The Applied Use of Technology in Phenotyping Apple (*Malus* × *domestica* Borkh.) Fruit and Trees

---

A THESIS

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

BY

Joshua Donel Anderson

Advised by

Dr. James J. Luby, Dr. Cindy B. Tong

July 2019

ACKNOWLEDGMENTS

Appreciation is expressed to my advisors Dr. James J. Luby and Dr. Cindy B. Tong for their patience and discretion as they took me on as a graduate student. I have only developed my current confidence and capabilities because of their support and continued guidance. They were both interested in my development as a graduate student and as a person. My family and I are very thankful for the funding that made my degree possible.

I extend gratitude to my lab mates who assisted me in my data collection and were excellent sounding-boards who providing valuable feedback. Namely Jack Tillman, Ashley Powell, Seth Wannemuehler, Elizabeth Blissett, Baylee Miller, Hsueh-Yuan Chang, Nicole Marshall, Stephen Opyio, Soon Li Teh, Laise Moreira, Anna Underhill, Lu Yin, and Nicholas Howard.

I wish to thank Dr. I. Volkan Isler for his positive support on my committee and providing me the opportunity to collaborate with his Robotic Sensor Network Lab, of which I acknowledge the hard work of Wenbo Dong for the image-derived data.

I am grateful for my parents for their prayers, encouragement, and support. I am thankful for Laura and Samantha Anderson for the spark in their eyes and playing with me. Words cannot express how much I appreciate the endurance, love, and support of my wife Leisa. My family is the purpose for which I do everything.

ABSTRACT

Improvements in phenotyping methods for apple (Malus × *domestica* Borkh.) are necessary to harness the potential information contained genomic data. Current phenotyping methods are destructive, time and labor intensive and can be subjective. The practical utility of technology needs to be determined to increase the quality and quantity of phenomic data.

Near infrared spectrometers are now available as portable handheld instruments capable of use outside of a laboratory. A handheld NIR spectrometer was used to acquire spectra along with traditional phenotyping of several fruit quality traits. Results showed that two trait spectral models (starch pattern index, and soluble solids concentration) had the ability to discriminate between low and high groups across a dataset of 15 cultivars. The models for other fruit traits examined, firmness, TA, skin traits, etc., either had low predictive abilities, or would not be useful beyond the ability to discriminate between cultivars with extreme trait values. Temperature and outdoor limitations of the F-750 spectrometer were determined to be minimal compared to the importance of collecting more than a single scan per fruit to control for individual fruit heterogeneity.

The architecture of apple trees is difficult to characterize, but it influences the efficiency of orchard systems. Growers depend on tree architecture research for decision making for example in tree training schemes, and rootstock-scion performance. Low cost 3D sensors are available, yet their practical horticultural application needs to be developed. A red-green-blue-depth (RGB-D) sensor was used to image a rootstock experiment. The relationship between image-derived metrics and hand-measured was highest for tree height R^2=0.93, and TCA R^2=0.71. With improvements in image processing time, image-derived tree metrics are a feasible replacement to manual measures. The predictive ability of cumulative yield by image based-tree volume was lower than by manual-measured tree volume. Tree volume in general did not improve upon the mixed model containing TCA and tree height when estimating cumulative yield.

TABLE OF CONTENTS

TABLE OF ABBREVIATIONS

Abbreviations	Definitions
A1	Average pressure in first 8 mm of fruit flesh (N)
A2	Average pressure between first 8 mm of fruit flesh to core (N)
BRusset	Presence of body russet
C0	Creep distance test (mm)
Cn	MDT-2 crispness test
Diameter	Fruit diameter (mm)
DM	Dry matter (%)
E2	Pressure just above the core (N)
F-750	Felix Instruments F-750 handheld spectrometer
Gcolor	At-harvest ground color (green-yellow)
LRusset	Presence of lenticel russet
LOO	Leave one out
M1	Maximum pressure (N) in first 8 mm of fruit flesh (≈Magness-Taylor & EPT-1)
M2	Maximum pressure in fruit flesh, 8 mm to core (N)
NIRS	Near infrared spectroscopy
OAH	Overall average hardness (N)
OMH	Overall maximum hardness (N)
%Ocolor	Percent of fruit covered by over-color
Ocolor	Over-color color (orange-pink-red)
Ocolor type	Type of over-color (blush-stripe)
pH	pH (potential hydrogen) of fruit juice
Flav	Total flavonoids (mg catechin/L), measured at 300 nm
Phen	Total phenolic content (mg catechin/L), measured at 765 nm
PHweight	Weight of fruit after storage (g)
PLSR	Partial least squares regression
QF	Quality factor (uses: M1, A2, E2, C0, & CN)
R^2	Coefficient of determination

Abbreviations	Definitions
RMSE (C, P)	Root mean square error (of Calibration, or Prediction)
RPD	Residual predictive deviance
%Russet	Overall percent russet
SCiO	Consumer Physics Molecular Sensor, handheld spectrometer
SD	Standard deviation
SPI	Starch pattern index (scale of 1-8)
SSC	Soluble solids content (%)
%Stained	Percent of apple cross section stained with I_2-KI solution
TA	Titratable acidity (mg malic acid/L)
TCA	Trunk cross sectional area (cm^2)
Weight	Weight of fruit at-harvest (g)
α	Alpha level of statistical significance

LIST OF FIGURES

LIST OF TABLES

Tables of Appendices

INTRODUCTION

Apple is an economically important crop globally, with an estimated 76.7 million metric tons produced in the 2016 to 2017 marketing year (USDA FAS, 2018). The USA is the third top producing country with 5.0 million metric tons following China and the European Union with 40.4 and 12.7 million metric tons, respectively (USDA FAS, 2018). The value of such a crop is worth $3.5 billion dollars in the USA alone (USDA NASS, 2017).

Fruit Quality

A significant increase in the percentage of US apple fruit crop utilized (kg) in the fresh market over the last few decades (Seetin, 2017) has put pressure on the industry to produce fruit meeting high quality standards that are demanded through the fresh market supply chain and ultimately by consumers (Kupferman, 2005). Some newer cultivars are also under trademark expectations requiring fruit sold under the trademark name to meet a certain high level of quality (Luby and Bedford, 2015). Breeding is the first step in meeting quality standards, since the potential quality of any fruit is limited by the cultivar characteristics. Visually appealing cultivars that are storage- and transportation-friendly have dominated the fresh market over the years. However, consumers have been driving portions of the market toward improved flavor (Baldwin, 2002), texture (F. Roger Harker et al., 2002; Harker et al., 1997), production methods (Cerda et al., 2012), and nutrition or health-promoting properties of fruit (Beghi et al., 2013; Endrizzi et al., 2015; A. Pissard et al., 2013).

The perceived qualities of apple fruit are complex and observed holistically by consumers (Charles et al., 2017). Some current methods to quantify apple fruit traits are at best approximations for the perceived traits values (Charles et al., 2019); despite this, the "proxy methods" are still utilized due to tradition, ease, and or a lack of methods that better represent the perceived traits. Other challenges that current trait measurement methods face is that they are destructive, time and labor intensive, and can be user-biased. Improved quality measurements methods are constantly needed to overcome

those acquisition challenges and to better reflect how quality is defined by the consumers (Magwaza and Opara, 2015).

Apple is a long-lived tree species with a lengthy juvenile period (Hanke et al., 2007). In order for breeders to evaluate fruit quality they must wait 4-7 years using current germplasm development methods (Wannemuehler, 2018). Genomic technologies have advanced and become exponentially cheaper (Andersen and Lübberstedt, 2003), however phenotyping technologies (characterizing fruit quality) have lagged in their cost and efficiency (Cobb et al., 2013; Else, 2017; Furbank and Tester, 2011). As soon as a plant has enough tissue for DNA extraction, genomic tools can then be used to generate trait predictions. In order to turn genetic data into information that can be utilized by breeders, phenotyping tools (sensors, scanners, penetrometers, sensory panels, etc.) must provide fruit quality data to establish the genetic association.

Researchers have been increasingly interest in chemometrics, an interdisciplinary field involving chemistry and multivariate data analysis (Wehrens, 2011). Near infrared spectroscopy is one of the many tools of 'chemometricians' (Pasquini, 2018), and has had much success in estimating fruit quality nondestructively (Nicolaï et al., 2007a). Some of the recent spectrometers are even handheld and portable (Teixeira dos Santos et al., 2015; Walsh, 2016a). This thesis describes testing the feasibility of using handheld spectrometers for estimating fruit quality traits, perhaps improving or supplementing current trait measurement methods.

Tree Quality

In order for growers to obtain high prices, they must constantly manage fruit quality within every cultivar and their respective cultivar diversity to represent prospective fruit quality trends. There is more to managing fruit quality than the ability to characterize fruit traits and improve thereon with breeding (Travers, 2013). Before consumers discover new cultivars and change preferences, the burden is on growers and pomologists (fruit researchers) to create 'best practices' for managing specific cultivars' tree quality, efficiency, and fruit quality. Individual cultivars each have their own nuances, and the

performance of one planting system may differ from location to location (Autio et al., 2017).

To manage fruit quality, tree quality and the orchard environment also should be evaluated (Musacchi and Serra, 2018). Quality in this respect involves tree health, yield, and the efficiency of the overall orchard system. Improvement of tree training systems and other cultural practices have allowed the production of more high quality fruit per acre than has ever been the case (Robinson et al., 2013). However, the classic methods used to characterize tree yield efficiency are antiquated. For instance, yield efficiency has typically been estimated by using a ratio of yield to trunk cross-sectional area (TCA), calculated from trunk diameter (Westwood and Roberts, 1970). This is a relatively quick and easy index, but provides no detail on canopy structure and may not approximate the relationship researchers seek. The connection of yield with TCA is variable with tree age and is even less associated after trees have filled their allotted space. At that point the canopy begins to operate in a semi-independent nature to the trunk (Kikuchi and Shiozaki, 2007) due to intense tree manipulation from pruning.

Just as precision agriculture has proved its worth with mapped yield metrics for many agronomic field crops (Zude-Sasse et al., 2016), emerging sensor and imaging technology (integrated into automated spraying, harvesting, weeding, and pruning machines) has become an exciting area of research for orchardists. Technology supporting precision horticulture is a newer field and is defined as site-specific management of horticultural crops (Colaço and Escolà, 2018). Horticultural production adds challenges because plants are often handled individually, and some perennial horticultural crops are grown in hedgerows, adding a third dimension to consider in order to fully characterize plant aspects that affect produce quality. Precision horticulture-driven sensing has improved such that most characteristics of plants can be calculated from images, including three-dimensional reconstructions. The state of the art in precision agriculture and horticulture is at a level of precision previously unavailable (Vázquez-Arellano et al., 2016).

In this thesis, I examine the feasibility of using low cost red-green-blue-depth RGB-D sensors to characterize apple trees. Imaging technology can provide data not available

from classic methods of measuring trees. Three dimensional tree reconstructions from RGB-D data could help estimate the yield efficiency of trees better than manual measurement methods.

Summary

Fruit quality is tied to tree quality and both need to be carefully recorded to make improvements thereon. Consumers purchase fruit based on visual qualities, but repeat purchases depend on internal fruit quality. Classic methods to measure fruit and tree quality are slow, destructive, subjective, and may not accurately approximate the trait. Improved methods to measure fruit and tree quality would benefit breeders, pomologists, postharvest horticulturists, and eventually the consumer. Breeders select fruit that have a good combination of qualities they predict that consumers will favor (Byrne, 2012; Callahan, 2003). Pomologists help growers create environments where quality fruit can be grown. Improved methods could even benefit wholesalers that manage postharvest fruit quality. Here I explore methods to improve the measurement of both fruit quality and tree efficiency.

Overall Goal

The overall goal of this thesis is to test the ability of new technology to phenotype fruit-quality-traits and tree architecture. Specific objectives for fruit quality are to explore the practical use a handheld NIR spectrometer by assessing the predictive ability of multivariate models for several fruit quality traits, and determining the limitations of the instrument's practical use. The objectives for tree architecture are to explore the horticultural application of data acquired from red-green-blue-depth (RGB-D) sensors by determining the correlation of image derived tree measures to manual measures and to examine whether image derived measures can explain variation in tree yield efficiency.

In this thesis, the first chapter will discuss fruit quality and the use of spectroscopic, measures. I will first provide an overview of some fruit quality traits that I think could benefit from technological advances in measurement methods. Then I will give a brief review of spectroscopy research with fruit quality and present an experiment exploring the practical use spectroscopy to estimate fruit quality traits. In chapter two, I discuss

areas of fruit tree research that I think could benefit from technological advances in assessing tree architecture. Then I will discuss the application of image-derived architectural metrics within the context of a rootstock evaluation study.

1 CHAPTER 1. FRUIT QUALITY TRAITS

Fruit quality is defined broadly in this thesis as the collective status of several fruit traits that can be measured by researchers to approximate what is perceived by consumers in terms of appearance, flavor, or taste (Callahan, 2003). Quality traits influence a fruit's ability to withstand the process of reaching the consumer and being attractive with pleasing taste and texture. Other studies have reported consumer perception of fruit quality (Baldwin, 2002; Harker et al., 2003; Péneau et al., 2006; Wismer, 2014). This thesis will focus on estimating parameters of fruit quality often measured by researchers. Below is an overview of some fruit quality traits and the typical methods used by apple researchers to quantify the traits.

Soluble solids content (SSC) is used to estimate the sugar content of fruit juice and is typically measured using either handheld or digital refractometers in units of ºBrix. Apple juice is expressed onto the refractometer lens, and the device computes the refractive index of the solution, or how much light is slowed down (or bent) in the solution (Mitcham et al., 1996). Although SSC is regularly used to estimate sugar content, the refractive index measures any dissolved solids in a liquid as compared to water. Simple sugars such as sucrose, fructose, and glucose are the main sugars in apple fruit (Aprea et al., 2017; Zhang et al., 2010), but other compounds, such as sugar alcohols, amino acids, organic acids, soluble pectins, and phenolic compounds also contribute to a juice's refractive index (Mitcham et al., 1996). SSC as estimated by ºBrix has been shown to be a problematic predictor ($r = 0.3$-0.1) of chemically determined sugars (Aprea et al., 2017). The perceived sweetness of fruit (Aprea et al., 2017; F. R. Harker et al., 2002) also has a relatively low correlation with SSC. Despite the drawbacks, SSC remains as an industry standard for fruit quality, perhaps due to the ease of measurement.

Dry matter content (DM) has recently been examined by researchers as a proxy for total carbohydrates and showed correlation to consumer liking in some cultivars (McGlone et al., 2003; Palmer et al., 2010). DM is measured as a percentage by taking the dry weight of a sample divided by the fresh weight before drying. This metric is important because of its relationship to insoluble carbohydrates, though it may not be able to distinguish

between insoluble and soluble carbohydrates (Travers, 2013). Robinson et al. (2017) noted that in 'Honeycrisp', DM content was not related to consumer liking; thus, there could be genetic differences in the constituents of DM patterns affecting the usefulness of the metric.

Starch pattern index (SPI) is a measure of maturity usually scored on a visual scale (Blanpied and Silsby, 1992). Fruit maturity is estimated because of its effects on quality traits (Watkins, 2003). Generally, as an apple fruit matures, the amount of starch in the flesh decreases. Fruit are scored for their SPI by cutting the fruit equatorially and dipping or spraying the exposed surface with an iodine-potassium iodide (I_2-KI) solution. Iodine binds to the starch in the fruit slice, turning it a dark purple color. The less stained the fruit slice is, the more mature it is expected to be. However, the conversion rate of starch to sugar may not be a linear process (Doerflinger et al., 2015) and it is dependent on cultivar.

Titratable acidity (TA) refers to the concentration of acids in apple juice, expressed as a percentage of a standard acid. Apple juice has five types of acids in detectable amounts (Lea, 1999), with the most abundant being malic (derived from the Latin name for apple: *Malus*) acid (Lea, 1999). An average value of malic acid has been reported to be approximately 5 $mg \bullet g^{-1}$ fresh weight, and is typically 90% of the total organic acids in a fruit (Ma et al., 2015). TA is measured by titrating juice with sodium hydroxide (NaOH) until the solution reaches a pH of ~eight. Apple TA is typically reported as milliequivalents of malic acid (Mitcham et al., 1996). The acidity level of apple fruit is an important driver of consumer preference (Harker et al., 2008b).

Total phenolics refers to the combined amount of polyphenols found in fruit flesh or skin, depending on the sampling method (Wargovich et al., 2012; Wojdyło, 2008). An increasing number of studies have focused on phenolic acids as well as on anthocyanin content of fruit, due to putative human health benefits (Wargovich et al., 2012). In apple, total phenolics can be commonly expressed as milligrams of gallic acid equivalent per gram of fresh weight and are commonly measured using the Folin–Ciocalteu method (Giovanelli et al., 2014; Audrey Pissard et al., 2013; Sánchez-Rangel et al., 2013).

Fruit firmness (FF) can be described in terms of force, with units in Newtons (N). The measurement can be taken using a handheld or digital penetrometer (Mitcham et al., 1996). The fruit skin is often removed at the test location, and then a probe is pushed through the fruit flesh. Automated digital penetrometers offer results that are more user independent than manual instruments. This allows the measure of several flesh firmness characteristics with one test (Evans et al., 2010). Fruit texture traits such as firmness, along with flavor, have the greatest influence on consumer preference (Harker et al., 2008b, 1997).

Crispness has been cited as the most salient texture trait for apple fruit (Di Guardo et al., 2017; Hampson et al., 2000; Manalo, 1990). Yet at the same time apple fruit crispness seems imponderable (Chang et al., 2018; Mann et al., 2005; Trujillo et al., 2012). Crispness has been defined as the fracturability of the flesh tissue (Mann et al., 2005), or as the sound and sensations apparent as a bite is taken into the fruit (Demattè et al., 2014). Crispness is related to perceived freshness by consumers (Péneau et al., 2006). Even though the accuracy has been debated (Mann et al., 2005), objective methods of estimating perceived crispness have been developed using models from digital instruments (Chang et al., 2018; Evans et al., 2010; Schmitz et al., 2013). Since the definition of crispness remains debated, no industry standard exists to express it.

Flavor/Taste is the collective perception of ascribed traits like juiciness, sweetness, sourness, fruitiness, and aroma. Some volatiles, which are components of flavor and aroma, can be measured in a lab setting (Rowan et al., 2009). The consumer palate is the driving factor to repeat purchases of cultivars in groceries (F. Roger Harker et al., 2002; Kupferman, 2002). The perception of flavor can be quite subjective in nature. Nevertheless, ascribed flavor/taste traits are best measured by panelists (Corollaro et al., 2014) in sensory evaluations or by trained personnel (Hampson et al., 2000).

Appearance includes several traits, such as fruit color, size and freedom from imperfections caused by disease or russeting that are highly important to growers (Iglesias et al., 2008) and consumers (Saure, 1990), and therefore to breeders. Consumers prefer fruit that are highly colored (Saure, 1990). Various methods have been used to

assess color. As part of the RosBREED project (USDA - Specialty Crop Research Initiative) a standardized protocol (Evans et al., 2012) for assessing apple fruit color in breeding germplasm considers: Gcolor, Ocolor type (blush pattern), and %Ocolor (blush coverage).

The majority of these traits are not static throughout growth or storage. Fruit quality needs to be controlled along each step of the apple supply chain for retailers to sell fruit of reliable quality (E.Watada et al., 1989; F. Roger Harker et al., 2002).

In order to select a genotype suitable for release as a commercial 'cultivar', breeders must perform a variety of tests. For example, breeders screen individuals for the desired visual characteristics, perform subjective taste tests, assess the chemical composition by wet laboratory tests, and place samples in long storage trials.

These methods can be destructive, susceptible to bias, and require extended processes (Diaz-Garcia et al., 2016; E.Watada et al., 1989; Iezzoni et al., 2017; Peace, 2017; Peirs et al., 2002). Laboratory tests sometimes must be done far away from the orchard where many decisions need to be carried out in regards to the fate of a genotype and its fruit.

Phenotyping, or the characterization of an individual's traits, has been considered a bottleneck to the application of plant genetic data (Cobb et al., 2013; Else, 2017; Furbank and Tester, 2011). Ideally, improved phenotyping methods would (1) be done without destructive sampling, (2) generate data throughout fruit growth and development (allowing repeated measures), (3) provide expeditious information outside of the typical laboratory setting, and (4) relay information significantly correlated to the specific trait in question, to provide a breeder greater confidence and efficiency in culling decisions.

Improving the prediction of trait values (over time or at specific points) is especially needed when the trait cannot be easily measured. As improved phenotyping methods are established, it is important to recognize the risk associated with estimating trait values in place of measuring trait values. Similarly, there are risks in marker-assisted breeding where marker tests estimate an individual's genotype at loci that should affect phenotype. Genomic tools are used because of time, labor, or space constraints on observing the phenotype. Genomic tools may provide the only source of information for culling at a

particular point in the breeding stage. Selection based on genomic tools is not exempt from errors, but breeders attempt to minimize errors by using data from many situations (years, environments, and genetic backgrounds) to inform the selection model.

The improvements of one specific technology, spectroscopy, have the potential to increase the speed and accuracy of estimating fruit trait values.

2 SPECTROSCOPY USE IN APPLE

Research in fruit spectroscopy has been ongoing for decades (Birth and Norris, 1965; Walsh, 2016a). Spectroscopy involves measuring reflectance, absorbance, or transmittance of light over a range of wavelengths in a sample. Spectroscopy can be nondestructive depending on the sample preparation methods, and may allow repeated measurements of the same sample. One set of spectra can be used to predict multiple traits. Interest in using spectroscopy to measure fruit traits has been increasing since its inception, as evidenced by the growing number of research articles published in peer-reviewed journals (Lin and Ying, 2009).

Spectral data, collected from large benchtop spectrometers, have successfully been used to predict the level of several chemical characteristics of apple fruit, predominantly soluble solids content (Lu et al., 2000; Moons et al., 2000; Murakami et al., 1994; Peirs et al., 2000; Temma et al., 2002b; Ventura et al., 1998; Zude et al., 2006). Models of processed spectral data have been used with varying degrees of success to predict other traits, like dry matter, titratable acidity, firmness, and starch pattern index (Harker et al., 2008b; Lu et al., 2000; McGlone et al., 2003; Moons et al., 2000; Murakami et al., 1994; Peirs et al., 2000; Zude et al., 2006; Zude and Herold, 2002). The status of quality traits influenced by storage, such as SSC, total starch content, and dry matter have been predicted using pre- and/or post-storage spectra by several researchers (Giovanelli et al., 2014; Kumar et al., 2015; McGlone et al., 2003; Pérez-Marín et al., 2009; Travers et al., 2014; Zude et al., 2006). As spectral data is modeled to predict fruit traits accurately, spectrometers potentially provide a quick source of information that otherwise would be

observed in a subjective manner, such as traits that are visually scored; or traits that can only be destructively measured in a sensory, instrumental, or wet lab setting.

Although laboratory-based machines have provided useful information on apple fruit characteristics, recent research has focused on acquiring similar spectral information in a field setting with portable devices. Many authors claimed to have used 'portable' near infrared (NIR) spectrometers. The convenience for many of these 'portable' spectrometers is limited, however, and insufficient for outdoor use. Most of the described 'portable' spectrometers are composed of several boxed components connected by cables, and require a connection to a notebook or laptop computer (Beghi et al., 2013; Herold et al., 2005; Nagy et al., 2016; Paz et al., 2009; Sun et al., 2009; Temma et al., 2002a; Torres et al., 2016; Travers et al., 2014; Ventura et al., 1998; Yuan et al., 2015; Zude and Herold, 2002). 'Handheld' is the term used in the literature to describe a single component instrument that does not depend on wired connections in order to acquire spectra. Few studies have used handheld spectrometers to predict apple fruit traits (Das et al., 2016; Kaur et al., 2017; Kumar et al., 2015; M. Li et al., 2018; Seifert et al., 2014; Zhang et al., 2019). Two of these studies (Das et al., 2016; Seifert et al., 2014) describe the development of a new handheld spectrometer using relatively small samples for preliminary testing of the commercially unavailable devices.

Spectrometer use in an orchard environment (Beghi et al., 2013; Herold et al., 2005; Kumar et al., 2015; Torres et al., 2016) has also received little attention. Herold et al. (2005) tested the usability of a custom 'portable' spectrometer to observe spectral changes throughout growth using 101 'Elstar' fruit from two trees. Spectral sampling was done on the same fruit through 12 sampling dates over 7 weeks. More recently, Torres et al. (2016) used a more 'portable' spectrometer to monitor sun injury from overexposure of 'Granny Smith' fruit. They collected spectra at 14 different sampling dates across 14 weeks for 500 fruit from 25 trees.

Beghi et al. (2013) also used a custom 'portable' spectrometer built from components housed in a backpack. It was operated with a fiber optic cable and laptop computer in each hand. The aim was to predict the level of several phytochemical traits (ascorbic acid,

total polyphenols, anthocyanins, flavonoids, and carotenoid content) as well as many of the quality traits typically modeled with NIR spectroscopy, like acidity, firmness, soluble solid content, and chlorophyll content. Beghi et al. (2013) sampled 20 fruit from 'Stark Red Delicious' and 'Golden Delicious' fruit over 6 and 8 weeks, respectively, for a total of 280 fruit. Each trait had varied success when modeling fruit from separate cultivars, with coefficients of determination (R^2 of cross validation) of 0.09 for total phenolics and 0.86 for chlorophyll and flavonoids.

Kumar et al. (2015) used spectroscopy to evaluate breeding populations. Over three years, spectra and trait data were collected in a lab setting on approximately 1400 fruit from 279 apple genotypes representing 67 full-sib families. The samples were pooled at the seedling level and the models were used to predict DM, firmness, and SSC at different storage periods. An additional test involved 100 unique genotypes that they scanned in the orchard setting. The postharvest prediction values were high for both SSC and DM (both $r > 0.9$). Of all the published studies conducted in an orchard setting, only Kumar et al. (2015) tested a 'handheld' spectrometer (Nirvana, Spectronics Inc. Australia). Handheld spectrometers are the result of recent improvements in computational power, memory storage, sensing capabilities, portability, and weight. These improvements are exciting for all types of plant phenotyping research (Araus and Cairns, 2014; Fiorani and Schurr, 2013; Lee et al., 2010).

While technology has improved the acquisition rate of phenotypic data, the ability to apply the information in 'real-time' is still lagging. Plant breeders could benefit from instant phenotypic information to make decisions in the field about whether to select individuals for extended trials, storage, or culling. Although technology has improved and become more portable, the quality of the data still must fulfill the "practical need of a given application"(O'Brien et al., 2012). In breeding, this practical need depends on the trait and how the trait is used to rate an individual genotype's fruit. Predictions may only need to have R^2 values of ~0.60 in order to classify fruit into categories for non-critical traits like vitamin C content (A. Pissard et al., 2013).

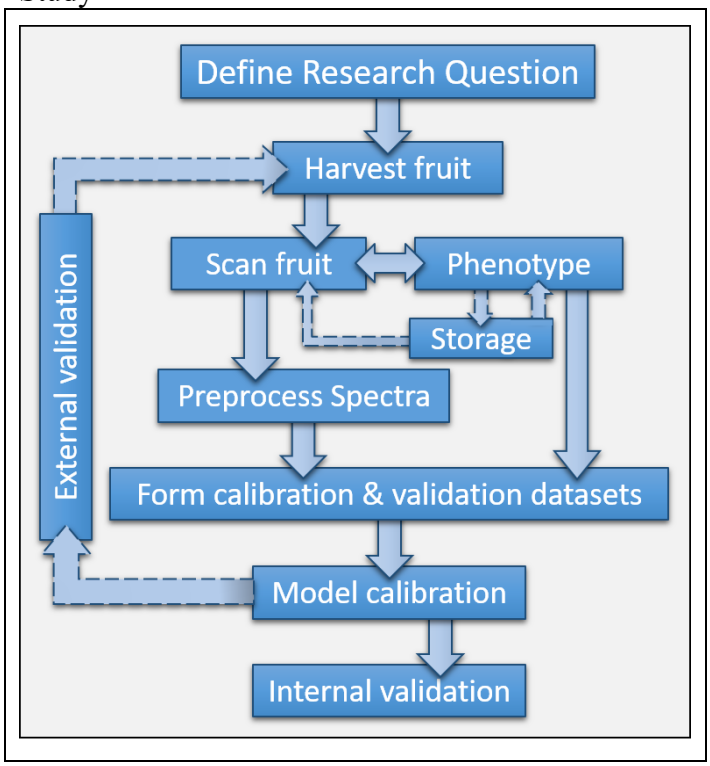
Predictive models using spectral data require a technical and extended process of data pre-processing, statistical modeling, and model validation (Nicolaï et al., 2007a). Yet many spectrometers provide no method of incorporating models into the user interface of the machine. Handheld spectrometers that have a response screen or that interact wirelessly with mobile devices have the potential for models to be built into their software that is often open source (Rateni et al., 2017). In order to make critical decisions instantly away from a laboratory setting, handheld devices need to operate in a user-friendly fashion, and immediately provide prediction results (perhaps for several traits at once).

Spectral Models

Incorporating robust models into handheld devices is crucial to the successful implementation of the technology by researchers, growers, and industry professionals. Besides unique software interfaces in the spectral model development workflow, data from handheld spectrometers is not different from portable or bench-top spectrometers. The typical process used to obtain models in fruit spectroscopy studies is diagrammed in Figure 2-1 (see similar flow charts in Lovász et al. 1994; Schmutzler et al. 2013). Almost every published NIR spectroscopy study has followed these steps.

As with any research endeavor, the hypotheses to be tested guide the researcher on which questions to ask, the types of data needed, the experiments to perform, and the approach for which analyses need to be executed. Most NIR spectroscopy research has been in the form of feasibility studies, for which the only goal is to prove feasibility for a specific trait, spectrometer, or calibration method (Williams et al., 2017). Few studies have been published for the industrial or commercial application of NIR spectroscopy. Yet another example is NIR spectroscopy geared at exploring a biological question, such as the composition of DM, its NIR absorption range (Travers et al., 2014), and whether DM correlates with post- versus at-harvest SSC (McGlone et al., 2003).

Figure 2-1. Flowchart Typical Fruit Spectroscopy Study



For a single point of observation, fruit samples are generally harvested directly from an orchard where growing conditions and other external variables can be recorded. Fruit are also often collected from vendors for this type of analysis. For long term analysis, to capture seasonal changes like fruit maturity, researchers either complete several harvests (Peirs et al., 2000) or follow tagged fruit while harvesting neighboring fruit for direct measurement, or "ground truthing", throughout the growing season (Herold et al., 2005; Torres et al., 2016). Following tagged fruit allows the spectral analyses to be performed at an individual fruit level, but relies on ground truthing from neighboring fruit; therefore, the utility of the results depend on the sample size being large enough to capture variation among fruits. Herold et al. (2005) found high variability in weekly spectral measurements among trees, which generally decreased as the fruit matured. Management decisions are never made based on a single fruit, therefore the extra work of following tagged fruit is most appropriate when researching a phenological or biological question rather than trying to estimate the condition of a crop.

The genetic material in the sample also influences the spectral variability. Most apple spectroscopy studies have utilized only a few cultivars (normally 1-8), and generally predict traits based on cultivar-specific models (Peirs et al., 2005). Only Kumar et al. (2015) and Pissard et al. (2013) have done experiments with diverse breeding germplasm, studying 279 and 32 genotypes, respectively. Little is known about spectral variation among cultivars besides the observation that the accuracy of models is impacted. This is partly because researchers' sampling plans have been limited by sample size when the number of cultivars increases. In these cases, the sample sizes per cultivar are not large enough to capture a cultivar's potential variation at any one time point (Peirs et al., 2005). Peirs et al. (2003) suggested that cultivar effects are similar to seasonal variation, and Pissard et al. (2013) suggested that spectral models for breeding populations would be best for placing cultivars into classes rather than predicting precise concentrations. However Zhang et al. (2019) found that SSC and DM models were improved and less sensitive to noise by including more cultivars, due to the increased range of trait values that was observed between the cultivar samples. The robustness of a predictive model may be improved by including more phenotypic and genotypic variation in the training datasets; however, the influence of genetic diversity on NIR absorption and scattering (affecting the predictive accuracy of a model) is not answered well in the literature.

Bench-top and portable spectrometers typically have boxes or seals around the sensors to avoid entry of any external ambient light (Herold et al., 2005; Nagy et al., 2016; Ventura et al., 1998). As apple fruit are complexly-organized and heterogeneous by nature (Dever et al., 1995; Harker et al., 2008a, 2005; Peiris et al., 1999), it is common for multiple scans to be made on an individual fruit (Lovász et al., 1994; Rowe et al., 2014; Travers et al., 2014). Scans are then averaged for each individual.

After fruit are scanned, they are phenotyped following conventional methods to measure the traits of interest. This step is sometimes referred to as collecting the 'ground truth' since the traits are observed (phenotyped) directly instead of estimated from spectra. Since an individual fruit has a heterogeneous composition, and it is assumed that spectra react to subtle changes in fruit composition, many researchers collect the portions of fruit that had been scanned for the ground truth measurements (Mendoza et al., 2014). Ground

truthing also occurs postharvest to understand changes that occur throughout storage, as apple fruit are stored for long periods commercially, to provide a continuous annual supply of fruit in retail stores (Giovanelli et al., 2014; Kumar et al., 2015; Lovász et al., 1994; McGlone et al., 2002a).

Preprocessing of data is important to avoid spurious correlations in spectral models. Preprocessing spectral data removes noise and baseline distractions, scales data to a standardized measure, and deals with additive and multiplicative effects (Nicolaï et al., 2007a). While not every type of preprocessing is necessary, choices are made based on operator preference, popularity among researchers, and type of spectrometer output (Nicolaï et al., 2007a). The most popular preprocessing procedure is multiple scatter correction (MSC) and derivative transformation with the Savitzky–Golay algorithm to an order of two (Nicolaï et al., 2007a).

How the spectral and phenotypic data are combined and organized depends on the research question. The same data can be combined in several manners to answer different questions. For example, instead of predicting a trait at the time of harvest based on harvest spectra, McGlone et al. (2003) used harvest spectra to predict post-storage SSC. The main body of NIR spectroscopy research in apple has focused on the general ability of spectral data to predict fruit traits and qualities, some more specific research areas include:

- Different methods of collecting fruit and spectral samples, and sample preparation (Lu et al., 2000)
- Different methods of ground truth collection (Peirs et al., 2000)
- Comparing different spectrometers (Kaur et al., 2017; Paz et al., 2009)
- Different statistical modeling methods (Martínez Vega et al., 2013) and
- Sorting fruit into categories (Moons et al., 2000)

The number of analyses are not limited by a project's initial research question, but the research question guides proper protocol development for downstream processing.

To avoid overfitting, bias, and inflation of model statistics, the best practice is to employ cross validation. The purpose of cross validation is to estimate a model's prediction error on data that was not used to train the model (Hastie et al., 2009). Cross validation sets a portion of the sample aside while fitting a model, and then use the model to predict the "left out" sample's characteristics. This process is repeated until every portion has been left out, and the combined variance of each portions' residuals is calculated. Leave one out (LOO) is a common method of cross validation, but could lead to over-optimistic models and less accurate estimation of the expected error (Giovanelli et al., 2014; Hastie et al., 2009). Multifold cross validation was used by Kumar et al. (2015) to predict seedling performance in a breeding program. Kumar et al. (2015) had five fruit from individual seedlings that were grouped together, and each seedling was left out one at a time to develop SSC and DM prediction models. Giovanelli et al. (2014) used fivefold cross validation groups.

A model can be compared to similar models by its root mean square error of prediction (RMSEP) and root mean square error of cross-validation (RMSECV).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_p}(\hat{y}_i - y_i)^2}{n_p}}$$

Kumar et al. (2015) found seedling level root mean square errors of prediction (RMSEPs) for soluble solids content and dry matter that were approximately 80% and 70%, respectively. It is common to perform internal cross validation, to subsample the data taken from the same set in the same year (Martínez Vega et al., 2013), however external validation (using samples from different orchards or years) would be the best proof of a model's robustness.

A general overall model accuracy statistic is the residual predictive deviance (RPD), or the ratio of the standard deviation of the response variable to the RMSEP or RMSECV. McGlone & Kawano (1998) stated that RPD can more directly indicate the performance of a model than the RMSEP or R^2 statistics when the variance explained or error are used alone. RPD can theoretically be considered equal to both RMSEP and R^2 for large data sets that are normally distributed (McGlone and Kawano, 1998). Higher RPD values

indicate better predictive performance. Nicolai et al. (2007) suggested ranges for fruit NIR model comparisons (shown in Table 2-1).

Table 2-1. Suggested* accuracy ranges for RPD used in ranking spectral models

RPD	Accuracy
1.5-2.0	Discriminate low to high
2.0-2.5	Coarse quantitative predictions
2.5 or greater	Good to excellent predictions

*(Nicolaï et al., 2007a)

In order to find the most accurate spectral models, researchers often go through lengthy experimentation with multiple modeling methods, testing prediction errors (Kumar et al., 2015; Martínez Vega et al., 2013; Moons et al., 2000; Torres et al., 2016; Yuan et al., 2015). While accurate results can be obtained from singling out the statistical method that has the highest predictive ability, the accuracy of a model with the collected training data should not be the only focus when models are built. Nicolai et al. (2007) suggest that an equal amount of attention should be placed on the robustness of a model for predicting data unseen by the model, and that there may be a trade-off with prediction accuracy. Model robustness is related to how insensitive a model is to changes in external factors, like the use of a model on different spectrometers or fruit samples of different years or origins.

3 STATUS OF OUTDOOR AND PORTABLE SPECTROSCOPY

The objective of Kumar et al. (2015) was to test harvest time spectral models (built from 279 genotypes) to predict poststorage traits within a breeding population (100 genotypes). They used a handheld instrument (Nirvana, Spectronics Inc., Australia) to collect spectral data indoors on fruit over three years to create models. The models were validated by predicting after-storage fruit quality of a fourth year's fruit scanned on the tree. RPD values (that I calculated using their data) for DM were modest, between 2.27 and 2.31, and much lower for SSC, 1.92-1.15. SSC RMSEP values on a seedling level were between 0.71 and 0.90, and on a fruit level between 1.04 and 1.15.

Kaur et al. (2017) created NIRS models predicting DM using two F-750 (Felix Instruments Camas, USA) handheld instruments, along with three other handheld-type instruments, two Nirvana (Spectronics, Inc. Australia), one H-100C (Sunforest, Korea), and one SCiO (Consumer Physics, Israel), comparing them to an in-house built benchtop instrument as a control. Their objective was to evaluate the three types of instruments capable of on-tree in orchard use; however, the spectra were not acquired outdoors. DM RPD results ranged from 1.57 to 2.30 for the handheld instruments and 3.88 for the benchtop control. The RPD was considerably different (0.33) between the two F-750 instruments. ANOVA results comparing instrumental performances showed overlapping ranks between the different instruments. Kaur et al. (2017) suggested that, "Choice of instrument then may likely come down to matters other than those of instrument precision and accuracy, not addressed here, such as reliability and robustness in field operation under conditions of varying ambient light and/or temperature, etc.".

Li et al. (2018) tested the feasibility of predicting several traits for several fruit crops using the SCiO spectrometer. All spectra were collected in a lab on a single apple cultivar 'Braeburn'. The RPDs reported by Li et al. (2018) for apple fruit were low; 1.7 for DM, 0.9 for SSC, and 0.7 for fruit firmness .

The objective of Zhang et al. (2019) was to test the F-750 spectrometer, compare both individual cultivar and multi-cultivar models, and to use data from another year to externally validate models. In the first year, fruit of eight cultivars were harvested from research orchards, and in the second year, fruit were additionally harvested from commercial orchards. The external validation RPD (my calculation) values for individual cultivar models ranged from 1.06 to 2.04 for SSC and 1.01 to 2.03 for DM. For multi-cultivar models, the RPD for SSC was 3.13 and 3.62 for DM.

Research on the utility of benchtop spectrometers is fairly mature and established, but the utility and flexibility of currently available portable and handheld spectrometers is still far from being conclusively proved. One aspect of handheld spectrometer performance that still needs assessment is how well trait models perform when spectra are collected in sub-optimal lighting or temperature situations. Using handheld spectrometers to assess

breeding populations and commercial orchards outdoors is an exciting foreseeable application (Kumar et al., 2015; Audrey Pissard et al., 2013).

One objective of this thesis was to assess the overall performance of models derived from F-750 spectra obtained under indoor laboratory compared to outdoor orchard conditions, using cultivars available in University of Minnesota apple germplasm. Experiments examined the extent to which: 1) daily and seasonal temperature fluctuation could be controlled, 2) the ability of spectra to capture sufficiently useful information from a single point on a whole apple fruit, and 3) the ability of the handheld spectrometer to control noise present in an outdoor scanning.

In the main experiment, fruit from 15 cultivars were harvested, and spectral data obtained at-harvest and after 10 weeks of refrigerated storage were compared to visual, physical, and chemical data obtained from the fruit. The specific goals of this experiment were to 1) verify the NIR predictability of traits described in the literature using a handheld spectrometer (Felix Instruments, F-750) in multiple settings, 2) discover other traits that could be predicted using the handheld NIR spectrometer, and 3) explore limitations of handheld spectrometers.

Another objective was to determine the difference in utility between two handheld spectrometers (Felix Instruments F-750 and Consumer Physics SCiO) used in an outdoor setting to characterize several apple fruit traits of several cultivars.

4 CHAPTER 1. MATERIALS AND METHODS

4.1 Detection of Several Fruit Quality Traits in 15 Apple Cultivars Using NIRS (2016)

4.1.1 Fruit Samples and Harvest Protocol

In 2016, fruit of 15 apple cultivars (see Table 4-1) were harvested from the University of Minnesota (UMN) Horticultural Research Center in Chanhassen (44°52'03.8"N 93°38'03.8"W 310m). Each cultivar was harvested twice usually one week apart. At each

time point, 50 fruit per cultivar were harvested into polypropylene bins, avoiding fruit with noticeable defects. The fruit were then directly transported to the UMN St. Paul campus for further analyses.

Table 4-1. Harvests and sample size by processing day per cultivar for 2016 experiment

Cultivar*	Harvests	Post 10 Week Storage			
		Day 0	Day 1	Day 7	Day 7 Discard
Chestnut	2	41	20	20	25
Cortland	2	44	20	20	20
Fireside	1	19	10	10	10
Frostbite‡	2	40	20	20	20
Haralson	2	40	20	20	20
Honeycrisp	2	40	20	20	20
Honeygold	2	40	20	20	20
Keepsake	2	40	20	20	20
McIntosh	2	39	20	20	20
Minneiska$^{\Delta}$	2	40	20	20	20
MN1942	2	40	20	20	19
MN55$^{\chi}$	1	30	0	35	3
Wildung$^{\phi}$	2	40	20	20	20
Sweet Sixteen	2	39	20	20	20
Minnewashta$^{\psi}$	2	40	20	20	20
	Total	572	270	305	277

Total number of fruit harvested in 2016 was 1424

*Trademark names: ‡= Frostbite™; Δ= SweeTango®; χ= Rave®/First Kiss®; ϕ= SnowSweet®; ψ= Zestar!®

4.1.2 Spectral Measurement Protocol

A handheld spectrometer, the F-750 Produce Quality Meter (Felix Instruments Camas, WA, USA), was used to collect spectral data in 2016. The F-750 uses active illumination from a xenon tungsten lamp and an automatic shutter for calibration/reference scans (the back of the shutter facing the sensor contains a standardized white paint). The instrument's shutter design compensates for ambient light and dark current by taking measurements with the 'light source off shutter closed' and the 'light source off shutter open' (Felix Instruments, 2015). The spectral range starts in the near-UV at 310 nm, includes the visual spectrum, and includes a portion of the near infrared up to 1100 nm,

with a sampling interval of ~three nm. The optical geometry of the F-750 is designed for interactance (i.e., sensor is on the same side as the sample) measurements, rather than reflectance (sensor at an angle to sample) or transmission (sensor measuring light through a sample). The instrument's software calculates the diffuse reflectance value by "subtracting the light reflected from the reference shutter from the light reemitted by the subject". The reflectance values are then used by the software to calculate the reciprocal absorbance:

$$absorbance = Log(\frac{1}{Reflectance})$$

and the "first and second derivative spectra by applying Savitzky-Golay (Savitzky and Golay, 1964; Schafer, 2011) coefficients" (Felix Instruments, 2015).

The number of scans at a single position of the fruit can often be adjusted by the spectrometer's software settings and should be set according the signal to noise ratio of the particular spectrometer (Greensill and Walsh, 2000; Nicolaï et al., 2007b). The setting used in this experiment was the average of two scans at a single position, except that the setting was inadvertently changed after the first 17 % of observations at-harvest, (i.e., the first 17% observations had four scans to average) and the first 20% of postharvest observations (also had four scans to average). This change in settings likely happened when the installed battery cap had a loose connection and the memory card had to be formatted, with a new copy of the instrument settings transferred over after the loss of battery connection. A new battery cap design solved the issue of connectivity.

On the day of harvest, each fruit was randomly numbered from 1-50 using a black permanent marker. Circles (39 mm diameter) were drawn on two sides of the fruit 180^{o} apart along the center equatorial line, once on the sun-exposed side and the other on the opposite (shaded) side of the fruit. The fruit were then scanned, with the spectrometer (angled away from the operator) in one hand and the fruit in the other, on both sides (sun and shade) in two environments - once in a lab under fluorescent lighting and once outside beneath a crabapple tree to simulate orchard light conditions. The surface

temperatures of six fruit in both outdoor and indoor scanning situations were recorded using an infrared thermometer (62MAX, Fluke).

4.1.3 Fruit Trait Data Collection Protocol

After scanning each side of the fruit in both environments (outdoor and indoor), each fruit was visually scored for the following skin traits: Ground-color (Gcolor), Over-color type (Ocolor type), over-color color (Ocolor), percentage of Ocolor (%Ocolor), percentage of russeting (%Russet), and location of russeting. Table 4-2 describes the various categories for each trait. The color and color type categories were considered mutually exclusive.

Individual whole fruit weight was recorded (SPE402 scale, Ohaus; SD = 0.01 g), after which fruit (numbered 1-30) were placed in stackable polypropylene bins for cold storage at ~2 °C ± 1 ° C for 10 weeks. The remaining fruit (numbered 31 to 50) were destructively processed within 1 day of harvest.

Flesh firmness and related texture measurements were captured from a computerized penetrometer (Mohr® Digi-Test MDT-1; Mohr and Associates, Richland, WA) using an 11 mm probe. Firmness variables were measured once at a small area (peel had been removed using a mandoline) along the fruit equator between the sun-exposed and shaded sides. See Table 4-3 for a description of Digi-Test variables. The M1 variable corresponds to the Magness-Taylor firmness measurement commonly used by the apple industry.

Starch pattern index was evaluated after texture analysis using a ~12 mm thick transverse section cut from the equator of the fruit. A sheet of butcher paper was labeled and numbered to identify the individual 20 fruit from each cultivar's harvest. Each section was dipped in potassium iodine solution (8.8 g KI + 2.2 g I_2/liter water) and allowed to rest on the paper for approximately 5 min. The percentage of stained flesh area was scored visually following a 'McIntosh' 8-point scale (Blanpied and Silsby, 1992).

Table 4-2. Skin traits that were visually scored on apple fruit

Skin trait	Phenotype/rating	Value
Gcolor	green, pale green, yellow-green, pale-yellow, or yellow	(1-5)
Ocolor type	blush, blush-stripe, stripe-blush, or stripe	(1-4)
Ocolor	orange, orange-red, pink, red, or dark-red	(1-4)
%Ocolor	intervals of ten*, average rating from two judges	0-10
%Russet	intervals of ten*, average rating from two judges	0-10
Location of russet	stem cavity, shoulders, body, lenticels, and/or calix	(0/1) per location

*The intervals were defined as 10% = 1, 20% = 2, etc.

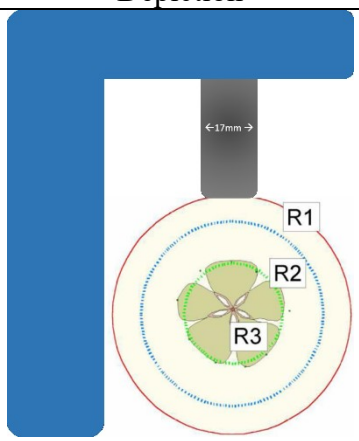
Percent dry matter was assessed by drying a peeled and cored sample of the distal portion of the fruit flesh. The starting average wet weights of the samples were between 10 and 20 g. The samples were placed in aluminum weigh boats to dry in convection ovens (multiple oven versions) at 60 ˚C for a minimum of 48 hours. The formula used for percent dry matter was

$$\%DM = \frac{cd - (cw - w)}{cw - (cw - w)} * 100$$

where *cw* is the combined weight of matter was assessed by drying a peeled and cored sample of the distal portion of the fruit flesh. The starting average wet weights of the samples were between 10 and 20 g. The samples the wet sample and weigh boat, *w* is wet weight of sample, and *cd* is combined weight of the dry sample and weigh boat.

Following dry matter sampling, the remaining portions of fruit were juiced using a Juice Extractor (Waring PJE401, Conair Corporation, Stamford, CT). Pulp caught in a paper filter was discarded. Labeled vials of juice from each individual fruit were frozen at -80 ˚C until chemical analysis was performed.

Table 4-3. Definitions of texture and firmness related variables measured using MDT-1 Mohr® Digi-Tester

Depiction	Abbreviation	Description
 Adapted from Schmitz et al. (2013)	M1*	Max Pressure (N) in first 8 mm of flesh (≈Magness-Taylor & EPT-1)
	A1*	Average Pressure in first 8 mm (N)
	M2‡	Max Pressure 8 mm to Core (N)
	A2‡	Average Pressure 8 mm to Core (N)
	C0	Creep Test Distance (mm)
	E2	Pressure just above the Core (N)
	OMH	Overall Maximum Hardness (N)
	OAH	Overall Average Hardness (N)
	Cn	MDT-2's Crispness test
	QF	Quality Factor (Uses: M1, A2, E2, C0, & Cn)
	Diameter	Fruit Diameter (mm)
	Weight	Weight (mass, g) of Fruit at-harvest

* Region 1 is from the fruit surface to ≈8 mm inward;
‡ Region 2 is from ≈8 mm to ≈ 3/5ths of the fruit radius.

A digital handheld refractometer (PAL-1 3810, Atago Co. Ltd., Tokyo, Japan; ± 0.2 % Brix) was used to estimate the SSC in the thawed juice. An automated titrator (T50A Mettler Toledo, Columbus, OH) was used to measure TA, recording the final volume of NaOH to represent the total amount of acid in the juice, which is reported as malic acid equivalents per liter. The pH of each raw juice sample was recorded using a titrator probe (DG 115-SC, Mettler Toledo, Columbus, OH). Total phenolics content was measured following the Folin-Ciocalteau method (Singleton et al., 1999), using absorbance at 765 nm instead of 760 nm, and absorbance at 300 nm was also recorded (similar to the work of Beghi et al. (2013) that used 280 nm), with only minor deviations in the calibration volumes.

4.1.3.1 Postharvest Evaluation

In the 10th week of cold storage, stored fruit were removed from the cooler and placed at room temperature (~ 23°C) for one day. Presence of any storage disorders and fruit decay was noted, and all fruit were scanned at the same positions marked on the day of harvest. Fruit weight after storage was measured, and Gcolor was re-scored. Ten sound fruit were processed with all destructive measures described above with the exception of SPI. The

remaining 20 fruit were stored at room temperature for an additional 6 days. After 7 days from removal from cold storage, all remaining fruit were scanned again, after which 10 sound fruit were destructively processed and the remaining unneeded 10 fruit were discarded.

4.1.4 Data and Analysis

4.1.4.1 Fruit Trait Data Overview

Nondestructive data were collected from 1424 fruit and destructive trait data were collected on 1147 fruit. Sixteen of the 25 traits were quantitatively measured: A1, A2, SSC, C0, MDT crispness, diameter, DM, E2, M1, M2, OAH, OMH, pH, QF, TA, and weight. The nine categorical or qualitatively measured traits included the presence of russeting on the body of fruit (BRusset), fruit damage, Gcolor, presence of lenticel russet (LRusset), Ocolor, Ocolor type, %Ocolor, %Russet, and SPI. Only weight, skin-related traits, and damage scores were measured nondestructively. Fruit diameter was measured at the time of texture analysis. Descriptive plots of each of the 25 traits are shown in Appendix A.

Outlier Detection and Removal

Outliers were detected in several variables. A complete list of detected outliers and reasons for their removal are provided in Table 4-4 (Olivieri, 2015; Williams et al., 2017).

Table 4-4. Outliers in Ground Truth Variables from 2016 experiment

Variable	Number Removed	Reason for Removal
C0	67	Machine error/small fruit (all 'Chestnut' crabapple fruit and fruit with creep distance above 4 mm were removed)
Cn	105	Machine error/small fruit; 5 observations above the value of 750 units were removed
M1	87	Visually deviant from sample distribution, values below 10 N or above 155 N were removed), 6 observations were due to machine error
A1	87	Visually deviant from sample distribution, values below 10 N or above 155 N were removed), 6 observations were due to machine error
M2	47	Machine error; Visually deviant, 5 observations below 10 N were removed)
A2	47	Machine error; Visually deviant, 5 observations below 10 N were removed)
OMH	10	Machine error; Visually deviant, 7 observations below 10 N were removed)
OAH	12	Machine error; Visually deviant, 9 observations below 10 N were removed)
E2	13	Machine error; Visually deviant, 4 observations below 10 N were removed)
QF	8	Machine error; Visually deviant, 6 observations below negative 200 units were removed
SSC	15	Sample contamination/transcription error; Visually deviant, 1 observation of less than 5 % was removed
DM	34	Transcription error; Visually deviant, 10 observations with; values less than 9 % or greater than 30 % were removed.

Outliers were easily detected for the variables originating from the MDT-1 when scatter plots were examined, and were designated as due to machine error (outliers that did not adjust the y-axis scale were included in the distribution plots). Machine errors occurred relatively infrequently with fruit over ~60 mm diameter. When errors did occur in such fruit, other simultaneously measured variables were also erroneous. Outliers were removed before performing any modeling of traits. The default settings of the MDT-1 penetrometer are best suited for measuring firmness on normal sized (60-90 mm in diameter) apple fruit. Two cultivars, 'Chestnut' and Frostbite, typically had fruit that were smaller than the limits of what the machine could measure without adjusting the

probe extension distance. When modeling several MDT-1 variables, data from 'Frostbite' and 'Chestnut' were not included in the analyses.

4.1.4.2 Spectral Data Preprocessing

Spectral outliers were selected subjectively by visually analyzing spectral line plots by cultivar, Table 4-5. Individual fruit with spectral values that were out of place compared to their companion fruit was designated as an outlier. The outliers usually had abnormal spectral absorbance patterns and values between wavelengths of 400 nm and 600 nm, although no obvious abnormalities were detected for the rest of the spectral wavelengths. More objective methods to select outliers include plots of Cook's distance that estimate a data point's influence on the regression (Nicolaï et al., 2007a), or plots showing the squared residual distances (Q^2) versus Hotelling T^2 values (Travers et al., 2014). In such methods, significance limits can be used to classify an observation as an outlier. These methods were not used because of operator unfamiliarity with their software operations. It is not known whether the resulting models would be different, but these methods could be tested in future research.

The spectral data from this experiment, the 2016 multi-varietal trial, included scans of two locations on the fruit (sun and shade sides). Models of individual sides and averages were created and compared.

Centering and Scaling

When spectrometer and fruit-surface temperature variables were included as x-variables, two preprocessing methods mean centering and scaling, were considered in order to ascertain the influence of having temperature naturally being on a different scale compared to the spectral values. Peirs et al. (2003a) suggested scaling all variables when including temperature variables. Scaling was done by subtracting the standard deviation of each variable from each observation. Mean centering was done by subtracting the mean of each variable from each observation using the default setting in the regression method used for model creation as discussed below in section 4.1.4.3.

Broad Wavelength Range Selection

The appropriate wavelength range to include for modeling spectra from the F-750 was initially chosen by scanning six store-bought fruit repeatedly in the same location 50 times, and calculating the standard deviation of each wavelength. Wavelengths below 500 nm and above 1000 nm were excluded from model building due to large random variation in this test (Figure 4-1).

Table 4-5. Spectral outliers counts in the 2016 experiment

Cultivar	Outdoor	Indoor	Postharvest
Chestnut	1	7	3
Cortland	2	-	1
Fireside	3	-	-
Frostbite	2	-	2
Haralson	(50*)	-	-
Honeycrisp	5	2	1
Honeygold	8	2	2
Keepsake	1	-	1
McIntosh	2	-	-
Minneiska	2	8	1
MN1942	3	2	-
MN55	1(35‡)	4	1
Wildung	1	-	1
Sweet Sixteen	-	11	-
Minnewashta	3	4	1
Total	34	40	14

Clerical errors:
*Haralson fruit were only removed when fruit temperature was used as an x-variable.
‡MN55 uncertainty in spectral filenames

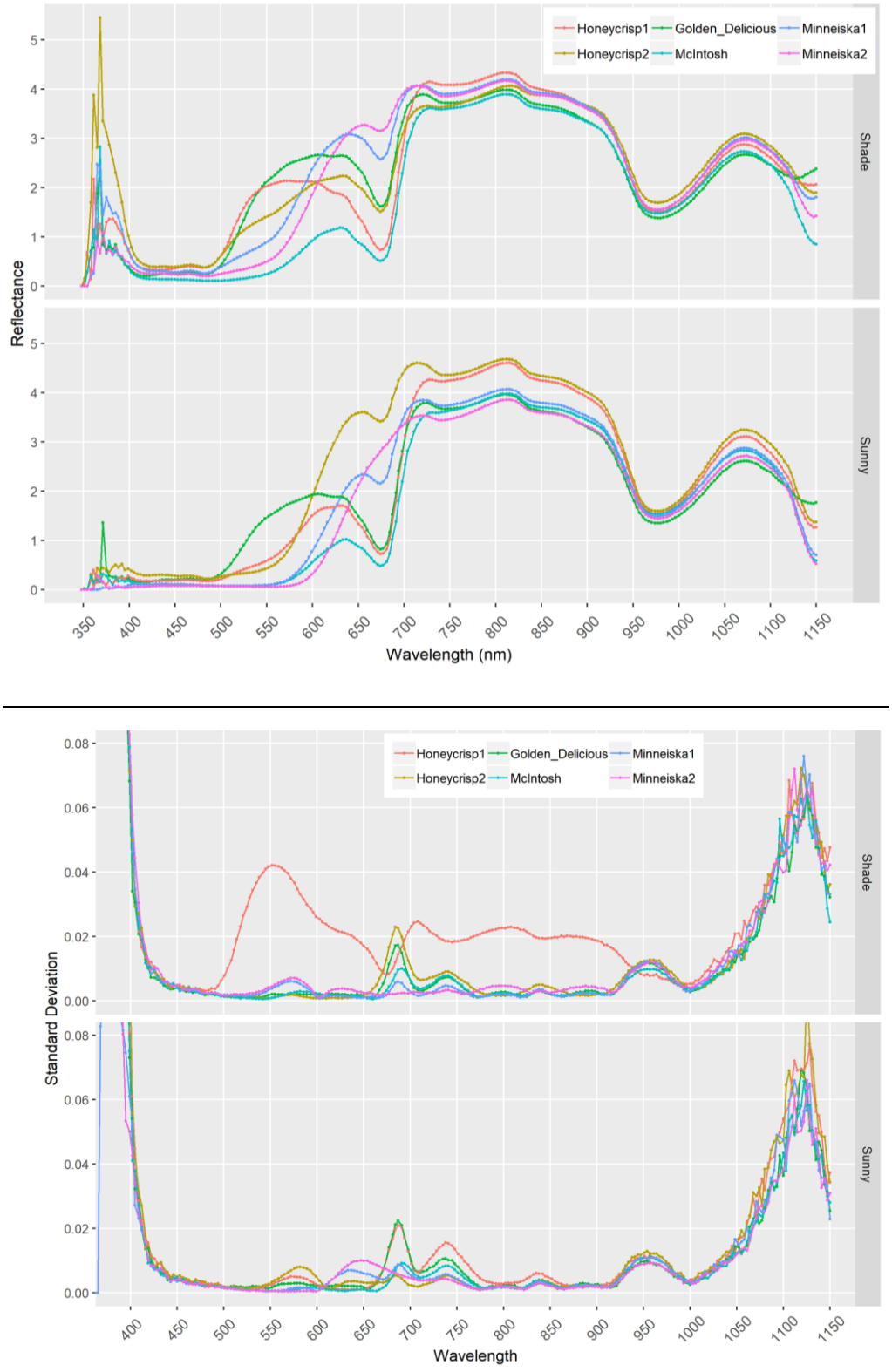


Figure 4-1. Mean reflectance (arbitrary units) and standard deviation of F-750 sensor at each wavelength from ~50 repeated scans on selected fruit.

Wavelength Selection Genetic Algorithm

Partial least squares regression (PLSR) is widely considered to be a suitable modeling procedure for the 'small n, large p problem', where often there are many more predictive variables than sample observations (Gottardo et al., 2016; Mehmood et al., 2012). Nevertheless, statisticians suggest that there are limits to the amount of noise that can be included in the form of extraneous predictor-variables for calibrating PLSR models. When the number of p (predictive variables) is much larger than the number of dependent variable observations (n), procedures such as PLSR and principal components regression (PCR) have difficulty ignoring the subspace attributable to noisy variables and correctly assigning the subspace to informative variables (Faber et al., 1995; Gottardo et al., 2016; Helland, 2001; Mehmood et al., 2012). A large number of potential methods are available to reduce the number of variables used as inputs through variable selection (Gottardo et al., 2016; Mehmood et al., 2012; Pasquini, 2018).

Most empirical methods of selecting variables for PLSR focus on improving the models through statistical comparisons such as reducing the RMSEPs. The chosen variables improve the model but may not have exact chemical relationships to the items being predicted (Pasquini, 2018).

Nevertheless, when dealing with complex information from complex objects (e.g., heterogeneous fruit), empirical selection of variables may be the only option (Pasquini, 2018). Therefore, use of a variable selection algorithm was investigated by integrating it into the analytical pipeline. Mehmood et al. (2012) reviewed several methods of variable selection, and some of the authors gathered several methods together into the R software package 'plsVarSel'. The genetic algorithm was the main selection algorithm used for the fruit trait data collected in this study, since the wavelengths selected by this algorithm were in the range of 800-1000 nm, where SSC has historically been well predicted (Nicolaï et al., 2007a; Wang et al., 2015) (see Figure 4-2).

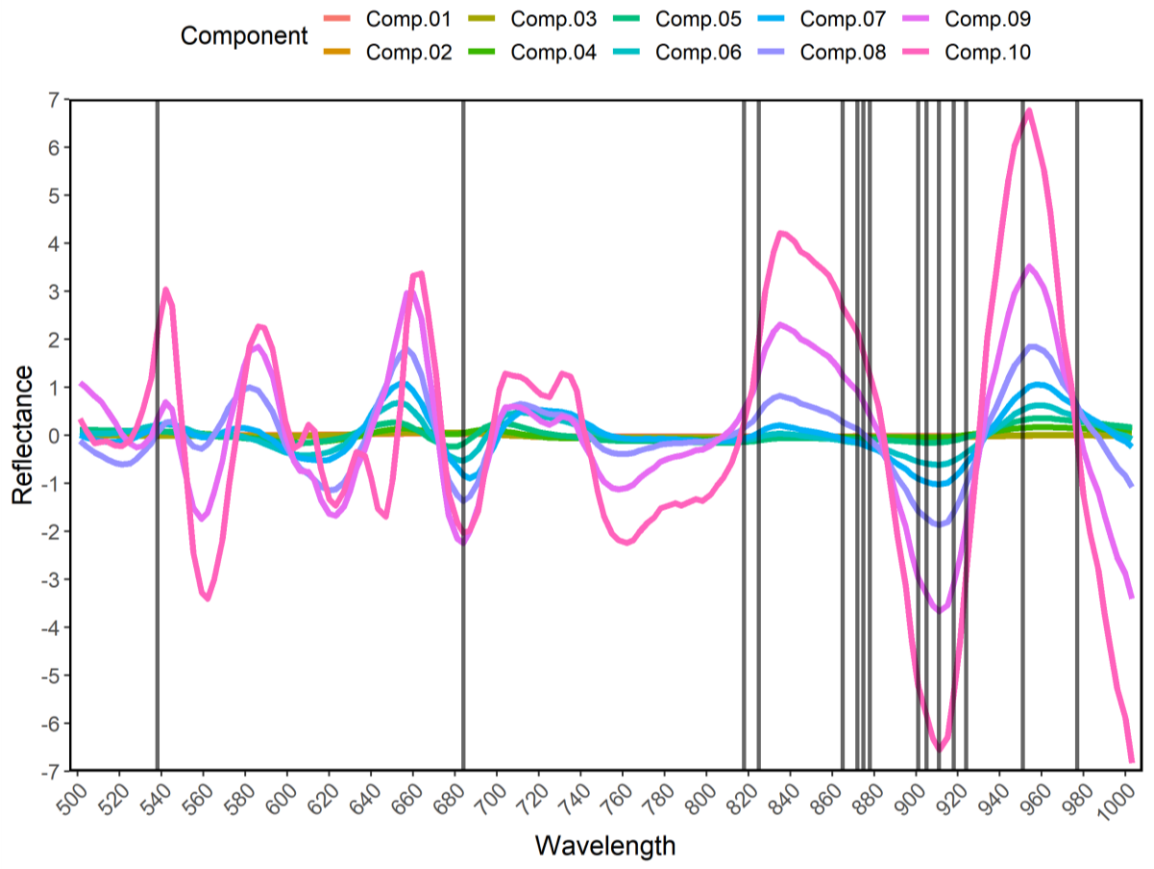


Figure 4-2. Example of specific wavelengths selected using the genetic algorithm of plsVarSel to produce a PLSR model for SSC.
Colored lines are regression coefficient values corresponding to components (i.e., latent variables) 1-10. Dark grey lines are the selected wavelengths.

A genetic selection algorithm works by assigning a portion of x variables to the initial population, then a PLSR model is fitted followed by variable evaluation on model performance. Variables have a greater chance of being kept in future models if they have a higher influence on the model performance. The next step is called ‘crossover and mutation’ in which the next population of x variables are selected based on crossover (performance of previous model fits) and random mutation, switching some variables for different ones at random. Variables surviving crossover and mutation advance to a new round starting with a new fitting of a PLSR model, with the cycle repeated for a set number of iterations (Mehmood et al., 2012). The ‘plsVarSel’ settings used for genetic

algorithm selection function 'ga_pls' were five for a threshold of the ratio of non-selected variables for each chromosome, 100 iterations, and a population size set to 150.

Calibration/Validation Data Divisions

Models were built using R software, version 3.4.4 with the RStudio GUI (R Core Team, 2018; RStudio Team, 2016). In preparation for model building for individual traits, the dataset was divided into two portions - 75% for calibration and 25% for internal validation. The data subsets were optimized to have a low variance between subsets but high variance within subsets using the 'smooth fractionator' approach following Martínez Vega et al. (2013). This approach avoids creating ill-conditioned models due to chance unequal population coverage/representation from random sampling. The 'smooth fractionator' method sorts data according to (trait) values and divides the data systematically, so that the subsets have variances representative to the overall observed sample (Gundersen, 2002; Martínez Vega et al., 2013; Wulfsohn et al., 2004). The systematic method creates four equally sized groups. For each trait, four models were run, using each of the four groups as the validation subset, one at a time.

Although this method of data division does not create independent calibration and validation datasets (Walsh, 2016b), it should be useful for testing the feasibility of the use of spectra from the handheld spectrometer.

The mean statistic values of the four different internal validations were used to represent the overall model performance for a particular calibration method per trait. The mean descriptive statistics of traits measured at-harvest and postharvest are provided in Appendix A. As described below, the statistics from the four models were used for comparing different calibration methods.

4.1.4.3 Multivariate Statistical Analyses

Classification methods (qualitative variables)

After dividing the data into calibration and validation subsets as described above, the categorical variables, including the presence/absence of BRusset, presence of LRusset, russet incidence (collected on a 0-10 scale, analyzed as low = 0-3 and high = 4-10), fruit

damage, Gcolor, Ocolor, Ocolor type, %Ocolor, and SPI, were modeled using linear discriminate analysis (LDA) from the 'MASS' package in R (Venables and Ripley, 2002). Default settings for function 'lda' were used (prior probabilities were estimated from the data, with no cross-validation). Each variable category was coded using a four-lettered assignation, and attached to the beginning of the spectral data. The resulting statistics for these models included the percentage of correct classifications per category and overall percentage of correct classifications for the variable. The entered categories involved only rational or pre-existing groups, i.e., no grouping procedure was done to search for the best model statistics.

LDA is similar to principle component analysis (PCA) except that, in the creation of latent variables (variables that are created from modeling as compared to directly observed variables such as fruit SSC), LDA attempts to explain the greatest variation between categories, instead of the greatest variation among the data itself, as with PCA. LDA is a parametric machine-learning method used to create linear discriminants (similar to components and latent variables) from the data, in order to reduce dimensionality (number of attributes of a dataset used in a model) and for later classification of the data (Ballabio and Todeschini, 2009; Wehrens, 2011). LDA assumes that data classes have normal distributions, and that the variance is the same for each class. It still performs well when these assumptions are only approximately met (Ballabio and Todeschini, 2009), and has a good track record (Hastie et al., 2009). The following publications cite successful use of LDA when classifying apple fruit traits with high dimensional data (Baranowski et al., 2012; Giovanelli et al., 2014; Khatiwada et al., 2015; Mendoza et al., 2014).

Model Creation and Component Selection (Quantitative Variables)

PLSR was implemented after data division using a kernel algorithm comparable to Nonlinear Iterative Partial Least Squares (NIPALS) from the 'pls' package in R (Mevik et al., 2018). A default setting of the pls function is to mean center both the y and x variables, in order to understand the results in regards to variation around the mean (Nicolaï et al., 2007a).

The optimal number of components (i.e. latent variables) for the PLSR models was determined by calculating each component's root mean square error (RMSE) with a LOO method, where each single observation was left out one at a time. This was done by component up to k components, with k = the total number of wavelengths used as x variables (see Figure 4-3). To avoid overfitting the model, the number of components was selected by identifying the model with the lowest root mean square error of calibration (RMSEC). Then a permutational comparison was done to find the simplest model with the fewest components not significantly (α=0.01) different from the global minimum RMSEC. That model, with the number of components found through permutation, was then validated internally using the test data subset, containing 25% of the observations that were not used in calibrating the model, to calculate several statistics, including the RMSEP, model bias, and residual predictive deviation (RPD) recorded for every model.

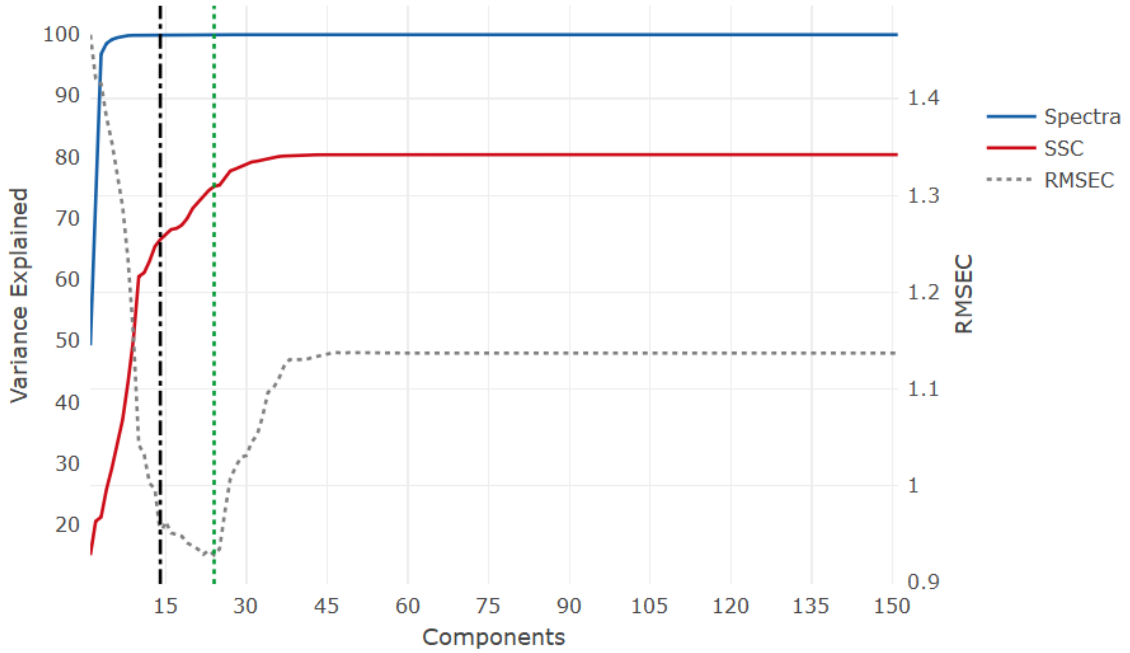


Figure 4-3. An example of the component selection method using a SSC calibration.
The selected number of components was determined by finding the model with the lowest RMSEC, and then selecting the model with fewest components not significantly different from the lowest RMSEC.
Black vertical line is the selected number of components, green vertical line is the component with the lowest RMSEC

Comparing partial least squares regression model calibrations

To understand how different settings and parameters affected the predictive abilities of the models, multiple different PLSR calibrations were analyzed. In the comparisons between any two selected calibrations, the RMSEPs from each model were compared using a standard two way ANOVA with fixed effects without interactions (Naes et al., 2002). In each calibration method there were four models created per trait, with each of the four versions corresponding to the four different data divisions as described above. In comparing a calibration method to another, the four RMSEP values (per trait) from each calibration method were used in an ANOVA using the calibration method as a fixed effect. Tukey's Honest Significant Difference test was used to create a set of confidence intervals based on the differences between the mean RMSEPs for each trait with $\alpha = 0.05$. The model calibrations and their purposes are described in Table 4-6.

Table 4-6. Calibration methods and purposes

Model ID	Description	Purpose
A	Spectra taken indoors at the time of harvest averaging both sunny and shade sides to predict at-harvest trait values	The basic model used as a benchmark
B	At-harvest indoor spectra using only the shaded side of the fruit to predict at-harvest traits	Test the use of spectra from a shaded side of the fruit
C	At-harvest indoor spectra using only the sunny side of the fruit to predict at-harvest traits	Test the use of spectra from the sunny side of the fruit
D	At-harvest indoor spectra from individual sides (sunny and shaded) treated as individual fruit	Test the use of spectra from the shaded side of the fruit
E	Spectra taken outdoors on the same fruit as in model A predicting at-harvest trait values	To test the spectrometer's outdoor use
F	Spectra, sides averaged, taken both indoors and outdoors on the same fruit but treated as different fruit	To test the potential benefit more than two scanning locations on the same fruit
G	Spectra taken in both indoors and outdoors locations on the shaded and sunny positions of the same fruit but position and location treated as different fruit	To test the potential benefit of including un-averaged spectra from both positions and both locations on the same fruit
H	Spectra averaged (sunny and shaded) taken indoors but with a fraction of the wavelengths selected to predict at-harvest traits	Test the potential benefit of using fewer wavelengths for predictions
I	Postharvest trait values were predicted using spectra taken poststorage indoors with positions averaged. If fruit were scanned on two days (day 1 removal from	Test the postharvest use of the spectrometer on predicting postharvest traits

Table 4-6. Calibration methods and purposes

Model ID	Description	Purpose
	storage and day 7), those scans were also used in producing the average spectra per fruit.	
J	Indoor at-harvest time spectra were used to predict poststorage trait values. Spectra were averaged by position.	Test if at-harvest indoor spectra could predict poststorage trait values
K	Outdoor at-harvest time spectra were used to predict poststorage trait values. Spectra were averaged by position.	Test if outdoor at-harvest spectra could predict poststorage trait values
L	At-harvest indoor and outdoor spectra were both used to predict poststorage trait values. Spectra were averaged by position.	Test if combining indoor and outdoor spectra could have higher predictive ability than when used separately for predicting poststorage trait values
M	Average indoor and average outdoor spectra used with temperature variables for fruit and spectrometer	Test if improvements could be made upon model F
N	Average indoor and average outdoor spectra used with only the fruit surface temperature variable	Test if there was a decline in predictive ability upon model F
O	Average indoor and average outdoor spectra used with the spectrometer temperature variable	Test if there was a decline in predictive ability upon model F

5 CHAPTER 1. RESULTS AND DISCUSSION

5.1 Comparison of Calibration Methods

5.1.1 PLSR Calibration Results

Presented below are concise descriptions and comparisons of the predictive abilities of different model calibrations. The dataset collected for this experiment allows the comparison of models created with spectral data collected indoors vs outdoors, at-harvest vs. postharvest, using sun-exposed vs. shaded-side measurements, with or without temperature variables, as well as other combinations of the aforementioned conditions. Predictive abilities of the models for individual traits are covered in the Discussion, section 0. In addition, detailed tables listing all traits by different calibrations methods are in Appendix section A.3 (Table A 3 to Table A 17).

In Table 5-1, the parameters and settings of the different calibration methods are presented, along with average statistics per calibration. The averaged values are calculated by considering all traits (e.g., a calibration's mean RPD is the average RPD for

all 23 traits models). The order in which the calibration methods are described in this Results section follows the order found in Table 5-1. For this Results section, I limit the description of the results to the range of RPD values (of all traits) found in a particular calibration method, identifying specific traits with RPD values above 1.5, and brief comparisons to other model calibrations. The purpose for each model calibration is covered in section 4.1.4.3, and is further considered in the Discussion, section 5.1.1.

Table 5-1. Partial least squares regression calibration parameters and average statistical results

Calibration ID	Acquisition Location	Fruit Sides	Temperature Vars	Spectra	Trait Time	Selection Algorithm	Training (N)	Test (N)	Wavelengths (N)	Mean Adj. R^2	Mean Abs. Bias	Mean Adj. RMSEP	Mean RPD	Calibration Comparison: Traits with p<0.05
A	Indoor	Avg	-	H	H	-	404	135	151	0.39	0.62	24.61%	1.33	
B	Indoor	Shade	-	H	H	-	404	135	151	0.29	0.55	26.96%	1.22	**A:** OMH, %Ocolor, SSC, SPI
C	Indoor	Sunny	-	H	H	-	404	135	151	0.31	0.60	26.16%	1.23	**A:** %Ocolor, SSC, Diam; **B:** nd
D	Indoor	Both	-	H	H	-	788	263	151	0.35	0.39	25.68%	1.27	**A:** %Ocolor, SSC
E	Outdoor	Avg	-	H	H	-	409	136	151	0.38	0.67	25.26%	1.31	**A:** A1, M2
F	Both	Avg	-	H	H	-	813	271	151	0.44	0.38	23.73%	1.39	**A:** DM; **E:** A1, M1, M2, DM
G	Both	Both	-	H	H	-	1584	528	151	0.39	0.24	24.84%	1.30	**A:** %Ocolor, DM, SSC; **F:** %Ocolor, SSC, Diam
H	Indoor	Avg	-	H	H	GA	404	135	21	0.32	0.60	26.16%	1.26	**A:** OMH, SPI, Diam
I	Indoor	Avg	-	PH	PH	-	412	137	151	0.48	0.53	14.50%	1.45	**A:** C0, QF, %Russet, DM, TA, Diam, Flav, Weight
J	Indoor	Avg	-	H	PH	-	404	135	151	0.41	0.76	15.19%	1.36	**I:** A1, A2, M2, OAH, QF, pH
K	Outdoor	Avg	-	H	PH	-	401	134	151	0.37	0.80	15.35%	1.31	**I:** A1, A2, C0, E2, M1, M2, OAH, OMH, QF, SSC, PHweight; **J:** M1
L	Both	Avg	-	H	PH	-	805	268	151	0.46	0.44	14.64%	1.41	**A:** A1, C0, Cn, %Russet, DM, TA, **I:** QF, Flav
M	Both	Avg	B	H	H	-	798	266	151	0.45	0.43	23.61%	1.39	**A:** DM, TA, Diam; **F:** nd
N	Both	Avg	F	H	H	-	798	266	151	0.44	0.47	23.70%	1.39	**A:** DM, TA, Diam; **F:** nd; **M:** nd
O	Both	Avg	S	H	H	-	813	271	151	0.44	0.34	23.77%	1.38	**A:** DM, Diam; **F:**nd; **M:** nd; **N:** nd

Abreviations: Avg=average of shaded and sunny sides, B= both fruit and spectrometer temperatures included, F=fruit temperature only, S=spectrometer temperature only, H= analysis used data collected at-harvest, PH= analysis used data collected 10 weeks postharvest, GA=genetic algorithm, nd=no difference p<0.05

Highlighted statistics come from the mean value for 23 fruit trait models for the given calibration method. The only variation in traits included is with the PH models that do not contain SPI, but do contain PHweight.

Traits in red font have a higher RMSE-CV compared to the listed model at p<0.05. The shading of the PH Mean Adj. RMSEP is independent of the rest of the column.

Bias values are the mean of the absolute values.

Phen and Flav sample size values were omitted in the calculation of Training and Test (N) due to a smaller set of available data ($N_{Training}$=291, N_{Test}=97)

Indoor Measurements (Calibration A (cal-A), Table 5-1**)**

The RPDs for individual traits predicted using spectra taken indoors ranged from 1.07 to 2.12 (Table A 3). Traits with RPD values above 1.5, the threshold that Nicolaï et al. (2007) suggested would be useful for discriminating groups of observations into low and high categories, included %Ocolor, SPI, SSC, and diameter. For other traits, the models did not have high predictive ability (all model RPDs were <1.5); however, to determine the best calibration method, this first calibration was used as the benchmark for comparing other methods.

Indoor Single-Sided Spectral Measurements (Calibration B and C), **vs Averages of Sun- and Shade-side** (cal-A) **Measurements**

Models calibrated with spectra exclusively from the shaded side (cal-B) of the fruit had RPDs ranging from 1.05 to 1.79 (Table A 4). %Ocolor and diameter were the only traits with RPD above 1.5. The average RPD value (all traits) using the shaded spectra for calibration was much lower compared to the calibration using averaged spectra from both sides of the fruit (cal-A).

Models calibrated with spectra of the sun-exposed side of the fruit (cal-C) had slightly higher average RPDs than calibrations using scans of the shaded side of the fruit (cal-B). The range of RPD values for cal-C was 1.04 to 1.83 (Table A 5). %Ocolor was the only trait with an RPD above 1.5. None of the trait RMSEP differed between the separate fruit sides ($p > 0.07$).

The models for %Ocolor, SSC, and diameter from calibrations with the sunny spectra (cal-C) had higher (i.e., worse) RMSEPs than models calibrated with averaged spectra (cal-A; all p-values=0.01). The traits models calibrated with the shade-side spectra (cal-B) with higher RMSEPs than cal-A were OMH, %Ocolor, SSC, and SPI ($p < 0.04$)

Overall, the calibration using average spectra from both sides (cal-A) had a much higher RPD than calibrations using spectra from only a single side (cal-B or C). Calibrations from each separate side were not statistically different from one another.

Indoor Scans with Both Sides Considered as Separate Fruit (cal-D)

Cal-D used indoor spectra from the shaded and sunny sides of the fruit with each side considered as if from a separate fruit. The trait data were copied and each copy was assigned to a separate side (i.e., position) of the fruit. Essentially this doubled the numbers of observations and each fruit was included twice, once with spectra from the shaded side of the fruit and again with sunny side spectra.

The traits %Ocolor and SSC had higher RMSEP values than the models calibrated from averaged indoor spectra (cal-A, $p<0.0039$). The traits Diameter, %Ocolor, and SPI had RPD values above 1.5. This calibration method, treating sides as separate fruit (cal-D), produced an average RPD that was higher than calibrations using data only from single sides (shaded or sunny sides exclusive) of the fruit, but still lower than the calibration with averaged spectra. This calibration is method is therefore preferred to that using only a single scan alone, but inferior to using two averaged scans.

Indoor (cal-A) **vs. Outdoor Measurements** (cal-E)

The RPDs for individual traits using models calibrated with outdoor spectra (cal-E) ranged from 1.06 to 1.99 (Table A 7). The traits diameter, %Ocolor, SSC, and SPI had RPDs above 1.5. With cal-E, models for two traits, A1 and M2, had higher RMSEP values compared to their respective indoor calibrated models ($p < 0.03$). Generally, calibrating prediction models with outdoor spectra resulted in lower average RPDs than the calibration method using indoor spectra (cal-A).

Both Average Indoor and Outdoor Measurements Inclusive (cal-F)

In this model calibration method, spectra from both outdoor and indoor scans were used to create a single model (cal-F). The same fruit were considered different fruit per location, (indoors or outdoors). Shaded and sunny sides were averaged in each location. The trait data was copied and each copy was assigned to separate scanning locations. Essentially the numbers of observations were doubled and each fruit was included twice, once with outdoor scans and again with indoor scans.

Cal-F improved the RMSEP for one trait, DM ($p = 4.19*10^{-03}$), compared to the calibration using only indoor spectra (averaged sides; cal-A). When compared to the calibration using only outdoor spectra (averaged sides; cal-E), the cal-F improved the

RMSEPs for four traits, A1, M1, M2, and DM ($p < 0.04$). Models from cal-F had RPDs ranging between 1.13 and 2.09 (Table A 8), with RPDs for diameter, %Ocolor, SSC, SPI, and weight being above 1.5. Considering all calibrations shown in Table 5-1, cal-F had one of the highest average RPD. However, no test was performed to determine the statistical significance of differences in average RPD values between calibrations.

Indoor, Outdoor, Shaded, Sunny Together Considered as Mutually Exclusive (cal-G)

This calibration method (cal-G) combines concepts from two of the previous methods (cal-F, and cal-D). Each side of the fruit was considered a different observation in both indoor and indoor locations. Essentially the number of observations was increased fourfold.

The models for %Ocolor and SSC from cal-G had higher RMSEP values when compared to both cal-A (with averaged indoor spectra; $p < 4.11*10^{-3}$) and cal- F (with averaged indoor and outdoor spectra; $p < 2.48*10^{-4}$). DM was the only trait to have a lower RMSEP, but only compared to DM's RSMEP in cal-A ($p=0.0084$). The traits modeled with RPD values above 1.5 were diameter, %Ocolor, and SPI (Table A 9). Cal-G had the lowest absolute mean bias of all the calibrations, but cal-G generally did not improve the predictive abilities of the trait models.

Indoor calibration (cal-A) **vs. Selection algorithm** (cal-H)

One calibration method used a genetic algorithm (cal-H) to select specific wavelengths from indoor spectra to create trait models. The traits modeled in cal-H had RPD values ranging between 1.07 and 2.11, with only %Ocolor and SSC having an RPD value above 1.5 (Table A 10). The traits OMH, SPI, and diameter had higher RMSEP values than the indoor model (cal-A, $p< 0.05$) employing all spectral wavelengths. The genetic algorithm did not result in simpler models (fewer wavelengths) that had predictive abilities equivalent to cal-A.

Postharvest (cal-I)

Stored fruit were scanned on either one or two occasions after storage. All stored fruit were scanned once before a subset of 10 fruit were destructively processed on the day after removal from storage, and the remaining 10 fruit were scanned again 7 days later and processed. To have sufficient sample sizes (approximately 500) to split for calibration and internal validation destructive-trait models, data from the 1- and 7-day measurements were pooled and fruit that had spectra from both dates were averaged.

Fruit can change significantly between days 1 and 7 after removal from storage so an additional calibration used only the spectral data from day 7 for the fruit processed on day 7 (i.e. the day 1 spectra for these fruit was not included, results not shown). With this additional calibration the average RPD was lower, the mean adjusted RMSEP was higher, and there was one statistically higher RMSEP (A2, $p=0.04$) compared to the first method discussed (cal-I); all other trait models showed no differences ($p > 0.22$). Therefore, the method of averaging additional spectra taken on day 1 with day 7 poststorage fruit was utilized (cal-I).

The RPDs for models calibrated with poststorage spectra (cal-I) ranged from 1.10-2.19, (Table A 11). The traits SSC, C0, Wt/Diam, Diameter, Weight, %Ocolor, and PHweight had RPD values above 1.5. SPI was not measured poststorage. The effect of storage (cal-I) on the predictive abilities of some of the models was two-way, with the RMSEP for C0, QF, and Flav higher than the indoor at-harvest calibration (cal-A, $p < 0.03$), and the RMSEP for %Russet, DM, TA, diameter, and weight lower ($p < 0.03$) than for cal-A. Cal-I had a higher mean RPD and lower mean RMSEP than the at-harvest calibrations (cal-A).

Postharvest Traits Predicted with At-Harvest Spectra (cal- J, K, and L)

The feasibility of using harvest-time spectra to predict poststorage trait values (Kumar et al., 2015; McGlone et al., 2003, 2002a) was tested by training trait model calibrations with at-harvest spectra. Indoor and outdoor spectra were used separately and once in a combined calibration.

Indoor

Using at-harvest indoor spectra for the poststorage trait calibrations (cal-J) produced a higher average RPD (all traits combined) relative to some of the other at-harvest calibration methods. Nevertheless, the traits A1, A2, M2, OAH, QF, and pH all had higher RMSEP with cal-J, Table A 12, compared to the calibration with poststorage spectra (cal-I, $p < 0.0467$). The RPD range went from 1.10 to 2.10, and SSC, diameter, at-harvest weight, %Ocolor, and PHweight had RPD values above 1.5.

Outdoor

When at-harvest outdoor spectra was used for the poststorage trait calibrations, (cal-K) the trait models' RPD values ranged from 1.12 to 2.10 (Table A 13). The traits Wt/Diam, weight, %Ocolor, and PHweight had RPD values above 1.5. The traits Wt/Diam, weight, %Ocolor, and PHweight had RPD values above 1.5. The RMSEPs were greater with cal-K for the traits A1, A2, C0, E2, M1, M2, OAH, OMH, QF, SSC, and PHweight compared to the calibrations with poststorage spectra (cal-I, $p< 0.0345$). In a comparison of cal-K (outdoor-at-harvest spectra) with the poststorage trait predictions of cal-J (indoor-at-harvest spectra), only the model for M1 had a statistically different RMSEP, which was higher with cal-K than with cal-J ($p=0.0370$).

Both indoor-outdoor

Based on the good predictive ability of cal-F (both indoor and outdoor average spectra) for at-harvest trait predictions, the same method was also used to predict poststorage fruit traits (cal-L). This at-harvest indoor-outdoor calibration predicting postharvest traits (cal-L) provided the second highest RPD and second lowest RMSEP relative to all other calibrations tested in this thesis. Only two traits were statistically different from cal-I (postharvest spectra-postharvest traits) - QF that had a higher RMSEP ($p=5.46*10^{-3}$) and Flav that had a lower RMSEP ($p=0.02$). The range of RPD values for trait models in cal-L was 1.14 to 2.18.

The average RPD value for cal-J, at-harvest indoor spectra, was expectedly higher than for cal-K, which used outdoor spectra for predicting postharvest traits. Cal-L (i.e. cal-F but with postharvest traits) had higher predictive abilities than both cal-J and cal-K.

However, the calibration with the best predictive ability, greatest mean RPD and lowest RMSEP, for postharvest traits used postharvest spectra (cal-I, mean RPD=1.45, mean adjusted RMSEP=14.50%), which was also the best calibration overall.

Fruit and Spectrometer Temperature (cal-M, N, O)

The influence of temperature was examined by including fruit surface temperature and spectrometer internal temperature as x-variables into the calibration with the highest RPD for at-harvest traits, cal-F (both average-outdoor and average-indoor spectra). The temperature variables were included together (cal-M) and separately (cal-N and O). There were no statistical differences in trait model RMSEP values compared to the calibration without temperature variables (cal-F, $p>0.1632$) for all three calibrations with temperature variables (M, N, O).

The calibrations with temperature variables (cal-M, N, O) were also compared to the indoor calibration (cal-A) without outdoor spectra. The models for DM and diameter models had lower RMSEPs when either temperature variable was included ($p< 0.0094$), while the model for TA had lower RMSEP only when the fruit temperature variable was included ($p< 0.0450$). The temperature calibrations, cal-M, N, and O had mean RPD values higher than cal-A with only indoor spectra, but when compared to cal-F with both indoor and outdoor spectra, the RPD values were similar (Table A 15, Table A 16, Table A 17).

5.1.2 Linear Discriminate Analysis Results

Ordinal and Categorical Traits

Linear discriminate analysis (LDA) was used to classify several ordinal or categorical traits that were initially recorded as potential covariates to the quantitative traits. The traits with relatively balanced numbers of observations in their corresponding categories were Gcolor, Ocolor, Ocolor type, and SPI, which had classification accuracies of 46%, 67%, 52%, and 61 %, respectively.

However, the experiment was not designed to sufficiently test NIRS on the categorical and ordinal traits and the predictive classification accuracy was very low. Table 5-3

provides an overview of the classification accuracies and the percent of the observations within each category. Cross-tabulated confusion matrices with predicted category classification for each trait are shown in Appendix B. In addition to the Discussion in this section, an additional brief synopsis of the categorical and ordinal trait results is provided in Appendix B.

Classification of Cultivars

LDA was also used to attempt to classify fruit by cultivar (Table 5-2). In this analysis, the predictions had an average error of 0.94%. A few fruit were inaccurately classified in only one of the four different cross validations from the 'fractionator' data divisions. Cultivars clearly had distinct spectral reflectance values.

Table 5-2. Confusion matrix for the classification of cultivars by linear discriminate analysis (LDA).

Cultivar	**CH**	**CT**	**FI**	**FR**	**HA**	**HC**	**HG**	**KP**	**MC**	**MI**	**42**	**55**	**WD**	**SS**	**MW**
Chestnut (CH)	24.75														
Cortland (CT)		25.25							0.25					0.25	
Fireside (FI)		0.25	12.25												
Frostbite (FR)				25							0.25				
Haralson (HA)					24.5			0.25			0.25	0.25			
Honeycrisp (HC)					0.25	24.5				0.25					
Honeygold (HG)							24.5								
Keepsake (KP)								24.75							
McIntosh (MC)		0.5							24.5						
Minneiska (MI)										22.75					
MN1942 (42)											23.75				
MN55 (55)												15.75			
Wildung (WD)													25		
Sweet Sixteen (SS)														21.75	
Minnewashta (MW)					0.25						0.25				24

Table values are the average count of fruit classified from 4 validations.

Counts following the left column are the observed cultivar counts. Values following the top row are the counts of classified fruit by the LDA model.

Overall classification accuracy was 99.06 % (0.55% SD).

Table 5-3. Linear discriminate analysis classification accuracy of apple fruit traits using spectra taken with the F-750 indoors on the day of harvest, 2016*

Variable	Category						Overall Accuracy (%)
	Prior Percentage						
	Accuracy						
	Average for 15 Cultivars						
Cultivar	0.07						0.99 (±0.01)
	0.99						
	abio	**biot**	**both**	**none**			
Damage Type	0.08	0.02	0.01	0.90			0.85 (±0.02)
	0.23	**0.00**	**0.00**	**0.92**			
	gree	**plgr**	**plyl**	**yell**	**ylgr**		
Ground color	0.09	0.1	0.17	0.36	0.29		0.46 (±0.03)
	0.56	**0.30**	**0.36**	**0.56**	**0.45**		
	dkrd	**none**	**oran**	**ored**	**pink**	**red**	
Over-color Color	0.3	0.04	0.02	0.02	0.08	0.55	0.67 (±0.01)
	0.68	**0.68**	**0.78**	**0.00**	**0.34**	**0.73**	
	blst	**blus**	**none**	**stbl**	**stri**		
Over-color Type	0.48	0.36	0.01	0.15	0.00		0.52 (±0.04)
	0.59	**0.55**	**0.42**	**0.21**	**0.00**		
	bico	**red**	**yell**				
Over-color Percent	0.19	0.73	0.08				0.85 (±0.05)
	0.62	**0.91**	**0.97**				
	none	**pres**					
Body Russet	0.71	0.29					0.73 (±0.02)
	0.82	**0.54**					
	none	**pres**					
Lenticel Russet	0.34	0.66					0.71 (±0.03)
	0.58	**0.78**					
	high	**low**					
Russet Percent	0.04	0.97					0.94 (±0.01)
	0.30	**0.98**					
	fres	**inmt**	**over**	**stor**			
Starch Pattern Index	0.51	0.07	0.21	0.21			0.61 (±0.04)
	0.69	**0.46**	**0.54**	**0.53**			

The rounded N_{cal} =413, and N_{val} =138, 15 cultivars and 151 wavelengths were used. Accuracy values are calculated from the 4 data splits, the bracketed values are the standard deviation or standard error. Abbreviations (definitions): abio=abiotic (mechanical injury), bio=biotic (insect or pathogen), gree=green, plgr=pale-green, plyl=pale-yellow, yell=yellow, ylgr=yellow-green, dkrd=dark-red, oran=orange, ored=orange-red, blst=blush-stripe, blus=blush, stbl=stripe-blush, stri=stripe, bico=bicolor (65-35% overcolor), pres=presence, fres=fresh (SPI 5-7) , inmat=immature (SPI 1-2), over=over-mature (SPI 8), stor=storage (SPI 3-4), low (russet percent 0-30%) and high (35-65%)

5.2 Discussion of PLSR Calibration Parameters and Comparison to Published Results

This section further explores the impact of different PLSR calibration parameters and compares them to the results of other published studies involving NIR spectroscopy with apple fruit. The primary statistics used to compare studies are the RPD and RMSEP (or RMSE-CV). If the RPD value was not immediately provided and if both the trait's standard deviation of the validation dataset and RMSE were available, then the RPD was calculated to facilitate comparisons. This Discussion first examines the performance of the separate calibration methods, and then delves into the overall ability for individual traits to be predicted by NIR spectra.

5.2.1 The Performance of Different Calibrations

The effects of different settings and parameters on the general performances of calibration methods are discussed below. The following points that influence trait predictability are discussed: the use of the spectrometer outdoors, using spectra from a single side of the fruit, treating repeated scans on the same fruit as individual observations, postharvest use of the spectrometer, and the influence of temperature on the spectral models.

5.2.1.1 Outdoor Measurements

Prediction of fruit traits using data collected outdoors would be useful for breeders, growers, and researchers. In my experiment, two thirds of the trait models had lower mean RPD values when predicted using outdoor spectra rather than indoor spectra. Only models of two traits (A1 and M2) of cal-E with outdoor scans showed statistical differences with higher RMSEPs compared to models with the indoor calibration (cal-A, $p < 0.03$). The outdoor environmental conditions of this experiment were semi-controlled, in that scanning was done the same day of harvest, at a location at least a 45 min drive away from the orchard, under the same crabapple tree. The fruit had at least 45 minutes to equilibrate in temperature and to dry surface moisture from the time point of harvest. However, the mix of time of day, sunlight conditions, and daily temperature varied on different harvest days. Since only two trait models had lower predictive ability, it appears

that the use of this particular spectrometer outdoors is feasible, but slightly lower performance could be expected. To the best of my knowledge, this experiment could be the first report of comparing the same fruit with different model calibrations created from spectra taken outdoors and indoors on the same day.

Using spectrometers in an orchard environment has received little attention, possibly due to significant challenges of outdoor NIRS use that involves variable temperature, light, and moisture conditions, all of which considerably influence the NIR spectrum (Nicolaï et al., 2007a). Few instruments used outdoors are designed specifically to control for sunlight. Herold et al. (2005), used a custom built instrument with a grommet to seal out ambient light. Walsh et al. (2000) noted that upon removing a spectrometer's shroud, performance decreased with their models. The spectrometer used in this report, the F-750, purportedly controls for ambient light effect (spectrometer readings due to light not from the reference bulb) and dark current (spectrometer readings due to random electron fluctuations within silicon structure of the spectrometer) after taking several different scans and internal calculations for each sample.

Studies that used NIRS outdoors with apple fruit include the reports of Herold et al. (2005), Zude et al. (2006), Beghi et al. (2013), Kumar et al. (2015), and Torres et al. (2016). Zude et al. (2006) used a modular spectrometer system in a 19" rack and a laptop computer to take measurements in an orchard to estimate SSC at-harvest and poststorage. The prediction value for SSC was low to moderate with R^2 of 0.05-0.41 and SECV $\left(SECV = \sqrt{\frac{\sum_{i=1}^{n_p}(\hat{y}-y_i-bias)^2}{n_p}} \right)$ values of 0.94-1.85. Beghi et al. (2013) collected spectra in the orchard using a 'portable' spectrometer in a backpack with a fiber optic cable and laptop computer in each hand. They sampled two cultivars over 6 and 8 weeks, and had moderate model performance with RPD values of 2.00-2.25, 1.49-1.81, and 1.54-1.43 for SSC, TA, and firmness, respectively. Their models were based on data from single cultivars. Torres et al. (2016) used a more 'portable' spectrometer than that of Beghi et al. (2013) to classify sun injury of 'Granny Smith' fruit with partial least squares discriminate analysis (PLS-DA). Their results showed errors of approximately 30% in the prediction of specific sunburn categories. The LDA classifications errors in my study

were much higher using indoor spectra, and it is assumed they would be higher using outdoor spectra. Among balanced categorical traits the overall errors ranged from 54-39 %, (Table 5-3).

Kumar et al. (2015) used a model calibrated from three years of spectra collected on fruit from over 200 genotypes scanned indoors to predict after-storage SSC of a fourth year's fruit scanned on the tree. The RPD value for SSC on a seedling level was 1.55, and on a fruit level between 0.90 to 1.35 for two different calibrations (other postharvest results from Kumar et al. (2015) are discussed below in section 0). In my study, the fruit level RPD for the SSC model calibrated using outdoor spectra (cal-E) was 1.57, similar to the findings of Kumar et al. (2015) on a seedling level, but higher compared to their results on a fruit level. This comparison, however, is not completely analogous, since my models calibrated from outdoor spectra predicted at-harvest SSC, while Kumar et al. (2015) only modeled after-storage SSC. This is important because respiration of apple fruit continues after harvest through storage until climacteric when nearly all starch has been converted to sugars (Duque and Arrabaça, 1999). Each fruit scanned at harvest could have a different state of conversion of starch to sugars, but after storage, fruit SSC are more stable. Therefore, it would be expected to have lower predictive ability for SSC model of cal-E than the models of Kumar et al. (2015) due to greater fruit-to-fruit variation of at-harvest SSC. The more appropriate comparison would be to validate my SSC model from cal-A (at-harvest indoor spectra), with new spectra acquired from fruit outdoors and SSC values measured poststorage.

Again, to the best of my knowledge, my experiment could be the first report of comparing the same fruit with different model calibrations from outdoor and indoor spectra. Since only two models showed a decrease ($p < 0.03$) in predictive ability with outdoor data, the spectrometer seems fully capable of being used outdoors, though the predictions may be expected to be slightly less accurate than when it is used in controlled conditions indoors.

5.2.1.2 Separate shade and sun exposed scans

When calibrating models, it is essential to capture the range of variation both in fruit quality (Zhang et al., 2019) and in environmental conditions (Peirs et al., 2003a). Bobelyn et al. (2010) reported on the significant influence of cultivar, origin, and season on the performance of spectral models. The influences of canopy environment (Peirs et al., 2003b; Sestras et al., 2009) and sun exposure (Kuckenberg et al., 2008) have also been reported. The more the variability is accounted for, the more robust the prediction models will be. For this reason, several calibration methods were designed to include spectral variations related to the same fruit with static quality metrics. The different calibration methods that were examined here included spectra from different sides of the fruit, outdoor and indoor spectra (as discussed above), and spectra at different temperatures. This section discusses the impact of different calibrations with spectra from different sides of the fruit.

The number of scans taken on different positions of a fruit plays an important role in capturing the variability found within a single fruit. Peiris et al. (1999) found that for SSC and DM, the least amount of variability is found around the fruit equatorial circumference. The average of two (Paz et al., 2009) or more scans (Pissard et al., 2018) is typically used in the literature. Thus, in apple spectroscopy, researchers commonly scan fruit equatorially on the shaded and sun-exposed sides (Bureau et al., 2009), as was done in this experiment.

In my data collection, the trait phenotypic values measured from the juice, such as SSC, TA, and pH, represented the whole fruit and not just of the scanned positions. Therefore, collecting spectral variation in the fruit from the proximal to distal (stem to calix) ends could be necessary to calibrate robust spectral models for accurate trait predictions as they were phenotyped (on the whole fruit). This experiment is not capable of addressing that question and it is suggested for future research. However, the difference between the sun and shade sides and their average can be compared.

The single-sided (sun-exposed or shaded) model calibration methods (cal-B and C) had fewer models with RPD values above 1.5. Two traits had higher RMSEP, %Ocolor and SSC, when using both single sides separately for model calibrations compared to

calibration with the averaged spectra (cal-A, $p < 0.01$). The traits OMH, SPI, and diameter had higher RMSEP values in only one of the single-sided calibrations compared to cal-A ($p < 0.02$, see Table 5-1). The chemical data collected for SSC, TA, and pH represented the whole fruit, and not just that of the scanned positions. The use of incomplete spectral information when calibrating models for traits that involve the whole fruit could affect the models' predictive abilities.

Although calibrations from shaded-side spectra (cal-B) produced models with slightly lower RPD and higher RMSEP values than those calibrated with the sunny-side spectra (cal-C) for many traits, they were not significantly different ($p > 0.07$) from the cal-C models. Since one side of the fruit appears not to be any more important, data collection from either side can be recommended. Spatial variability of traits in fruit could be cultivar-dependent. Spatial variability among the 15 cultivars in this study might mask the spatial differences found within fruit of individual cultivars. Kuckenberg et al. (2008) reported large SSC differences between sunny- and shaded-sides in 'Jonagold' but not in 'Golden Delicious' fruit.

My ANOVA tests only showed differences between calibrations using single sides (cal-B and C) and the calibration using averaged sides (cal-A). Costa et al. (2011) also found that extra scans on the same fruit improved model predictive ability to a great degree. Therefore, it is advisable to scan fruit from more than one side, and even more than twice would be advisable. However, if only one scan can be done (e.g. on the tree measurements) due to battery, time, or convenience limitations, more research is needed (perhaps by cultivar) to recommend the best position on the fruit for spectral acquisition.

5.2.1.3 Using collective scans from non-averaged sides as well as outdoor and indoor scans

The next question approached was, do both sides (shade and sunny) need to be averaged, or could the separate sides be used together to create trait models of comparable accuracy to cal-A? Cal-D used un-averaged indoor spectra from both shaded and sunny sides together (i.e., each fruit was included twice once per side). The predictive performances of all models using single-sided calibrations (cal-B, cal-C, and cal-D) were negatively

affected compared to models calibrated from averaged indoor spectra. On average, the RPD was slightly higher with cal-D than calibrations using only single sides (shaded or sunny sides), but still lower than the calibration with averaged spectra (cal-A). I theorize that the lower predictive ability occurred because the trait data was not collected from separate sides of the fruit corresponding to the sunny- or shaded-side of the fruit. An area for more research could be thoroughly testing the spatial (positional) sampling on the fruit for trait and spectral data.

The subsequent calibration method combined the previous questions, would sunny- and shaded-side spectra, if used separately in both locations, outdoor and indoor, improve the predictive abilities of the trait models? Trait models of cal-G were built using all spectra taken at harvest time, using the un-averaged spectra from both sides of the fruit in both indoor and outdoor locations, thus including more variability of the spectra. Cal-G had an average RPD that was lower than cal-A (indoor averaged shade and sunny) and cal-F (average-indoor and average-outdoor spectra). Cal-G did have the lowest absolute mean bias of all the calibrations, but cal-G generally did not improve the predictive abilities of trait models.

The resulting suppositions are that the predictive ability of trait models will decrease when spectra from separate sides of the same fruit are not averaged, and increase when collective scans of the same fruit in different situations or locations are included as an average. The calibration with the highest RPD for at-harvest time trait predictions combined averaged indoor spectra with averaged outdoor spectra (cal-F). Therefore, separate scans of individual sides of the fruit should be averaged for model development. Variation in the fruit is likely confounding the model-fitting algorithm when sides are used separately. Also, more scans from more locations could help in controlling measurement error inherent to the spectrometer (Esbensen et al., 2013) or due to location (e.g., indoor or outdoor).

5.2.1.4 Postharvest Trait Prediction Models Using Before- and After-Storage Spectra

Some apple cultivars can be stored up to several months before reaching consumers. Therefore, understanding the quality of fruit throughout storage is essential. In this thesis, poststorage traits were predicted using models calibrated with spectra acquired before or after storage.

General postharvest results

The results of calibrations predicting postharvest traits were promising. All calibrations for poststorage traits (cal- I, J, K, L) had lower mean-adjusted RMSEP values relative to the at-harvest trait predictions (Table 5-1). This is probably due to the changes in distribution of texture and firmness traits of stored fruit compared to at-harvest fruit. The expected decrease in texture quality for stored fruit was shown by the lower mean values for all MDT traits, except C0, which had the expected increase. Also important to consider is that the postharvest standard deviation increased for all destructive traits, except DM, TA, and Phen.

Postharvest spectra

The calibration that used poststorage spectra (cal-I) had an average RPD much higher than the RPD of the average indoor at-harvest calibration (cal-A). Five of the postharvest trait models had lower RMSEP values ($p < 0.03$), and three had higher RMSEP values ($p < 0.02$), compared to the at-harvest models (cal-A).

The RMSEPs for C0 and QF in cal-I were greater compared to those with cal-A, but the RPDs for C0 and QF in cal-I also were greater compared to with cal-A. The opposite was true for TA. At first glance, these results give conflicting impressions; however, as is also noted above, the standard deviations of most traits destructively assayed increased poststorage. In this case, the increase in model error was accompanied by an increase at a different rate for the standard deviation. The RMSEP for the poststorage C0 model increased 48% while and the standard deviation of C0 measurements of the poststorage fruit increased 83% over the at-harvest model.

The RPD metric considers both SD and RMSE, and for the postharvest model, the larger increase in the standard deviation had more effect than the RMSE. A change of 40-80% in the spectral absorptions is unlikely. The spectral values were not likely more attuned in postharvest models but, rather, the increase in postharvest model RPD over the at-harvest model is more likely related to changes in the trait data. For example, the amount of variance explained by the postharvest model (cal-I) increased by 50% over the at-harvest model (cal-A). Furthermore, the C0 trait is more easily predicted poststorage when there is sufficient variation in the trait than at harvest, while the opposite would be true for TA.

At-harvest spectra

McGlone et al. (2003) suggested using models built from spectra taken at-harvest to predict trait values poststorage, specifically for SSC. Kumar et al. (2015) also built models using spectra at-harvest time to predict poststorage trait values. Kumar et al. (2015) also used outdoor spectra taken in the orchard at-harvest to validate their models. Therefore, to test the feasibility of using harvest-time spectra to predict poststorage trait vales, at-harvest indoor and outdoor spectra (as separate calibrations) were used to calibrate poststorage trait prediction models (McGlone et al., 2003, 2002a). At-harvest indoor and outdoor spectra were used separately as well as in a combined calibration.

At-harvest spectra did not produce postharvest trait models with higher predictive abilities compared to models calibrated with postharvest spectra. The calibration using at-harvest indoor spectra for poststorage trait predictions (cal-J) had a lower mean RPD (all traits) and five traits had higher RMSEP values ($p < 0.0467$) than cal-I (postharvest spectra). However, cal-J did have a higher average RPD (all traits) relative to at-harvest calibrations. The calibration with outdoor spectra (cal-K) expectedly had an average RPD lower than the calibrations with at-harvest indoor spectra (cal-J) and poststorage spectra (cal-I).

Since the calibration using both averaged indoor and averaged outdoor spectra (cal-F) had good predictive ability for at-harvest traits, the same method was also used to predict poststorage fruit traits (cal-L). This at-harvest indoor-outdoor calibration predicting postharvest traits provided the second highest RPD and second lowest RMSEP relative to all other calibrations. There were also fewer trait models (2) that were statistically

different to cal-I (postharvest spectra) relative to both cal-J (indoor, 5) and cal-K (outdoor, 11).

Literature comparison

Postharvest NIRS prediction has received considerable attention in the literature (Giovanelli et al., 2014; Kumar et al., 2015; McGlone et al., 2002a). The study of Kumar et al. (2015) had poststorage RPD values on a seedling level for SSC between 1.92 and 1.11, and DM between 2.31 and 2.27. On a fruit level, my postharvest SSC RPD values were 1.44 to 1.63 and 1.28 to 1.46 for DM. My SSC results were higher, and DM results lower, than those of Kumar et al. (2015).

Zhang et al. (2019) calibrated SSC and DM models using the F-750 spectrometer for each of eight individual cultivars, as well as for a mixture of cultivars. The calibrations were validated internally with the same year's data and externally with a second year's data (summarized in Table 5-4). The predictive abilities for their calibrations were much higher for SSC and DM in general, except for the lower values in single cultivar calibrations where the SSC and DM results in this thesis were similar. The higher predictive abilities of the Zhang et al. (2019) models are likely due in part to the acquisition of trait data from fruit tissue at the scanning positions. Zhang et al. (2019) pooled at-harvest and postharvest spectra and trait values for the calibration dataset, therefore it is problematic to compare their calibrations to those in this thesis where all calibrations in this thesis were created using separate at-harvest or postharvest trait datasets.

Table 5-4. Summarized RMSE and RPD* values for soluble solids content (SSC) and dry matter content (DM) reported by or calculated based on Zhang et al. (2019)

Calibration-validation	SSC		DM	
	RMSE	RPD*	RMSE	RPD*
Mixed cultivars-Internal	0.54	2.96	4.40	4.45
Single cultivars-Internal	0.39-0.58	1.79-2.76	3.29-4.8	1.64-3.36
Mixed cultivars-External	0.67	3.13	5.88	3.62
Single cultivars-External	0.49-1.32	1.06-2.04	5.31-12.81	1.01-2.03

*RPD values calculated based on information in Zhang et al. (2019)

Giovanelli et al. (2014) collected 280 'Golden Delicious' fruit out of a commercial storage room over several storage time periods and over two seasons. They found very reliable predictions of SSC with an RPD of 3.33 (my calculation), though their study used a very small sample size for model calibration. Giovanelli et al. (2014) also used LDA to classify fruit according to storage time, and found an average accuracy in cross validation of 93.7%. McGlone et al. (2002a) created models from spectra taken by a custom-built spectrometer over two years from a single cultivar (approximately n=900). Between the five measured traits, the resulting RPDs ranged from 1.5 to 2.35 for at-harvest models, and 1.6-1.8 for storage models. The results I present here are similar to those of McGlone et al. (2002a) for overall postharvest calibrations, but my trait models taken individually have lower predictive abilties compared to those of previous authors.

All the same, postharvest trait model calibrations were generally as high or higher in predictive ability relative to at-harvest calibrations. The average adjusted R^2 and RPD for the poststorage calibration using poststorage spectra (cal-I) were the highest compared to all other calibrations, which is expected since the fruit are in a less dynamic state after weeks of cold storage. Also the high predictive abilities for poststorage trait models could be due to the different distributions of trait values in postharvest fruit. Stored fruit appear to have greater variation in some traits, allowing the calibrations to better model the

observations. Thus, the use of this particular spectrometer for postharvest trait analysis is certainly feasible.

5.2.1.5 Influence of Temperature

The work of Peirs et al. (2003a) showed that without controlling for temperature when calibrating models, either with temperature-specific models or inclusion of temperature as an x variable (or y variables as in PLS2), predictions of SSC could be affected by up to 4 units (%). In the case of this thesis, temperature could therefore affect approximatly 30 % of the mean for the SSC value. Temperature affects how NIR is absorbed and scattered in fruit, most specifically to wavelengths related to water (reflectance peak near 970 nm) and chlorophyll (peak near 670 nm) . Peirs et al. (2003a) showed that with increasing temperature, the relative reflection for the chlorophyll-related wavelength band increased, while the OH-related bands decreased. OH- spectral peaks are also expected to shift (Hansen et al., 2000) with temperature changes, but Peirs et al. (2003a) explained that perhaps the complexity of fruit tissue masked such shift effects.

In this thesis, calibrations with temperature variables, cal-M (both temperature variables), cal-N (fruit temperature) and cal-O (spectrometer temperature) and the calibration without temperature variables (cal-F) were not statistically different. Since this experiment was not designed to test the predictive performances at a range of temperatures, it could not be shown that the temperature variables (spectrometer and fruit surface temperature) affected the predictive abilities of the trait models. To thourghly determine the influence of temperature, spectral scans of the fruit at a range of controlled temperatures would need to accompany the inclusion of temperature variables.

A possible reason why the temperature variables had no effect on the trait models is that the fruit surface temperature was recorded for every sixth fruit and the rest of the fruit temperature values were imputed using the value from the nearest neighbor. As such, the temperatures not acquired empirically could be incorrect. Another possiblity for not seeing a temperature effect is that surface temperature might be different than the flesh temperature at the full depth of NIR penetration.

However, the spectrometer may inherently control for temperature effects, as it is used in variable temperatures, and by the design of its calibration or standardization scans of the reference shutter and dark current (Felix Instruments, 2015). Kumar et al. 2015 used a handheld spectrometer (predecessor to the F-750) outdoors in variable temperature conditions between 15°C and 23 °C, and due to their validation results, assumed the spectrometer was robust against temperature variation. Their good validation results could also reasonably be due to the sufficient distribution of the range of temperatures when the spectrometer was used. This reasoning was also used by Beghi et al. (2013), as they also collected spectra outdoors between the temperatures of 15 °C and 25 °C. To the best of my knowledg,e the work of Peirs et al. (2003a) is the only study to thoroughly control for temperature via inclusion of it as an x-variable in apple spectroscopy modeling. All other studies attempted to control fruit temperature by allowing the fruit to acclimate to room temperature after harvest, or did not control for possible temperature effects. Overall the F-750 spectrometer utilized in this experiement appears to sufficiently control for varying temperature conditions to allow outdoor use.

5.2.2 Performance of calibration models for traits

In this section, the predictive abilities of models for individual traits is discussed and compared to the results of other studies. The most common traits other than SSC to be estimated by NIRS in the literature are firmness, TA, DM, phenolics or flavonoids, chlorophyll, maturity indicators, individual sugars or acids, and pH. Many of the traits measured here are commonly studied, but others, such as fruit weight, russeting, fruit diameter, and unique MDT firmness traits have not been previously studied.

Of the traits modeled, only diameter, %Ocolor, SSC, SPI, C0, Wt/Diam, and weight had RPD values above 1.5. Models in the range of 1.5 to 2 predicted trait values with accuracy sufficiently to have confidence in discriminating observations into low and high trait groups (Nicolaï et al., 2007a). Models for most of the traits did not reach the suggested RPD levels to be useful for coarse quantitative predictions (Table 2-1), except for %Ocolor, which had an RPD above two in most calibrations.

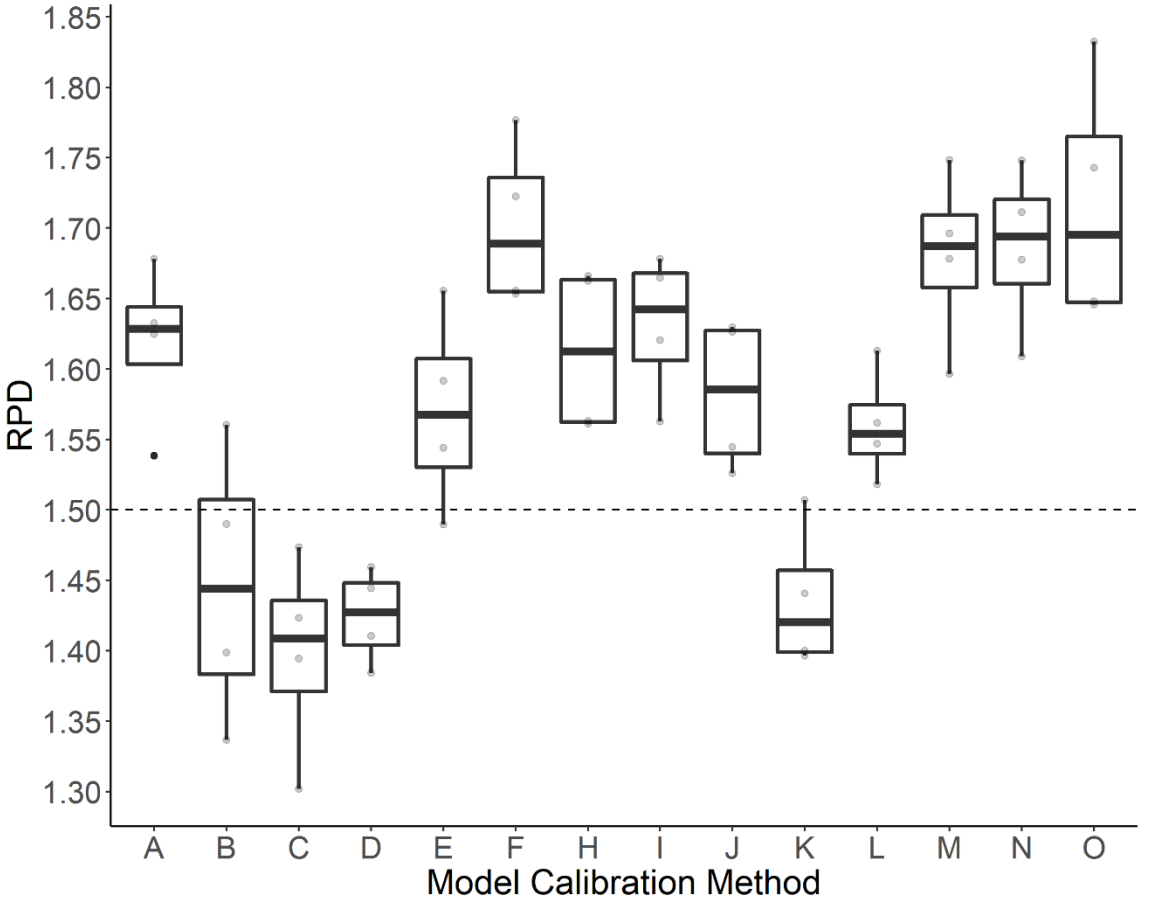
Over-color (%Ocolor)

Despite having the highest RPD, the model for %Ocolor is not useful in a practical sense. Rating the amount of the fruit covered by red is not a difficult or essential task until large quantities of fruit need to be sorted at the same time in a packinghouse. The prediction plot (Figure A- 20) also shows one nuance of using this model, specifically that the high R^2 was based on two extreme groups, the yellow fruit of 'Honeygold' and all other red-fruited cultivars. When 'Honeygold' fruit were not included in the data, the performance of the %Ocolor model was poor (data not shown). Large commercial packinghouses currently equip their sorting lines with sensors that successfully sort fruit by color. It is doubtful that an additional predictive model from a NIR sensor on a handheld device could be of much practical use.

SSC

In general, the F-750 spectral models performed at a lower level than reported in the literature for lab-based instruments. The RMSEP for SSC reported in previous studies ranged from 0.23 to 0.50 in previous studies (Liu and Ying, 2005; Lu et al., 2000; McGlone et al., 2003; OUYANG et al., 2012; Walsh et al., 2004). The RMSEP ranged from 0.91 to 1.13 for SSC in this thesis, which was approximately 7 to 9 % of the validation mean. The lowest error was found in the model calibrated with both averaged indoor and averaged outdoor at-harvest spectra (cal-F). The RMSEP results for SSC presented here are close to what was reported by Ventura et al. (1998) as 1.10-1.18, and Sun et al. (2009) as 0.82 using portable spectrometers, and by Bureau et al. (2012) as 1.4 using a lab mid-infrared instrument. Beghi et al. (2013) reported lower RMSEVC values of 0.72-0.78 for SSC of two cultivars using orchard spectra obtained from their "backpack' spectrometer compared to values found in this study.

Figure 5-1. Performance of SSC by model calibration method



*RPD values from the four data splits are summarized in the box plots.

Dashed line is RPD=1.5.

For definition of the different calibration, methods on the x-axis see Table 5-1.

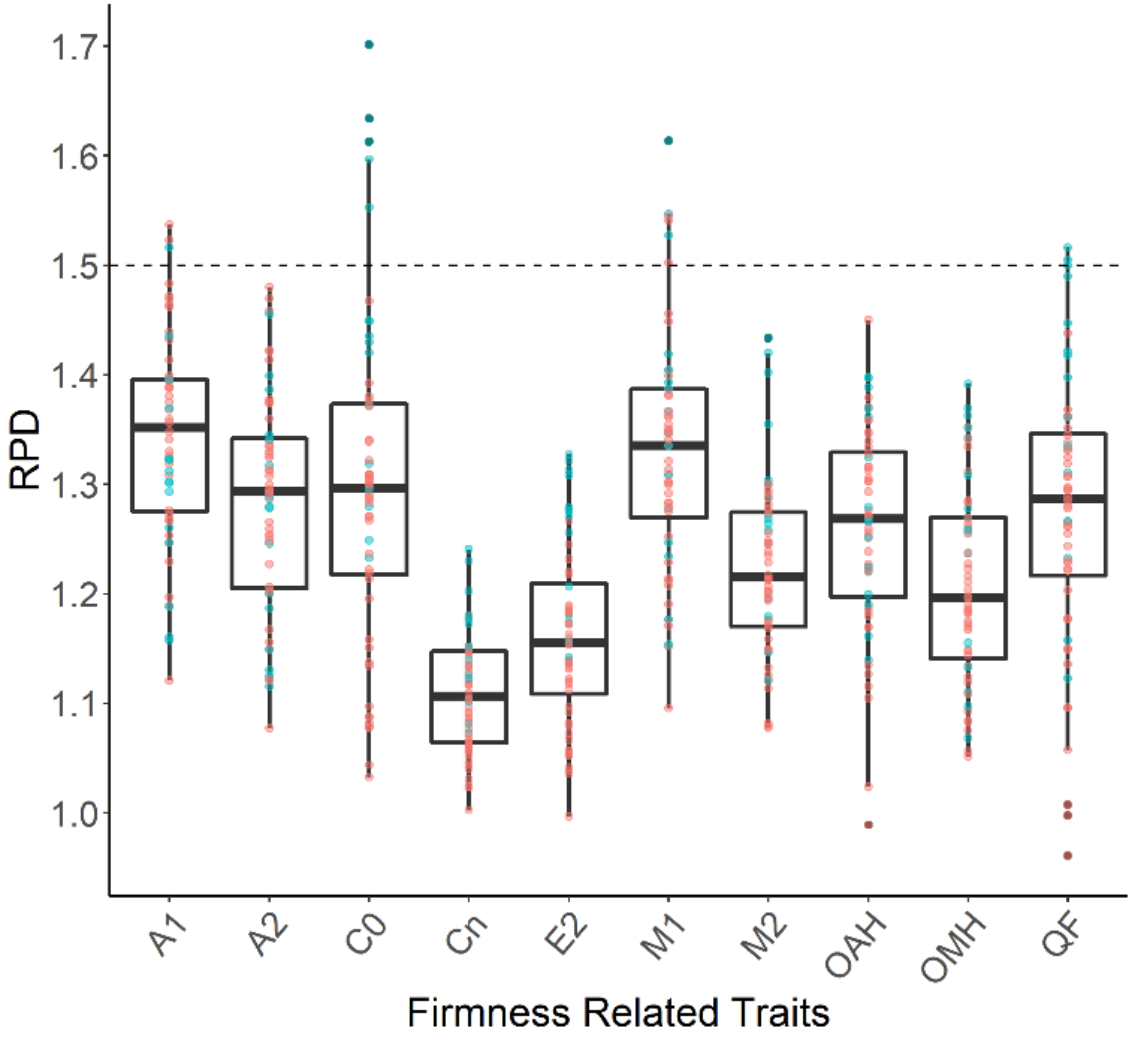
Using the F-750 handheld spectrometer, Zhang et al. (2019) presented SSC RMSEPs for individual cultivar models of between 0.39 with poststorage internal validation, and 1.32 with poststorage external validation, Table 5-4. Kumar et al. (2015) used a Nirvana spectrometer, an instrument preceding the F-750, and published a RMSEP on a seedling level for breeding germplasm between 0.71 and 0.90, and on a fruit level between 1.04 and 1.15. Considering that Zhang et al. (2019) used 8 individual cultivars in their models, and Kumar et al. (2015) used results at a seedling level for 279 genotypes, the SSC results I present here performed comparatively well, given that the model statistics were on a fruit level using 14 cultivars.

Firmness

Apple fruit firmness and texture-related traits are among the most salient traits after color and flavor to influence consumer preference (Harker et al., 2008b, 1997). Estimating firmness by NIRS has been plagued in the past by inconsistent or poor performance. This may partly be due to the imprecision of handheld penetrometers. McGlone et al. (2002a) asserts that "a credible theory or mechanism does not exist on which to base" a NIR model on fruit firmness, though many authors claim relationships of fruit texture with the scattering properties of fruit tissue (Nicolaï et al., 2007a). Despite the issues and debates found with firmness and textural traits in NIRS models, research will likely continue for a nondestructive method to reliably estimate these traits due to their important role in fruit quality assessments.

The several firmness and texture measures obtained from the digital penetrometer MDT-1 were poorly predicted in this study. In all of the model calibration methods, most of the texture traits did not have an average RPD value above 1.5 (Figure 5-2). The C0 trait models had the highest predictive ability of the texture trait models, but it also had a large range in RPD values. A1 and M1 models had higher RPDs on average than models for A2, M2, and E2, showing that perhaps the variables that are measured furthest inside the fruit had models of lowest predictive ability due to less NIR penetration at the depth of the MDT-1's region two (8 mm from fruit surface to the core).

Figure 5-2. Range of RPD values for firmness related traits



Ranges include 4 individual RPD values from each model calibration method. Dashed line = RPD of 1.5, blue dots are RPD values from models predicting postharvest traits, red dots are for models predicting at-harvest trait RPD values

Peirs et al. (2000) predicted firmness from a collection of two years of NIR spectra from seven cultivars and three orchards. They reported SEPs between 0.90 and 1.13; however, the Magness-Taylor penetrometer utilized in their study had a low maximum threshold of measurement, which is probably why the errors were so low. Paz et al. (2009) compared the predictive abilities of three spectrometers for firmness measured with a 1000 N load cell. The RPD values for the three instruments ranged from 1.11-1.40, 1.42-0.99, and 1.31-1.46. The firmness model by Beghi et al. (2013) had an RPD of 1.54. Giovanelli et al. (2014) had an abnormally high RPD for firmness of 3.52, though from a small sample

size of only a single cultivar. McGlone et al. (2002a) reported an RPD of 1.6 for firmness.

More research is needed to identify mechanisms affecting the absorbance (and light scattering) to firmness (and texture) relationship. Perhaps by knowing the mechanism, more targeted firmness models could have higher predictive abilities. Conceivably, like the innovative ideas described by Costa et al. (2011) or Mendoza et al. (2012), using a combined acoustic-mechanical-spectral method could be used to study fruit texture.

Diameter

Diameter was measured by the MDT-1 instrument at the same time as the other texture traits. The predictive ability for fruit diameter was surprisingly high when compared to the other MDT-1 variables. The range for the average diameter RPD for model calibrations was between 1.35 and 1.80. There is no immediately apparent reason that diameter of an object could be observed by the spectra. On a theoretical level, perhaps the curvature of the fruit, as a function of diameter, affects the spectral acquisition because the skin at the center of the scan circle will be closer or further from the instrument depending on the diameter see Figure 5-3 for an illustration.

Another, more plausible reason why diameter had better relative performance is that it could be correlated with some other fruit characteristic that was correlated with the spectra. Cultivar differences could likely drive the regression of diameter, especially since two cultivars, Frostbite and Chestnut, were smaller than all other cultivars used. Moons et al. (2000) attempted to predict fruit diameter using a mock in-line grading NIR and reported a SECV of between 0.14-0.17, though among the other variables tested (SSC, TA, pH, firmness, and DM), diameter had the least consistent variation of the standard error when scanned multiple times. Their reasoning for creating a model for diameter was that there was enough background area visible to the spectrometer to discriminate the illuminated fruit from the background.

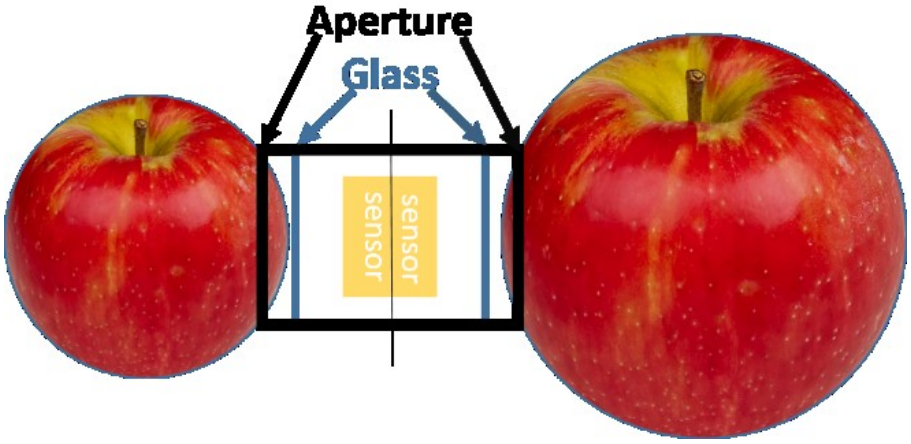


Figure 5-3. Simplified diagram of spectrometers with fruit of different diameters. Note the distance between the fruit and the glass & sensor.

Weight

The RPD for the fruit weight model ranged from 1.22 to 1.70, having a similar pattern but lower predictive ability than that for diameter. I am not aware of any other reports of spectral models created for fruit weight. The relatively accurate predictions of weight and diameter suggest that those traits most likely co-vary with other fruit attributes. Chestnut and Frostbite are two cultivars that are much smaller on average than other cultivars. Therefore, the spectral characteristics of those two cultivars likely generated the high correlations in the models.

In considering the two traits of weight and diameter together, perhaps the density of the fruit affected the NIR light scattering or penetration. McGlone et al.(2002b) used NIR and density to predict kiwi fruit dry matter, and Schaare and Fraser (2000) used NIR to estimate kiwi fruit density with an R^2 of 0.59 - 0.74 and a SEP between 3.6 - 4.8 (kg m^{-3}). To test whether fruit density interacted with apple NIR spectra, two mock variables were calculated. Density was calculated by taking the fruit weight divided by the fruit's spherical volume$\left(v = \frac{\pi * diameter^3}{6}\right)$. Weight diameter ratio was additionally calculated. The predictive ability for density was very low (data not shown), while the RPD for the weight diameter ratio ranged from 1.22 to 1.68 with the different calibration methods. Texture and firmness traits were not predicted with sufficient accuracy by NIR nondestructive spectral models, but prediction of fruit density is perhaps worth more investigation.

TA

The average RPDs for different model calibrations for TA presented in this thesis ranged from 1.10 to 1.29. These results for TA were expected, since the acids in apple fruit are found in low concentrations. NIR predictive models for TA have been evaluated often, but results are usually low in published studies. Using bench top spectrometers, Peirs et al. (2000) reported SEPs between 1.73 and 2.30 for acidity. Liu et al. (2005) reported RPD values of 1.36-1.43 (my calculation). MartínezVega et al. (2013) created models for TA using two cultivars with RPDs ranging from 1.11-1.25 (my calculation). Beghi et al. (2013) developed models from portable spectrometers for two cultivars with RPDs for TA from 1.49 -1.81, and ascorbic acid from 2.55-8.43 (very high due to the low SD of ascorbic acid in 'Stark Red Delicious').

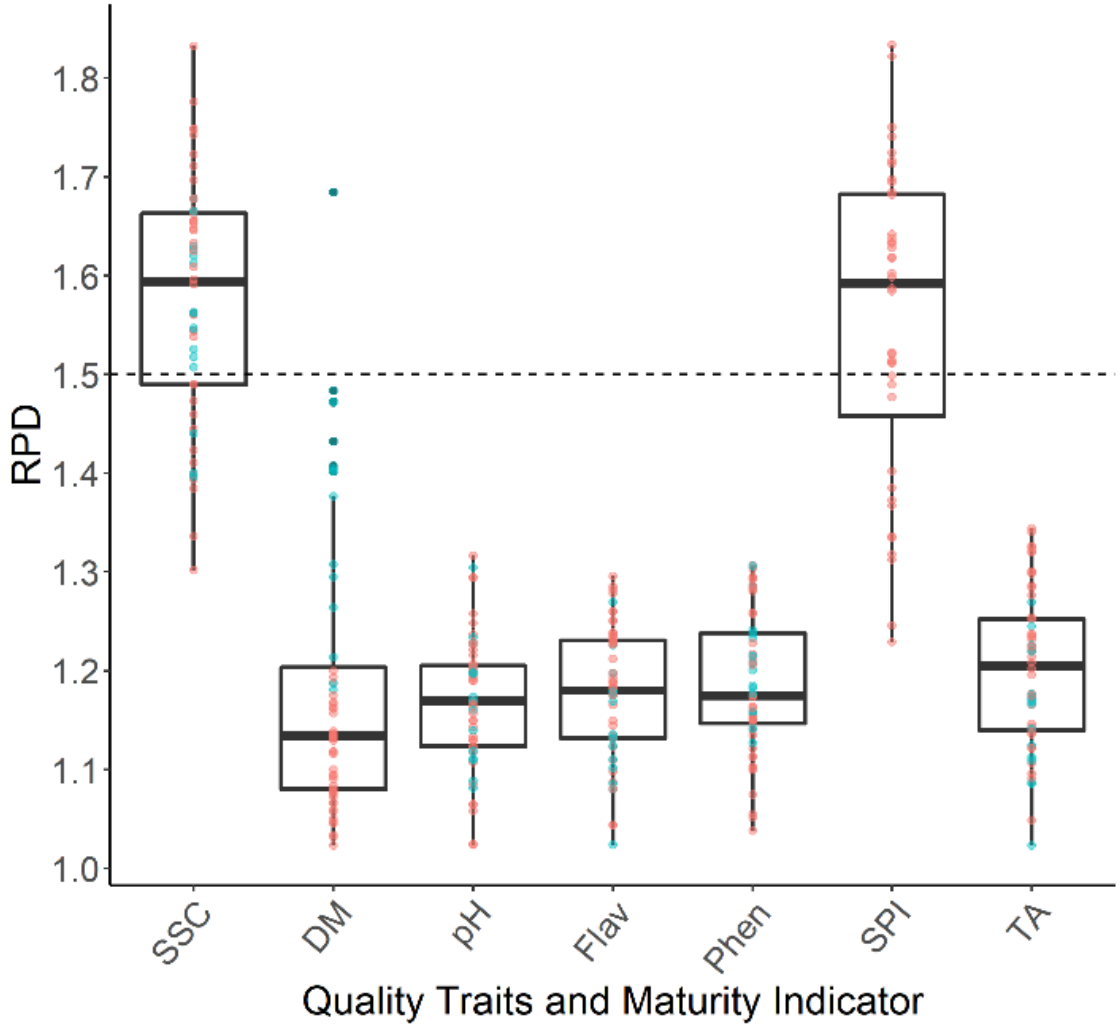
There is evidence that after concentrating fruit into juice or liquid, the prediction ability for an acid model increases. Bureau et al (2012) reported model RMSEPs of 3.74 (mequiv kg^{-1}) for TA, and 0.23 (g kg^{-1}) for malic acid, using a bench top spectrometer to scan vials containing liquid samples. Modeling citric acid concentration was attempted, but the concentration was below the measureable amount, so predictive ability of the model was not assessed. Acidity remains an apple fruit trait for which accurate nondestructive estimation is unattainable.

pH

The pH of apple fruit is the ratio of H^+ to OH^-, usually easily measured from juice along with TA. The pH of apple fruit influences the perceived fruit flavor. Dever et al. (1995) found the highest correlation between pH and sourness when examining different sensory (juiciness, crispness, firmness, fruit flavor, sweetness and sourness) and analytical variables (TA, SSC, flesh and skin colors, weight, and internal ethylene). Lower pH values may also lessen the rate of browning of the fruit flesh and juice (Huang et al., 2018; Sun-Waterhouse et al., 2011). Although pH could be a relatively important trait to estimate nondestructively, the models evaluated in this thesis had low predictive ability for pH. The RPDs for pH models were slightly lower than for TA models, with a range of 1.07-1.23, and RMSEPs of 0.23-0.18. This predictive ability was a large degree lower than the RPD values for pH reported by Liu et al. (2005) between1.59 to 1.78 (my

calculation) and RMSEPs of 0.0044-0.0042. Moons et al. (2000) and Lammertyn et al. (1998) reported better predictive abilities than Liu et al. (2005) with SECVs between 0.10-0.14, and SEPs between 0.062-0.079, respectively. The data in this thesis lead to the conclusion that NIR estimation of pH is similarly impractical as modeling TA.

Figure 5-4. RPD values for quality traits and starch index



Includes 4 individual RPD values from each model calibration method. Dashed line = RPD of 1.5, blue dots are RPD values from models predicting postharvest traits, red dots are for models predicting at-harvest trait RPD values

SPI

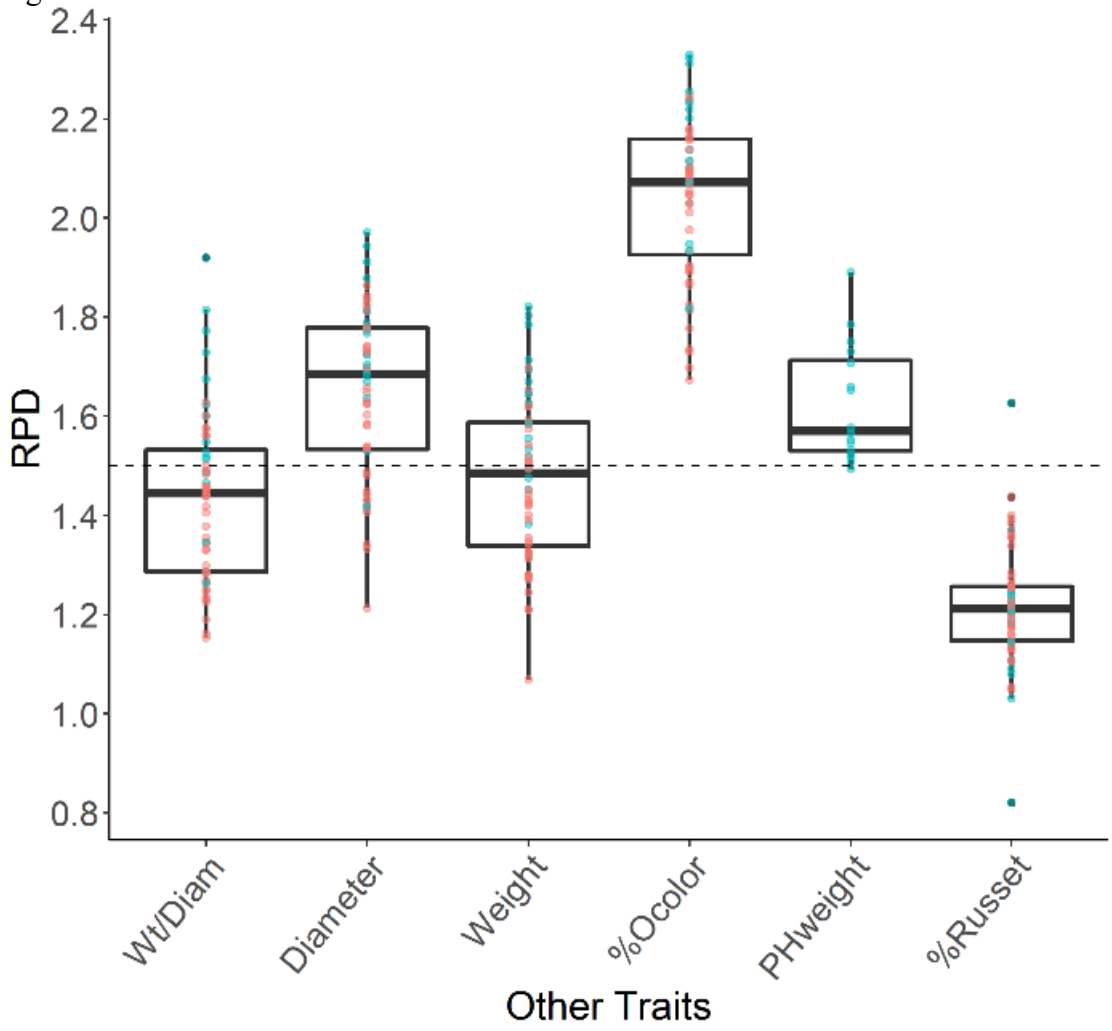
The starch pattern index (SPI) is primarily a maturity indicator based on colorimetric estimation of starch distribution in fruit. It is among the few tools available to estimate apple fruit maturity. The decision of when to harvest fruit is paramount to estimate postharvest fruit quality. The predictive abilities of models evaluated in this thesis for SPI were very encouraging, with an RPD range between 1.31 and 1.69. SPI was usually among the most accurately predicted traits for each model calibration method. The lowest RPD values of 1.31, 1.36, and 1.48 were from the models built with the genetic algorithm, and the single shaded and sun-exposed sides. In all other model calibration methods, SPI had RPDs above 1.51. The RMSEPs ranged from as low as 1.17 to 1.5 units on the scale of one= immature to eight= over mature. An approximate 95% confidence interval ($\pm\ 1.96\ \times\ RMSEP$) for the best predictions was within ±2.30 units. Furthermore, despite the moderately high RPD values, this confidence interval was probably too large for basing economically important harvest decisions. This result is discouraging, considering the potential benefit of nondestructive maturity estimation. For example, more accurate methods of estimating fruit maturity could lead to substantial reductions in food waste throughout the supply chain (B. Li et al., 2018).

McGlone et al. (2002a) reported SPI model performance with RPD of 2 and RMSEP of 0.95, very similar to the results reported in this thesis. The study of Peirs et al. (2000) claims to be the first report of predicting harvest date by NIRS, as the SEP for time before the optimal harvest date was between 5.81-9.05 (days). Bertone et al (2012) used NIRS along with SSC, firmness, and starch content (Lugol test) to estimate the optimal harvest date, but their sample size was small (~300 fruit) and only of a single cultivar. The RMSEPs were 1.7 (on a 1-10 scale) for prediction of starch content, and 2.8 days for the optimal harvest date.

A major issue with using SPI in NIRS models is the high degree of subjectivity and bias from the person doing the scoring (Peirs et al., 2002). Another issue is that starch patterns are cultivar-specific, sometimes requiring their own scale due to biological differences in starch conversion distribution and rates (Doerflinger et al., 2015). The often-used scale by Blanpied and Silsby (1992) is based on ‘McIntosh’, and scales for other cultivars are

available in the literature, including ‘Granny Smith’ (Reid et al., 2017), ‘Delicious’ and ‘Northern Spy’ (Smith et al., 1979). More accurate methods digitally quantifying starch stained in a fruit slice could perhaps improve performance in NIRS models (Doerflinger et al., 2015; Menesatti et al., 2009).

Figure 5-5. RPD values for other traits



Includes 4 individual RPD values from each model calibration method. Dashed line = RPD of 1.5, blue dots are RPD values from models predicting postharvest traits, red dots are for models predicting at-harvest trait RPD values

DM

In apple, interest in predicting dry matter (DM) content is fairly new, and the theory of its use as a quality trait is driven mostly by a research group from New Zealand (McGlone et al., 2003; Palmer et al., 2010). They proposed that at harvest, DM relates to postharvest SSC due to starch hydrolysis during storage, converting insoluble to soluble solids. At harvest, unconverted starch is measured as DM, therefore the relationship is assumed to best be suited to poststorage SSC. However, there have been relatively few cultivars studied in regards to the utility of DM. The New Zealand studies involved 'Royal Gala' and one of its progeny, 'Scifresh'. More recently, Travers et al. (2014) studying 'Elshof', and Robinson et al. (2017) studying 'Honeycrisp', found DM was not a useful predictor of SSC or consumer likability.

The protocols for DM measurement in this study were not consistently followed, resulting in data that was too unreliable to permit credible interpretation of results. For the first month of data collection, the main issues were inconsistent position on the fruit where the sample was collected, size of the sample, and samples with and without skin. However, the largest issue was the use of different drying ovens where one particular oven did not remove moisture sufficiently from inside the oven. The model performance was poor for DM, which had RPDs ranging from 1.06 to 1.46 and average RMSEP values between 1.45 and 2.18. The models were not sufficient to be of much utility. It is likely that the sampling method affected the results, as well as the inconsistency of the drying ovens. For future research it would be interesting to explore the use of DM and its NIR predictive ability among cultivars sharing similar 'uniqueness-es' (Robinson et al., 2017) to 'Honeycrisp', and compared to 'Gala' and its relatives.

6 CHAPTER 1. SUMMARY

The aim of this experiment was to explore the utility of a handheld NIR spectrometer by assessing the predictive ability of multivariate models for several fruit quality traits, and to determine limitations of the instrument's practical use. The fruit traits investigated, common to those in published literature were soluble solids content (SSC), dry matter content (DM), and titratable acidity (TA). Traits investigated with few reports in the

literature were fruit diameter, pH, total phenolics (Phen), total flavonoids (Flav), and starch pattern index (SPI). The traits investigated that are not already reported in previous investigations, to the best of my knowledge, were percent over-color (%Ocolor), fruit weight (weight), weight-diameter ratio, percent russet (%Russet), and the firmness and texture traits (A1, M1, A2, M2, E2, OAH, OMH, C0, Cn, QF).

The F-750 handheld spectrometer showed utility for predicting the values of SSC and SPI across a dataset of 15 cultivars with RPD values indicating the ability to discriminate between low and high groups. Individual cultivars were classified with an accuracy of 99.06%. Consequently, other traits with average RPD values above 1.5, likely had higher RPD values attributable to cultivar-driven differences (e.g. small vs. large average size, or the presence vs. absence of over-color). The distribution of trait values changed poststorage, which improved the training of poststorage trait models to have similar predictive abilities to those described in the literature. The calibration with the highest at-harvest predictive abilities for trait models combined outdoor and indoor spectra. Temperature and outdoor limitations of the F-750 spectrometer were determined to be minimal compared to the importance of collecting more than a single scan per fruit to control for individual fruit heterogeneity.

7 CHAPTER 2. TREE ARCHITECTURE TRAITS

Introduction

Apple breeding is mainly focused on quality, and not primarily yield-driven; however, in cultivar commercialization, pomologists and growers conduct considerable research on the efficiency of cultivation (Palmer, 2011). An orchard is a long-term investment, and it is essential to make the best use of land (Bruce H. Barritt, 2000; Hoying and Robinson, 2000) that will be committed to the tree crop for potentially 2-3 decades (Lordan et al., 2018).

Tree architecture (defined in this introduction as tree shape, size, and configuration as determined by the spatial location of branches and trunks) is linked to the cultivation efficiency of fruit trees. Despite substantial efforts by researchers to thoroughly capture the related fruit and tree metrics, comprehensively phenotyping or characterizing tree architecture traits can be a difficult task due to the size, complexity, and long life of the plant (Costes and Gion, 2015). Owing to the difficulty of comprehensively measuring tree parameters, simple relatively quick measures such as tree height and trunk cross sectional area (TCA) have been relied on for years (Autio et al., 2017; Westwood and Roberts, 1970). These simple tree measures are limiting because they fail to provide more than a static one- or two-dimensional (i.e., more comprehensive) status of the architecture of individual trees.

Researchers are harnessing new technology to tackle the complex task of comprehensively assessing tree architecture. Some of the technology includes a variety of sensors, cameras, drones, and robotics. The technology and research has widespread applications in horticulture (Ahn et al., 2018; Lee et al., 2010; Zude-Sasse et al., 2016). Including the estimations of yield (Gong et al., 2013), harvest date (Wendel et al., 2018), produce quality (Halstead et al., 2018), plant health (Usha and Singh, 2013), physiological traits (Virlet et al., 2015), plant size (Rosell Polo et al., 2009), plant position for dynamic spraying (Llorens et al., 2010), and machine harvesting (Bac et al., 2014).

In this thesis, I provide an overview of some areas in fruit tree research that could benefit from technological advances in assessing tree architecture including, yield and efficiency metrics, canopy light dynamics, rootstock-scion interactions, training systems, and plant protection products.

Second, an experiment exploring image-derived tree volume and tree height with the specific horticultural application of assessing the yield efficiency of different rootstocks is discussed.

7.1 Areas of apple tree architecture research

Architecture

Tree architecture, particularly the spatial location tree branches and trunks, is a complicated area of tree physiology (Dejong, 2017; Jones, 2017; Lakso, 2017; Palmer, 2017). Studies of tree architecture commonly focus on one of two themes 1) the effects of management practices on tree architecture and the resulting yield, cost, and efficiency associated and 2) the genetics or basic physiological science behind how trees grow with minimal management (Atay et al., 2018; Kikuchi and Shiozaki, 2007; Lauri et al., 1995; Segura et al., 2006). The fundamental goal is to select individuals (scions, rootstocks and / or their combinations) with favorable architectural genetics and establish the best environmental situation for production.

Some areas of research related to tree architecture include: balanced seasonal fruit precocity and vegetative growth (Hirst and Ferree, 1995; Lauri et al., 2006), equal canopy light distribution (Willaume et al., 2004), favorable branching patterns (Costes and Guédon, 2002) for efficient management (i.e. pruning and training), and good carbon partitioning (Dejong, 2017; Stephan et al., 2008) between fruit and whole tree dry matter.

Yield

The encompassing theme of fruit tree research is understanding and improving yield efficiency (Robinson and Lakso, 1991) of high quality fruit. In pomological research, two examples of yield efficiency are 1) the yield in weight per tree unit (e.g. dry weight, height, leaf area, TCA, or volume), or 2) the yield per unit of a management input (e.g.

pruning, pesticide, irrigation, or fertilization regime). The yield efficiency definition examined in this thesis is the former, since pomological traits associated with tree size are related to the material, time, and energy resource inputs essential for the cultivation of fruit.

Fully characterizing mature trees with measurement parameters such as length, width, and height, is slow and problematic if not impossible using hand measurement devices. Consequently, yield efficiency has been estimated for decades as fruit yield (weight) per cross sectional area of the trunk (TCA) (Westwood and Roberts, 1970). This measure is quick and simple to acquire for large experiments where no other rapid alternative has existed to estimate efficiency. Researchers have found strong correlations between TCA and tree weight (Westwood and Roberts, 1970), tree dry weight (Strong and Azarenko, 2000), cumulative fruit yield (Barden and Marini, 2001), and pruning time (Barden and Marini, 2001). However, TCA-based yield efficiency models have limitations such as the degree of utility being dependent on tree age, and a less accurate relationship to yield when tree treatments involve excessive pruning.

TCA is highly accurate with regard to whole tree dry matter content, but after the trees reach their allotted size within a row, it becomes less informative (Kikuchi and Shiozaki, 2007; Robinson and Lakso, 1991; Strong and Azarenko, 2000). Once a tree has an established canopy, the canopy operates in a semi-independent nature to the trunk (Kikuchi and Shiozaki, 2007). Since fruit trees are a perennial plant in a confined space, annual pruning affects the relationship of TCA with yield efficiency (Nesme et al., 2005). Fruit trees are highly manipulated in their cultivated environments and allometric (mathematical relationships among body size) relationships may or may not hold true as manipulation increases with tree age (Brym, 2016; Lauri et al., 2006).

Light

Light dynamics in a canopy are important in providing the most efficient situations for photosynthesis and resulting fruit development (Robinson, 2017). The most limiting factor concerning yield is the amount of intercepted light (Palmer, 2017; Tustin, 2014; Wunsche et al., 1996). Light that is captured or intercepted in a canopy can be different from how that light is distributed among the canopy's available leaves (Palmer, 1989;

Sansavini and Corelli, 1992; Wagenmakers and Callesen, 1995). For example, if the tree has a very dense foliar canopy it may intercept a similar amount of light as a less dense canopy but the denser canopy will likely distribute the energy from that light to sustain different tree parts and functions than the less dense canopy. One method to measure light interception is to quantify the amount of light that penetrates a canopy. There are multiple sensors available to measure photosynthetically active radiation (PAR; range of 400- 700 nm) in a canopy. One sensor measures the photon flux density of PAR in units of μmol $m^{-2}s^{-2}$ using a ceptometer light meter (Rosell and Sanz, 2012).

Rootstock

Selecting the best scion-rootstock combination for an orchard is an essential decision for apple growers. Dwarfing rootstocks are critical to the commercial high-density orchard. A rootstock imparts what is referred to as a faster "physiological aging" (Lauri et al., 2006) to scions for earlier fruit bearing, reduced vegetative vigor, and height control. Resistance to soil diseases is another essential factor of rootstocks. A tree's architectural growth is largely influenced by rootstock, while the distribution of shoot growth is controlled primarily by the scion (Lauri et al., 2006). Before the 1970s, rootstock breeding and development took 20 to 30 years, due to the number of experiments required to collect sufficient data to support confident decision making by growers (Ferree and Perry, 1988). Each rootstock experiment might have had different objectives and complicating factors like soil and climate. These differences confounded comparisons that could be made between sites, which limited the ability for researchers to advise growers. A collaboration, which started in 1976 (Ferree, 1976), between North American research sites called the North Central 140 project (NC-140) facilitated the standardization of synchronous experiments across North America, dramatically shortening the time to acquire data on rootstock-scion interactions. Rootstock trials attempt to measure efficiency region-wide using yield and tree architecture aspects while holding cultural practices and training systems constant.

Training systems

The methods of arranging trees in space has received considerable attention recently (Robinson et al., 2013; Robinson, 2003; Tustin, 2014), but has been experimented with

for centuries (Lauri et al., 2009; Mika, 1992). Dutch and German farmers were among the first to plant orchards with high tree density a practice currently incorporated in orchard systems world-wide (Robinson, 2003). An orchard system can be arranged in an unlimited number of ways. Some of the recent most popular training systems include

- Palmette (Corelli-Grappadelli, 2000)
- Vertical Axis (Lauri and Lespinasse, 2000; Perry, 2000)
- Solaxe (Lauri and Lespinasse, 2000)
- V-Shaped (Robinson, 2000)
- Slender Pyramid (Tustin, 2000)
- Super Spindle (Weber, 2000)
- HYTEC (B. H. Barritt, 2000)
- Tall Spindle (Robinson et al., 2011)
- Bi-Axis (Dorigoni et al., 2011)

Among many other measures, training system trials attempt to measure efficiency by cost of installation, yield, planting density, light interception, and cultural management upkeep (i.e. time required for pruning, thinning, etc.).

Plant protection

The accurate characterizing of tree architecture is an essential step in developing efficient modern spray methods. Essentially, the rate of spray can be varied with modern technology to match the observed tree architecture measurements. Numerous substances are applied on or near a tree for seasonal management in both organic and conventional cropping systems. Orchard amendment examples include fertilizer/nutrients, plant growth regulators (PGR; bloom, vegetative, fruit thinning, and drop/maturity control), pesticides, and herbicides. Studies have been done (Dussi, 2011; Reganold et al., 2001) on methods to apply these products in a responsible and sustainable manner. In weed science for example, sprayer technology has received significant attention in determining the best nozzle types and droplet size on the effectiveness of herbicides (Miller and Tuck, 2014). Another example is a recognized goal of a 2009 Directive of European Parliament to research agricultural sustainability through the accurate application of pesticides at variable rates according to variable plant tissues in space (Rosell and Sanz, 2012). The

modern application of these substances involves the integration of the geometric characterization of tree architecture and spray science. The focus of plant protection research is searching for the minimum effective spray quantity with new emphasis on variable rates.

7.2 3D phenotyping of fruit trees

All aspects of fruit tree research recognize the necessity to quantify the efficiency of the orchard system. Many efficiency measures are very time-intensive (Stephan et al., 2008), require destructive sampling (Rosell Polo et al., 2009), and are imprecise (Best et al., 2008). The most widely used, quick and nondestructive measure involves TCA for comparing different treatments affecting tree size. While TCA may be a convenient denominator in efficiency ratios, it has limitations and accuracy issues, such as the degree of utility being dependent on tree age, and a less accurate relationship to yield when tree treatments involve excessive pruning.

Development of emerging sensor and imaging technology has been an exciting area of research for orchardists and has the potential to not only replace manual measurement, but also provide a level of precision previously unavailable (Vázquez-Arellano et al., 2016). Imaging systems have evolved from large complicated machines mounted on planes and tractors to simple devices connected by a USB port to wireless devices integrated into drones. Software is being developed that can acquire information from 3D objects from a smartphone camera (Confalonieri et al., 2013).Table 7-1, adapted from Rosell and Sanz (2012), reviews technologies used to characterize tree crops.

Table 7-1. Summary of 3D imaging technology for phenotyping tree crops, adapted from Rosell and Sanz (2012).

Sensor	Principle	Pros	Cons
Radar	Uses radiation in the microwave range	Independent of weather, measure large scale differences	Limited in spatial resolution, not accurate in height and volume
X-ray computed tomography	Uses radiation in the x-ray range	Accurately view plant in 3D	Plants <1m tall, expensive, x-ray source is a health hazard
Photography	Converts light via charge coupled device (CCD) or a complementary metal oxide semiconductor (CMOS)	Low cost, and reasonably accurate	Slow to analyze, 3D is difficult
Light sensor	Measures the light/shadow of canopy, can obtain 3D traits	Customizable, can measure size, shape and exposed area of canopy	Weather dependent, slow to acquire measures, 3D only after processing
Stereovision	Binocular camera, 3D point cloud built from algorithms	Most realistic and direct measure of size, volume; can obtain spectra, and GIS	Less accurate than laser, affected by variable light & weather, large data files
Ultrasonic	Uses sound waves	Low cost, robust, & easy to use	Limited resolution and accuracy, need many units to cover area
Light Detection and Ranging (LIDAR)	Uses laser (UV, VIS, or NIR) emitter, measures by phase shift or time of flight	Quick, high precision and accuracy, sufficient 3D images 3D Flash LIDAR is faster, smaller, measures to ± mm	Limited by dust, fog, wet; airborne expensive, needs additional sensors to correct errors, large data files

Sensing and image analysis can improve tree architecture phenotyping since it has the potential to:

1) be nondestructive
2) generate data throughout a growing season
3) relay information that is significantly correlated to architecture and canopy traits
4) provide real-time information in the orchard (as technology and software improve)

In order to compare treatment effects in many pomological studies some level of geometrical characterization is required. More objective phenotyping methods could improve the impact of fruit tree research. Training systems and rootstock effects could be

evaluated more accurately, estimation of the genetic basis of architectural traits could be improved through precise phenotyping, light interception and distribution could be tracked with high-resolution canopy parameters, and the rate of application of plant protection products could be variably applied according to dynamic canopy volumes.

Estimation of Tree Architecture Using Computer Vision

The aim of this research is to explore the horticultural application of data acquired from red-green-blue-depth (RGB-D) sensors. For that purpose, an experiment was conducted using an RGB-D camera on a study evaluating the effects of rootstocks on a common scion cultivar, Honeycrisp, at the University of Minnesota Horticultural Research Center to evaluate computer vision methods for measuring apple tree architecture traits. In this experiment, 247 ‘Honeycrisp’ trees grafted on rootstocks that potentially affected tree size and cropping efficiency were imaged several times throughout the 2017 growing season. Hand measures of tree architecture were compared to similar measures extracted from images. The specific objectives of this experiment were to 1) determine if image-extracted tree measurements were accurate and could replace manual measurements, and 2) test whether image-relayed information can explain variation for yield among the rootstock treatments. The important tree metrics were obtained by Wenbo Dong and colleagues from the University of Minnesota’s Robotic Sensor Network (RSN) lab. The RSN algorithms were developed to measure canopy height, width, and volume for groups and individual trees.

8 CHAPTER 2. MATERIALS AND METHODS

8.1 Manual tree measurement methods and materials

In 2017, tree architecture and yield data were collected on 247 ‘Honeycrisp’ trees grafted onto 31 different rootstocks at the UMN Horticultural Research Center (44°51'51.9"N 93°38'02.9"W; 316m). This rootstock trial was one of 13 across North America as part of a coordinated NC140 experiment planted in 2010 (Autio et al., 2017). The experiment was planned as a randomized complete block design with four blocks (replications) at each site. Each block contained one plot per rootstock with one to three trees per plot.

The trees were trained to the tall spindle system (Robinson et al., 2011), and in Minnesota the trees were planted in three rows running north to south with 1.2 m between trees and 3.6 m between the rows. All trees for the 13 locations came from a single nursery and, due to shortages, in Minnesota there were two rootstocks that lacked a plot in the fourth block.

The following architecture data were collected per tree at the end of the growing season on October 23 in 2017: trunk diameter at the graft union (approximately 76 mm above the ground), widest scaffold span in the row, widest scaffold span among rows, height of lowest scaffold, and total height of the tree. All traits discussed in the text, tables, and figures of this chapter are hand-measured unless it is described otherwise (as image-derived).

Total harvested fruit yield and average fruit weight were recorded for each tree. Yield per tree and average fruit weight were estimated by weighing and counting all fruit from a tree with less than 25 fruit. For trees having greater than 25 fruit, yield was recorded as weight of all fruit and average fruit weight was estimated by weighing a 25 fruit sample. The number of fruit that dropped was estimated by counting the number of fruit closest to the trunk base of each tree. Yield data were available for the years 2012-2017, and the main dependent variable used in the analyses was the weight of the cumulative 6-year yield for each tree.

8.2 Imaging Methods

An Intel® RealSense™ R200 sensor (Intel Corp., Santa Clara, CA) was used to acquire images of both east and west sides of three rows of ‘Honeycrisp’ trees in the rootstock trial. The R200 RBG-D sensor was used because it was inexpensive compared to LIDAR systems. The R200 contains two stereoscopic infrared sensors for depth, a red-green-blue (RGB) color camera, and a laser projector for scene perception. It operates by USB connection to a computer. Image collection in the orchard was done by holding the sensor attached to an extended ‘selfie stick’ roughly 2.5 m high. The sensor was angled down to capture the top view of the canopy. Images were collected separately on each side of a row, producing about 8,000 images per side or approximately 16,000 per row. The date of tree image collection was on day prior to harvest (09/17/2019).

Collaborators at the RSN Lab (Dong and Isler, 2018) developed the computer vision system to three dimensionally reconstruct orchard rows for extracting individual tree architecture measures. First, a sequence of RGB (color) and depth images were collected to separately reconstruct individual sides of a row. A single side reconstruction can be obtained very easily using established methods (Roy and Isler, 2016; Sturm et al., 2012); however, merging the reconstructions of two sides was difficult due to a lack of overlap between the two partial reconstructions.

Semantic mapping was used to merge the rows and extract the desired information. Semantic mapping in computer vision is a method of segmentation or assigning meta-information (e.g. ground, sky, fruit, tree trunk, or canopy) to their corresponding group of pixels in images. The human eye readily recognizes a group of pixels as a tree, but a computer reads images as data points and the pixels are just points until the computer is taught how to tag the meta-information to those pixels.

To merge the two sides, common semantic features of each side were identified (i.e. mapped) and used as constraints. The features included tree silhouettes, below canopy tree trunks, and the local ground plane surround each tree trunk. Using the merged 3D model from both sides, a tree segmentation algorithm was applied to compute canopy volume per tree. Tree height was estimated by fitting a bounding box on each segmented tree model after detecting the local ground plane surrounding each tree model. An image-derived trunk diameter metric was also estimated by first fitting an elliptic cylinder to the 3D points semantically assigned to each trunk, and the diameter was recorded as an average value at the height of approximately 200 to 250 mm above the ground plane.

8.3 Tree Architecture Data Curation and Analysis

Plots describing the data and distributions and correlations among all the tree architecture variables are presented in Appendix C12Appendix C. Mean rootstock cumulative yield, tree height, and TCA per in 2017 is in Table 8-1.

Data Curation

Seven individual trees had substantially deviant TCA values in 2017 compared to 2016. An examination of the trend line of diameter growth by year showed that this must be due

to measurement errors (trunk position or caliper related). Hence, diameter values for those trees were replaced with the average diameter of other trees on the same rootstock.

Not every rootstock was imaged in every block (Table 8-1). The imaging dataset lacked measurements on three rootstocks (B.7-3-150, B.70-20-20, and B.71-7-22), the two tallest rootstocks, and the shortest, respectively. In addition, 63 (247-184) trees were not imaged with sufficient quality to acquire morphology measures. As a result, the rootstock block interaction groups were reduced by 27 (124-97).

Table 8-1. Summary of rootstocks per block, and 2017 cumulative yield, tree height, and TCA.

Rootstock	Trees in Block (Trees Imaged)				Mean (sd)		
	1	2	3	4	Yield	Height	TCA
B-10	3 (3)	2 (2)	2 (2)	2 (2)	61.6 (18.0)	259.6 (19.2)	219.3 (20.4)
B-64-194	2 (0)	2 (0)	2 (0)	1 (1)	115.2 (16.0)	355.6 (56.3)	363.1 (62.8)
B-67-5-32	3 (0)	3 (1)	2 (0)	2 (1)	84.2 (16.4)	373.4 (47.2)	364.2 (47.6)
B-7-20-21	3 (3)	3 (3)	3 (0)	3 (3)	86.5 (22.5)	333.4 (64.8)	330.6 (51.9)
B-7-3-150	3 (0)	3 (0)	2 (0)	2 (0)	114.9 (15.1)	386.1 (27.6)	379.0 (24.6)
B-70-20-20	3 (0)	3 (0)	3 (0)	3 (0)	122.3 (33.0)	435.0 (34.3)	474.5 (25.4)
B-70-6-8	3 (0)	3 (3)	3 (0)	3 (0)	103.5 (23.4)	372.5 (25.0)	352.8 (42.1)
B-71-7-22	2 (0)	2 (0)	1 (0)	1 (0)	18.5 (2.6)	167.2 (14.8)	119.3 (10.1)
B-9	3 (3)	3 (3)	3 (3)	3 (3)	40.1 (12.1)	220.1 (14.7)	187.1 (22.4)
CG-2034	2 (2)	1 (1)	1 (1)	1 (1)	50.8 (15.1)	233.7 (36.6)	171.7 (17.6)
CG-3001	1 (1)	1 (1)	1 (1)	0 (0)	83.4 (4.4)	292.1 (12.7)	258.3 (12.4)
CG-4003	2 (2)	1 (1)	1 (1)	1 (1)	49.9 (6.8)	221.0 (11.4)	190.5 (21.4)
CG-4004	2 (0)	1 (1)	1 (0)	1 (1)	115.4 (28.0)	350.5 (52.8)	315.2 (12.5)
CG-4013	1 (1)	1 (1)	1 (1)	1 (1)	68.4 (20.1)	285.8 (26.4)	215.9 (35.7)
CG-4214	2 (2)	2 (2)	2 (2)	2 (2)	92.6 (12.6)	282.6 (21.2)	236.3 (21.3)
CG-4814	2 (2)	2 (2)	2 (2)	2 (2)	90.9 (14.8)	284.2 (25.3)	244.3 (23.1)
CG-5087	1 (1)	1 (1)	1 (0)	1 (1)	85.5 (19.9)	295.3 (30.0)	241.8 (32.0)
CG-5222	2 (2)	2 (2)	2 (2)	1 (1)	68.7 (22.3)	274.0 (25.2)	252.7 (21.8)
G-11	3 (3)	3 (3)	3 (3)	2 (2)	84.0 (22.0)	260.9 (32.3)	211.7 (20.5)
G-202-N	2 (1)	2 (2)	1 (0)	1 (1)	97.3 (26.3)	321.7 (36.5)	307.6 (66.6)
G-202-TC	2 (2)	2 (2)	1 (1)	1 (1)	80.2 (21.2)	277.3 (37.2)	241.8 (26.1)
G-41-N	3 (3)	3 (3)	3 (3)	2 (2)	83.9 (11.8)	273.6 (30.2)	234.6 (17.3)
G-41-TC	1 (1)	1 (1)	1 (1)	1 (1)	79.6 (18.4)	311.2 (24.3)	239.3 (19.1)
G-935-N	3 (3)	3 (3)	2 (2)	2 (2)	70.1 (12.4)	273.1 (22.6)	239.8 (22.8)
G-935-TC	1 (1)	1 (1)	1 (1)	0 (0)	71.1 (8.5)	245.5 (29.3)	227.0 (17.8)
M-26-EMLA	3 (3)	2 (2)	2 (1)	2 (2)	56.3 (16.3)	270.9 (20.1)	242.4 (37.7)
M-9-Pajam-2	3 (3)	3 (3)	3 (3)	3 (3)	62.9 (21.1)	259.3 (22.6)	223.6 (22.8)
M-9-T-337	3 (3)	3 (3)	3 (3)	3 (3)	66.6 (17.4)	260.4 (21.3)	218.4 (23.4)
PiAu-51-11	3 (3)	3 (3)	3 (2)	2 (2)	66.9 (18.6)	273.1 (18.2)	276.8 (28.5)
PiAu-9-90	2 (1)	2 (2)	2 (2)	1 (1)	31.0 (14.5)	259.4 (70.3)	236.2 (74.0)
Supp-3	2 (2)	2 (2)	1 (1)	1 (1)	46.1 (19.5)	251.9 (23.3)	204.8 (43.8)

Statistical Analyses

Several relationships of cumulative fruit yield to tree size variables, including image-derived variables, were investigated to determine the best relationship that explained the most variance in cumulative yield. The relationships were assessed in two ways: 1) considering all trees as independent from any experimental design, like a forest of individual trees, and 2) considering the randomized complete block design of the rootstock study. In the first method, the linear relationship for different tree size variables were compared using linear regression. For the second method, Autio et al. (2017) described the multi-location mixed model as

$$E = r + s + s{:}b$$

where yield efficiency (E, defined as the ratio of cumulative yield to TCA) was the main determinant of performance. Rootstock (r) and site (s) were fixed main effects, and block (b) was a random nested effect within site. In our situation, this model was simplified to only a single site in Minnesota, with a resulting mixed model as

$$Y = r + r{:}b$$

where Y represented either cumulative yield or yield efficiency, site was dropped, block was treated as a random effect, and a rootstock block interaction was added. In the analysis done by Autio et al. (2017) the rootstock replicates within a plot were averaged. For both types of analyses in this thesis, no averaging was done, and each tree within a plot was considered individually. The method to determine the best relationship to explain cumulative yield, tree size variables were added one at a time and each sequential model was compared in a stepwise fashion, where the first model included a tree size variable x_1 and the final model had n size variables.

$$Y = r + r{:}b + x_1 + x_2 + \cdots + x_n$$

Each additional model was attempted to improve the relationship by using a different combination of tree size variables ($x_1 + x_2 + \ldots + x_n$). Therefore a base (or reduced) model was used to compare to an expanded model containing additional variables (Holland et al., 2003). The mixed models were run in R using the 'lmer' function of the 'lme4' package (Bates et al., 2015). The 'lsmeans' function of the 'emmeans' package (Lenth, 2019) was used to compare each pair of sequential mixed models. In the comparisons, the estimated marginal means (emmeans, or least square means) of each rootstock were

calculated using maximum-likelihood equations in order to use a likelihood ratio test for comparison. This was done because methods for comparing one REML mixed model to another do not calculate mean squares of random effects.

9 CHAPTER 2. RESULTS

9.1 Correlation of Image-derived Metrics with Hand-Measured Metrics

Tree height

The image-derived tree metrics varied in their level of association with analogous hand-measured tree metrics. Coefficients of determination (R^2) varied between 0.46 and 0.93. The hand-measured to image-derived tree height relationship had the highest R^2 of 0.93, Figure 9-1.

TCA

Since small changes in trunk diameter are exponentially magnified in calculating TCA, it was not surprising that the relationship for image-derived to hand-measured TCA had a lower R^2 of 0.71 (Figure 9-1), than did the relationship for hand-measured to image-derived relationship for tree height.

Tree volume

Hand-measured tree volume was calculated as an ellipsoid cone according to the equation:

$$v = \left[\left(\frac{1}{3}\right) * \pi * a * w * h\right] + t$$

where t = the cylindrical volume of the below canopy trunk that is estimated with trunk diameter to the height of the lowest scaffold branch, a = the tree's widest point across rows, w = the canopy's widest point within its row, and h= the tree height. Figure 9-3 shows the relationship between image-derived and hand-measured tree volume (v). The R^2 of 0.46 indicated that hand-based volume only partially captured variation in image-derived tree volume measurements. Representations of tree volume calculated as

rectangular, pyramidal, and cylindrical shapes using hand measures had relationships to image-derived canopy volume that were similar to the ellipsoid cone representation (data not shown).

_

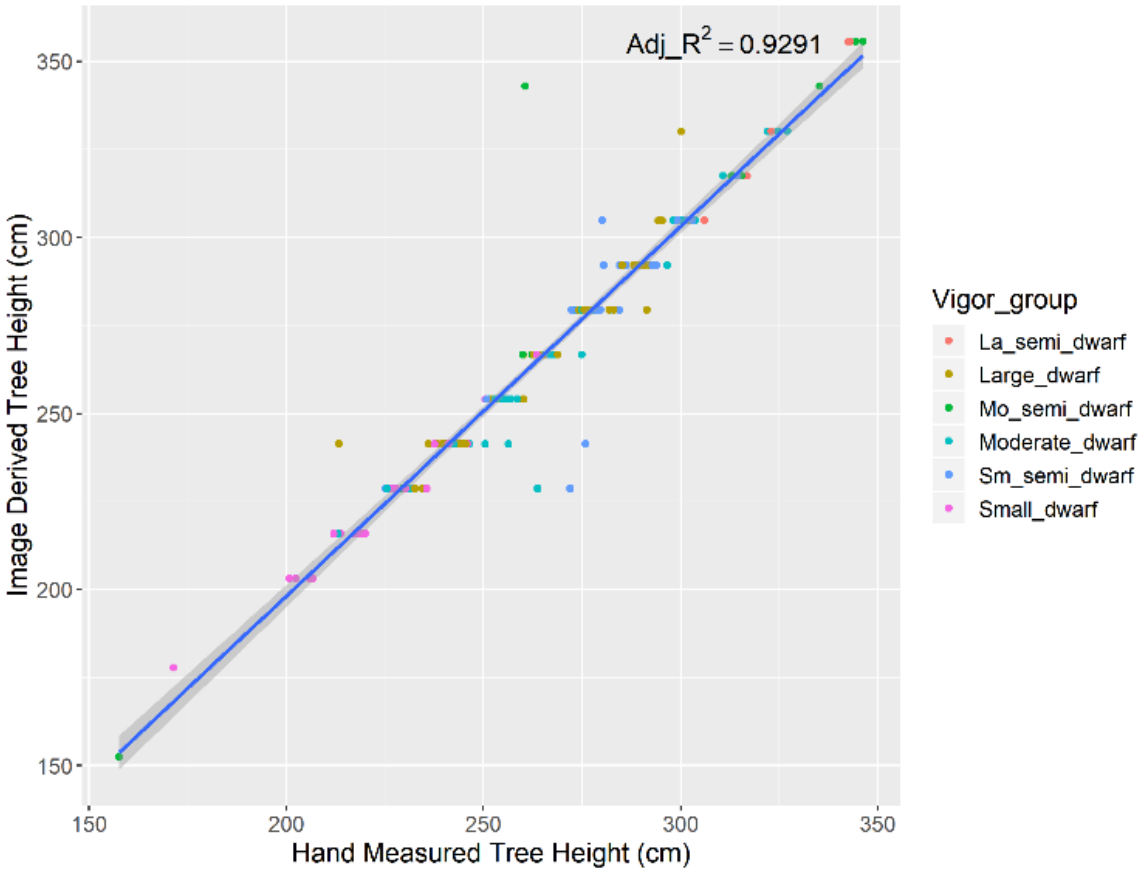


Figure 9-1. Image- vs hand- measured tree height;
Blue line is the calculated regression, while the gray shading is an estimated 95% confidence interval around the regression line.
Color is from vigor group assignations by Autio et al. (2017). Abbreviated vigor groups: La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

_

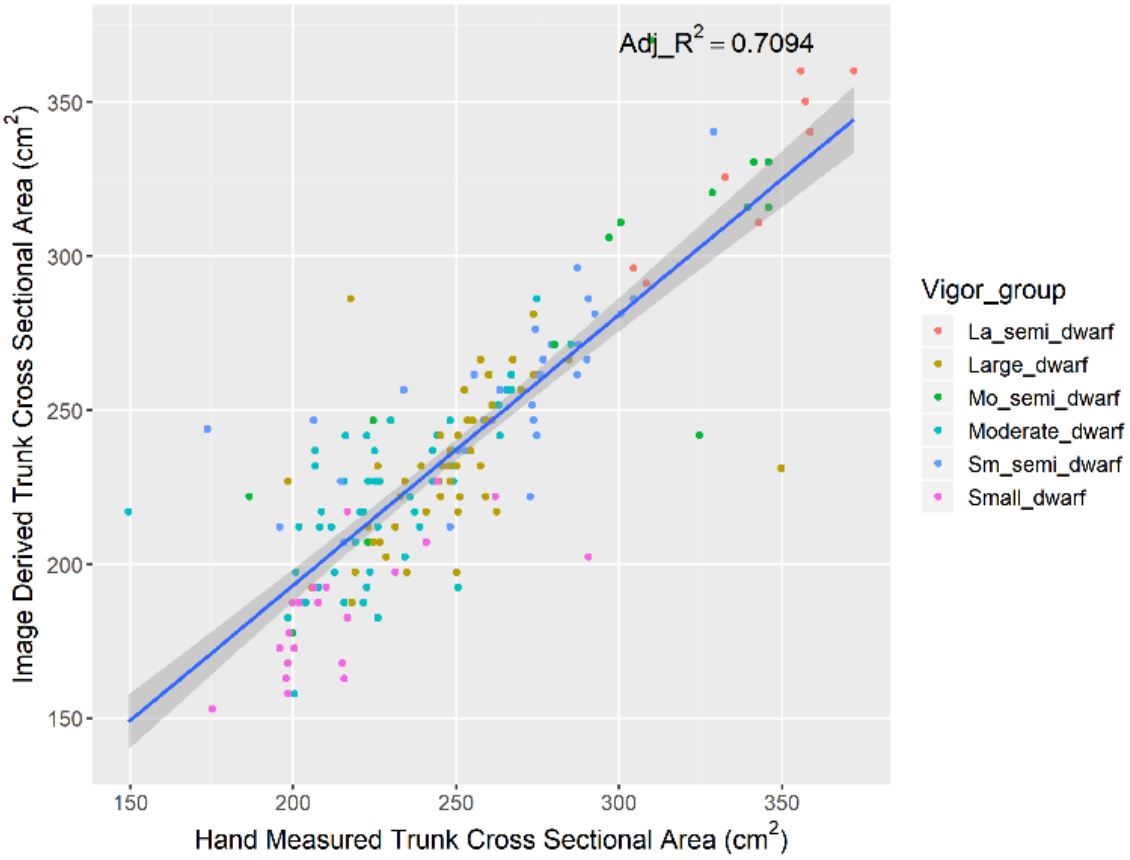


Figure 9-2. Image- vs hand- measured TCA

Blue line is the calculated regression, while the gray shading is an estimated 95% confidence interval around the regression line.

Color is from vigor group assignations by Autio et al. (2017). Abbreviated vigor groups: La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

_

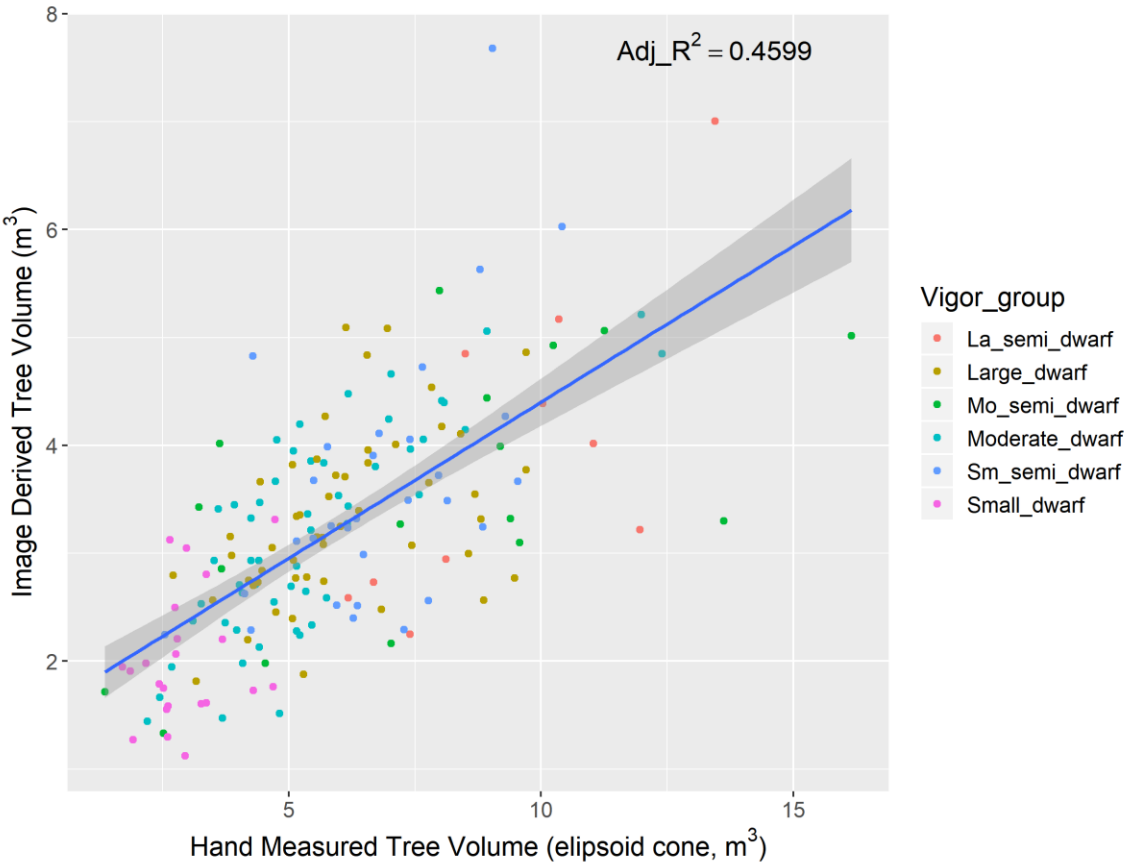


Figure 9-3. Imaged-derived tree volume by ellipsoid cone volume calculated from hand-measured parameters.

Blue line is the calculated regression, while the gray shading is an estimated 95% confidence interval around the regression line.

Color is from vigor group assignations by Autio et al. (2017). Abbreviated vigor groups: La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

_

9.2 Correlations between Tree Architecture Traits

Autio et al. (2017) described the Minnesota NC140 2010 'Honeycrisp' rootstock trial in its experimental context relative to other sites; here the Minnesota site is considered separately. In this first set of analyses, all trees were considered as individuals, independent from rootstock and experimental design factors. In this section, several variable relationships are examined. The purpose for the comparisons and the order that they are discussed is shown in Table 9-1.

Table 9-1. Relationships of tree architecture measures examined (without considering experimental design)

Dependent	Independent	Purpose
TCA	Imaged-derived tree volume	Possible correlation of TCA with a more exact method of tree volume calculation
Cumulative yield	TCA	Setting a benchmark relationship with a tree size variable with cumulative yield
Cumulative yield	Image-derived tree volume	Possible relationship with new, more exact method of tree volume calculation

The relationship between image-derived tree volume and TCA (Figure 9-4) was statistically significant with a p-value= $2.45*10^{-15}$, however the amount of variation explained was low at 29%.

The relationship of image-derived tree volume was greater with hand-measured tree volume ($R^2 = 0.4599$, $p=2.20*10^{-16}$) than the relationship of image-derived tree volume with hand-measured TCA ($R^2 = 0.2881$, $p=2.45*10^{-15}$). This may be because hand-measured volume increases the tree volume, as empty spaces between branches were not subtracted from total volume, and the allometric relationship of TCA with tree size would be of highest agreement for unpruned trees that would fill the canopy space more than pruned trees (Brym and Ernest, 2018).

Cumulative yield is an economically important trait and is the common factor for efficiency models. There are multiple issues biologically (e.g., biennial bearing), environmentally (e.g., severe weather events), and culturally (e.g., intensity of pruning or thinning) that can affect the yield of trees in an individual season. In order to control for these issues most tree-fruit studies report cumulative yield over several years instead of the yield of a single year. In this experiment, a significant hailstorm affecting the quality of nearly all fruit in the experiment occurred in June 2017. The correlations of the yield in 2017 to both image-derived tree volume ($R^2= 0.04$, $p= 0.0065$) and TCA ($R^2= 0.23$, $p= 7.287*10^{-11}$) were low (Figure C- 6 and Figure C- 7). Therefore, the cumulative yield per tree was used for analysis in place of the yield in the year of 2017. Figure 9-6 shows cumulative yield plotted against hand measured TCA. This relationship accounted for 31.42% of the variation in cumulative yield. The utility of image-derived measures can be assessed by comparison to this benchmark relationship of cumulative yield and TCA.

The relationship of image-derived tree volume with cumulative yield (R^2 = 0.1963, p = $1.808*10^{-10}$) was weaker than that of hand-measured TCA with cumulative yield (R^2 = 0.3142, p = $2.20*10^{-16}$).

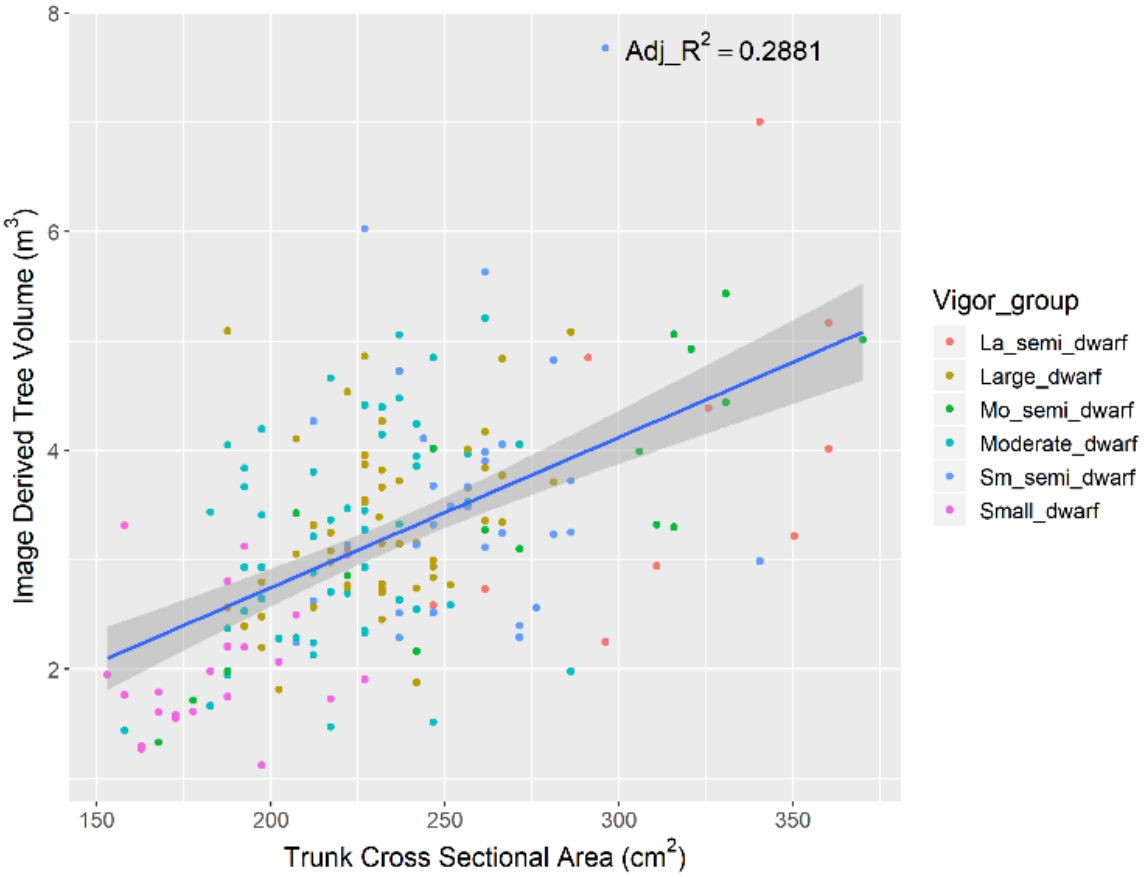


Figure 9-4.Imaged-derived tree volume by TCA.
Color is from vigor group assignations by Autio et al. (2017). Blue line is the calculated regression, while the gray shading is an estimated 95% confidence interval around the regression line. La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

_

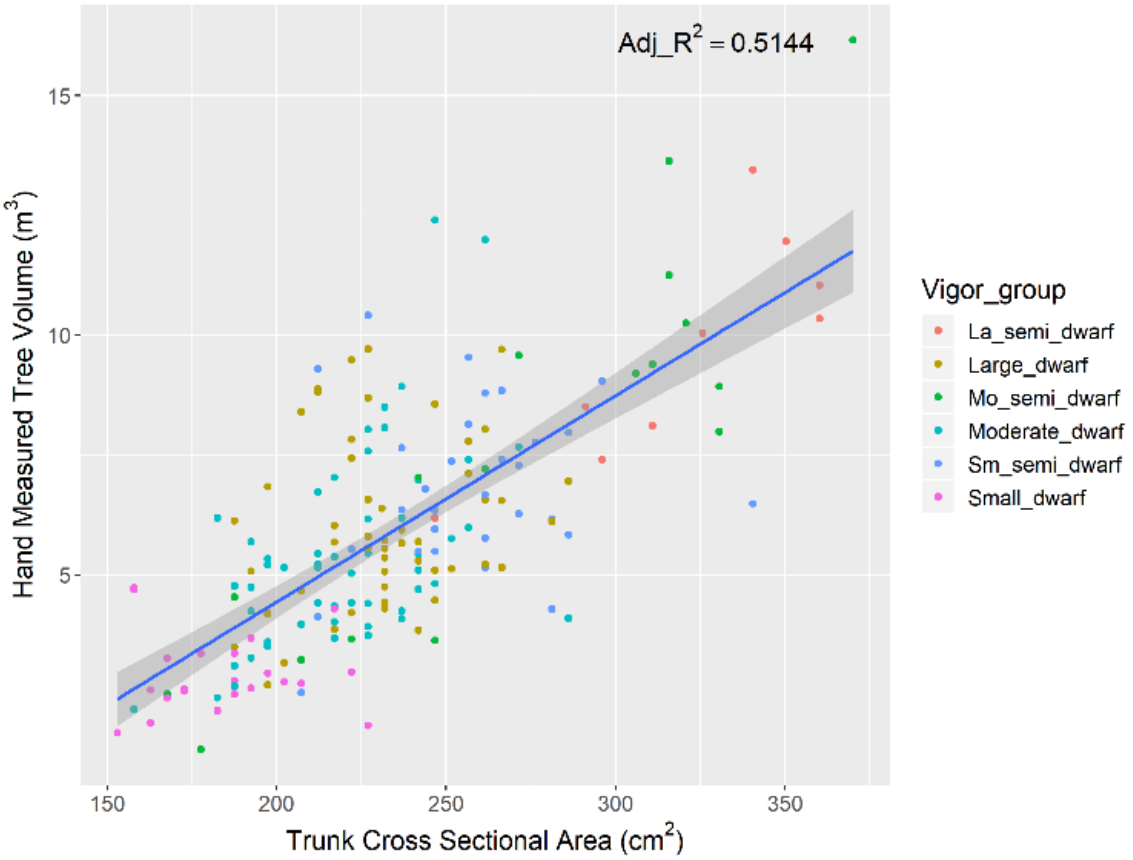


Figure 9-5. Hand-measured tree volume by TCA.

Color is from vigor group assignations by Autio et al. (2017). Blue line is the calculated regression, while the gray shading is an estimated 95% confidence interval around the regression line. La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

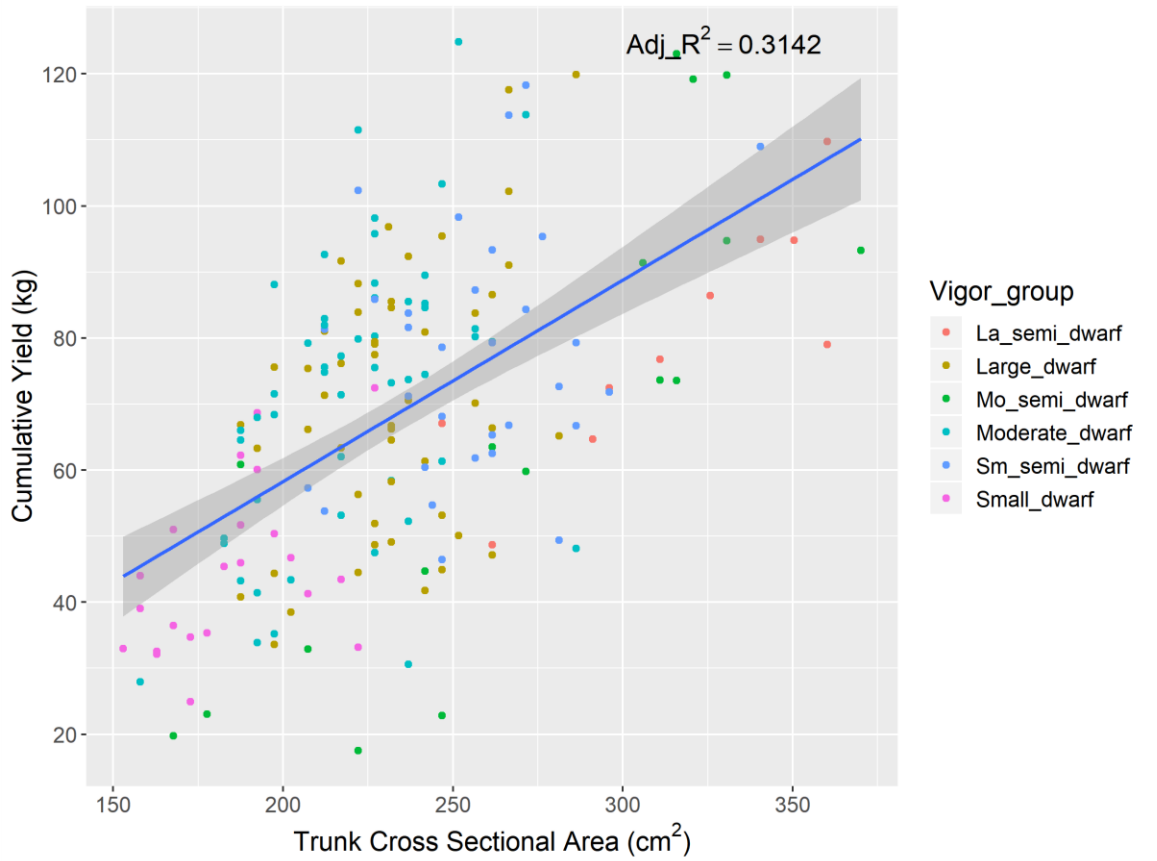


Figure 9-6. Six-year cumulative yield by TCA.

Color is from vigor group assignations by Autio et al. (2017). Blue line is the calculated regression, while the gray shading is an estimated 95% confidence interval around the regression line. La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

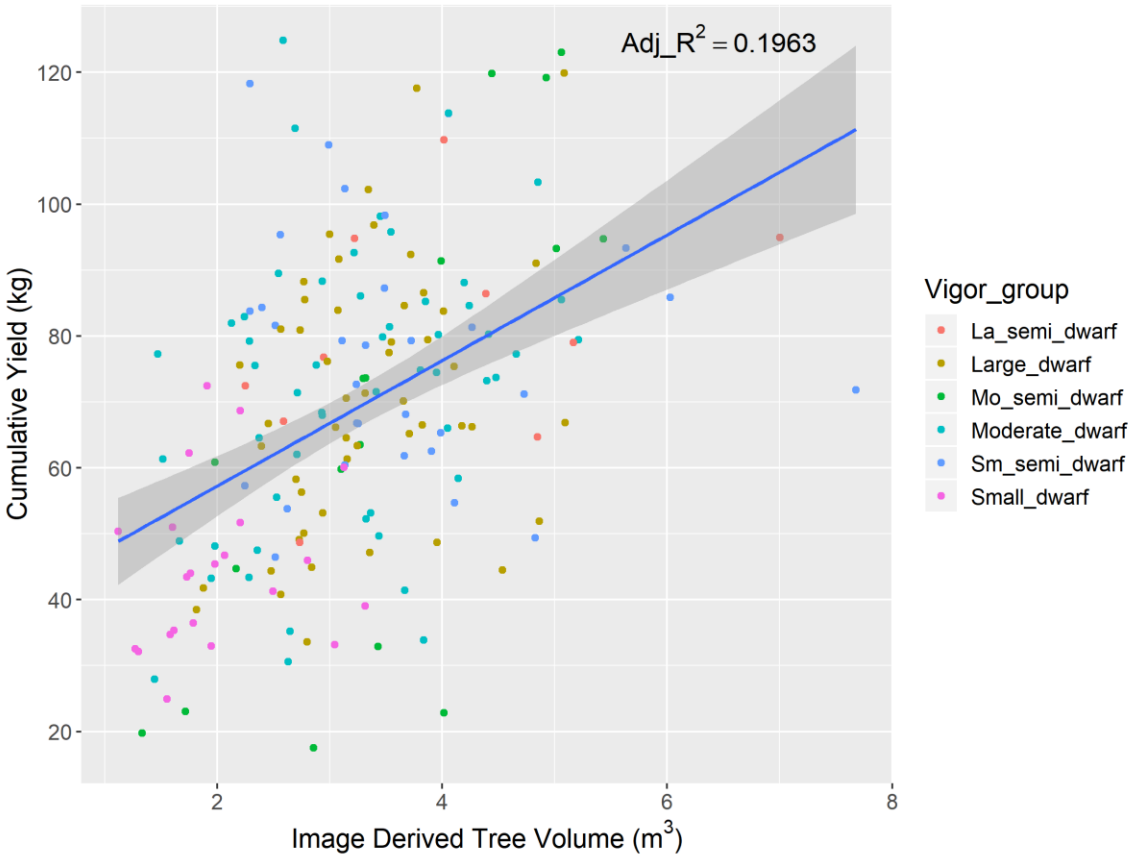


Figure 9-7. Six-year cumulative yield by image-derived tree volume.
Color is from vigor group assignations by Autio et al. (2017). Blue line is the calculated regression, while the gray shading is an estimated 95% confidence interval around the regression line. La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

_

9.3 Comparison of Models Estimating Cumulative Yield (Including Rootstock and Blocking Effects): Hand- and Image-Derived Traits

In this section, cumulative yield (6 years, 2012-2017) was estimated with mixed models, incorporating experimental design and various architectural features. Rootstock replicates within plots were not averaged to keep the sample size as large as possible, although averaging by plot was done in previous analyses by Autio et al. (2017). Mixed models were used to incorporate the experimental design and handle missing data due to the lack

of complete rootstock-block interactions and insufficient imaging of all rootstocks. In the base model,

$$Cumulative\ Yield \sim Rootstock\ +\ (1\ |\ Block)\ +\ (1\ |\ Rootstock{:}Block)$$

rootstock was set as a fixed effect, while block and a rootstock-block interaction were set as random effects.

The rootstock-block interaction was added because the experimental design had a large number of rootstock treatments (31) and low number of trees per rootstock per block (0-3). The rootstock effect could be different in different blocks due to the very large range of tree sizes. Some plots with large-tree plots from non-dwarfing rootstocks crowded adjacent plots, potentially affecting the tree vigor (especially of very small trees).

When the variance due to the rootstock block interaction was accounted for, the prediction of cumulative yield improved significantly compared to a model without the interaction (p=0.04729, table not shown).

A reduced model retaining the rootstock-block interaction but removing the random effect of block performed similarly to the model including both effects (type I ANOVA, p-value = 0.9302). Variance due to the block effects is likely accounted for some variance in the interaction term. When the block term was included in equivalent models along with the interaction, as discussed below, it led to singular fits (i.e. model parameters were too close to the tolerance level for having variances of linear combinations of the effects close to zero). Thus, the block effect was not used for remaining comparisons.

The purpose in building each new model, outlined below in Table 9-2, was to identify the variable(s) having the largest correlation with cumulative yield and to determine whether image-derived measures could be used in place of hand-measured variables. The best model would include the variable(s) that account for the largest amount of variation in cumulative yield. Table 9-2 also shows the log likelihood per model and the chi squared probability [Pr(>Chisq)] per model comparisons.

Table 9-2. Steps in mixed model analysis for determine the best model for predicting cumulative yield.

Model ID	Additional Variables	Log Likelihood	logLik df	Compared with	Purpose	Pr(>Chisq)
A (Basic)	Rootstock + (1 | Block :Rootstock)	-744.8	30	-	-	-
B	TCA	-718.4	31	A	Does TCA improve the prediction of cumulative yield?	3.82E-13
C	TCA + Height	-714.8	32	B	Will tree height improve upon model B?	0.01
C.1 to C.4	C + 1st Scaffold Height; C + In-Row Spread; C + Across-Row Spread; C + Avg Canopy Spread	-713.6 -713.3 -714.7 -713.7	33	C	Do any other canopy metrics help improve upon model C?	0.12 0.09 0.66 0.13
D	TCA + Tree Volume	-716.2	32	B	Will hand-measured tree volume improve upon model B?	0.03
E	Tree Volume	-735.4	31	A	Does hand-measured tree volume improve upon the basic model?	1.25E-05
F	TCA + Image-derived Tree Volume	-717.5	32	B	Will image derived tree volume improve upon the model B and is it different form model D?	0.19
G	Imaged Derived Tree Volume	-740.6	31	A	How does a model with image derived tree volume compare to models E and B?	3.51E-03
H	Image-derived Tree Volume and Image-derived TCA	-717.2	32	G	How do image derived metrics compare to hand-measured TCA and tree volume?	7.13E-12

In Table 9-3 the rootstock coefficients for the basic mixed model are shown with the standard error values. In this model, eight rootstocks with significant ($p<0.05$) coefficients were identified as having effects on cumulative yield (positive or negative) from the mean of all rootstocks (intercept = 62.86 kg).

Table 9-3. Basic mixed model cumulative yield ~ Rootstock + (1 | Block) + (1 | Rootstock:Block)

Rootstock	Coefficient	Std Error	df	t value	Pr(>|t|)	Sig
B-64-194	32.09	17.89	110	1.79	0.08	
B-67-5-32	3.85	13.48	102	0.29	0.78	
B-7-20-21	14.89	9.71	48	1.53	0.13	
B-70-6-8	13.3	13.98	45	0.95	0.35	
B-9	-22.75	9.03	49	-2.52	0.02	*
CG-2034	-13.67	10.24	78	-1.33	0.19	
CG-3001	20.54	11.65	96	1.76	0.08	
CG-4003	-10.93	11.06	77	-0.99	0.33	
CG-4004	58.26	13.48	102	4.32	3.6E-05	***
CG-4013	5.56	10.62	91	0.52	0.60	
CG-4214	29.74	9.45	59	3.15	2.6E-03	**
CG-4814	28.04	9.45	59	2.97	4.3E-03	**
CG-5087	17.6	11.65	96	1.51	0.13	
CG-5222	5.52	9.67	63	0.57	0.57	
G-11	19.08	9.22	53	2.07	4.3E-02	*
G-202-N	38.81	12.4	74	3.13	2.5E-03	**
G-202-TC	18.85	9.93	69	1.90	0.06	
G-41-N	21.06	9.12	51	2.31	2.5E-02	*
G-41-TC	16.73	10.62	91	1.58	0.12	
G-935-N	7.17	9.22	53	0.78	0.44	
G-935-TC	8.25	11.65	96	0.71	0.48	
M-26-EMLA	-12.28	9.53	59	-1.29	0.20	
M-9-Pajam-2	0.03	9.03	49	0.00	1.00	
M-9-T-337	3.78	9.03	49	0.42	0.68	
PiAu-51-11	7.05	9.22	53	0.76	0.45	
PiAu-9-90	-34.89	9.93	69	-3.51	7.9E-04	***
Supp-3	-14.88	9.93	69	-1.50	0.14	
(Intercept)	62.86	6.6	56	9.52	2.6E-13	***
AIC				1383.93		
BIC				1479.89		
Log Likelihood				-661.97		
Num. obs.				181		
Num. groups: Rootstock:Block				95		
Var: Rootstock:Block (Intercept)				89.75		
Var: Residual				186.77		

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

The basic mixed model was first compared to a model containing TCA as an additional fixed effect (likelihood ratio test results are in Table 9-4). The p-value was very low, signifying that these two models produced results significantly different from each other. The Akaike Information Criterion (AIC, an estimator of relative quality of a model for the dataset) value was lower by 59 units for the model with the addition of TCA, denoting an improvement in predicting cumulative yield.

The image-derived trunk diameter was used to calculate an image-derived TCA measure, however a model with it added upon the basic model did not have as high of a log likelihood (-711.0) as a hand-measured TCA model (-706.6).

Table 9-4. Likelihood ratio test between basic mixed model and one with the addition of TCA to estimating cumulative yield

Models	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
A	30	1547	1643	-744	1487			
B	31	1497	1596	-717	1435	52.73	1	3.82E-13

Formulas:
A: cum_yield ~ Rootstock + (1 | Rootstock:Block)
B: cum_yield ~ Rootstock + (1 | Rootstock:Block) + TCA_cm2

Adding hand-measured tree height to the model containing TCA significantly improved prediction of cumulative yield.

Table 9-5 Likelihood ratio test between model containing TCA to a model with the tree height variable added, for estimating cumulative yield

Models	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
B	31	1497	1596	-717	1435			
C	32	1491	1594	-714	1427	7.28	1	0.01

Formulas:
B: cum_yield ~ Rootstock + (1 | Rootstock:Block) + TCA_cm2
C: cum_yield ~ Rootstock + (1 | Rootstock:Block) + TCA_cm2 + Height_cm

Adding other architectural dimensions, (within-row canopy spread, across-row canopy spread, and height of first scaffold) did not significantly improve prediction compared to the model containing TCA and tree height ($p > 0.09$). When compared to the basic mixed

model, the in-row spread and across-row spread did not influence the predictive ability of the model, when in combination, but each did so when each was used alone (data not shown)

Hand-measured tree volume (calculated as an ellipsoid cone; with the below-canopy trunk volume included) did improve upon the basic model ($p=1.25*10^{-5}$). When included with TCA as a fixed effect, a small but significant improvement over the model containing TCA alone was achieved ($p = 0.03$; AIC difference of three units). Tree height was not included because it is a mathematical component of canopy volume.

Table 9-6. Likelihood ratio test between the model containing TCA to a model with hand-measured tree volume added, for estimating cumulative yield.

Models	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
B	31	1497	1596	-717	1435			
D	32	1494	1596	-715	1430	4.55	1	0.03

Formulas:
B: cum_yield ~ Rootstock + (1 | Rootstock:Block) + TCA_cm2
D: cum_yield ~ Rootstock + (1 | Rootstock:Block) + TCA_cm2 + TreeVolume_Elp_m3

Adding image-derived tree volume to the model containing TCA did not improve predictive ability ($p = 0.19$). However, the model including only image-derived tree volume improved the basic model ($p = 3.51*10^{-3}$, table not shown).

Table 9-7. Likelihood ratio test between model containing TCA to a model with image-derived tree volume added, for estimating cumulative yield.

Models	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
B	31	1497	1596	-717	1435			
F	32	1497	1599	-716	1433	1.73	1	0.19

Formulas:
B: cum_yield ~ Rootstock + (1 | Rootstock:Block) + TCA_cm2
F: cum_yield ~ Rootstock + (1 | Rootstock:Block) + TCA_cm2 + I_Canopy_volume_m3

Only hand-measured tree volume improved the cumulative yield prediction beyond the model with only hand-measured TCA. In addition, in models with TCA not included, hand-measured tree volume resulted in a higher log likelihood at -735.4 (DF= 31)

compared to -740.6 for image-derived volume. Thus, the hand-measured tree volume model may have approximated the data better than the image-derived tree volume model.

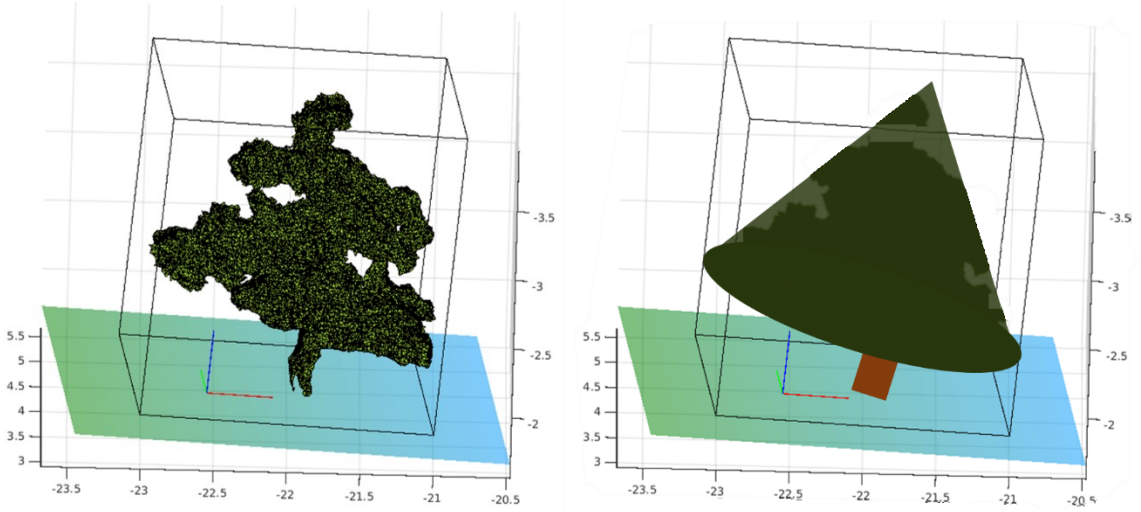
The utility of using image-derived canopy volume in place of TCA when added to a basic model was examined by comparing the log likelihood of the models. The log likelihood was higher, by approximately 3% for the model using only TCA compared to the model with image-derived tree volume. The predictive ability for cumulative yield was lower using image-derived volume than TCA, though imaged-derived volume still improved upon the basic model.

10 CHAPTER 2. DISCUSSION

Correlations between Tree Architecture Traits

In this experiment, hand-measured tree height exhibited strong correlations with image-derived tree height and TCA (R^2=0.93 and 0.71, respectively). The relationship between image-derived and manual measurement of tree volume was much lower (R^2 =0.46). The lower correlation for tree volume is expected given that apple trees do not normally grow to fill a cone as a pine or fir tree would, see Figure 10-1. To estimate a tree's true volume using manual measurements, an almost infinite number of measurements would be needed to "measure" the true volume (e.g., calculating the sum of the volume for each leaf, twig, and branch). Even with the current state of technology, image-derived tree volume should be the standard for comparing methods of manually estimating tree volume, though this is still just an estimate of volume in a statistical sense.

Figure 10-1. An image-derived and manually measured representation of a tree's volume.



Given that the image-derived metrics are closely related to manual methods of measuring tree architecture, other factors need to be considered such as the labor and time intensity of acquiring the measurements. The time needed for data collection is an important consideration given that resources are limited in research. Collecting tree dimensions manually in this 'Honeycrisp' rootstock experiment, required several hours up to several days annually depending on the number of variables being measured. Collecting the images of the experiment takes a maximum of 30 min. However processing the images with the extraction algorithms is still time and computationally intensive. If the processing of large amounts of image data can become rapid and automated, then image-derived measures such as TCA, tree volume, and even yield estimation (Roy et al., 2018; Wang et al., 2013) would be a time and labor saving alternative to manual methods.

Mixed models with experimental design

The best predictive model for cumulative yield was the basic model (rootstock and rootstock x block interaction effects) with hand-measured TCA and tree height. Image derived tree volume, though accounting for significant variation in a model by itself, did not improve upon models containing TCA. This leads to the conclusion that with this dataset, TCA is still useful to explain variation for cumulative yield, despite the many debates over its usefulness in the literature (Kikuchi and Shiozaki, 2007; Nesme et al.,

2005; Robinson and Lakso, 1991; Strong and Azarenko, 2000). This NC140 experiment used only measures of trees at 7 years after planting, so the utility of TCA in estimating relative tree size, as debated in the literature, might not be a problem until the trees became older. However, upon investigating the trend in variance explained by a per year 'cumulative yield'-TCA linear model the variation explained increased each year, Table 10-1. The adjusted R^2 improved each year to reach the value of 0.56 in year 2017 (for log transformed TCA). This confirms that the utility of TCA in predicting relative yield varies with tree age, as it was especially low for the early years (2012-2015) of this planting.

Tree height, TCA, and tree volume can be derived from RGB-D data using current algorithms but it remains to be shown that the extracted information will be as useful as information from hand-measured metrics. Perhaps estimations of tree volume can be improved by collecting additional years of data and / or using improved algorithms and sensors. Although it can be argued that the exact volume of a tree should be highly related to the yield production of that tree, assuming a constant relationship between fruit and tree volume, the model results using image-derived tree volume of this report do not support such argument.

Yield - TCA Relationship

TCA is an objective, easy-to-measure trait to use in calculating yield efficiency. A problem with tree volume is that it can be easily manipulated through the particular pruning schema; therefore, tree volume could change to a greater degree (negatively or positively) each year than TCA (should only be positive). The yearly standard deviation is likely higher for tree volume than TCA. On the other hand, to calculate yield efficiency particularly with imaged-derived tree volume, the best method would be an average value of (imaged-based tree volume-) yield efficiency over years. This would better control for yearly volume differences that are caused by various factors (e.g. one year's heavy pruning or biennial bearing), which may or may not also affect yield. Therefore, to test this an annual image-derived tree volume metric should be included in future long-term research measuring yield efficiency.

Table 10-1. Model statistics for the linear relationship between cumulative yield and TCA over years (not considering rootstock or blocking effects)

Version	Year	Adjusted R^2	Residual SE	Statistic	p-value	logLik	Deviance	DF Residual
log TCA	2012	0.00	4.3	1.9	0.1647	-707	4435	245
log TCA	2013	0.07	4.0	18.1	2.91E-05	-693	3964	245
log TCA	2014	0.01	8.0	3.5	0.0612	-865	15870	245
log TCA	2015	0.23	14.1	75.4	5.55E-16	-1004	48889	245
log TCA	2016	0.43	15.7	181.4	2.94E-31	-1022	60160	243
log TCA	2017	0.56	20.2	313.3	1.17E-45	-1087	99211	244
TCA	2012	0.00	4.3	0.0	0.9993	-708	4470	245
TCA	2013	0.03	4.1	8.3	0.0043	-698	4118	245
TCA	2014	0.00	8.1	0.2	0.6308	-866	16084	245
TCA	2015	0.16	14.8	48.8	2.67E-11	-1014	53313	245
TCA	2016	0.36	16.6	138.7	1.24E-25	-1035	66883	243
TCA	2017	0.53	21.0	272.1	1.42E-41	-1096	107127	244

Yield Efficiency

Cumulative yield is an important trait in itself, but it is also a commonly used numerator in yield efficiency ratios (Autio et al., 1996; Hampson et al., 2004). When a grower has an option between rootstocks, an important factor to consider is the productivity of the fruit tree in its allotted space as part of the tree training scheme. Therefore, to show rootstock differences in yield efficiency of this experiment, least squares pairwise tests were used to find distinct yield efficient groups among rootstocks types.

In the rootstock experimental investigation analysis (section 9.3), imaged-based tree volume had lower predictive ability than TCA when estimating cumulative yield. Imaged-based tree volume may also not be useful in describing efficiency ratios, at least with this dataset. However, for a demonstrative comparison to TCA, image-derived tree volume (as part of yield efficiency) is used for separating rootstocks (Autio et al., 2017). The comparisons must be understood only relatively since this report uses data up to the year 2017 and only from the Minnesota site, whereas Autio et al. (2017) had 10 sites and used data through 2014.

Therefore, the two ways yield efficiency was calculated are 1) with cumulative yield divided by TCA (units of $kg \bullet cm^{-2}$) and 2) cumulative yield divided by image-derived tree volume (units of $kg \bullet m^{-3}$). The estimated marginal mean (Lenth, 2019) yield efficiency values for each rootstock were calculated using mixed models to predict cumulative yield efficiency following the formula

$$\text{Cumulative Yield} \sim \text{Rootstock} + (1 \mid \text{Rootstock:Block}).$$

Table 11-1 shows the mean rootstock cumulative yield efficiencies using TCA, image-derived tree volume, and compares the rankings with the results of Autio et al. (2017).

11 CHAPTER 2. SUMMARY

The aim of this experiment was to explore the horticultural application of data acquired from red-green-blue-depth (RGB-D) sensors. In this experiment, tree architecture data was extracted from images of 247 'Honeycrisp' trees grafted on 31 rootstocks. The relationship between image-derived metrics and hand-measured was highest for tree height R^2=0.93, and TCA R^2=0.71. With improvements in image processing time, image-derived tree metrics are a feasible replacement to manual measures. The predictive ability of cumulative yield by image based-tree volume was lower than by manual-measured tree volume. Tree volume in general did not improve upon the mixed model containing TCA and tree height when estimating cumulative yield. Future research including improvements in image-merging and data extraction algorithms, experimental designs more adapted to imaging, and data over years is necessary to better examine the utility of image-derived tree morphology metrics.

Table 11-1. Pairwise comparison of rootstock yield efficiency least square means of trees in the 2010 NC-140 Honeycrisp Apple Rootstock Trial

		Image-derived volume (kg/m^3)		TCA (kg/cm^2)		TCA 2011-2014 (Autio et al. 2017)	
Vigor category	Rootstock	Least square mean ψ	Rank	Least square mean ψ	Rank	kg/cm^2	Rank
Semi-standard	B.70-20-20	nd	nd	nd	nd	0.7	31
Large semi-dwarf	B.7-20-21	0.0473 a	19	0.2480 abcde	23	1.3	27
	B.64-194	0.0738 ab	2	0.2788 abcdef	18	1.2	28
Moderate semi-dwarf	CG.4004	0.0412 ab	23	0.3806 bcdef	3	2	18
	G.202N	0.0500 ab	11	0.2960 abcdef	13	1.8	21
	B.70-6-8	0.0468 ab	20	0.2594 abcdef	21	1.4	25
	B.7-3-150	nd	nd	nd	nd	1.4	25
	B.67-5-32	0.0483 ab	15	0.2267 abcdef	25	1	29
	PiAu 9-90	0.1015 b	1	0.1310 a	28	0.9	30
Small semi-dwarf	CG.5087	0.0483 ab	16	0.3483 bcdef	6	2.2	11
	CG.4814	0.0393 a	24	0.3725 def	4	2.2	11
	CG.3001	0.0369 a	27	0.3231 bcdef	9	2.1	15
	CG.5222	0.0586 ab	6	0.2702 abcdef	19	1.8	21
	PiAu 51-11	0.0568 ab	8	0.2506 abcd	22	1.4	25
Large dwarf	G.935N	0.0546 ab	9	0.2957 bcdef	14	2.4	5
	CG.4214	0.0338 a	28	0.3913 f	1	2.4	5
	G.935TC	0.0481 ab	18	0.3131 bcdef	11	2.3	8
	G.202TC	0.0454 a	21	0.3329 bcdef	7	2.1	15
	M.9 Pajam 2	0.0483 a	17	0.2823 bcdef	17	2	18
	M.26 EMLA	0.0672 ab	3	0.2190 ab	26	1.7	22
	CG.4013	0.0580 ab	7	0.3138 bcdef	10	1.6	23
Moderate dwarf	G.11	0.0379 a	26	0.3822 ef	2	2.6	2
	M.9 NAKBT337	0.0493 a	14	0.3031 bcdef	12	2.4	5
	G.41N	0.0422 a	22	0.3581 cdef	5	2.4	5
	B.10	0.0495 a	13	0.2909 bcdef	15	2.2	11
	Supp.3	0.0587 ab	5	0.2307 abc	24	2.2	11
	G.41TC	0.0610 ab	4	0.3324 bcdef	8	2	18
Small dwarf	CG.4003	0.0388 a	25	0.2637 abcdef	20	2.6	2
	B.9	0.0508 a	10	0.2126 ab	27	2.3	8
	CG.2034	0.0497 a	12	0.2858 bcdef	16	2.2	11
Sub-dwarf	B.71-7-22	nd	nd	nd	nd	2	18

Rank of 1 is highest yield efficiency.

ψ Based only on Minnesota data from the mixed model formula: (Yield efficiency ~ Rootstock + (1 | Rootstock:Block).

All values are least-square means (arbitrary units) using R software package `emmeans` that were derived from the mixed model using function `lmer` from `lme4` package.

Lowercase letter represents the assigned pairwise group of least square means using `CLD` function of `emmeans` package with alpha=0.05.

12 THESIS CONCLUSION

The applied use of technology in phenotyping apple fruit and trees was explored in this thesis. The aim of Chapter 1 was to explore the utility and practical use of a handheld NIR spectrometer in estimating several fruit quality traits. The F-750 handheld spectrometer showed utility for predicting the values of SSC and SPI across a dataset of 15 cultivars with RPD values indicating the ability to discriminate between low and high groups. The models for other fruit traits examined, firmness, TA, skin traits, etc., either had low predictive abilities, or would not be useful beyond the ability to discriminate between cultivars with extreme trait values. Temperature and outdoor limitations of the F-750 spectrometer were determined to be minimal compared to the importance of collecting more than a single scan per fruit to control for individual fruit heterogeneity.

The aim of Chapter 2 was to explore the horticultural application of data acquired from red-green-blue-depth (RGB-D) sensors. The relationship between image-derived metrics and hand-measured was highest for tree height R^2=0.93, and TCA R^2=0.71. With improvements in image processing time, image-derived tree metrics are a feasible replacement to manual measures. The predictive ability of cumulative yield by image based-tree volume was lower than by manual-measured tree volume. Tree volume in general did not improve upon the mixed model containing TCA and tree height when estimating cumulative yield. Future research including improvements in image-merging and data extraction algorithms, experimental designs more adapted to imaging, and data over years is necessary to better examine the utility of image-derived tree architecture metrics.

REFERENCES

Ahn, H.S., Dayoub, F., Popovi´c, M.P., Macdonald, B.A., Siegwart, R., Sa, I., 2018. An Overview of Perception Methods for Horticultural Robots: From Pollination to Harvest. arXiv.

Andersen, J.R., Lübberstedt, T., 2003. Functional markers in plants. Trends Plant Sci. 8, 554–560. https://doi.org/10.1016/j.tplants.2003.09.010

Aprea, E., Charles, M., Endrizzi, I., Laura Corollaro, M., Betta, E., Biasioli, F., Gasperi, F., 2017. Sweet taste in apple: The role of sorbitol, individual sugars, organic acids and volatile compounds. Sci. Rep. 7, 1–10. https://doi.org/10.1038/srep44950

Araus, J.L., Cairns, J.E., 2014. Field high-throughput phenotyping: The new crop breeding frontier. Trends Plant Sci. 19, 52–61. https://doi.org/10.1016/j.tplants.2013.09.008

Atay, A.N., Atay, E., Lauri, P.-E., Kunter, B., Kantoglu, K.Y., 2018. Phenotyping gamma-ray-induced mutant population of ‘Amasya’ apple for architectural traits, precocity, floral phenology and fruit characteristics. Sci. Hortic. (Amsterdam). 233, 195–203. https://doi.org/10.1016/j.scienta.2018.01.003

Autio, W., Robinson, T., Black, B., Blatt, S., Cochran, D., Cowgill, W., Hampson, C., Hoover, E., Lang, G., Miller, D., Minas, I., Quezada, R.P., Stasiak, M., 2017. Budagovsky , Geneva , Pillnitz , and Malling Apple Rootstocks Affect ’ Honeycrisp ’ Performance Over the First Five Years ... Rootstocks Affect ’ Honeycrisp ’ Performance Over the First Five Years of the 2010 NC-140. J. Am. Pomol. Soc. 71, 149–166.

Autio, W.R., Greene, D.W., Lord, W.J., 1996. Performance of ‘ McIntosh ’ Apple Trees on Seven Rootstocks and a Comparison of Methods of Productivity Assessment. HortScience 31, 1160–1163.

Bac, C.W., Henten, E.J. van, Hemming, J., Edan, Y., 2014. Harvesting Robots for High-value Crops: State-of-the-art Review and Challenges Ahead. J. F. Robot. 31, 888–911. https://doi.org/10.1002/rob.21525

Baldwin, E.A., 2002. Fruit flavor, volatile metabolism and consumer perceptions, in: Fruit Quality and Its Biological Basis. CRC Press, pp. 89–106.

Ballabio, D., Todeschini, R., 2009. Multivariate Classification for Qualitative Analysis, in: Sun, D.-W. (Ed.), Infrared Spectroscopy for Food Quality Analysis and Control. Elsevier Inc., pp. 83–104. https://doi.org/10.1016/B978-0-12-374136-3.00004-3

Baranowski, P., Mazurek, W., Wozniak, J., Majewska, U., 2012. Detection of early bruises in apples using hyperspectral data and thermal imaging. J. Food Eng. 110, 345–355. https://doi.org/10.1016/j.jfoodeng.2011.12.038

Barden, J.A., Marini, R.P., 2001. Comparison of methods to express growth, size, and productivity of apple trees. J. Am. Pomol. Soc. 55, 251.

Barritt, Bruce H., 2000. Selecting an Orchard System for Apples. Compact Fruit Tree 33, 89–92.

Barritt, B. H., 2000. The HYTEC (Hybrid Tree Cone) orchard system for apples. Acta Hortic.

Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting Linear Mixed-Effects Models using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Beghi, R., Spinardi, A., Bodria, L., Mignani, I., Guidetti, R., 2013. Apples Nutraceutic Properties Evaluation Through a Visible and Near-Infrared Portable System. Food Bioprocess Technol. 6, 2547–2554. https://doi.org/10.1007/s11947-012-0824-7

Bertone, E., Venturello, A., Leardi, R., Geobaldo, F., 2012. Prediction of the optimum harvest time of ' Scarlet ' apples using DR-UV-Vis and NIR spectroscopy. Postharvest Biol. Technol. 69, 15–23. https://doi.org/10.1016/j.postharvbio.2012.02.009

Best, S., Salazar, F., Bastías, R., Leon, L., 2008. Crop Load Estimation Model to Optimize Yield – Quality Ratio in Apple Orchards , Malus Domestica Borkh , Var . Royal Gala. J. Inf. Technol. Agric. 11–18.

Birth, G.S., Norris, K.H., 1965. The Difference Meter for Measuring Interior Quality of Foods and Pigments in Biological Tissues, USDA ARS Technical Bulletin.

Blanpied, G.D., Silsby, K.J., 1992. Predicting Harvest Date Window for Apples. Cornell Coop. Ext. Publ.

Bobelyn, E., Serban, A.S., Nicu, M., Lammertyn, J., Nicolai, B.M., Saeys, W., 2010. Postharvest quality of apple predicted by NIR-spectroscopy: Study of the effect of biological variability on spectra and model performance. Postharvest Biol. Technol. 55, 133–143. https://doi.org/10.1016/j.postharvbio.2009.09.006

Brym, Z.T., 2016. An Allometric Approach to Evaluate Physiological and Production Efficiencies in Tree Size for Tart Cherry and Apple Orchard Systems. Utah State Univ. Digit.

Brym, Z.T., Ernest, S.K.M., 2018. Process-based allometry describes the influence of management on orchard tree aboveground architecture. PeerJ 6, 1–16. https://doi.org/10.7717/peerj.4949

Bureau, S., Ruiz, D., Reich, M., Gouble, B., Bertrand, D., Audergon, J.M., Renard, C.M.G.C.G.C., 2009. Rapid and non-destructive analysis of apricot fruit quality using FT-near-infrared spectroscopy. Food Chem. 113, 1323–1328. https://doi.org/10.1016/j.foodchem.2008.08.066

Bureau, S., Ścibisz, I., Le Bourvellec, C., Renard, C.M.G.. C.M.G.C., Scibisz, I., Le Bourvellec, C., Renard, C.M.G.. C.M.G.C., 2012. Effect of sample preparation on the measurement of sugars, organic acids, and polyphenols in apple fruit by mid-infrared spectroscopy. J. Agric. Food Chem. 60, 3551–3563. https://doi.org/10.1021/jf204785w

Byrne, D.H., 2012. Trends in Fruit breeding, in: Badenes, M.L., Byrne, D.H. (Eds.), Fruit Breeding. Springer, pp. 3–36. https://doi.org/10.1007/978-1-4419-0763-9

Callahan, A.M., 2003. Breeding for fruit quality. Acta Hortic. 622, 295–302. https://doi.org/10.17660/ActaHortic.2003.622.27

Cerda, A.A., García, L.Y., Ortega-Farías, S., Ubilla, Á.M., 2012. Consumer preferences and willingness to pay for organic apples. Cienc. e Investig. Agrar. 39, 47–59. https://doi.org/10.4067/S0718-16202012000100004

Chang, H.-Y.Y., Vickers, Z.M., Tong, C.B.S., 2018. The use of a combination of instrumental methods to assess change in sensory crispness during storage of a 'Honeycrisp'' apple breeding family.' J. Texture Stud. 49, 228–239. https://doi.org/10.1111/jtxs.12325

Charles, M., Aprea, E., Gasperi, F., 2019. Factors Influencing Sweet Taste in Apple, in: Bioactive Molecules in Food, Reference Series in Phytochemistry. Springer, Cham, pp. 1673–1694. https://doi.org/https://doi.org/10.1007/978-3-319-78030-6_80

Charles, M., Endrizzi, I., Aprea, E., Zambanini, J., Betta, E., Gasperi, F., 2017. Dynamic and static sensory methods to study the role of aroma on taste and texture: A multisensory approach to apple perception. Food Qual. Prefer. 62, 17–30. https://doi.org/10.1016/j.foodqual.2017.06.014

Cobb, J.N., DeClerck, G., Greenberg, A., Clark, R., McCouch, S., 2013. Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor. Appl. Genet. 126, 867–887. https://doi.org/10.1007/s00122-013-2066-0

Colaço, A.F., Escolà, A., 2018. Application of light detection and ranging and ultrasonic sensors to high-throughput phenotyping and precision horticulture : current status and challenges. Hortic. Res. https://doi.org/10.1038/s41438-018-0043-0

Confalonieri, R., Foi, M., Casa, R., Aquaro, S., Tona, E., Peterle, M., Boldini, A., De Carli, G., Ferrari, A., Finotto, G., Guarneri, T., Manzoni, V., Movedi, E., Nisoli, A., Paleari, L., Radici, I., Suardi, M., Veronesi, D., Bregaglio, S., Cappelli, G., Chiodini, M.E., Dominoni, P., Francone, C., Frasso, N., Stella, T., Acutis, M., 2013. Development of an app for estimating leaf area index using a smartphone. Trueness and precision determination and comparison with other indirect methods. Comput. Electron. Agric. 96, 67–74. https://doi.org/10.1016/j.compag.2013.04.019

Corelli-Grappadelli, L., 2000. The palmette training system. Acta Hortic. https://doi.org/10.17660/ActaHortic.1998.513.39

Corollaro, M.L., Aprea, E., Endrizzi, I., Betta, E., Demattè, M.L., Charles, M., Bergamaschi, M., Costa, F., Biasioli, F., Corelli Grappadelli, L., Gasperi, F., 2014. A combined sensory-instrumental tool for apple quality evaluation. Postharvest Biol. Technol. 96, 135–144. https://doi.org/10.1016/j.postharvbio.2014.05.016

Costa, F., Cappellin, L., Longhi, S., Guerra, W., Magnago, P., Porro, D., Soukoulis, C., Salvi, S., Velasco, R., Biasioli, F., Gasperi, F., 2011. Assessment of apple

(Malus×domestica Borkh.) fruit texture by a combined acoustic-mechanical profiling strategy. Postharvest Biol. Technol. 61, 21–28. https://doi.org/10.1016/j.postharvbio.2011.02.006

Costes, E., Gion, J.M., 2015. Genetics and Genomics of Tree Architecture, Advances in Botanical Research. Elsevier Ltd. https://doi.org/10.1016/bs.abr.2015.05.001

Costes, E., Guédon, Y., 2002. Modelling branching patterns on 1-year-old trunks of six apple cultivars. Ann. Bot. 89, 513–524. https://doi.org/10.1093/aob/mcf078

Das, A.J., Wahi, A., Kothari, I., Raskar, R., 2016. Ultra-portable, wireless smartphone spectrometer for rapid, non-destructive testing of fruit ripeness. Sci. Rep. 6, 32504. https://doi.org/10.1038/srep32504

Dejong, T.M., 2017. The understanding of carbohydrate budgets in fruit trees made easy: what we know and ideas about what we need to know. Acta Hortic. 1177, 29–40. https://doi.org/10.17660/ActaHortic.2017.1177.3

Demattè, M.L., Pojer, N., Endrizzi, I., Corollaro, M.L., Betta, E., Aprea, E., Charles, M., Biasioli, F., Zampini, M., Gasperi, F., 2014. Effects of the sound of the bite on apple perceived crispness and hardness. Food Qual. Prefer. 38, 58–64. https://doi.org/10.1016/j.foodqual.2014.05.009

Dever, M.C., Cliff, M.A., Hall, J.W., 1995. Analysis of variation and multivariate relationships among analytical and sensory characteristics in whole apple evaluation. J. Sci. Food Agric. 69, 329–338. https://doi.org/10.1002/jsfa.2740690309

Di Guardo, M., Bink, M.C.A.M., Guerra, W., Letschka, T., Lozano, L., Busatto, N., Poles, L., Tadiello, A., Bianco, L., Visser, R.G.F., van de Weg, E., Costa, F., 2017. Deciphering the genetic control of fruit texture in apple by multiple family-based analysis and genome-wide association. J. Exp. Bot. 68, 1451–1466. https://doi.org/10.1093/jxb/erx017

Diaz-Garcia, L., Covarrubias-Pazaran, G., Schlautman, B., Zalapa, J., 2016. GiNA, an Efficient and high-throughput software for horticultural phenotyping. PLoS One 11, 1–12. https://doi.org/10.1371/journal.pone.0160439

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 Biol. Technol. 110, 86–95. https://doi.org/10.1016/j.postharvbio.2015.07.012

Dong, W., Isler, V., 2018. Tree Morphology for Phenotyping from Semantics-Based Mapping in Orchard Environments. arXiv.

Dorigoni, A., Lezzer, P., Dallabetta, N., Serra, S., Musacchi, S., 2011. Bi-axis: An alternative to slender spindle for apple Orchards. Acta Hortic. 903, 581–588. https://doi.org/10.17660/ActaHortic.2011.903.80

Duque, P., Arrabaça, J.D., 1999. Respiratory metabolism during cold storage of apple fruit. I. Sucrose metabolism and glycolysis. Physiol. Plant. 107, 14–23. https://doi.org/10.1034/j.1399-3054.1999.100104.x

Dussi, M.C., 2011. Sustainable use of plant bioregulators in pear production. Acta Hortic. 909, 353–368.

E.Watada, A., Watada, A.E., E.Watada, A., Watada, A.E., 1989. NonDestructive Methods of Evaluating Quality of Fresh Fruits and Vegetables. Acta Hortic. 258, 321–329.

Else, M.A., 2017. Developing a vision system to enhance phenotyping in apples (Pomevision) [WWW Document]. Res. Counc. UK Gatew. to Res. Grant Arch. URL http://gtr.rcuk.ac.uk/projects?ref=BB%2FL017466%2F1 (accessed 2.20.18).

Endrizzi, I., Torri, L., Corollaro, M.L., Demattè, M.L., Aprea, E., Charles, M., Biasioli, F., Gasperi, F., 2015. A conjoint study on apple acceptability: Sensory characteristics and nutritional information. Food Qual. Prefer. 40, 39–48. https://doi.org/10.1016/j.foodqual.2014.08.007

Esbensen, K.H., Geladi, P., Larsen, A., 2013. The Replication Myth 1. NIR news 24, 17–20. https://doi.org/10.1255/nirn.1349

Evans, K., Brutcher, L., Konishi, B., Barritt, B., 2010. Correlation of Sensory Analysis with Physical Textural Data from a Computerized Penetrometer in the Washington State University Apple Breeding Program. Horttechnology 20(6), 1026–1029.

Evans, K., Guan, Y., Luby, J., Clark, M., Schmitz, C., Brown, S., Orcheski, B., Peace, C., Van De Weg, E., Iezzoni, A., 2012. Large-scale standardized phenotyping of apple in RosBREED. Acta Hortic. 945, 233–238.

Faber, N.M., Meinders, M.J., Geladi, P., Sjöström, M., Buydens, L.M.C., Kateman, G., 1995. Random error bias in principal component analysis. Part I. derivation of theoretical predictions. Anal. Chim. Acta 304, 257–271. https://doi.org/10.1016/0003-2670(94)00585-A

Felix Instruments, 2015. F-750 Produce Quality Meter Operations Manual.

Ferree, D.C., 1976. Multi-State Cooperative Apple lnterstem Planting lnterstem Established in 1976. Fruit Var. J. 36, 2–7.

Ferree, D.C., Perry, R.L., 1988. Rootstock Evaluation Through the NC-140 Project. HortScience 23, 102–104.

Fiorani, F., Schurr, U., 2013. Future Scenarios for Plant Phenotyping. Annu. Rev. Plant Biol. 64, 267–291. https://doi.org/10.1146/annurev-arplant-050312-120137

Furbank, R.T., Tester, M., 2011. Phenomics - technologies to relieve the phenotyping bottleneck. Trends Plant Sci. 16, 635–644. https://doi.org/10.1016/j.tplants.2011.09.005

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

Gong, A., Yu, J., He, Y., Qiu, Z., 2013. Citrus yield estimation based on images

processed by an Android mobile phone. Biosyst. Eng. 115, 162–170. https://doi.org/10.1016/j.biosystemseng.2013.03.009

Gottardo, P., Penasa, M., Lopez-Villalobos, N., De Marchi, M., 2016. Variable selection procedures before partial least squares regression enhance the accuracy of milk fatty acid composition predicted by mid-infrared spectroscopy. J. Dairy Sci. 99, 7782–7790. https://doi.org/10.3168/jds.2016-10849

Greensill, C. V., Walsh, K.B., 2000. Optimization of instrumentation precision and wavelength resolution for the performance of NIR calibrations of sucrose in a water-cellulose matrix. Appl. Spectrosc. 54, 426–430. https://doi.org/10.1366/0003702001949528

Gundersen, H.J.G., 2002. The smooth fractionator. J. Microsc. 207, 191–210. https://doi.org/10.1046/j.1365-2818.2002.01054.x

Halstead, M., McCool, C., Denman, S., Perez, T., Fookes, C., 2018. Fruit Quantity and Quality Estimation using a Robotic Vision System. IEEE Robot. Autom. Lett. 3, 2995–3002. https://doi.org/10.1109/LRA.2018.2849514

Hampson, C.R., Quamme, H.A., Kappel, F., Brownlee, R.T., 2004. Varying Density with Constant Rectangularity : II . Effects on Apple Tree Yield , Fruit Size , and Fruit Color Development in Three Training Systems over Ten Years. HortScience 39, 507–511.

Hampson, C.R.C.C.R., Quamme, H.A.H.H.A., Hall, J.W.J.W., MacDonald, R.A.R.A., King, M.C.M.M.C., Cliff, M.M.A.A., 2000. Sensory evaluation as a selection tool in apple breeding. Euphytica 111, 79–90. https://doi.org/10.1023/A:1003769304778

Hanke, M. V, Flachowsky, H., Peil, A., Hättasch, C., 2007. No Flower no Fruit – Genetic Potentials to Trigger Flowering in Fruit Trees. Genes, Genomes and Genomics 1, 1–20.

Hansen, W.G., Wiedemann, S.C.C., Snieder, M., Wortel, V.A.L., 2000. Tolerance of near infrared calibrations to temperature variations; a practical evaluation. J. Near Infrared Spectrosc. 8, 125–132. https://doi.org/10.1255/jnirs.272

Harker, F.R., Amos, R.L., White, A., Petley, M.B., Wohlers, M., 2008a. Flavor differences in heterogeneous foods can be detected using repeated measures of consumer preferences. J. Sens. Stud. 23, 52–64. https://doi.org/10.1111/j.1745-459X.2007.00141.x

Harker, F. Roger, Gunson, F.A., Brookfield, P.L., White, A., 2002. An apple a day: the influence of memory on consumer judgment of quality. Food Qual. Prefer. 13, 173–179. https://doi.org/https://doi.org/10.1016/S0950-3293(02)00013-7

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 Biol. Technol. 28, 333–347. https://doi.org/10.1016/S0925-5214(02)00215-6

Harker, F.R., Kupferman, E.M., Marin, A.B., Gunson, F.A., Triggs, C.M., 2008b. Eating

quality standards for apples based on consumer preferences. Postharvest Biol. Technol. 50, 70–78. https://doi.org/10.1016/j.postharvbio.2008.03.020

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 Biol. Technol. 24, 241–250. https://doi.org/10.1016/S0925-5214(01)00157-0

Harker, F.R., Norquay, C., Amos, R., Jackman, R., Gunson, A., Williams, M., 2005. The use and misuse of discrimination tests for assessing the sensory properties of fruit and vegetables. Postharvest Biol. Technol. 38, 195–201. https://doi.org/10.1016/j.postharvbio.2005.06.007

Harker, F.R., Redwell, R.J., Hallett, I.C., Murray, S.H., Carter, G., 1997. Textures of Fresh Fruit, in: Horticultural Reviews. John Wiley & Sons, pp. 121–224.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, 2nd ed, Springer Series in Statistics. Springer. https://doi.org/10.1007/b94608

Helland, I.S., 2001. Some theoretical aspects of partial least squares regression. Chemom. Intell. Lab. Syst. 58, 97–107. https://doi.org/https://doi.org/10.1016/S0169-7439(01)00154-X

Herold, B., Truppel, I., Zude, M., Geyer, M., 2005. Spectral measurements on “Elstar” apples during fruit development on the tree. Biosyst. Eng. 91, 173–182. https://doi.org/10.1016/j.biosystemseng.2005.03.005

Hirst, P., Ferree, D., 1995. Rootstock effects on shoot morphology and spur quality of Delicious’ apple and relationships with precocity and productivity. J. Am. Soc. Hortic. Sci. 120, 622–634.

Holland, J.B., Nyquist, W.E., Cervantes-Martínez, C.T., 2003. Estimating and Interpreting Heritability for Plant Breeding: An Update, in: Janick, J. (Ed.), Plant Breeding Reviews. John Wiley & Sons, pp. 9–112. https://doi.org/10.1002/9780470650202.ch2

Hoying, S.A., Robinson, T.L., 2000. The apple orchard planting systems puzzle. Acta Hortic. 513, 257–260. https://doi.org/10.1017/CBO9781107415324.004

Huang, Z., Hu, H., Shen, F., Wu, B., Wang, X., Zhang, B., Wang, W., Liu, L., Liu, J., Chen, C., Zhang, R., Chen, R., Wang, Y., Wu, T., Xu, X., Han, Z., Zhang, X., 2018. Relatively high acidity is an important breeding objective for fresh juice-specific apple cultivars. Sci. Hortic. (Amsterdam). 233, 29–37. https://doi.org/10.1016/j.scienta.2018.01.026

Iezzoni, A., Peace, C., Main, D., Bassil, N., Coe, M., Finn, C., Gasic, K., Luby, J., Hokanson, S., McFerson, J., Norelli, J., Olmstead, M., Whitaker, V., Yue, C., 2017. RosBREED2: Progress and future plans to enable DNAinformed breeding in the Rosaceae. Acta Hortic. 1172, 115–118. https://doi.org/10.17660/ActaHortic.2017.1172.20

Iglesias, I., Echeverría, G., Soria, Y., 2008. Differences in fruit colour development, anthocyanin content, fruit quality and consumer acceptability of eight “Gala” apple strains. Sci. Hortic. (Amsterdam). 119, 32–40. https://doi.org/10.1016/j.scienta.2008.07.004

Jones, H.G., 2017. Future opportunities for crop physiology in fruit production. Acta Hortic. 1177, 59–72. https://doi.org/10.17660/ActaHortic.2017.1177.6

Kaur, H., Künnemeyer, R., McGlone, V.A., 2017. Comparison of hand-held near infrared spectrophotometers for fruit dry matter assessment. J. Near Infrared Spectrosc. 25, 267–277. https://doi.org/10.1177/0967033517725530

Khatiwada, B.P., Subedi, P., Hayes, C., Walsh, K., 2015. Sorting for internal flesh browning in apple using visible-shortwave near infrared spectroscopy. Proc. ICNIRS 1, 60–63. https://doi.org/10.17648/nir-2015-34209

Kikuchi, T., Shiozaki, Y., 2007. Apple Canopies as Populations of Branches : a New Concept for Measuring Tree Vigor. Acta Hortic. 732, 675–680.

Kuckenberg, J., Tartachnyk, I., Noga, G., 2008. Evaluation of fluorescence and remission techniques for monitoring changes in peel chlorophyll and internal fruit characteristics in sunlit and shaded sides of apple fruit during shelf-life. Postharvest Biol. Technol. 48, 231–241. https://doi.org/10.1016/j.postharvbio.2007.10.013

Kumar, S., McGlone, V.A., Whitworth, C., Volz, R., 2015. Postharvest performance of apple phenotypes predicted by near-infrared (NIR) spectral analysis. Postharvest Biol. Technol. 100, 16–22. https://doi.org/10.1016/j.postharvbio.2014.09.021

Kupferman, E., 2005. We are in The Food Business: Providing Consumers with The Quality Apples they want, in: Washington Tree Fruit Postharvest Conference. Wenatchee, WA, pp. 1–7.

Kupferman, E., 2002. Critical Aspects of Harvest and Quality Management. Postharvest Inf. Network. 1–5.

Lakso, A.N., 2017. Perspectives on physiology in support of fruit production and future opportunities: lessons from 40 years of fruit research. Acta Hortic. 1177, 1–12. https://doi.org/10.17660/ActaHortic.2017.1177.1

Lammertyn, J., Ooms, K., Smedt, V. De, Baerdemaeker, J. De, 1998. Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR spectroscopy. Trans. ASAE 41, 1089–1094.

Lauri, P.-É.P., Trottier, C., Maguylo, K., Itier, C.T.R.O., Trottier, C., 2006. Architecture and Size Relations: An Essay on the Apple (Malus x domestica , Rosaceae) Tree. Am. J. Bot. 93, 357–368.

Lauri, P.É., Costes, E., Regnard, J.L., Brun, L., Simon, S., Monney, P., Sinoquet, H., 2009. Does knowledge on fruit tree architecture and its implications for orchard management improve horticultural sustainability? An overview of recent advances in the apple. Acta Hortic. 817, 243–250.

https://doi.org/10.17660/ActaHortic.2009.817.25

Lauri, P.E., Lespinasse, J.M., 2000. The Vertical Axis and Solaxe Systems in France. Acta Hortic. 513, 287–297.

Lauri, P.E., Térouanne, E., Lespinasse, J.M., Regnard, J.-L.L., Kelner, J.-J.J., Terouanne, E., Lespinasse, J.M., Regnard, J.-L.L., Kelner, J.-J.J., 1995. Genotypic differences in the axillary bud growth and fruiting pattern of apple fruiting branches over several years-an approach to regulation of fruit bearing. Sci. Hortic. (Amsterdam). 64, 265–281. https://doi.org/10.1016/0304-4238(95)00836-5

Lea, A.G.H.G.H., 1999. Apple juice, in: Ashurst, P.R. (Ed.), Production and Packaging of Non-Carbonated Fruit Juices and Fruit Beverages. Springer Science+Business Media, New York, pp. 153–196. https://doi.org/https://doi.org/10.1007/978-1-4757-6296-9_6

Lee, W.S.S., Alchanatis, V., Yang, C., Hirafuji, M., Moshou, D., Li, C., Alchanatisbvl, V., Yang, C., Hirafujid, M., Moshou, D., Li, C., 2010. Sensing technologies for precision specialty crop production. Comput. Electron. Agric. 74, 2–33. https://doi.org/10.1016/j.compag.2010.08.005

Lenth, R., 2019. emmeans: Estimated Marginal Means, aka Least-Squares Means. cran.r-project.org.

Li, B., Lecourt, J., Bishop, G., 2018. Advances in Non-Destructive Early Assessment of Fruit Ripeness towards Defining Optimal Time of Harvest and Yield Prediction—A Review. Plants 7, 1–20. https://doi.org/10.3390/plants7010003

Li, M., Qian, Z., Shi, B., Medlicott, J., East, A., 2018. Evaluating the performance of a consumer scale SCiO^{TM} molecular sensor to predict quality of horticultural products. Postharvest Biol. Technol. 145, 183–192. https://doi.org/10.1016/j.postharvbio.2018.07.009

Lin, H., Ying, Y., 2009. Theory and application of near infrared spectroscopy in assessment of fruit quality: A review. Sens. Instrum. Food Qual. Saf. 3, 130–141. https://doi.org/10.1007/s11694-009-9079-z

Liu, Y., Ying, Y., 2005. Use of FT-NIR spectrometry in non-invasive measurements of internal quality of “Fuji” apples. Postharvest Biol. Technol. 37, 65–71. https://doi.org/10.1016/j.postharvbio.2005.02.013

Llorens, J., Gil, E., Llop, J., Escolà, A., 2010. Variable rate dosing in precision viticulture: Use of electronic devices to improve application efficiency. Crop Prot. 29, 239–248. https://doi.org/10.1016/j.cropro.2009.12.022

Lordan, J., Francescatto, P., Dominguez, L.I., Robinson, T.L., 2018. Long-term effects of tree density and tree shape on apple orchard performance , a 20 year study — Part 1 , agronomic analysis. Sci. Hortic. (Amsterdam). 238, 303–317. https://doi.org/10.1016/j.scienta.2018.04.033

Lovász, T., Merész, P., Salgó, A., 1994. Application of near infrared transmission

spectroscopy for the determination of some quality parameters of apples. J. Near Infrared Spectrosc. 2, 213. https://doi.org/10.1255/jnirs.47

Lu, R., Guyer, D.E., Beaudry, R.M., 2000. Determination of Firmness and Sugar Content of Apples Using Near-Infrared Diffuse Reflectance. J. Texture Stud. 31, 615–630. https://doi.org/10.1111/j.1745-4603.2000.tb01024.x

Luby, J.J., Bedford, D.S., 2015. Cultivars as consumer brands: Trends in protecting and commercializing apple cultivars via intellectual property rights. Crop Sci. 55, 2504–2510. https://doi.org/10.2135/cropsci2014.10.0684

Ma, B., Chen, J., Zheng, H., Fang, T., Ogutu, C., Li, S., Han, Y., Wu, B., 2015. Comparative assessment of sugar and malic acid composition in cultivated and wild apples. Food Chem. 172, 86–91. https://doi.org/10.1016/j.foodchem.2014.09.032

Magwaza, L.S., Opara, U.L., 2015. Analytical methods for determination of sugars and sweetness of horticultural products-A review. Sci. Hortic. (Amsterdam). 184, 179–192. https://doi.org/10.1016/j.scienta.2015.01.001

Manalo, A., 1990. Assessing the importance of apple attributes: an agricultural application of conjoint analysis. Northeast. J. Agric. Resour. Econ. 19, 118–124.

Mann, H., Bedford, D., Luby, J., Vickers, Z., Tong, C., 2005. Relationship of instrumental and sensory texture measurements of fresh and stored apples to cell number and size. HortScience 40, 1815–1820.

Martínez Vega, M. V., Sharifzadeh, S., Wulfsohn, D., Skov, T., Clemmensen, L.H., Toldam-Andersen, T.B., 2013. A sampling approach for predicting the eating quality of apples using visible-near infrared spectroscopy. J. Sci. Food Agric. 93, 3710–3719. https://doi.org/10.1002/jsfa.6207

McGlone, V.A., Jordan, R.B., Martinsen, P.J., 2002a. Vis/NIR estimation at harvest of pre- and post-storage quality indices for “Royal Gala” apple. Postharvest Biol. Technol. 25, 135–144. https://doi.org/10.1016/S0925-5214(01)00180-6

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 Biol. Technol. 28, 431–435. https://doi.org/10.1016/S0925-5214(02)00207-7

McGlone, V.A., Jordan, R.B., Seelye, R., Martinsen, P.J., 2002b. Comparing density and NIR methods for measurement of Kiwifruit dry matter and soluble solids content. Postharvest Biol. Technol. 26, 191–198. https://doi.org/10.1016/S0925-5214(02)00014-5

McGlone, V.A., Kawano, S., 1998. Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. Postharvest Biol. Technol. 13, 131–141. https://doi.org/10.1016/S0925-5214(98)00007-6

Mehmood, T., Liland, K.H., Snipen, L., Saebo, S., Sæbø, S., 2012. A review of variable selection methods in Partial Least Squares Regression. Chemom. Intell. Lab. Syst. 118, 62–69. https://doi.org/10.1016/j.chemolab.2012.07.010

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. J. Food Eng. 125, 59–68. https://doi.org/10.1016/j.jfoodeng.2013.10.022

Mendoza, F., Lu, R., Cen, H., 2012. Comparison and fusion of four nondestructive sensors for predicting apple fruit firmness and soluble solids content. Postharvest Biol. Technol. 73, 89–98. https://doi.org/10.1016/j.postharvbio.2012.05.012

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 Bioprocess Technol. 2, 308–314. https://doi.org/10.1007/s11947-008-0120-8

Mevik, B.-H., Wehrens, R., Liland, K.H., 2018. pls: Partial Least Squares and Principal Component Regression.

Mika, A., 1992. Trends in Fruit Tree Training and Pruning Systems in Europe. Acta Hortic. 322, 29–35.

Miller, P.C.H., Tuck, C.R., 2014. Factors Influencing the Performance of Spray Delivery Systems : A Review of Recent Developments. J. ASTM Int. 2, 1–13. https://doi.org/10.1520/JAI12900

Mitcham, B., Cantwell, M., Kader, A., 1996. Methods for determining quality of fresh commodities, Perishables handling newsletter. Davis, CA.

Moons, E., Sinnaeve, G., Dardenne, P., 2000. Non destructive visible and nir spectroscopy measurement for the determination of apple internal quality. Acta Hortic. 517, 441–448.

Murakami, M., Himoto, J.-I., Itoh, K., 1994. Analysis of Apple Quality by Near Infrared Reflectance Spectroscopy. J. Fac. Agric. Hokkaido Univ. 66, 51–61.

Musacchi, S., Serra, S., 2018. Apple fruit quality: Overview on pre-harvest factors. Sci. Hortic. (Amsterdam). 234, 409–430. https://doi.org/10.1016/j.scienta.2017.12.057

Naes, T., Isakson, T., Fearn, T., Davies, T., 2002. 13.9 How accurate can you get?, in: A User-friendly Guide to Multivariate Calibration and Classification. NIR Publications, Chichester, UK, pp. 172–175.

Nagy, A., Riczu, P., Tamás, J., 2016. Spectral evaluation of apple fruit ripening and pigment content alteration. Sci. Hortic. (Amsterdam). 201, 256–264. https://doi.org/10.1016/j.scienta.2016.02.016

Nesme, T., Plenet, D., Hucbourg, B., Fandos, G., Lauri, P.E., 2005. A set of vegetative morphological variables to objectively estimate apple (Malus x domestica) tree orchard vigour. Sci. Hortic. (Amsterdam). 106, 76–90. https://doi.org/10.1016/j.scienta.2005.02.017

Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007a. Nondestructive measurement of fruit and vegetable quality by means of

NIR spectroscopy: A review. Postharvest Biol. Technol. 46, 99–118. https://doi.org/10.1016/j.postharvbio.2007.06.024

Nicolaï, B.M., Theron, K.I., Lammertyn, J., 2007b. Kernel PLS regression on wavelet transformed NIR spectra for prediction of sugar content of apple. Chemom. Intell. Lab. Syst. 85, 243–252. https://doi.org/10.1016/j.chemolab.2006.07.001

O'Brien, N.A., Hulse, C.A., Friedrich, D.M., Van Milligen, F.J., von Gunten, M.K., Pfeifer, F., Siesler, H.W., 2012. Miniature near-infrared (NIR) spectrometer engine for handheld applications, in: SPIE Defense, Security, and Sensing . Next-Generation Spectroscopic Technologies V. Baltimore, Maryland, pp. 837404-1–8. https://doi.org/10.1117/12.917983

Olivieri, A.C., 2015. Practical guidelines for reporting results in single- and multi-component analytical calibration: A tutorial. Anal. Chim. Acta 868, 10–22. https://doi.org/10.1016/j.aca.2015.01.017

OUYANG, A., XIE, X., ZHOU, Y., LIU, Y., 2012. Partial Least Squares Regression Variable Screening Studies on Apple Soluble Solids NIR Spectral Detection. Spectrosc. Spectr. Anal. 32, 2680-2684(5). https://doi.org/https://doi.org/10.3964/j

Palmer, J., 2017. Forty years of scientific research , a journey of exploration ? Acta Hortic. 1177, 13–28. https://doi.org/10.17660/ActaHortic.2017.1177.2

Palmer, J.W., 2011. Changing Concepts of Efficiency in Orchard Systems. Acta Hortic. 903, 41–50.

Palmer, J.W., 1989. The effects of row orientation, tree height, time of year and latitude on light interception and distribution in model apple hedgerow canopies. J. Hortic. Sci. 64, 137–145. https://doi.org/10.1080/14620316.1989.11515937

Palmer, J.W., Harker, F.R., Tustin, D.S., Johnston, J., 2010. Fruit dry matter concentration: A new quality metric for apples. J. Sci. Food Agric. 90, 2586–2594. https://doi.org/10.1002/jsfa.4125

Pasquini, C., 2018. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. Anal. Chim. Acta 1026, 8–36. https://doi.org/10.1016/j.aca.2018.04.004

Paz, P., Sánchez, M.T., Pérez-Marín, D., Guerrero, J.E., Garrido-Varo, A., 2009. Evaluating NIR instruments for quantitative and qualitative assessment of intact apple quality. J. Sci. Food Agric. 89, 781–790. https://doi.org/10.1002/jsfa.3512

Peace, C.P., 2017. DNA-informed breeding of rosaceous crops: Promises, progress and prospects. Hortic. Res. 4, 1–13. https://doi.org/10.1038/hortres.2017.6

Peiris, K.H.S., Dull, G.G., Leffler, R.G., Kays, S.J., 1999. Spatial variability of soluble solids or dry-matter content within individual fruits, bulbs, or tubers: Implications for the development and use of NIR spectrometric techniques. HortScience 34, 114–118.

Peirs, A., Lammertyn, J., Ooms, K., Nicola, B.M., 2000. Prediction of the optimal picking date of different apple cultivars by means of VIS/NIR-spectroscopy. Postharvest Biol. Technol. 21, 189–199. https://doi.org/10.1016/S0925-5214(00)00145-9

Peirs, A., Scheerlinck, N., Nicolaï, B.M., 2003a. Temperature compensation for near infrared reflectance measurement of apple fruit soluble solids contents. Postharvest Biol. Technol. 30, 233–248. https://doi.org/10.1016/S0925-5214(03)00118-2

Peirs, A., Scheerlinck, N., Perez, A.B., Jancsók, P., Nicolaï, B.M., 2002. Uncertainty analysis and modelling of the starch index during apple fruit maturation. Postharvest Biol. Technol. 26, 199–207. https://doi.org/10.1016/S0925-5214(02)00038-8

Peirs, A., Schenk, A., Nicolaï, B.M., 2005. Effect of natural variability among apples on the accuracy of VIS-NIR calibration models for optimal harvest date predictions. Postharvest Biol. Technol. 35, 1–13. https://doi.org/10.1016/j.postharvbio.2004.05.010

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 Biol. Technol. 28, 269–280. https://doi.org/10.1016/S0925-5214(02)00196-5

Péneau, S., Hoehn, E., Roth, H.R., Escher, F., Nuessli, J., 2006. Importance and consumer perception of freshness of apples. Food Qual. Prefer. 17, 9–19. https://doi.org/10.1016/j.foodqual.2005.05.002

Pérez-Marín, D., Sánchez, M.-T.T., Paz, P., Soriano, M.-A.A., Guerrero, J.-E.E., Garrido-Varo, A., 2009. Non-destructive determination of quality parameters in nectarines during on-tree ripening and postharvest storage. Postharvest Biol. Technol. 52, 180–188. https://doi.org/10.1016/j.postharvbio.2008.10.005

Perry, R.L., 2000. The Vertical Axis System in North America. Acta Hortic. 513, 297–301.

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. Biotechnol. Agron. Soc. Environ. 2018 22, 1–10.

Pissard, A., Bastiaanse, H., Baeten, V., Sinnaeve, G., Romnee, J.M., Dupont, P., Mouteau, A., Lateur, M., 2013. Use of NIR spectroscopy in an apple breeding program for quality and nutritional parameters. Acta Hortic. 976, 409–414. https://doi.org/10.17660/actahortic.2013.976.56

Pissard, Audrey, Fernández Pierna, J.A., Baeten, V., Sinnaeve, G., Lognay, G., Mouteau, A., Dupont, P., Rondia, A., Lateur, M., 2013. Non-destructive measurement of vitamin C, total polyphenol and sugar content in apples using near-infrared spectroscopy. J. Sci. Food Agric. 93, 238–244. https://doi.org/10.1002/jsfa.5779

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Found. Stat. Comput.

Rateni, G., Dario, P., Cavallo, F., 2017. Smartphone-based food diagnostic technologies: A review. Sensors (Switzerland) 17, 1–22. https://doi.org/10.3390/s17061453

Reganold, J.P., Glover, J.D., Andrews, P.K., Hinman, H.R., 2001. Sustainbility of three apple production systems. Nature 410, 927–929.

Reid, M.S., Padfield, C.A.S., Watkins, C.B., Harman, J.E., 2017. Starch iodine pattern as a maturity index for Granny Smith apples. New Zeal. J. Agric. Res. 25, 229–237. https://doi.org/10.1080/00288233.1982.10420918

Robinson, T., Hoying, S., Sazo, M.., DeMarree, A., Dominguez, L., 2013. A vision for apple orchard systems of the future. New York Fruit Q. 21, 11–16.

Robinson, T.L., 2017. Can we manage light interception levels above 70% in apple orchards? Acta Hortic. 79–86. https://doi.org/10.17660/ActaHortic.2017.1177.8

Robinson, T.L., 2003. Apple-orchard Planting Systems, in: Ferree, D.C., Warrington, U. (Eds.), Apples: Botany, Production and Uses. CAB International, pp. 346–407.

Robinson, T.L., 2000. V-shaped apple planting systems. Acta Hortic.

Robinson, T.L., Hoying, S.A., Reginato, G.H., 2011. The tall spindle planting system: Principles and performance. Acta Hortic. 903, 571–579.

Robinson, T.L., Lakso, A.N., 1991. Bases of yield and production efficiency in apple orchard systems. J. Am. … 116, 188–194.

Robinson, T.L., Rufato, A.D., Rufato, L., Dominguez, L.I., 2017. Is fruit dry matter concentration a useful predictor of ‘ Honeycrisp ’ apple fruit quality after storage ? Acta Hortic. 1177, 151–156. https://doi.org/10.17660/ActaHortic.2017.1177.20

Rosell, J.R., Sanz, R., 2012. A review of methods and applications of the geometric characterization of tree crops in agricultural activities. Comput. Electron. Agric. 81, 124–141. https://doi.org/10.1016/j.compag.2011.09.007

Rosell Polo, J.R., Sanz, R., Llorens, J., Arnó, J., Escolà, A., Ribes-Dasi, M., Masip, J., Camp, F., Gràcia, F., Solanelles, F., Pallejà, T., Val, L., Planas, S., Gil, E., Palacín, J., 2009. A tractor-mounted scanning LIDAR for the non-destructive measurement of vegetative volume and surface area of tree-row plantations: A comparison with conventional destructive measurements. Biosyst. Eng. 102, 128–134. https://doi.org/10.1016/j.biosystemseng.2008.10.009

Rowan, D.D., Hunt, M.B., Dimouro, A., Alspach, P.A., Weskett, R., Volz, R.K., Gardiner, S.E., Chagné, D., 2009. Profiling fruit volatiles in the progeny of a “royal gala” x “granny smith” apple (Malus x domestica) cross. J. Agric. Food Chem. 57, 7953–7961. https://doi.org/10.1021/jf901678v

Rowe, P.I., Künnemeyer, R., Mcglone, V.A., Talele, S., Martinsen, P., Seelye, R., 2014. Relationship between tissue firmness and optical properties of “Royal Gala” apples from 400 to 1050nm. Postharvest Biol. Technol. 94, 89–96. https://doi.org/10.1016/j.postharvbio.2014.03.007

Roy, P., Isler, V., 2016. Surveying apple orchards with a monocular vision system. IEEE Int. Conf. Autom. Sci. Eng. 2016-Novem, 916–921. https://doi.org/10.1109/COASE.2016.7743500

Roy, P., Kislay, A., Plonski, P.A., Luby, J., Isler, V., 2018. Vision-Based Preharvest Yield Mapping for Apple Orchards. arXiv 1–21.

RStudio Team, 2016. RStudio: Integrated Development Environment for R.

Sánchez-Rangel, J.C., Benavides, J., Heredia, J.B., Cisneros-Zevallos, L., Jacobo-Velázquez, D. a., 2013. The Folin–Ciocalteu assay revisited: improvement of its specificity for total phenolic content determination. Anal. Methods 5, 5990. https://doi.org/10.1039/c3ay41125g

Sansavini, S., Corelli, L., 1992. Canopy Efficiency of Apple as Affected by Microclimatic Factors and Tree Structure. Acta Hortic. 322, 69–77.

Saure, M.C., 1990. External control of anthocyanin formation in apple. Sci. Hortic. (Amsterdam). 42, 181–218. https://doi.org/10.1016/0304-4238(90)90082-P

Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047

Schaare, P.N., Fraser, D.G., 2000. Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (Actinidia chinensis). Postharvest Biol. Technol. 20, 175–184. https://doi.org/10.1016/S0925-5214(00)00130-7

Schafer, R.W., 2011. What is a savitzky-golay filter? IEEE Signal Process. Mag. 28, 111–117. https://doi.org/10.1109/MSP.2011.941097

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: An open-source platform for biological-image analysis. Nat. Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019

Schmitz, C.A., Clark, M.D., Luby, J.J., Bradeen, J.M., Guan, Y., Evans, K., Orcheski, B., Brown, S., Verma, S., Peace, C., 2013. Fruit texture phenotypes of the RosBREED U.S. apple reference germplasm set. HortScience 48, 296–303.

Schmutzler, M., Lutz, O.M.D., Huck, C.W., 2013. Analytical pathway based on non-destructive nirs for quality control of apples. Infrared Spectrosc. Theory, Dev. Appl. 1–22.

Seetin, M., 2017. 2017 U.S. Apple Crop Outlook and Overview.

Segura, V., Cilas, C., Laurens, F., Costes, E., 2006. Phenotyping progenies for complex architectural traits: A strategy for 1-year-old apple trees (Malus x domestica Borkh.). Tree Genet. Genomes 2, 140–151. https://doi.org/10.1007/s11295-006-

0037-1

Seifert, B., Pflanz, M., Zude, M., 2014. Spectral shift as advanced index for fruit chlorophyll breakdown. Food Bioprocess Technol. 7, 2050–2059. https://doi.org/10.1007/s11947-013-1218-1

Sestras, A., Sestras, R., Lazar, V., 2009. The Influence of Fruit Position in the Crown of Trees on the Sugar Content and Morphological Traits of Apple Fruits. Bull. UASVM Hortic. 66, 170–176.

Singleton, V.L., Orthofer, R., Lamuela-Raventós, R.M., 1999. Analysis of total phenols and other oxidation substrates and antioxidants by means of folin-ciocalteu reagent. Methods Enzymol. 299, 152–178. https://doi.org/10.1016/S0076-6879(99)99017-1

Smith, R.B., Lougheed, E.C., Franklin, E.W., McMillan, I., 1979. STAGE OF MATURATION IN APPLES apparently the same stage of. Can. J. Plant Sci. 725–735.

Stephan, J., Sinoquet, H., Donès, N., Haddad, N., Talhouk, S., Lauri, P.É., 2008. Light interception and partitioning between shoots in apple cultivars influenced by training. Tree Physiol. 28, 331–342. https://doi.org/10.1093/treephys/28.3.331

Strong, D., Azarenko, A.N., 2000. Relationship Between Trunk Cross-sectional Area, Harvest Index, Total Tree Dry Weight and Yield Components of 'Starkspur Supreme Delicious' Apple Trees. J. Am. Pomol. Soc. 54, 22–27. https://doi.org/10.23736/S0392-9590.16.03730-5

Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D., 2012. A benchmark for the evaluation of RGB-D SLAM systems. IEEE Int. Conf. Intell. Robot. Syst. 573–580. https://doi.org/10.1109/IROS.2012.6385773

Sun-Waterhouse, D., Waterhouse, G.I.N., Jin, D., Wibisono, R., Luberriaga, C., Wadhwa, S.S., 2011. Juices, Fibres and Skin Waste Extracts from White, Pink or Red-Fleshed Apple Genotypes as Potential Food Ingredients. Food Bioprocess Technol. 6, 377–390. https://doi.org/10.1007/s11947-011-0692-6

Sun, X., Zhang, H., Pan, Y., Liu, Y., 2009. Nondestructive measurement soluble solids content of apple by portable and online near infrared spectroscopy. Photonics Optoelectron. Meet. POEM 2009 - Fiber Opt. Commun. Sensors 7514. https://doi.org/10.1117/12.843390

Szalay, L., Ordidge, M., Ficzek, G., Hadley, P., Tóth, M., Battey, N.H., 2013. Grouping of 24 apple cultivars on the basis of starch degradation rate and their fruit pattern. Hortic. Sci. 40, 93–101. https://doi.org/10.1111/jfpp.12544

Teixeira dos Santos, C.A., Páscoa, R.N., Lopo, M., Lopes, J.A., Teixeira dos Santos, C.A., Páscoa, R.N., Lopo, M., Lopes, J.A., 2015. Applications of Portable Near-infrared Spectrometers. Encycl. ofAnalytical Chem. https://doi.org/10.1002/9780470027318.A9455

Temma, T., Hanamatsu, K., Shinoki, F., 2002a. Development of a portable near infrared

sugar-measuring instrument. J. Near Infrared Spectrosc. 10, 77–83. https://doi.org/10.1255/jnirs.324

Temma, T., Hanamatsu, K., Smnoki, F., 2002b. Measuring the Sugar Content of Apples and Apple Julce by Near Infrared Spectroscopy. Opt. Rev. 9, 40–44.

Torres, C.A., León, L., Sánchez-contreras, J., Leon, L., Sánchez-contreras, J., 2016. Spectral fingerprints during sun injury development on the tree in Granny Smith apples: A potential non-destructive prediction tool during the growing season. Sci. Hortic. (Amsterdam). 209, 165–172. https://doi.org/10.1016/j.scienta.2016.06.024

Travers, S., 2013. Dry Matter and Fruit Quality: Manipulation in the Field and Evaluation with NIR Spectroscopy. Aarhus Univ. Aarhus University.

Travers, S., Bertelsen, M.G., Kucheryavskiy, S. V., 2014. Predicting apple (cv. Elshof) postharvest dry matter and soluble solids content with near infrared spectroscopy. J. Sci. Food Agric. 94, 955–962. https://doi.org/10.1002/jsfa.6343

Trujillo, D.I., Mann, H.S., Tong, C.B.S., 2012. Examination of expansin genes as related to apple fruit crispness. Tree Genet. Genomes 8, 27–38. https://doi.org/10.1007/s11295-011-0417-z

Tustin, D.S., 2014. Future orchard planting systems - Do we need another revolution? Acta Hortic. 1058, 27–36. https://doi.org/10.17660/ActaHortic.2014.1058.1

Tustin, D.S., 2000. The slender pyramid tree management system - In pursuit of higher standards of apple fruit quality. Acta Hortic. https://doi.org/10.17660/ActaHortic.1998.513.37

USDA FAS, 2018. Fresh Deciduous Fruit: World Markets and Trade (Apples, Grapes, & Pears), Production, Supply and Distribution (PS&D).

USDA NASS, 2017. Apples: Value of Utilized Production, US.

Usha, K., Singh, B., 2013. Potential applications of remote sensing in horticulture-A review. Sci. Hortic. (Amsterdam). 153, 71–83. https://doi.org/10.1016/j.scienta.2013.01.008

Vázquez-Arellano, M., Griepentrog, H.W., Reiser, D., Paraforos, D.S., 2016. 3-D imaging systems for agricultural applications—a review. Sensors (Switzerland) 16. https://doi.org/10.3390/s16050618

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Fourth. ed, Statistics and Computing. Springer-Verlag, New York. https://doi.org/10.1007/978-0-387-21706-2

Ventura, M., De Jager, A., De Putter, H., Roelofs, F.P. m. m., 1998. Non-destructive determination of soluble solids in apple fruit by near infrared spectroscopy (NIRS). Postharvest Biol. Technol. 14, 21–27. https://doi.org/10.1016/S0925-5214(98)00030-1

Virlet, N., Costes, E., Martinez, S., Kelner, J.J., Regnard, J.L., 2015. Multispectral

airborne imagery in the field reveals genetic determinisms of morphological and transpiration traits of an apple tree hybrid population in response to water deficit. J. Exp. Bot. 66, 5453–5465. https://doi.org/10.1093/jxb/erv355

Wagenmakers, P.S., Callesen, O., 1995. Light distribution in apple orchard systems in relation to production and fruit quality. J. Hortic. Sci. 70, 935–948. https://doi.org/10.1080/14620316.1995.11515369

Walsh, K.B., 2016a. The evolution of spectrophotometers used in fruit quality assessment. Acta Hortic. 1119, 203–208. https://doi.org/10.17660/ActaHortic.2016.1119.28

Walsh, K.B., 2016b. Detection of attribute XXX in fruit YYY using NIRS. Acta Hortic. 1119, 141–146. https://doi.org/10.17660/ActaHortic.2016.1119.19

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. J. Near Infrared Spectrosc. 12, 141–148. https://doi.org/10.1255/jnirs.419

Walsh, K.B., Guthrie, J.A., Burney, J.W., 2000. Application of commercially available, low-cost, miniaturised NIR spectrometers to the assessment of the sugar content of intact fruit. Aust. J. Plant Physiol. 27, 1175–1186.

Wang, H., Peng, J., Xie, C., Bao, Y., He, Y., 2015. Fruit Quality Evaluation Using Spectroscopy Technology: A Review. Sensors 15, 11889–11927. https://doi.org/10.3390/s150511889

Wang, Q., Nuske, S., Bergerman, M., Singh, S., 2013. Automated Crop Yield Estimation for Apple Orchards. Exp. Robot. 745–758. https://doi.org/10.1007/978-3-319-00065-7

Wannemuehler, S.D., 2018. Cost-Benefit Analysis of Marker-Assisted Selection in Rosaceous Fruit Crop Breeding Programs. University of Minnesota. https://doi.org/http://hdl.handle.net/11299/200133

Wargovich, M.J., Morris, J., Moseley, V., Weber, R., Byrne, D.H., 2012. Developing Fruit Cultivars with Enhanced Health Properties, in: Byrne, D.H., Badenes, M.L. (Eds.), Fruit Breeding. Springer, New York, pp. 37–68. https://doi.org/10.1007/978-1-4419-0763-9

Watkins, C.B., 2003. Principles and practices of postharvest handling and stress., in: Ferree, D.C., Warrington, I.J. (Eds.), Apples: Botany, Production and Uses. CAB International, pp. 585–614.

Weber, M.S., 2000. The super spindle system. Acta Hortic.

Wehrens, R., 2011. Chemometrics with R: Multivariate Data Analysis in the Natural Sciences and Life Sciences, Applied Spatial Data Analysis with R. Springer, Berlin Heidelberg. https://doi.org/10.1007/978-1-4419-7976-6

Wendel, A., Underwood, J., Walsh, K., 2018. Maturity estimation of mangoes using hyperspectral imaging from a ground based mobile platform. Comput. Electron. Agric. 155, 298–313. https://doi.org/10.1016/j.compag.2018.10.021

Westwood, M., Roberts, A., 1970. The relationship between trunk cross-sectional area and weight of apple trees. J. Am. Soc. Hortic. Sci. 95, 28–30.

Willaume, M., Lauri, P.É., Sinoquet, H., 2004. Light interception in apple trees influenced by canopy architecture manipulation. Trees - Struct. Funct. 18, 705–713. https://doi.org/10.1007/s00468-004-0357-4

Williams, P., Dardenne, P., Flinn, P., 2017. Tutorial: Items to be included in a report on a near infrared spectroscopy project. J. Near Infrared Spectrosc. 25, 85–90. https://doi.org/10.1177/0967033517702395

Wismer, W. V., 2014. Consumer Eating Habits and Perceptions of Fresh Produce Quality, Postharvest Handling: A Systems Approach. Elsevier Inc. https://doi.org/10.1016/B978-0-12-408137-6.00003-X

Wojdyło, A., 2008. Polyphenolic compounds and antioxidant activity of new and old apple varieties. J. Agric. Food Chem. 56, 6520–6530.

Wulfsohn, D., Maletti, G.M., Toldam-Andersen, T.B., 2004. Unbiased estimator for the total number of Flowers on a tree. 7th Int. Syposium Comput. Model. Fruit Res. Orchard Manag. 20–24.

Wunsche, J.N., Lakso, A.N., Robison, T.L., Lenz, F., Denning, S.S., 1996. The Basis of Apple Productivity in Apple Production Systems: The role of Light Interception by Different Shoot types. J. Amer. Soc. Hort. Sci 121, 886–893.

Yuan, L. ming, Cai, J. rong, Sun, L., Han, E., Ernest, T., 2015. Nondestructive Measurement of Soluble Solids Content in Apples by a Portable Fruit Analyzer. Food Anal. Methods 9, 785–794. https://doi.org/10.1007/s12161-015-0251-2

Zhang, Y., Li, P., Cheng, L., 2010. Developmental changes of carbohydrates, organic acids, amino acids, and phenolic compounds in “Honeycrisp” apple flesh. Food Chem. 123, 1013–1018. https://doi.org/10.1016/j.foodchem.2010.05.053

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 Biol. Technol. 151, 111–118. https://doi.org/10.1016/J.POSTHARVBIO.2019.01.009

Zude-Sasse, M., Fountas, S., Gemtos, T.A., Abu-Khalaf, N., 2016. Applications of precision agriculture in horticultural crops. Eur. J. Hortic. Sci. 81, 78–90. https://doi.org/10.17660/eJHS.2016/81.2.2

Zude, M., Herold, B., 2002. Optimum harvest date determination for apples using spectral analysis. Die Gartenbauwiss. 67, 199–204.

Zude, M., Herold, B., Roger, J.-M.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. J. Food Eng. 77, 254–260. https://doi.org/10.1016/j.jfoodeng.2005.06.027

APPENDIX A FRUIT TRAIT STATISTICS, DESCRIPTIVE PLOTS, AND PLSR PREDICTION PLOTS

A.1 Descriptive Statistics and Trait Distributions

Table A 1. Whole dataset trait statistics for fruit processed at-harvest time in 2016

Trait	Cultivars (N)	Calibration (N)	Validation (N)	Min	Max	SD	Mean	Median
A1	14	383	128	18.5	58.0	5.5	37.4	37.1
A2	15	404	135	33.4	110.2	11.2	73.6	73.4
C0	14	381	127	0.0	3.7	0.5	0.5	0.4
Cn	14	377	126	4.6	619.1	101.2	189.7	171.0
E2	15	412	137	27.2	162.6	14.7	86.2	87.1
M1	14	383	128	28.5	96.0	9.9	60.1	59.0
M2	15	404	135	40.3	182.0	13.8	91.0	90.5
OAH	15	412	137	27.7	86.4	8.7	60.3	60.3
OMH	15	412	137	40.3	182.0	13.7	90.9	90.5
QF	15	412	137	-178.9	135.1	44.3	40.0	43.5
%Ocolor	15	413	138	0.0	10.0	2.2	7.1	8.0
%Russet	15	413	138	0.0	6.5	0.8	1.2	1.0
DM	15	407	136	9.6	27.2	2.3	16.3	16.1
pH	15	406	135	3.1	4.2	0.2	3.5	3.5
SSC	15	405	135	8.5	17.1	1.6	12.7	12.7
SPI	15	413	138	1.0	8.0	2.0	5.9	7.0
TA	15	406	135	1.2	7.9	1.3	3.9	3.9
Flav	14	291	97	0.2	0.8	0.1	0.4	0.4
Phen	14	291	97	0.1	0.8	0.1	0.4	0.3
Diameter	15	412	137	45.1	108.6	10.2	72.7	73.4
Weight	15	413	138	26.6	430.9	61.8	169.0	166.3
Wt/Diam	15	412	137	0.3	4.4	0.6	2.3	2.3

Table A 2. Whole dataset trait statistics for fruit processed poststorage in 2016

Trait	Cultivars (N)	Calibration (N)	Validation (N)	Min	Max	SD	Mean	Median
A1	14	396	132	13.7	43.9	5.9	30.0	30.6
A2	15	404	135	22.9	78.8	11.2	55.8	56.3
Creep	14	396	132	0.0	3.8	1.0	1.3	0.9
Crispness	14	389	130	1.8	622.5	105.3	133.5	101.0
E2	15	420	140	21.3	103.5	16.0	62.2	61.5
M1	14	396	132	23.5	75.5	10.0	48.2	50.1
M2	15	404	135	22.9	113.3	15.4	66.7	66.0
OAH	15	421	140	11.1	66.5	10.1	45.4	45.4
OMH	15	422	141	11.8	113.3	15.9	66.2	65.9
QF	15	424	141	-190.8	85.5	65.2	-35.1	-27.1
%Ocolor	14	404	135	0.0	9.5	2.2	7.0	7.5
%Russet	14	404	135	0.0	6.5	0.6	1.2	1.0
DM	15	411	137	11.0	24.5	2.1	16.3	16.3
pH	15	427	142	3.3	4.6	0.2	3.8	3.8
SSC	15	427	142	8.4	17.5	1.6	12.8	12.9
TA	15	427	142	0.8	6.6	1.0	2.8	2.7
Flav	15	314	105	0.2	0.7	0.1	0.5	0.4
Phen	15	314	105	0.1	0.7	0.1	0.4	0.3
Diameter	15	429	143	41.2	110.6	10.5	72.0	73.4
Wt/Diam	14	404	135	0.3	4.3	0.5	2.2	2.3
PHweight	15	427	142	35.2	388.7	57.3	166.1	166.6

Trait distribution plots

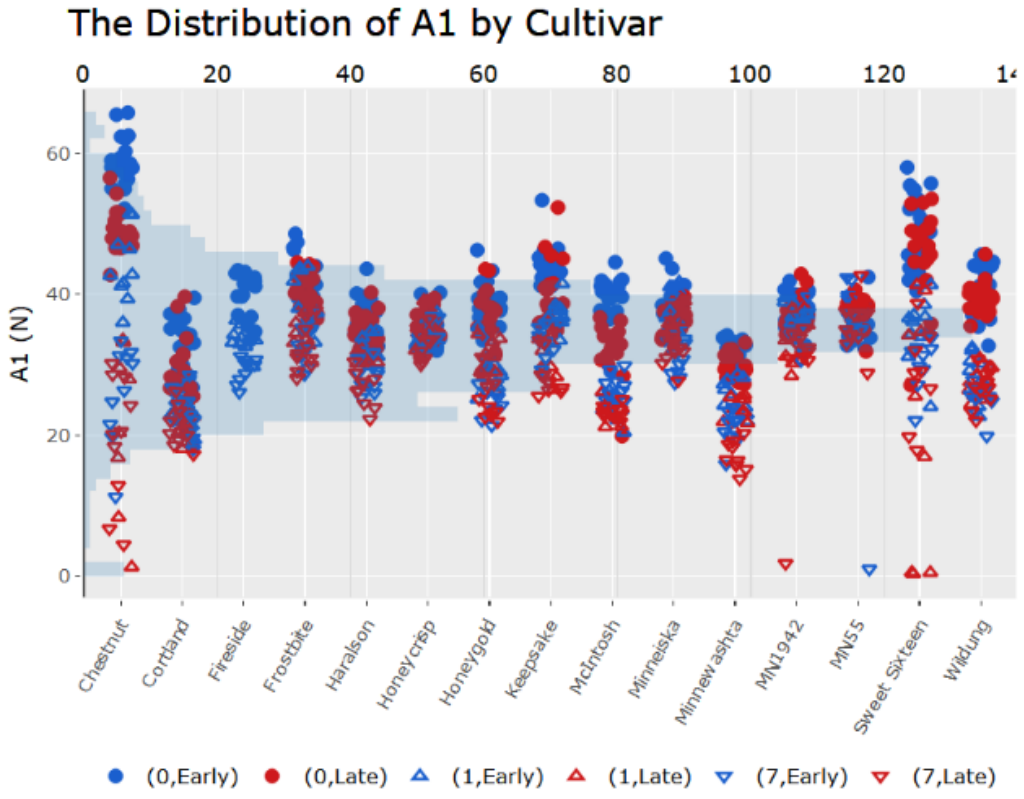
The plots below show the distributions for the 25 trait values by cultivar (listed here, but described in detail elsewhere). There were 16 quantitatively measured traits A1, A2, SSC, C0, MDT crispness, diameter, dry matter, E2, M1, M2, OAH, OMH, pH, QF, TA, and weight. The categorical or ordinal traits include the presence of BRusset, fruit damage, Gcolor, presence of lenticel russet, Ocolor, Ocolor type, %Ocolor, %Russet, and SPI.

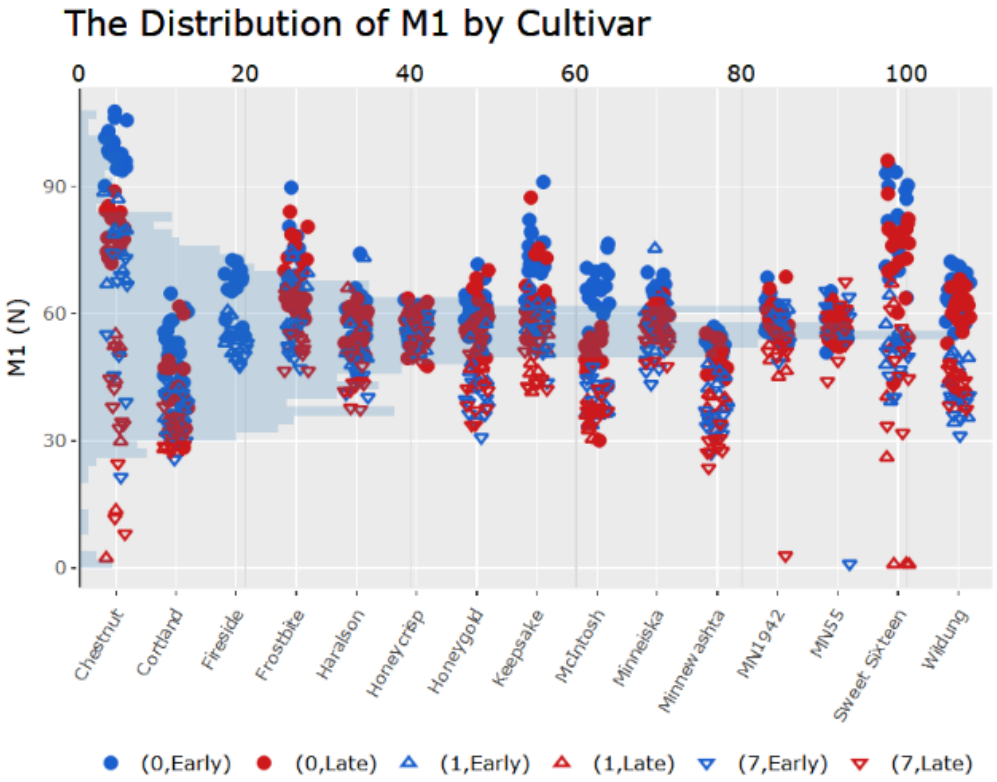
A1, M1, C0

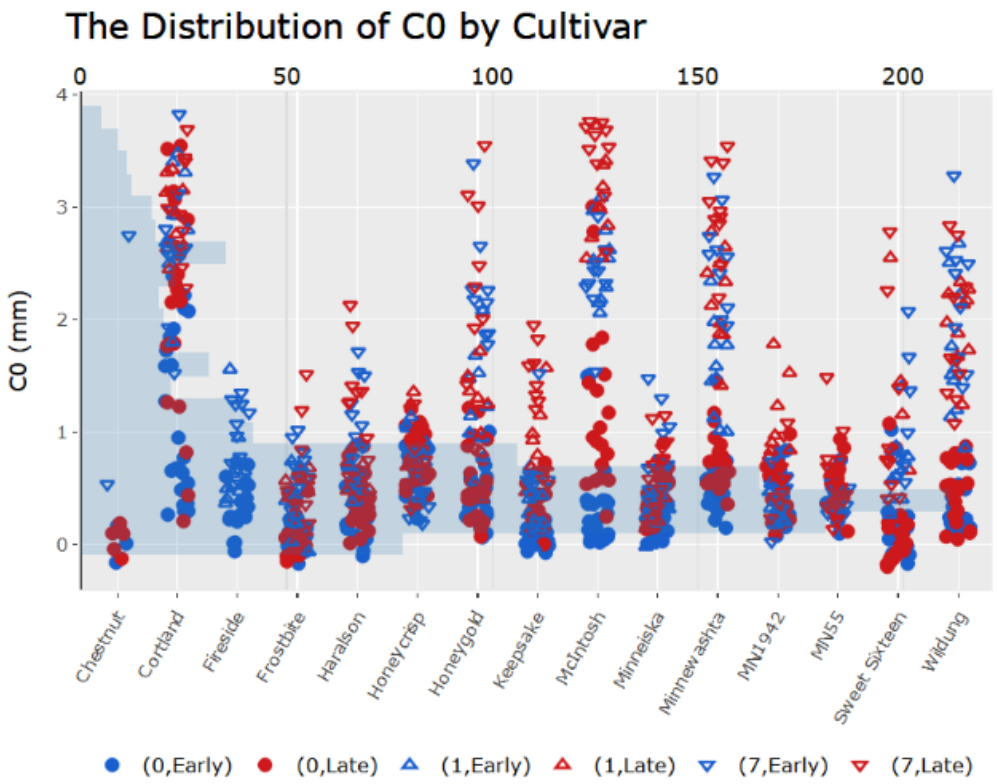
The distribution of average (A1) and maximum (M1) firmness observations measured within region one have expected changes in firmness, in that the stored fruit and fruit from the later harvest had lower firmness values. The creep distance (C0) increased for fruit that were harvested later and for fruit that were stored. The distribution of the measured values depended on the cultivar. Fruit from Chestnut crabapple (n=48) and Frostbite (n=2), that were too small for the machine to correctly measure C0 and were removed before plotting.

Note for all distribution plots below: *lower x-axis lists cultivars, dots are jittered (horizontally and vertically) and colored according harvest week, and shape is according to processing day. The upper x-axis is the counts for the histogram distributions for the blue colored histogram bins. The y-axis is the respective value for each trait.*

Figure A- 1. Distribution plots for fruit firmness in region 1, and creep distance.



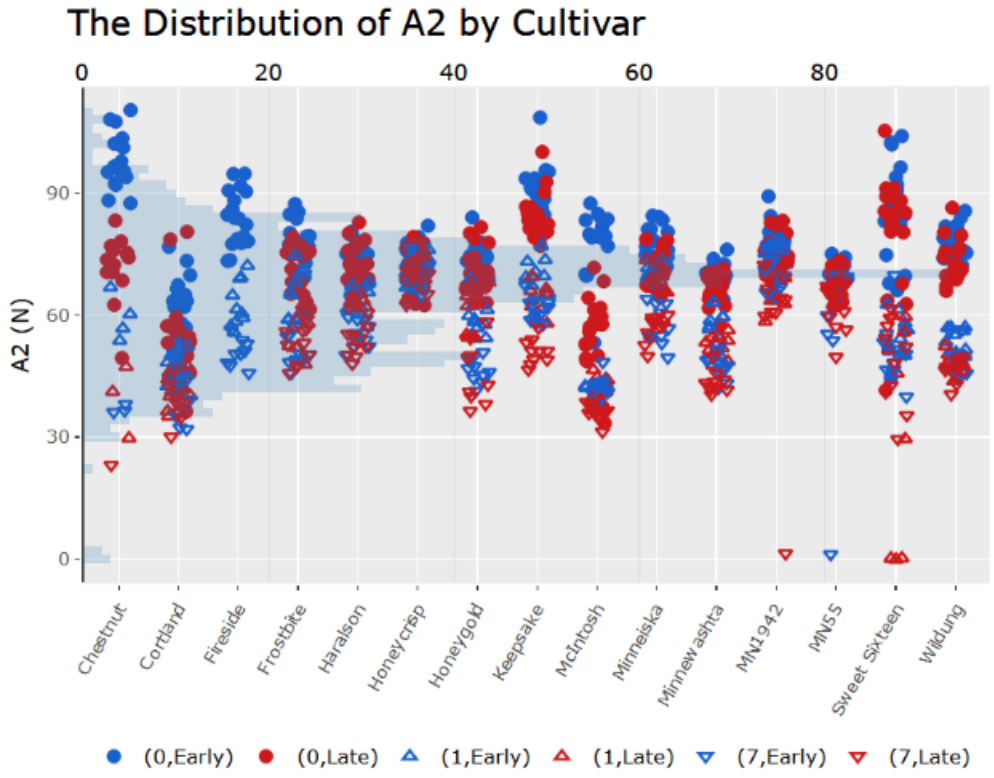

The Distribution of M1 by Cultivar
M1 (N)
0
20
40
60
80
100
90
60
30
0
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
Honeygold
Keepsake
McIntosh
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(0,Late)
(1,Early)
(1,Late)
(7,Early)
(7,Late)

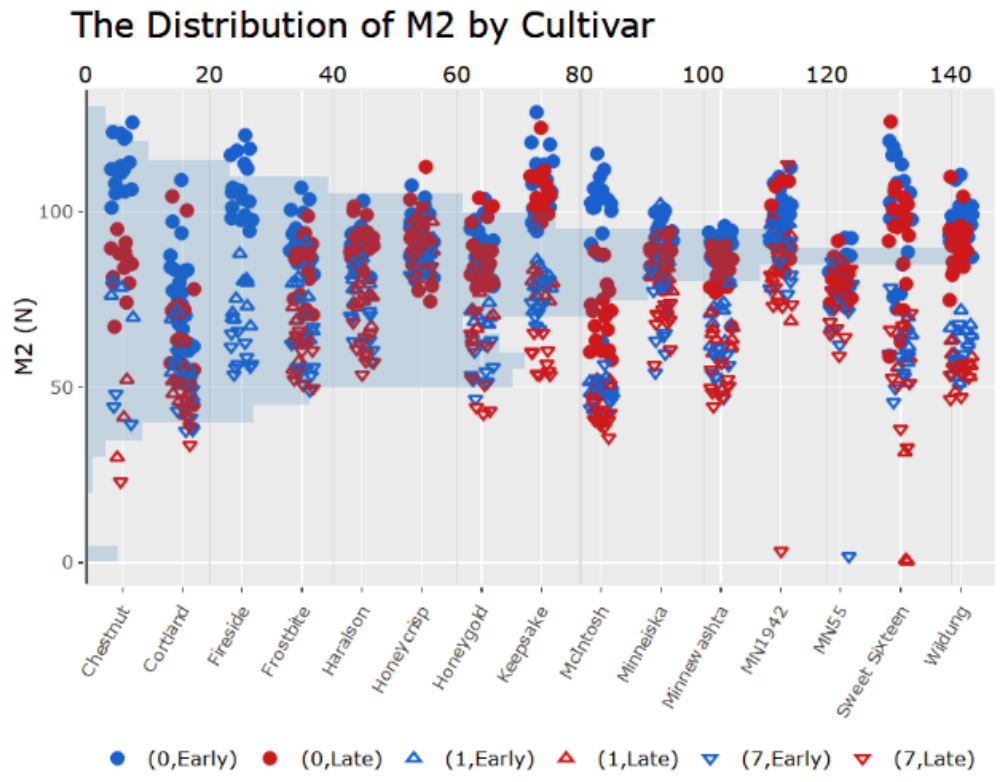

The Distribution of C0 by Cultivar
C0 (mm)
0
50
100
150
200
4
3
2
1
0
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
Honeygold
Keepsake
McIntosh
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(0,Late)
(1,Early)
(1,Late)
(7,Early)
(7,Late)

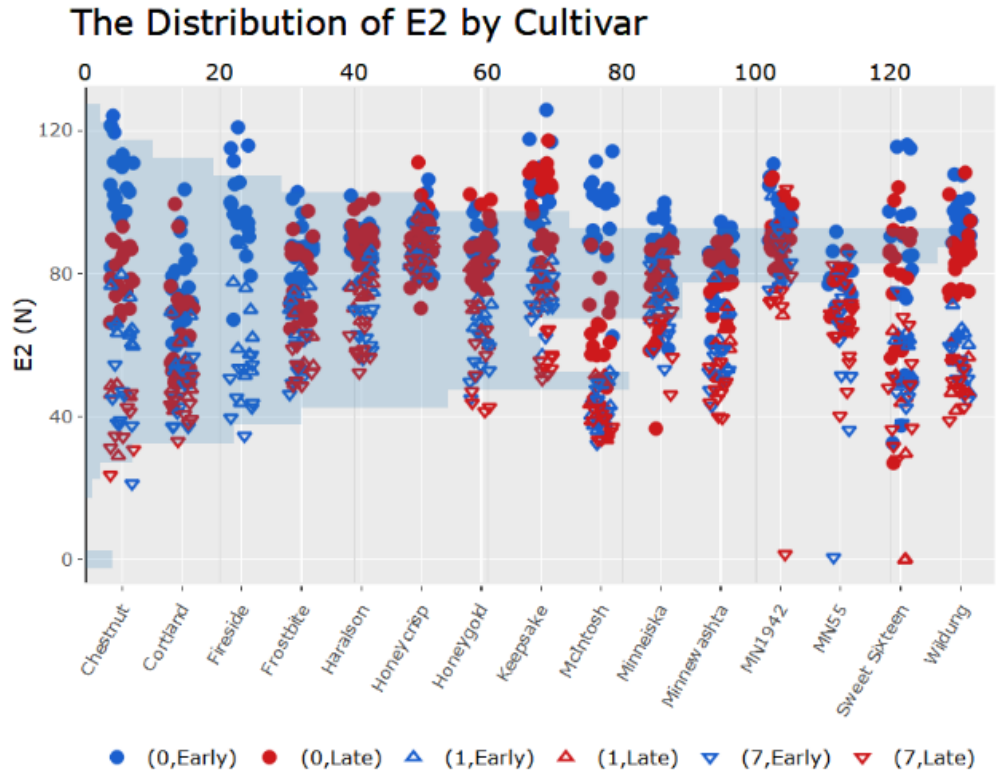
A2, M2, & E2

The distribution of the firmness at the end of region two (E2) with the average (A2) and maximum (A2) firmness observations measured within region two also had the general expected changes in firmness, in that the stored fruit and fruit from the later harvest had lower firmness values.

Figure A- 2: Distribution plots for fruit firmness traits in region 2 .



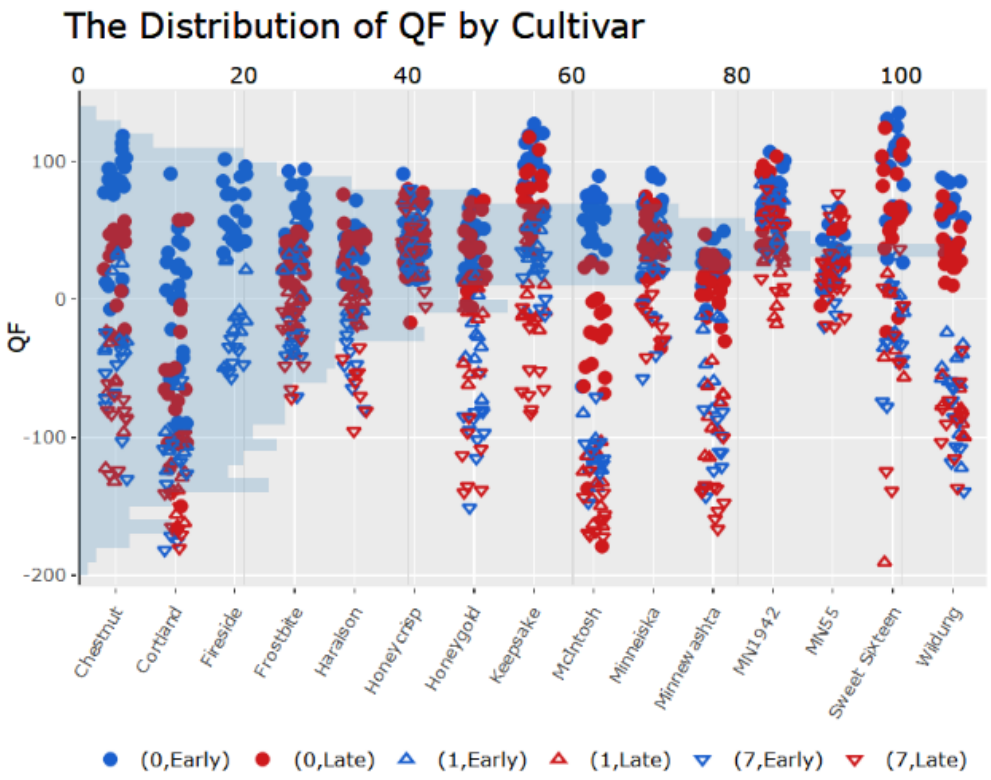

The Distribution of M2 by Cultivar
M2 (N)
0
20
40
60
80
100
120
140
0
50
100
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
Honeygold
Keepsake
McIntosh
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(0,Late)
(1,Early)
(1,Late)
(7,Early)
(7,Late)

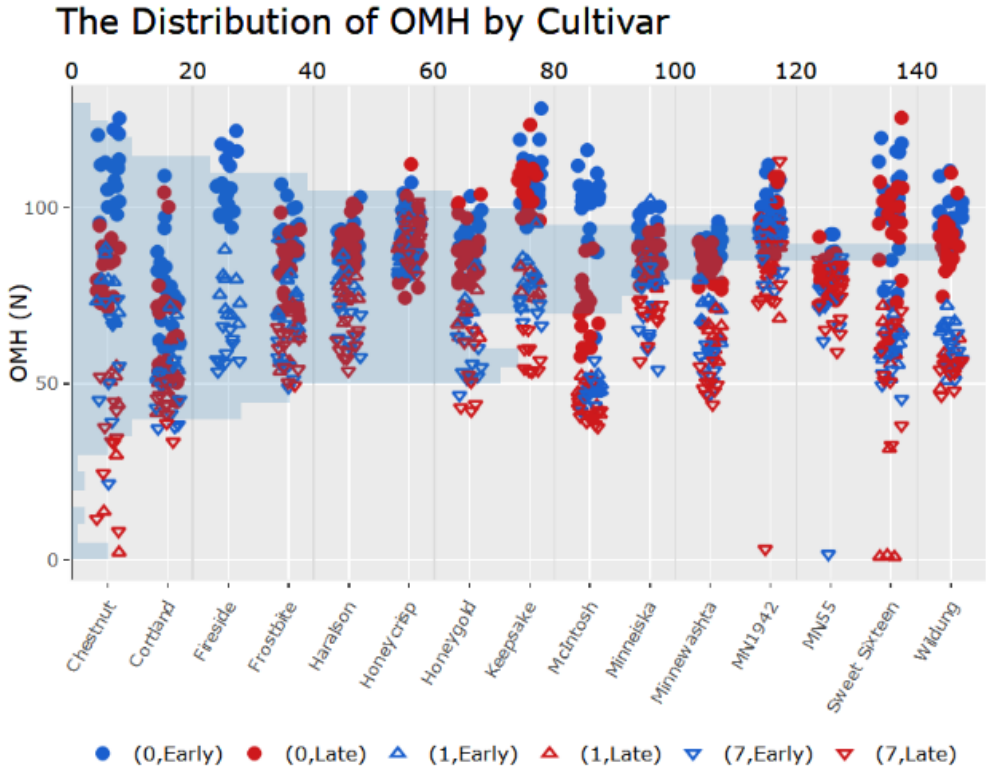

The Distribution of E2 by Cultivar
E2 (N)
0
20
40
60
80
100
120
0
40
80
120
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
Honeygold
Keepsake
McIntosh
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(0,Late)
(1,Early)
(1,Late)
(7,Early)
(7,Late)

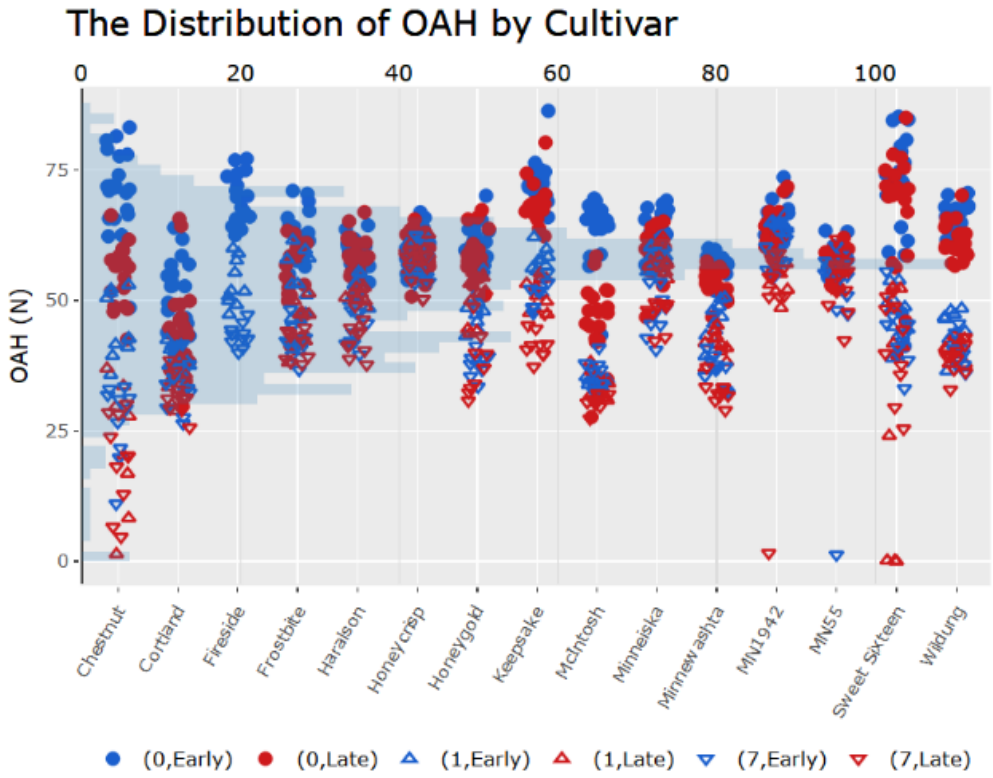
OMH, OAH, & QF

Cultivars with softer flesh generally showed distinct differences in firmness (OMH and OAH) between harvest dates and between stored and non-stored fruit. When assessing the firmness of a fruit cultivar the average and tightness of distribution should be considered. A tighter distribution would indicate more consistent quality is expected. The Mohr's quality factor (QF) is one variable that shows this fairly well.

Figure A- 3. Distribution plots for fruit texture traits

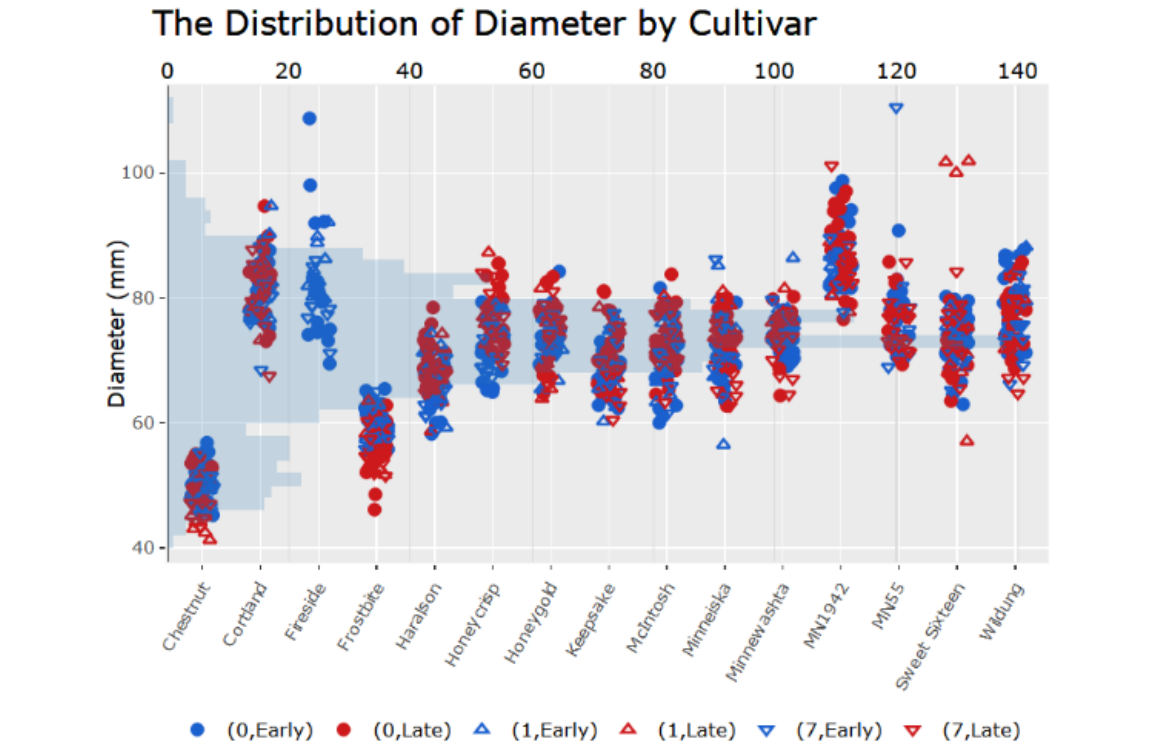


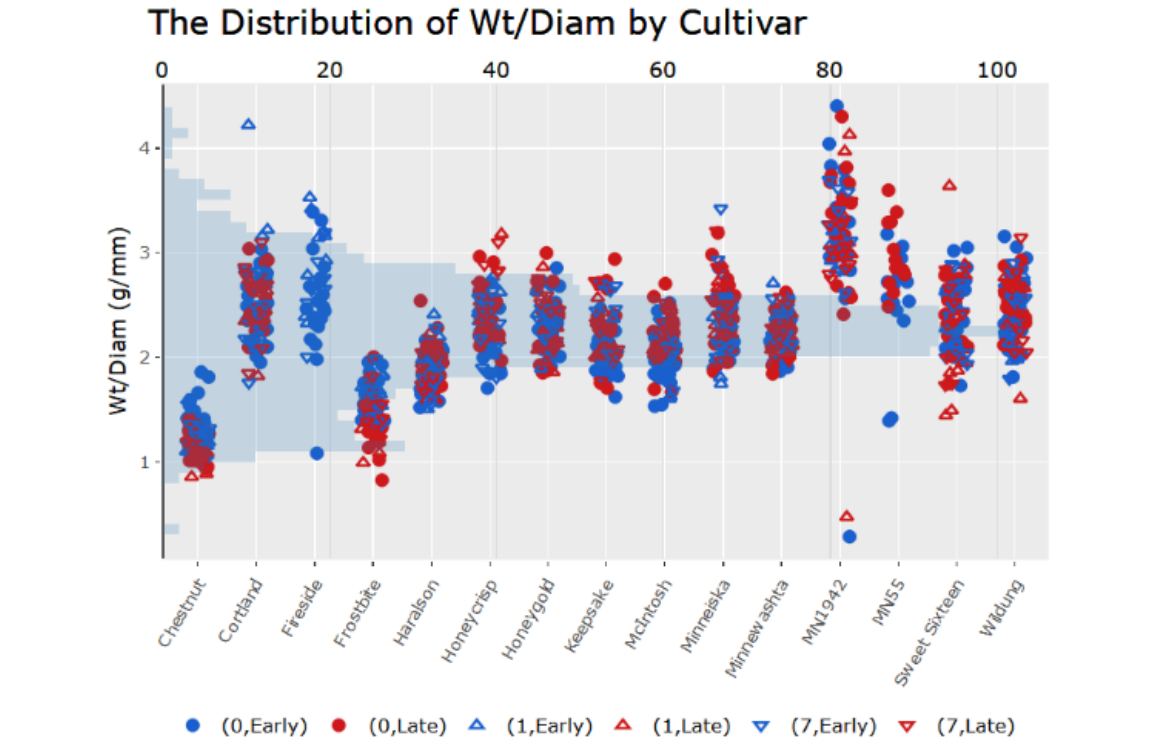

The Distribution of OMH by Cultivar
OMH (N)
0
20
40
60
80
100
120
140
0
50
100
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
Honeygold
Keepsake
McIntosh
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(0,Late)
(1,Early)
(1,Late)
(7,Early)
(7,Late)

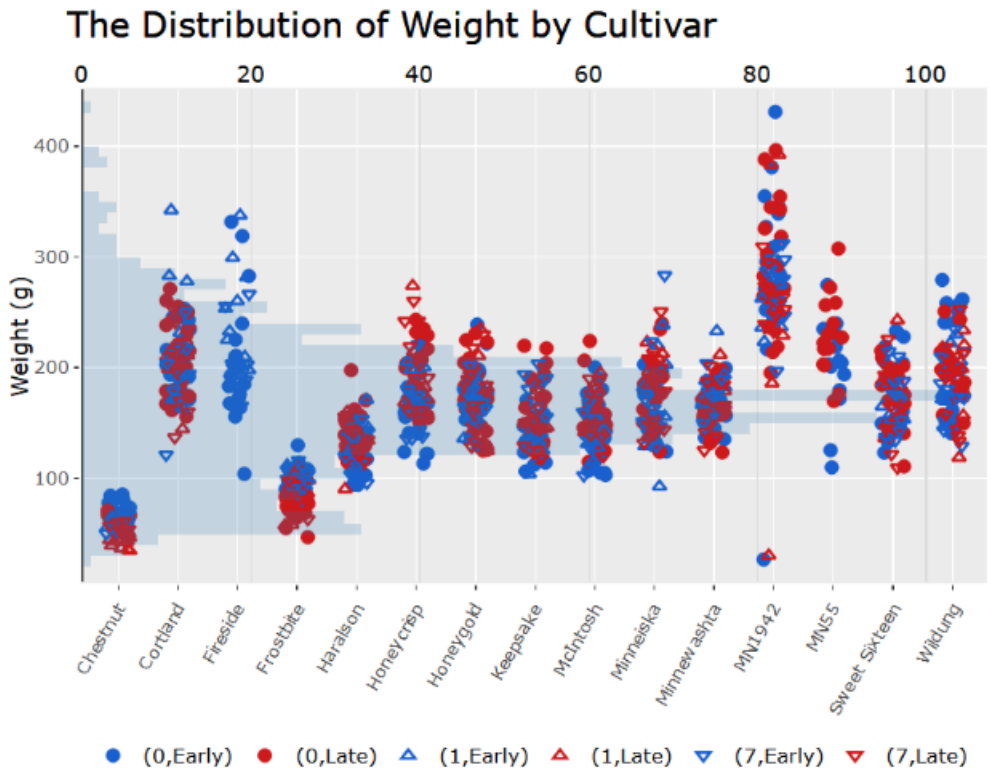

The Distribution of OAH by Cultivar
OAH (N)
0
20
40
60
80
100
0
25
50
75
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
Honeygold
Keepsake
McIntosh
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(0,Late)
(1,Early)
(1,Late)
(7,Early)
(7,Late)

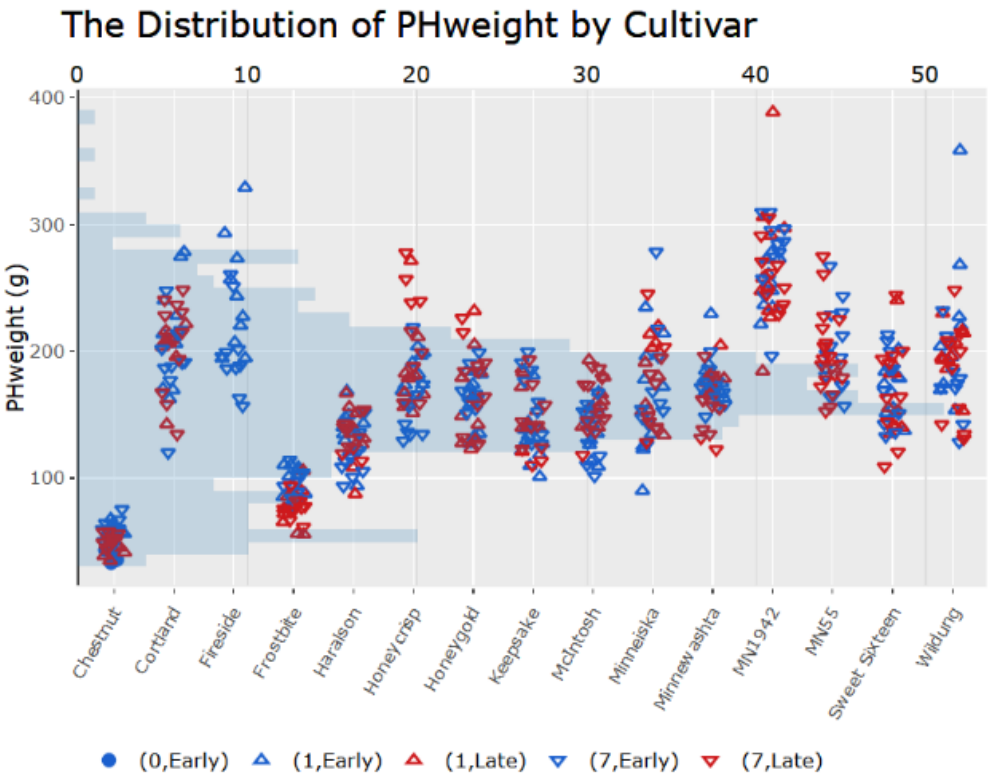
Diameter and Weight

Figure A- 4. Distribution plots for fruit weight and diameter






The Distribution of Weight by Cultivar
0
20
40
60
80
100
400
300
200
100
Weight (g)
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
Honeygold
Keepsake
McIntosh
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(0,Late)
(1,Early)
(1,Late)
(7,Early)
(7,Late)


The Distribution of PHweight by Cultivar
0
10
20
30
40
50
400
300
200
100
PHweight (g)
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
Honeygold
Keepsake
McIntosh
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(1,Early)
(1,Late)
(7,Early)
(7,Late)

Skin Traits

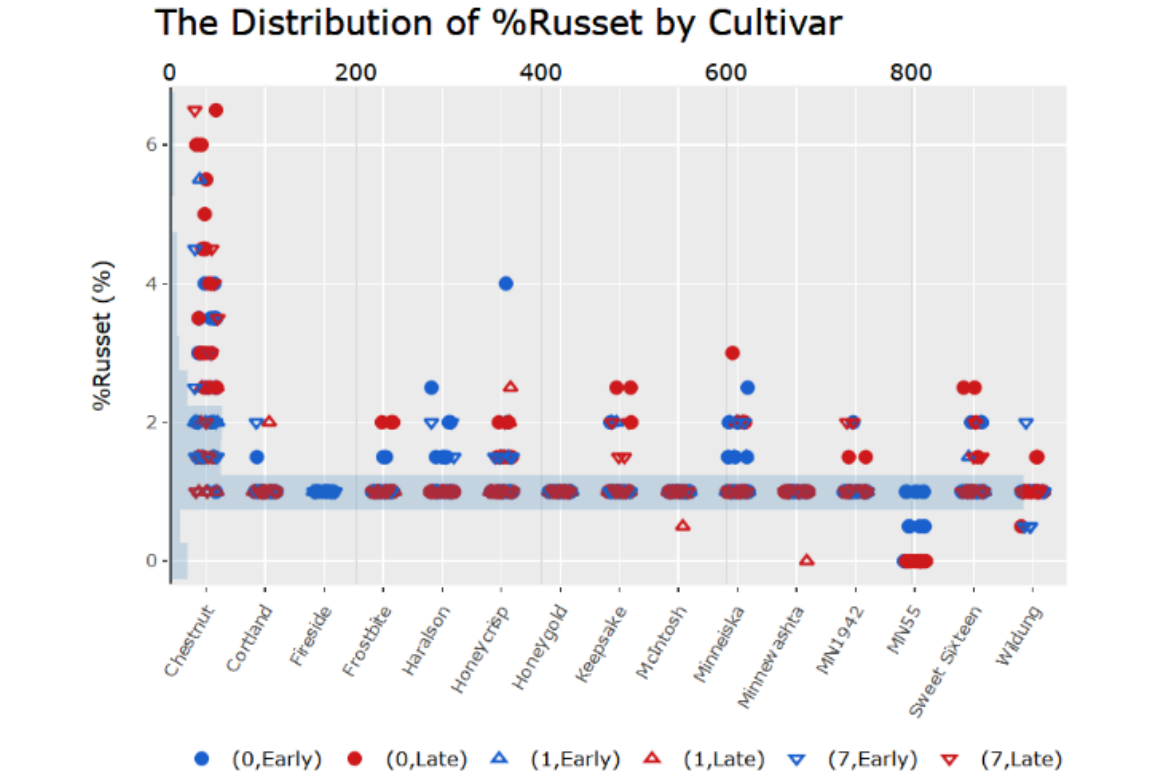
Russet

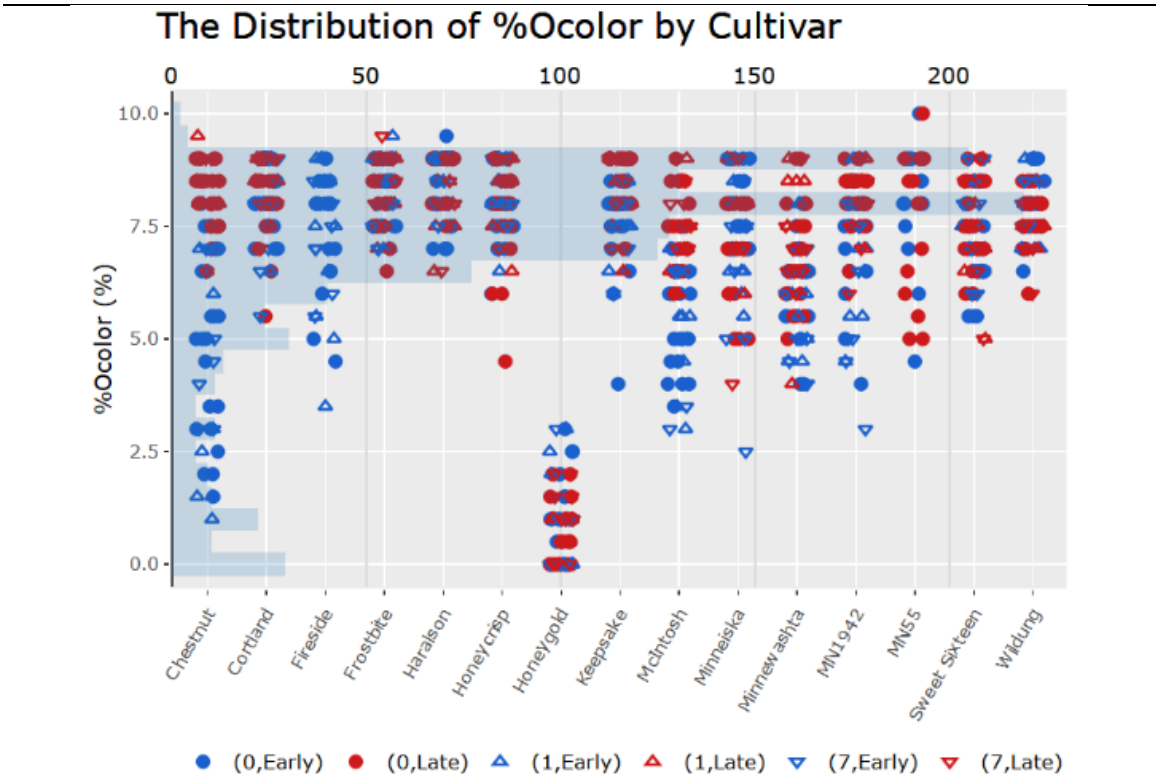
Six russet different scores were collected however only three were used to create models. Around 30% of the fruit did not have body russet (BRusset) present, while 66% of the fruit exhibited lenticel russeting (LRusset). Almost all fruit (82%) had a rating of 10% russet (%Russet), which rating was given if any type of russet occurred on the fruit.

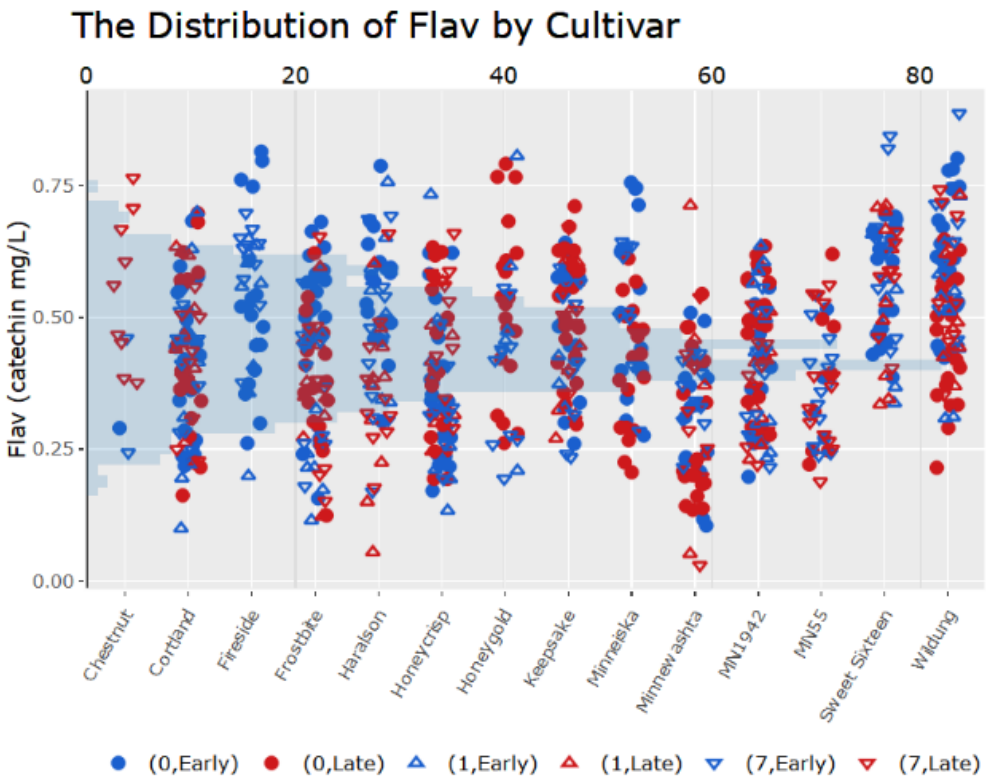
Color

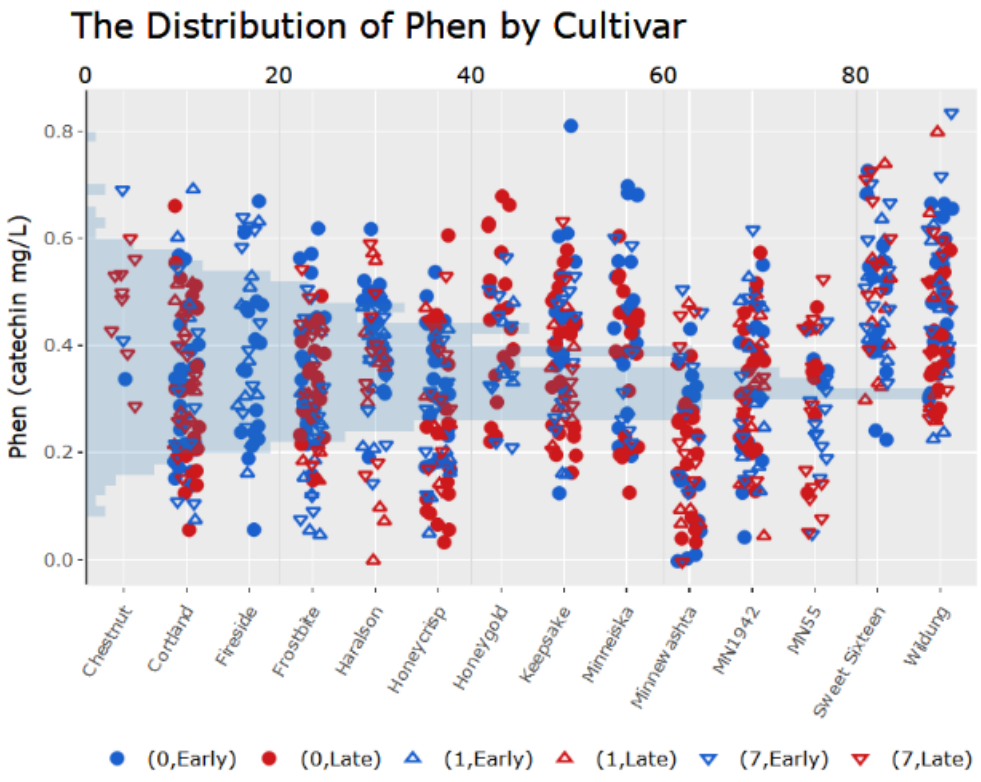
Our method of scoring fruit color is subjective and categorical with the intention to provide a general idea of patterns among cultivars. For ground-color (Gcolor) more fruit were in the categories yellow-green and yellow then other colors. Over-color (Ocolor) for most fruits was considered red as compared to dark/pink/orange red. Most fruit were categorized as having over-color types (Ocolor type) of blush-stripe or stripe-blush than the separate categories of stripe or blush alone. The cultivar Honeygold had fruit completely lacking anthocyanin Ocolor.

Figure A- 5. Distribution plots for skin traits





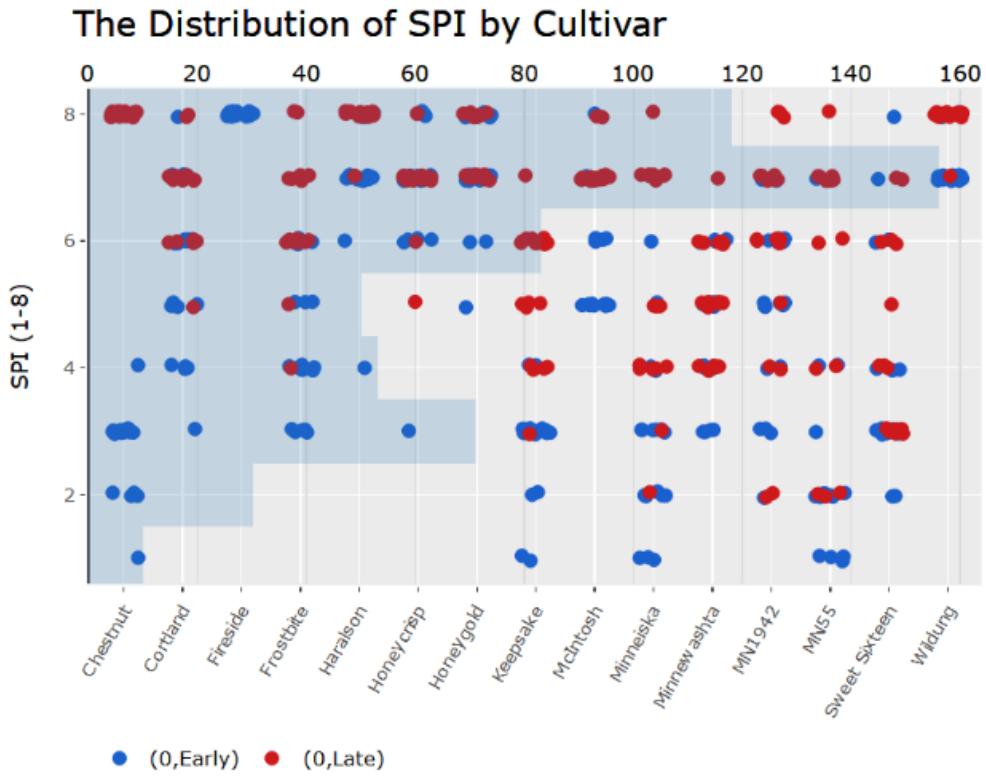

The Distribution of Flav by Cultivar
Flav (catechin mg/L)
0.00
0.25
0.50
0.75
0
20
40
60
80
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
Honeygold
Keepsake
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(0,Late)
(1,Early)
(1,Late)
(7,Early)
(7,Late)

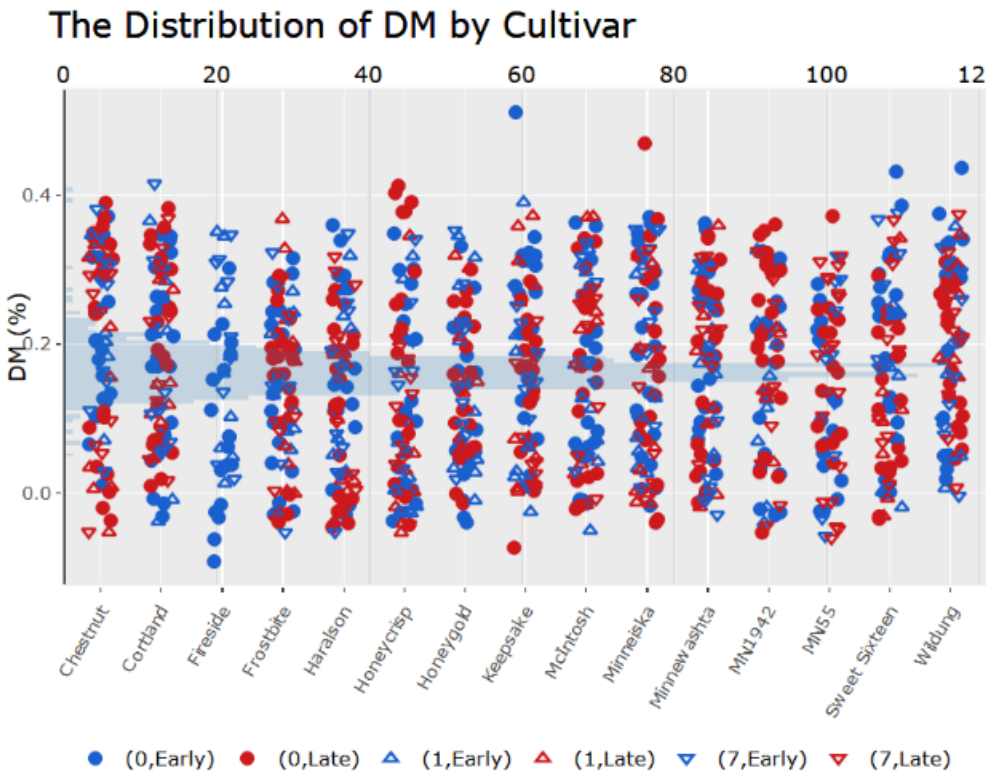

The Distribution of Phen by Cultivar
Phen (catechin mg/L)
0.0
0.2
0.4
0.6
0.8
0
20
40
60
80
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
Honeygold
Keepsake
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(0,Late)
(1,Early)
(1,Late)
(7,Early)
(7,Late)

Maturity and Dry Matter

The starch pattern index (SPI) was only measured on fruit destructively processed at-harvest time. Despite the attempt to obtain an equal number of fruit in the 3-4 range as the 6-7 range, ~62.5% of observations fell in the 8-6 SPI range, with ~30.5% between 5-3.

Figure A- 6. Distribution plot for SPI and DM

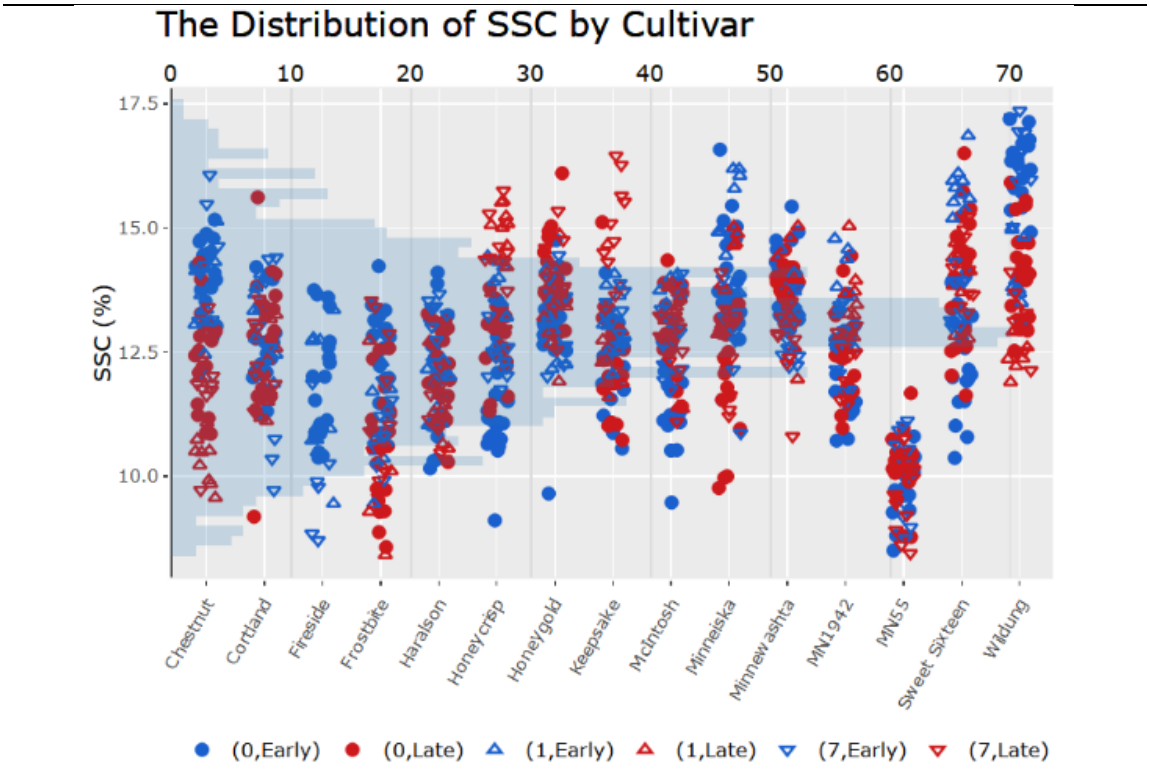


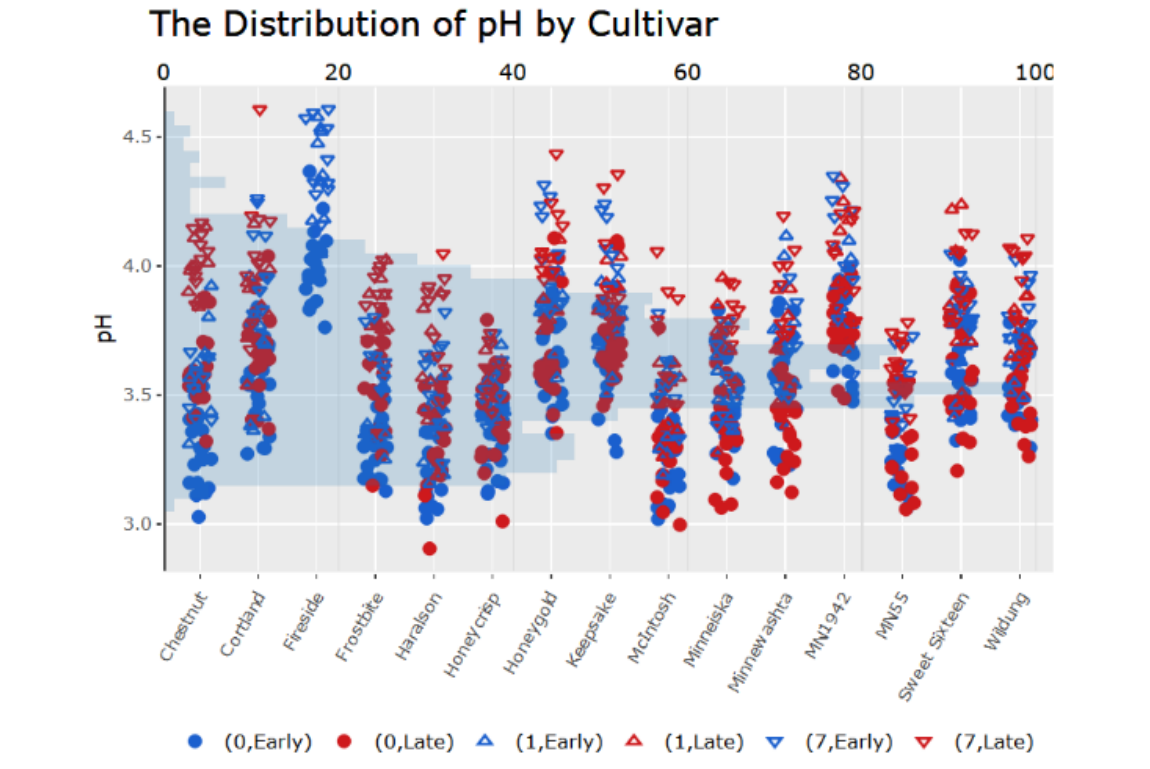


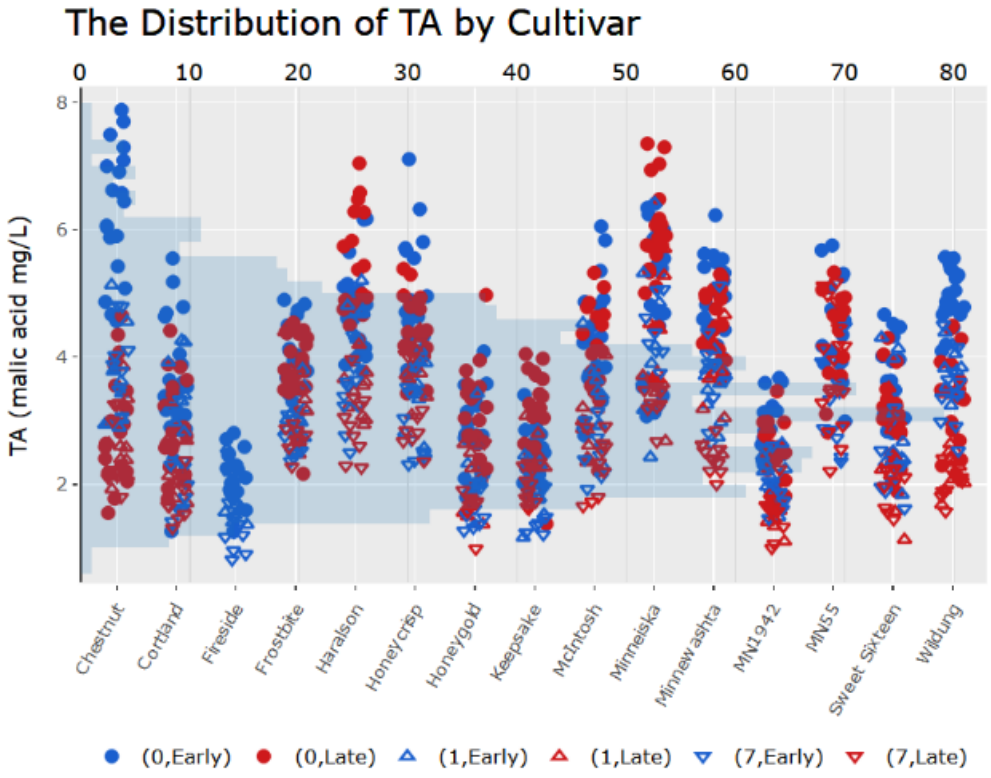
Fruit juice traits

The SSC trait has been the focus of most NIRS studies involving fruit. Observations in this dataset approximate a normal distribution. Patterns in this dataset are dependent on cultivar. For example, the late harvests of some cultivars (e.g. Chestnut and Snowsweet) have lower SSC values when compared to the earlier harvest, but that trend switches (e.g. Honeygold and Honeycrisp) among cultivars or there is no pattern (e.g. MN1942 and MN55).

Figure A- 7. Distribution plots for fruit juice related traits

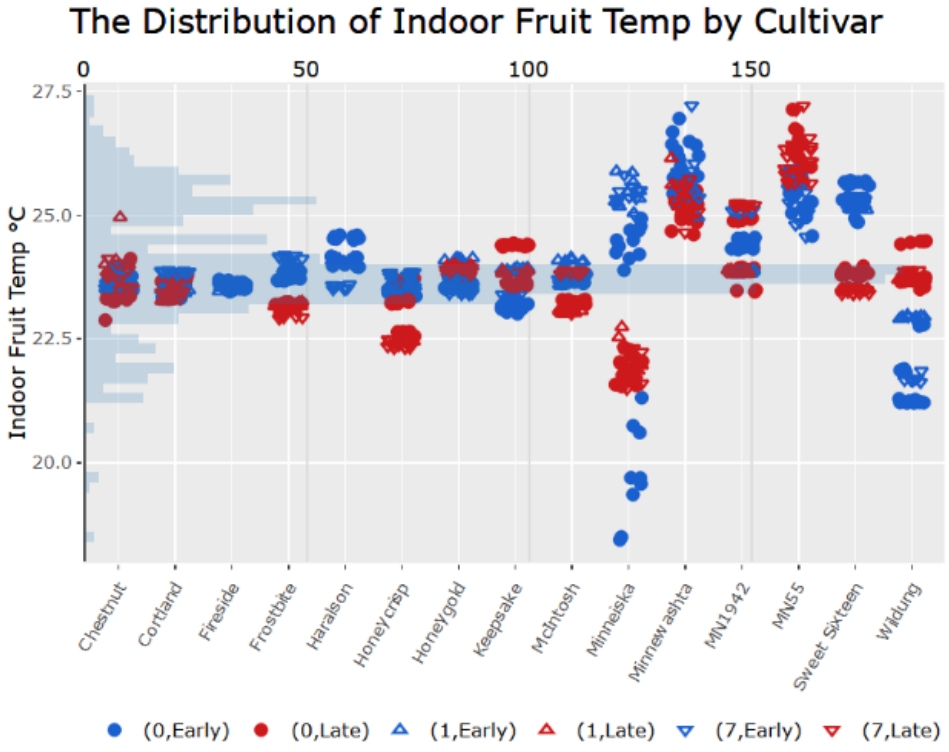


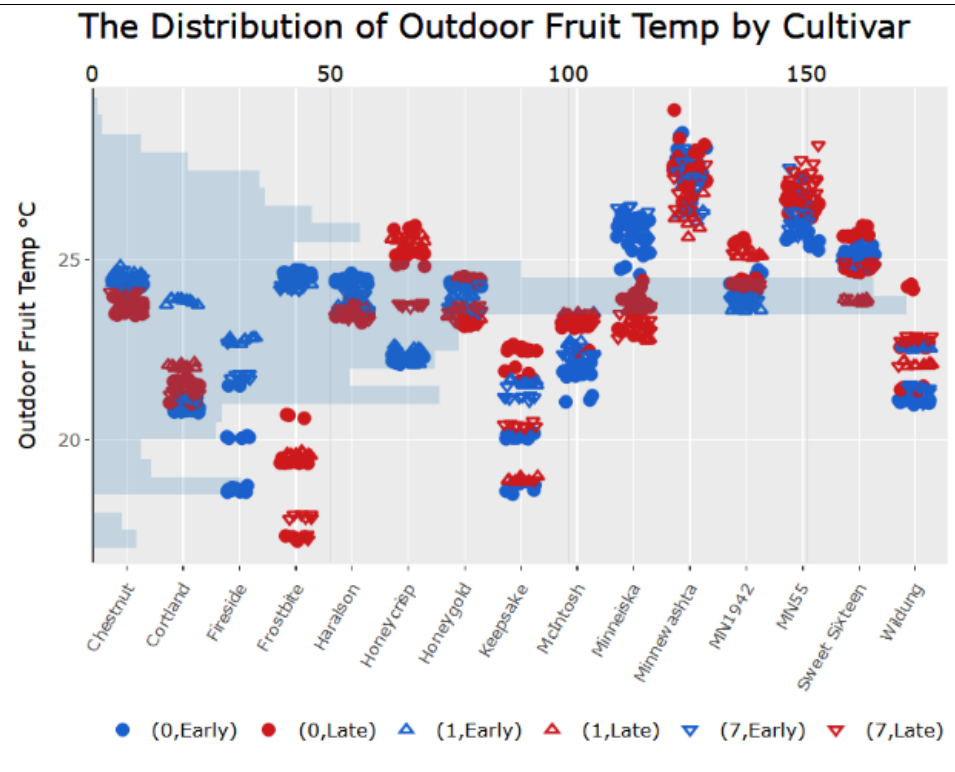


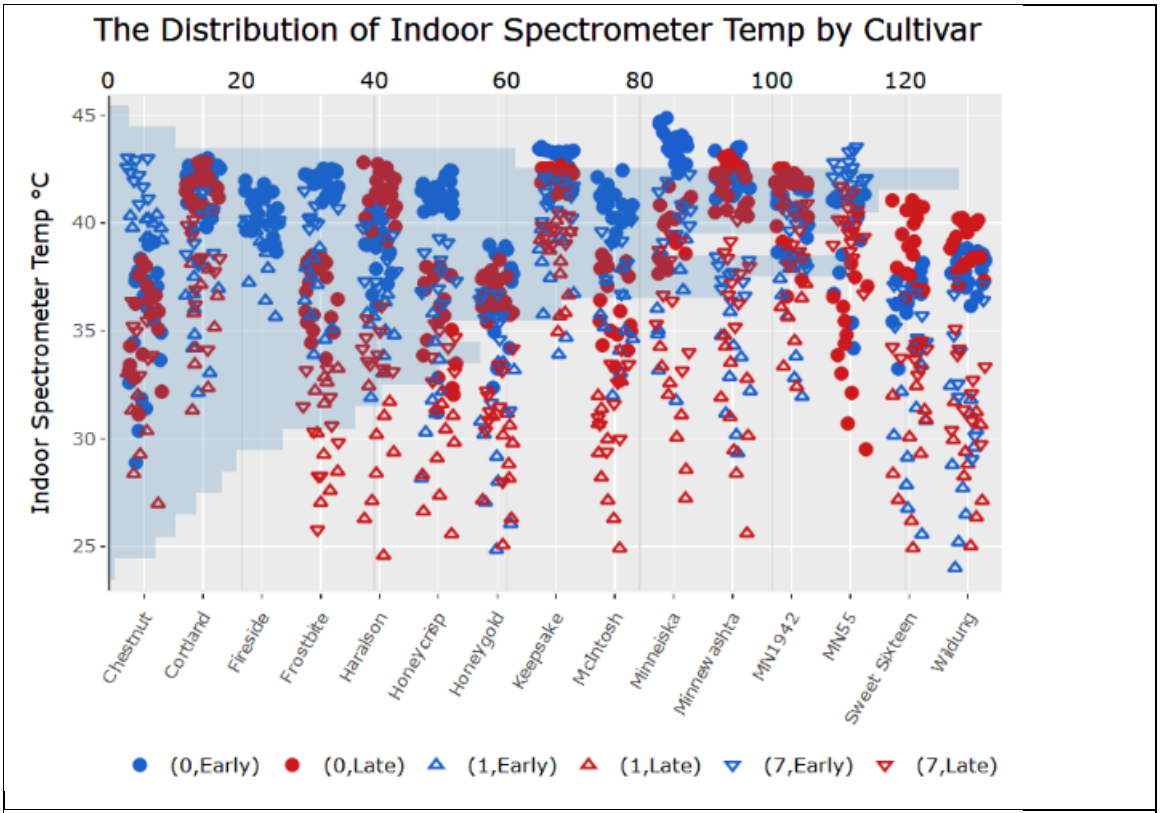

The Distribution of TA by Cultivar
TA (malic acid mg/L)
0
10
20
30
40
50
60
70
80
2
4
6
8
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
Honeygold
Keepsake
McIntosh
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(0,Late)
(1,Early)
(1,Late)
(7,Early)
(7,Late)

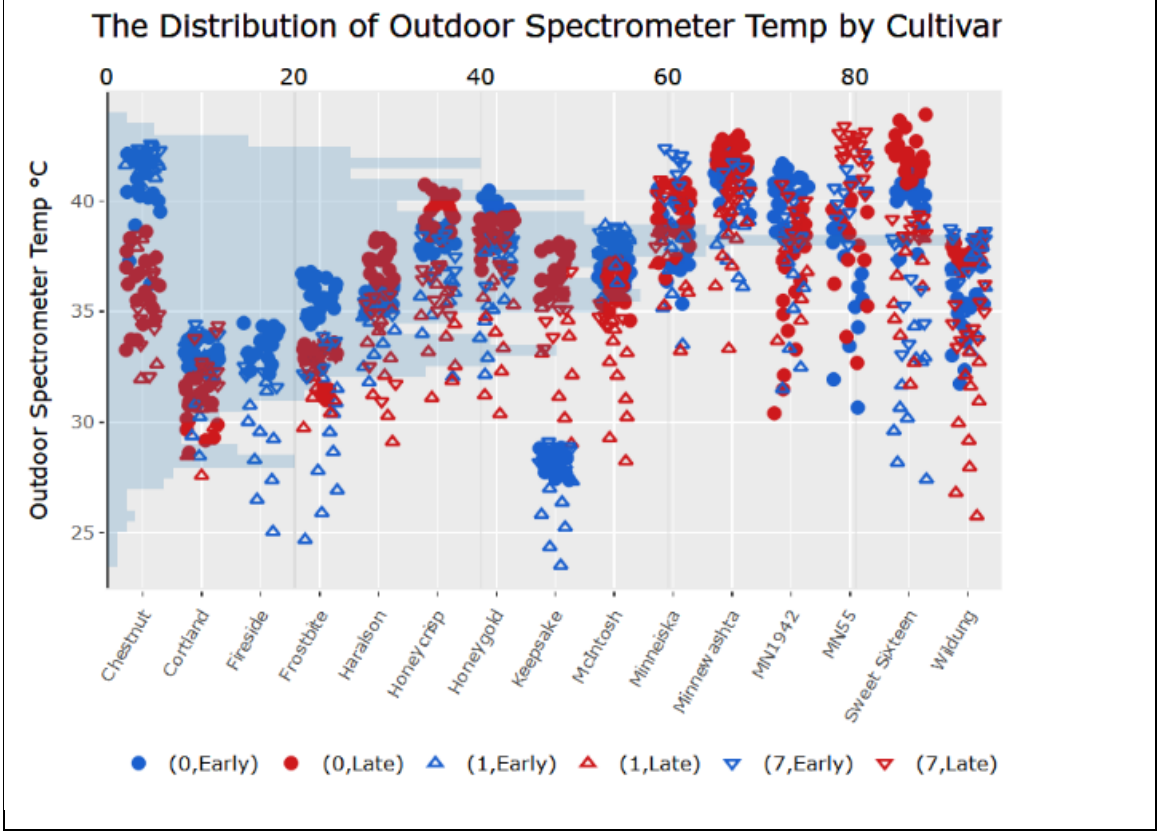
Temperature Variables

Figure A- 8. Plots corresponding to outdoor and indoor: fruit surface temperature and spectrometer temperature (°C)



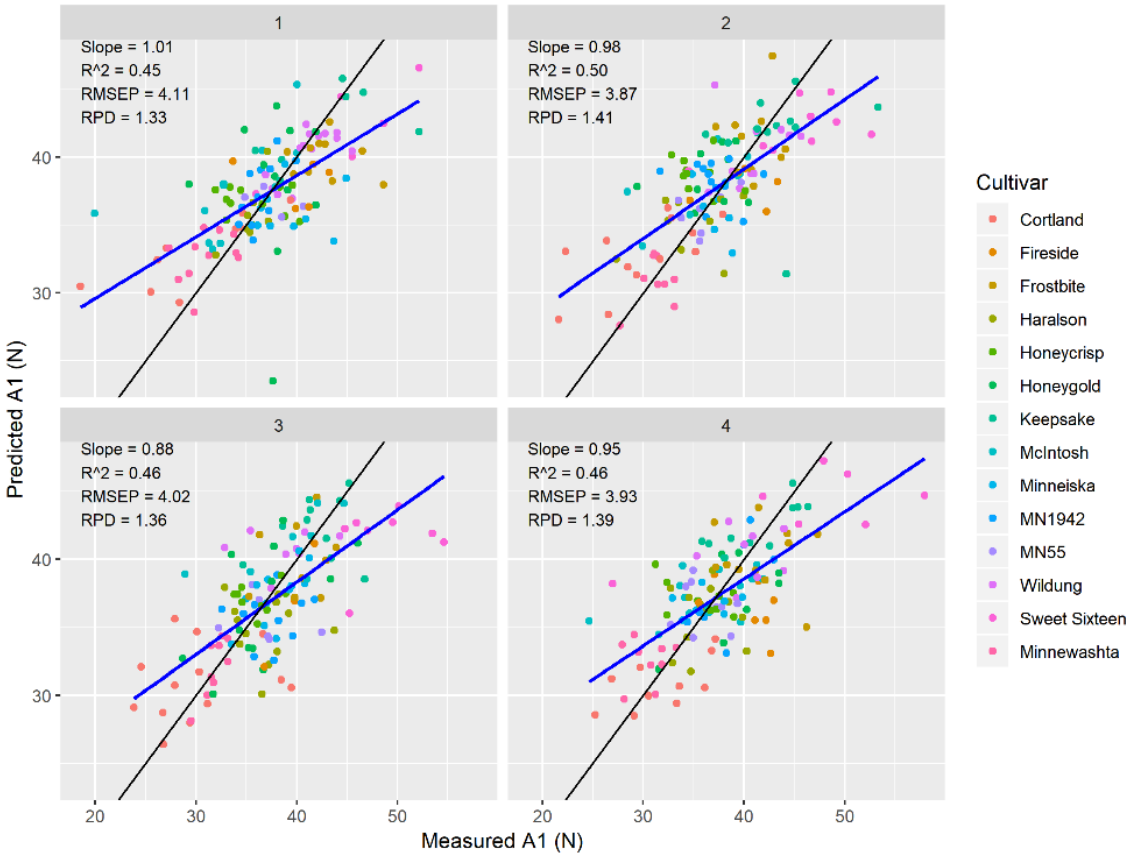



The Distribution of Indoor Spectrometer Temp by Cultivar
Indoor Spectrometer Temp °C
0
20
40
60
80
100
120
45
40
35
30
25
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
HoneyGold
Keepsake
McIntosh
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(0,Late)
(1,Early)
(1,Late)
(7,Early)
(7,Late)


The Distribution of Outdoor Spectrometer Temp by Cultivar
Outdoor Spectrometer Temp °C
0
20
40
60
80
40
35
30
25
Chestnut
Cortland
Fireside
Frostbite
Haralson
Honeycrisp
HoneyGold
Keepsake
McIntosh
Minneiska
Minnewashta
MN1942
MN55
Sweet Sixteen
Wildung
(0,Early)
(0,Late)
(1,Early)
(1,Late)
(7,Early)
(7,Late)

A.2 Prediction Plots

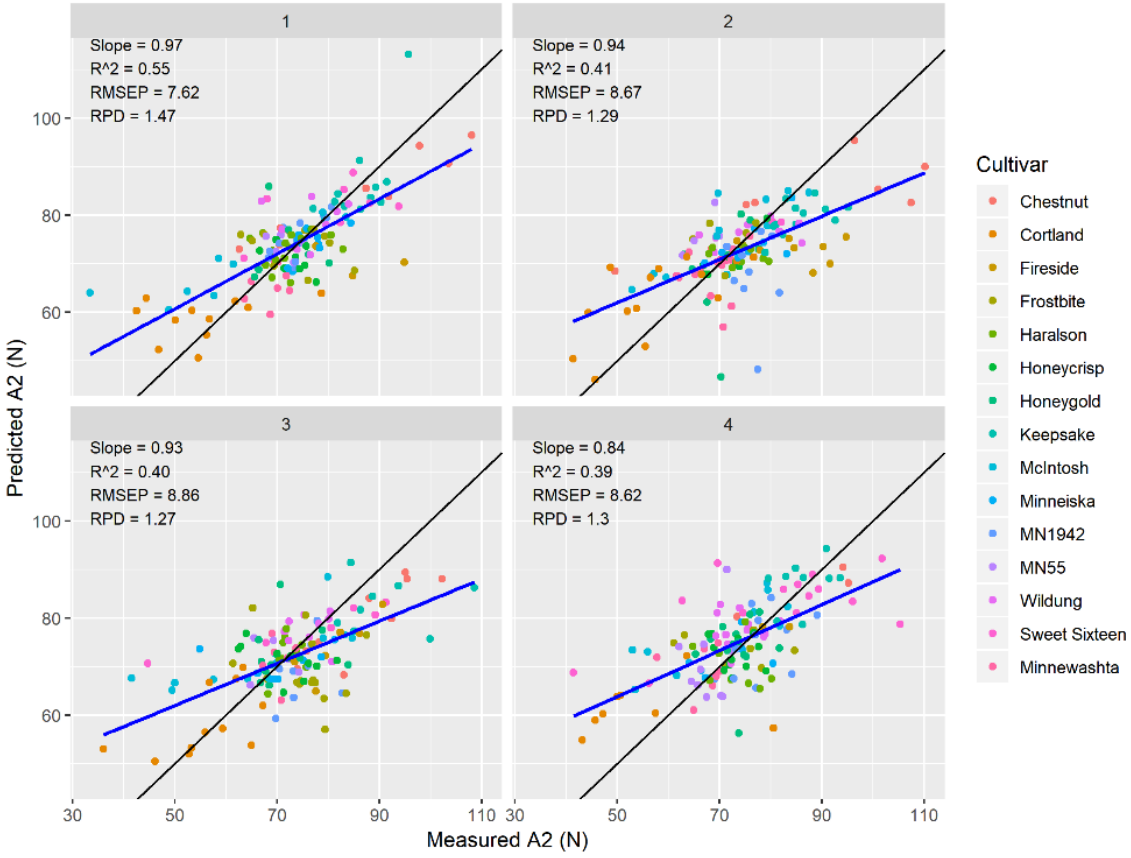
Figure A- 9. Indoors average spectra at-harvest prediction plots for the trait A1



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

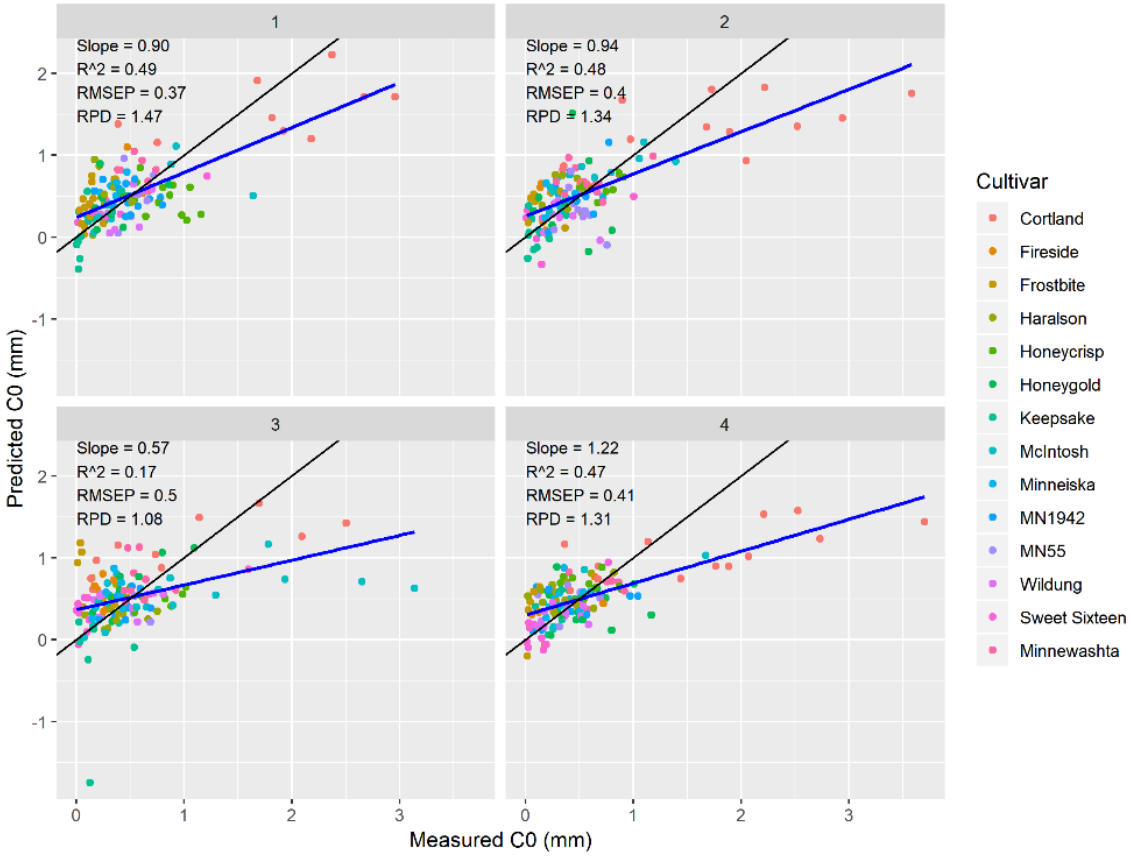
_

Figure A- 10. Indoors average spectra at-harvest prediction plots for the trait A2



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

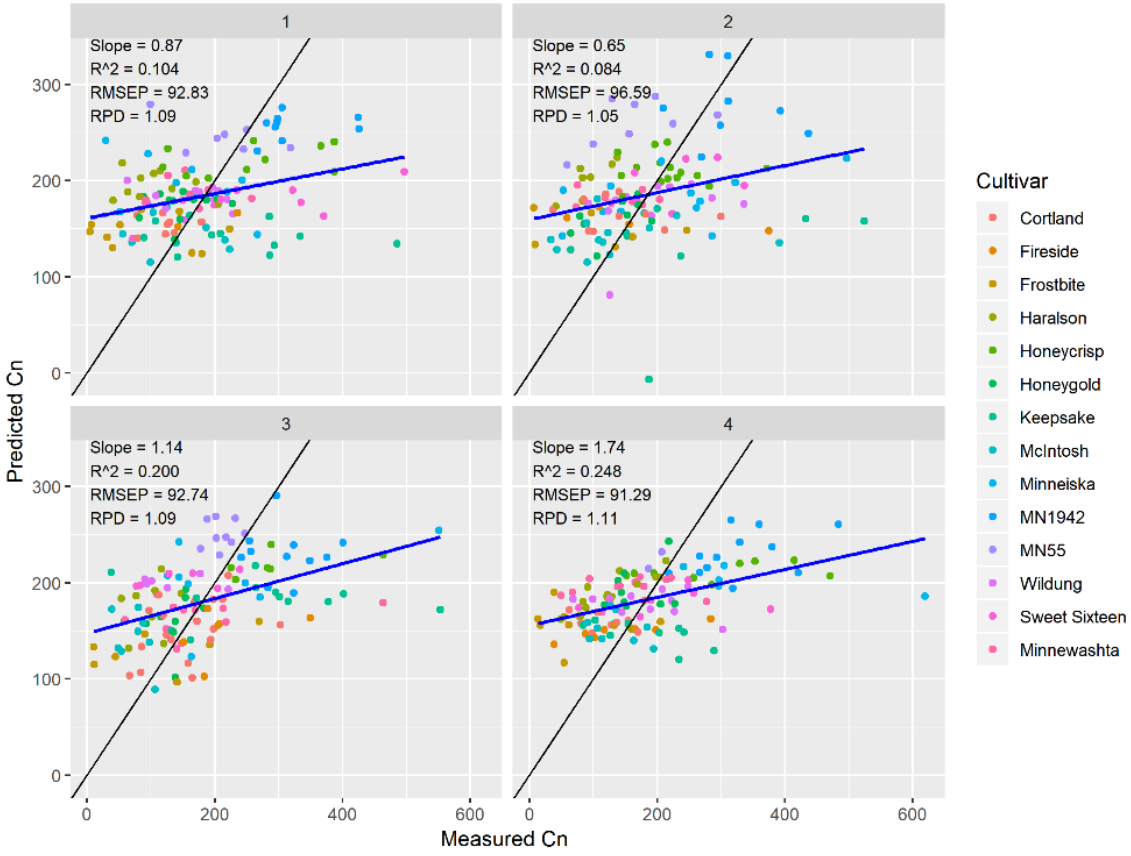
Figure A- 11. Indoors average spectra at-harvest prediction plots for the trait Creep Distance



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

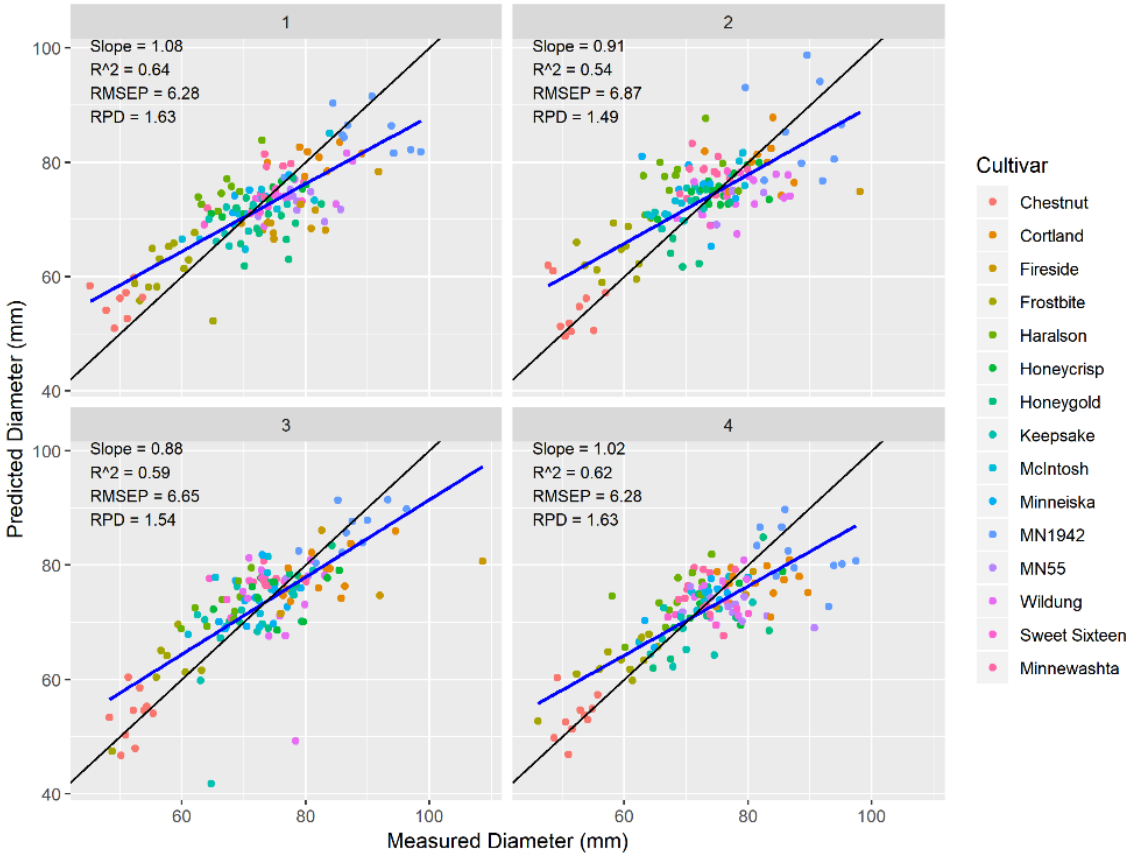
_

Figure A- 12. Indoors average spectra at-harvest prediction plots for the trait Crispness



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

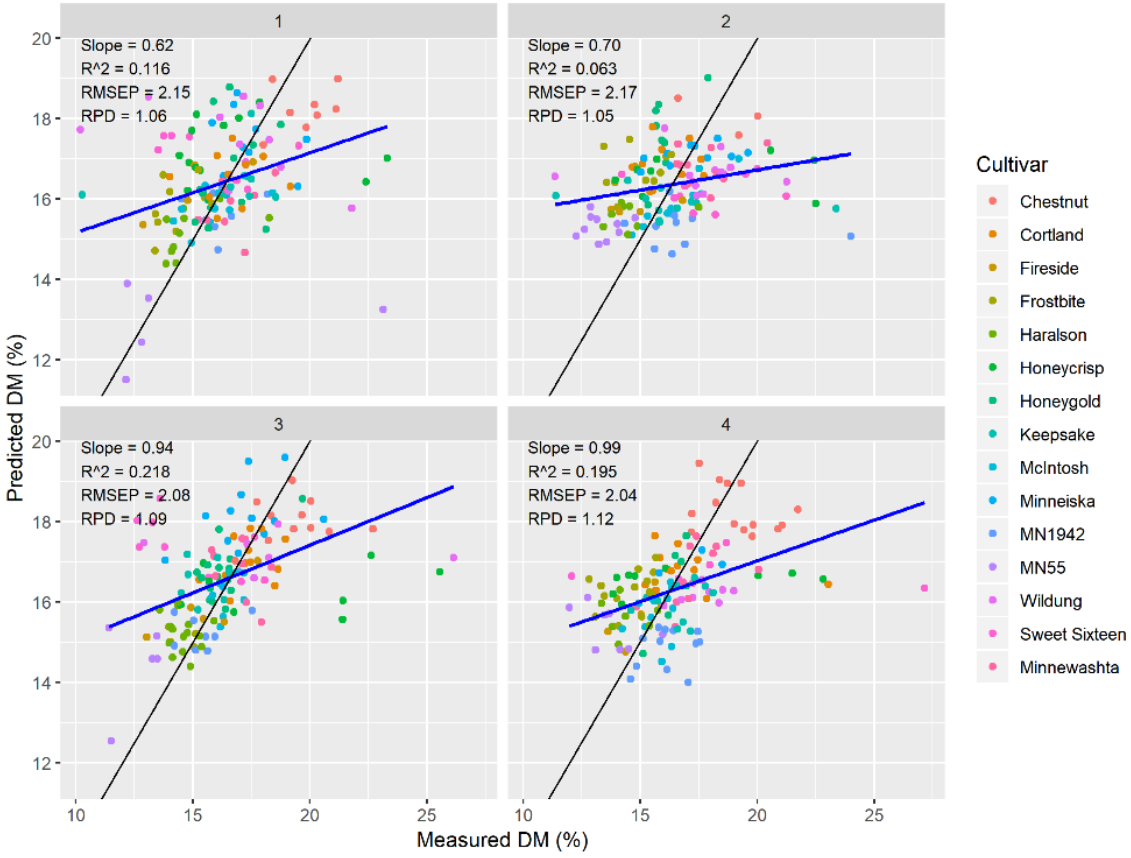
Figure A- 13. Indoors average spectra at-harvest prediction plots for the trait Diameter



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

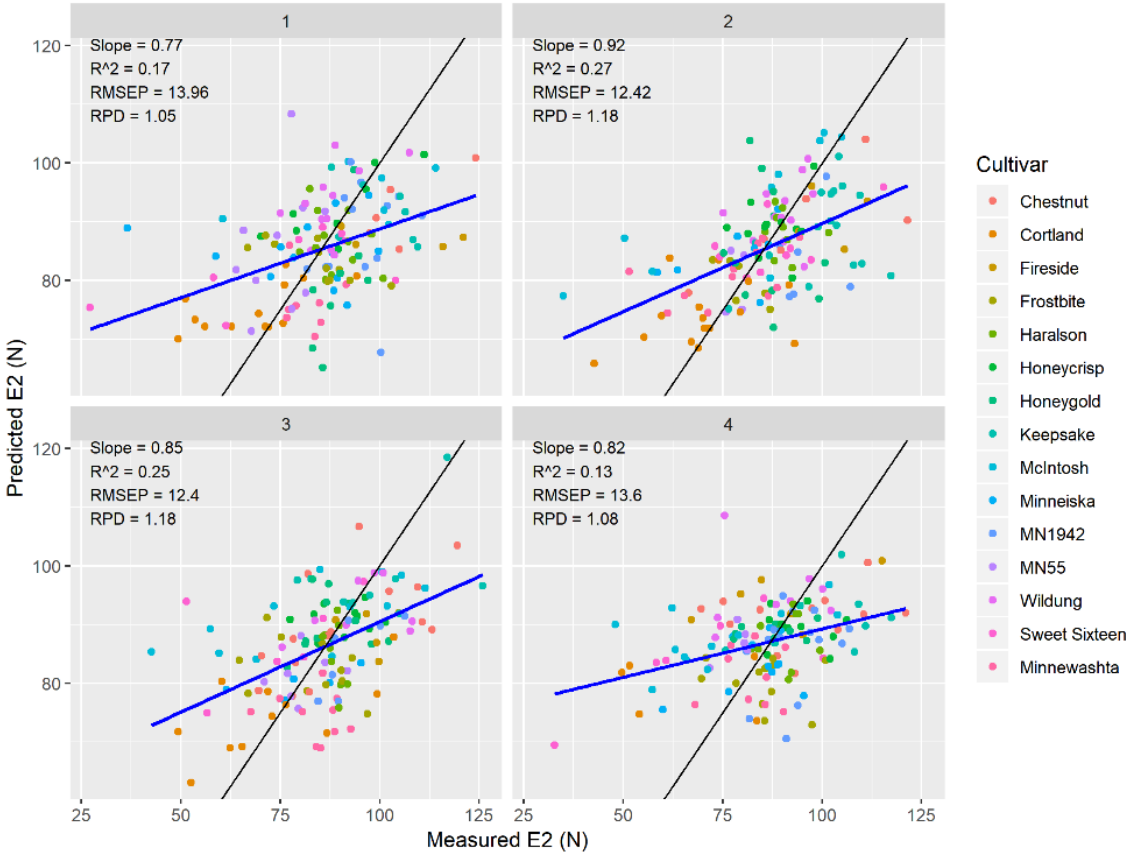
_

Figure A- 14. Indoors average spectra at-harvest prediction plots for the trait Dry Matter



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

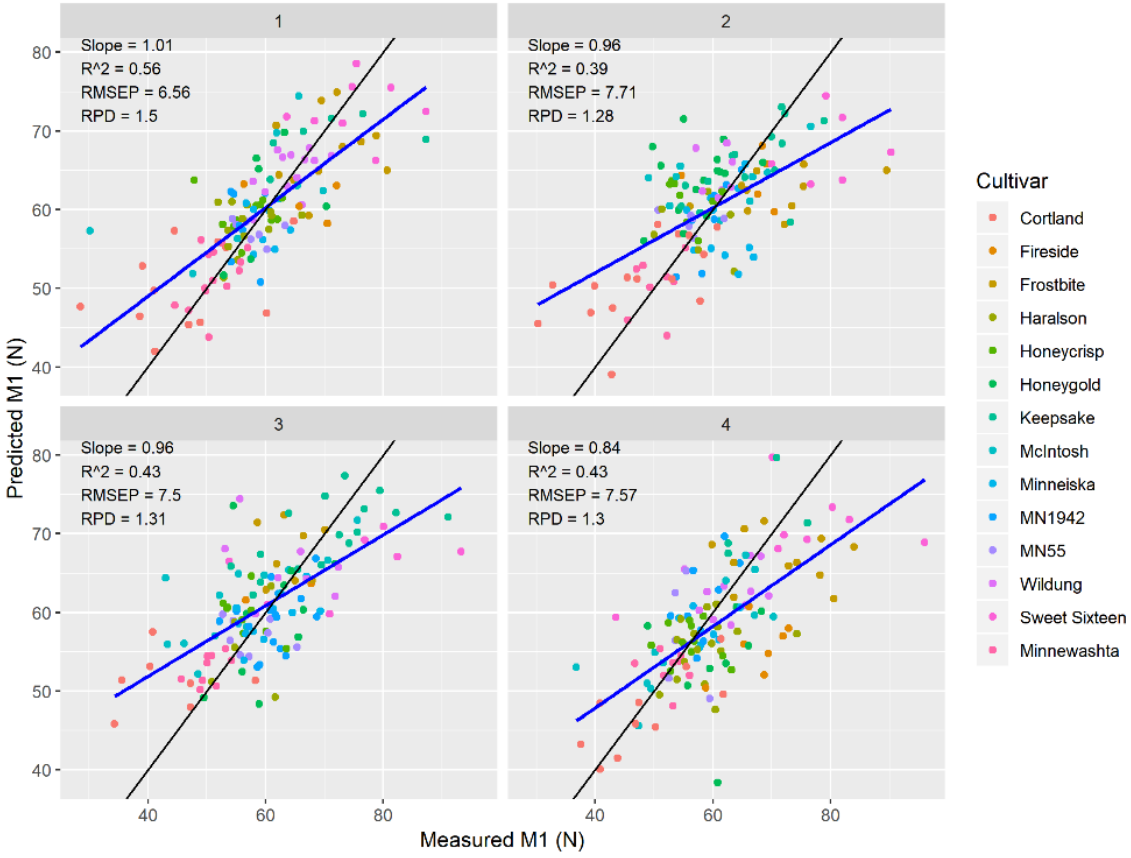
Figure A- 15. Indoors average spectra at-harvest prediction plots for the trait E2



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

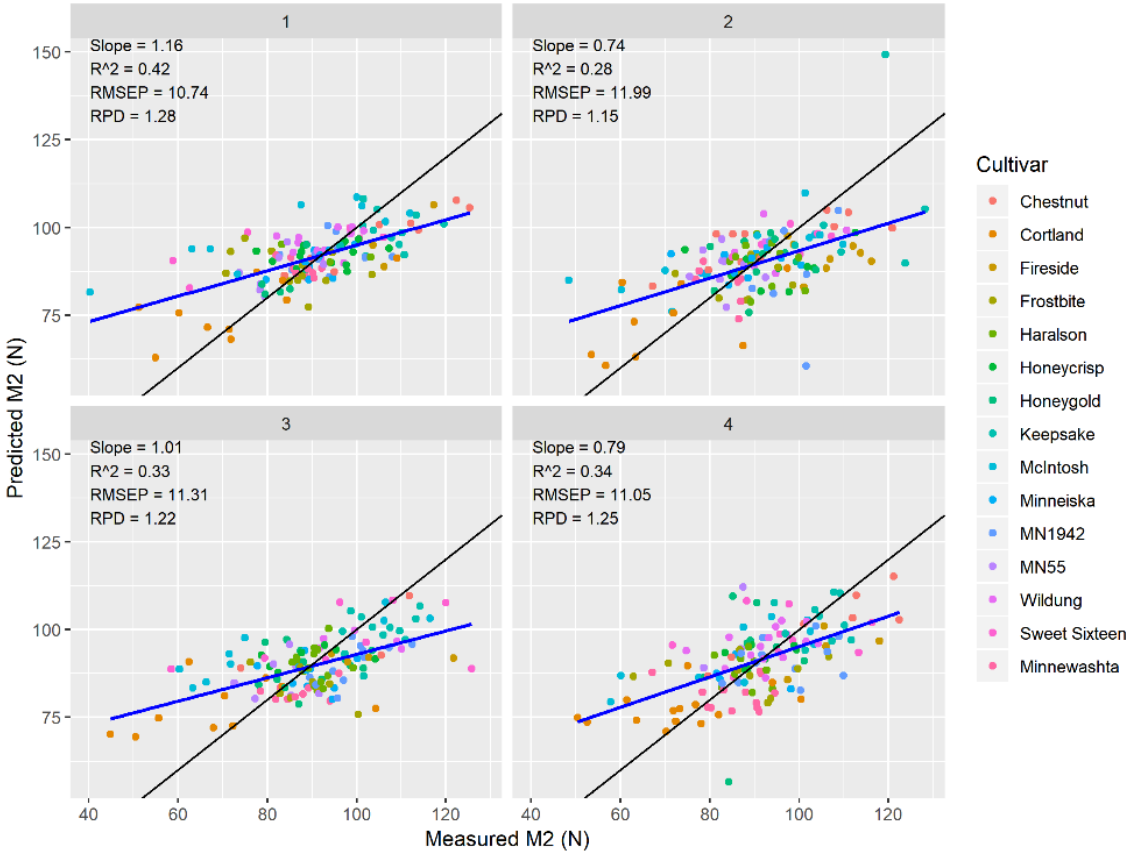
_

Figure A- 16. Indoors average spectra at-harvest prediction plots for the trait M1



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

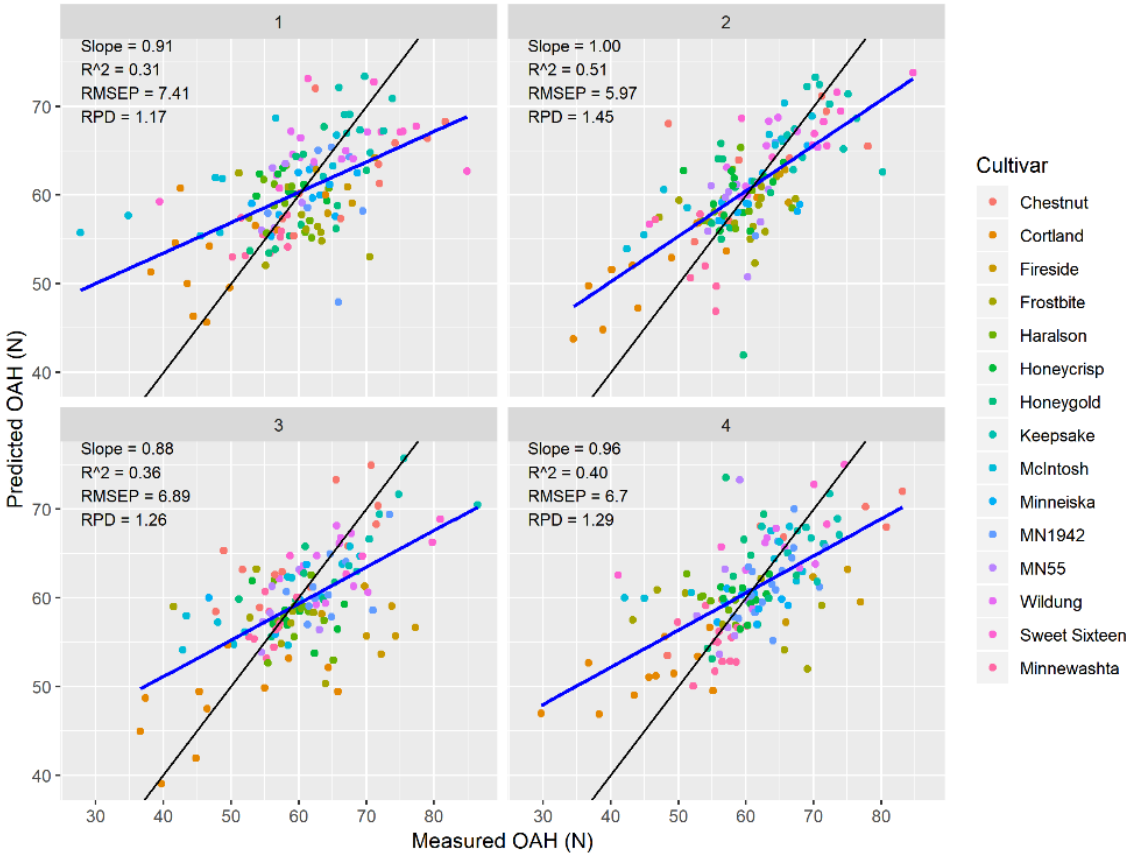
Figure A- 17. Indoors average spectra at-harvest prediction plots for the trait M2



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

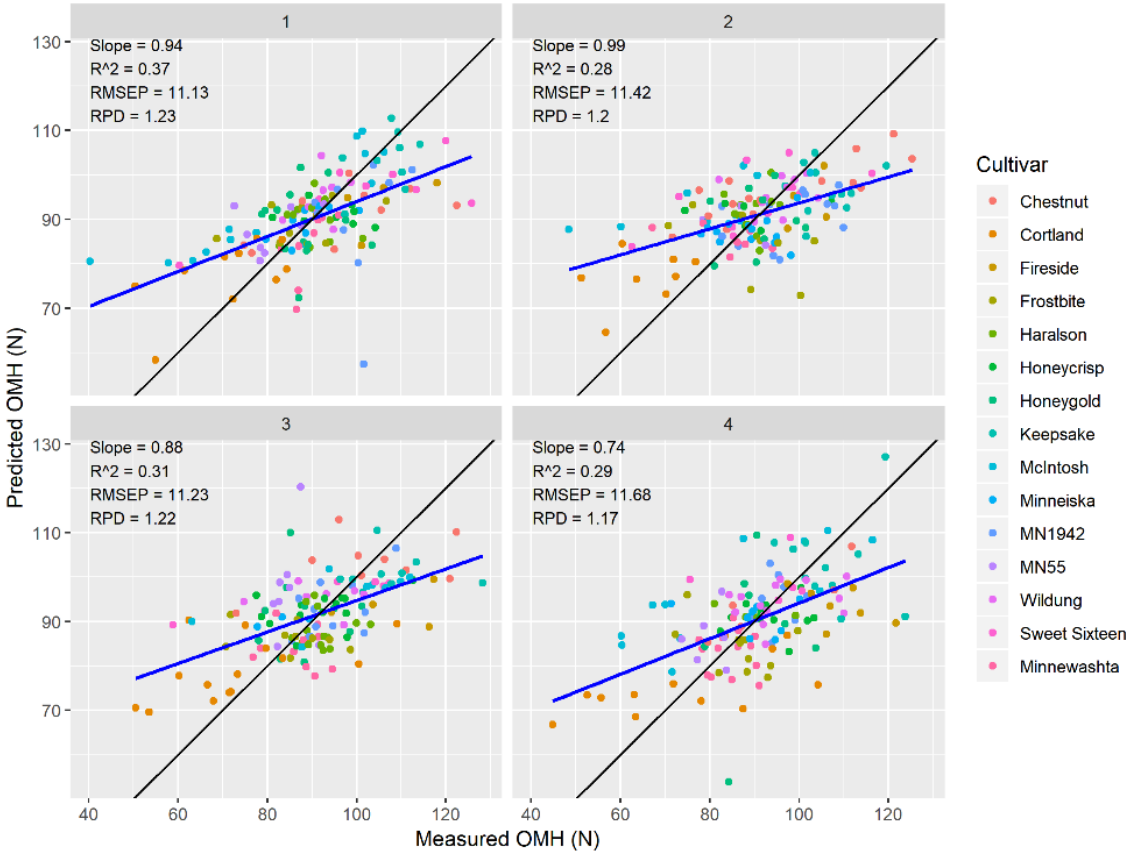
_

Figure A- 18. Indoors average spectra at-harvest prediction plots for the trait OAH



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

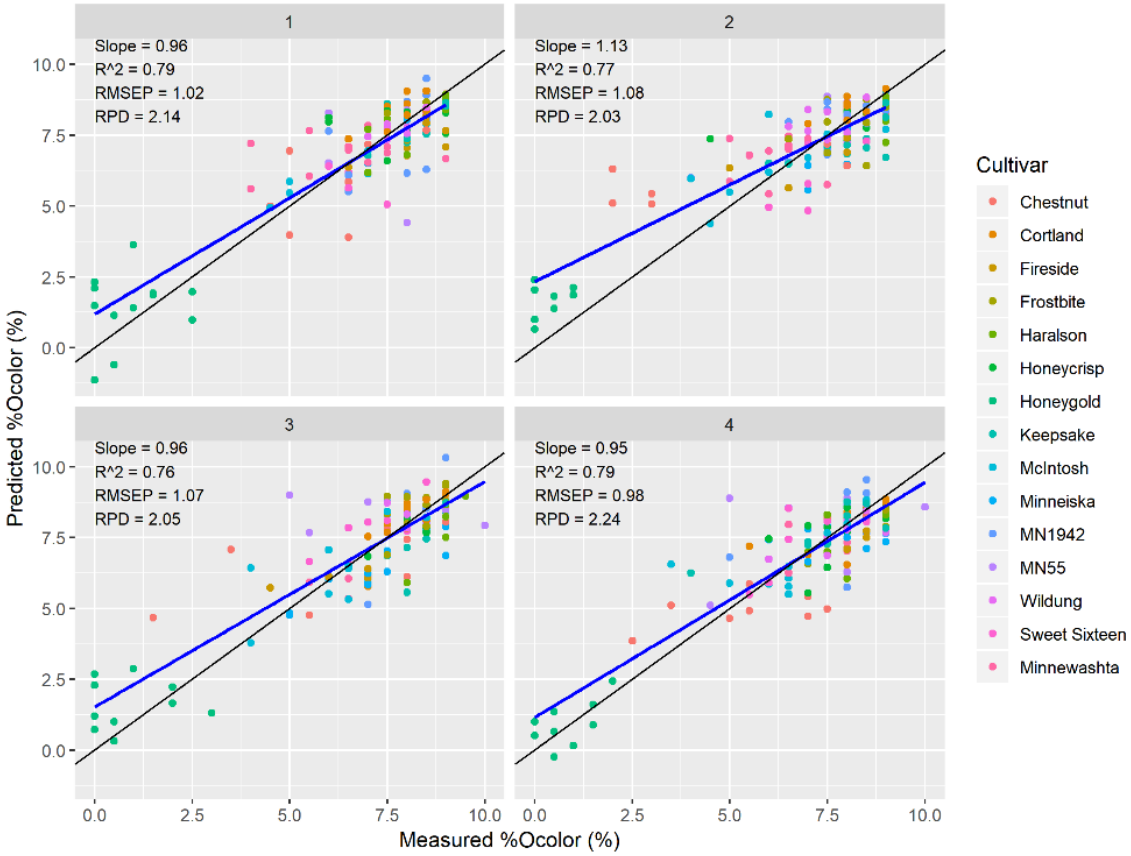
Figure A- 19. Indoors average spectra at-harvest prediction plots for the trait OMH



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

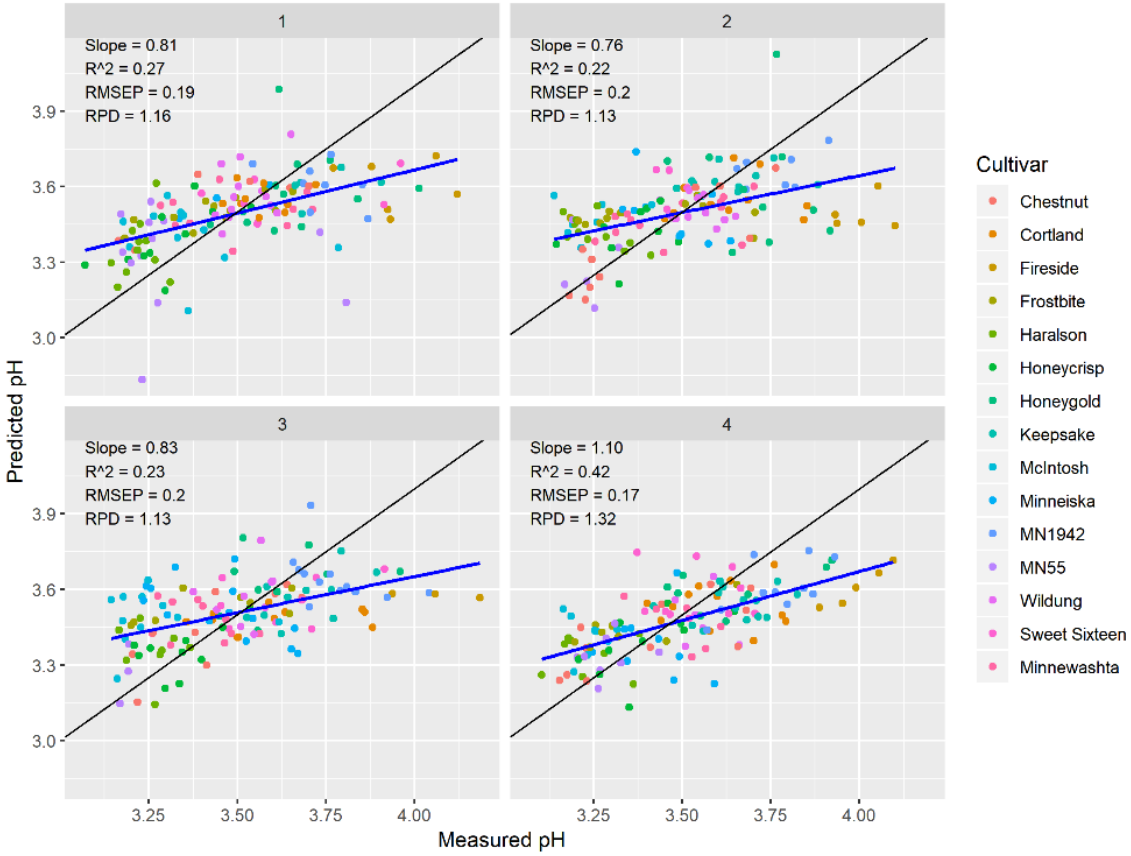
_

Figure A- 20. Indoors average spectra at-harvest prediction plots for the trait %Ocolor



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

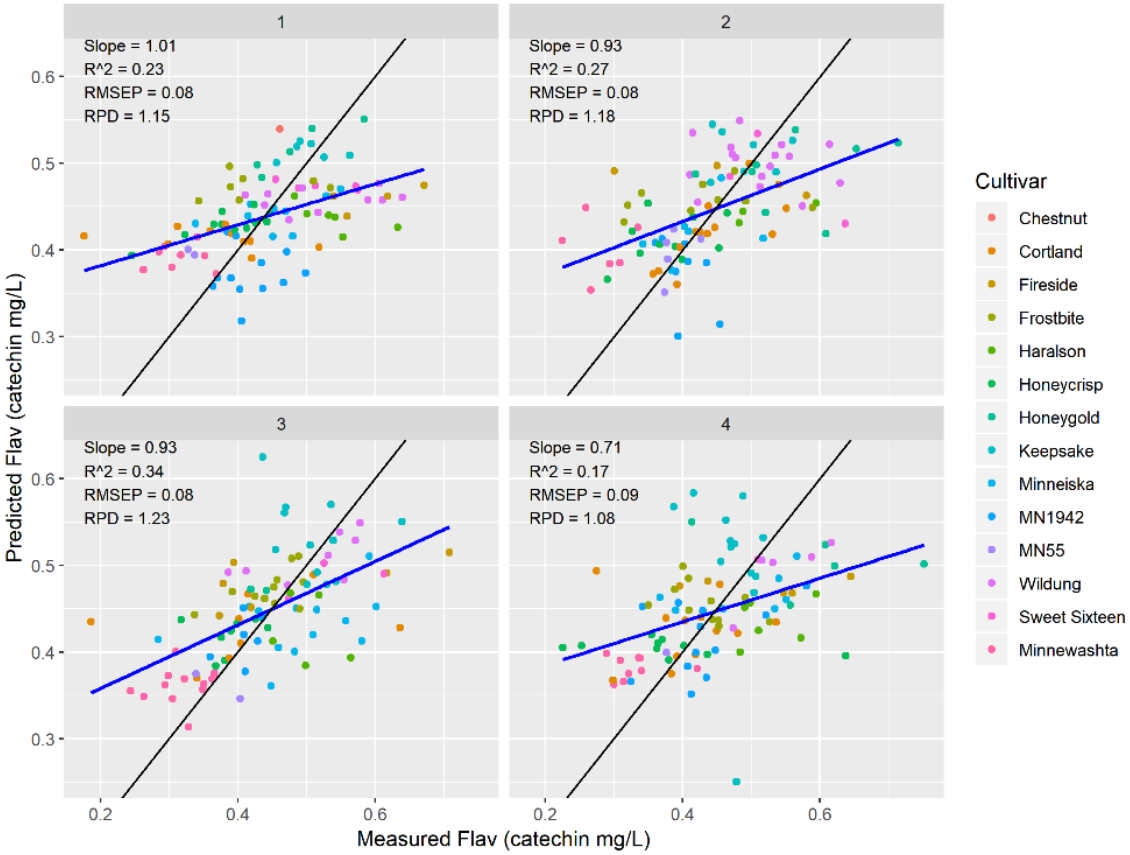
Figure A- 21. Indoors average spectra at-harvest prediction plots for the trait pH



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

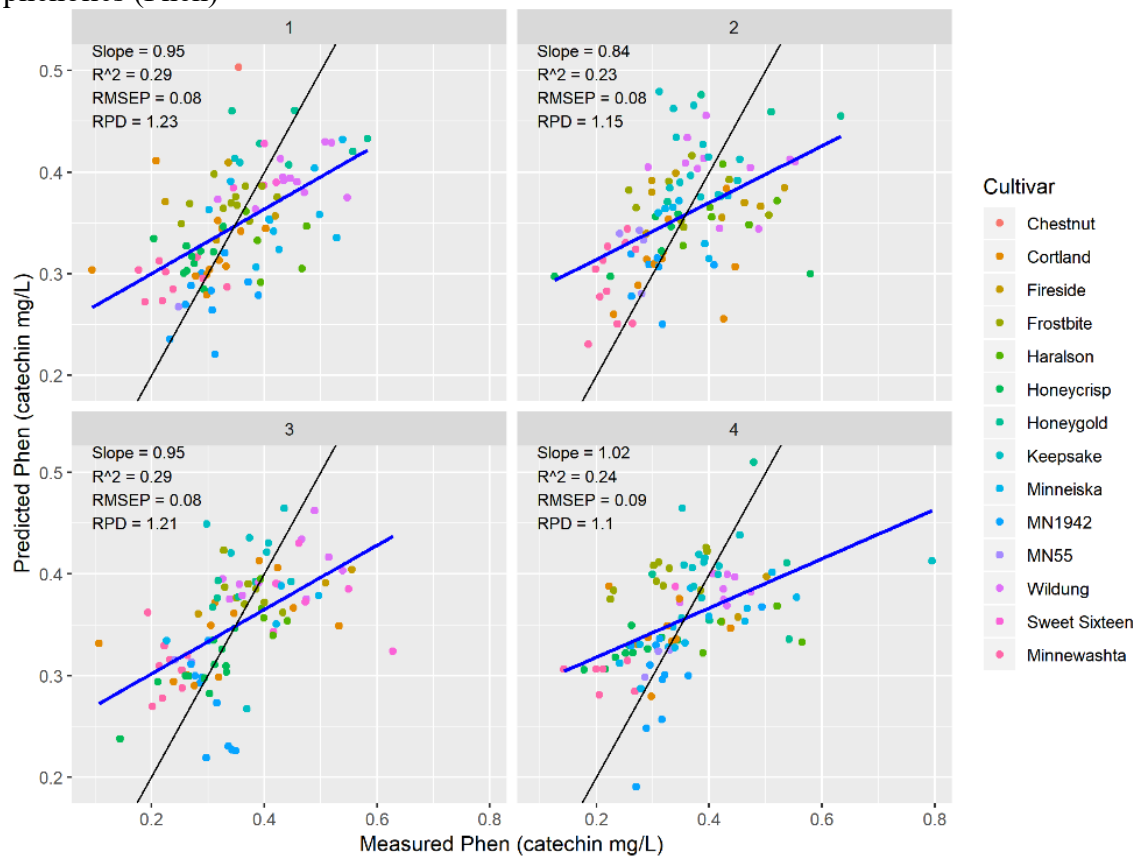
_

Figure A- 22. Indoors average spectra at-harvest prediction plots for the trait Flavonoids (Flav)



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

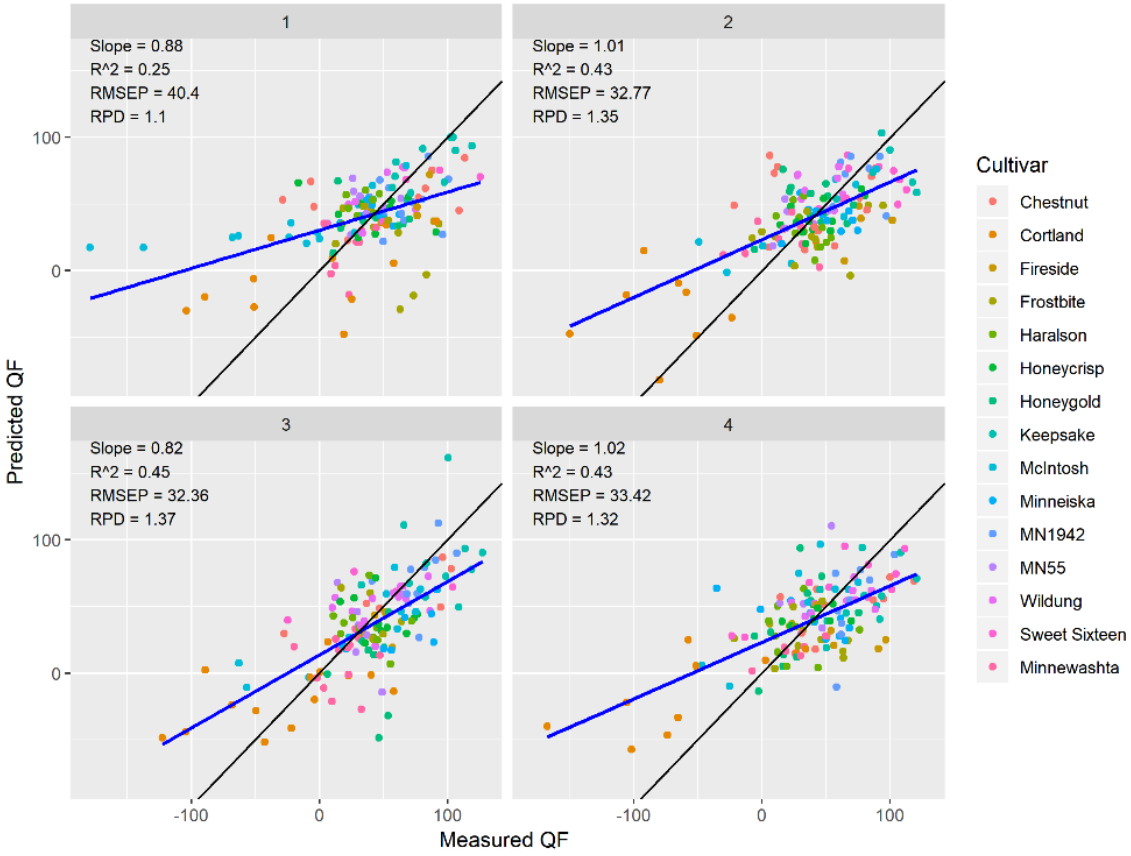
Figure A- 23. Indoors average spectra at-harvest prediction plots for the trait total phenolics (Phen)



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

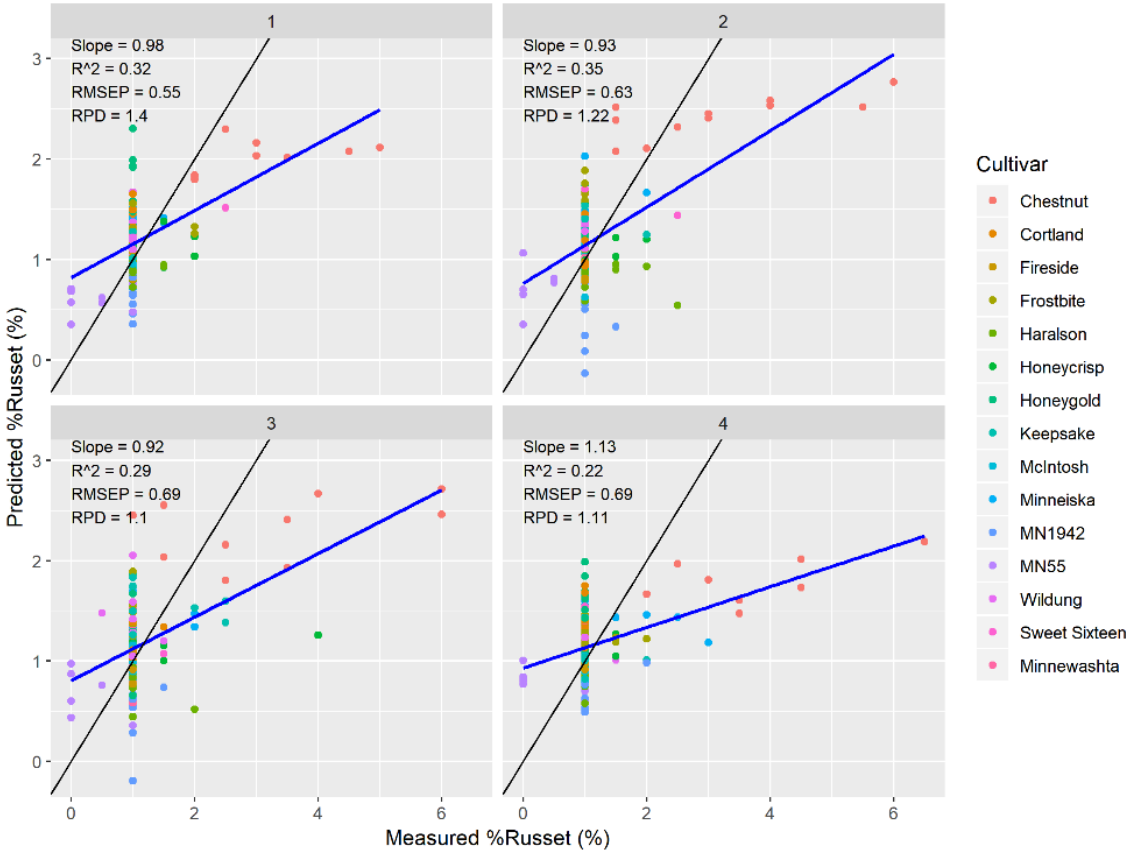
_

Figure A- 24. Indoors average spectra at-harvest prediction plots for the trait QF



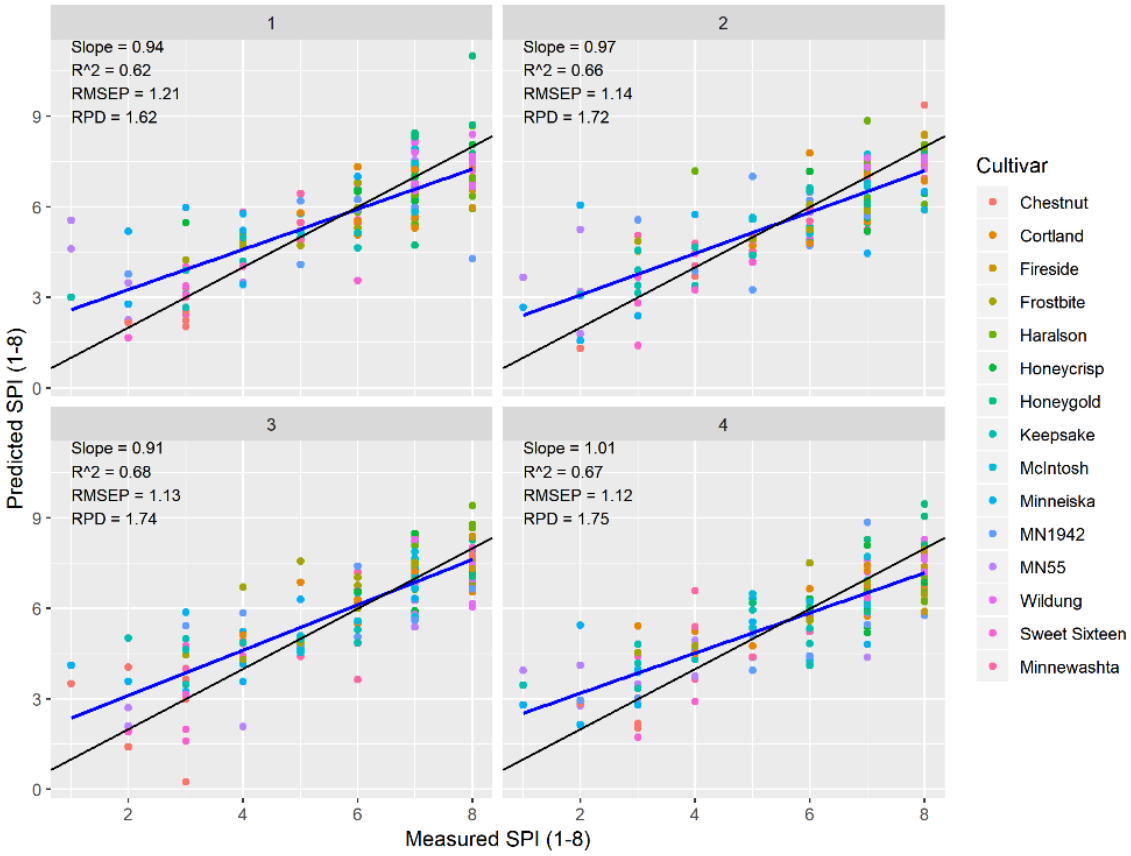
*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

Figure A- 25. Indoors average spectra at-harvest prediction plots for the trait Russet



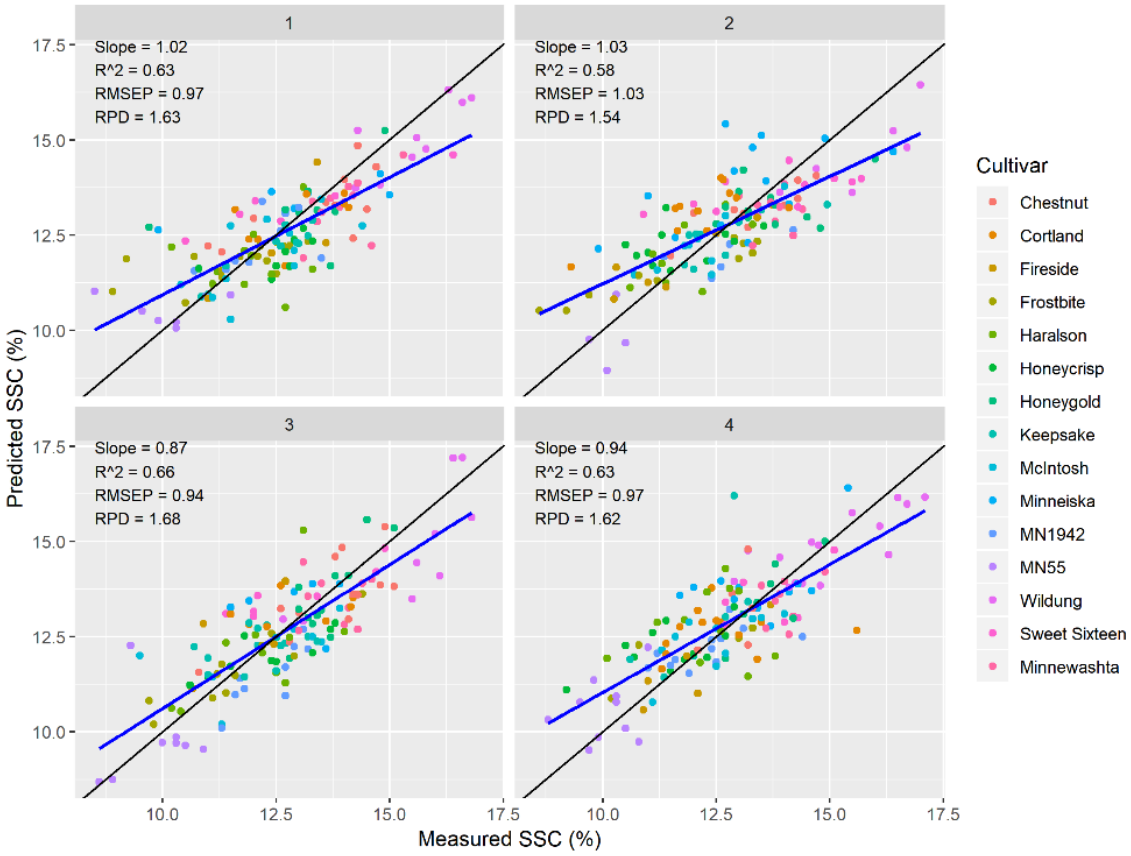
*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

Figure A- 26. Indoors average spectra at-harvest prediction plots for the trait SPI



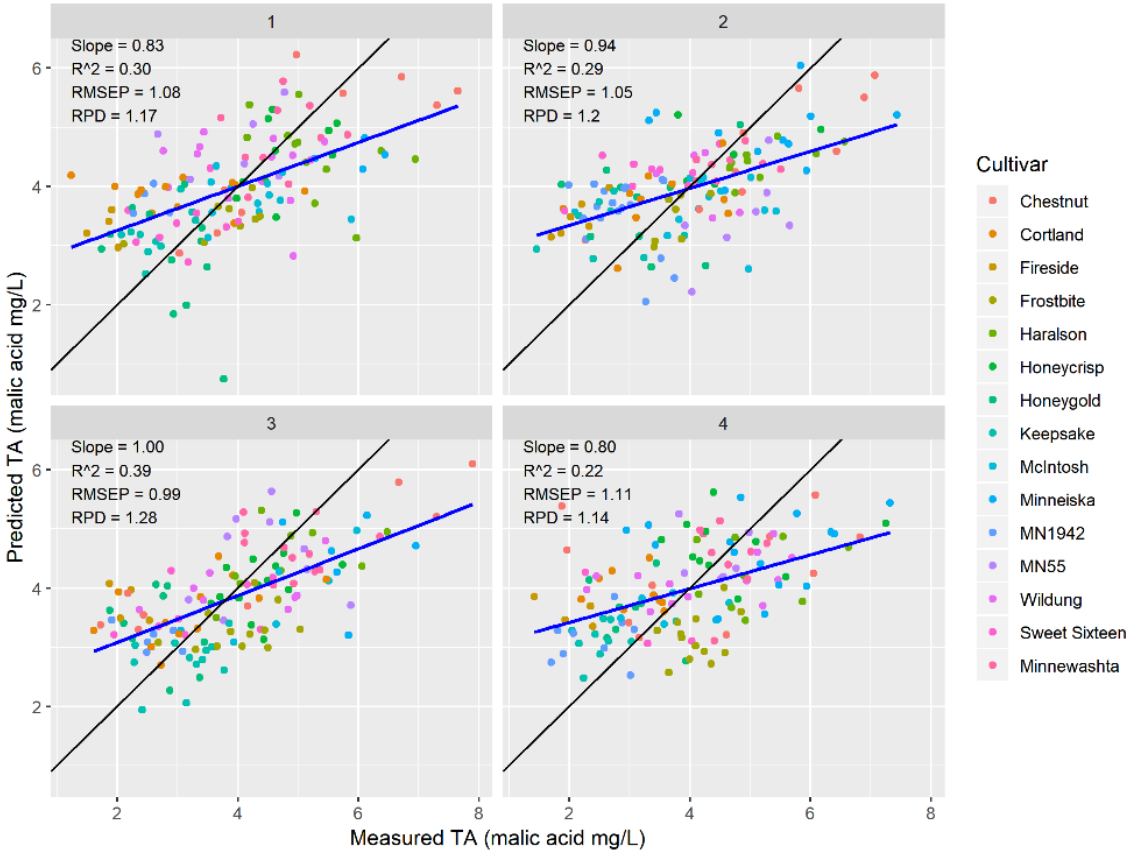
*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

Figure A- 27 Indoors average spectra at-harvest prediction plots for the trait SSC



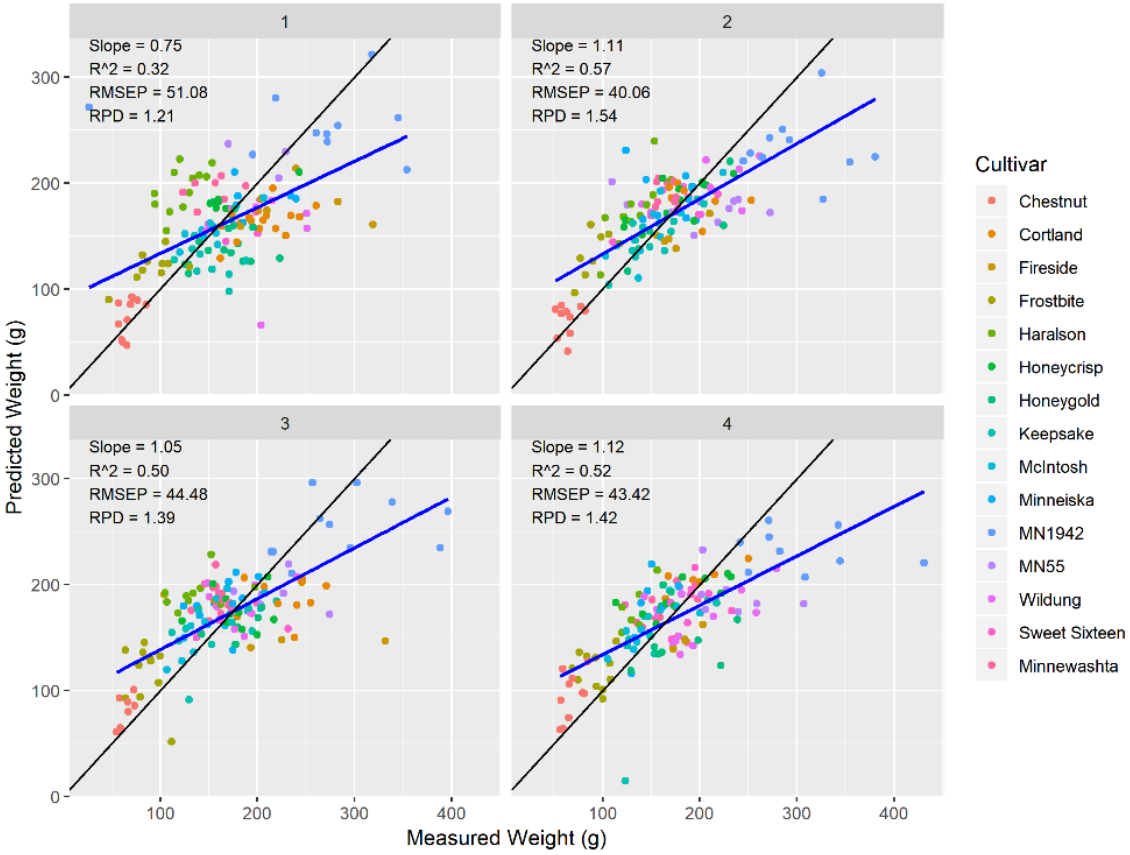
*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

Figure A- 28. Indoors average spectra at-harvest prediction plots for the trait TA



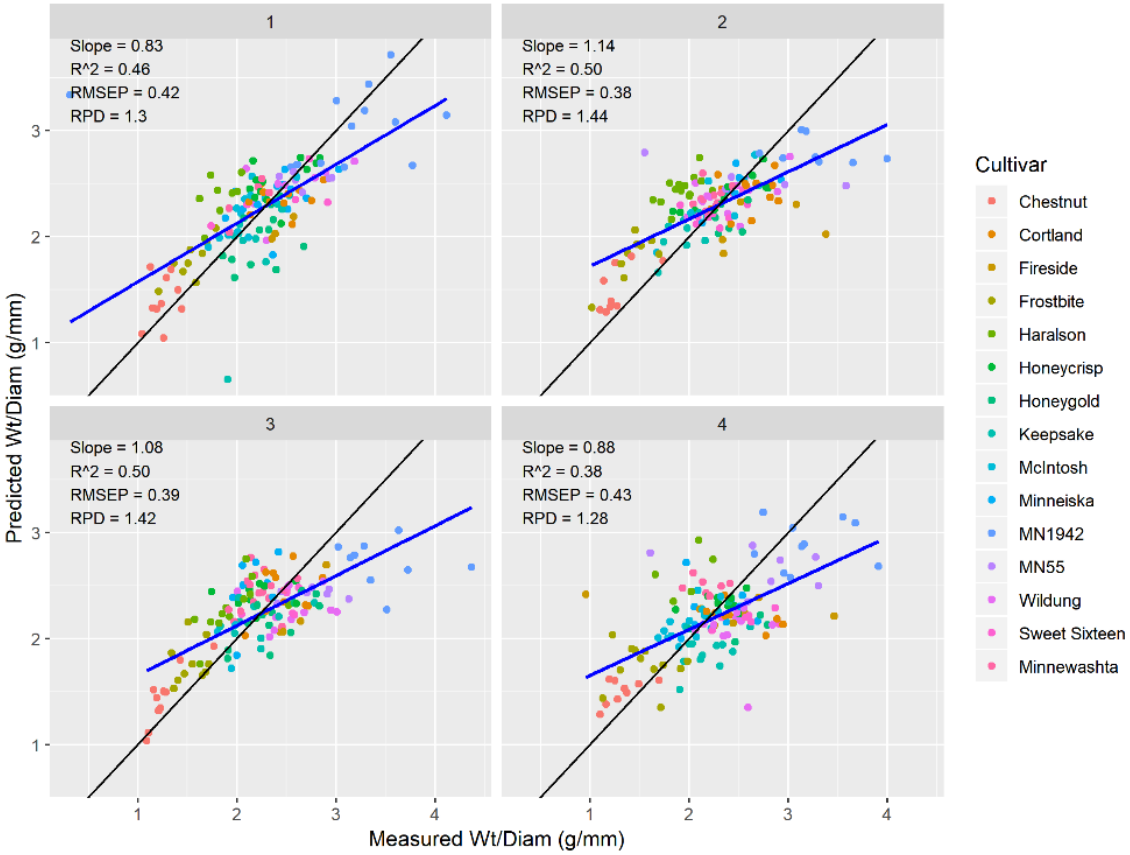
*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

Figure A- 29. Indoors average spectra at-harvest prediction plots for the trait Weight



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

Figure A- 30. Indoors average spectra at-harvest prediction plots for the trait Weight Diameter Ratio



*black line follows 1:1; blue line is regression line; numbers 1:4 represent the different data divisions

A.3 Tables of Partial Least Squares Regression Model Calibrations

_

Table A 3. Calibration-A. Partial least squares regression using spectra taken indoors with shade and sunny side averaged on the day of harvest, 2016.

Trait	N_{cal}	N_{val}	r^2_{cal}	RMSEC	LV	R^2_{val}	RMSEP	RPD
A1	383	128	0.55	3.99 (±0.11)	14	0.48	3.95 (±0.15)	1.38
A2	404	135	0.54	8.58 (±0.16)	15	0.43	8.54 (±0.55)	1.32
Creep	381	127	0.44	0.43 (±0.01)	15	0.40	0.42 (±0.06)	1.30
Crispness	377	126	0.34	94.19 (±2.12)	3	0.15	94.36 (±2.72)	1.07
E2	412	137	0.37	13.05 (±0.24)	12	0.19	13.22 (±0.76)	1.11
M1	383	128	0.54	7.30 (±0.19)	13	0.44	7.41 (±0.57)	1.34
M2	404	135	0.47	11.34 (±0.33)	14	0.35	11.22 (±0.53)	1.23
OAH	412	137	0.47	6.80 (±0.28)	14	0.39	6.74 (±0.59)	1.29
OMH	412	137	0.45	11.21 (±0.26)	14	0.31	11.42 (±0.27)	1.20
QF	412	137	0.46	34.99 (±1.27)	15	0.39	34.71 (±3.82)	1.29
%Ocolor	413	138	0.81	1.05 (±0.01)	6	0.78	1.04 (±0.05)	2.12
%Russet	413	138	0.50	0.65 (±0.04)	3	0.31	0.63 (±0.06)	1.22
DM	407	136	0.33	2.06 (±0.07)	7	0.15	2.11 (±0.06)	1.08
pH	406	135	0.39	0.19 (±0.01)	16	0.27	0.19 (±0.01)	1.18
SSC	405	135	0.70	0.97 (±0.02)	14	0.62	0.98 (±0.04)	1.62
SPI	413	138	0.73	1.18 (±0.02)	23	0.65	1.17 (±0.08)	1.68
TA	406	135	0.46	1.06 (±0.03)	15	0.30	1.06 (±0.04)	1.19
Flav	291	97	0.52	0.08 (±0.00)	9	0.26	0.08 (±0.00)	1.16
Phen	291	97	0.50	0.08 (±0.00)	8	0.24	0.08 (±0.00)	1.16
Diameter	412	137	0.69	6.54 (±0.19)	12	0.60	6.48 (±0.24)	1.58
Weight	413	138	0.61	45.49 (±0.61)	9	0.48	44.77 (±4.60)	1.39
Wt/Diam	412	137	0.59	0.41 (±0.02)	9	0.47	0.41 (±0.03)	1.37

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , and LV are rounded to whole integers.

Table A 4. Calibration-B. Partial least squares regression using spectra taken indoors on the shaded side of the fruit at-harvest, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	LV	r^2_{VAL}	RMSEP	RPD
A1	383	128	0.53	4.42 (±0.11)	9	0.37	4.36 (±0.20)	1.25
A2	404	135	0.48	9.56 (±0.29)	11	0.28	9.56 (±0.47)	1.17
Creep	381	127	0.38	0.48 (±0.01)	7	0.19	0.49 (±0.03)	1.11
Crispness	377	126	0.26	96.43 (±0.60)	7	0.11	95.34 (±5.60)	1.06
E2	412	137	0.32	14.00 (±0.38)	7	0.10	14.02 (±0.64)	1.05
M1	383	128	0.52	7.88 (±0.27)	10	0.37	7.88 (±0.55)	1.26
M2	404	135	0.41	12.57 (±0.19)	9	0.21	12.25 (±0.58)	1.13
OAH	412	137	0.43	7.65 (±0.38)	11	0.24	7.68 (±0.80)	1.14
OMH	412	137	0.39	12.63 (±0.23)	6	0.14	12.69 (±0.18)‡	1.08
QF	412	137	0.39	40.33 (±1.61)	10	0.18	40.93 (±2.74)	1.09
%Ocolor	413	138	0.74	1.25 (±0.01)	10	0.69	1.23 (±0.07)‡	1.79
%Russet	413	138	0.49	0.65 (±0.03)	6	0.31	0.63 (±0.08)	1.23
DM	407	136	0.30	2.06 (±0.04)	7	0.17	2.09 (±0.09)	1.09
pH	406	135	0.31	0.21 (±0.00)	10	0.15	0.21 (±0.01)	1.08
SSC	405	135	0.62	1.08 (±0.02)	12	0.52	1.10 (±0.07)‡	1.45
SPI	413	138	0.63	1.43 (±0.03)	22	0.47	1.44 (±0.03)‡	1.36
TA	406	135	0.40	1.15 (±0.03)	12	0.19	1.14 (±0.05)	1.11
Flav	291	97	0.43	0.08 (±0.00)	8	0.21	0.08 (±0.00)	1.12
Phen	291	97	0.36	0.08 (±0.00)	8	0.19	0.09 (±0.00)	1.12
Diameter	412	137	0.68	6.82 (±0.07)	10	0.56	6.82 (±0.44)	1.50
Weight	413	138	0.62	45.75 (±1.88)	7	0.44	46.16 (±2.24)	1.34
Wt/Diam	412	137	0.63	0.42 (±0.02)	6	0.39	0.43 (±0.03)	1.28

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , and LV are rounded to whole integers.

‡ indicates that the RMSEP is statistically different from models calibrated using the average indoor spectra when α=0.05.

Table A 5. Calibration-C. PLSR prediction of apple fruit traits using F-750 from spectra taken indoors on sun exposed side of the fruit at-harvest, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	LV	r^2_{VAL}	RMSEP	RPD
A1	383	128	0.45	4.41 (±0.10)	9	0.37	4.39 (±0.39)	1.25
A2	404	135	0.45	9.08 (±0.32)	11	0.28	9.27 (±0.81)	1.22
Creep	381	127	0.41	0.46 (±0.02)	7	0.19	0.46 (±0.04)	1.17
Crispness	377	126	0.27	94.50 (±2.92)	7	0.11	95.08 (±2.03)	1.07
E2	412	137	0.34	13.47 (±0.51)	7	0.10	13.48 (±0.52)	1.09
M1	383	128	0.41	8.17 (±0.11)	10	0.37	8.31 (±0.61)	1.19
M2	404	135	0.41	11.82 (±0.28)	9	0.21	11.96 (±0.70)	1.15
OAH	412	137	0.43	6.99 (±0.35)	11	0.24	7.14 (±0.51)	1.22
OMH	412	137	0.41	11.70 (±0.46)	6	0.14	11.91 (±0.78)	1.15
QF	412	137	0.43	36.42 (±1.88)	10	0.18	35.77 (±5.75)	1.26
%Ocolor	413	138	0.75	1.21 (±0.03)	10	0.69	1.20 (±0.06)‡	1.83
%Russet	413	138	0.41	0.67 (±0.05)	6	0.31	0.66 (±0.07)	1.18
DM	407	136	0.27	2.18 (±0.05)	7	0.17	2.18 (±0.04)	1.04
pH	406	135	0.37	0.19 (±0.00)	10	0.15	0.19 (±0.00)	1.16
SSC	405	135	0.59	1.12 (±0.01)	12	0.52	1.13 (±0.06)‡	1.40
SPI	413	138	0.67	1.35 (±0.03)	22	0.47	1.34 (±0.16)	1.48
TA	406	135	0.40	1.09 (±0.01)	12	0.19	1.08 (±0.04)	1.17
Flav	291	97	0.49	0.08 (±0.00)	8	0.21	0.08 (±0.00)	1.16
Phen	291	97	0.49	0.08 (±0.00)	8	0.19	0.08 (±0.01)	1.15
Diameter	412	137	0.60	7.55 (±0.20)	10	0.56	7.61 (±0.58)‡	1.35
Weight	413	138	0.50	50.71 (±1.51)	7	0.44	51.06 (±4.55)	1.22
Wt/Diam	412	137	0.49	0.46 (±0.01)	6	0.39	0.45 (±0.02)	1.22

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , and LV are rounded to whole integers.

‡ indicates that the RMSEP is statistically different from models calibrated using the average indoor spectra when α=0.05.

Table A 6. Calibration-D. Partial least squares regression using F-750 spectra of both sides (sunny and shade not averaged) of the fruit taken indoors on the day of harvest, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	LV	r^2_{VAL}	RMSEP	RPD
A1	747	249	0.50	4.16 (±0.08)	15	0.42	4.19 (±0.14)	1.32
A2	789	263	0.47	9.06 (±0.19)	16	0.36	9.04 (±0.28)	1.25
Creep	743	248	0.40	0.45 (±0.01)	16	0.31	0.45 (±0.03)	1.21
Crispness	735	245	0.27	96.05 (±0.86)	5	0.11	96.30 (±1.72)	1.06
E2	804	268	0.33	13.20 (±0.20)	13	0.18	13.43 (±0.57)	1.10
M1	747	249	0.47	7.70 (±0.09)	15	0.41	7.68 (±0.06)	1.30
M2	789	263	0.40	11.79 (±0.18)	15	0.29	11.72 (±0.42)	1.18
OAH	804	268	0.43	7.13 (±0.14)	15	0.34	7.09 (±0.58)	1.24
OMH	804	268	0.40	11.61 (±0.36)	15	0.27	11.78 (±0.32)	1.17
QF	804	268	0.42	36.89 (±0.78)	16	0.31	37.08 (±1.80)	1.21
%Ocolor	806	269	0.73	1.21 (±0.02)	21	0.69	1.22 (±0.06)‡	1.81
%Russet	806	269	0.41	0.64 (±0.02)	7	0.28	0.64 (±0.01)	1.18
DM	794	265	0.27	2.04 (±0.01)	13	0.16	2.05 (±0.05)	1.09
pH	791	264	0.33	0.20 (±0.00)	18	0.23	0.20 (±0.01)	1.14
SSC	789	263	0.56	1.08 (±0.04)	17	0.51	1.08 (±0.03)‡	1.42
SPI	804	268	0.65	1.27 (±0.02)	24	0.57	1.27 (±0.04)	1.52
TA	791	264	0.40	1.09 (±0.02)	17	0.28	1.08 (±0.03)	1.18
Flav	582	194	0.43	0.08 (±0.00)	13	0.28	0.08 (±0.01)	1.18
Phen	582	194	0.41	0.08 (±0.00)	10	0.25	0.08 (±0.01)	1.15
Diameter	804	268	0.61	6.84 (±0.21)	20	0.57	6.83 (±0.44)	1.51
Weight	806	269	0.54	46.16 (±0.86)	16	0.44	46.60 (±2.68)	1.33
Wt/Diam	804	268	0.52	0.42 (±0.01)	13	0.42	0.42 (±0.03)	1.31

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , and LV are rounded to whole integers.

‡ indicates that the RMSEP is statistically different from models calibrated with average sides taken indoors when α=0.05.

Table A 7. Calibration-E. Partial least squares regression using F-750 from spectra taken outdoor on the day of harvest, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	LV	r^2_{VAL}	RMSECV	RPD
A1	387	129	0.55	4.32 (±0.06)	12	0.46	4.23 (±0.08)‡	1.37
A2	409	136	0.56	9.23 (±0.16)	13	0.38	8.99 (±0.84)	1.28
Creep	385	128	0.48	0.44 (±0.01)	13	0.32	0.45 (±0.03)	1.21
Crispness	381	127	0.28	96.45 (±0.92)	4	0.13	96.42 (±4.10)	1.07
E2	418	139	0.39	13.05 (±0.39)	12	0.20	13.30 (±0.28)	1.11
M1	387	129	0.55	8.07 (±0.13)	10	0.39	8.15 (±0.32)	1.28
M2	409	136	0.48	11.88 (±0.40)	12	0.24	12.10 (±0.29)‡	1.14
OAH	418	139	0.50	7.41 (±0.12)	13	0.33	7.30 (±0.36)	1.22
OMH	418	139	0.46	11.95 (±0.11)	12	0.27	11.71 (±0.53)	1.17
QF	418	139	0.46	37.74 (±1.21)	13	0.31	37.44 (±2.70)	1.20
%Ocolor	418	139	0.80	1.11 (±0.03)	6	0.75	1.11 (±0.06)	1.99
%Russet	418	139	0.45	0.65 (±0.03)	2	0.27	0.64 (±0.02)	1.17
DM	411	137	0.36	2.06 (±0.08)	8	0.13	2.10 (±0.06)	1.06
pH	410	137	0.46	0.19 (±0.00)	16	0.29	0.19 (±0.00)	1.18
SSC	410	137	0.69	1.00 (±0.02)	17	0.59	0.98 (±0.04)	1.57
SPI	417	139	0.66	1.32 (±0.02)	22	0.57	1.29 (±0.10)	1.53
TA	410	137	0.53	1.04 (±0.03)	15	0.35	1.03 (±0.03)	1.25
Flav	295	98	0.51	0.08 (±0.00)	10	0.30	0.08 (±0.01)	1.20
Phen	295	98	0.46	0.08 (±0.00)	10	0.26	0.09 (±0.01)	1.14
Diameter	418	139	0.75	5.84 (±0.09)	18	0.66	5.96 (±0.42)	1.73
Weight	418	139	0.64	42.24 (±1.71)	11	0.54	41.54 (±4.02)	1.49
Wt/Diam	418	139	0.63	0.38 (±0.02)	10	0.54	0.37 (±0.04)	1.48

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , and LV are rounded to whole integers.

‡ indicates that the outside RMSEP is statistically different from models calibrated using indoor average spectra when α=0.05.

Table A 8. Calibration-F. Partial least squares regression using F-750 average indoor and average outdoor spectra on the day of harvest, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	LV	r^2_{VAL}	RMSEP	RPD
A1	770	257	0.57	3.96 (±0.08)	16	0.52	3.90 (±0.21)‡	1.44
A2	813	271	0.55	8.47 (±0.07)	18	0.46	8.28 (±0.25)	1.37
Creep	766	255	0.48	0.41 (±0.01)	18	0.43	0.41 (±0.01)	1.32
Crispness	758	253	0.31	91.09 (±1.30)	11	0.22	90.75 (±3.91)	1.13
E2	830	277	0.40	12.53 (±0.24)	15	0.28	12.50 (±0.58)	1.18
M1	770	257	0.56	7.34 (±0.05)	15	0.49	7.24 (±0.21)‡	1.40
M2	813	271	0.48	11.06 (±0.18)	17	0.36	11.10 (±0.54)‡	1.24
OAH	830	277	0.50	6.78 (±0.09)	18	0.41	6.77 (±0.34)	1.30
OMH	830	277	0.47	10.98 (±0.19)	17	0.36	10.95 (±0.78)	1.26
QF	830	277	0.47	35.14 (±0.37)	17	0.38	35.16 (±1.46)	1.27
%Ocolor	831	277	0.81	1.03 (±0.03)	13	0.77	1.05 (±0.01)	2.09
%Russet	831	277	0.48	0.60 (±0.01)	9	0.37	0.60 (±0.00)	1.26
DM	818	273	0.36	1.96 (±0.02)	16	0.27	1.92 (±0.05)‡ψ	1.17
pH	816	272	0.44	0.18 (±0.00)	20	0.32	0.18 (±0.01)	1.21
SSC	815	272	0.69	0.93 (±0.02)	21	0.66	0.92 (±0.03)	1.70
SPI	830	277	0.68	1.21 (±0.02)	24	0.63	1.20 (±0.07)	1.64
TA	816	272	0.50	0.98 (±0.02)	21	0.39	0.99 (±0.04)	1.28
Flav	586	195	0.48	0.08 (±0.00)	12	0.34	0.08 (±0.00)	1.23
Phen	586	195	0.46	0.08 (±0.00)	11	0.33	0.08 (±0.00)	1.22
Diameter	830	277	0.72	5.84 (±0.16)	21	0.67	5.87 (±0.30)	1.75
Weight	831	277	0.63	41.09 (±0.87)	18	0.57	40.19 (±2.44)	1.53
Wt/Diam	830	277	0.62	0.37 (±0.01)	18	0.54	0.37 (±0.02)	1.47

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , and LV are rounded to whole integers.

‡ indicates that the RMSEP is statistically different from the models calibrated using only outdoor spectra when α=0.05.

ψ indicates that the RMSEP is statistically different from the models calibrated using only indoor spectra when α=0.05.

Table A 9. Calibration-G. Partial least squares regression using F-750 spectra of both sides (sunny and shade) of the fruit taken indoors and outdoors on the day of harvest, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	LV	r^2_{VAL}	RMSEP	RPD
A1	1499	500	0.50	4.17 (±0.03)	9	0.37	4.21 (±0.10)	1.34
A2	1585	528	0.46	8.80 (±0.10)	11	0.28	8.77 (±0.26)	1.29
Creep	1490	497	0.40	0.44 (±0.01)	7	0.19	0.44 (±0.02)	1.24
Crispness	1475	492	0.26	93.34 (±0.18)	7	0.11	93.09 (±1.26)	1.10
E2	1618	539	0.32	12.83 (±0.22)	7	0.10	12.85 (±0.16)	1.15
M1	1499	500	0.48	7.76 (±0.06)	10	0.37	7.85 (±0.30)	1.30
M2	1585	528	0.39	11.42 (±0.12)	9	0.21	11.48 (±0.42)	1.20
OAH	1618	539	0.43	7.08 (±0.04)	11	0.24	7.01 (±0.14)	1.26
OMH	1618	539	0.39	11.38 (±0.09)	6	0.14	11.33 (±0.30)	1.21
QF	1618	539	0.41	36.34 (±0.40)	10	0.18	36.50 (±1.41)	1.22
%Ocolor	1619	540	0.72	1.20 (±0.01)	10	0.69	1.20 (±0.03)‡ψ	1.83
%Russet	1619	540	0.41	0.61 (±0.01)	6	0.31	0.61 (±0.01)	1.23
DM	1594	531	0.28	1.98 (±0.02)	7	0.17	1.98 (±0.01) ψ	1.12
pH	1589	530	0.33	0.19 (±0.00)	10	0.15	0.19 (±0.00)	1.17
SSC	1586	529	0.54	1.08 (±0.01)	12	0.52	1.07 (±0.02)‡ψ	1.42
SPI	1616	539	0.61	1.28 (±0.01)	22	0.47	1.28 (±0.05)	1.51
TA	1589	530	0.41	1.03 (±0.01)	12	0.19	1.03 (±0.01)	1.24
Flav	1162	387	0.40	0.08 (±0.00)	8	0.21	0.08 (±0.00)	1.21
Phen	1162	387	0.38	0.08 (±0.00)	8	0.19	0.08 (±0.00)	1.19
Diameter	1618	539	0.64	6.52 (±0.05)	10	0.56	6.53 (±0.19)‡	1.58
Weight	1619	540	0.55	43.41 (±0.61)	7	0.44	43.60 (±2.06)	1.41
Wt/Diam	1618	539	0.53	0.39 (±0.01)	6	0.39	0.39 (±0.00)	1.41

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , and LV are rounded to whole integers.

‡ indicates that the RMSEP is statistically different from models calibrated with average indoors and average outdoor spectra when α=0.05.

ψ indicates that the RMSEP is statistically different from models calibrated with average indoor spectra when α=0.05.

Table A 10. Calibration-H. Genetic algorithm (GA) wavelength selection with partial least squares regression using F-750 average spectra taken indoor* on the day of harvest, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	W_{Num}	LV	r^2_{VAL}	RMSEP	RPD
A1	383	128	0.46	4.29 (±0.09)	20	6	0.39	4.27 (±0.27)	1.28
A2	404	135	0.38	9.26 (±0.17)	22	11	0.33	9.16 (±0.61)	1.23
Creep	381	127	0.33	0.47 (±0.01)	20	4	0.22	0.47 (±0.03)	1.14
Crispness	377	126	0.23	93.61 (±1.79)	25	3	0.17	93.32 (±4.03)	1.09
E2	412	137	0.25	13.33 (±0.19)	18	8	0.14	13.59 (±0.62)	1.08
M1	383	128	0.43	7.89 (±0.25)	19	8	0.34	8.04 (±0.42)	1.23
M2	404	135	0.34	11.69 (±0.15)	22	13	0.29	11.55 (±0.45)	1.19
OAH	412	137	0.40	7.15 (±0.33)	25	10	0.29	7.33 (±0.90)	1.19
OMH	412	137	0.31	11.98 (±0.72)	22	7	0.20	12.23 (±0.28)‡	1.12
QF	412	137	0.35	37.08 (±0.33)	22	10	0.28	37.98 (±5.45)	1.18
%Ocolor	413	138	0.79	1.04 (±0.02)	17	5	0.77	1.04 (±0.06)	2.11
%Russet	413	138	0.40	0.65 (±0.04)	26	3	0.31	0.64 (±0.07)	1.22
DM	407	136	0.27	2.06 (±0.07)	20	5	0.17	2.08 (±0.05)	1.10
pH	406	135	0.23	0.21 (±0.01)	17	7	0.12	0.21 (±0.01)	1.07
SSC	405	135	0.63	0.99 (±0.06)	16	9	0.61	0.98 (±0.04)	1.61
SPI	413	138	0.48	1.51 (±0.07)	23	13	0.41	1.50 (±0.07)‡	1.31
TA	406	135	0.33	1.11 (±0.03)	26	9	0.21	1.12 (±0.06)	1.13
Flav	291	97	0.37	0.08 (±0.00)	28	7	0.23	0.08 (±0.01)	1.14
Phen	291	97	0.35	0.08 (±0.00)	24	7	0.27	0.08 (±0.01)	1.18
Diameter	412	137	0.59	6.94 (±0.21)	23	4	0.53	7.06 (±0.40)‡	1.45
Weight	413	138	0.51	46.00 (±1.18)	22	4	0.44	46.61 (±3.47)	1.33
Wt/Diam	412	137	0.49	0.42 (±0.02)	17	4	0.40	0.43 (±0.01)	1.28

*Inside spectra models included 151 wavelengths

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , W_{num}, and LV are rounded to whole integers.

‡ indicates that the RMSEP is statistically different from models calibrated with all 151 wavelengths of indoor average spectra when α=0.05.

Table A 11. Calibration-I. Partial least squares regression using F-750 spectra taken indoors post-storage, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	LV	r^2_{VAL}	RMSEP	RPD
A1	396	132	0.60	4.14 (±0.13)	16	0.51	4.19 (±0.25)	1.41
A2	404	135	0.59	8.13 (±0.28)	15	0.47	8.11 (±0.31)	1.38
Creep	396	132	0.68	0.62 (±0.02)	18	0.60	0.62 (±0.04)‡	1.59
Crispness	389	130	0.44	88.76 (±2.11)	9	0.27	89.58 (±3.69)	1.18
E2	420	140	0.54	12.36 (±0.21)	14	0.40	12.38 (±0.23)	1.30
M1	396	132	0.61	6.91 (±0.10)	16	0.54	6.77 (±0.47)	1.48
M2	404	135	0.58	11.32 (±0.45)	15	0.46	11.31 (±0.63)	1.36
OAH	421	140	0.58	7.34 (±0.14)	15	0.46	7.43 (±0.34)	1.36
OMH	422	141	0.56	11.93 (±0.26)	14	0.43	12.02 (±0.46)	1.32
QF	424	141	0.64	44.05 (±0.61)	17	0.55	43.76 (±0.90)‡	1.49
%Ocolor	404	135	0.83	1.00 (±0.06)	18	0.79	1.01 (±0.05)	2.19
%Russet	404	135	0.40	0.49 (±0.02)	14	0.29	0.50 (±0.03)‡	1.17
DM	411	137	0.56	1.47 (±0.06)	12	0.52	1.45 (±0.17)‡	1.46
pH	427	142	0.48	0.20 (±0.01)	17	0.32	0.21 (±0.01)	1.22
SSC	427	142	0.69	0.99 (±0.02)	12	0.62	0.99 (±0.03)	1.63
TA	427	142	0.47	0.83 (±0.04)	16	0.25	0.84 (±0.08)‡	1.16
Flav	314	105	0.31	0.09 (±0.00)	4	0.17	0.09 (±0.00)‡	1.10
Phen	314	105	0.40	0.08 (±0.00)	8	0.29	0.08 (±0.01)	1.19
Diameter	429	143	0.76	5.86 (±0.24)	18	0.69	5.83 (±0.32)‡	1.80
Weight	404	135	0.74	34.83 (±1.65)	19	0.66	34.73 (±2.17)‡	1.70
Wt/Diam	404	135	0.72	0.34 (±0.02)	18	0.62	0.33 (±0.07)	1.68
PHweight	427	142	0.74	33.13 (±0.64)	19	0.67	33.02 (±2.64)	1.74

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , and LV are rounded to whole integers.

‡ indicates that the RMSEP is statistically different from models calibrated using the average indoor spectra when α=0.05.

Table A 12. Calibration-J. Partial least squares regression of postharvest apple fruit traits using F-750 spectra taken indoor at-harvest, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	LV	r^2_{VAL}	RMSEP	RPD
A1	389	130	0.59	4.48 (±0.15)	15	0.40	4.58 (±0.12)‡	1.29
A2	397	132	0.55	8.89 (±0.33)	14	0.35	9.02 (±0.63)‡	1.24
Creep	389	130	0.65	0.69 (±0.03)	19	0.50	0.70 (±0.07)	1.43
Crispness	383	128	0.41	92.57 (±1.72)	4	0.21	93.40 (±3.67)	1.13
E2	412	137	0.49	13.38 (±0.39)	10	0.31	13.27 (±0.90)	1.20
M1	389	130	0.58	7.36 (±0.25)	17	0.46	7.33 (±0.60)	1.37
M2	397	132	0.56	12.42 (±0.23)	11	0.32	12.59 (±0.58)‡	1.22
OAH	413	138	0.56	7.85 (±0.38)	14	0.35	8.03 (±0.22)‡	1.23
OMH	414	138	0.54	12.75 (±0.47)	15	0.33	12.87 (±1.35)	1.23
QF	416	139	0.60	47.74 (±1.51)	19	0.47	47.40 (±2.78)‡	1.38
%Ocolor	396	132	0.81	1.07 (±0.03)	6	0.77	1.06 (±0.07)	2.10
%Russet	396	132	0.30	0.47 (±0.03)	10	0.26	0.45 (±0.14)	1.25
DM	403	134	0.55	1.57 (±0.06)	13	0.46	1.57 (±0.15)	1.36
pH	417	139	0.41	0.23 (±0.00)	13	0.19	0.23 (±0.01)‡	1.11
SSC	417	139	0.65	1.03 (±0.02)	15	0.60	1.02 (±0.04)	1.58
TA	417	139	0.44	0.87 (±0.02)	13	0.18	0.88 (±0.01)	1.10
Flav	308	103	0.41	0.09 (±0.00)	8	0.25	0.08 (±0.00)	1.16
Phen	308	103	0.47	0.08 (±0.00)	9	0.31	0.08 (±0.00)	1.20
Diameter	420	140	0.72	6.42 (±0.20)	12	0.62	6.29 (±0.73)	1.65
Weight	396	132	0.68	37.21 (±1.33)	15	0.59	37.80 (±3.81)	1.56
Wt/Diam	395	132	0.66	0.37 (±0.01)	10	0.55	0.37 (±0.03)	1.49
PHweight	419	140	0.69	36.43 (±1.34)	13	0.58	36.84 (±1.68)	1.54

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , and LV are rounded to whole integers.

‡ indicates that the RMSEP is statistically different from models calibrated using postharvest average indoor spectra when α=0.05.

Table A 13. Calibration-K. Partial least squares regression of postharvest apple fruit traits using F-750 spectra taken outdoors on the day of harvest, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	LV	r^2_{VAL}	RMSEP	RPD
A1	382	127	0.53	4.89 (±0.10)	5	0.29	4.91 (±0.19) ψ	1.19
A2	390	130	0.49	9.65 (±0.25)	4	0.24	9.81 (±0.31) ψ	1.15
Creep	382	127	0.58	0.78 (±0.01)	8	0.39	0.78 (±0.02) ψ	1.27
Crispness	375	125	0.39	90.58 (±3.80)	5	0.20	93.12 (±2.69)	1.12
E2	407	136	0.45	13.82 (±0.36)	8	0.28	13.73 (±0.31) ψ	1.18
M1	382	127	0.52	8.19 (±0.18)	7	0.31	8.33 (±0.31)‡ψ	1.20
M2	390	130	0.49	12.90 (±0.21)	8	0.31	12.78 (±0.85) ψ	1.21
OAH	407	136	0.49	8.50 (±0.15)	4	0.28	8.52 (±0.27) ψ	1.18
OMH	409	136	0.47	13.65 (±0.29)	4	0.26	13.74 (±1.18) ψ	1.17
QF	410	137	0.52	51.94 (±3.11)	13	0.32	54.58 (±3.02) ψ	1.19
%Ocolor	405	135	0.82	1.05 (±0.01)	6	0.77	1.06 (±0.11)	2.10
%Russet	405	135	0.33	0.49 (±0.01)	15	0.31	0.49 (±0.03)	1.20
DM	398	133	0.52	1.61 (±0.05)	12	0.38	1.61 (±0.12)	1.28
pH	413	138	0.41	0.21 (±0.00)	17	0.26	0.21 (±0.01)	1.16
SSC	413	138	0.61	1.07 (±0.02)	15	0.52	1.07 (±0.04) ψ	1.44
TA	413	138	0.43	0.83 (±0.03)	13	0.24	0.83 (±0.02)	1.15
Flav	303	101	0.34	0.09 (±0.00)	8	0.23	0.09 (±0.00)	1.13
Phen	303	101	0.39	0.08 (±0.00)	8	0.26	0.08 (±0.00)	1.17
Diameter	416	139	0.75	6.23 (±0.28)	14	0.66	6.20 (±0.64)	1.73
Weight	405	135	0.67	38.25 (±1.86)	11	0.57	38.83 (±3.21)	1.53
Wt/Diam	404	135	0.65	0.37 (±0.02)	9	0.56	0.36 (±0.05)	1.54
PHweight	413	138	0.68	36.94 (±0.82)	13	0.58	37.42 (±0.58) ψ	1.54

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction. RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , and LV are rounded to whole integers.

‡ indicates that the RMSEP is statistically different from models calibrated using at-harvest average indoor spectra to predict postharvest traits when α=0.05.

ψ indicates that the RMSEP is statistically different from models calibrated using postharvest average indoor spectra when α=0.05.

Table A 14. Calibration-L. Partial least squares regression of postharvest apple fruit traits using average indoor and average outdoor spectra taken on the day of harvest, 2016.

Trait	N_{cal}	N_{val}	r^2_{cal}	RMSEC	LV	R^2_{val}	RMSEP	RPD
A1	771	257	0.56	4.45 (±0.08)	19	0.43	4.42 (±0.09)‡	1.33
A2	787	262	0.53	8.34 (±0.14)	22	0.40	8.75 (±0.56)	1.28
Creep	771	257	0.61	0.68 (±0.00)	21	0.53	0.68 (±0.03)‡	1.46
Crispness	758	253	0.41	87.80 (±1.63)	8	0.29	88.37 (±2.82)‡	1.18
E2	818	273	0.47	12.42 (±0.30)	22	0.40	12.50 (±0.23)	1.28
M1	771	257	0.56	7.37 (±0.07)	20	0.47	7.26 (±0.19)	1.38
M2	787	262	0.52	11.65 (±0.14)	22	0.43	11.67 (±0.58)	1.32
OAH	820	273	0.52	7.55 (±0.18)	22	0.44	7.44 (±0.28)	1.34
OMH	823	274	0.50	12.21 (±0.35)	22	0.39	12.41 (±0.95)	1.28
QF	826	275	0.57	46.93 (±0.83)	22	0.49	46.49 (±0.92) ψ	1.40
%Ocolor	801	267	0.82	1.03 (±0.03)	7	0.79	1.02 (±0.08)	2.18
%Russet	801	267	0.33	0.49 (±0.02)	3	0.21	0.49 (±0.04)‡	1.14
DM	800	267	0.53	1.51 (±0.03)	15	0.48	1.49 (±0.08)‡	1.40
pH	830	277	0.40	0.21 (±0.00)	18	0.29	0.21 (±0.01)	1.19
SSC	830	277	0.64	1.02 (±0.03)	17	0.59	1.01 (±0.03)	1.56
TA	830	277	0.44	0.80 (±0.01)	20	0.33	0.79 (±0.02)‡	1.22
Flav	611	204	0.41	0.08 (±0.00)	9	0.30	0.08 (±0.00) ψ	1.20
Phen	611	204	0.47	0.08 (±0.00)	11	0.35	0.08 (±0.00)	1.24
Diameter	836	279	0.73	5.85 (±0.17)	21	0.68	5.90 (±0.35)	1.78
Weight	801	267	0.68	35.88 (±0.54)	19	0.63	35.90 (±1.49)	1.64
Wt/Diam	800	267	0.66	0.35 (±0.01)	16	0.59	0.35 (±0.02)	1.56
PHweight	832	277	0.69	34.71 (±0.58)	20	0.64	34.42 (±1.82)	1.67

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , and LV are rounded to whole integers.

‡ indicates that the RMSEP is statistically different from models calibrated using average indoor spectra for predicting at-harvest traits when α=0.05.

ψ indicates that the RMSEP is statistically different from models calibrated using postharvest spectra for predicting postharvest traits when α=0.05.

Table A 15. Calibration-M. Partial least squares regression using F-750 average indoor and average outdoor spectra on the day of harvest with fruit surface and spectrometer temperature as x-variables, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	LV	r^2_{VAL}	RMSEP	RPD
A1	755	252	0.58	3.95 (±0.07)	18	0.51	3.98 (±0.10)	1.42
A2	798	266	0.56	8.32 (±0.11)	19	0.47	8.30 (±0.43)	1.38
Creep	751	250	0.49	0.41 (±0.01)	19	0.43	0.41 (±0.01)	1.32
Crispness	743	248	0.31	90.54 (±1.12)	13	0.19	92.33 (±2.73)	1.11
E2	815	272	0.40	12.60 (±0.08)	16	0.30	12.41 (±0.37)	1.20
M1	755	252	0.56	7.27 (±0.21)	18	0.48	7.33 (±0.48)	1.40
M2	798	266	0.48	10.97 (±0.11)	18	0.37	11.05 (±0.32)	1.26
OAH	815	272	0.51	6.77 (±0.04)	19	0.44	6.65 (±0.23)	1.33
OMH	815	272	0.48	11.00 (±0.15)	18	0.37	11.02 (±0.27)	1.25
QF	815	272	0.48	34.28 (±0.33)	19	0.42	34.12 (±0.84)	1.32
%Ocolor	816	272	0.81	1.05 (±0.02)	10	0.78	1.04 (±0.03)	2.12
%Russet	816	272	0.48	0.60 (±0.01)	11	0.39	0.60 (±0.04)	1.28
DM	803	268	0.36	1.98 (±0.02)	13	0.23	1.99 (±0.04) ψ	1.14
pH	801	267	0.43	0.18 (±0.00)	21	0.32	0.18 (±0.01)	1.22
SSC	800	267	0.71	0.92 (±0.00)	23	0.64	0.93 (±0.04)	1.68
SPI	815	272	0.70	1.17 (±0.01)	26	0.64	1.17 (±0.07)	1.68
TA	801	267	0.51	0.99 (±0.02)	21	0.40	0.98 (±0.03) ψ	1.29
Flav	586	195	0.48	0.08 (±0.00)	14	0.35	0.08 (±0.00)	1.24
Phen	586	195	0.46	0.08 (±0.00)	13	0.33	0.08 (±0.01)	1.23
Diameter	815	272	0.73	5.89 (±0.05)	21	0.69	5.78 (±0.14) ψ	1.79
Weight	816	272	0.64	40.59 (±1.37)	21	0.56	41.16 (±4.17)	1.52
Wt/Diam	815	272	0.62	0.37 (±0.01)	16	0.55	0.37 (±0.02)	1.49

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

Ncal , Nval , and LV are rounded to whole integers.

‡ indicates that the RMSEP is statistically different from models calibrated using average indoor and average outdoor spectra when α=0.05.

ψ indicates that the RMSEP is statistically different from models calibrated using average indoor spectra when α=0.05.

Table A 16. Calibration-N. Partial least squares regression using F-750 average indoor and average outdoor spectra on the day of harvest with fruit surface temperature as x-variables, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	LV	r^2_{VAL}	RMSEP	RPD
A1	755	252	0.58	3.93 (±0.09)	18	0.51	3.98 (±0.09)	1.42
A2	798	266	0.56	8.40 (±0.12)	19	0.45	8.42 (±0.45)	1.36
Creep	751	250	0.49	0.41 (±0.01)	18	0.42	0.41 (±0.01)	1.31
Crispness	743	248	0.31	90.73 (±1.22)	12	0.20	92.22 (±3.00)	1.11
E2	815	272	0.40	12.73 (±0.08)	15	0.28	12.58 (±0.35)	1.18
M1	755	252	0.56	7.22 (±0.24)	18	0.49	7.31 (±0.46)	1.40
M2	798	266	0.48	11.10 (±0.09)	17	0.35	11.19 (±0.28)	1.24
OAH	815	272	0.50	6.89 (±0.11)	18	0.41	6.79 (±0.17)	1.30
OMH	815	272	0.47	11.03 (±0.16)	19	0.37	11.01 (±0.35)	1.25
QF	815	272	0.48	34.70 (±0.39)	18	0.41	34.50 (±0.90)	1.30
%Ocolor	816	272	0.81	1.05 (±0.02)	9	0.77	1.04 (±0.04)	2.11
%Russet	816	272	0.48	0.60 (±0.01)	11	0.39	0.59 (±0.04)	1.29
DM	803	268	0.36	1.98 (±0.01)	12	0.22	1.99 (±0.03) ψ	1.14
pH	801	267	0.43	0.18 (±0.00)	20	0.32	0.18 (±0.01)	1.21
SSC	800	267	0.70	0.92 (±0.01)	21	0.65	0.93 (±0.03)	1.69
SPI	815	272	0.70	1.17 (±0.02)	25	0.65	1.17 (±0.07)	1.69
TA	801	267	0.51	0.99 (±0.02)	20	0.40	0.98 (±0.03) ψ	1.29
Flav	586	195	0.48	0.08 (±0.00)	13	0.34	0.08 (±0.00)	1.23
Phen	586	195	0.46	0.08 (±0.00)	12	0.34	0.08 (±0.00)	1.23
Diameter	815	272	0.73	5.91 (±0.04)	20	0.68	5.81 (±0.13) ψ	1.78
Weight	816	272	0.64	40.61 (±1.32)	20	0.56	40.91 (±4.14)	1.52
Wt/Diam	815	272	0.62	0.38 (±0.00)	14	0.54	0.37 (±0.02)	1.48

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

Ncal , Nval , and LV are rounded to whole integers.

ψ indicates that the RMSEP is statistically different from models calibrated using average indoor spectra when α=0.05.

Table A 17. Calibration-O. Partial least squares regression using F-750 average indoor and average outdoor spectra on the day of harvest with spectrometer temperature as x-variables, 2016.

Trait	N_{cal}	N_{val}	R^2_{CAL}	RMSEC	LV	r^2_{VAL}	RMSEP	RPD
A1	770	257	0.57	3.93 (±0.09)	18	0.53	3.86 (±0.2)	1.46
A2	813	271	0.55	8.41 (±0.09)	19	0.46	8.29 (±0.24)	1.36
Creep	766	255	0.48	0.41 (±0.01)	18	0.42	0.41 (±0.01)	1.32
Crispness	758	253	0.31	90.91 (±0.98)	11	0.21	91.01 (±3.4)	1.12
E2	830	277	0.41	12.59 (±0.17)	15	0.28	12.55 (±0.55)	1.18
M1	770	257	0.56	7.35 (±0.09)	16	0.49	7.25 (±0.19)	1.40
M2	813	271	0.48	11.05 (±0.12)	17	0.36	11.02 (±0.44)	1.25
OAH	830	277	0.51	6.79 (±0.13)	18	0.40	6.79 (±0.35)	1.30
OMH	830	277	0.47	10.92 (±0.2)	18	0.37	10.89 (±0.76)	1.26
QF	830	277	0.48	34.61 (±0.3)	17	0.38	35.05 (±1.5)	1.27
%Ocolor	831	277	0.81	1.04 (±0.03)	11	0.77	1.05 (±0)	2.08
%Russet	831	277	0.48	0.6 (±0.01)	9	0.36	0.6 (±0.01)	1.25
DM	818	273	0.36	1.97 (±0.02)	16	0.26	1.93 (±0.04)‡	1.17
pH	816	272	0.44	0.18 (±0)	21	0.33	0.18 (±0.01)	1.23
SSC	815	272	0.69	0.92 (±0.01)	22	0.66	0.91 (±0.05)	1.72
SPI	830	277	0.68	1.2 (±0.02)	25	0.63	1.2 (±0.07)	1.64
TA	816	272	0.50	0.98 (±0.02)	21	0.39	1 (±0.04)	1.28
Flav	586	195	0.48	0.08 (±0)	13	0.34	0.08 (±0)	1.23
Phen	586	195	0.46	0.08 (±0)	12	0.32	0.08 (±0.01)	1.22
Diameter	830	277	0.72	5.9 (±0.09)	20	0.67	5.9 (±0.23)‡	1.74
Weight	831	277	0.63	41.12 (±0.82)	19	0.57	40.42 (±1.99)	1.52
Wt/Diam	830	277	0.62	0.37 (±0.01)	18	0.53	0.38 (±0.03)	1.46

Abbreviations: cal = calibration; val= validation; LV= latent variable (i.e. component); RMSE (C, P) = root mean square error of calibration, and prediction.

RMSE values are in the units of the variable; the bracketed value is the standard error calculated from the 4 data splits.

N_{cal} , N_{val} , and LV are rounded to whole integers.

‡ indicates that the RMSEP is statistically different from models calibrated using average indoor and average outdoor spectra when α=0.05.

APPENDIX B CLASSIFICATION CONFUSION MATRICIES

Table B- 1. Confusion matrices for the prediction of the categories of Ocolor and SPI by linear discriminate analysis (LDA).

Ocolor: Overall classification accuracy was 0.52 (±0.04)

	dkrd	none	oran	ored	pink	red
dkrd	23	0	0	1	1	12
none	0	5	1	0	0	2
oran	0	1	2	0	0	0
ored	1	0	0	0	1	1
pink	1	0	0	1	2	7
red	14	0	0	1	5	53

Abbreviations (definitions): dkrd=dark-red, oran=orange, ored=orange-red, none (no-overcolor)

SPI: Overall classification accuracy was 0.61 (±0.04)

	fres	inmt	over	stor
fres	48	1	11	8
inmt	1	2	0	4
over	10	0	17	1
stor	10	4	0	15

Abbreviations (definitions): , fres=fresh (SPI 5-7) , inmat=immature (SPI 1-2), over=over-mature (SPI 8), stor=storage (SPI 3-4)

Ocolor

The categorization results for Ocolor (64%) were higher than Gcolor; however, the categories were not well balanced. There were too few fruit in the categories of dark-, pink-, or orange-red.

Starch pattern index

Starch pattern was divided into four categories with accuracies ranging from 28 to 71%. The low predictive accuracy could also be due to the subjective nature of the trait acquisition, or maybe low proportions of fruit in some categories.

Table B- 2. Confusion matrices for the prediction of the categories of Gcolor and Ocolor type by linear discriminate analysis (LDA).

Gcolor:
Overall classification accuracy was 0.46 (±0.03)

	gree	plgr	plyl	yell	ylgr
gree	5	3	1	0	4
plgr	2	3	2	1	2
plyl	1	1	7	10	3
yell	1	1	9	26	11
ylgr	3	3	4	11	19

Abbreviations (definitions): gree=green, plgr=pale-green, plyl=pale-yellow, yell=yellow, ylgr=yellow-green

Ocolor type:
Overall classification accuracy was 0.52 (±0.04)

	blst	blus	none	stbl	stri
blst	38	15	1	9	0
blus	16	25	1	5	0
none	0	1	1	0	0
stbl	8	6	0	7	0
stri	0	0	0	0	0

Abbreviations (definitions): blst=blush-stripe, blus=blush, stbl=stripe-blush, stri=stripe

_

Ground color

Gcolor was measured subjectively. It was a trait with more balanced data, i.e. samples were distributed more evenly among categories. However, fruit were still not well classified, with an overall accuracy below 50%. An alternative method, yet still subjective, that has been used is a based on ten color swatches from yellow to green and rated on a 1-10 scale (McGlone et al., 2002a; Palmer et al., 2010). The ideal method would be to have an instrumental reading measurement of the actual color in color spaces such as RGB, CMYK, or CIELAB.

Ocolor type

Highly subjective and difficult to score, Ocolor type was not well-classified (54%), and categories were not well-proportioned.

Table B- 3. Confusion matrices for the prediction of the categories of fruit damage, and %Ocolor by linear discriminate analysis (LDA)

Fruit damage: Overall classification accuracy was 0.85 (±0.02)

	abio	biot	both	none
abio	2	1	0	9
biot	0	0	0	2
both	0	0	0	0
none	9	2	1	106

Abbreviations (definitions): abio=abiotic (mechanical injury), bio=biotic (insect or pathogen)

%Ocolor: Overall classification accuracy was 0.85 (±0.05)

	bico	red	yell
bico	15	12	2
red	9	84	0
yell	0	0	10

Abbreviations (definitions): bico=bicolor (65-35% overcolor), red (70-100%), yell=yellow (0-30%)

Fruit Damage

Damage was recorded after fruit were harvested (harvest was done to intentionally avoid damaged fruit). The percentage of fruit that were damaged to some extent (even minor damage) was 11%, so that less than 6% of the damaged fruit could be categorized correctly. Therefore, even though the total overall accuracy for the trait was 82%, the spectra could correctly identify a good apple, but bad ones were not readily identified.

%Ocolor

When the 1-10 scale %Ocolor trait was treated as a three-category trait, the low-proportioned class (yellow fruit) was predicted with near 100% accuracy. Models had difficulty in correctly classifying bicolored fruit. %Ocolor is more accurately predicted as a quantitative trait.

Table B- 4. Confusion matrices for the prediction of the categories of BRusset, LRusset, and %Russet by linear discriminate analysis (LDA)

Presence of BRusset: Overall classification accuracy was 0.73 (±0.02)

	none	**pres**
none	70	18
pres	21	22

Presence of LRusset: Overall classification accuracy was 0.71 (±0.03)

	none	**pres**
none	24	19
pres	16	71

Overall %Russet: Overall classification accuracy was 0.94 (±0.01)

	high	**low**
high	3	5
low	2	122

Abbreviations (definitions): pres=presence, low (russet percent 0-30%) and high (35-65%)

BRusset and LRusset

If fruit russeting could be estimated accurately on a commercial fruit sorting line, then consumers could find fewer bad looking fruit in the stores. Our results were not promising, with overall accuracy between 71 and 73%. The low-proportioned classes (high incidence) of both body and lenticel, which are important to accurately classify, were correctly classified only ~50% of the time.

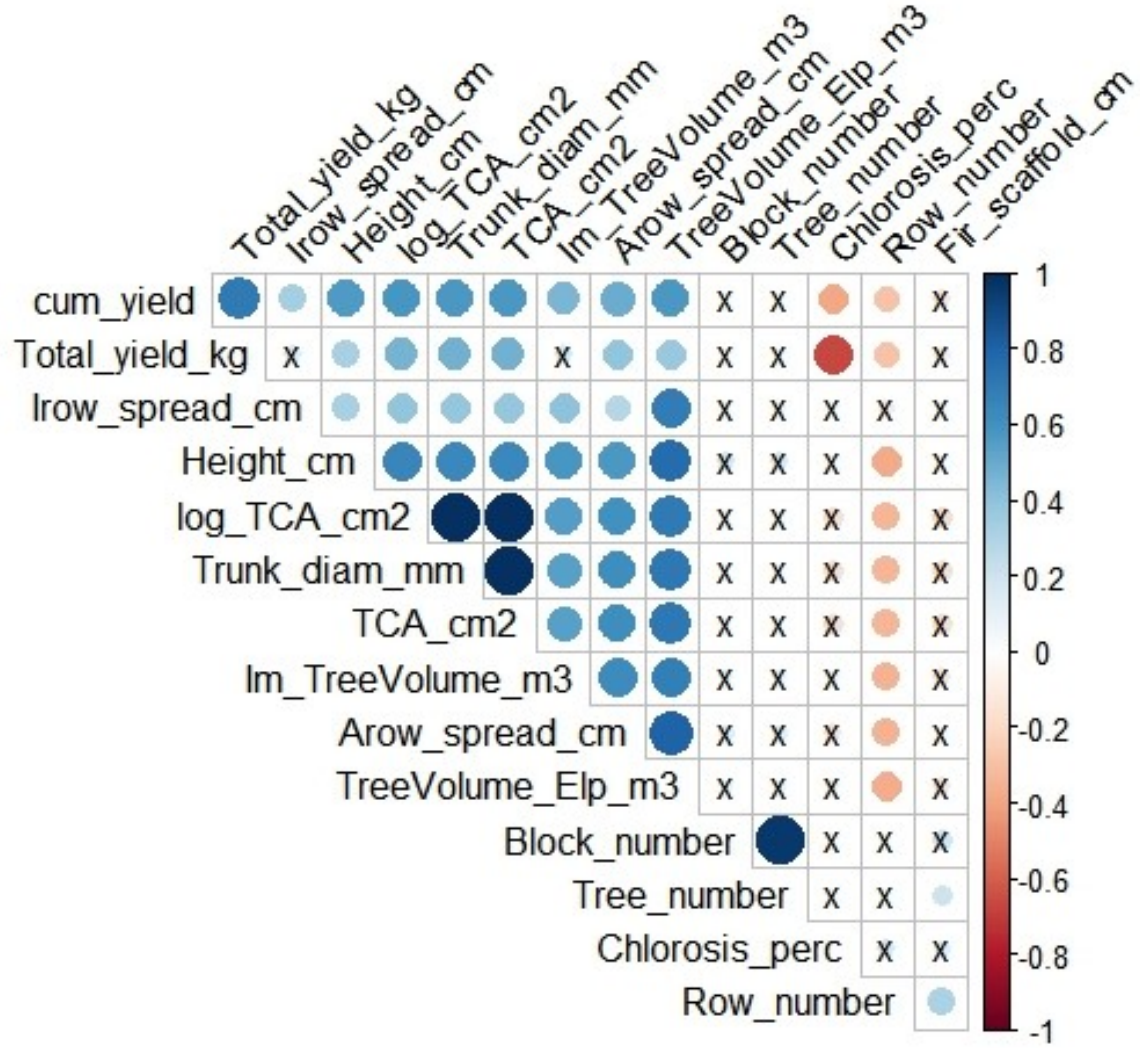
%Russet

This trait was initially rated on a 1-10 with the highest average rating at 6.5. A binomial trait was developed where a mean %Russet score of 3 and lower was considered low incidence, and 3.5 and above was considered high incidence. The high category was estimated correctly less than 40% of the time.

APPENDIX C TREE ARCHITECTURE DESCRIPTIVE PLOTS

The correlation of the tree architecture measures is shown in Figure C- 1. The correlations not significant at $p<0.01$ are marked with an "x".

Figure C- 1. Correlation plot of tree architecture measures.



_

The distribution of TCA values by 28 rootstock is show in Figure C- 2. Large dwarf, small semi-dwarf, and moderate dwarfing rootstocks are not easily distinguished from each other in this trial. Small dwarf, moderate semi-dwarf, and large semi-dwarfing trees do appear in better groupings except for PiAu-9-90, which seems to join the previous group of rootstocks.

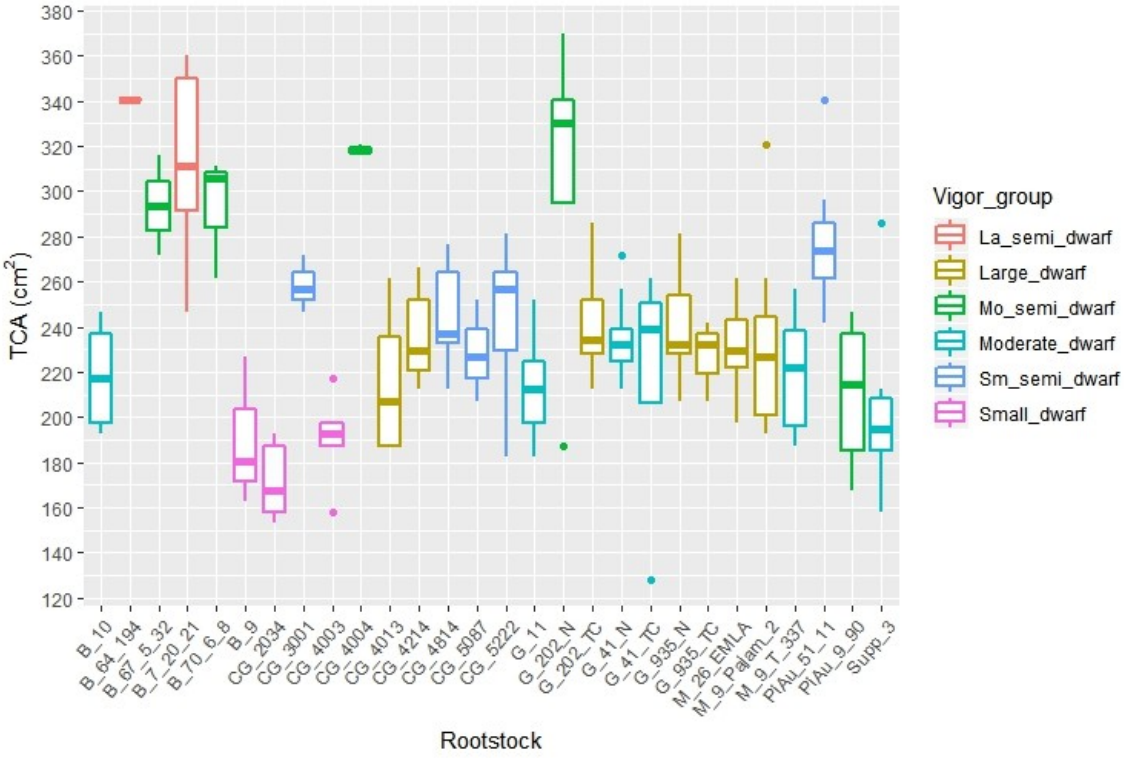


Figure C- 2. Boxplot of trunk cross sectional area (cm2) by rootstock.
Median = center bold line, box extremes = first and third quartiles, whiskers extend up to 1.5* inter-quartile range.
Color is from vigor group assignations by Autio et al. (2017). Abbreviated legend items La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

The distribution of tree height by rootstock is show in Figure C- 3. The pattern is similar with tree height as it was with TCA. There is one PiAu-9-90 individual with less than ordinary tree height.

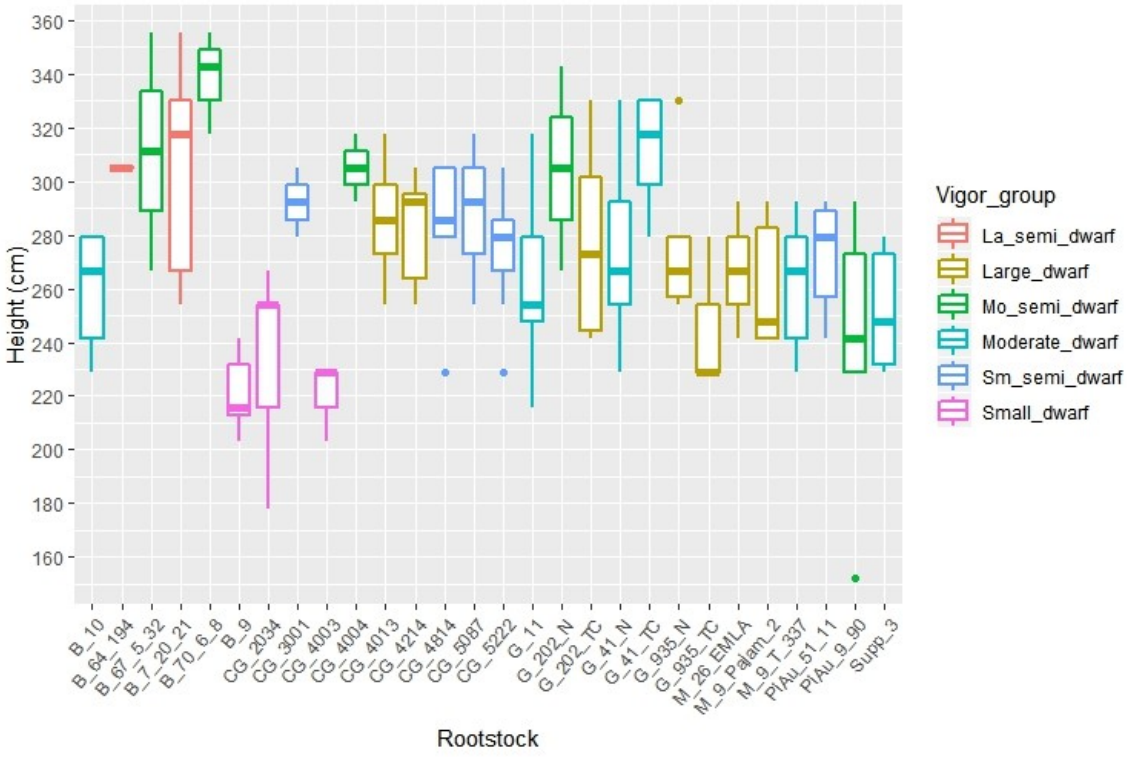


Figure C- 3. Boxplot of tree height cm by rootstock

Median = center bold line, box extremes = first and third quartiles, whiskers extend up to 1.5* inter-quartile range. Color is from vigor group assignations by Autio et al. (2017). Abbreviated legend items La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

There is much less of a pattern with the six year cumulative yield (Figure C- 4) compared to height and TCA. Small dwarf, Supp-3, and PiAu-9-90 are the only rootstocks that clearly appear different from the rest.

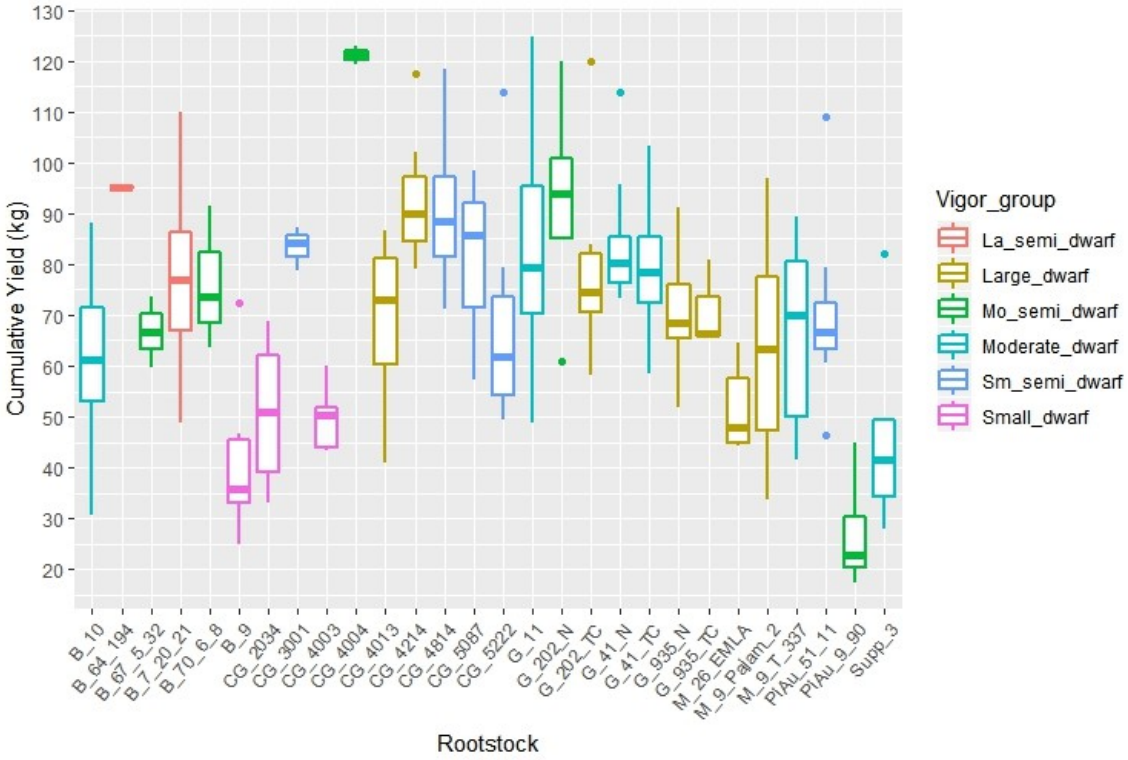


Figure C- 4. Boxplot of 6-year cumulative yield kg by rootstock.

Median = center bold line, box extremes = first and third quartiles, whiskers extend up to 1.5* inter-quartile range. Color is from vigor group assignations by Autio et al. (2017). Abbreviated legend items La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

The image-derived tree volume (m^3) follows a pattern similar to tree height except for an outlying PiAu-51-11 individual with larger volume, Figure C- 5.

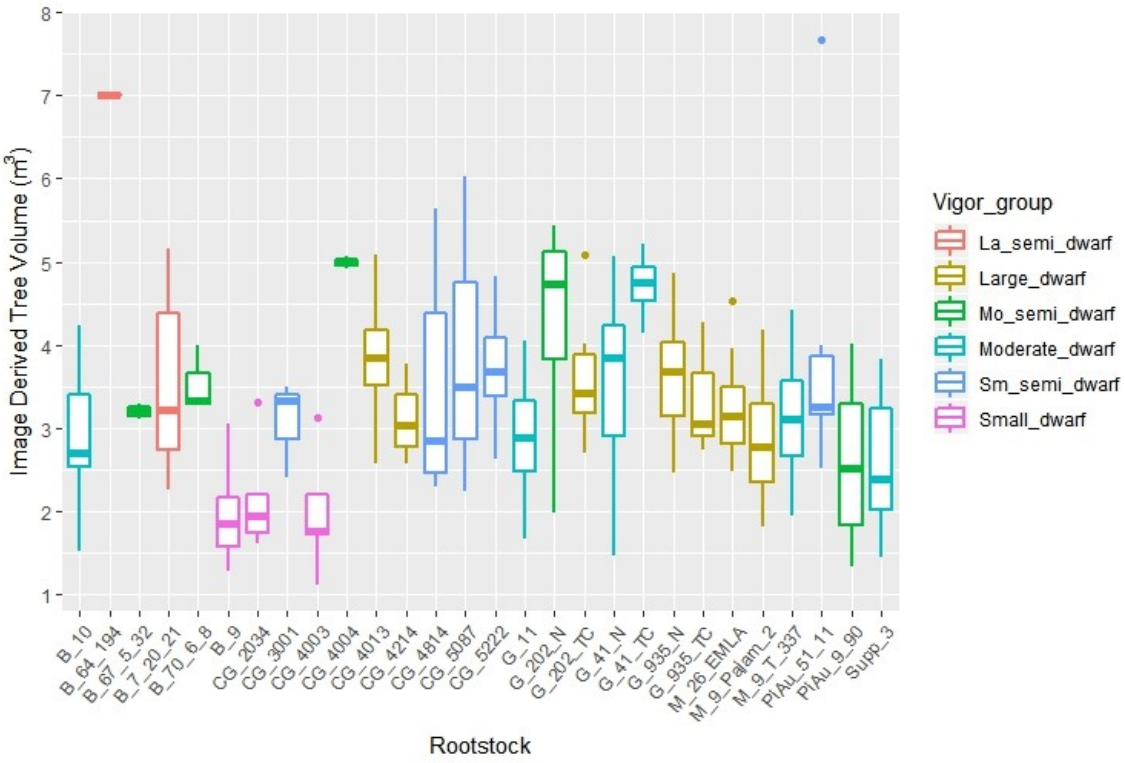


Figure C- 5. Boxplot of image-derived tree volume by rootstock.
Median = center bold line, box extremes = first and third quartiles, whiskers extend up to 1.5* inter-quartile range.
Color is from vigor group assignations by Autio et al. (2017). Abbreviated legend items La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

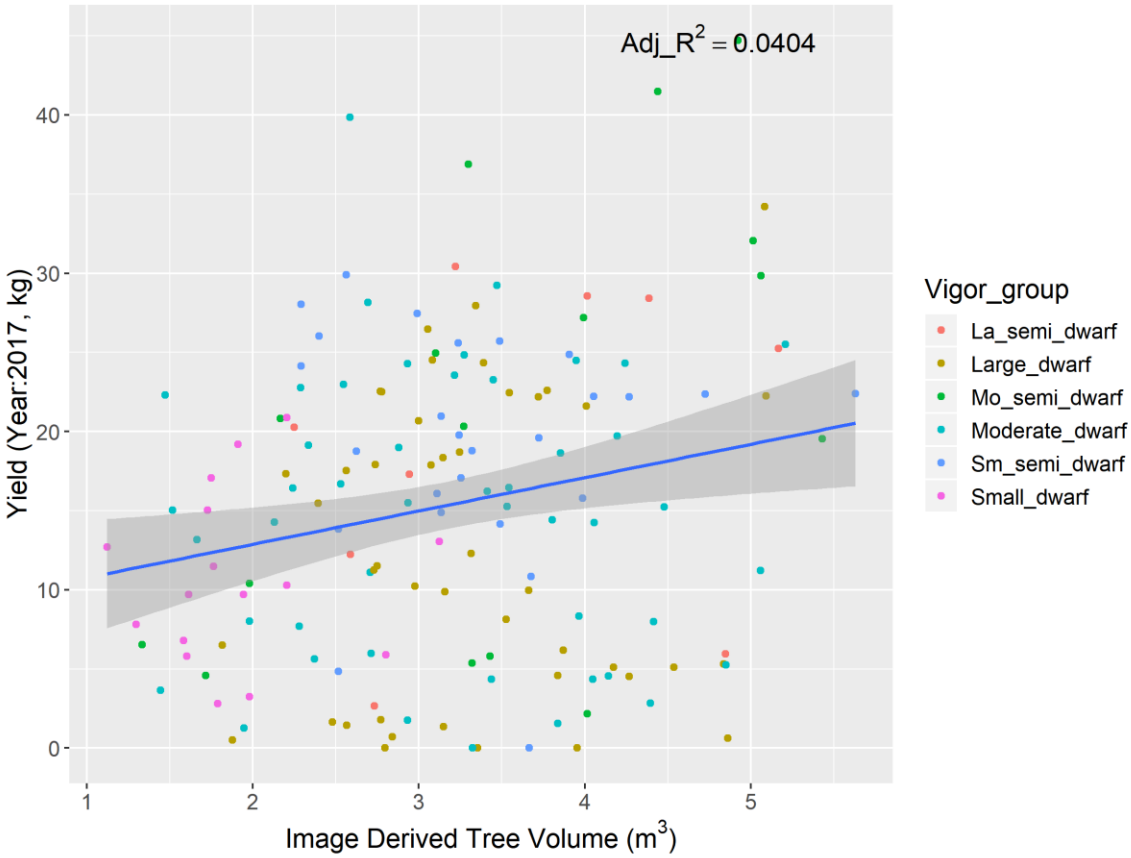


Figure C- 6. Yield of year 2017 by image-derived tree volume

Blue line is the calculated regression, while the gray shading is an estimated 95% confidence interval around the regression line.

Color is from vigor group assignations by Autio et al. (2017). Abbreviated vigor groups: La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

Figure C- 7. Yield of year 2017 by hand-measured trunk cross sectional area

Blue line is the calculated regression, while the gray shading is an estimated 95% confidence interval around the regression line.

Color is from vigor group assignations by Autio et al. (2017). Abbreviated vigor groups: La_semi_dwarf = large semi-dwarf, Mo_semi_dwarf = moderate semi-dwarf, Sm_semi_dwarf = small semi-dwarf.

APPENDIX D EXPERIMENT ON THE RIPENING OF THREE APPLE CULTIVARS USING THE SCiO AND F750 HANDHELD SPECTROMETERS (2017)

In this preliminary experiment, fruit were harvested from three cultivars at several time points during maturation. Spectral data were collected while the fruit were on the tree were compared to maturity and quality traits. The specific goals of this experiment were to 1) further test the use of spectrometers in an outdoor setting, 2) compare the predictive performance between two handheld spectrometers, and 3) test if fruit maturity could be predicted using the handheld spectrometers.

D.1 Methods and Materials

Fruit Samples and Harvest Protocol

In 2017, a weekly set of 25 fruit from each of the 'Honeycrisp', 'McIntosh', and 'Minneiska' cultivars at the UMN Horticultural Research Center were numbered 1- 25 while still on the tree. The fruit were scanned once (on the same location on the fruit) with two different spectrometers (F-750 and the SCiO). The surface temperatures of the fruit were recorded prior to scanning using an infrared thermometer. Fruit were harvested weekly for 5 to 6 weeks until fruit were past commercial maturity based on SPI score of eight. The fruit were then transported to the UMN St. Paul campus for further analyses done that same day.

Table D-1. Harvest dates, total harvests, and sample sizes per cultivar for the 2017 experiment

Cultivar	First Harvest	Last Harvest	Harvests	Total Fruit
'Honeycrisp'	Aug-21	Sept-23	5	125
'McIntosh'	Sept-4	Oct-12	6	150
'Minneiska'	Aug-10	Sept-14	6	150

Overall total fruit was 425

Spectral Measurement Protocol

The spectrometers used in 2017 were the Felix F-750, and the SCiO Molecular Sensor, a small handheld spectrometer sold by Consumer Physics (Tel Aviv, Israel). The SCiO spectrometer operates using active illumination from an LED light source. The spectrometer sensor is set back behind an optical filter and focusing element (Kaur et al., 2017), in an interactance optical design. The noteworthy differences of the SCiO spectrometer compared to the F-750 spectrometer are that it uses a narrower range in wavelengths, claims a finer resolution of 1 nm, costs significantly less, and is a fraction of the weight of the F-750 (see Table D- 2 for more details).

The battery life of the SCiO is sufficient for performing approximately 250 scans successively before needing to be recharged. However the battery life is shorter when scans are not immediately successive, therefore in this experiment, two SCiO instruments were used to acquire the required scans in the orchard. The SCiO comes with a magnetically attachable shade that is ~15 mm deep. However, to better control for spectral noise due to ambient light, a collapsible silicon cup (grey X-Cup, Sea To Summit, Boulder, CO) was modified to fit around the SCiO spectrometer (Figure D- 1).

Figure D- 1. Modifications for outdoor use of the SCiO spectrometer

Table D- 2. Comparison of spectrometer specifications

Spectrometer Specifications	F-750 Produce Quality Meter	SCiO (v1.2)
Dispersive element	Grating	Info not available
Optical geometry	Interactance	Interactance
Spectrometer Detector	Si (256) photodiode array ZEISS MMS1 VIS-NIR	Si 1.2M pixel CMOS (ON Semiconductor, USA)
Wavelength Range	310-1100 nm ±10 nm (265 variables)	740-1070 nm (331 variables)
Spectral Sampling (Resolution)	3 nm (8-13 nm)	<1 nm (resolution claimed to be less relevant)*
Repeatability (avg. wavelength SD)‡	0.0110 (reflectance units, >400nm)	0.0018 (reflectance units)
Illumination Source	Xenon Tungsten Lamp	White LED
Illumination area	~37 mm diameter spot	~20 mm diameter spot
Scan rate or Integration time	4-6 s	1.5 s
Thermometer	Internal instrument temperature	Not available (v1.2)
Calibration Standard	White painted shutter	White plaque inside device cover
Integrated GPS	48 Channel GPS	None
Battery Life	~1600 Scans	~250 Scans
Data exchange/storage	Wi-Fi/ Internal storage	Bluetooth
Weight	1.0500 kg	0.0318 kg
Research Cost (at time of writing)	~$7,000	~$1,500

*FAQ section of consumer physics website
‡ See section 0 for F-750 measurement protocol, and 0 for SCiO

Fruit Trait Data Protocol

The following skin traits were visually measured: Gcolor, number of scab infections (though difficult to differentiate from other damage), and the number of hail marks (due to a significant hailstorm June 11th, 2017). Texture analyses was performed using either the MDT-1 or MDT-2, at two locations on each fruit 180^{o} apart and just above the equator of the fruit. Immediately after the probe retracted for each side, juice from the probe entry was placed on a refractometer (Atago PAL-1 3810) to record SSC of juice

from individual sides (later averaged) of each fruit. After texture analysis, a transverse section (~12 mm) was cut from the equator of the fruit and placed on a labeled (1-25) sheet of butcher paper.

Each cut section was dipped in an iodine-potassium iodide (I_2-KI) solution (Blanpied and Silsby, 1992) and allowed to rest on the paper for approximately 5 min. Photographs were taken using a digital camera (PowerShot S100, Cannon) beneath two fluorescent tube lights (~1 m above fruit). The starch pattern index were visually scored following the 'McIntosh' 8-point scale (Blanpied and Silsby, 1992). The percentage of flesh stained for individual fruit was digitally calculated in the R software (R Core Team, 2018) from pixel color segmentations of the stained vs unstained proportions of the fruit. RBG values for training the R software segmentation algorithm were extracted from images using the Fiji Image J software (Schindelin et al., 2012).

D.2 Data Analyses

Fruit Trait Data Overview

Data were collected on 450 fruit across all cultivars and harvest weeks. The traits SSC, and percentage flesh stained (%Stained) with (I_2-KI) solution were quantitatively measured. The trait SPI was categorical or qualitative measured. Descriptive plots of the traits SSC, SPI, and %Stained are shown in below.

Spectra Data Preprocessing

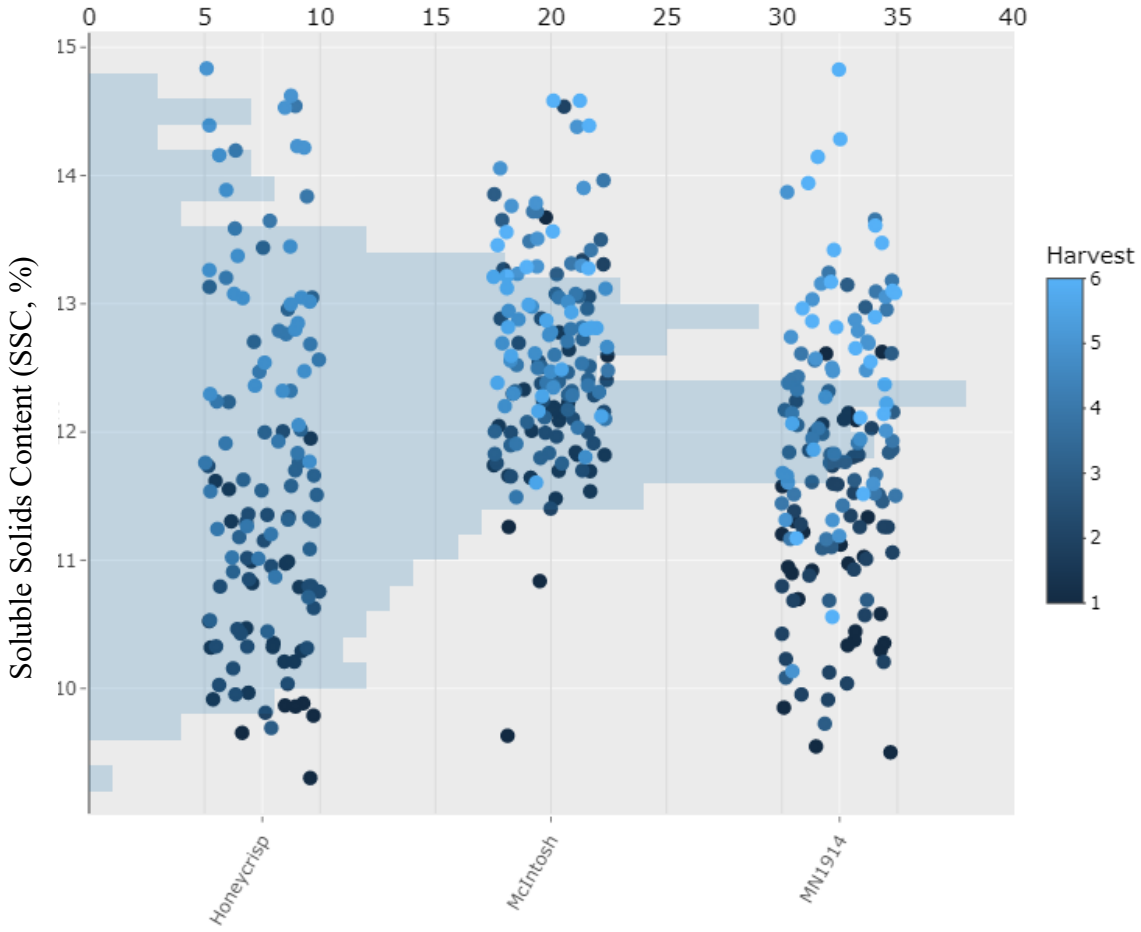
All spectral data from the F-750 spectrometer for this experiment were processed as in the experiment of Chapter 1. The raw SCiO spectral data were not preprocessed. No wavelengths from the SCiO were removed due to a high standard deviation (SD), as all SD were below 0.006. This decision was based on a plot of the SDs from repeatedly scanning several samples (see Figure D- 5).

Multivariate Analyses

The spectral data were analyzed in the same manner as in the experiment in chapter 1, using the same quantitative calibration/validation data divisions, as well as PLS algorithm

and settings. The qualitative data were also handled in the same manner as in chapter 1, except with the inclusion of extra variables, e.g., percentage of stained flesh.

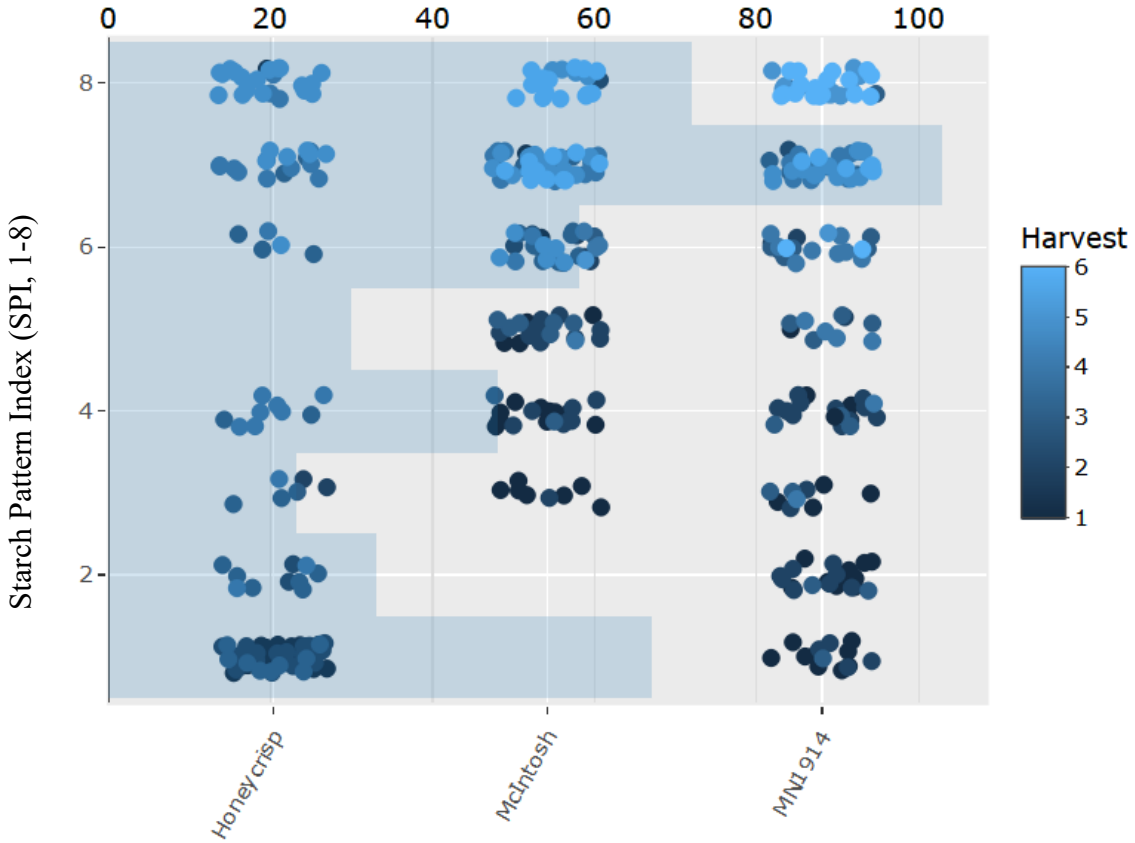
Figure D- 2. Distribution plot for SSC by cultivar and harvest week



The lower x-axis lists cultivars. Dots are jittered (horizontally and vertically) and colored according harvest week. The upper x-axis is the counts for the histogram distributions for the blue colored histogram bins. The y-axis is the respective value for each trait.

–

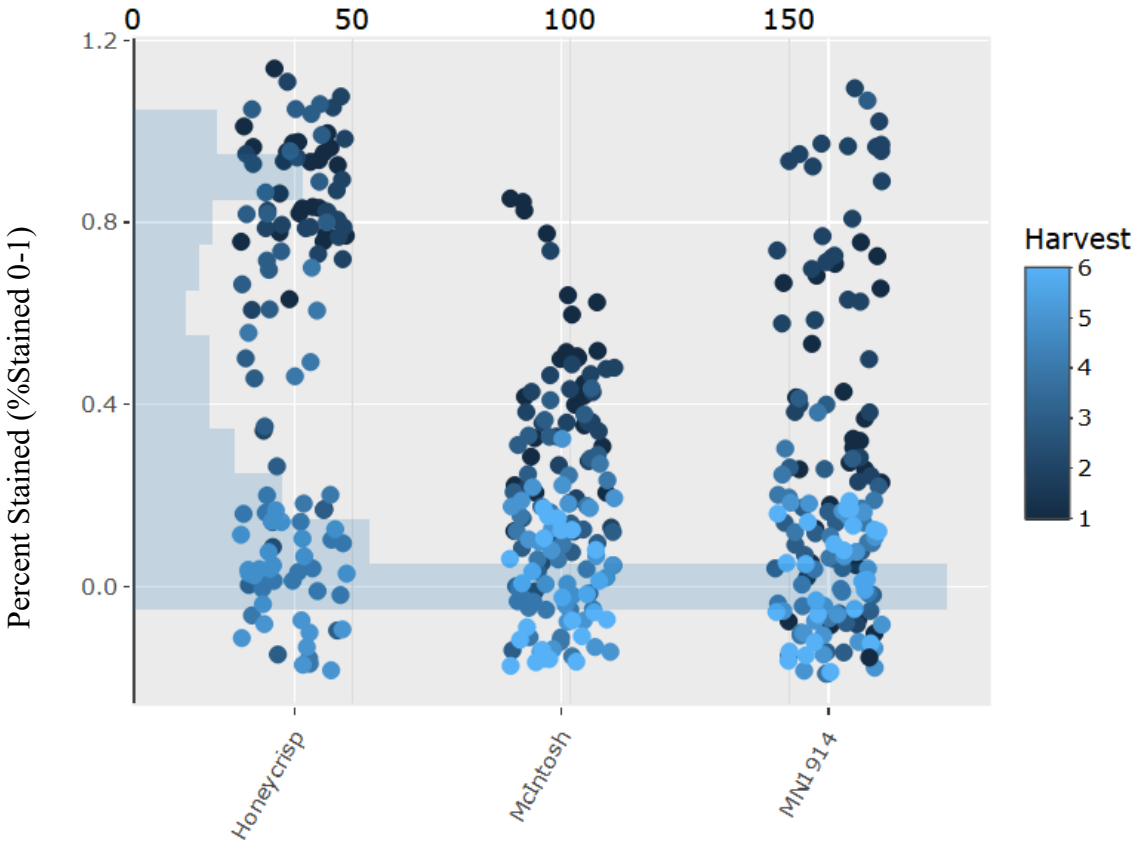
Figure D- 3. Distribution plot for SPI by cultivar and harvest week



The lower x-axis lists cultivars. Dots are jittered (horizontally and vertically) and colored according harvest week. The upper x-axis is the counts for the histogram distributions for the blue colored histogram bins. The y-axis is the respective value for each trait.

_

Figure D- 4. Distribution plot for %Stained by cultivar and harvest week



The lower x-axis lists cultivars. Dots are jittered (horizontally and vertically) and colored according harvest week. The upper x-axis is the counts for the histogram distributions for the blue colored histogram bins. The y-axis is the respective value for each trait.

–

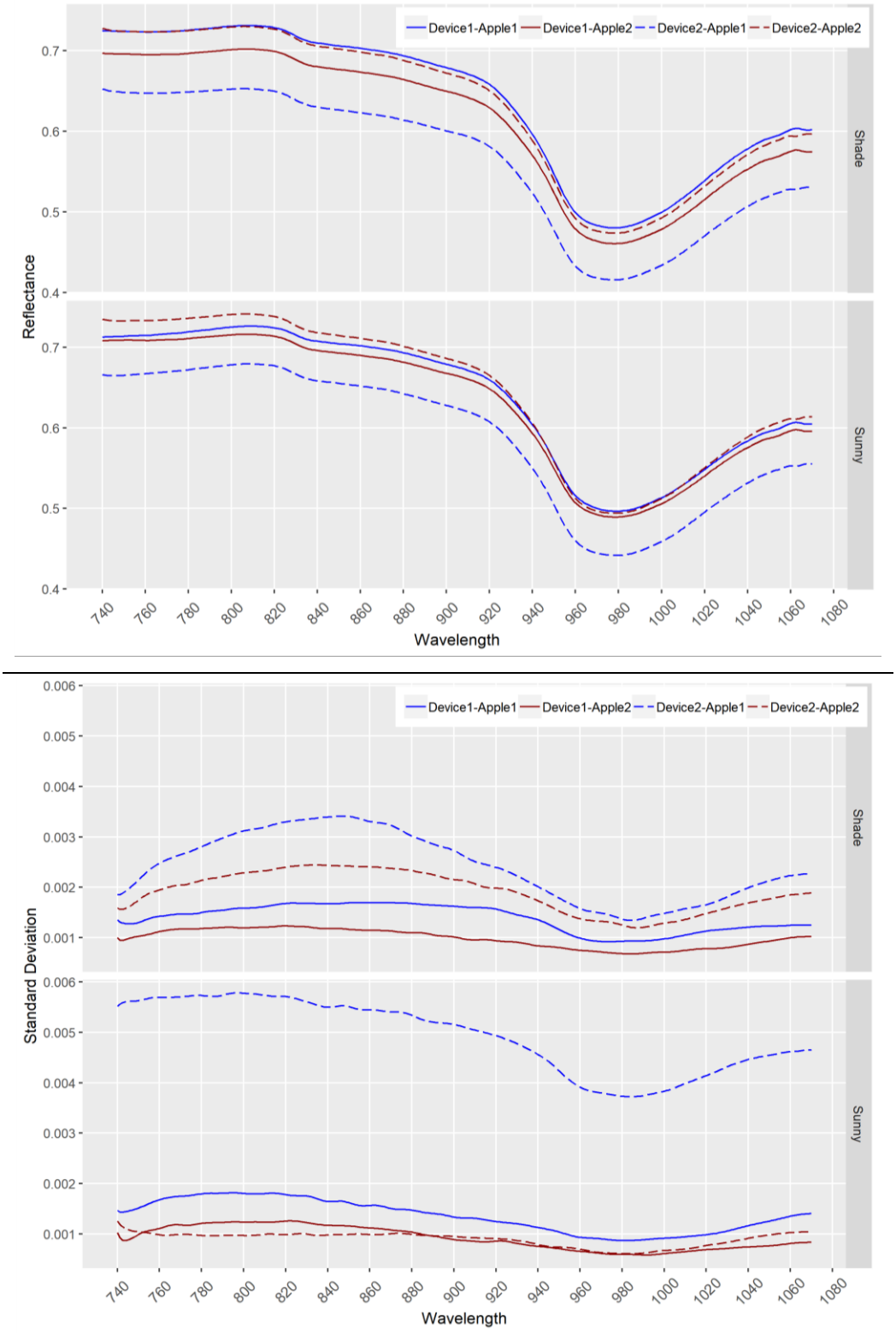


Figure D- 5. Averages and standard deviations of two SCiO sensors at each wavelength from ~50 repeated scans on two apple fruit.

D.3 Comparison of the Model Results for SSC, SPI, and %Stained from F-750 and SCiO Spectrometers

This experiment was designed to compare the predictive ability of models calibrated from spectra of the F-750 or SCiO spectrometers and to explore the estimation of maturity. Three traits related to maturity, SSC, SPI and %Stained, were assessed through a greater range of harvests per cultivar compared to the experiment in chapter 1. Descriptions of model results, by trait, for different calibration methods are presented in Table D- 4. Temperature effect is discussed, and the validation of SSC and SPI models from Chapter 1 with this preliminary experiment's data. Lastly, a brief overview, in a phenotyping context is given considering the image-derived metric %Stained and how it compares with SPI.

A fruit was scanned only once in this experiment, which contributed to much lower predictive abilities for fruit trait models compared to Chapter 1. The predicative abilities of trait models in this experiment were very low. SSC had RPD values between 1.19 and 1.35, SPI between 1.27 and 1.48, and %Stained between 1.25 and 1.38. The results suggest that there will be lower performance when using the instruments outdoors,

Table D- 3. Descriptive statistics for traits preliminary experiment

	SSC	SPI	%Stained
Calibration (N)	323	325	325
Validation (N)	108	108	108
Min	9.3	1	4.4E-05
Max	14.8	8	0.97
SD	1.1	2.5	0.34
Mean	12.1	5.0	0.28
Median	12.1	6	0.11

Table D- 4. Model calibration results for SSC, SPI, and %Stained from 2017

	Calibration ID	V	W	X	Y	Z
Calibration Parameters	Spectrometer	F-750	F-750	SCiO	SCiO	F-750
	Training Spectra	2017	2017	2017	2017	2016
	Validation Spectra	2017	2017	2017	2017	2017
	Fruit Temperature	No	Yes	No	Yes	No
	Spec. Temperature	No	Yes	No	Yes	No
	Training (N)	324	324	325	325	1096
	Test (N)	108	108	108	108	432
	Wavelength Range (nm)	501-1103	501-1103	740-1070	740-1070	501-1103
	Wavelengths (N)	151	151	331	331	151
SSC	Latent Variables	10	13	16	21	22
	RMSEC	0.81	0.81	0.90	0.91	0.92
	Mean Adj. R^2_{val}	0.41	0.45	0.31	0.31	0.15
	Mean Abs $Bias_{val}$	0.05	0.04	0.08	0.09	0.09
	RMSEP	0.84	0.82	0.93	0.93	1.22
	Mean RPD	1.32	1.35	1.19	1.19	0.90
SPI	Latent Variables	11.50	13.75	12.50	13.25	24.25
	RMSEC	1.69	1.70	1.94	1.95	1.19
	Mean Adj. R^2_{val}	0.51	0.54	0.39	0.38	0.17
	Mean Abs $Bias_{val}$	0.08	0.08	0.12	0.12	1.72
	RMSEP	1.72	1.67	1.93	1.95	2.95
	Mean RPD	1.44	1.48	1.28	1.27	0.84
%Stained	Latent Variables	11.50	13.75	11.00	12.00	
	RMSEC	0.25	0.24	0.27	0.27	
	Mean Adj. R^2_{val}	0.47	0.50	0.35	0.36	
	Mean Abs $Bias_{val}$	0.02	0.01	0.01	0.01	
	RMSEP	0.25	0.24	0.27	0.27	
	Mean RPD	1.38	1.42	1.25	1.26	

Except for the validation of Exp-1 models in calibration E, the results are based on the average of four calibrations using the four data splits as described in the methods. Training, test, and latent variables were rounded to whole integers.

F-750 vs SCiO Calibrations

The calibrations using spectral data from the F-750 (cal-V, Table D- 4) produced models with higher predictive abilities compared to the calibrations using spectral data from the SCiO (cal-X). The RMSEP for both SSC and SPI were both higher in cal-X compared to cal-V, but only SPI was statistically higher ($p=0.03$). For all three traits (SSC, SPI, %Stained) on average, the F-750 models had RPD values that were 0.16 higher than the SCiO models.

Considering the spectrometer designs, the F-750 trait models were expected to have higher predictive abilities than the SCiO models. For example, one reason for the superiority of the F-750 for SPI in this experiment could be the larger range of wavelengths used for F-750 calibrations including data from 501 to 1103 nm whereas the SCiO only records data from 740 to 1070 nm.

Temperature Effect

The addition of fruit and spectrometer temperature variables, comparing cal-W to cal-V and cal-Y to cal-X, did not have any effect ($p > 0.62$) on the predictive ability of the models. The range of fruit surface-temperatures for the experiment was 10 to 35 °C, thus perhaps by taking a representative number of samples at the different temperatures as other authors have suggested (Beghi et al., 2013; Kumar et al., 2015) helped to control for temperature effects in the model calibrations.

Validating F-750 SSC and SPI Calibration Models from Chapter 1

F-750 spectra and trait data from 2017 in this preliminary experiment were used to validate the spectral models developed from data collected in 2016 from Chapter 1. The calibration method validated was similar to cal-F, where both the average-indoor and average-outdoor spectra were used without any extra temperature variables. The only difference was that no samples from Chapter 1 were held out of the calibration resulting in a calibration size of approximately 1096 (Table D- 4).

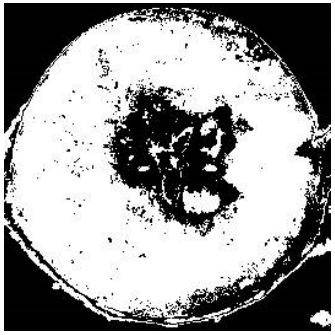
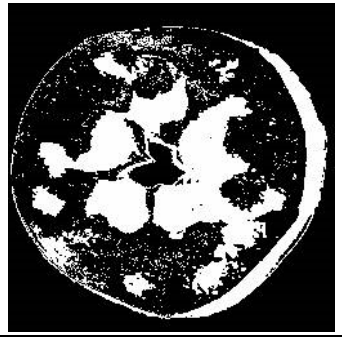
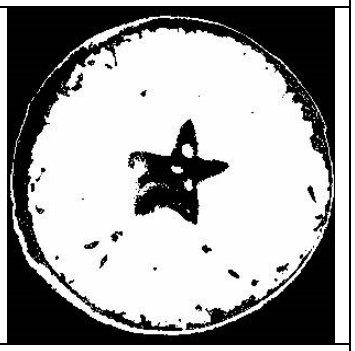
This experiment was not designed to collect data as a validation set for the experiment of chapter 1, but this analysis was done to understand the data sets' compatibility. The only

traits examined with this calibration method are SPI and SSC, both of which have very low predictive abilities in the validation. There were 432 fruit in the validation set from this preliminary experiment (2017), where the SPI and SSC trait models had RPD values of 0.84 and 0.90, respectively. The low results are not surprising since trait models for cal-Z were calibrated using only relatively mature fruit and they were used to predict values of fruits from five to six harvests with many immature fruit. Thus, the data collected in the two experiments were not similar enough for validation.

%Stained in a phenotyping context

Figure D- 6 shows the relative ability for the employed random forest model to discriminate between stained and unstained portions of the fruit cross section. The trained model and the display of the masked images could have been improved. For example, where the masked image for the first sample H2 #3 contains pixels that are colored nearly white, it is hard to distinguish whether this image was correctly classified. However, the values representing the nearly white pixels were supposedly designed to represent stained pixels. As a proof of concept, %Stained suits its purpose showing that this method is capable of approximating the percentage of cross section that is stained.

Figure D- 6. Original and masked images of ‘Minneiska’ starch patterns by random forest model

Harvest# ID	H2 #3	H3 #7	H4 #12	H6 #25
At-harvest image				
SPI (1-8)	2	4	6	8
Masked image				
%Stained (0-100)	88.34%	50.21%	9.96%	0.87%

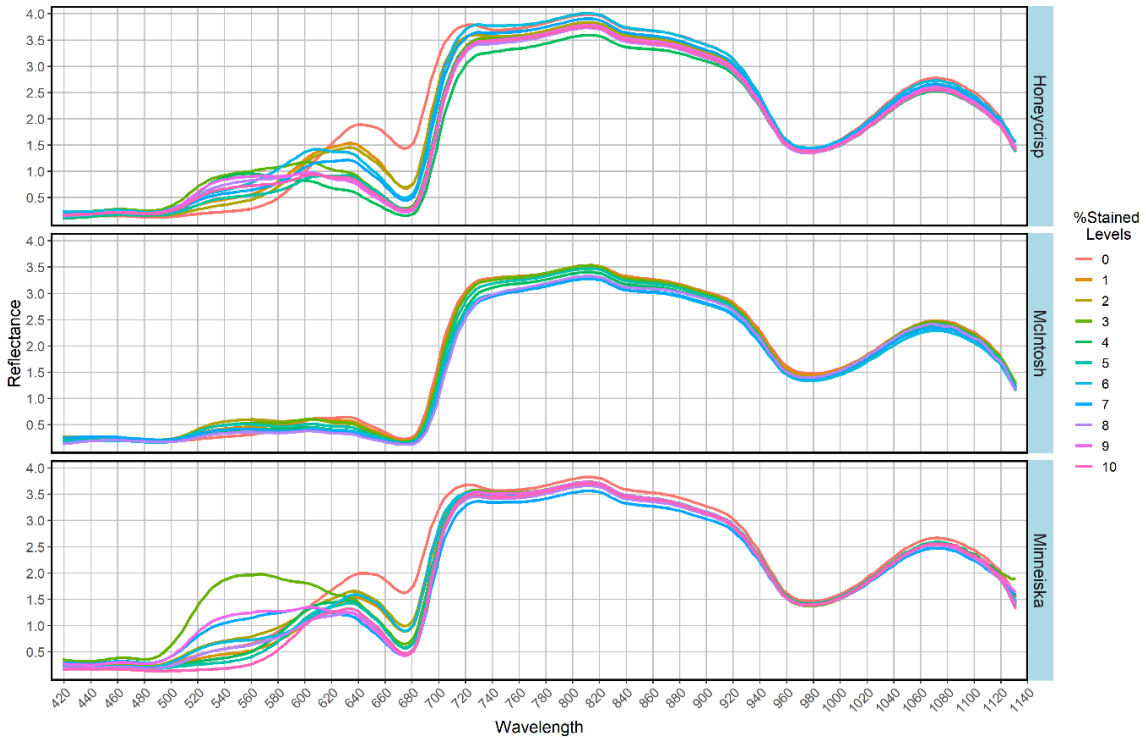


Figure D- 7. Comparing the average F-750 spectra for different %Stained levels by cultivar

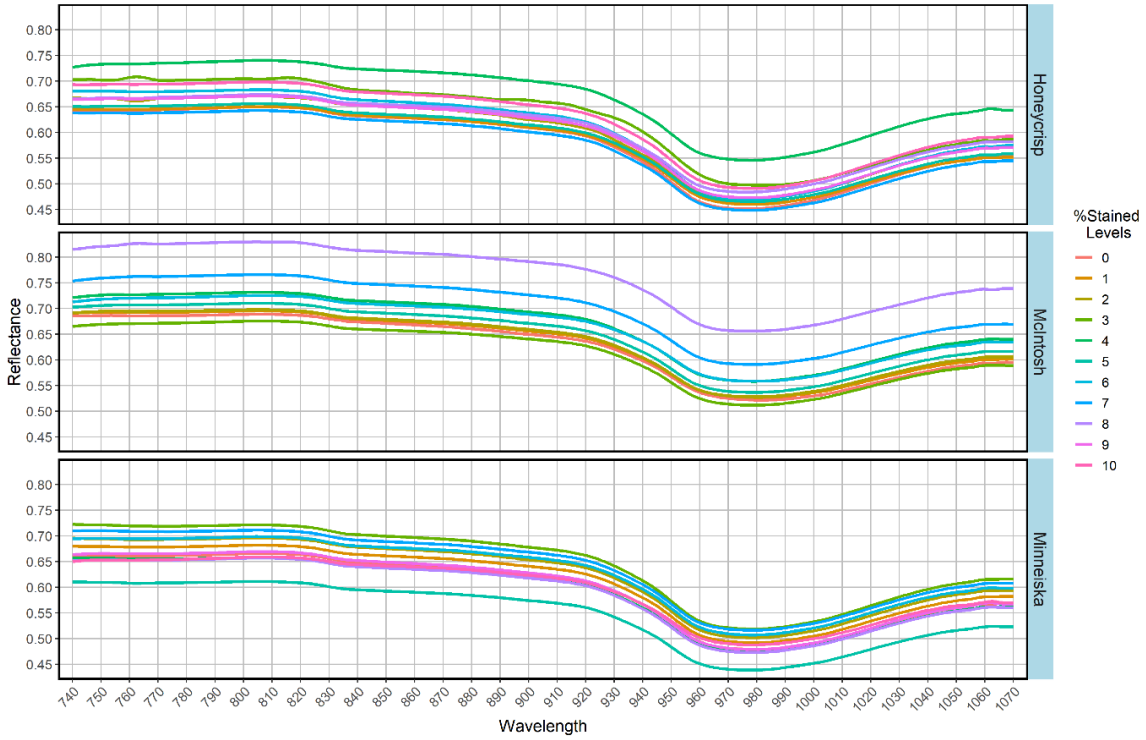


Figure D- 8. Comparing the average SCiO spectra for different %Stained levels by cultivar.

–

Maturity Estimation

Fruit maturity is linked to many quality traits; therefore, it would be valuable to estimate maturity rapidly and nondestructively. The models for SPI and %Stained had low predictive ability on a single fruit basis, but the spectra contained useful information. In Figure D- 7 and Figure D- 8, the change in average F-750 and SCiO spectra from grouped %Stained levels (converted to an integer scale of 1-10) are shown. In Figure D-7, the reflectance patterns from the F-750 spectra are similar between cultivars, but the rates of change are markedly different at specific wavelengths. McIntosh is a cultivar that does not lose the green ground-color with increasing maturity. This is obvious in its spectral patterns since the changes with %Stained levels is very minimal compared to those of 'Honeycrisp' and 'Minneiska'.

Herold et al. (2005) presented a similar figure to Figure D- 7 with spectral data from 'Elstar' fruit. They showed that as fruit matured reflectance increased at approximately 670 nm due to decreasing ground color, and the reflectance at approximately 570nm decreased due to increasing over-color. Figure D- 7 shows that the pattern is visible using spectra from the F-750 (420-1130 nm). The spectra from the SCiO spectrometer (Figure D- 8), however, is of a reduced wavelength range (740-1070 nm) omitting the spectral changes relative to the visible spectrum. The only clear pattern among SCiO spectra in %Stained levels is with the McIntosh cultivar as it appears to have increasing reflectance for all wavelengths with increasing %Stained, but with the other two cultivars, any pattern is more difficult to distinguish.

This pattern may be explained further in Figure D- 9 with the distribution of %Stained observations over harvest weeks. There is a clear pattern with 'McIntosh', but with the other two cultivars, the earlier weeks do not follow the expected pattern of decreasing %Stained. Thus for future research it is important to plan data collection specific to a cultivar's rate of maturation. Enough fruit needs to be collected with increased number of harvests to capture the range of maturity especially during the immature to mature transition. Biological differences in starch distribution may also be different between cultivars. For example, Szalay et al. (2013) was able to assign 24 cultivars to four types

of starch patterns. The accuracy of a multi-cultivar predictive model may not have the accuracy as a cultivar specific model.

_

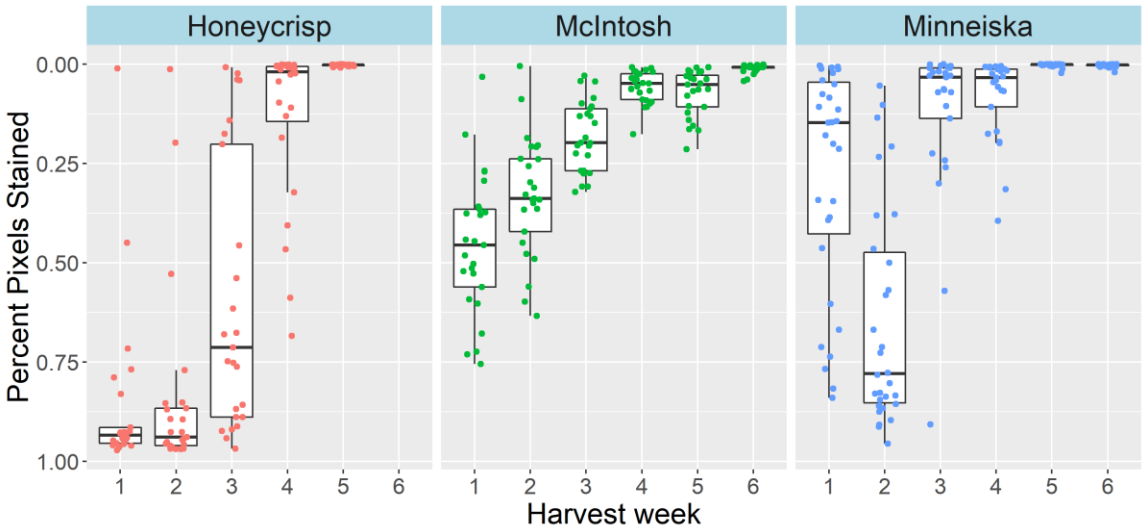


Figure D- 9. Boxplot of %Stained by harvest week by cultivar.

The y-axis is reversed to represent the natural direction of change in %Stained, compare to Fig. 5 of Piers et al. (2002).

D.4 Summary

The overall performance of both spectrometers in capturing predictive spectra from a single outdoor scan was low (RPD <1.48). More data and analysis would be needed to thoroughly compare the devices. The outdoor location where the spectra were acquired was likely a contributing factor, but the results of Chapter 1 suggest that the major factors to the low performance are 1) the insufficient spectral information per volume of fruit, and 2) phenotypic data that was not specific to the exact scan position on the fruit. To more comprehensively test and compare the two spectrometers, it is proposed to do full fruit juicing and collecting two or more scans per fruit.