THESIS

IMPACT OF VARIOUS FACTORS ON PARTIAL LEAST SQUARES MODEL ROBUSTNESS FOR NONDESTRUCTIVE PEACH FRUIT QUALITY ASSESSMENT

Submitted by

Jakob Pott

Department of Horticulture and Landscape Architecture

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2023

Master's Committee:

Advisor: Ioannis Minas

Joe Eakes
Matt Koslovsky

Copyright by Jakob Pott 2023

All Rights Reserved

ABSTRACT

IMPACT OF VARIOUS FACTORS ON PARTIAL LEAST SQUARES MODEL ROBUSTNESS FOR NONDESTRUCTIVE PEACH FRUIT QUALITY ASSESSMENT

Given declining fruit consumption due to poor fruit quality and large amounts of waste, peach growers have continuously suffered from financial loss and the industry has seen a sharp decline in recent decades. Due to the time consuming and destructive nature of conventional fruit quality assessment, many peach growers prioritize fruit characteristics conducive to shipping and storage over characteristics which correlate with consumer acceptance. This prioritization has resulted in the poor-quality fruit which consumers have grown to associate with fresh peaches and contributed to large annual waste. A potential solution is the use of near-infrared spectroscopy (Vis-NIRS) paired with partial least squares (PLS) modeling, as a field deployable method that can be used to measure preharvest internal fruit quality to produce information quickly and non-destructively. These qualities offer an answer to declining fruit quality and waste. Although promising, the technology is only as good as the data used to train the models. Quality data is hard to collect as it requires the consideration of many factors including the temperature of the sample and the inclusion of biological variability impacted by seasonal changes, cultivar differences, fruit maturity, and many management factors such as crop load, rootstocks, irrigation regimes, and training systems to capture the relationships needed for good model performance.

In tree fruit research, handheld Vis-NIRS devices have been used to predict internal quality parameters such as sweetness (dry matter content, DMC; soluble solids concentration, SSC) and fruit physiological maturity related to chlorophyll content (index of absorbance difference, I_{AD}).

Although accurate, the statistical models used to make such predictions often struggle with robustness across cultivars and growing seasons and regions due to a lack of biological variability, or a lack of representative data from factors like temperature. These challenges have led to slow industry adoption. To address this issue, models for 13 distinct peach cultivars were constructed by combining data from two seasons (2016 and 2021) followed by external validation with data from a third season (2022). The data from 2016 was collected over a range of preharvest factors, fruit development stages and temperatures, and the inclusion of 2021 data added additional biological variability. External validation produced error rates of 0.36 - 0.42%, 0.59 - 0.63%, and 0.05 - 0.04 for DMC, SSC and I_{AD}, respectively, across the 13 peach cultivars indicating the models trained in 2021 were robust and performing at an acceptable level to impact grower decision making. It was observed that the additional inclusion of data from different cultivars and growing environments, as well as a third growing season (2017) did not significantly impact model performance. The lack of improvement suggests that the data from each year contain enough covariate variability to cover a broad range of measurements (i.e. input values) that growers and researchers are likely to observe when collecting data to predict peach quality in different orchards or seasons. This insensitivity to various environmental and growing conditions, generally referred to as external factors, due to the variability captured in the data used to build model is characteristic of a robust model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

BACKGROUND INFORMATION

1.1.Introduction

Consumption of peaches is driven largely by the consumers' sensory experience when eating the fruit. Peach consumption rate has seen a decline in recent decades linked with reduced fruit quality. Fruit characteristics that are linked to consumer acceptance include soluble sugars, organic acid concentration, and flesh texture. These attributes are traditionally quantified with methods that are slow and destructive. Due to the time consuming and inherently wasteful nature of destructive quality analysis, the use of visible-near-infrared spectroscopy (Vis-NIRS) has been explored as a non-destructive alternative. Research has been conducted to determine the efficacy of this technology for non-destructive measurements as well as large data acquisition to aid in decision making throughout the fresh fruit supply chain offering applications to growers, packers, and distributors.

Since each fruit quality component contains chemical bonds which are uniquely responsive to various electromagnetic wavelengths, they can be measured indirectly by way of spectral absorbance measurements. Electromagnetic radiation that is administered to a sample from a light source, which when paired with a spectrometer to measure the amount of light reflected, transmitted, or absorbed, can produce information of the sample composition based on the response of the water, lipids, organic acids, and carbohydrates to the various wavelengths applied from the light source.

There are several specific fruit quality indices commonly referenced in the literature due to the relationship they have with consumer acceptance which characterize the quantity of carbohydrates, acids, pigments, or other constituents. Some of the most important quality indices include dry mater content (DMC), soluble solids concentration or total soluble solids (SSC and TSS) or °Brix and titratable acidity (TA). Maturity is directly linked to quality and is a critical parameter for decision making throughout the fresh fruit supply chain from harvest and storage to shipping and retail marketing decisions. Maturity has been quantified with different indices and measurements including fruit flesh firmness (FF), background color and index of absorbance difference (I_{AD}). For apple, where the relationship between maturity and quality is similarly important to peach, the Streif Index, and a ripening index (RPI), two multiparameter indices characterize the relationship between fruit firmness, acidity, and sugar content with a single value. These indices are a composite of various physical and chemical properties representative of fruit genotypes and the environments they are grown in. Since these indices encapsulate so many different attributes that exist at the cellular and environmental levels, it becomes very difficult or impossible to measure each contributing factor at once. Vis-NIRS technology has the ability, when coupled with advanced statistical modeling techniques, to make indirect measurements of these metrics and indices of interest to producers, consumers, or researchers (e.g., DMC or SSC), which can act as a proxy for those attributes. In this way the slow and destructive direct method of measurement is replaced with the fast indirect Vis-NIRS method. The ability to make these measurements and collect large amounts of data has opened the door for the exploration of the relationships between factors that constitute fruit quality metrics/parameters, and the relationships those factors have with various spectra.

1.2. Fruit Quality and Maturity Parameters of Interest

Depending on the fruit crop, the primary parameters of interest used to characterize fruit quality and maturity vary. In general, the parameters of interest focus on quantifying carbohydrates in the fruit, since these are what consumers primarily detect when they taste the fruit. The carbohydrates of interest vary between fruit crops, but their chemical signatures for detection, both destructively and non-destructively, are often similar. For different fruit species, these parameters are complexly related to one another, and their relationships to cultural practices in the orchard, vineyard, etc., are also convoluted. A desire to understand these complex relationships has led to extensive research, probing the impact of different preharvest factors on fruit quality to maximize quality and efficiency in the orchard (Minas et al., 2018). These research goals have yielded information regarding how quality attributes can vary between species and between cultivars. As these studies are conducted, a profile of the fruit comes into focus as values are assigned to fruit considered mature or immature, of high quality or low quality. Not only is a quality/maturity profile developed, but also a better understanding of the specific conditions which produce fruit that fit those categories. DMC is described at length by Palmer et al. (2010) as a metric for fruit quality linked with consumer acceptance, due in part to the high correlation DMC has with soluble sugars in the fruit. Soluble sugars constituting roughly 80% of the dry matter in many apple cultivars (Palmer et al., 2010). Because of this, it is also noted that the general trend is higher DMC is met with greater consumer acceptance.

There are now many original research studies and review papers that exist exploring fruit quality and maturity and their measurement (Anthony et al., 2023; Costa et al., 2002; Crisosto et al., 2018; Lin and Ying, 2009; Minas et al., 2018; 2021; Nicolai et al., 2007; Palmer et al., 2010; Wang et al., 2015; Saeys et al., 2019). Of these studies, fruit quality is described for mangos (Anderson et al., 2020; Rungpichayapichet et al., 2016; Sun et al., 2020), pears (Li et al., 2019; Mishra et al., 2021; Wang et al., 2017), cherries (Overbeck et al., 2017), grapes (Zeiter et al., 2006), plums (Louw and Theron et al., 2010), apples (Kumar et al., 2015; Luo et al., 2018; Palmer et al., 2010; Peirs et al., 2003a; 2003b; 2004; Teh et al., 2020; Zhang et al., 2019), and nectarine/peaches (Anthony et al, 2023; Costa et al., 2002; Fu et al., 2008; Minas et al., 2018; 2021; 2023; Mukarev and Walsh, et al., 2012; Nascimento et al., 2016; Uwadaira et al., 2018; Ziosi et al., 2008), where quality is either directly measured or predicted using some combination of statistical modeling and spectroscopy. Most of this work is primarily concerned with quantifying carbohydrates such as simple sugars and/or starch or DMC (Anderson et al., 2020; Minas et al., 2021; 2023; Palmer et al., 2010; Sun et al., 2020; Teh et al., 2020; Walsh et al., 2004; Zhang et al., 2019) and SSC (Li et al., 2019; Luo et al., 2018; Mishra et al., 2021; Mukarev et al., 2012; Nascimento et al., 2016; Peirs et al., 2003a; 2003b, Walsh et al., 2004; Wang et al., 2017; Zhang et al., 2019). DMC refers to everything in the fruit flesh that is not water and is calculated as the percentage of dry matter over total fresh weight (Palmer et al., 2010). SSC, measured as ˚Brix, is an estimate of the percent of soluble sugars in a given sample of fresh fruit juice. These studies use these parameters both to indicate the impact of various treatments in their studies (e.g., crop load studies in apple and peach) (Grossman and DeJong, 1995; Anthony et al., 2020) and to explore the use of these parameters as predictors for storage potential and consumer acceptance (Palmer et al., 2010; Crisosto and Costa, 2008). Conversely, some studies have focused on moisture content in tandem with SSC as a quality parameter (Mishra et al., 2021). Quantifying the moisture content of a fruit shares the same fundamental approach as measuring DMC, both destructively and non-destructively. Where DMC is the percent of dried fruit tissue over the fresh weight of that sample, moisture content would be determined as 1 – (dry weight/fresh weight).

Another key attribute used to characterize fruit quality and consumer acceptance is organic acid content measured and reported as titratable acidity (TA). Although important for characterizing the flavor profile of a fruit, the non-destructive measurement of organic acids is less common in the literature compared to sugar content. Acid content is often used to help explain the impact of preharvest factors on final fruit quality and is also used to help determine an appropriate maturity index for harvesting non-climacteric fruit like citrus and grapes (Laminakra et al., 1995).

Other important studies that have been conducted on fruit are those concerned with fruit maturity, often described as ripeness. Maturity is inherently more difficult to quantify as it is not the accumulation of one thing, but generally observed as the development of reproductive physiological functions, which in peach, include several distinct stages. Examining peaches as an example of fruit development, the stages of development post-bloom follow an early establishment of cellular material (stage S1: cell division), seed development and endocarp lignification (stage S2: pit hardening), and then the final stages of development where fruit size increases, background color develops in the skin (stage S3: cell enlargement), and the fruit firmness begins to decrease until commercial harvest (stage S4I: pre-climacteric maturity) or 'on-tree ripening' stage (S4I: climacteric maturity) (Minas et al., 2023). Throughout these stages, carbohydrates are accumulating, and the concentration of organic acids decreases while other flavor compounds and secondary metabolites accumulate (Minas et al., 2023). Since the definition of fruit maturity includes the shift of many different fruit qualities, studies have delved into defining and measuring maturity based on different combinations of these traits. These combinations are then reported as indices which provide a more comprehensive summary of fruit development and include RPI, a ripening index representing fruit firmness, titratable acidity, and soluble sugars (Rungpichayapichet et al., 2016), and the Streif index, a combination of starch, sugars, and firmness levels to determine apple maturity (Peirs et al., 2005), or the difference in chlorophyll-A and chlorophyll-B absorbance of visual light beneath the fruit skin as the index of absorbance difference (I_{AD}) (Ziosi et al., 2008; Costa et al., 2009; Minas et al., 2021). These studies have helped elucidate the role of maturity in the development of the quality parameters listed above as they better summarize the interplay of those parameters during development, across the season. It is clear that these parameters of interest are tightly linked with one another and the manipulation of one of them often has a profound effect on others. Anthony and Minas (2022) explained the need to account for the confounding effect of fruit maturation by controlling for equal fruit maturity in the studies that aim to determine the direct impact of preharvest factors on quality parameters. In addition, understanding fruit development and maturation is critical for growers as the parameters which define fruit maturity lend themselves to the prediction of optimal harvest time and yield (Li et al., 2017). Fruit firmness is used as a fruit maturity index as well as an indicator of quality. Studies which explain the cross-section of maturity and quality include Anthony et al. (2020; 2021), Fu et al. (2008), Nascimento et al. (2016), Uwadaira et al. (2017), Wang et al. (2017), elucidating the relationship between the two.

Ultimately, these parameters are important to researchers for the sake of quantifying the impact of different treatments in preharvest fruit research studies, but also for growers who depend on the quality of their fruit to maintain the commercial viability of their produce and the longevity of their industries. Many of the methods for quantifying fruit quality and maturity are time consuming. For this reason, research has been conducted to determine alternative means of measurement. Applications of visible-near-infrared spectroscopy (Vis-NIRS) has been one of the most popular methods, particularly when paired with partial least squares modeling both of which will be the focus of the remainder of this chapter.

1.3. Chemometrics

Chemometrics is a discipline of science that utilizes mathematical and statistical approaches to extract information from the complex chemical composition of various samples. Chemometrics has recently become more popular in fruit research due to the arrival of accessible equipment, hardware, and software with the computational power needed to process the large amounts of data produced in these studies (e.g., spectral data). Chemometrics as a discipline includes both the data preprocessing step and statistical modeling portion of experiments.

The studies listed in this chapter describing the interaction of spectra, fruit tissues, and specific parameters connected to both, are all chemometric studies. Relevant to this body of work, experiments focused on the use of the visible and near infrared spectrum (380 – 2500 nm) to detect and estimate the concentrations of various compounds in fruit tissue, and particularly the extraction of meaningful information from the collected spectral data, are examples of chemometrics. An example of a chemometric application is in the detection of O-H bonds, common in fruit tissue due to the abundance of water and carbohydrates. These dominant and broad peaks produce a low signal to noise ratio when detected, making it more difficult to identify chemical structures (Wang et al., 2015).

One of the challenges of modeling spectral data is its high dimensionality. Depending on the spectral resolution of the spectrometer being used, over 1,000 measurements can be collected for each sample. This creates the need for various statistical modeling techniques, including dimension reduction, and requires high computational power for data processing and inference. In regression modeling, high dimensionality typically refers to a dataset in which the number of covariates (p) is greater than the number of samples (n) (i.e., $n < p$) from which observations were made for each covariate of interest. This poses mathematical issues for traditional linear regression models. Other common statistical challenges in chemometric research are high collinearity among the covariates, non-linear relationships between the outcome and covariates, and potential overfitting. In predictive settings, it is imperative to build generalizable models that avoid overfitting. To address these challenges, chemometrics researchers typically apply PLS for developing predictive models (Nicolai et al., 2007).

1.4. Partial Least Squares (PLS)

Partial least squares (PLS) regression is a form of multivariate statistical modeling that is particularly well suited for modeling large amounts of often highly correlated, high dimensional chemometric data (Wold et al., 2001). This modeling technique, which was developed in the 1970's, was created to better handle issues of complexity and dimensionality in data from studies concerned with chemical composition of various samples. It is often the case in horticultural studies that chemical structure and quantity are of interest, carbohydrates in fruit tissues as an example, where spectral data is measured as an indirect measurement of sugar content. Large amounts of highly correlated data are characteristic of such spectral data due to broad overlapping peaks produced by the absorbance of different compounds present in the fruit tissues, particularly the carbohydrates and water (Saeys et al., 2019). Much of the value of this data and the ability to model it accurately is in the application of models for prediction, particularly when coupled with spectrometry as a non-destructive data generating technique used on intact fruit tissues. In the case of horticultural studies this is centered on gathering information from fruit flesh, the portion of interest to consumers.

As a statistical method, PLS is often compared to principal component analysis (PCA) based on the way they each decompose the variance in a given matrix of observations. PLS goes further than PCA however, since PCA explains the relationship between covariates, represented as an n x p matrix, $\mathbf{X}$, where n is the number of observations and p is the number of covariates, and PLS explains the relationship between covariates as well as the relationship with the outcomes. Outcome values are represented as $\mathbf{Y}$, a n x m matrix where n is the number of observations and m is the number of outcome variables. In PLS, **Y** can either be a single outcome variable or multiple variables. This allows for the detection of relationships in the data which are impactful to the outcome variables captured as latent variables (see below) (Saeys et al., 2019). In practice, PLS can capture information about how the chemical structures being measured interact with each other and combine to yield measurements of the outcome of interest.

Briefly, the non-linear iterative partial least squares (NIPLS) algorithm is used in PLS for estimation of relationships in the data as bivariate regression slopes (Wold et al., 2001). NIPLS iteratively regresses each column of the **X** matrix onto the column of outcome values. When there is a single outcome variable, the **X** matrix column values are regressed onto values for that variable and the process ends. When there are multiple outcome variables, X-scores (initial regression coefficients from the first regression) are used to produce Y-weights when columns of the **Y** matrix are regressed onto the X-scores and subsequently Y-scores are calculated. The new Y-scores are used to produce new X-weights, and the process is reiterated until changes in the X-scores become sufficiently small (a difference of 10^{-6} or 10^{-8} (Wold et al., 2001)).

This iterative process is the first step in "deflating" **X**. After the scores and weights are determined, a new matrix (the same dimensions as the initial **X)** of predicted values is estimated and the estimates are subtracted from the original values. This removal of predicted values is what it is meant when **X** is deflated. After several rounds of deflation, it is assumed that all the relationships in the data are captured, and thereafter this process begins to model noise in the data. The modeling of noise is monitored and quantified as Q^2, the cross-validation statistic estimating how predictive the model is. Modeled noise can be observed when Q^2 fails to approach 1 and instead begins to get smaller. Each time **X** is deflated after multiple rounds of calculating scores, weights, and loadings, a latent variable (LV) is extracted and stored. The number of iterations required to produce the highest Q^2 is the number of LVs in that model (Saeys 2019; Wold et al., 2001). LVs derived from PLS are analogous to principal components in PCA but are not considered principal components since they capture covariance in **X** and **Y** (Saeys et al., 2019).

Both X-scores and Y-scores (when multiple outcome variables are present), and weights, can be used to interpret which variables are contributing the most information to the model. Weights with large values correspond to variables which contribute large amounts of information, while weights with low values and similar to other weights indicate variables which are not imparting a large impact on the model. Although the values produced by the model offer information both on relationships within the data (X-scores) as well as the variables (X-weights), in the context of predictive modeling for applications, these interpretations are typically not of interest. However, this information may be valuable when attempting to identify key variables for future model building. For the purposes of building a predictive model, the chief interpretation of interest is that of prediction accuracy, particularly when making predictions on samples outside the training data. Ultimately, accuracy is most critical when utilizing these models as they are implemented to optimize the data gathering process and increase the amount of information growers and researchers are utilizing to make decisions.

The maximization of covariance and minimization of correlation between variables through the construction of LVs helps to make PLS more robust against overparameterization and robust against some amount of non-linearity even when assumptions of linearity are held (Wold et al., 2001). This is critical for spectral data from fruit flesh where there are many overlapping absorption peaks at the same wavelength for different parameters (Naes et al., 2002). The estimated orthogonal variables also make the inclusion of large amounts of data per sample feasible and informative as opposed to mathematically cumbersome and detrimental to model performance.

Through the deflation of the measured **X** values, the underlying absorbance peaks lost in the overlapping absorbance profile can be identified due to the LV approach to handling covariance (Mishra et al., 2021). Despite this feature, it is still possible to overfit the model from the addition of too many LVs (Deng et al., 2021). Models that are in danger of being overfit are often referred to in horticultural literature as “complex” and up to nine LVs has been suggested in several studies as a guideline for the limit of complexity (Peirs et al., 2003b). Concerns of overly complex models can seemingly be dispelled when Q^2 is utilized to avoid the inclusion of noise in the data, and a low root mean square error of prediction (RMSEP) are produced, regardless of the number of LVs. Complexity jeopardizes model prediction performance on samples that are not present in the training data. Internal validation is used in PLS to compare the effect of different LVs on model performance (Jung, 2017). Several statistics (R^2 and Q^2) exist to characterize and quantify this performance and subsequently guide the chosen number of LVs in the model (Deng et al., 2021; Wold et al., 2001).

Another key metric for the final iteration of the model, and a diagnostic to determine the appropriate number of LVs is RMSEP. As stated, at the end of each round of regressing scores and weights on row and column values of **X** and **Y**, a set of predictions for both the **X** and **Y** matrices are produced for deflation. The predictions of **Y** can then also be used to generate the RMSEP for that round of the model. It is expected that with each set of predictions, from each LV, the RMSEP will decrease. These error values can then be plotted with the corresponding LV which will produce a distinct graph where there is a dramatic drop in error for the first several LVs, a characteristic “knee” in the trend line when the error begins to be reduced less dramatically, and then finally a flat line where the error ceases to decrease (**Figure 1.1**). There is often some amount of variation in error where the error can increase again after declining, particularly when the noise in the data is captured in the latter LVs (Deng et al., 2021). Like the tipping point in Q^2, where Q^2 begins to fall back toward 0, this flat line, or slight increase in error is an indicator of the successful extraction of all explainable variation in the data. Various methods of cross validation are used to determine the number of LVs (Teh et al., 2020). Some studies will focus on SEP compared to RMSEP to report prediction error although they function similarly (Porep et al., 2015). For making predictions on the internal quality of fruit from outside the calibration/training data set, it is often the case that models with fewer LVs are more generalizable and generally make better predictions on fruit in the future (Anderson et al., 2020; Porep et al., 2015; Saeys et al., 2019; Wang et al., 2017). Peirs et al. (2003) expressed that models with more than 9 LVs are effectively overfit and points to this as the source of high error (>1% ^{o}Brix) in the predictions made in their study. This is likely an example of spectral noise being modeled with successive LVs (low Q^2) more than it is a product of the sheer number of LVs. Bureau et al. (2009), showed that 9 LVs was a sufficient number of variables to capture the variation present in 8 apricot cultivars across variable maturity status and still produce a RMSEP <1% ˚Brix when externally validating. Sun et al. (2020) set the number of LVs in their temperature compensated mango DMC models to 8, based on recommendations from Anderson et al. (2017) and Acharaya et al. (2014), with the lowest RMSEP reported at 1.05%. Kumar et al. (2015) built models all with the optimum number of LVs being between 6 and 9. These studies lend credence to the prescription of approximately 9 LVs when modeling and predicting fruit quality. Once the number of LVs is selected, the final scores are used as beta (β) estimates for each covariate. The model can now be represented as a standard linear equation with an intercept (β_0) and regression coefficients, β_p, for each covariate in the model. Regression coefficients for each covariate in this linear representation allows for estimates of dependent variable(s) (**Y**) given new observations.

Another strength of the PLS model is the interpolation of missing data. Peirs et al. (2003) showed that their temperature adjusted PLS model was able to interpolate for temperatures not represented in the training data producing more accurate results than models trained at one temperature and validated with another. In addition, some alternative modeling approaches have been explored which yield better prediction accuracy for ˚Brix in peach using least squares – support vector machine regression (Mukarev and Walsh, 2012).

1.5. Effect of Data Preprocessing for Model Training

A foundational element in spectroscopy is Beer-Lambert's Law. This principle explains the linear relationship between the concentration of compounds absorbing applied spectral energy, the amount of spectral energy applied and the length the energy is traveling. Beer-Lambert's Law operates on the assumption that there is minimal or no light scattering in the material that spectral energy is being applied to. Here lies one of the primary obstacles for researchers modeling spectral relationships. Fruit tissue violates this assumption as there are many sources of light scattering present in the tissue (Anderson and Walsh, 2021). Because of this light scattering, the apparent absorption of the applied wavelengths is higher than it would be in the absence of light scattering. This introduces non-linearity into the relationship which must be corrected for by preprocessing the spectral data.

Variables with large variance will have a greater impact on the model than those with less variance, and so they are often scaled to allow each variable to contribute similar amounts of influence (Saeys et al., 2019). However, Saeys et al. (2019) do not recommend scaling when data are collected on the same instrument or when variables are measured in the same units as this can lead to the loss of information since scaling variables with low signal to noise ratio will now contribute equally to the total variance in the model as those with high signal to noise ratio.

Beyond manipulation of raw data, different spectral processing techniques are considered to increase the amount of relevant information present in the dataset and to decrease the amount of noise. What is considered relevant or superfluous is dependent on the physical properties of the sample that measurements are taken from and the compounds of interest present in that sample. Commonly considered properties of samples are their light scattering properties including reflectance, transmittance, and absorbance. The objective of data preprocessing is to improve the model fit and thus improve model performance (Rinnan et al., 2009). Changes in model performance can be seen as an increase or decrease of the error, calculated as the average difference between actual and predicted values. The sum of the square root of the differences between actual and predicted values is calculated and reported as the root mean square error (RMSE) and can be calculated both during cross validation (RMSECV) when predictions are made on the training data, and during external validation when predicted values are made using new observations (RMSEP). The comparison of these errors is made between models fit to preprocessed data and raw data. Data processing does not always improve model fit and performance. An example of data preprocessing failing to improve performance was observed in a previous study (Peirs et al., 2005) predicting optimal fruit harvesting constructed models using various data preprocessing (smoothing, scatter correction, and first and second derivatives). The impact of these various preprocessing methods on out-of-sample prediction accuracy was evaluated by comparing the RMSEP of models trained with different preprocessing methods, and no significant improvement in accuracy was seen (Peirs et al., 2005).

The first or second derivative of raw absorbance data is often taken as a common form of preprocessing for data smoothing and is often used in conjunction with scaled and centered values (Anderson et al., 2020; Bobelyn et al., 2010; Fu et al., 2008; Nascimento et al., 2016; Mishra et al., 2021; Mukarev and Walsh, 2021; Sun et al., 2020; Walsh et al., 2004; Wang et al., 2017). These preprocessing techniques are a response to the scattering effect that fruit tissue has on different wavelengths, violating Beer-Lambert's Law which introduces non-linearity into the data, and increasing the apparent absorption of wavelengths. This violation is largely due to the increased pathlength the radiation travels due to scattering before absorption resulting in additive and multiplicative effects in the reported absorbance data (Fearn and Davies, 2003). To correct this, different derivatives are calculated, or other scattering correction techniques such as multiplicative scatter correction and standard normal variate (MSV and SNV, respectively) are applied (Nascimento et al., 2016; Rinnan et al., 2009; Rungpichayapichet et al., 2016). Smoothing acts to help with the signal to noise ratio and thereby increases spectral resolution (Lin and Ying, 2009). As an example of additive and multiplicative effects, one can see clearly how these effects are caused by changes in temperature and can be visualized as shifts in the plotted spectra of a single sample where two different temperatures produce two distinct spectra absorbances despite being the same sample, which should theoretically have the same absorbance. Scatter correction can then be seen when the two spectra from the same sample are treated, and the shifts in the plot are then compressed. After this compression, the two spectra more closely overlap (Peirs et al., 2003a). This principle is applied beyond the effect of temperature on a single fruit sample and is used to correct for the effect of other physical or chemical influences which cause distinct scattering (e.g., skin thickness, cell density, degree of cell wall degradation, cultivar differences) in different fruit making it difficult to compare the spectra of a sample population. Minimal skin thickness and homogenous fruit flesh have been credited for some of the higher accuracy predictive models particularly for apples, with peaches being similarly accurate due to the thin skin. A lack of flesh texture homogeneity in peaches compared to apples is described as a reason peach models have been less accurate in predicting various quality and maturity attributes (Walsh et al., 2004). Changing sample temperatures adds non-linearity to data, but given the accuracy that the mango models produced, the use of MSV indicates the potential for the correction of non-linearity in spectral data when predicting the temperature of a sample. Bobelyn et al. (2010) selected SNV over the second derivative due to better observed performance when comparing the two approaches for SSC prediction in apples.

Beyond preprocessing techniques, dimension reduction methods are commonly applied to help reduce the amount of redundant and superfluous information and ultimately improve prediction. Variable selection, one example of dimension reduction techniques, is referred to as the (manual or automatic) determination of active or influential covariates in model. Limiting the amount of information gathered and processed in practical applications via variable selection also improves the rate at which predictions can be made and the number of sample predictions can be made on (Li et al., 2019). Li et al. (2019), showed that there are benefits to variable selection and was able to make accurate SSC predictions as low as 0.23-0.30° Brix in pear. It has been demonstrated across disciplines implementing Vis-NIRS that not all wavelengths have an equal role to play in elucidating the composition of a sample and the inclusion of all the variables can hurt performance when the signal to noise ratio is low (Saeys et al., 2019). It is rather the case that only a few narrow spectral ranges carry most of the pertinent information regarding sample composition indicating that many spectral regions can be excluded without reducing model performance. Although spectral regions can be reasonably excluded, the inclusion of large swaths of the Vis-NIR regions similarly does not reduce model performance. For this reason, many studies utilizing PLS, a dimension reduction technique which does not explicitly perform variable selection, focus on a broad range of spectra spanning the entirety of the NIR spectrum and much of the visible spectrum since PLS performance is not as dramatically penalized by the inclusion of broad covariates and is instead capable of extracting more nuanced information of a sample (Wold et al., 2001) contributing to the variability in the model as well as overall accuracy and robustness.

Despite both preprocessing and dimension reduction techniques being important statistical considerations for researchers with large amounts of spectral data, the application of these techniques, namely preprocessing, is not always necessary. It has been suggested that the use of raw data free of any preprocessing can be the most informative since it allows certain spectral areas to contribute more information in the model (Peirs et al., 2003a). Walsh et al. (2004) saw little improvement between the use of second derivative and raw absorbance, but also noted that this may be due to the use of a single population of fruit for each model, removing some of the need for spectral correction as a solution to the variability of light scattering in different fruit populations where reduced homogeneity might be expected. Similarly, when modeling firmness, a textural trait known for its diverse light scatter properties, Bobelyn et al. (2010) opted to use raw data instead of preprocessing to retain scattering information in their data. Fu et al. (2007) also modeled firmness and applied MSV and was able to predict firmness in peaches more accurately with the pretreatment than Bobelyn et al. (2010) was in apples (RMSEP 5.42 N vs. 5.9 – 8.8 N). Wang et al. (2017) also applied preprocessing to their spectral data when predicting pear firmness and produced an error or 8.18 N. Peirs et al. (2003) followed this approach in a later study on the impact of biological variability on model performance and reported high RMSEP (1.91%) when predicting ˚Brix in apples with external validation. This approach means forgoing some of the help preprocessing provides for correcting for non-linearities and relying on a very homogenous set of calibration samples producing smooth spectral data. Peirs's no preprocessing suggestion also acts as a counternarrative to the signal/noise relationship. High RMSEP of sugar concentration in apples may be indicative of issues presented when preprocessing is forgone. Instances of comparable RMSEPs when predicting firmness between models containing preprocessed data and models containing raw data indicate that relevant information is not lost when preprocessing is applied. It also indicates to some extent that less accuracy than expected is gained through preprocessing and speaks to the ability of PLS to deconvolute spectra through dimensional reduction. Between examples of preprocessing both improving and reducing model performance, and the ability of PLS to handle many covariates, the application of these techniques cannot be generally applied in all instances but should be explored in studies to produce the best results.

1.6. Spectral Ranges Related to Fruit Quality Parameters

It has been established that PLS is appropriate when modeling spectral data, but it is also critical that the appropriate spectrum be utilized. The near infrared spectrum is the neighboring portion of the electromagnetic spectrum to visible light and ranges from 780 to 2500 nm. Because it is so close to the visible spectrum, it is not uncommon that the range of wavelengths administered to a sample in research studies will also include some amount of the visible spectrum. The best example of this in horticulture is the use of visible light to help quantify the concentration of chlorophyll, a compound highly reactive to specific wavelengths of the visible light for photosynthesis, specifically 670 and 720 nm (Ziosi et al., 2008; Costa et al., 2009). The visible spectrum was intentionally excluded by Peirs et al. (2005) in their study on apples due to the relationship between the changing color of the fruit skin and visible light reflectance, to explore the maturity information present in the near infrared spectrum. Their total spectrum spanned from 800 – 2000 nm, and they concluded that the near infrared spectrum was able to capture internal changes occurring during maturation based on the amount of variation (42%) that was explained by the first principal component in a PCA analysis conducted on the spectral data from their apples.

Similarly, Zhang et al. (2019) excluded all portions of the visible spectrum to exclude information deemed “irrelevant” for SSC and DMC prediction. Interestingly, Anderson et al. (2020) saw a decrease in RMSEP for DMC prediction when extending the range into the visible spectrum. The authors indicated that it is likely the inverse relationship of chlorophyll degradation and DMC accumulation that might be contributing to the model, however they do not believe that the relationship observed in their data between chlorophyll and DMC to be reliable enough in future populations to contribute this same amount of accuracy in future predictions and so it will not add to the robustness of the model. It is the case that the spectral profiles of fruit tissue samples are dominated by water absorption bands, and so it is often the case that much of the information that is gathered regarding the internal state of a given sample is reflective of the amount and state of water in the sample (Nicolai et al., 2007).

Another component that has been investigated is how well different ranges penetrate biological tissues. Walsh et al. (2004) mentioned that the use of reflectance optics is capable of penetrating between 4 and 20 mm into the fruit tissue. It has been shown that lower ranges of NIR wavelengths are more able to penetrate deeper into fruit tissue than higher ranges and a more accurate picture of that tissue is produced. While higher ranges may yield strong absorbance peaks, these wavelengths do not penetrate deeply into the tissue and result in values that indicate the presence of compounds only near the surface, leaving the interior unknown (Lin and Ying, 2009). It has been demonstrated that there are significant differences between values nearest the exocarp and those closer to the endocarp, making the lower ranges of the near-infrared spectrum, roughly 700 – 900 nm, of greater interest for analyzing fruit tissues (Lammertyn et al., 2000). There are also instances where the quality of the skin is of primary interest, generally for the sake of monitoring chlorophyll content (Ziosi et al., 2008; Costa et al., 2009) and in these instances it reasons one would focus on the wavelengths reactive in that area without introducing additional noise to the data from deeper in the sample.

Beyond tissue penetration depth, the near-infrared spectrum is utilized due to the documented absorbances of those wavelengths by O-H and C-H bonds in water and carbohydrates. Peirs et al. (2005) recognized the important role of the typical overtone absorbance bands due to O-H bonds near 970, 1450, and 1940 nm, when using spectral profiles from fruit of variable maturity. It has been consistently recorded that the range of 729 – 975 nm is effective range for DMC as it contains O-H absorbance information as well as sugar and other carbohydrate absorbances (Anderson et al., 2020; Mishra et al., 2021; Nicolai et al., 2007; Sun et al., 2020; Teh et al., 2020; Zhang et al., 2019; Minas et al., 2021; 2023). This is in keeping with the goal of capturing typical overtone bands of O-H bonds from deeper in the fruit tissue by using a lower range of the NIR spectrum, acting as a "manual" form of variable selection as described by Mishra et al. (2021). Other prominent peaks were observed at 1170, 1400, and 1800 nm by Peirs et al. (2003) when inspecting the impact of temperature on apparent absorbance. Sun et al. (2020) showed peak shifts in mango caused by temperature increases in the second derivative of absorbance to peaks at 740, 840, and 963 nm where the peaks decreased and at 920 nm where the peak increased.

By selecting narrower ranges of spectra, researchers are effectively performing model selection since more precise information is gathered with less noise when the spectra are thoughtfully selected (Mishra et al., 2020). Mishra et al. (2020) presented yet another instance where the balance of two attainable elements is of interest, these being the gathering of enough information that each sample contributes unique information to the model and not adding too much information that the model is bogged down by noise and overfit with too many LVs (Deng et al., 2015; Mishra et al., 2021). Mishra et al. (2021) utilized two variable selection approaches in a study on pears, using interval partial least squares regression (iPLS2R) and CovSel, a covariate selection package. Using the iPLS2R approach, they were able to identify the range of wavelengths for SSC as 709 – 759 nm and 789 – 999 nm and the two ranges for moisture content as 743 – 779 nm, and 879 – 939 nm. It seems appropriate to revisit the typical range of 729 – 975 nm for these parameters as it is similar to those ranges. CovSel was able to more precisely pick out 736, 709, 961, 1109, 1125, 816, 912, and 879 nm as the informative wavelengths for both SSC and moisture content. Several of these are in the neighborhood of well-known NIR absorbance peaks for both O-H and C-H bonds. Using these selected ranges as an alternative to the entire spectrum, prediction error was reduced from 1.31% to 0.19% for moisture and from 1.44 to 0.58 for SSC. Selecting the range of 729 – 975 nm out of the total NIR spectrum can be considered as a form of variable selection given that it focuses on an informative range of wavelengths within the total Vis-NIR spectrum. Golic and Walsh (2004) focused on the wavelength region of 734 – 931 nm for SSC and DMC, noting the weighting of 910 nm as a primary contributor of information when measuring sugar. Other key absorbance peaks observed in fruit are at 840 and 960 nm, also connected with O-H bonds, and C-H bonds near 910 nm (Anderson et al., 2020; Subedi et al., 2007; Walsh and Golic, 2004). Rungpichayapichet et al. (2016) noted the O-H bonds seen as a broad band peak around 960 – 990 nm, both from water as well as from carbohydrates like sugars and starch.

The conclusion from this collection of studies is highlighting the value and relevance of the near-infrared spectrum particularly between 700 – 1000 nm. These studies also highlight the use and application of neighboring spectra in both the visible and infrared regions and the methods to process and handle the data for maximal model performance. These studies have validated that there is a large amount of relevant information in much of the NIR spectrum, and that while the inclusion of large portions of the spectrum may yield good results, the selection of specific wavelength regions can yield improved performance.

1.7. Biological Variation and Spectral Data

Large amounts of variability in chemical and physical composition of samples introduce more light scattering due to a lack of homogeneity in the material. This makes accounting for scattering and non-linearity in the system more difficult to account for (Rinnan et al., 2009; Wang et al., 2015). With that, different parameters of interest may relate more to absorbance or to scattering. Depending on if absorbance or scattering is more relevant, the sample variability corresponding with those characteristics should be considered based on biological variation. This has serious implications for the selection of appropriate spectral data preprocessing (Bobelyn et al., 2010).

As mentioned earlier, different materials have different interactions with the visible and near infrared spectrum. This includes both physical properties and the chemical compositional properties of that object. Physical properties impact the way in which light passes through an object (scattering) while chemical properties impact wavelength absorbance based on the presence and abundance of chemical structures which are responsive to the wavelengths applied to the sample (Nicolai et al., 2007). It has been established in horticultural literature that different fruit cultivars are genotypically and thus phenotypically distinct. These phenotypic distinctions can be observed both in the physical traits (skin color, flesh color, flesh type, fruit shape, fruit size, and fruit firmness) as well as in the organoleptic traits (DMC, SSC or ˚Brix, TA, and fruit texture) (Minas et al., 2018). Ultimately all these traits, and variations of them between fruit specimens, impact the way in which certain spectral regions utilized in Vis-NIRS interact with the fruit (Nascimento et al., 2016; Ziosi et al., 2008). In short both physical and chemical properties of the fruit flesh, as the material of interest, are determined by biological and environmental factors. Variation in these factors is captured as variations in spectral absorbance in Vis-NIRS studies and made present in the experimental design. For this reason, from a statistical modeling perspective, it becomes of value to capture physical and biological variation for the purpose of spectral variations that, when modeled, make a more robust model (Nicolai et al., 2007).

Factors that determine physical variation at the species level include rootstock, cultivar, maturity, and growing environment including canopy position/architecture, crop load, irrigation, mineral nutrition, and climate (Minas et al., 2018; Peirs et al., 2005). Fruit flesh density, also related to fruit firmness, and skin thickness are two attributes that have been linked with these factors (Bobelyn et al., 2010). Not only is it the case that there is a wide range of variation between samples, but Fu et al. (2008) reported significant differences in the spectral properties of different latitudinal scans of a single fruit. They also reported improved prediction accuracy for predicting fruit firmness when the multiple scans of a single fruit were averaged and modeled as opposed to the use of a single scan (average of scans RMSEP: 5.96 vs. individual scans RMSEP: 7.04 – 7.82 N).

Although many of these fruit characteristics are genetically predetermined, the greater cause of phenotypic variability in fruit comes from cultural practices followed during cultivation, and the environmental and orchard factors that they will be exposed to during development (Anthony and Minas, 2022; Minas et al., 2018). It is of additional interest to distributors and sellers what the environmental impacts of postharvest handling and conditions might be on the physical and chemical properties of each fruit.

At every stage of development and handling, fruits are exposed to factors which will have some effect on them (crop load, harvest date, bruising, chilling injury, continued respiration during storage). These effects can be miniscule or profound, as in the case of different crop loads (Grossman and DeJong, 1995; Minas et al., 2021). Minas et al. (2021) observed statistically significant ($p < 0.05$) differences in DMC between a "commercial" (15 cm spacing between fruit) and "heavy" (5 cm spacing between fruit) crop load, being 14% and 11.9% respectively and found shifts in the rate of fruit maturation within the same cultivar. With the inclusion of crop load variability in the training data for a PLS model, fit using the 729 nm – 975 nm range with DMC and SSC as two separate outcomes, they were able to generate prediction models with low error rates of 0.41% for DMC and 0.58% for SSC when externally validated. A PCA bi-plot from the study indicates spectral variation attributed to canopy position and crop load.

Since different cultivars can produce statistically significant differences in sugar and water content, the same way of thinking about environmental variability can and should also be applied to the way in which cultivars are modeled. Anderson et al. (2020), and Anthony et al. (2023), both reported that locally calibrated, cultivar specific models, for mangoes and peaches, respectively, are superior to global cultivar models based on the homogeneity of the samples by limiting the presence of variability based on phenotypic differences. Zhang et al. (2019) reported individual apple cultivar models outperform multi-cultivar models, however, they report lower RMSEP when externally validating both SSC and DMC using the multi-cultivar models compared to the individual cultivar models (0.47 – 0.78% vs. 0.49 – 1.32% SSC and 4.83 – 7.03 g kg^{-1} vs. 5.31 – 12.81 g kg^{-1} DMC). This indicates some disparity in how model performance is being characterized between studies. Emphasis is occasionally placed on R^2 values over prediction error (RMSEP), but error is more important in these kinds of studies. Better performance in single cultivar models is reinforced by the findings of Peirs et al. (2003), Louw and Theron (2010), and Wang et al. (2017), which indicated that much of the variability in samples stemmed from cultivar and prescribed the use of cultivar specific models over global cultivar models. Wang et al. (2017) acknowledged this variability and applied additional processing (orthogonal signal correction with second derivation and smoothing) to the data to reduce spectral differences between cultivars while retaining pertinent information for making predictions on SSC. Bureau et al. (2009) saw sufficient accuracy and robustness in their apricot models when predicting SSC (RMSEP: 0.99% ˚Brix) when 8 cultivars were included in the model. The removal of an "atypical" cultivar did result in a reduction of error to 0.91% indicating there may have been a lack of homogeneity introduced by that atypical cultivar. Anthony et al. (2023) indicated that cultivars of similar phenotypic characteristics and harvest times could potentially be lumped together to produce models performing with adequate accuracy across cultivars when predicting DMC and I_{AD}. Bureau et al. (2009) supports the notion of aggregating similar cultivars when producing models for groups of cultivars based on phenotypic attributes with reports of improved model performance after a particularly distinct cultivar was removed from a global cultivar model. However, Bureau et al. (2009) expressed concern that this reduction in error may come at the cost of some robustness. Li et al. (2019) saw comparable performance between a multi-cultivar model and an individual cultivar models when predicting SSC in pears, with prediction errors being 0.2 and 0.35 ˚Brix for both model types, respectively. Wang et al. (2017) determined that with additional processing, a multi-cultivar mode for SSC in pears performed just as well as many reported single cultivar models with RMSEP of 0.46 ˚Brix.

Given the unique microstructures present in each cultivar and the impact microstructures are known to have on spectral absorbance, it is intuitive to assume maximal model accuracy would be achieved with cultivar specific models. This then becomes at its core a discussion of global vs. local models which is determined by the inclusion or exclusion of certain samples in the model training data, much like the number of samples of a given temperature for a temperature compensated DMC model (Sun et al., 2020).

Cell wall deterioration is another consideration when attempting to capture maturity variability. The more mature a fruit becomes; the more water molecules are redistributed in intercellular spaces while pectin and cellulose change (Louw and Theron, 2010). With this shift in the location of water, and the change in light scattering due to changing cell wall conditions, greater differences in apparent absorbance occur (Peirs et al., 2005; Nicolai et al., 2007). Changes in cellular spacing, intercellular water, and cell wall composition which can be linked with firmness is also responsible for differences in light scattering, and these differences can be seen in hard green mangoes when compared to more mature and soft fruit (Anderson et al., 2020) as well as in Japanese plums (Louw and Theron, 2010).

The impact of other environmental factors has also been widely explored and is considered heavily when inspecting "robustness" particularly when making predictions on samples from different seasons or orchards (Bobelyn et al., 2010; Minas et al., 2021; 2023; Rungpichayapichet et al., 2016; Teh et al., 2020). Both orchards and seasons represent two distinct environments, distinct due to microclimates and to seasonal variability between years, that with intrinsic species or cultivar variation, produce unique compound variations in the fruit that needs to be accounted for and represented in the data to construct a model robust against as many of these conditions as possible. Rungpichayapichet et al. (2016) showed better model performance with the inclusion of multiple seasons of data, attributing the improvement to the inclusion of greater variability. Although important to consider, Peirs et al. (2003) showed that in their models, variation between cultivar and season had a larger effect than orchard location. Teh et al. (2020), also expound on the impact that location and season have on DMC predictions as another source of variability and found that prediction was better within year performed than between year (0.48% vs. 0.79%), and better prediction within orchard compared to between orchard with error ranges of 0.55 – 0.63% and 0.49 – 0.85% between orchards for the years 2015 and 2016, respectively. Teh et al. (2020) also determined that a model calibrated at an orchard that shared regional/environmental characteristics (representative of other orchards in the area) was consistently more accurate than models trained with samples from an environmentally distinct orchard. This supports the notion that locally trained models perform with better accuracy, but at the cost of robustness and reinforces the need to be mindful of where training data originated from. It also indicates that representative samples can produce functional models robust against location. Perhaps the most significant conclusion from this study was that models were able to be constructed simultaneously robust against orchard and seasonal variation.

Interestingly, tree age has also been seen to have an impact on model accuracy (Teh et al., 2020). It was discovered that models calibrated with samples from more mature trees made more accurate cultivar specific models which may be related to a mature trees less uniform canopy contributing to variability captured in the model, as well as a mature tree's ability to carry a heavier crop load.

In the search of accurate and robust models, it has been thoroughly explored and determined that there is a need for variability in the training data to produce generalized models which predict accurately and are insensitive to external variability when encountering new data (Sun et al., 2020; Teh et al., 2020; Wang et al., 2017; Zhang et al., 2019). Accuracy and robustness are achieved simultaneously when there is a balance of variation in the data and variables in the final model, as there are instances when large amounts of variation are well modeled, but the model becomes overly complex, and robustness is lost resulting in larger errors. This is particularly true when atypical data are included in the model training population (Peirs et al., 2003b). This can be summarized as the balance between underfit and overfit models where an underfit model is plagued by bias in the model and overfit models are plagued with an increasing amount of variance which becomes an issue when the data are noisy. This is the bias/variance tradeoff described by Deng et al. (2015) and emphasizes the value and need of applying various model diagnostics to ensure the proper tradeoff is reached.

The final consideration when balancing variability and homogeneity is the number of samples used to train the models. The larger the number of distinct populations used to train models risks the introduction of heterogenous characteristics in the training data resulting in reduced linearity and reduced model performance. Connected to the issue of combining different sample populations is the number of samples present in those populations. Luo et al. (2018) demonstrated the reduction of RMSEP in their apple models, characteristic of increased variability in the training data with the inclusion of a broad range of samples, but also showed that at times, as the number of samples increases, RMSEP will either plateau, or increase due to a lack of homogeneity. They discussed some of the issues related to the discrepancies was the starch hydrolysis process taking place during apple maturation as a source of heterogeneity.

1.8. Temperature Variability and Influence in Model Training

Temperature influence on spectral absorbance is noted throughout the literature assessment related to Vis-NIRS fruit modeling efforts due to the strong effect temperature has on hydrogen bonds. Objects with approximately 85% water content, like fruit, are considered to have a high moisture content while other reports indicate that any water level greater than 80% are subject to the strong influence of temperature on H bonds and apparent spectral absorption (Anderson and Walsh, 2021; Hansen et al., 2000). This phenomenon is the result of reduced water molecule cluster size in the sample due to increased temperatures causing an increase in absorbance and a reduction in reflectance (Peirs et al., 2003a). This raised absorbance is due to the breakage of hydrogen bonds between water molecules caused by the increase in vibrational energy due to the elevated temperature. This is the cause of the reduced clusters described earlier.

Peirs et al. (2003a) showed that models trained at a single temperature point (local temperature) were sensitive to samples at a different temperature, with an increasing prediction error as the temperature difference increased. When models were temperature compensated (global temperature model), a lower RMSEP was produced due to the increased variability in these models that were less sensitive to small changes in new samples. Overall, the global model outperformed the local model even when the local model made predictions on an external sample at the same temperature as the calibration data. With the conclusion that the broader the temperature range used to train the model, the more accurate the predictions were.

Kumar et al. (2015) collected spectral data from the orchard to account for environmental variability. This included temperature as their measurements were collected while orchard temperatures fluctuated between 15 and 23 ˚C and saw in their model validations that their models were insensitive to temperature (consistent RMSEP < 1% between internal and external validation) when predicting SSC. These models indicate the ability for model compensation by capturing variability during data collection.

Like biological variation, temperature is another source of variation impacting model accuracy and robustness. Unlike biological variation which is represented across the training population as individual samples each with a unique spectral profile and a quality parameter of interest, temperature is represented as multiple spectral profiles for the same parameter. This is another example of the strength and ability of PLS modeling to handle highly correlated data.

1.9. Conclusion

The information collected and presented here offers a brief overview of the history and applications of PLS and Vis-NIRS for the modeling and prediction of a variety of parameters and indices used across different fruit crops. It has been thoroughly documented that different research projects for different fruit crops have seen different levels of success which generally can be attributed to the methods of data collection and preprocessing while PLS has generally been validated as a statistically capable framework for approaching and capturing the nuanced variance in complex data. Key takeaways include the influence of biological variation and temperature on model prediction.

1.10. Figures

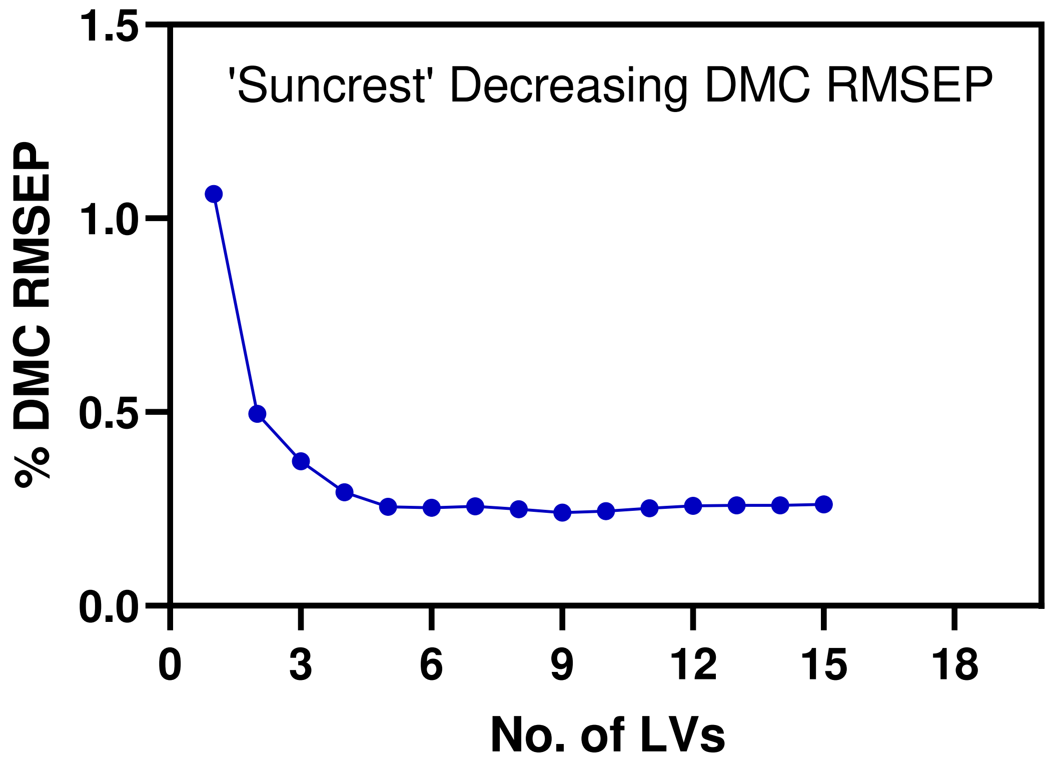

Figure 1.1. RMSEP of DMC decline and plateau with an increasing number of latent variables (LVs) from PLS model.

1.11. References

1. Acharya, U., Walsh, K., Subedi, P. (2014). Robustness of partial least-squares models to change in sample temperature: II. Application to fruit attributes. Journal of Near Infrared Spectroscopy, 22(4), 287-295. https://doi.org/10.1255/jnirs.1119
2. Anderson, N. T., & Walsh, K. B. (2022). Review: The evolution of chemometrics coupled with near infrared spectroscopy for fruit quality evaluation. Journal of Near Infrared Spectroscopy, 30, 3–17. https://doi.org/10.1177/09670335211057235
3. Anderson, N. T., Walsh, K. B., Subedi, P. P., & Hayes, C. H. (2020). Achieving robustness across season, location and cultivar for a NIRS model for intact mango fruit dry matter content. Postharvest Biology and Technology, 168, 111202 https://doi.org/10.1016/j.postharvbio.2020.111202
4. Anthony, B. M., Chaparro, J. M., Prenni, J. E., & Minas, I. S. (2020). Early metabolic priming under differing carbon sufficiency conditions influences peach fruit quality development. Plant Physiol. Biochem., 157, 416–431. https://doi:10.1016/j.plaphy.2020.11.004
5. Anthony, B. M., Sterle, D. G., & Minas, I. S. (2023). Robust non-destructive individual cultivar models allow for accurate peach fruit quality and maturity assessment following customization in phenotypically similar cultivars. Postharvest Biology and Technology, 195, 112148. https://doi.org/10.1016/J.POSTHARVBIO.2022.112148
6. Balabin, R. M., & Smirnov, S. V. (2012). Interpolation and extrapolation problems of multivariate regression in analytical chemistry: Benchmarking the robustness on near-infrared (NIR) spectroscopy data. Analyst, 137(7), 1604–1610. https://doi.org/10.1039/c2an15972d
7. 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 Biology and Technology, 55(3), 133–143. https://doi.org/10.1016/j.postharvbio.2009.09.006

8. Bureau, S., Ruiz, D., Reich, M., Gouble, B., Bertrand, D., Audergon, J. M., & Renard, C. M. G. C. (2009). Rapid and non-destructive analysis of apricot fruit quality using FT-near-infrared spectroscopy. Food Chemistry, 113(4), 1323–1328. https://doi.org/10.1016/j.foodchem.2008.08.066
9. Costa, G., Noferini, M., Fiori, G., Miserocchi, O., & Bregoli, A. M. (2002). NIRS evaluation of peach and nectarine fruit quality in pre-and post-harvest conditions. Acta Horticulturae, 592, 593–599. https://doi.org/10.17660/ActaHortic.2002.592.81
10. Deng, B. C., Yun, Y. H., Liang, Y. Z., Cao, D. S., Xu, Q. S., Yi, L. Z., & Huang, X. (2015). A new strategy to prevent over-fitting in partial least squares models based on model population analysis. Analytica Chimica Acta, 880, 32–41. https://doi.org/10.1016/j.aca.2015.04.045
11. Fearn, T., & Davies, A. M. C. (2003). Locally-biased regression. Journal of Near Infrared Spectroscopy, 11(6), 467-478. https://doi:10.1255/jnirs.397
12. Fu, X. P., Ying, Y. Bin, Zhou, Y., Xie, L. J., & Xu, H. R. (2008). Application of NIR spectroscopy for firmness evaluation of peaches. Journal of Zhejiang University: Science B, 9(7), 552–557. https://doi.org/10.1631/jzus.B0720018
13. Golic, M., Walsh, K., & Lawson, P. (2003). Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature. Applied spectroscopy, 57(2), 139–145. https://doi.org/10.1366/000370203321535033
14. Grossman, Y. L., & DeJong, T. M. (1995). Maximum fruit growth potential following resource limitation during peach growth. Annals of Botany, 75(6), 561–567. https://doi.org/10.1006/ANBO.1995.1059

15. 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. Journal of Near Infrared Spectroscopy, 8(2), 125-132.

16. Jung, Y. (2018). Multiple predicting K-fold cross-validation for model selection. Journal of Nonparametric Statistics, 30(1), 197–215. https://doi.org/10.1080/10485252.2017.1404598

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

18. Lammertyn, J., Peirs, A., De Baerdemaeker, & J., Nicolaï, B. (2000). Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. Postharvest Biology and Technology, 18(2), 121-132. https://doi.org/10.1016/S0925-5214(99)00071-X.

19. 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). https://doi.org/10.3390/plants7010003

20. Li, J., Zhang, H., Zhan, B., Wang, Z., & Jiang, Y. (2019). Determination of SSC in pears by establishing the multi-cultivar models based on visible-NIR spectroscopy. Infrared Physics and Technology, 102. https://doi.org/10.1016/j.infrared.2019.103066

21. Lin, H., & Ying, Y. (2009). Theory and application of near infrared spectroscopy in assessment of fruit quality: A review. Sensing and Instrumentation for Food Quality and Safety, 3(2), 130–141. https://doi.org/10.1007/s11694-009-9079-z

22. Luo, X., Ye, Z., Xu, H., Zhang, D., Bai, S., & Ying, Y. (2018). Robustness improvement of NIR-based determination of soluble solids in apple fruit by local calibration. Postharvest Biology and Technology, 139, 82–90. https://doi.org/10.1016/j.postharvbio.2018.01.019

23. Louw, E. D., & Theron, K. I. (2010). Robust prediction models for quality parameters in Japanese plums (Prunus salicina L.) using NIR spectroscopy. Postharvest Biology and Technology, 58(3), 176–184. https://doi.org/10.1016/j.postharvbio.2010.07.001

24. Mayer, N. A., Bianchi, V. J., Feldberg, N. P., & Morini, S. (2017). Advances in peach, nectarine and plum propagation. Revista Brasileira de Fruticultura, 39(4). https://doi.org/10.1590/0100-29452017355

25. Minas, I. S., Anthony, B., Pieper, J., Sterle, D. (2023). Large-scale and accurate non-destructive visual to near infrared spectroscopy-based assessment of the effect of rootstock on peach fruit internal quality, European Journal of Agronomy, 143, 1161-0301, https://doi.org/10.1016/j.eja.2022.126706

26. Minas, I. S., Blanco-Cipollone, F., & Sterle, D. (2021). Accurate non-destructive prediction of peach fruit internal quality and physiological maturity with a single scan using near infrared spectroscopy. Food Chemistry, 335. https://doi.org/10.1016/j.foodchem.2020.127626

27. Minas, I. S., Tanou, G., & Molassiotis, A. (2018). Environmental and orchard bases of peach fruit quality. Scientia Horticulturae, 235, 307–322. https://doi.org/10.1016/j.scienta.2018.01.028

28. Mishra, P., & Nikzad-Langerodi, R. (2020). Partial least square regression versus domain invariant partial least square regression with application to near-infrared spectroscopy of fresh fruit. Infrared Physics and Technology, 111. https://doi.org/10.1016/j.infrared.2020.103547

29. Mishra, P., Woltering, E., Brouwer, B., & Hogeveen-van Echtelt, E. (2021). Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach. Postharvest Biology and Technology, 171. https://doi.org/10.1016/j.postharvbio.2020.111348

30. Mukarev, M., Walsh, K. (2012). Prediction of brix values of intact peaches with least squares-support vector machine regression models. Journal of Near Infrared Spectroscopy, 20(6), 647-655. https://doi:10.1255/jnirs.1026

31. Næs, T., Isaksson, T., Fearn, T., & Davies, T. (2002). A user-friendly guide to multivariate calibration and classification (Vol. 6, p. 347). Chichester: NIR.

32. Nascimento, P. A. M., Carvalho, L. C. de, Júnior, L. C. C., Pereira, F. M. V., & Teixeira, G. H. de A. (2016). Robust PLS models for soluble solids content and firmness determination in low chilling peach using near-infrared spectroscopy (NIR). Postharvest Biology and Technology, 111, 345–351. https://doi.org/10.1016/j.postharvbio.2015.08.006

33. Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biology and Technology, 46(2), 99–118. https://doi.org/10.1016/j.postharvbio.2007.06.024

34. Overbeck, V., Schmitz, M., & Blanke, M. (2017). Non-destructive sensor-based prediction of maturity and optimum harvest date of sweet cherry fruit. Sensors (Switzerland), 17(2), 277. https://doi.org/10.3390/s17020277

35. Palmer, J. W., Harker, F. R., Tustin, D. S., & Johnston, J. (2010). Fruit dry matter concentration: A new quality metric for apples. Journal of the Science of Food and Agriculture, 90(15), 2586–2594. https://doi.org/10.1002/jsfa.4125

36. Peirs, A., Scheerlinck, N., & Nicolaï, B. M. (2003a). Temperature compensation for near infrared reflectance measurement of apple fruit soluble solids contents. Postharvest Biology and Technology, 30(3), 233–248. https://doi.org/10.1016/S0925-5214(03)00118-2

37. 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 Biology and Technology, 35(1), 1–13. https://doi.org/10.1016/j.postharvbio.2004.05.010

38. Peirs, A., Tirry, J., Verlinden, B., Darius, P., & Nicolaï, B. M. (2003b). Effect of biological variability on the robustness of NIR models for soluble solids content of apples. Postharvest Biology and Technology, 28(2), 269–280. https://doi.org/10.1016/S0925-5214(02)00196-5

39. Porep, J., Kammerer, D., & Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. Trends in Food Science & Technology, 46(2), 211-230. https://doi.org/10.1016/j.tifs.2015.10.002

40. Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. Trends in Analytical Chemistry, 28(10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

41. Rungpichayapichet, P., Mahayothee, B., Nagle, M., Khuwijitjaru, P., & Müller, J. (2016). Robust NIRS models for non-destructive prediction of postharvest fruit ripeness and quality in mango. Postharvest Biology and Technology, 111, 31–40. https://doi.org/10.1016/j.postharvbio.2015.07.006

42. Saeys, W., Nguyen Do Trong, N., Van Beers, R., & Nicolaï, B. M. (2019). Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: A review. Postharvest Biology and Technology, 158. https://doi.org/10.1016/j.postharvbio.2019.110981

43. Sun, X., Subedi, P., & Walsh, K. B. (2020). Achieving robustness to temperature change of a NIRS-PLSR model for intact mango fruit dry matter content. Postharvest Biology and Technology, 162. https://doi.org/10.1016/j.postharvbio.2019.111117

44. Teh, S. L., Coggins, J. L., Kostick, S. A., & Evans, K. M. (2020). Location, year, and tree age impact NIR-based postharvest prediction of dry matter concentration for 58 apple accessions. Postharvest Biology and Technology, 166. https://doi.org/10.1016/j.postharvbio.2020.111125

45. Uwadaira, Y., Sekiyama, Y., & Ikehata, A. (2018). An examination of the principle of non-destructive flesh firmness measurement of peach fruit by using VIS-NIR spectroscopy. Heliyon, 4, 531. https://doi.org/10.1016/j.heliyon.2018

46. Wang, H., Peng, J., Xie, C., Bao, Y., & He, Y. (2015). Fruit quality evaluation using spectroscopy technology: A review. In Sensors (Switzerland), 15(5), 11889–11927. https://doi.org/10.3390/s150511889

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

48. Wold, S., Sjostrom, M., Eriksson, L., & Sweden˚˚, S. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58. www.elsevier.comrlocaterchemometrics

49. Zeaiter, M., Roger, J. M., & Bellon-Maurel, V. (2006). Dynamic orthogonal projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations. Chemometrics and Intelligent Laboratory Systems, 80(2), 227–235. https://doi.org/10.1016/j.chemolab.2005.06.011

50. Zhang, Y., Nock, J. F., Al Shoffe, Y., & Watkins, C. B. (2019). Non-destructive prediction of soluble solids and dry matter contents in eight apple cultivars using near-infrared spectroscopy. Postharvest Biology and Technology, 151, 111–118. https://doi.org/10.1016/j.postharvbio.2019.01.009

51. Ziosi, V., Noferini, M., Fiori, G., Tadiello, A., Trainotti, L., Casadoro, G., & Costa, G. (2008). A new index based on vis spectroscopy to characterize the progression of ripening in peach fruit. Postharvest Biology and Technology, 49(3), 319–329. https://doi.org/10.1016/j.postharvbio.2008.01.017

CHAPTER TWO

TRAINING ROBUST NON-DESTRUCTIVE MODELS ACROSS 13 CULTIVARS FOR ACCURATE PEACH FRUIT QUALITY AND MATURITY ASSESSMENT

2.1. Introduction

Peach production in the U.S. faces many barriers to success. Poor peach quality has contributed to a decline peach consumption nationally, and poor management has led to considerable waste at all levels of the fresh fruit supply chain. The last two decades have seen peach consumption in steady decline, and the peach industry generates annual waste equating to roughly $428 million per year (Minas et al., 2018; Manganaris et al., 2022). Despite trends in peach consumption, there is strong evidence of consumers' willingness to pay for high quality fruit (Anthony and Minas, 2022). Declining consumption, increased operation costs, and lost profits due to waste have resulted in a diminished and an ever-shrinking industry. Due to the willingness of consumers to pay for high quality fruit and the need to reduce waste, an emphasis has been placed on methods for establishing high fruit quality in the orchard, maintaining high quality after harvest, and developing methods and technology to determine fruit quality and inform management decisions at all stages of the fresh fruit supply chain to improve peach consumption and restore a sustainable industry. Despite the importance of fruit quality, growers are stuck facing the limitations of the slow and destructive methods common in the industry resulting in the lack of knowledge for optimal fruit quality that satisfies both the consumers expectations and shipping/storage needs. A proposed solution to these issues is the use of handheld visible near infrared spectroscopy (Vis-NIRS) sensors to non-destructively predict both maturity and quality in the orchard and after harvest. If this technology proves to be accurate and available, it presents an opportunity to remedy many of these concerns.

Peach quality and storage potential are determined by myriad factors in the orchard which include management of irrigation, fertilization, tree architecture, and crop load, as well as differences in physiological factors such as cultivar and maturity (Minas et al., 2018). Good performance and appropriate utilization of Vis-NIRS technology is dependent upon the consideration of these factors when attempting to train statistical models used in the sensors to generate valuable information for growers.

We hypothesize that given the distinct genetic and phenotypic characteristics of peach cultivars which impact fruit microstructure and other fruit quality characteristics of interest, and the impact of management and environmental factors on maturity/quality that data collected from individual peach cultivars, containing broad variability, will contribute to the robustness of predictive models when combined with data from previous seasons. By training cultivar specific models with variable fruit sample data, accuracy, and robustness of these models will be improved and Vis-NIRS technology will be further validated as a solution to issues that plague growers worldwide.

2.2. Materials and methods

2.2.1. Experimental approach for Vis-NIRs model training, calibration, and validation

Thirteen peach [*Prunus persica* (L.) Batsch.) cultivars were assessed for physiological maturity and internal quality at three maturity stages using destructive methods for reference data collection and non-destructive Vis-NIRS model calibration. The Vis-NIRS model predictions were then validated with destructive fruit quality reference measurements. Internal fruit quality was measured both destructively and non-destructively as dry matter content (DMC) and soluble solids concentration (SSC). DMC was measured as a percentage of the total fresh fruit weight. An initial fresh weight of peach mesocarp was measured before entering a 65 ˚C oven for ~72 hours (Minas et al., 2021). Peach fruit maturity was assessed and measured as the index of absorbance difference (I_{AD}, A_{670}-A_{720}) using a DA-meter® (T.R. Turoni srl, Forlì, Italy) (Ziosi et al., 2008). All Vis-NIRS spectral data scans for model construction and prediction were collected using a Felix F-750 produce quality meter (Felix Instruments, Inc., Camas, WA, USA). With this instrument, scans for non-destructive prediction and future model training were collected simultaneously.

The 13 peach cultivars used for model building and validation in order of harvest date were 'Redhaven', 'Galaxy Donut', 'Newhaven', 'Starfire', 'Glowhaven', 'PF-19', 'Suncrest', 'Glowingstar', 'Blushingstar', 'PF-23', 'PF-24C', 'Angelus', and 'O'Henry'. From each cultivar, the population used for reference data acquisition and spectra collection for model construction and validation was broken into three maturity classes classified by fruit firmness (FF) and measured in Newtons (N): immature fruit (FF > 50 N), commercial harvest (FF = 30-50 N), and tree-ripe (FF < 30 N). Although the non-destructive maturity metric of interest is based on I_{AD}, fruit firmness (FF) is still a relevant metric for fruit maturity in the industry. Minas et al. (2021), observed that I_{AD} is not strongly related to FF. Although I_{AD} and FF are not strongly correlated, I_{AD} can be used to indicate physiological development and can be predicted. FF itself is yet to be well modeled and predicted non-destructively. For this reason, FF was used to sort the fruit initially and I_{AD} was used moving forward.

Fruit populations were sampled from five trees of each cultivar grown in the Colorado State University's (CSU) Experimental Orchard at the Western Colorado Research Center in Orchard Mesa, CO, USA in 2021. Trees grafted onto either 'Lovell' rootstock or 'Krymsk®86' rootstock were planted in 2016 and 2017 at approximately 908 trees per acre, trained to a Perpendicular-V system, and managed in accordance with local commercial standards. From the five selected trees, 100 fruit were randomly sampled in one or two picks and segregated into maturity classes based upon the measurement of FF. Upon segregation, each maturity class consisted of roughly 25 immature fruit, 50 commercial fruit, and 25 tree-ripe fruit.

Shortly after harvest, and prior to destructive quality assessment, two scans per fruit were taken from each side (cheek) of the fruit near the equatorial region. Flesh samples for DMC and FF/SSC were taken at the marked location of each scan. The side of the fruit selected for each sample alternated between the sun exposed side and the shaded side, determined by coloration on either side of the peach suture, for each parameter. By alternating sides, varying internal quality and maturity caused by different amounts of sun exposure (Gullo et al., 2014) was captured. Pre-existing models trained in 2016 were then used to make predictions with the collected scans from each cultivar. These 2016 models consisted of a DMC, SSC, and I_{AD} model for each cultivar used to train the models to produce nine total models (Minas et al., 2021; 2023). The cultivars used in 2016 included the early season cultivar 'Redhaven' (RH), a late-season bi-color cultivar 'Cresthaven' (CH), and an early-season full red-overcolor cultivar 'Sierra Rich' (SR). All 2016 models for each parameter will hereto be referred to as the base models as the training data from those models were incorporated in the training data for models trained for the 13 cultivars in 2021. Following the methodology outlined by Anthony et al. (2023), the 200 scans from each of the 13 cultivars were preprocessed using Savitzky-Golay smoothing and exported as the second derivative of absorbance for the whole spectra (400-1100 nm) at spectral intervals of 3 nm (3 nm spectrometer resolution). All absorbance processing was done using the Vis-NIRS Data Viewer software (Felix Instruments, Inc., Camas, WA, USA). Using the regression coefficients from the base models (RH, CH, SR), the collected spectra were post-processed to produce predicted values for internal quality: DMC and SSC, and maturity: I_{AD}. Model performance and accuracy was then determined via comparison of predicted values to actual values and assessed for goodness of fit, reported as R^2, and root mean square error of prediction (RMSEP). Of the RH, CH, and SR models, the model predicting with the lowest RMSEP was then combined with the destructive and non-destructive data collected from each of the 13 cultivars to create 39 accurate and robust predictive models in total, three models (one for each parameter) per cultivar.

2.2.2. Development of robust multivariate Vis-NIRS prediction models for internal quality and physiological maturity across 13 cultivars of distinct phenotype and harvest time

It was well established that DMC, SSC, and I_{AD} could be accurately predicted with a single scan using the RH, CH, and SR base models from 2016 (Anthony et al., 2023; Minas et al., 2021; 2023). The aim of this study is to determine how PLS models trained with data from 2016 combined with data gathered in 2021 can improve model performance by capturing distinct genetic and phenotypic characteristics as well as distinct environmental factors such as canopy microclimates and seasonal weather. Models trained in 2021 consist of data from one of the 2016 base models and 2021 data from each cultivar. These models are termed cultivar specific models. This definition differs slightly from previous uses of "cultivar specific" in the literature as the models do include data from two cultivars, they are however trained to make predictions on specific cultivars. Fruit utilized for model training of the base models in 2016, were selected from a diverse population to represent a range of internal quality, physiological maturity, and growing conditions. This diverse variation in the training data was selected to add robustness to the models in the prediction of internal quality and maturity of fruit from the same cultivar at various times in the season and across seasons (Minas et al., 2021; 2023). Training data from each of the base models included temperature compensation to address the impact of variable temperatures on wavelength absorbance for different spectra. Temperature compensation was achieved by scanning each fruit twice at three different temperatures (0, 20, 30 ˚C) for a total of six scans per fruit (Minas et al., 2021; 2023). The use of this temperature adjusted data set was used to make models robust against changes in temperature while allowing for focus on quality data from individual cultivars which was not temperature adjusted. In addition, a separate set of internal quality and maturity reference values, and spectral data were collected in 2017 from 'Redhaven', 'Glohaven', 'Suncrest', and 'Angelus', and these data were used for training a three-season model.

For all models, regression coefficients were estimated for each 3 nm spectra intervals between 729 nm to 975 nm for both DMC and SSC (83 total covariates and corresponding coefficients excluding the intercept), while physiological maturity utilized 3 nm spectra intervals range spanning 600 nm to 750 nm (51 covariates). These spectral ranges were selected based on previous studies seeking to determine which spectral ranges best captured absorbance peaks and adequately penetrated fruit flesh. Selecting wavelength ranges from the total near infrared spectrum follows an internal wavelength selection process described by Mishra et al., 2020. The coefficients were estimated using the partial least squares (PLS) model using k-fold cross validation and the NIPLS algorithm. PLS was selected for this study following previous applications of PLS in other non-destructive studies (Minas et al., 2021; 2023; Mishra et al., 2020; Nascimento et al., 2016). PLS has been historically selected for this application for the ability to accommodate highly correlated variables and maximizes covariance in the predictors, particularly when there are a large number of predictors. This approach allows for incorporating large amounts of information in the model without overfitting (Nicolai et al., 2007).

The cultivar specific PLS models were trained with 12 – 15 LVs determined via k-fold cross validation and constructed using the NIPLS algorithm. The optimized number of latent variables was determined based on the number of LV that yielded the highest Q^2, the model statistic used in part to determine the number of LVs, and $R^2_{cv.}$

2.2.3. Assessing quality and maturity of 13 cultivars

Upon completion of harvesting 100 fruit per given cultivar, the F-750 produce quality meter (Felix Instruments, Inc., Camas, WA, USA), a handheld spectrometer with a 3 nm resolution, was used to scan and record absorbance values for each group of 3 nm from 400 nm to 1100 nm. Narrower ranges of 600 – 750 nm and 729 – 975 nm, would later be used as effective ranges for maturity (I_{AD}) and quality (SSC and DMC) respectively in the model construction and the absorbance values within these ranges would then be used to generate predictions.

With the initial scans taken, reference values for I_{AD} were collected before destructive fruit quality analysis was conducted for all other reference values. Taken as close to the center of the portion of the fruit where the Vis-NIRS scan was taken, I_{AD} reference values were taken with the DA-meter® (T.R. Turoni srl, Forlì, Italy). This area where both scans had been taken would become the location where destructive reference internal quality analysis occurred.

Alternating between the sun exposed portion of the fruit and the shaded half of the fruit a section of fresh mesocarp that was previously scanned was taken using a cork borer of 25 mm in diameter and the exocarp was removed. After drying the samples for three days at 65˚ C, dry matter content (DMC) was calculated as the percentage of dry weight over fresh weight. From the other side of the fruit, fruit firmness (FF) was first measured using the fruit texture analyzer (FTA; Guss Manufacturing (Pty) Ltd., Strand, South Africa) after the removal of 1 mm thick exocarp at the location where FF was measured. With the exocarp removed, the FTA inserted an 8mm diameter plunger to a depth of 10 mm. Once FF was measured, the portion of the fruit where firmness was recorded was removed and pressed to extract the fruit juice using a garlic press. This juice was extracted onto a digital refractometer to measure the soluble solids concentration of (SSC) and produce the reference values. This process was used for all reference values to which predictions from each model would be compared.

2.2.4. Statistical analysis of cultivar maturity and quality

To determine the cultivar effect on differences in harvest time, internal quality, and maturity, each parameter was compared between each cultivar. Mean values for each parameter were tested using an ANOVA with the Tukey adjustment at P = 0.05. The JMP Pro 15 (SAS Inc., Cary, NC, USA) statistical software was used for all mean comparisons. JMP output for each PLS model yielded the $R^2_{CV,}$ Q^2, and number of LVs.

The Prism v8.2.1 software (Graph Pad Inc., San Diego, CA, USA) was used to regress the actual and predicted values for DMC, SSC, and I_{AD}. Using this same software, graphs were plotted to visualize the relationship between actual and predicted values. These regressions generated the reported R^2 and RMSEP values used to determine model performance.

2.2.5. Further data set additions to cultivar specific model training sets for three season model

Additional Vis-NIRS scans and reference values for DMC, SSC, and I_{AD}, for the four cultivars, 'Redhaven', 'Glohaven', 'Suncrest', and 'Angelus' from the 2017 season were added to the training data sets of the cultivar specific models for those four cultivars to train new, three season models. These new training data sets for the three season models include representative data from the 2016, 2017, and 2021 seasons. These data were then modeled with PLS in the same manner as the cultivar specific models using the NIPALS algorithm and k-fold cross-validation The estimated regression coefficients were exported from the linear model equation and used to make predictions on fruit scans from the 2022 season.

2.2.6. External model validation from 2022 season

After comparing the predictive accuracy of all three base models and the accuracy of the cultivar specific models, data from the 2022 season were used to externally validate model robustness. 30 fruit for validation was harvested from all 13 cultivars and processed identically to the fruit in 2021. This 2022 fruit was used to validate all models and to assess model robustness by emulating a field application of these predictive models by simulating a grower's or researcher's use of the models to make predictions on fruit from seasons or orchards not represented in the training data. The validation performance of the base, cultivar-specific, and three-season models, was compared and used to assess the impact of multiple seasons, various maturity, and temperatures.

2.3. Results and Discussion

2.3.1. Fruit physiology

The 13 cultivars used in this study span the season from early ripening cultivars beginning with 'Redhaven' to late ripening varieties ending with 'O'Henry'. Eleven other cultivars reached commercial and tree ripe maturity between these two to represent most of the harvest season in western Colorado. It was observed after the classification of fruit, based on maturity defined in this study as the "commercial harvest" stage (FF = 30 – 50 N), that, even when fruit are determined to be of the same maturity class, there are significant differences (at the 0.05 alpha-level) in other quality parameters such as DMC and SSC across cultivars (**Table 2.1**). Given the comparable tree age, training system, crop load, and overall horticultural management of the trees of each cultivar, it is believed that the variability in these quality parameters between cultivars is linked to phenotype and ripening time.

The statistical differences observed between cultivars support the hypothesis that cultivar-specific models may be most appropriate for maximizing the accuracy of a model, by tailoring that model to the physiological variability of the cultivars while keeping growing conditions constant as to not introduce confounding variables from external orchard factors (training system, canopy position, and tree age).

2.3.2. Model complexity

The number of LVs determined using k-fold cross validation for all models regardless of the number of data sets combined as the training data sets ranged from 11 – 15, 10 – 13, and 12 – 15 for DMC, SSC, and I_{AD} respectively for all cultivars (**Table 2.2**). Ten to fifteen LVs are often expected to overfit the data (Peirs et al., 2003; Deng et al., 2021). Concerns of non-linearity are also legitimate with so many LVs needed to model the data. Non-linearity is a possibility with the combination of each distinct dataset from different seasons or cultivars resulting from a lack of homogeneity in the data used to train the model. Concerns of overfitting and non-linearity have implications regarding how generalizable the training data are and determines how well it will perform when used to make predictions for new data (Peirs et al., 2003). Although a legitimate concern based on previous reports, the issue of overfit models and non-linearity appears to be a non-issue in this data based on model accuracy when externally validated.

2.3.3. Base model performance on the 13 new cultivars

With the balance of model accuracy and robustness in mind, the 2016 base model, trained as a global temperature model with three temperature classes (Minas et al., 2021), was selected for each cultivar based on model accuracy (low RMSEP). For each model outcome, DMC, SSC, and I_{AD}, one of the base models was selected and then the data used to fit the model was later included in the cultivar specific model training data set. Of the base global temperature models, the base model trained with 'Redhaven' (Minas et al., 2023) fruit had the lowest error on average for DMC (RMSEP = 0.33 – 1.08%), 'Sierra Rich' (Minas et al., 2021) was the lowest for SSC (RMSEP =

0.44 – 0.98 ˚Brix), and the 'Cresthaven' model had the lowest error when predicting I_{AD} (RMSEP = 0.04 – 0.08) (**Figure 2.1**). An emphasis has been placed on the base models containing the impact of various temperatures on the absorbance of the selected wavelengths, but this data also contains the impact of different crop loads, different rootstocks as a known source of chemical variation in fruit, as well as the impact of different canopy positions on fruit quality (Minas et al., 2021; Peirs et al., 2005). Ultimately the base model offers firstly, global temperature data, and secondly crop load, rootstock and canopy position data, four critical sources of variability for maturity and quality that when accounted for offer robustness to the models when predicting in different orchards and future seasons (Minas et al., 2021). These factors are presumed to contribute to the variation captured with additional LVs after it was seen that the first two LVs capture the majority of the variability from the reference values and the temperatures. Determining good performing models and adding data to update those models, generally yields improvement (Anthony et al., 2023) and follows the methodology outlined by Peirs et al. (2003) and Bobelyn et al. (2010). This breadth of information from the base models was then improved upon with the addition of the 2021 cultivar data adding "depth" in the training data set.

The inclusion of base models was essential for the addition of temperature adjustments as well as rootstock, canopy position and crop load information captured in the fruit samples. The predictions from these base models on 2021 data are external validations. These validations demonstrate that the base models are in themselves robust except in the case of two cultivars 'PF-24' and 'Angelus'. The average prediction errors of these predictions were 0.51% DMC, 0.64 ˚Brix, and 0.06 I_{AD} for all 13 cultivars. This robustness is a likely a product of the inclusion of temperature adjustments (Anderson et al., 2017; Minas et al., 2021), canopy position (Minas et al., 2021), rootstock (Minas et al., 2023) and crop load (Grossman and DeJong, 1995; Anthony et al., 2020; Minas et al., 2021). Without added cultivar-specific information these models will be robust; however, the concern is that they are limited in their robustness against other additional cultivars and seasons of data. Cultivar-specific models were shown to outperform global cultivar models in studies on peach, apple, and mango (Anthony et al., 2023; Zhang et al., 2019; Anderson et al., 2020). Anthony et al. (2023) described the addition of cultivar-specific data to existing models as a viable means of model improvement. This approach was followed for these cultivars in hopes of improving both accuracy and robustness.

2.3.4. Training new cultivar-specific models

Following the approach of Anthony et al. (2023) the best performing base models were combined with ~100 samples of each cultivar to train the new models. After the addition of the 100 samples per cultivar to either of the selected base models and the new models were trained, the R^2 of cross-validation (R^2_{cv}) and RMSE of cross validation (RMSECV) increased and decreased, respectively, compared to the initial R^2 and RMSEP of the initial base model predictions. This is unsurprising given that models will generally perform better when making predictions on data that is contained in the training data compared to data from outside the initial training set. These model performance metrics do not verify that the cultivar-specific models are more accurate or robust, but it does indicate that the new models fit the data well and predict below an error of 1%, a threshold determined by what has been reported as difference in sugar concentration detectable to consumers (Harker et al., 2002).

Cross validation of DMC, SSC, and I_{AD} produced a R^2_{cv} ranging from 0.90 – 0.97, 0.85 – 0.94, and 0.98 – 0.99 respectively. The Q^2 for DMC, SSC, and I_{AD} ranged from 0.88 – 0.96 %, 0.83 – 0.93 ˚Brix, and 0.98 – 0.99 respectively (**Figure 2.2**). Given the absorbed energy by the insoluble carbohydrates ignored when measuring SSC, the PLS models struggle to find the strongest relationship between absorbance and SSC (Kumar et al., 2015). It is also worth noting that the refractometer used to measure the SSC reference value is less accurate than the scale used to measure DMC. With less precision in the initial reference value for SSC, overall accuracy of models predicting SSC may suffer. Since I_{AD} relies on the amount of chlorophyll-A and chlorophyll-B, which have specific absorbance peaks (Ziosi et al., 2008), and therefore less overlapping information in the spectral data, it reasons that this parameter outperforms the others due to less noise and fewer overlapping peaks in the data for I_{AD}. I_{AD} data are unique compared to DMC and SSC in the way that absorbances values will follow a linear trend (R^2=0.98) as chlorophyll degrades during peach fruit maturation (Chalmers and Van den Ende, 1975; Ziosi et al., 2008). Thorough representation of fruit maturity in the 2021 data neatly captures a decline in chlorophyll and for this reason, I_{AD} stands apart from DMC and SSC in the way it, as a physiological phenomenon, has a higher signal to noise ratio and is modeled so well. As observed in other studies, DMC has outperformed SSC in prediction accuracy. As described by Kumar et al. (2015) this is likely due to the inability to distinguish between absorbance values for soluble carbohydrates and insoluble carbohydrates. Making prediction on both soluble and insoluble sugars simultaneously as constituents of dry matter content, the improved model performance of DMC vs. SSC can be seen and understood. Mishra et al. (2021) use NIR to predict moisture content and use the same range of wavelengths. Due to the O-H absorbance bands at these wavelengths from water in the fruit flesh, and that water is most of peach fresh weight, (DMC being the ratio of dry weight to fresh weight), the strong signal from water when scanning fruit for DMC prediction also aids in improved DMC accuracy. From this, it has been concluded that DMC is the stronger parameter for NIRS- based prediction and an appropriate metric for describing fruit quality. Given the strong positive relationship between DMC and SSC described by Palmer et al.

(2010), it is also concluded that information regarding sweetness and consumer acceptance is not being sacrificed by prioritizing DMC over SSC.

Despite some cultivar-specific models sharing training data from the base models, it was observed that some cultivar-specific models performed better than others (**Figure 2.1).** It is likely that the models which performed best were more generalized, which is to say, the training set contained variability in fruit composition which is consistent across cultivars. Looking at the model output, there does not appear to be an obvious trend indicating why certain cultivar specific models outperformed others based on the number of LVs.

2.3.5. Additional data inclusion for model training

To further inspect the influence of season on model robustness and performance, cultivar data from ‘Redhaven’, ‘Glohaven’, ‘Suncrest’, and ‘Angelus’, collected during the 2017 was added to the 2016 and 2021 data to fit a three season DMC and I_{AD} model for each of the four cultivars to produce 8 models representing three growing seasons. These models did not yield notable differences in RMSECV and performed nearly identically to the cultivar-specific models that included only two season of data. For the four cultivars which were refit with the addition of 2017 data, the RMSECV’s from predictions made on 2021 data for DMC with the cultivar specific models were 0.36%, 0.33%, 0.37% and 0.34%. After refitting the models, the three-season model RMSECV predicting on 2021 data for DMC was 0.35%, 0.33%, 0.43%, and 0.54%. The small increase in error is potentially linked to the phenomenon of increasing error as sample size increases (Luo et al., 2018; Teh et al., 2020), but for intents and purposes these error rates when making internal prediction are practically the same. A lack of model improvement with the addition of new seasons of data was observed in a previous study (Peirs et al., 2003) and likely indicates that all relevant variability in the data necessary to capture a large amount of variability in future seasons was present in the initial data set. This indicates that the initial data was already robust against change in season and the addition of the third season may not be necessary.

2.3.6. 2022 External validation

Following the training of the 13 models (for each parameter) in 2021 and additional data collection from 30 fruit from each of the 13 cultivars in 2022, each model was validated with the external 2022 fruit data. These 30 fruit were determined to be at a commercial maturity via I_{AD} values based on the I_{AD} values from the fruit with a 30 – 50 N firmness in 2021. Commercial harvest maturity was selected to assess how applicable these models could be for growers in a commercial environment. Using the regression coefficients from the 2021 cultivar specific models, predictions were made for these 30 fruit (**Figure 2.3**). Prediction accuracy indicated that the cultivar-specific models are robust with average RMSE changing from 0.36% to 0.42 % DMC, 0.59 to 0.63 ˚Brix, and 0.05 to 0.04 of I_{AD} values across the 13 cultivars. The initial error values are the RMSECV from the internal cross validation (2021), and later errors are the RMSEP values they are being compared to are from the external validation of the 2022 season.

During the base model selection phase of model building, it was determined that the base temperature compensated models, were generally robust when making predictions on the 2021 cultivar data (RMSEP on 2021 fruit: 0.41%, 0.64%, and 0.06, for DMC, SSC and I_{AD}, respectively). The 0.41% DMC reported here excludes 'PF-24' and 'Angelus' from the average given the relatively high error observed in 2021 to help show the conserved RMSE between internal and external validation. The exclusion of 'PF-24' and 'Angelus' from this average was justified based on the unprecedented drop in error in 2022 from internal to external validation in the same models. The improved performance of 'PF-24' and 'Angelus' indicates scans taken incorrectly may have been processed causing an initial high error during cross validation but when 2022 scans were used to make predictions and the error dropped. The drop in error suggests that an issue with the scans arose either during collection or processing and highlights the importance of handling scans carefully. An indication that cultivar-specific models were trained well, but a bad scan, or scans, was used to make predictions was again demonstrated when making the final comparison of all models using the data gathered from the 13 cultivars in 2022 as purely external validation (RMSEP on 2022 fruit: 0.42%, 0.68%, and 0.06, for DMC, SSC and I_{AD}, respectively). These error rates show that the model itself is performing well, but some issues existed in some of the 2021 scans. This model accuracy over two seasons demonstrates insensitivity to season and leads to the conclusion that the models that include different arrangements of data which represent a broad range of external factors like temperature and maturity, will be similarly insensitive and thus robust (**Figure 2.4**).

Given the large amount of variability in the data, the number of LVs needed to capture that variability is considered large when compared to the nine LVs prescribed in the literature to avoid overfitting (Deng, 2021; Peirs et al., 2003). The result of this round of external validation, however, indicates that even the DMC and I_{AD} models fit with 15 LVs were not overfit and predicted well (**Table 2.2**). Of the DMC models with 15 LVs, 'Galaxy' was the worst performing with an error of 0.54%. The best performing DMC model with 15 LVs was 'Newhaven' with an error of 0.32%, an error rate which outperformed all cultivars except 'PF-23' (RMSEP: 0.30%) in the external validation of DMC. SSC was modeled with fewer LVs than DMC and I_{AD} (10 – 13 LVs). SSC never performed as strongly as DMC or I_{AD} for reasons discussed previously. All the I_{AD} models were fit with 15 LVs, and all I_{AD} models performed exceptionally well compared to the initial cross validation with the average error decreasing from 0.05 to 0.04 across the cultivars. This decrease in error is assumed to be related to the external validation consisting of the single commercial harvest maturity classification allowing the 2022 data to fall within the spread of the training data.

It is concluded from the performance of these models when predicting quality and maturity parameters in a new season that models trained with a “large” number of LVs are not necessarily overfit simply because of the number of LVs and that models with more than 9 LVs are viable models for making prediction. This does not mean that there is not the potential for models to be overfit with the inclusion of more LVs, but it does underscore the importance of tuning the model to the proper number of LVs with elements such as maximum Q^2. Since Q^2 aids in the determination of the explained variability and indicates when noise in the data is modeled, it offers more clear insight into the internal workings of the PLS model more than the number of LVs does.

Although the models, excluding the base temperature models, lack a second season of external validation to lend additional confidence to the robustness of the models, it is seen that the RMSEP is similar between all of them. Although the considerable improvement of model prediction performance was not observed in the way that was anticipated, it is believed that the models are still robust, and a ceiling of model performance was reached with the amount of variability and the number of samples provided to train the models. Given the range of samples used to train the models and the consistent RMSEP, it is not assumed that the addition of any more samples would necessarily increase model performance, although it is reassuring that a strong decrease in accuracy was not observed with added fruit samples as previously described (Luo et al., 2018). Luo et al. (2018) described this phenomenon in apples, so it is believed that the consistent RMSEP after large increases in sample size is possibly due to the absence of starch in the peaches. This absence of starch in peach produces a more homogenous fruit tissue across the different physiological and environmental factors. Expanding on this line of thinking leads to the thought that homogeneity in the fruit flesh might equate to less non-linearity and better model performance. This is an area that needs to be further investigated in the future for peach and other fruit crops.

2.3.7. Comparison of cultivar-specific model robustness against other cultivars

After it was determined that the cultivar-specific models were shown to be robust against changes in season, the models were used to make predictions on scans from the other cultivars. This was done to explore how robust the models were when considering a situation where growers might use these models in the field to assess cultivars other than the 13 cultivars studied. For this assessment, the four three-season cultivar models were used to make predictions on DMC for the three other cultivars which were not included in the training data for those models. **Table 2.4** shows how the models performed and demonstrates that the models predicted consistently across cultivar. 'Redhaven' had the lowest performance (RMSEP 0.60% - 0.69% DMC) while 'Glohaven', 'Suncrest', and 'Angelus' all predicted with errors below 0.5% DMC. The fact that the models perform consistently irrespective of the cultivar that scans were collected from, suggests that the variability contained in the data set for each model has an impact on how the model performs, and if cultivars are similar enough predictions will be similar (Teh et al., 2020). Ultimately the prediction performance is still accurate enough across both season and cultivar to indicate that the three season models are thoroughly robust and insensitive to a wide range of external factors which might impact predictions including season and cultivar.

2.3.8. Explained variance from PLS based on temperature, maturity, cultivar, and season

It is understood that inclusion of variability in the training data is essential for producing accurate models (Wang et al., 2017; Zhang et al., 2019). It for this reason imperative that researchers designing experiments and collecting data strive to understand what variability in the data is the most informative and influential to reduce needless labor and maximize model performance. For this reason, relationships within each distinct data set were inspected to help determine what the model was able to discern as important variance captured in each LV. These distinct data sets were those that were later used to compose the models for each of the 13 cultivars.

To assess variance and attempt to determine how much each factor (temperature, maturity, cultivar, and season) could be contributing to the model, score plots were made from PLS models that were fit to the specific data sets designed to represent each factor (**Figure 2.5**). These plots show the percentage of explained variance for each LV and a visualization of the position of each data point relative to the others. These score plots are similar to PCA bi-plots although unlike PCA they feature the impact of the output data (DMC, SSC, I_{AD}). This spread helps to indicate what or if factors cause any segregation of data points and to what degree or if some parameters contribute to the distribution of data. Attaching the percent variation to the spread of the data points allows for interpretation of which factors have a meaningful impact on the variation in the data relevant to model robustness.

Previous studies have ordered external factors of significant influence as orchard, season, and temperature, from most to least impactful (Peirs et al., 2003; Teh et al., 2020). From the data included in this study, clear separation can be observed in the effect of temperature in the spectral data from 2016 and some discernable separation when observing maturity, cultivar, and season. The least obvious separation based on PLS score values comes from maturity. Taking a closer look at the explained variation from individual PLS models for each factor shows typical trends in the way in which most of the relationships between the measured variables **X** and the outcomes **Y** are captured within the first few LVs before a smaller percentage of variation is explained by subsequent LVs, and the model begins to incorporate noise in the model. Looking at 'Redhaven' data from 2016, 2017, and 2021 ('Glohaven', 'Suncrest', and 'Angelus' analyzed, data not shown), it can be observed that temperature causes distinct separation contributing to 55%, 61%, and 56% of the variation explained by the second latent variable for DMC, SSC, and I_{AD} respectively. It is worth noting that large amounts of variation in the data is captured by the second LV while the first captures variation in the range of DMC values (**Figure 2.6**). This extends to SSC and I_{AD} where PLS is initially identifying the relationship between the values of **X** and the values of **Y**. It is noteworthy that DMC itself contributes less direct information on the relationship between spectral data and DMC concentration than temperature (22% vs. 55%). In the case of DMC, this expresses the clear relationship between the spectral absorbance and the abundance of carbohydrates and water contributing to the primary absorbance peaks between 729 and 975 nm. Clearly key information does not come exclusively from the parameter of interest, but instead is largely influenced by other external factors.

Analyzing the data used to compose the cultivar models in this way offers a snapshot of the kind of variation PLS can find in the data, and which factors are driving the spread of the data. This is critical as it offers a better understanding of how the models are capturing relationships in the training data. With this understanding, it can be better understood how to apply the models in situations where physiological and environmental factors cannot be known but can be confidently assumed. Examples of these assumptions are instances of field applications of models used when the maturity of the new sample is generally understood to be within the range of maturities represented in the training data or the temperature of the orchard is known to be between the temperatures of the samples used to calibrate the temperature model. With broad representative ranges of the parameter of interest are captured, and possibly more importantly the ranges of physiological and environmental factors are represented, the likelihood that new samples from various temperatures, maturities, cultivars, and seasons fall within the ranges of the samples in the training data. When new samples fall within these ranges, or just outside these ranges, it is most likely that the predicted value of the parameter of interest will have a comparable RMSEP to that of the training data, and the models will then be sufficiently robust.

It is worth noting the likelihood of confounders within this data. Although few Vis-NIR studies mention the potential effects of confounding, it is safe to assume that a number of the selected environmental factors chosen for the variability they add to the data including fruit maturity, training system, canopy position, rootstock, and tree age will impact the carbohydrate content in the fruit and play a role in the way sugars are accumulating in the fruit and how the fruit is maturing (Minas et al., 2018; 2021; 2023). All these factors may cause confounding given the similar roles they play in fruit development are training system, canopy position, and tree age as well as the relationship they have with other physiological, namely vegetative growth processes e.g., canopy development/shading, tree size, ect., all relating to how the tree develops and impacts fruit development (Chalmers et al., 1978; Gullo et al., 2014). Despite the presence of confounders in the data, it is evident that they are not inhibiting the model from making accurate predictions and therefore are not of considerable concern where model performance for field application is concerned. These confounders do, however, make it impossible to know with certainty which of the data sets are primarily responsible for model performance. The low error rates have led to the conclusion that PLS is sufficiently identifying relationships in the data which act as good predictors of quality and maturity.

2.5. Conclusion

From the external validation of the 13 cultivars in 2022 and 2021 and the additional data from 2016 and 2017, it is determined that robust models for DMC, SSC, and I_{AD} can be trained with the inclusion of highly variable temperature, maturity, and cultivar data. Inspecting score plots made from the subsets of the training data revealed that variation in spectral data was determined firstly by factors such as temperature, maturity, cultivar, and season, and secondly by the abundance of DMC, SSC, or I_{AD}. It was then observed that PLS models trained with as many as 15 LVs to capture the spectral variability were robust across seasons despite concerns of overfitting and were in fact not overfit. Models were also tested to assess their ability to make predictions not only on data from new seasons, but also data from new cultivars. The result of this exploration indicated that the models were able to make accurate predictions on new seasons while also showing improvements when making predictions on specific cultivars.

From this exploration of variance within the spectral data from various factors there is room for future applications of artificial neural networks, and other multivariate statistics which may produce more accurate quality and maturity predictions. It may be that these alternative methods are more adept at capturing relationships between the DMC, SSC, and I_{AD} and the impact of environmental and physiological on NIR absorbance resulting in more accurate predictions. Before new models are fit, these PLS models should be further validated with data and reference values from other orchards near Grand Junction, CO, and other fruit growing regions to determine if the models are robust against location.

2.6. Tables

Table 2.1. Effect of cultivar and harvest date on internal peach fruit quality (dry matter content, DMC and soluble solids concentration, SSC) and physiological maturity (index of absorbance difference, I_{AD}) at commercial harvest maturity (fruit firmness, FF =30 – 50 N). ns, *, **, *** indicate no significance or significance at p-values of <0.05, 0.01, 0.0001;
DMC, SSC, and FF were measured on opposite fruit sides (DMC measured on one side, SSC and FF on the other side). Measurements alternated between sun exposed and shaded sides of the fruit; statistical significance was assessed with the Tukey test for mean comparisons; Means in columns with the same letter indicate non-significance at P = 0.05.

Cultivar	Harvest Date	DMC (%)	SSC (%)	I_{AD}	FF (N)
‘Red Haven’	7/27/2021	12.58 ef	11.23 ef	0.32 cd	40.29 a
‘Galaxy’	7/29 – 7/30/2021	15.81 b	15.12 a	0.54 a	40.61 a
‘Newhaven’	7/30 – 8/2/2021	11.85 fg	11.12 ef	0.41 bc	42.41 a
‘Starfire’	7/30 – 8/2/2021	11.49 g	10.03 g	0.26 de	42.71 a
‘Glohaven’	8/10/2021	12.10 fg	10.81 f	0.45 b	41.81 a
‘PF-19’	8/11/2021	13.07 de	11.58 def	0.23 de	41.01 a
‘Suncrest’	8/12 – 8/13/2021	13.82 cd	12.27 cd	0.50 ab	40.84 a
‘Glowingstar’	8/17/2021	12.96 de	12.65 bc	0.21 e	42.23 a
‘Blushingstar’	8/17 – 8/18/2021	12.58 ef	11.61 de	0.19 e	40.25 a
‘PF-23’	8/18/2021	13.56 d	12.44 c	0.26 de	40.99 a
‘PF-24C’	8/19/2021	13.41 d	12.46 c	0.23 de	42.46 a
‘Angelus’	8/25 - 8/30/2021	17.39 a	15.64 a	0.21 de	39.72 a
‘O’Henry’	9/8 - 9/13/2021	14.47 c	13.21 b	0.46 ab	41.11 a
Significance		***	***	***	ns

Table 2.2. 2021 Cultivar specific model parameters for internal peach fruit quality (dry matter content, DMC and soluble solids concentration, SSC) and physiological maturity (index of absorbance difference, I_{AD}). Each cultivar specific model's number of latent variables and the corresponding Q^2, the indicator for model predictive power. Q^2 presented in the table is maximum value (0 – 1) before noise in the data is included in the model as subsequent latent variables. Both the number of latent variables and the Q^2 indicate the amount of variability in the training data explained with each iteration of the NIPLS algorithm. R^2_{cv} indicates summarizes the model fit and k-fold is the number of folds set for the cross validation in order to produce the model value.

	DMC				SSC				I_{AD}			
Cultivar	# LV's	Q^2	R^2_{CV}	k-fold	# LV's	Q^2	R^2_{CV}	k-fold	# LV's	Q^2	R^2_{CV}	k-fold
'Redhaven'	11	0.88	0.90	10	12	0.90	0.92	10	15	0.99	0.99	10
'Galaxy'	15	0.92	0.94	10	13	0.91	0.93	10	12	0.99	0.99	10
'Newhaven'	15	0.90	0.92	10	10	0.85	0.87	10	15	0.99	0.99	10
'Starfire'	13	0.94	0.95	10	12	0.91	0.92	10	15	0.99	0.99	10
'Glohaven'	14	0.93	0.94	10	11	0.91	0.92	10	15	0.99	0.99	10
'Pf-19'	14	0.88	0.91	10	12	0.91	0.92	10	15	0.99	0.99	10
'Suncrest'	13	0.91	0.93	10	11	0.89	0.90	10	15	0.98	0.99	10
'Glowingstar'	15	0.92	0.94	10	13	0.91	0.93	10	15	0.99	0.99	10
'Blushingstar'	14	0.89	0.91	10	10	0.83	0.85	10	15	0.99	0.99	10
'Pf-23'	15	0.88	0.90	10	13	0.88	0.89	10	15	0.99	0.99	10
'Pf-24'	12	0.92	0.93	10	13	0.86	0.88	10	15	0.98	0.98	10
'Angelus'	14	0.96	0.97	10	11	0.93	0.94	10	15	0.99	0.99	10
'O'Henry'	15	0.91	0.93	10	13	0.92	0.93	10	15	0.99	0.99	10
Average	14	0.91	0.93	10	12	0.89	0.91	10	15	0.99	0.99	10

Table 2.3. Model fit and accuracy when making predictions on 100 fruit samples across 13 peach cultivars from 2021 growing season. R^2 and RMSEP for each cultivar and parameter indicate how accurate the predictions were from each model when making predictions on the 2021 cultivar data used to train the model with k-fold cross validation. These values are internal predictions and do not indicate model robustness but do offer insight into how well each model is fit to the data. Given the internal predictions all error rates are lower than would be expected when making predictions on data gathered in a different season. All predictions for DMC and SSC are well below the detectable by consumers 1% difference in internal quality metrics (DMC and SSC), and I_{AD} predictions characterize the physiological maturity of the fruit with precision to classify the fruit according to conventionally maturity classification.

	DMC		SSC		I_{AD}	
Cultivar	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP
‘Redhaven'	0.90	0.36	0.69	0.47	0.97	0.04
‘Galaxy'	0.87	0.42	0.70	0.55	0.97	0.04
‘Newhaven'	0.84	0.34	0.67	0.44	0.99	0.04
‘Starfire'	0.84	0.35	0.58	0.49	0.97	0.03
‘Glohaven'	0.87	0.33	0.69	0.38	0.96	0.05
‘Pf-19'	0.90	0.28	0.77	0.47	0.97	0.03
‘Suncrest'	0.95	0.37	0.90	0.49	0.97	0.06
‘Glowingstar'	0.93	0.32	0.65	0.61	0.89	0.05
‘Blushingstar'	0.92	0.30	0.70	0.61	0.94	0.05
‘Pf-23'	0.89	0.32	0.54	0.52	0.95	0.05
‘Pf-24'	0.85	0.36	0.40	1.12	0.84	0.07
‘Angelus'	0.87	0.54	0.65	0.93	0.96	0.05
‘O'Henry'	0.91	0.40	0.67	0.65	0.98	0.06
Average	0.89	0.36	0.66	0.59	0.95	0.05

Table 2.4. Comparison of 'three-season' models when making predictions on all four cultivar samples from the 2022 growing season. Each of the 'three-season' models was used to make predictions on the cultivar it was trained to measure as well as the other three cultivars used to train the other models. Performance of each model is very consistent across the four cultivars regardless of the model chosen to make predictions. This is interpreted as models being robust against change in cultivar, and that variation in the cultivar prediction data, not the model train data, causes differences in performance.

Model: Three-Season 'Redhaven'

'Redhaven' 2022		'Glohaven' 2022		'Suncrest' 2022		'Angelus' 2022	
R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP
0.67	0.60	0.84	0.29	0.98	0.35	0.95	0.47

Model: Three Season 'Glohaven'

'Redhaven' 2022		'Glohaven' 2022		'Suncrest' 2022		'Angelus' 2022	
R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP
0.64	0.69	0.79	0.37	0.98	0.36	0.97	0.40

Model: Three Season 'Suncrest'

'Redhaven' 2022		'Glohaven' 2022		'Suncrest' 2022		'Angelus' 2022	
R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP
0.66	0.60	0.81	0.34	0.97	0.38	0.97	0.40

Model: Three Season 'Angelus'

'Redhaven' 2022		'Glohaven' 2022		'Suncrest' 2022		'Angelus' 2022	
R^2	RMSEP	R^2	RMSEP	R^2	RMSEP	R^2	RMSEP
0.61	0.61	0.75	0.33	0.97	0.35	0.98	0.29

2.7. Figures

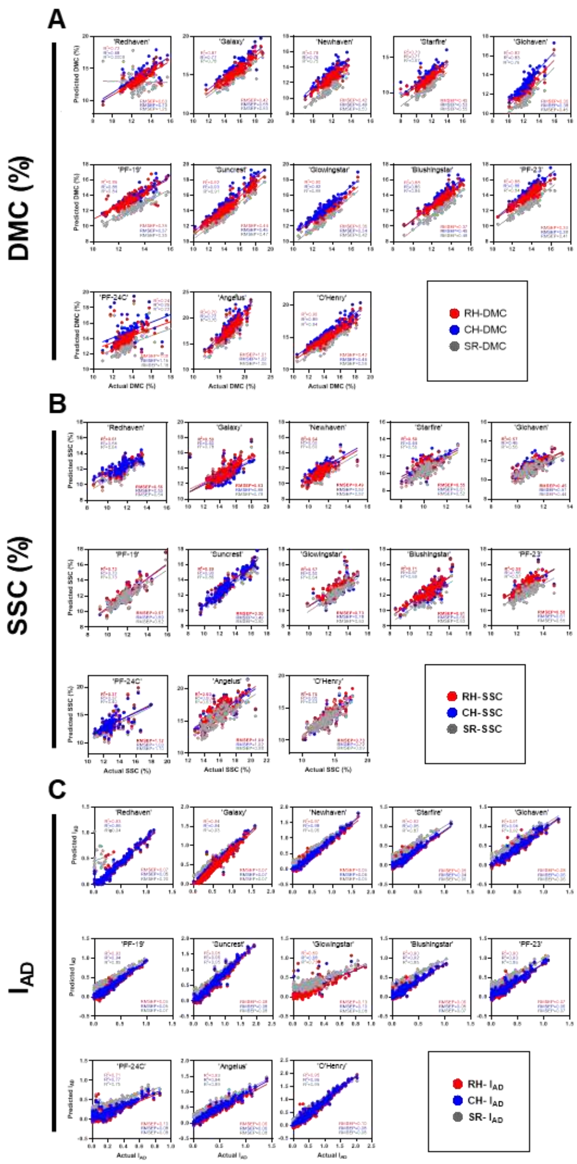

Figure 2.1. Base model (2016) external validation performance for non-destructive internal peach fruit quality (dry matter content, DMC (A) and soluble solids concentration, SSC (B)) and physiological maturity (index of absorbance difference, I_{AD} (C)) assessment across 13 peach cultivars in a different growing season (2021).

Each plot for each parameter and each cultivar shows the performance of each of the three base models built in 2016. The model with the best accuracy was selected as the foundation for cultivar specific models. The 2016 data and cultivar data from 2021 were combined as the training data for the initial cultivar specific models.

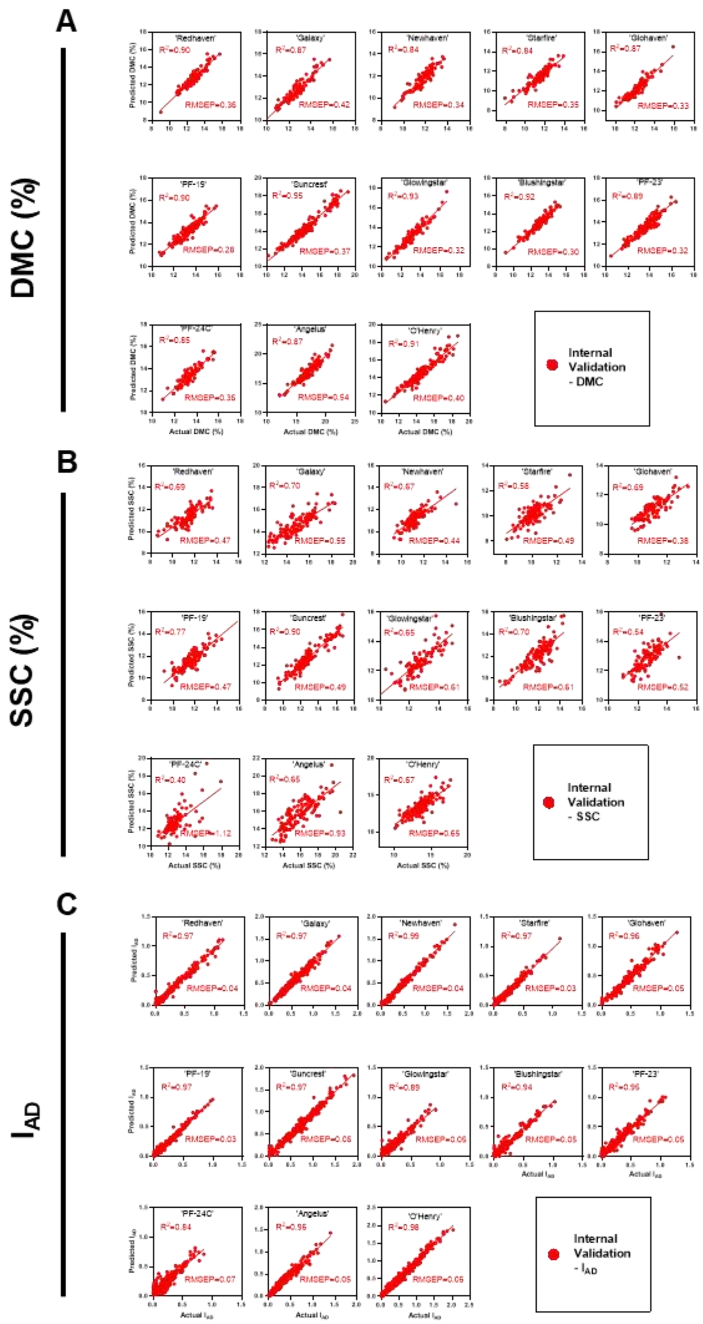

Figure 2.2. Cultivar specific model internal validation performance for non-destructive internal peach fruit quality (dry matter content, DMC (A) and soluble solids concentration, SSC (B)) and physiological maturity (index of absorbance difference, I_{AD} (C)) assessment across 13 peach cultivars during the 2021 growing season.

These plots indicate model accuracy and the relationships between actual and predicted values made with each cultivar model on the 2021 cultivar spectral data. The error from these predictions compared to the errors from the initial base model selection phase indicate improvement however these predictions are not external predictions and should not be compared directly. These predictions do indicate good model fit on the data based on cross validation.

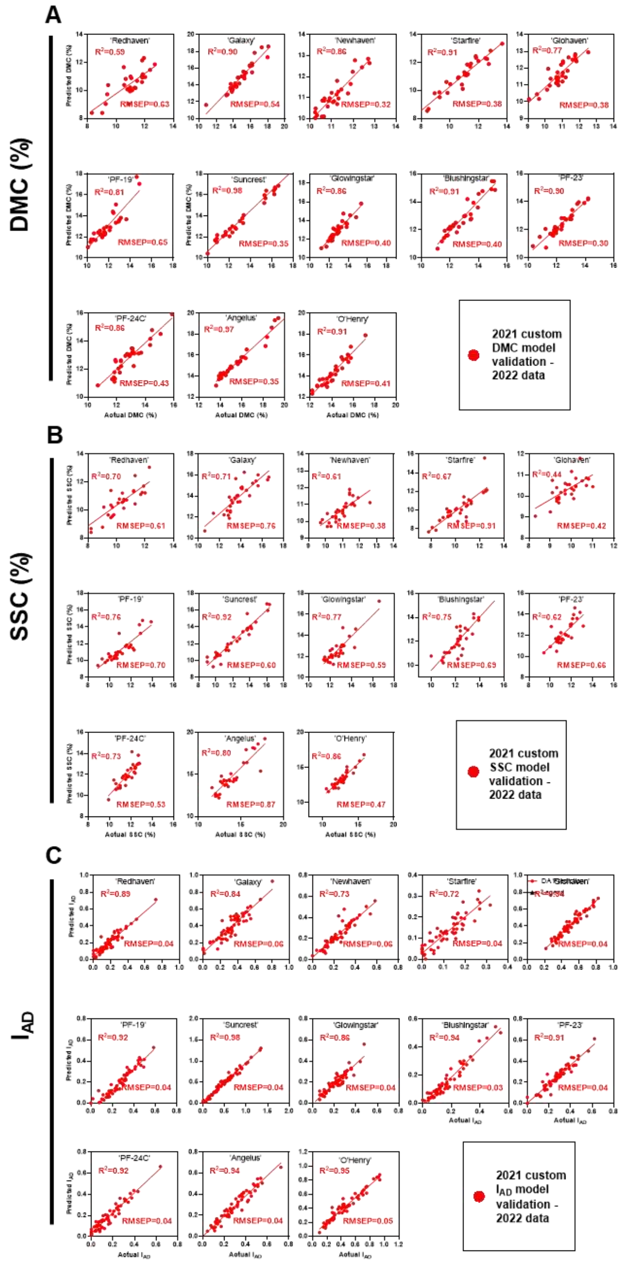

Figure 2.3. Cultivar specific model external validation performance for non-destructive internal peach fruit quality (dry matter content, DMC (A) and soluble solids concentration, SSC (B)) and physiological maturity (index of absorbance difference, I_{AD} (C)) assessment across 13 peach cultivars during the 2022 growing season that was not included in the training set.

Model robustness is shown with the low RMSEP of these actual vs. predicted value plots when models trained with 2016 and 2021 data are validated on new data from 2022. For DMC, SSC, and I_{AD}, error rates have slightly increased compared to internal predictions with 2021 data (**see Figure 2.2**), however, the margins are still within a range that indicate models are performing at a level that could be useful for grower management decision making.

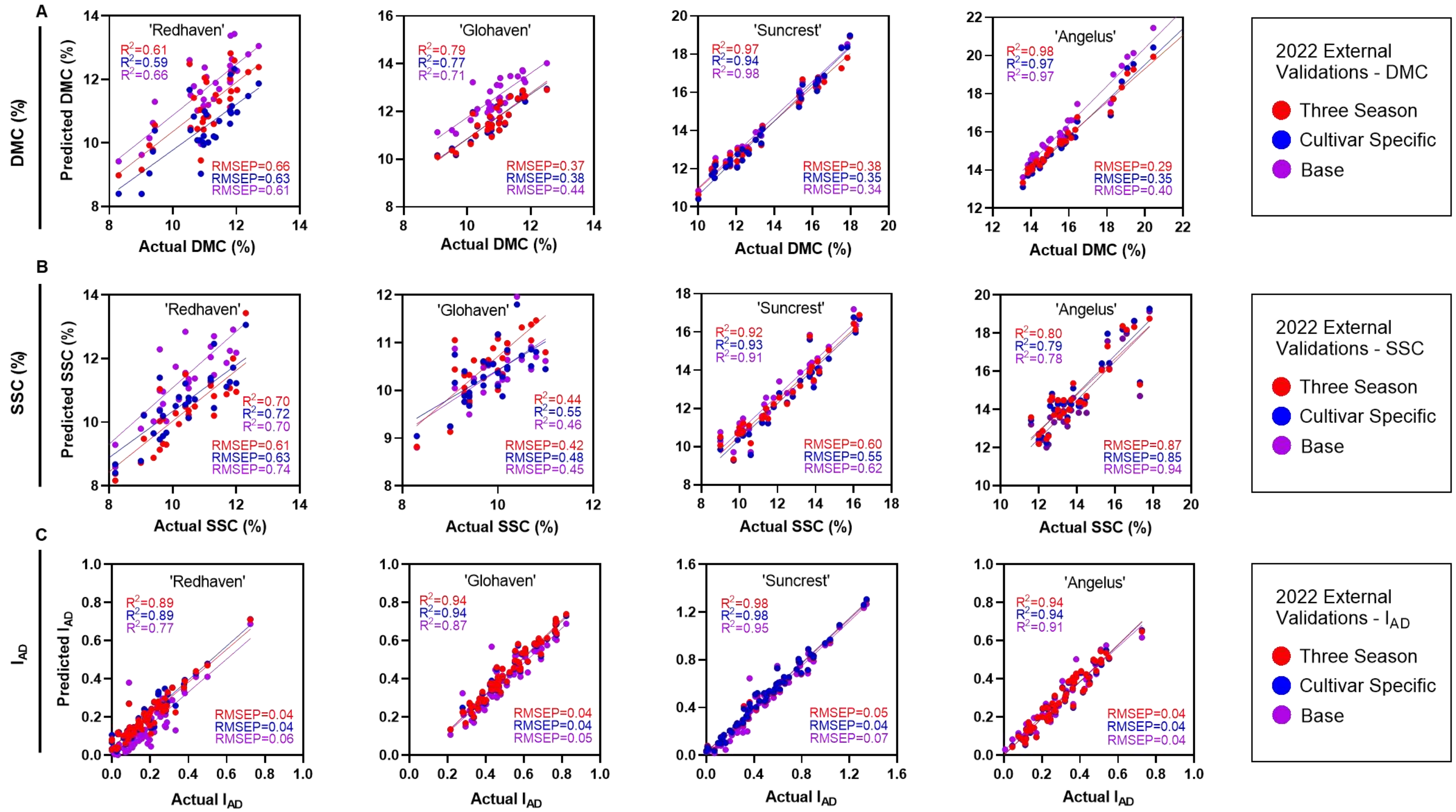

Figure 2.4. Comparison of multi-season (using one: base (2016); two: cultivar specific (2016 and 2021); or three seasons (2016, 2017 and 2021) of data) and cultivar specific model external validation performance for non-destructive internal peach fruit quality (dry matter content, DMC (A) and soluble solids concentration, SSC (B)) and physiological maturity (index of absorbance difference, I_{AD} (C)) assessment across four peach cultivars during the 2022 growing season that was not included in the training set.

Plots showing the similarities in model robustness for the base, cultivar specific, and three season models. The similarity in RMSEP indicates that all models are robust, and that there was not significant improvement from subsequent additions of new data.

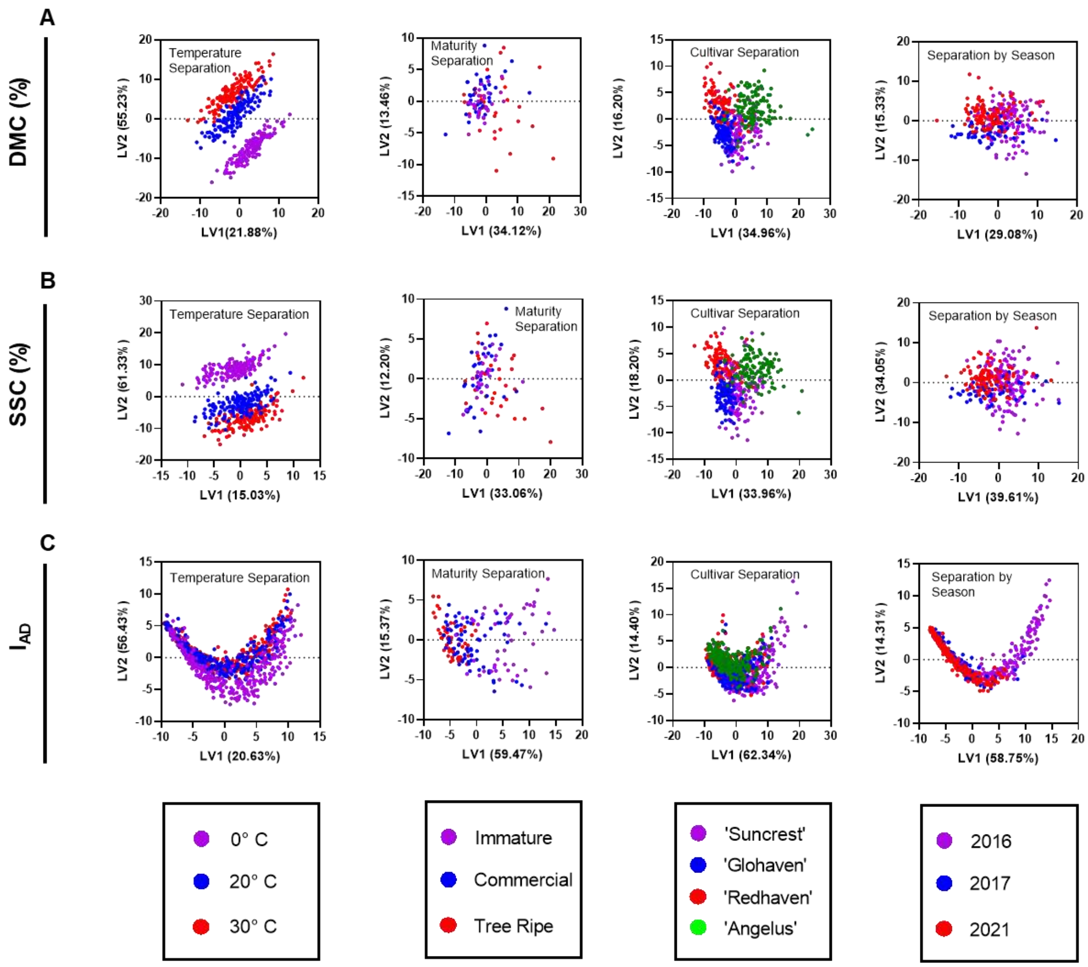

Figure 2.5. Visualization of the spread of data due to the effect of external factors such as temperature, maturity, cultivar and season in the data sets used to train robust models. Each of the base 2016 models was constructed with scans with spectral data at 0, 20, and 30 ˚C for each fruit sample to compensate for the effect of temperature on spectral absorbance. Each temperature plot is a visualization of the variance in the first two latent variables (LVs) of the 2016 'Redhaven' data for each parameter while the maturity and cultivar plots are from the 2021 season. 'Redhaven' was chosen as the representative cultivar. Each cultivar had a similar distributions of data points. The x and y axis indicates the percent of total variance explained by each LV. Points are color coded to indicate how the data is organized within the model between subsets of each factor (temperature, maturity, cultivar and season) based on the relationship between each model variable (600 – 750 nm and 729 – 975 nm) and the model outcome (DMC, SSC, I_{AD}) determined using the NIPLS algorithm. The wide range of variance in the data contributes to how well the model will perform when external data falls within the spread of data reducing or eliminating extrapolation. By inspecting the variance explained by each LV and by visualizing the spread, it can be seen how temperature contributes to the spread of data points along the y axis explaining the majority of data variance from 2016. From the 2021 data, maturity and cultivar variance contributes to the spread although the tight grouping indicates that other factors are likely contributing to the spread of the data. The final column of graphs indicates the variance caused by growing season. The tight grouping shows both that each season is well represented in the data and offers insight into why models are so robust against change in season.

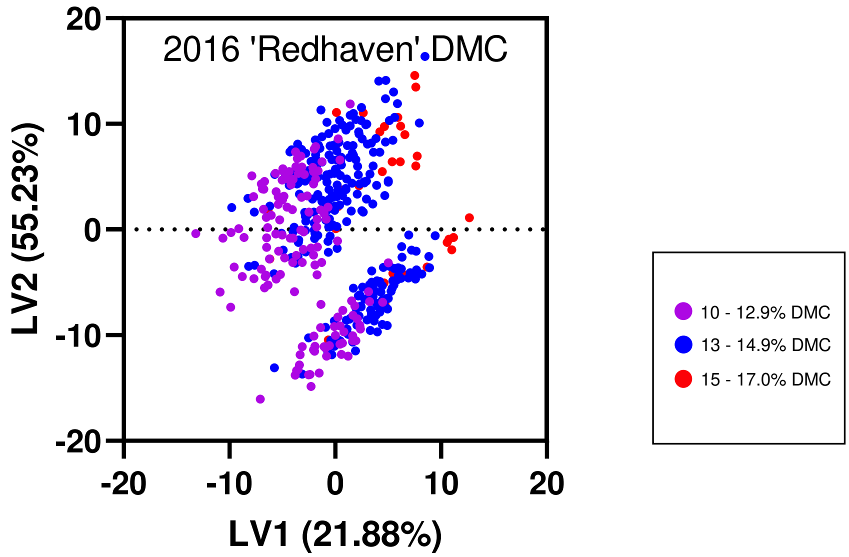

Figure 2.6. Visualization of impact range of DMC has variance and spread of data in the first latent variable (LV).

As seen in **Figure 2.5**, temperature causes a broad spread along the y axis. Here is the same data but now the color coding reflects DMC split into three categories from low to high DMC. Compared to the temperature separation in **Figure 2.5**, DMC is largely responsible for the variation along the x axis.

2.8. References

1. Anderson, N. T., Walsh, K. B., Subedi, P. P., & Hayes, C. H. (2020). Achieving robustness across season, location and cultivar for a NIRS model for intact mango fruit dry matter content. Postharvest Biology and Technology, 168. https://doi.org/10.1016/j.postharvbio.2020.111202

2. Anthony, B.M.; Chaparro, J.M.; Prenni, J.E.; Minas, I.S. (2020). Early metabolic priming under differing carbon sufficiency conditions influences peach fruit quality development. Plant Physiol. Biochem., 157, 416–431. https://doi:10.1016/j.plaphy.2020.11.004.

3. Anthony, B. M., Sterle, D. G., & Minas, I. S. (2023). Robust non-destructive individual cultivar models allow for accurate peach fruit quality and maturity assessment following customization in phenotypically similar cultivars. Postharvest Biology and Technology, 195, 112148. https://doi.org/10.1016/J.POSTHARVBIO.2022.112148

4. 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 Biology and Technology, 55(3), 133–143. https://doi.org/10.1016/j.postharvbio.2009.09.006

5. Chalmers, D.J. and Ende, B.V.D. 1975. A reappraisal of the growth and development of peach fruit. Func. Plant Biol. 2, 623-634.

6. Deng, B. C., Yun, Y. H., Liang, Y. Z., Cao, D. S., Xu, Q. S., Yi, L. Z., & Huang, X. (2015). A new strategy to prevent over-fitting in partial least squares models based on model population analysis. Analytica Chimica Acta, 880, 32–41. https://doi.org/10.1016/j.aca.2015.04.045

7. Grossman, Y. L., & DeJong, T. M. (1995). Maximum fruit growth potential following resource limitation during peach growth. Annals of Botany, 75(6), 561–567. https://doi.org/10.1006/ANBO.1995.1059

8. Gullo, G., Motisi, A., Zappia, R., Dattola, A., Diamanti, J., & Mezzetti, B. (2014). Rootstock and fruit canopy position affect peach [Prunus persica (L.) Batsch] (cv. Rich May) plant productivity and fruit sensorial and nutritional quality. Food chemistry, 153, 234–242. https://doi.org/10.1016/j.foodchem.2013.12.056

9. Harker, F.R., Marsh, K.B., Young, H., Murray, S.H., Gunson, F.A. and Walker, S.B. (2002). Sensory interpretation of instrumental measurements 2: sweet and acid taste of apple fruit. Postharvest biology and technology, 24(3), 241-250.

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

11. Luo, X., Ye, Z., Xu, H., Zhang, D., Bai, S., & Ying, Y. (2018). Robustness improvement of NIR-based determination of soluble solids in apple fruit by local calibration. Postharvest Biology and Technology, 139, 82–90. https://doi.org/10.1016/j.postharvbio.2018.01.019

12. Manganaris, G., Minas, I., Cirilli, M., Torres, R., Bassi, D., & Costa, G. (2022). Peach for the future: A specialty crop revisited. Scientia Horticulturae, 305, 111390, https://doi.org/10.1016/j.scienta.2022.111390

13. Minas, I. S., Anthony, B., Pieper, J., Sterle, D. (2023). Large-scale and accurate non-destructive visual to near infrared spectroscopy-based assessment of the effect of rootstock on peach fruit internal quality, European Journal of Agronomy, Volume 143, 1161-0301, https://doi.org/10.1016/j.eja.2022.126706

14. Minas, I. S., Blanco-Cipollone, F., & Sterle, D. (2021). Accurate non-destructive prediction of peach fruit internal quality and physiological maturity with a single scan using near infrared spectroscopy. Food Chemistry, 335. https://doi.org/10.1016/j.foodchem.2020.127626

15. Minas, I. S., Tanou, G., & Molassiotis, A. (2018). Environmental and orchard bases of peach fruit quality. In Scientia Horticulturae (Vol. 235, pp. 307–322). Elsevier B.V. https://doi.org/10.1016/j.scienta.2018.01.028

16. Mishra, P., Woltering, E., Brouwer, B., & Hogeveen-van Echtelt, E. (2021). Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach. Postharvest Biology and Technology, 171. https://doi.org/10.1016/j.postharvbio.2020.111348

17. Nascimento, P. A. M., Carvalho, L. C. de, Júnior, L. C. C., Pereira, F. M. V., & Teixeira, G. H. de A. (2016). Robust PLS models for soluble solids content and firmness determination in low chilling peach using near-infrared spectroscopy (NIR). Postharvest Biology and Technology, 111, 345–351. https://doi.org/10.1016/j.postharvbio.2015.08.006

18. Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. In Postharvest Biology and Technology (Vol. 46, Issue 2, pp. 99–118). https://doi.org/10.1016/j.postharvbio.2007.06.024

19. Palmer, J. W., Harker, F. R., Tustin, D. S., & Johnston, J. (2010). Fruit dry matter concentration: A new quality metric for apples. Journal of the Science of Food and Agriculture, 90(15), 2586–2594. https://doi.org/10.1002/jsfa.4125

20. Peirs, A., Scheerlinck, N., & Nicolaï, B. M. (2003). Temperature compensation for near infrared reflectance measurement of apple fruit soluble solids contents. Postharvest Biology and Technology, 30(3), 233–248. https://doi.org/10.1016/S0925-5214(03)00118-2

21. 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 Biology and Technology, 35(1), 1–13. https://doi.org/10.1016/j.postharvbio.2004.05.010

22. Peirs, A., Tirry, J., Verlinden, B., Darius, P., & Nicolaï, B. M. (2003). Effect of biological variability on the robustness of NIR models for soluble solids content of apples. Postharvest Biology and Technology, 28(2), 269–280. https://doi.org/10.1016/S0925-5214(02)00196-5

23. Teh, S. L., Coggins, J. L., Kostick, S. A., & Evans, K. M. (2020). Location, year, and tree age impact NIR-based postharvest prediction of dry matter concentration for 58 apple accessions. Postharvest Biology and Technology, 166. https://doi.org/10.1016/j.postharvbio.2020.111125

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

25. Wold, S., Sjostrom, M., Eriksson, L., & Sweden˚˚, S. (2001). PLS-regression: a basic tool of chemometrics. In Chemometrics and Intelligent Laboratory Systems (Vol. 58). www.elsevier.comrlocaterchemometrics

26. Zhang, Y., Nock, J. F., Al Shoffe, Y., & Watkins, C. B. (2019). Non-destructive prediction of soluble solids and dry matter contents in eight apple cultivars using near-infrared spectroscopy. Postharvest Biology and Technology, 151, 111–118. https://doi.org/10.1016/j.postharvbio.2019.01.009

27. Ziosi, V., Noferini, M., Fiori, G., Tadiello, A., Trainotti, L., Casadoro, G., & Costa, G. (2008). A new index based on vis spectroscopy to characterize the progression of ripening in peach fruit. Postharvest Biology and Technology, 49(3), 319–329. https://doi.org/10.1016/j.postharvbio.2008.01.017