L. E., Khambaty, F. M., Fry, F. S., Dubois, J., & Calvey, E. M. (2004). Detection and identification of bacteria in a juice matrix with fourier transform-near infrared spectroscopy and multivariate analysis. *Journal of Food Protection, 67*(11), 2555-2559.

Roger, J. M., & Boulet, J. C. (2018). A review of orthogonal projections for calibration. *Journal of Chemometrics, 32*(9).

Rojas-Candelas, L. E., Chanona-Pérez, J. J., Méndez Méndez, J. V., Perea-Flores, M. J., Cervantes-Sodi, H. F., Hernández-Hernández, H. M., & Marin-Bustamante, M. Q. (2021). Physicochemical, structural and nanomechanical study elucidating the differences in firmness among four apple cultivars. *Postharvest Biology and Technology, 171*.

Roman, M., Dobrowolski, J. C., Baranska, M., & Baranski, R. (2011). Spectroscopic Studies on Bioactive Polyacetylenes and Other Plant Components in Wild Carrot Root. *Journal of Natural Products, 74*(8), 1757-1763.

Roman, M., Marzec, K. M., Grzebelus, E., Simon, P. W., Baranska, M., & Baranski, R. (2015). Composition and (in)homogeneity of carotenoid crystals in carrot cells revealed by high resolution Raman imaging. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 136*, 1395-1400.

Root, W., & Barrett, D. (2005). Apples and apple processing. Processing Fruits: Science and Technology. CRC Press.

S

Saei, A., Tustin, D. S., Zamani, Z., Talaie, A., & Hall, A. J. (2011). Cropping effects on the loss of apple fruit firmness during storage: The relationship between texture retention and fruit dry matter concentration. *Scientia Horticulturae, 130*(1), 256-265.

Salguero-Chaparro, L., Palagos, B., Peña-Rodríguez, F., & Roger, J. M. (2013). Calibration transfer of intact olive NIR spectra between a pre-dispersive instrument and a portable spectrometer. *Computers and Electronics in Agriculture, 96*, 202-208.

Sanoner, P., Guyot, S., Marnet, N., Molle, D., Drilleau, J.-F., 1999. Polyphenol profiles of French cider apple varieties (Malus domestica sp.). *Journal of Agricultural and Food Chemistry, 47*, 4847-4853.

Schijvens, E., Van Vliet, T., & Van Dijk, C. (1998). Effect of processing conditions on the composition and rheological properties of applesauce. *Journal of Texture Studies, 29*(2), 123-143.

Schmutzler, M., & Huck, C. W. (2014). Automatic sample rotation for simultaneous determination of geographical origin and quality characteristics of apples based on near infrared spectroscopy (NIRS). *Vibrational Spectroscopy, 72*, 97-104.

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

Schulz, H., & Baranska, M. (2007). Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vibrational Spectroscopy, 43*(1), 13-25.

Schulz, H., Baranska, M., & Baranski, R. (2005). Potential of NIR-FT-Raman spectroscopy in natural carotenoid analysis. *Biopolymers, 77*(4), 212-221.

Schulz, H., Krähmer, A., Naumann, A., & Gudi, G. (2014). Infrared and Raman Spectroscopic Mapping and Imaging of Plant Materials. In *Infrared and Raman Spectroscopic Imaging* (pp. 225-294).

Sen, P. N. (2004). Diffusion and tissue microstructure. *Journal of Physics: Condensed Matter, 16*(44), S5213-S5220.

Serranti, S., Bonifazi, G., & Luciani, V. (2017). *Non-destructive quality control of kiwi fruits by hyperspectral imaging*: SPIE.

Shafie, K. A., Künnemeyer, R., McGlone, A., Talele, S., & Vetrova, V. (2015). An optimised six-wavelength model for predicting kiwifruit dry matter. *Journal of Near Infrared Spectroscopy, 23*(2), 103-109.

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*: Cambridge university press.

Shao, Y., Liu, Y., Xuan, G., Wang, Y., Gao, Z., Hu, Z., Wang, K. (2020). Application of hyperspectral imaging for spatial prediction of soluble solid content in sweet potato. *RSC Advances, 10*(55), 33148-33154.

Siddiqui, S., Brackmann, A., Streif, J., & Bangerth, F. (1996). Controlled atmosphere storage of apples: Cell wall composition and fruit softening. *Journal of Horticultural Science, 71*(4), 613-620.

Sila, D., Van Buggenhout, S., Duvetter, T., Fraeye, I., De Roeck, A., Van Loey, A., & Hendrickx, M. (2009). Pectins in processed fruits and vegetables: Part II Structure function relationships. *Comprehensive reviews in food science and food safety, 8*(2), 86-104.

Silva, A. C., Lourenço, A. S., & de Araujo, M. C. U. (2018). Simultaneous voltammetric determination of four organic acids in fruit juices using multiway calibration. *Food Chemistry, 266*, 232-239.

Sinnaeve, G., Dardenne, P., Agneessens, R., Lateur, M., & Hallet, A. (1997). Quantitative analysis of raw apple juices using near infrared, Fourier-transform near infrared and Fourier-transform infrared instruments: A comparison of their analytical performances. *Journal of Near Infrared Spectroscopy, 5*(1), 1-17.

Smith, E., & Dent, G. (2005). Modern Raman spectroscopy: a practical approach.

Smith, W. H. (1950). Cell-multiplication and cell-enlargement in the development of the flesh of the apple fruit. *Annals of Botany, 14*(1), 23-38.

Sofu, M. M., Er, O., Kayacan, M. C., & Cetişli, B. (2016). Design of an automatic apple sorting system using machine vision. *Computers and Electronics in Agriculture, 127*, 395-405.

Solomakhin, A. A., & Blanke, M. M. (2010). Mechanical flower thinning improves the fruit quality of apples. *Journal of the Science of Food and Agriculture, 90*(5), 735-741.

Song, W., Wang, H., Maguire, P., & Nibouche, O. (2016). Differentiation of organic and non-organic apples using near infrared reflectance spectroscopy—a pattern recognition approach. *2016 IEEE SENSORS* (pp. 1-3): IEEE.

Song, Y., Yao, Y.-x., Zhai, H., Du, Y.-p., Chen, F., & Shu-wei, W. E. I. (2007). Polyphenolic Compound and the Degree of Browning in Processing Apple Varieties. *Agricultural Sciences in China, 6*(5), 607-612.

Soukup, A. (2014). Selected Simple Methods of Plant Cell Wall Histochemistry and Staining for Light Microscopy. In V. Žárský & F. Cvrčková (Eds.), *Plant Cell Morphogenesis: Methods and Protocols*, (pp. 25-40). Totowa, NJ: Humana Press.

Stanley, C. J., Tustin, D. S., Lupton, G. B., McArtney, S., Cashmore, W. M., & De Silva, H. N. (2000).

Towards understanding the role of temperature in apple fruit growth responses in three geographical regions within New Zealand. *Journal of Horticultural Science and Biotechnology, 75*(4), 413-422.

Statista. (2020). Global leading countries of apple production in 2019-2020. https://www.statista.com/statistics/279555/global-top-apple-producing-countries/

Statista. (2021). Most popular apple varieties in France as of 2021. https://www.statista.com/statistics/784784/favorite-apples-varieties-france/

Stewart, D. (1996). Fourier Transform Infrared Microspectroscopy of Plant Tissues. *Applied Spectroscopy, 50*(3), 357-365.

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.

Stopar, M., Bolcina, U., Vanzo, A., & Vrhovsek, U. (2002). Lower Crop Load for Cv. Jonagold Apples (Malus × domestica Borkh.) Increases Polyphenol Content and Fruit Quality. *Journal of Agricultural and Food Chemistry, 50*(6), 1643-1646.

Su, W.-H., & Sun, D.-W. (2018). Multispectral Imaging for Plant Food Quality Analysis and Visualization. *Comprehensive Reviews in Food Science and Food Safety, 17*(1), 220-239.

Sudha, M. L., Baskaran, V., & Leelavathi, K. (2007). Apple pomace as a source of dietary fiber and polyphenols and its effect on the rheological characteristics and cake making. *Food Chemistry, 104*(2), 686-692.

Sugiyama, J. (1999). Visualization of Sugar Content in the Flesh of a Melon by Near-Infrared Imaging. *Journal of Agricultural and Food Chemistry, 47*(7), 2715-2718.

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

Sun, H., Zhang, S., Chen, C., Li, C., Xing, S., Liu, J., & Xue, J. (2019). Detection of the Soluble Solid Contents from Fresh Jujubes during Different Maturation Periods Using NIR Hyperspectral Imaging and an Artificial Bee Colony. *Journal of Analytical Methods in Chemistry, 2019*, 5032950.

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 Chemistry, 218*, 413-421.

Sun, X., Zhang, H., Pan, Y., & Liu, Y. (2009). Nondestructive measurement soluble solids content of apple by portable and online near infrared spectroscopy. *Photonics and Optoelectronics Meetings (POEM) 2009: Fiber Optic Communication and Sensors* (Vol. 7514, p. 75140P): International Society for Optics and Photonics.

Sun, Y., Xiao, H., Tu, S., Sun, K., Pan, L., & Tu, K. (2018). Detecting decayed peach using a rotating hyperspectral imaging testbed. *LWT- Food Science and Technology, 87*, 326-332.

Swinehart, D. F. (1962). The beer-lambert law. *Journal of chemical education*, 39(7), 333.

Szczesniak, A. S., & Kahn, E. L. (1971). Consumer awareness of and attitudes to food texture: I: Adults. *Journal of Texture Studies, 2*(3), 280-295.

Szuvandzsiev, P., Helyes, L., Lugasi, A., Szántó, C., Baranowski, P., & Pék, Z. (2014). Estimation of antioxidant components of tomato using VIS-NIR reflectance data by handheld portable spectrometer. *International Agrophysics, 28*(4), 521-527.

Szymanska-Chargot, M., Chylinska, M., Kruk, B., & Zdunek, A. (2015). Combining FT-IR spectroscopy and multivariate analysis for qualitative and quantitative analysis of the cell wall composition

changes during apples development. *Carbohydrate Polymers, 115*, 93-103.

Szymańska-Chargot, M., Chylińska, M., Pieczywek, P. M., Rösch, P., Schmitt, M., Popp, J., & Zdunek, A. (2016). Raman imaging of changes in the polysaccharides distribution in the cell wall during apple fruit development and senescence. *Planta, 243*(4), 935-945.

T

Tarea, S., Cuvelier, G., & Sieffermann, J. M. (2007). Sensory evaluation of the texture of 49 commercial apple and pear purees. *Journal of Food Quality, 30*(6), 1121-1131.

Tauler, R. (1995). Multivariate curve resolution applied to second order data. *Chemometrics and intelligent laboratory systems, 30*(1), 133-146.

Tauler, R. (2005). Multivariate curve Resolution: theory and applications. *Constraints*.

Telias, A., Bradeen, J. M., Luby, J. J., Hoover, E. E., & Allen, A. C. (2011). Regulation of Anthocyanin Accumulation in Apple Peel. In *Horticultural Reviews* (pp. 357-391): wiley.

Temma, T., Hanamatsu, K., & Shinoki, F. (2002). Measuring the sugar content of apples and apple juice by near infrared spectroscopy. *Optical Review, 9*(2), 40-44.

Tewari, J., Joshi, M., Gupta, A., Mehrotra, R., & Chandra, S. (1999). Determination of sugars and organic acid concentration in apple juices using infrared spectroscopy. *NISCAIR-CSIR*.

Thompson, A. K. (2008). *Fruit and vegetables: harvesting, handling and storage*: John Wiley & Sons.

Tian, X., Fan, S., Li, J., Xia, Y., Huang, W., & Zhao, C. (2019). Comparison and optimization of models for SSC on-line determination of intact apple using efficient spectrum optimization and variable selection algorithm. *Infrared Physics & Technology, 102*, 102979.

Tian, Y., Sun, J., Zhou, X., Wu, X., Lu, B., & Dai, C. (2020). Research on apple origin classification based on variable iterative space shrinkage approach with stepwise regression–support vector machine algorithm and visible-near infrared hyperspectral imaging. *Journal of Food Process Engineering, 43*(8).

Timilsena, Y. P., Vongsvivut, J., Tobin, M. J., Adhikari, R., Barrow, C., & Adhikari, B. (2019). Investigation of oil distribution in spray-dried chia seed oil microcapsules using synchrotron-FTIR microspectroscopy. *Food Chemistry, 275*, 457-466.

Ting, V. J. L., Silcock, P., Bremer, P. J., & Biasioli, F. (2013). X-Ray Micro-Computer Tomographic Method to Visualize the Microstructure of Different Apple Cultivars. *Journal of Food Science, 78*(11), E1735-E1742.

Tiwari, B. K., Norton, T., & Holden, N. M. (2013). *Sustainable food processing*: John Wiley & Sons.

Toledo-Martín, E. M., García-García, M. C., Font, R., Moreno-Rojas, J. M., Gómez, P., Salinas-Navarro, M., & Del Río-Celestino, M. (2016). Application of visible/near-infrared reflectance spectroscopy for predicting internal and external quality in pepper. *Journal of the Science of Food and Agriculture, 96*(9), 3114-3125.

Tomala, K., Andziak, J., Jeziorek, K., & Dziuban, R. (2008). Influence of rootstock on the quality of 'Jonagold' apples at harvest and after storage. *Journal of Fruit and Ornamental Plant Research, 16*, 31-38.

Travers, S., Bertelsen, M. G., & Kucheryavskiy, S. V. (2014). Predicting apple (cv. Elshof) postharvest dry matter and soluble solids content with near infrared spectroscopy. *Journal of the Science of Food and Agriculture, 94*(5), 955-962.

Tu, K., Nicolaï, B., & De Baerdemaeker, J. (2000). Effects of relative humidity on apple quality under simulated shelf temperature storage. *Scientia Horticulturae, 85*(3), 217-229.

Türker-Kaya, S., & Huck, C. W. (2017). A Review of Mid-Infrared and Near-Infrared Imaging: Principles, Concepts and Applications in Plant Tissue Analysis. *Molecules, 22*(1), 168.

U

USDA, Agricultural Research Service. (2019) Apples, raw, with skin (Includes foods for USDA's Food Distribution Program). Uploaded on January 4, 2019, from https://fdc.nal.usda.gov/ fdc-app.html#/food-details/171688/nutrients.

USDA, (2020). Fresh deciduous Fruit: World markets and trade (apples, grapes, & pears). Uploaded on January 4, 2019, from https://usda.library.cornell.edu/concern/publications/1z40ks800?locale =en.

V

Van Boekel, M. A. J. S. (2008). Kinetic Modeling of Food Quality: A Critical Review. *Comprehensive reviews in food science and food safety, 7*(1), 144-158.

Van Beers, R., Aernouts, B., Gutierrez, L. L., Erkinbaev, C., Rutten, K., Schenk, A., . . . Saeys, W. (2015). Optimal Illumination-Detection Distance and Detector Size for Predicting Braeburn Apple Maturity from Vis/NIR Laser Reflectance Measurements. *Food and Bioprocess Technology, 8*(10), 2123-2136.

Van der Sluis, A. A., Dekker, M., De Jager, A., & Jongen, W. M. F. (2001). Activity and concentration of polyphenolic antioxidants in apple: Effect of cultivar, harvest year, and storage conditions. *Journal of Agricultural and Food Chemistry, 49*(8), 3606-3613.

Varela, P., Salvador, A., & Fiszman, S. (2005). Shelf-life estimation of 'Fuji' apples: Sensory characteristics and consumer acceptability. *Postharvest Biology and Technology, 38*(1), 18-24.

Varela, P., Salvador, A., & Fiszman, S. (2007). Changes in apple tissue with storage time: Rheological, textural and microstructural analyses. *Journal of Food Engineering, 78*(2), 622-629.

Veberic, R., Schmitzer, V., Petkovsek, M. M., & Stampar, F. (2010). Impact of Shelf Life on Content of Primary and Secondary Metabolites in Apple (Malus domestica Borkh.). *Journal of Food Science, 75*(9), S461-S468.

Verboven, P., Kerckhofs, G., Mebatsion, H. K., Ho, Q. T., Temst, K., Wevers, M., . . . Nicolaï, B. M. (2008). Three-Dimensional Gas Exchange Pathways in Pome Fruit Characterized by Synchrotron X-Ray Computed Tomography. *Plant Physiology, 147*(2), 518-527.

Vidot, K., Devaux, M. F., Alvarado, C., Guyot, S., Jamme, F., Gaillard, C., . . . Lahaye, M. (2019). Phenolic distribution in apple epidermal and outer cortex tissue by multispectral deep-UV autofluorescence cryo-imaging. *Plant Science, 283*, 51-59.

Vidot, K., Rivard, C., Van Vooren, G., Siret, R., & Lahaye, M. (2020). Metallic ions distribution in texture and phenolic content contrasted cider apples. *Postharvest Biology and Technology, 160*, 111046.

Vincent, J., Wang, H., Nibouche, O., & Maguire, P. (2018). Differentiation of apple varieties and investigation of organic status using portable visible range reflectance spectroscopy. *Sensors (Switzerland), 18*(6).

Vishnu Prasanna, K., Sudhakar Rao, D., & Krishnamurthy, S. (2000). Effect of storage temperature on ripening and quality of custard apple (Annona squamosa L.) fruits. *The Journal of Horticultural Science and Biotechnology, 75*(5), 546-550.

Visser, T., Schaap, A. A., & De Vries, D. P. (1968). Acidity and sweetness in apple and pear. *Euphytica, 17*(2), 153-167.

Volz, R. K., & McGhie, T. K. (2011). Genetic Variability in Apple Fruit Polyphenol Composition in Malus × domestica and Malus sieversii Germplasm Grown in New Zealand. *Journal of Agricultural and Food Chemistry, 59*(21), 11509-11521.

Vongsvivut, J., Pérez-Guaita, D., Wood, B. R., Heraud, P., Khambatta, K., Hartnell, D., . . . Tobin, M. J. (2019). Synchrotron macro ATR-FTIR microspectroscopy for high-resolution chemical mapping of single cells. *Analyst, 144*(10), 3226-3238.

Vrhovsek, U., Rigo, A., Tonon, D., & Mattivi, F. (2004). Quantitation of polyphenols in different apple varieties. *Journal of Agricultural and Food Chemistry, 52*(21), 6532-6538.

W

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 Biology and Technology, 168*.

Walsh, K. B., Golic, M., & Greensill, C. V. (2004). Sorting of fruit using near infrared spectroscopy: Application to a range of fruit and vegetables for soluble solids and dry matter content. *Journal of Near Infrared Spectroscopy, 12*(3), 141-148.

Walsh, K. B., McGlone, V. A., & Han, D. H. (2020). The uses of near infra-red spectroscopy in postharvest decision support: A review. *Postharvest Biology and Technology, 163*.

Wang, H., Peng, J., Xie, C., Bao, Y., & He, Y. (2015). Fruit quality evaluation using spectroscopy technology: A review. *Sensors (Switzerland), 15*(5), 11889-11927.

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.

Wang, L., Tian, X., Tian, X., Wu, H., Li, R., Liu, J., & Ren, X. (2020). Effect of different O2/CO2 proportions on the physiological characteristics of 'Fuji' apple fruit during modified atmosphere storage. *Journal of Fruit Science, 37*(6), 909-919.

Wang, Z., Verboven, P., & Nicolai, B. (2017). Contrast-enhanced 3D micro-CT of plant tissues using different impregnation techniques. *Plant Methods, 13*(1), 105.

Warrington, I. J., Fulton, T. A., Halligan, E. A., & De Silva, H. N. (1999). Apple fruit growth and maturity are affected by early season temperatures. *Journal of the American Society for Horticultural Science, 124*(5), 468-477.

Watkins, R. (1985). Apple genetic resources. *Acta Hortic, 159*, 21-30.

Way, R. D., Aldwinckle, H. S., Lamb, R. C., Rejman, A., Sansavini, S., Shen, T., . . . Yoshida, Y. (1990). Apples (Malus). *Genetic Resources of Temperate Fruit and Nut Crops, 290*, 3-62.

Westwood, M. N., Batjer, L. P., & Billingsley, H. D. (1967). Cell size, cell number, and fruit density of apples as related to fruit size, position in cluster, and thinning method. *Proc. Amer. Soc. Hort. Sci., 91*, 51-62.

Wilson, R. F. (1941). New Zealand Production. *Economic Record, 17*(33), 252-255.

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

Wismer, P. T., Proctor, J., & Elfving, D. (1995). Benzyladenine affects cell division and cell size during apple fruit thinning. *Journal of the American Society for Horticultural Science, 120*(5), 802-

807.

Włodarska, K., Khmelinskii, I., & Sikorska, E. (2018). Evaluation of quality parameters of apple juices using near-infrared spectroscopy and chemometrics. *Journal of Spectroscopy, 2018*.

Wu, X., Wu, B., Sun, J., & Li, M. (2015). Rapid discrimination of apple varieties via near-infrared reflectance spectroscopy and fast allied fuzzy C-means clustering. *International Journal of Food Engineering, 11*(1), 23-30.

Wu, X., Zhou, H., Wu, B., & Fu, H. (2020). Determination of apple varieties by near infrared reflectance spectroscopy coupled with improved possibilistic Gath–Geva clustering algorithm. *Journal of Food Processing and Preservation, 44*(8), e14561.

Wu, Z., Pu, H., & Sun, D.-W. (2021). Fingerprinting and tagging detection of mycotoxins in agri-food products by surface-enhanced Raman spectroscopy: Principles and recent applications. *Trends in Food Science & Technology, 110*, 393-404.

Wünsche, J. N., Lakso, A. N., Robinson, T. L., Lenz, F., & Denning, S. S. (1996). The bases of productivity in apple production systems: The role of light interception by different shoot types. *Journal of the American Society for Horticultural Science, 121*(5), 886-893.

X

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.

Xiao, N., Bock, P., Antreich, S. J., Staedler, Y. M., Schönenberger, J., & Gierlinger, N. (2020). From the Soft to the Hard: Changes in Microchemistry During Cell Wall Maturation of Walnut Shells. *Frontiers in Plant Science, 11*.

Xiao, Q., Bai, X., & He, Y. (2020). Rapid Screen of the Color and Water Content of Fresh-Cut Potato Tuber Slices Using Hyperspectral Imaging Coupled with Multivariate Analysis. *Foods, 9*(1), 94.

Xie, L. J., Wang, A. C., Xu, H. R., Fu, X. P., & Ying, Y. B. (2016). Applications of Near-infrared systems for quality evaluation of fruits: A review. *Transactions of the ASABE, 59*(2), 399-419.

Xu, M.-L., Gao, Y., Han, X. X., & Zhao, B. (2017). Detection of pesticide residues in food using surface-enhanced Raman spectroscopy: A review. *Journal of Agricultural and Food Chemistry, 65*(32), 6719-6726.

Y

Yang, D., & Ying, Y. (2011). Applications of Raman spectroscopy in agricultural products and food analysis: A review. *Applied Spectroscopy Reviews, 46*(7), 539-560.

Yang, J., Liu, Q., Zhao, N., Chen, J., Peng, J., Pan, L., & Tu, K. (2020). Hyperspectral Imaging for Non-destructive Determination and Visualization of Moisture and Carotenoid Contents in Carrot Slices during Drying. *Shipin Kexue/Food Science, 41*(12), 285-291.

Yaseen, T., Sun, D., & Cheng, J. (2017). Raman imaging for food quality and safety evaluation: Fundamentals and applications. *Trends in Food Science & Technology, 62*, 177-189.

Ye, M., Gao, Z., Li, Z., Yuan, Y., & Yue, T. (2016). Rapid detection of volatile compounds in apple wines using FT-NIR spectroscopy. *Food Chemistry, 190*, 701-708.

Ye, M., Yue, T., Yuan, Y., & Li, Z. (2014). Application of FT-NIR Spectroscopy to Apple Wine for Rapid Simultaneous Determination of Soluble Solids Content, pH, Total Acidity, and Total Ester Content. *Food and Bioprocess Technology, 7*(10), 3055-3062.

Yuan, R. (2007). Effects of temperature on fruit thinning with ethephon in 'Golden Delicious' apples. *Scientia Horticulturae, 113*(1), 8-12.

Z

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*(12), 2099-2118.

Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., & Liu, C. (2014). Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Research International, 62*, 326-343.

Zhang, D., Pu, H., Huang, L., & Sun, D.-W. (2021). Advances in flexible surface-enhanced Raman scattering (SERS) substrates for nondestructive food detection: Fundamentals and recent applications. *Trends in Food Science & Technology, 109*, 690-701.

Zhang, D., Xu, L., Liang, D., Xu, C., Jin, X., & Weng, S. (2018). Fast Prediction of Sugar Content in Dangshan Pear (Pyrus spp.) Using Hyperspectral Imagery Data. *Food Analytical Methods, 11*(8), 2336-2345.

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.

Zhang, H., Cha, S., & Yeung, E. S. (2007). Colloidal Graphite-Assisted Laser Desorption/Ionization MS and MSn of Small Molecules. 2. Direct Profiling and MS Imaging of Small Metabolites from Fruits. *Analytical Chemistry, 79*(17), 6575-6584.

Zhang, Y., Nock, J. F., Al Shoffe, Y., & Watkins, C. B. (2019). Non-destructive prediction of soluble solids and dry matter contents in eight apple cultivars using near-infrared spectroscopy. *Postharvest Biology and Technology, 151*, 111-118.

Zhao, Y., Zhang, C., Zhu, S., Li, Y., He, Y., & Liu, F. (2020). Shape induced reflectance correction for non-destructive determination and visualization of soluble solids content in winter jujubes using hyperspectral imaging in two different spectral ranges. *Postharvest Biology and Technology, 161*, 111080.

Zhu, D., Ji, B., Qing, Z., Wang, C., & Zude, M. (2011). The detection of quality deterioration of apple juice by near infrared and fluorescence spectroscopy. *IFIP Advances in Information and Communication Technology* (Vol. 346 AICT, pp. 84-91).

Zhu, H., Chu, B., Fan, Y., Tao, X., Yin, W., & He, Y. (2017). Hyperspectral Imaging for Predicting the Internal Quality of Kiwifruits Based on Variable Selection Algorithms and Chemometric Models. *Scientific Reports, 7*(1), 7845.

Zou, X., Li, Y., & Zhao, J. (2007). Using genetic algorithm interval partial least squares selection of the optimal near infrared wavelength regions for determination of the soluble solids content of "Fuji" apple. *Journal of Near Infrared Spectroscopy, 15*(3), 153-159.

Zou, X., & Zhao, J. (2015). Machine vision online measurements. *Nondestructive measurement in food and agro-products*, 11-56.

Zou, X., Chen Q., Shi, Z., Li G., Zhang C., Zhao J. Multi-dimensional intelligent vision technology and application of characteristic food processing. (2019). *Second Prize of China National Technological Invention Award*.

Zude-Sasse, M., Herold, B., & Geyer, M. (2000). Comparative study on maturity prediction in 'Elstar' and 'Jonagold' apples. *Gartenbauwissenschaft, 65*(6), 260-265.

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

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

VII. Résumé

Cette thèse a montré comment la spectroscopie vibrationnelle (NIRS, MIRS, Raman et NIRS-HSI) couplée à des méthodes de chimiométrie (PCA, FDA, PLS, MCR-ALS et machine learning) permet de mettre en évidence la variabilité et l'hétérogénéité des pommes et des purées de pommes. Les essais expérimentaux ont été conçus pour obtenir une variabilité importante à la fois dans les pommes et les purées en modulant plusieurs facteurs, au verger (variétés, pratiques culturales avec l'éclaircissage des fruits), lors du stockage post-récolte (4°C pendant 0, 1, 3 et 6 mois) et lors de la transformation (températures de cuisson, vitesses de broyage et niveaux de raffinage). La cuisson des pommes a été réalisée à l'échelle du laboratoire avec un robot coupeur-cuiseur pour étudier, de façon maitrisée, le lien entre la qualité des pommes avant transformation et la qualité des purées après transformation. Les informations spectrales et les données de référence ont été systématiquement acquises sur l'ensemble des pommes et purées, afin d'explorer de nouvelles solutions pour gérer la variabilité et l'hétérogénéité lors de la transformation des pommes.

Nos travaux de recherche ont visé en particulier à répondre aux questions suivantes :

Comment identifier la variabilité et l'hétérogénéité des pommes et des purées à l'aide de différentes techniques de spectroscopie et d'imagerie ?

Les images hyperspectrales en proche infrarouge (NIRS-HSI) et les données de référence ont permis de mettre en évidence une grande hétérogénéité des teneurs en matière sèche (DMC), sucres totaux (TSC), sucres individuels, acide malique et polyphénols dans les pommes individuelles de quatre variétés (**Article III**). Cette variabilité a été observée dans toutes les directions à l'intérieur de chaque pomme, du proximal vers le distal, de l'intérieur vers l'extérieur et dans le sens équatorial. Par rapport à la littérature, notre étude a fourni une solution efficace pour établir les modèles de prédiction, en utilisant les données de référence mesurées sur 141 échantillons sélectionnés à partir de leurs données NIRS-HSI et représentatifs des 1056 échantillons préparés. Notre étude a ainsi permis d'illustrer la répartition des TSC et DMC dans les pommes, avec un nombre limité d'analyses complexes (sucres individuels mesurés par spectrophotométrie à l'aide de kits enzymatiques) et chronophages (au moins 24 heures pour la lyophilisation) nécessaires à la modélisation. Les TSC et DMC ont varié plus intensément dans les pommes Braeburn et Royal Gala que dans les pommes Granny

Smith et Golden Delicious. Cependant, notre approche n'a pas permis de décrire la distribution des autres variables étudiées dans les pommes (acide malique et polyphénols).

Dans la plupart de nos expérimentations (**Articles I, II, IV, V et VI**), une large gamme de variabilité des pommes a été considérée, tout en assurant la traçabilité entre les pommes et les purées après cuisson. La variabilité des pommes a entraîné des différences significatives de couleur (L*, a* et b*), de texture (viscosité et viscoélasticité), d'acidité titrable (TA), des teneurs en solides solubles (SSC), fructose, saccharose, acide malique et parois cellulaires (AIS) dans les purées (**Articles I, II et IV, tableau 28**).

La NIRS a discriminé les lots de pommes et de purées en fonction de la variété et de la durée de stockage, avec un niveau de classification supérieur à 82% (**Article IV**). De plus, la NIRS a permis de prédire de façon acceptable les paramètres globaux (SSC, DMC, TA) à la fois dans les pommes et les purées ($R^2 > 0,82$, RPD > 2,3) (**Articles I et IV**). Cependant, elle n'a pas permis de prédire la viscosité de la purée. Avec le couplage des deux régions spectrales, VIS et NIR, les paramètres de couleur (L*, a* et b*) ont été prédits de façon acceptable lorsque la variabilité des purées était suffisamment importante (**Article IV, Tableau 29**).

Les techniques VIS-NIRS, NIRS, MIRS, Raman et NIRS-HSI ont été comparées pour déterminer la qualité des purées (**Article I**). La MIRS a été la plus performante pour évaluer la variabilité des purées (variété, pratique d'éclaircissage et stockage) et les conditions de transformation (chauffage, broyage et raffinage), avec une précision de discrimination supérieure à 90%. La MIRS a permis d'évaluer les propriétés biochimiques de la purée (SSC, TA, DMC, teneurs en fructose, saccharose et acide malique) avec un RPD allant de 2,2 à 6,0 (**Articles I et VI**). Elle a également permis d'estimer les propriétés rhéologiques (viscosité et modules viscoélastiques) et texturales (taille et volume des particules) des purées (**Articles I et II**). En lien avec l'évolution de la texture au cours de la transformation, la MIRS a également permis de déterminer les teneurs en parois cellulaires (AIS) et de visualiser leur évolution (principalement la solubilisation des pectines et la perte de galactose) mais uniquement sur les purées lyophilisées et sur les extraits purifiés des parois cellulaires, respectivement.

Comment est liée la variabilité des pommes à la qualité des purées

transformées ? Et, est-il possible de prédire la qualité des purées transformées à partir des spectres VIS-NIR, NIR et MIR des pommes intactes ?

Deux expérimentations originales ont été menées pour explorer comment les variabilités intra-lots et inter-lots de pommes affectaient les propriétés des purées (**Articles IV et V**). Différentes purées ont été cuites, à l'aide d'un robot coupeur-cuiseur, à partir de lots de pommes présentant une grande variabilité inter-lots provenant de différentes variétés, pratiques d'éclaircissage et durées de conservation (**Article IV**). Une ACP a confirmé un impact similaire de ces facteurs à la fois dans les pommes et les purées (**Fig. 35**). Des corrélations ont ensuite été montrées : entre la texture des pommes (fermeté et croquant) et la viscosité des purées ($R^2 > 0,79$), de même qu'entre la composition des pommes et des purées pour TA ($R^2 > 0,91$), SSC ($R^2 > 0,79$) et DMC ($R^2 > 0,72$) (**Articles IV, Fig. 36**). Cependant, ces relations ont été obtenues sur des lots de pommes (au moins 2,5 kg représentant environ 10 à 15 pommes transformées dans le robot coupeur-cuiseur).

Afin de prendre en compte l'effet de la variabilité intra-lots apportée par chaque pomme, un essai a été conçu pour répondre à la définition absolue « d'une pomme pour une purée » en utilisant un procédé avec les micro-ondes permettant de cuire les pommes individuellement (**Article V**). La variabilité issue des variétés, observée précédemment, a été confirmée ici. De plus, certains critères de qualité se sont révélés être très variables entre les pommes et purées d'une même variété. Par exemple, la couleur de la purée a été beaucoup plus variable entre les pommes pour Braeburn et Royal Gala, de même que la SSC pour Golden Delicious et la viscosité pour Granny Smith. A l'inverse, certaines caractéristiques ont été assez stables au sein d'une même variété, comme l'acidité pour Royal Gala ou la couleur pour Golden Delicious (**Fig. 40 et 41**). La limitation de cette étude est qu'il n'était pas possible d'effectuer des mesures destructives de qualité sur les pommes individuelles car celles-ci étaient cuites intactes dans le four à micro-ondes. Brièvement, nous avons tout de même confirmé que les variabilités inter-lots et intra-lots des pommes impactaient les caractéristiques des purées. Les relations entre les spectres VIS, NIR et MIR des pommes et des purées ont été explorées en parallèle (**Articles IV, V et VI**). Concernant la cuisson des pommes individuelles à l'aide des micro-ondes, une méthode de corrélation bidimensionnelle (2D-COS) a mis en évidence de fortes corrélations dans les régions VIS et NIR, en particulier à 650-680 nm et 1125-1400 nm (**Article V, Fig. 44**). Les spectres VIS-NIR

des pommes ont permis une évaluation fiable de la viscosité des purées ($R^2 > 0,81$), de leur SSC ($R^2 = 0,78$) et TA ($R^2 = 0,87$) (**Tableau 34**). Concernant la cuisson des lots de pommes avec le robot coupeur-cuiseur, des prédictions satisfaisantes de la viscosité, SSC et TA de la purée ($R^2 > 0,82$, $> 0,80$ et $> 0,80$ respectivement) et même des teneurs en parois cellulaires ($R^2 > 0,81$) ont été obtenues à partir des spectres NIR moyennés des pommes (**Article IV, Tableaux 31 et 32**).

Ainsi, les techniques VIS-NIRS et NIRS seraient pertinentes pour l'industrie pour évaluer le goût et la texture des purées à partir de mesures non-destructives des pommes individuelles (**Fig. 59**). Cependant, la difficulté dans toutes ces expérimentations a été d'acquérir systématiquement les spectres et les données de référence sur les pommes fraiches et transformées, en prenant en compte une large variabilité et donc en caractérisant un nombre conséquent d'échantillons, pour la modélisation. Afin d'améliorer notre démarche, une expérimentation a été conçue dans laquelle différentes conditions de cuisson ont été comparées (3 températures de chauffage et 3 niveaux de broyage), et dans laquelle une méthode de normalisation directe a été appliquée afin d'établir les relations entre les spectres MIR des pommes et des purées (**Article VI**). Cette stratégie a permis de construire des spectres MIR de purées à partir des spectres MIR des pommes fraiches broyées. De cette façon, la MIRS couplée à des modèles PLS utilisant ces spectres reconstruits de purées a permis de prédire les critères suivants : TA ($R^2 > 0,86$), SSC ($R^2 > 0,85$), DMC ($R^2 > 0,84$) et taille moyenne des particules ($R^2 > 0,84$).

Comment améliorer la formulation des purées par spectroscopie infrarouge comme solution innovante pour gérer la variabilité des purées de pommes ?

Les articles IV, V et VI ont été principalement dédiés à l'exploration de solutions pour gérer la variabilité des pommes et des purées transformées en utilisant la spectroscopie infrarouge. L'accent a été mis sur le développement d'une méthode simple pour aider à la formulation de la purée (**Article VII**). Pour ce faire, quatre purées monovariétales ont été préparées puis mélangées deux à deux avec neuf proportions différentes variant de 5% à 95%. Une méthode chimiométrique innovante basée sur les profils de concentration (MCR-ALS) a été testée pour reconstruire les spectres des purées formulées à partir des spectres des purées monovariétales, dans un objectif d'optimiser les propriétés physico-chimiques des purées finales. A notre connaissance,

il s'agit du premier article concernant la prédiction de la qualité des purées mélangées en fonction de l'information spectrale des purées monovariétales en utilisant une approche de reconstruction spectrale. Cette approche permet une optimisation multi-paramètre de la texture et du goût des purées finales (principalement la viscosité, SSC, TA, acide malique, avec des valeurs RPD > 4,0) à partir de purées monovariétales variables. Il s'agit ici d'optimiser les mélanges de purées et contrôler la qualité du produit fini, voire d'aider au développement de nouveaux produits (**Fig. 61**).

En conclusion, cette thèse a identifié plusieurs solutions potentielles pour l'industrie de la transformation des pommes, afin de gérer la variabilité et l'hétérogénéité, qui tendent à augmenter en lien avec le changement climatique, la coexistence de différents systèmes de production et les procédés de transformation. C'est un premier pas vers une transformation des fruits plus durable et plus précise en utilisant la connaissance des matières premières pour prédire la qualité des produits transformés. Nos résultats ouvrent la possibilité d'identifier les gammes spectrales pertinentes de la région infrarouge qui pourraient être intégrées dans le PAT (Process Analytical Technology) de la chaine de transformation de la pomme (**Fig. 62**) avec notamment :

i) A la récolte des pommes, utiliser la spectroscopie VIS-NIR, NIR ou MIR sur les pommes fraiches pour déterminer objectivement les propriétés qualitatives des purées en fonction des propriétés initiales des pommes et des conditions de transformation, mieux gérer les pommes et réduire les pertes tout au long de la chaîne de transformation,

ii) Après la transformation des pommes, utiliser la spectroscopie MIR sur des purées monovariétales très variables pour guider la formulation des purées qui atteindraient alors la qualité standard, stable, appréciée des consommateurs ou au contraire produire des purées personnalisées avec des qualités spécifiques.

Résumé

Cette thèse visait à montrer comment la spectroscopie vibrationnelle incluant le proche infrarouge (NIR), le moyen infrarouge (MIR), le Raman et l'imagerie hyperspectrale (NIR-HSI) couplée à la chimiométrie pouvait mettre en évidence la variabilité et l'hétérogénéité des pommes et des purées après transformation. Des essais expérimentaux ont été conçus pour moduler plusieurs facteurs, au verger (variétés, pratiques agricoles), pendant le stockage post-récolte (4°C) et la transformation (température, broyage et raffinage) afin de faire varier les propriétés et la composition des pommes et des purées. Une approche efficace utilisant la NIRS-HSI a permis d'illustrer la répartition des sucres totaux et de la matière sèche à l'intérieur des pommes. Les variabilités inter-lots et intra-lots des pommes ont profondément modifié les purées. La spectroscopie MIR a été le meilleur outil pour détecter la variabilité des purées et évaluer leurs propriétés biochimiques (solides solubles, acidité, matière sèche, fructose, saccharose et acide malique), rhéologiques (viscosité et modules viscoélastiques) et texturales (taille et volume des particules). Des corrélations linéaires ont été trouvées entre la texture de la pomme et la viscosité de la purée, ainsi qu'entre les pommes et les purées pour l'acidité, les solides solubles et la matière sèche. Ainsi, les techniques VIS-NIR et NIR ont permis de prédire le goût et la texture des purées à partir des spectres acquis sur les pommes intactes. De plus, la spectroscopie MIR a pu guider la formulation des purées à partir des spectres des purées monovariétales. Nos approches innovantes pourraient fournir des données objectives pour mieux gérer les pommes et adapter les conditions de transformation en fonction de leurs propriétés initiales. L'objectif ultime est d'améliorer la qualité des fruits frais et transformés tout en réduisant les pertes.

Mots clés : *Malus domestica* Borkh., Purée, Variabilité, Hétérogénéité, Spectroscopie infrarouge, Imagerie hyperspectrale, Prédiction, Discrimination.

Abstract

This thesis aimed to show how vibrational spectroscopy including near infrared (NIR), mid infrared (MIR), Raman and NIR hyperspectral imaging (NIR-HSI) coupled with advanced chemometrics can highlight the variability and heterogeneity of both, raw apples and processed purees. Experimental trials were designed to modulate several factors in orchard (varieties, agricultural practices), during post-harvest storage (4°C) and processing (temperature, grinding and refining) in order to modify properties and composition of apples and purees. An efficient approach using NIR-HSI allowed illustrating the distribution of total sugars and dry matter inside apples. The inter-batch variability of apples and the intra-batch variability between individual apples intensively changed the cooked purees. MIR spectroscopy was the best tool to detect the variability of purees and assess their biochemical (soluble solids, acidity, dry matter, fructose, sucrose and malic acid), rheological (viscosity and viscoelastic moduli) and textural (particle size and volume) properties. Good linear correlations were found between apple texture and puree viscosity, as well as between apple and puree acidity, soluble solids and dry matter. Therefore, VIS-NIR and NIR techniques allowed to predict the taste and texture of purees from the non-destructive spectra of apples. Besides, MIR spectroscopy can guide puree formulation from spectra of single-variety purees. Our innovative approaches could provide objective data to better manage apples and to adapt processing conditions according to their initial properties. The ultimate goal is to improve the quality of fresh and processed fruits while reducing losses.

Keywords: *Malus domestica* Borkh., Puree, Variability, Heterogeneity, Infrared spectroscopy, Hyperspectral imaging, Prediction, Discrimination.

AVIGNON UNIVERSITÉ