

Using climatic and imaging data to predict apple phenology

Haidee Tang

BSc, MS

Thesis Submitted for the Degree of Doctor of Philosophy

School of Computer Science and Electronic Engineering

University of Essex

September 2025

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Haidee Tang

September 2025

Acknowledgments

The pursuit of my PhD has been extremely challenging, but it has equally been rich in learning, development and achievement. I have my supervisors, Prof. Xiangming Xu and Dr. Xiaojun Zhai, to thank for their wisdom, patience, support and guidance throughout the course of my PhD. Prof. Xu's abundant knowledge and ability to clearly explain complex concepts have been invaluable to my development. Despite his many commitments, he always made time for me, consistently motivating me to maintain momentum in my work. His thoughtful insights and steady encouragement provided clarity in moments of uncertainty. I will always appreciate the way he made everything sound so easy, turning challenges into manageable steps. Dr. Zhai's continuous encouragement and friendly attitude has supported me throughout my PhD. His ability to highlight my achievements helped me stay motivated and confident in my work. His approachable nature created a positive and uplifting environment, despite the remote nature of our communication. I am truly grateful for the optimism and reassurance that he brought to this experience.

I also want to thank my fellow PhD students, Hamish McLean and Katie Stewart (the apple folk), and Will Atkinson for their invaluable company and keeping me sane through the toughest and busiest times. To the colleagues in the Pests, Pathology and Entomology team (PPE), thank you for always looking out for me and providing me support and assisting me with my experiments whenever I needed it. In particular, a great thank you to Tom Passey, Jennifer Kingsnorth, Kirsty McLeary for looking after my flowering apple trees when I wasn't in the UK.

I would also like to thank the New Zealand Institute of Bioeconomy Science Institute (formerly Plant and Food Research) in Hawke's Bay, New Zealand, for allowing me to conduct part of my research at their facilities. It was an invaluable opportunity to expand my dataset. In particular, Nicolette Niemann and Jason Johnston for organising lab space and apple samples, Hannah Lloyd and Tattersall for their help and guidance with my work and giving me a place to stay, Anna Kokeny for being an awesome flatmate.

To my husband Jeffrey, who has been my greatest supporter, I thank you for your encouragements and affirmation during the highs and the lows of this journey. Thank

you for moving half way across the world with me, shouldering the responsibilities of our household and being a source of constant comfort and safety. Thank you for taking care of me by feeding me and ensuring I go to the gym. Without you the last four years would have been completely different. I can't wait to start the next chapter of our life together.

Thank you to my parents, for believing in me and being so supportive of me moving to the UK. To my in-laws, who have been so proud of my achievements. To my siblings, Teresa, Amy, Derek, Maggie and Carol, thanks for your kind words whenever we had a chance to connect.

Publications

Journal papers:

1. Tang, H., Zhai, X., & Xu, X. (2025). Flowering variation induces apple maturity variation at harvest. *Frontiers in Agronomy*, 7, 1545070.
2. Tang, H., Zhai, X., & Xu, X. (2024). Evaluating the performance of models predicting the flowering times of twenty-six apple cultivars in England. *European Journal of Agronomy*, 160, 127319.
3. To be submitted: Toward accurate prediction of apple firmness and Brix across countries, seasons and cultivars with hyperspectral imaging.
4. To be submitted: A multi-cultivar apple hyperspectral dataset: a resource for predicting apple firmness, brix and starch content

Summary

This thesis focuses on using phenology models and deep learning models on hyperspectral images to predict the optimal harvest window and the fruit maturity, respectively. Conventional maturity assessments require a lot of time and manpower, yet they are destructive and can only estimate the harvest window within seven days of harvest at the orchard-level. A need for a quicker, non-destructive method that can achieve maturity estimates per fruit will enable apples to be picked at their optimum maturity. Improved fruit maturity can improve shelf-life and the final fruit quality at market. Chapter 1 introduces the current methods in the literature and their limitations. In Chapter 2, this research aimed to evaluate the accuracy of the PhenoFlex phenology model and determine the best method of parameterisation when dealing with numerous cultivars. The 3rd chapter focused on the shortcomings of phenology models that predict only the average flowering date. The effect of flowering variation on fruit maturity was discussed. In Chapter 4, we explored the use of hyperspectral imaging with deep learning models to predict apple maturity features. The main contributions of our studies showed that phenology models can be applied at the species level, generalising the model across all apple cultivars. Flowering time can add a large degree of variation to fruit maturity and should be considered when estimating the harvest window. Lastly, the large diverse hyperspectral dataset collected across cultivars, seasons and countries was used to train deep learning models. Key wavelengths were identified in our study, which can be used to create simpler cameras, making imaging more accessible to growers. In Chapter 5, we discuss an alternative framework that uses phenology models and imaging to accurately and non-destructively predict the optimal harvest window for apples.

Abstract

This thesis investigates the biological and environmental factors influencing apple maturity, with the aim of developing a non-destructive framework for accurate harvest prediction. It begins by reviewing phenology models commonly used to estimate flowering and harvest timing and introduces imaging as a promising alternative to directly assess apple maturity.

Chapter 3 focuses on the parameterisation of multiple cultivars to determine the level of specificity during model fitting. This has not previously been done before, mainly due to the lack of training data. In this chapter, the PhenoFlex model is applied to a unique 85-year dataset of 26 apple cultivars grown at East Malling. The study demonstrates that generic (species-level) or grouped (based on flowering time similarity) parameterisation approaches can effectively predict flowering time, reducing the need for cultivar-specific calibration, especially with limited data.

Chapter 4 explores the impact of flowering time variation on fruit maturity. The temperature experienced by individual flower clusters was tracked until their corresponding fruit was harvested. The Growing Degree Hours model was applied to each flowering to harvest period to determine the influence of temperature on growth. The variation in maturity was assessed considering flowering time, season, tree and canopy region, and it was found that flowering time accounts for up to 20% of maturity variation, depending on the cultivar. Smaller effects were observed from seasonal and tree-level effects. These findings highlight the limitations of using average flowering dates in harvest models and support the need for more precise, fruit-level assessments starting from the earliest possible harvest date.

Chapter 5 evaluates hyperspectral imaging as a non-destructive method

for assessing apple maturity. A large and diverse dataset of 5,756 apples was collected from different cultivars, seasons and countries. This dataset is approximately 1800% larger than the datasets used in previous studies studying apple traits. The Vision Transformer architecture achieved the highest accuracy in predicting Brix and firmness. Key spectral and spatial features were identified, and cultivar-specific information improved model performance. Imaging a single side of the fruit was insufficient; the model that was trained on images from all four sides yielded better results.

The thesis concludes by discussing the results found in each chapter and proposes a way to integrate phenology and imaging into a framework for harvest window predictions.

Contents

List of Figures	xiii
List of Tables	xvii
1 Introduction	1
1.1 Background	1
1.2 Research motivation	2
1.3 Contributions and thesis structure	3
1.3.1 Chapter 2	3
1.3.2 Chapter 3	4
1.3.3 Chapter 4	4
1.3.4 Chapter 5	5
1.3.5 Chapter 6	7
2 Literature Review	8
2.1 Conventional methods for predicting apple maturity	9
2.1.1 Temperature effects on dormancy	10
2.1.2 Approaches of flowering time models	11
2.1.3 Temperature effects on fruit development	12
2.1.4 Approaches of harvest date models	13
2.1.5 Flowering time effects on fruit maturity	13
2.1.6 Maturity assessments	14
2.2 Imaging	15
2.3 Summary	17

3	Evaluating the performance of models predicting the flowering times of twenty-six apple cultivars in England	18
3.1	Introduction	18
3.2	Materials and Methods	23
3.2.1	Flowering data	23
3.2.2	Temperature data	23
3.2.3	Model formulation	26
3.2.4	Model optimisation and performance evaluation	28
3.2.5	Comparative modelling	29
3.3	Results	30
3.3.1	PhenoFlex models fitted to individual cultivars	30
3.3.2	PhenoFlex models fitted to groups of cultivars as identified by mean flowering dates and variation across years	33
3.3.3	A common PhenoFlex model fitted to all cultivars	36
3.3.4	Comparison of the common PhenoFlex and StepChill models	39
3.4	Discussion	40
3.5	Conclusion	44
4	Flowering variation induces apple maturity variation at harvest	46
4.1	Introduction	46
4.2	Methods	49
4.2.1	Plants	49
4.2.2	Flowering records	49
4.2.3	Tree canopy zones	49
4.2.4	Temperature records	51
4.2.5	Maturity measurements	51
4.2.6	Models describing temperature effects	52

4.2.7	Assessment of the relative importance of experimental factors in fruit maturity	54
4.3	Results	55
4.3.1	Variability of temperature and effects on flowering	55
4.3.2	Fruit maturity measurements	57
4.3.3	Temperature-based fruit development	57
4.3.4	Factors contributing to maturity variation	63
4.4	Discussion	65
4.4.1	Harvest predictions and flowering effect	65
4.4.2	Year, tree, region and cultivar effects on maturity variation	66
4.4.3	Evaluation of linear GDH, non-linear GDH and Thermodynamic models	67
4.4.4	Assumptions and limitations	69
4.4.5	Implementations	70
4.5	Conclusion	70
5	Toward accurate prediction of apple firmness and Brix across countries, seasons and cultivars with hyperspectral imaging	71
5.1	Introduction	71
5.2	Material and Methods	73
5.2.1	Fruit Samples	73
5.2.2	Obtaining fruit images and initial image processing	75
5.2.3	Image Processing	76
5.2.4	Deep Learning Models	78
5.2.5	Hyperparameter Tuning	78
5.2.6	Model Evaluation and Feature Selection	80
5.2.7	Hardware and software environment	81

5.2.8	Modelling strategies to assess model performance on maturity features	81
5.3	Results	83
5.3.1	Analysis of maturity parameters	83
5.3.2	Analysis of Spectral Data	83
5.3.3	Model Training Analysis	86
5.3.4	Seasonal Comparisons	91
5.3.5	Single-sided Analysis	91
5.3.6	Shapley Analysis	93
5.4	Discussion	96
5.4.1	Cultivar and Seasonal Information	98
5.4.2	Model Results	99
5.4.3	Important Channels	100
5.4.4	Regions of the Fruit	101
5.4.5	Single-sided Analysis	102
5.4.6	Pre-processing Method	103
5.4.7	Applications	103
5.4.8	Limitations	103
5.5	Conclusion	104
6	Conclusion and Discussion	105
6.0.1	Contributions	108
	References	111

List of Figures

- 2.1 Objective points alongside phenological periods from winter to harvest. 9
- 2.2 Starch iodine indexes in apple [1]. An iodine mixture is used to spray halved apples to determine fruit ripeness from 1 - least ripe to 8 fully ripe. In common practice, the optimal harvest dates for long-term storage are around stages 3 and 4. 15
- 3.1 Year range for each of the twenty-six cultivars, shown as black bars. The data highlighted in grey indicate simulated temperatures using latitude and maximum and minimum daily temperatures. The unhighlighted area represents temperature data from real recorded hourly temperatures, supplemented with simulated data where there are missing hourly values. The dashed lines indicate the specific years in which the three days and five hours of missing data occurs. 25
- 3.2 A comparison of the mean observed and predicted bloom dates using cultivar-specific parameters on the PhenoFlex model on A) training data and B) test data. The dashed line represents the line of equality or the $y = x$ relationship between the x and y coordinates. The square points in B represent cultivars which have RMSEs greater than 13 days. 32
- 3.3 K-means clustering presented in a PCA plot for mean flowering time and flowering variation across years for the twenty-six cultivars. B) Silhouette plot indicating three optimal clusters. 35

-
- 3.4 A comparison of the observed and predicted bloom dates of cultivars grouped by K-means clustering on mean flowering and variation on the PhenoFlex model for A) training data and B) test data. The dashed line represents the line of equality or the $y = x$ relationship between the x and y coordinates. 36
- 3.5 A comparison of the observed and predicted bloom dates using common parameters on the PhenoFlex model on A) training data and B) test data. The dashed line represents the line of equality or the $y = x$ relationship between the x and y coordinates. 37
- 3.6 Comparison of RMSE of the cultivar-specific, mean flowering grouped and common models for A) the training data and B) the test data. The graphs are separated by whether the specific (S), grouped by mean flowering date (M) or common (C) models perform better. When two of the approaches perform equally well, two letters are shown (M&S and C&M) or whether there is no difference between approaches (All). 38
- 3.7 A comparison of the observed and predicted bloom dates using common parameters on the StepChill model on A) training data and B) test data. The dashed line represents the line of equality or the $y = x$ relationship between the x and y coordinates. 39
- 4.1 The frequency of temperatures between 0 and 40°C at 0.5°C increments from the first flowering to harvest for the five apple cultivars in 2022 (black bars) and 2023 (grey bars). 56
- 4.2 Strief vs soluble solids (Brix), firmness and starch measurements of all apple cultivars collected in 2022 and 2023 from a single orchard. Note: Cox only produced fruit in 2023 and samples were collected from three trees per cultivar, except for Fuji which only had two. 58

4.3	Growth rate of apple cultivars for hourly exposure to temperatures between 0 and 40°C using A) linear GDH, B) non-linear GDH and C) Thermodynamic models.	59
4.4	Proportion of starch (1 – immature and 0 – mature) against relative growth units calculated using A) linear Growing Degree Hours, B) non-linear Growing Degrees Hours and C) Thermodynamic model for each apple cultivar across 2 years. The trend lines show the trends of the values from 2022 and 2023 for each cultivar on each model	62
5.1	Schematic of the hyperspectral imaging equipment setup. Two lights on stands were set up on either side of a camera on a tripod, facing the light diffusion tent containing the apples. The lights and tent formed a triangle with sides of ~1m length.	76
5.2	Example of a swapped quadrants image. The four corners were swapped diagonally to simulate whether apple regions are important.	77
5.3	2D-Convolutional Neural Network architecture.	79
5.4	3D-Convolutional Neural Network architecture.	79
5.5	Hybrid model architecture, consisting of 2D-CNN and Transformer blocks.	79
5.6	Vision Transformer model architecture.	80
5.7	Model training and testing workflow. Solid rectangles represent research questions, while diamonds represent the options tested for each query. Pink lines depict the experimental pathway, with annotations along each line representing the rationale behind successive choices. The last four boxes, connected by the yellow lines, show the outcome of the varied inputs.	82
5.8	Pairwise plot between Brix, firmness and starch, coloured by each cultivar.	85
5.9	Mean reflectance of six apple cultivars between 400-1000 nm.	86
5.10	Predicted vs Actual values of A) Brix, B) Firmness and C) Starch on the optimised trained models. The red dashed line represents the x=y line.	90

-
- 5.11 Prediction vs actual plot of A) Brix, B) Firmness and C) Starch models on a single side of apple data. The red dashed line represents the $x=y$ line. 93
- 5.12 Mean spectral importance. 1) Mean Shapley values across wavebands and 2) the top 20% most significant wavebands for the A) Brix, B) firmness and C) starch models. 94
- 5.13 Predicted vs Actual plot of A) Brix, B) Firmness and C) Starch models trained with 50% reduced wavelengths. The red dashed line represents the $x=y$ line. 95
- 5.14 Mean spectral importance (Shapley values) per pixel for all apples for A) Brix, B) firmness and C) starch models. The black margin in starch is due to the image size (40-by-40 pixels) being divided into patch sizes of 9-by-9 pixels. 97
- 5.15 Mean spectral importance (Shapley values) per pixel for all apples for A) Brix, B) firmness and C) starch models trained on swapped images. The black margin in starch is due to the image size (40-by-40 pixels) being divided into patch sizes of 9-by-9 pixels. 97
- 6.1 Graphical representation of applications. The phenological phases are represented in the top brown bar, and the applications and their applied range are shown below as red and blue bars. 109

List of Tables

3.1	Summary of the range of flowering time data available from East Malling, the total number of years and number of datapoints used to train and test each model for each of the twenty-six cultivars. The table is split by K-means clustering on mean flowering dates and variation across years. The last column represents the standard deviation of flowering.	24
3.2	The best fit parameters of the PhenoFlex model for twenty-six apple cultivars, each one optimised by generalised simulated annealing algorithm with 99 bootstraps.	31
3.3	Average RMSE and RPIQ for each model parameterisation approach from training and test datasets.	33
3.4	Analysis of bloom dates and training data across cultivar groups which showed either more accurate predictions with common, specific, or mean flowering time models or whether they are similar. Some only had one cultivar so no standard error was calculated.	34
3.5	The best fit parameters of the PhenoFlex model for the K-means groups and Common (all cultivars). The standard errors are calculated on 99 bootstraps and the original parameters.	43
4.1	Summary of fruit and cluster quantities across different trees and zones in 2022 and 2023. Three trees of each cultivar were used in the study, except for Fuji. Collection of fruit occurred in 2022 and 2023 but Cox was only collected in 2023. Cluster represents the total flower clusters collected from each cultivar, year and tree.	50

4.2	Summary of flowering data from 2022 and 2023 for 5 apple cultivars.	55
4.3	Linear Growing Degree Hour model parameters estimated by the best correlation (Kendall's Tau) to Starch. Where multiple combinations result in the best correlation, the parameters presented are the closest to the median values. The errors represent the standard deviation of the best correlated parameters.	60
4.4	Non-linear Growing Degree Hour model parameters estimated by the best correlation (Kendall's Tau) to Starch. Where multiple combinations result in the best correlation, the parameters presented are the closest to the median values. The errors represent the standard deviation of the best correlated parameters.	60
4.5	Thermodynamic model parameters estimated by the best correlation (Kendall's Tau) to Starch. Where multiple combinations result in the best correlation, the parameters presented are the closest to the median values. The errors represent the standard deviation of the best correlated parameters.	61
4.6	The percentage of the total variation explained by the linear GDH, non-linear GDH and Thermodynamic models for all cultivars. The values represent the percentage explained by each variable. Chi-squared test was used to determine the significance of each variable.	64
5.1	Image and apple count per season split by cultivar. Numbers represent pre-filtering counts.	74
5.2	Quality distribution table of apples by cultivar from 2023-2024. The columns represent the mean and standard deviation of the mean for firmness, soluble solids(Brix) and starch. Firmness is the average firmness of the two sides of the fruit. The rows represent the six apple cultivars in this study. Each cultivar is calculated for each season separately and with all seasons together.	84

5.3	Mean dimensions of apple cultivars in pixels	87
5.4	Model summary results for Brix, Firmness, and Starch	88
5.5	Bayesian Optimisation results for Brix, firmness and starch models. The default represents the parameters used prior to optimisation.	89
5.6	Floating Point Operations (FLOPs) calculated for each model before and after Bayesian optimisation	89
5.7	Cross-validation of seasonal data on other seasons data. The values are the average RMSE computed across five randomly sampled subsets of the test dataset.	92

Introduction

1.1 Background

The apple (*Malus domestica* (Suckow) Borkh), a widely consumed fruit from the Rosaceae family, is one of the most extensively cultivated fruits worldwide [2, 3]. In the United Kingdom, apples rank among the top 20 agricultural commodities, producing 407,770 tonnes with an estimated production value of £140,501,640 in 2023 [4]. An estimated 10-50% of apples produced go to waste [5, 6], resulting in significant financial losses and implications for food security. Fresh fruit ranks the lowest in the UK's food to supply ratio [7], showing a heavy reliance on imports to meet domestic demand. These suggest a need for interventions to improve fruit quality and reduce food loss in production.

Apple quality is crucial for market value and consumer satisfaction. Quality is influenced by physical traits and chemical traits (sugar content and bioactive compounds). The UK marketing standard guidelines indicate that apples should be intact, clean, free from pests, odd flavours, and external moisture; at most slightly damaged, and at an appropriate maturity when they reach the markets [8]. Growers can manage most of these guidelines at the post-harvest stage, but some guidelines require intervention at the harvest stage. Improving the quality of apples is important for minimising fruit loss in the production process [9] and can increase the market value [10]. Higher quality apples require fewer fruit to be harvested to achieve the same value as lower

quality fruit, occupy less space in cold storage, and are less likely to deteriorate during storage. These advantages offer both economic and ecological benefits, reinforcing the importance of enhancing apple fruit quality. Harvesting apples at the appropriate maturity influences the quality and shelf life of apples. Previous research has found that apples that have started ripening are more likely to develop post-harvest storage disorders such as rots and superficial scald due to softer fruit in long-term storage [11, 12]. However, if the fruit are harvested too early, they are physiologically and biochemically underdeveloped, resulting in smaller, flatter fruit with impaired flavours [11, 13].

Harvest maturity has been known to influence post-storage fruit quality. However, it is difficult to determine the optimal harvest maturity with conventional maturity assessments. The current methods are destructive, labour-intensive and limited in forecasting ability. They typically provide orchard-level insights only a week before harvest, which restricts efficient resource allocation and may include underdeveloped or overripe fruit. To overcome these limitations, there is a need for non-destructive approaches that can accurately forecast the optimal harvest window earlier than current practices.

1.2 Research motivation

Conventional apple maturity assessments rely on destructive sampling methods that are labour-intensive, time-consuming and often subjective, particularly for starch measurements. These methods typically offer forecasts within seven days before harvest and provide orchard-level insights that fail to capture intra-orchard variability. As a result, growers face challenges in efficiently allocating resources and may harvest fruit that are either not completely developed or overripe.

In contrast, phenology models and hyperspectral imaging offer promising non-destructive

alternatives to maturity predictions. Phenology models can predict flowering time and harvest windows based on temperature, while imaging can achieve objective, real-time assessments of apple traits (firmness and sugar content) without the need for destructive sampling.

The purpose of this study was to identify a non-destructive method of maturity prediction in apples more than seven days before harvest. By leveraging the well-defined phenological cycle—where flowering time can be predicted from dormancy, and the harvest date can be predicted from flowering time, primarily through temperature-driven models—phenology models provide early-season forecasts. Imaging techniques, applied closer to harvest, offer more precise assessments of fruit maturity. These approaches have the potential to improve orchard planning, enhance post-storage fruit quality and reduce food loss.

1.3 Contributions and thesis structure

Two of the technical chapters (3 and 4) presented in this thesis have been peer-reviewed and published as standalone research articles in the *European Journal of Agronomy* and *Frontiers in Agronomy*, respectively, highlighting the novel approaches and results in this thesis. Chapter 5 has been prepared for submission into a peer-reviewed journal. Lastly, the hyperspectral images are being prepared for submission as a data paper, allowing others to train models on the collected dataset.

1.3.1 Chapter 2

The objective of this chapter was to examine the biological and environmental factors influencing apple maturity and the limitations of current assessment methods. This chapter reviewed key phenology models used to estimate flowering time and harvest windows, focusing on those commonly applied for apple flowering time and the harvest

window predictions. Following this, imaging was evaluated as a non-destructive alternative to conventional maturity assessments.

1.3.2 Chapter 3

In Chapter 3, the PhenoFlex model is applied to a large flowering-time dataset. The key gap in the current research is the limited understanding of how accurately flowering time can be predicted across different apple cultivars. To address this, the PhenoFlex model was fitted on twenty-six apple cultivars grown at a single site in East Malling. This dataset is the largest collection available from a single site, encompassing multiple cultivars over a period of up to 85 years. Its consistency across a single location removes environmental variability as a confounding factor in model training. The model was chosen as it is theoretically the most biologically accurate model currently available, enabling a flexible amount of overlap through parameter tuning for each cultivar. Additionally, a combined parameterisation approach, based on similarities in flowering times, and a generic parameterisation approach (using all cultivars in the study) were conducted to determine whether parameterisation was required at the cultivar, grouped-flowering-time or at the species (common apple) level. From this study, the most promising method was identified as a generic or a flowering-time-based grouped parameterisation approach, especially with small datasets.

1.3.3 Chapter 4

Building on the previous objective, this chapter addresses the understudied impact of flowering time variation on fruit maturity; most harvest date predictions focus on the average flowering date, disregarding the effect of the spread of flowering time. To explore the impact of flowering time, flower clusters were tagged as they opened (April) and fruit maturity was processed at harvest (September – October) for five apple cultivars. This gave the specific days between flowering and harvest for each individual fruit. Maturity

data (starch, Brix and firmness) was combined with the calculated number of growing units, using the Thermodynamic, Growing Degree Days (GDD) and Growing Degree Hours (GDH) models, by applying temperature data recorded within the range of those specific dates. To understand the factors contributing to variation in apple maturity, cultivar-specific influences including flowering time, seasonal conditions, individual tree variability and canopy position were examined. Flowering time was hypothesised to play a significant role in determining harvest maturity. Seasonal and tree-level effects were considered to account for environmental and biological variability, while canopy region was included to assess potential microclimatic influences, particularly related to light exposure. This analysis aimed to evaluate the limitations of phenology-based harvest predictions—particularly those relying on average flowering dates—and to determine whether a more targeted, image-based approach closer to harvest could offer improved accuracy for non-destructive maturity assessment. The results show that the majority of variation (up to 20%) is explained by flowering time variation. Seasonal and individual tree effects had a smaller impact on fruit maturity, while canopy region showed no significant impact.

1.3.4 Chapter 5

In Chapter 5, the third objective is addressed, focusing on directly assessing apples using imaging techniques as a non-destructive approach to predict apple maturity. Compared to spectroscopy and multispectral imaging, hyperspectral imaging acquires both spatial and spectral information. Although hyperspectral cameras are relatively expensive, the technology can determine the significant wavelengths of interest to predict apple maturity. This insight allows for the practical use of simpler, more affordable imaging systems tailored to specific wavelengths associated with maturity predictions. An extensive dataset of hyperspectral images was collected from multiple regions, seasons, and cultivars, yielding the largest collection of apple hyperspectral images. This dataset

is important for training deep learning models, as model performance generally benefits from more diverse datasets, which enhance generalisation and predictive accuracy. Building on the promising outcomes of simpler models, deep learning techniques were applied to establish more complex relationships between image data and fruit maturity traits. Four different model architectures were tested to determine the best-performing model architecture. To optimise computational efficiency, without losing spatial information, input images were downsampled. Edge cropping was applied to isolate the key region of the apples, ensuring that the model focused only on the face of the apples. Cultivar information was embedded into the input layer to provide genetic information during training. Bayesian optimisation was employed to efficiently tune hyperparameters and reduce computational complexity. Shapley analysis was used to quantify the relative importance of spectral wavelengths and spatial regions within the images. Additionally, the necessity of multiple views of the fruit are required for accurate prediction was evaluated to assess the feasibility of simplified imaging setups for practical orchard use. Lastly, models were independently trained on seasonal data to observe seasonal-specificity in the data. These insights point toward the potential for a combined phenology model and imaging approach to guide a practical, non-destructive framework for reliable apple maturity predictions, one that could deliver reliable forecasts earlier than conventional methods. The greatest contribution was the large and diverse imaging dataset collected for the study. The dataset was collected from 6 cultivars, 3 seasons, and 2 countries, totalling 3636 images of 5756 apples. This is approximately 600% larger than any other apple dataset used to train deep learning models (approximately 800 apples [14]) for fruit maturity traits. This dataset provides a strong foundation for training deep learning models due to its size, diversity and consistency. The results show that imaging is a promising alternative to conventional maturity assessment methods for Brix and firmness. The most effective deep learning model found was the Vision Transformer model, due to its ability to process both spatial and spectral information. Cultivar information was found to be

important during model training, as distinct spectral values were found across cultivars. Key wavelengths for predicting Brix and firmness were identified in the study using Shapley analysis. Spatial importance was found for firmness but not soluble solids. Lastly, it was determined that a single side of the fruit was insufficient in making accurate maturity predictions; the highest accuracy was found by using all four sides of the apple images.

1.3.5 Chapter 6

In the discussion chapter, the key results from the studies are presented and evaluated for their applicability in developing a non-destructive framework for predicting apple maturity. The contributions and limitations of the approaches are highlighted, and the thesis is concluded by reflecting on the broader implications of the research.

Literature Review

Apples are a staple ingredient in any household, being a versatile product with various culinary applications (flavourings, juices, ciders, cooking and dessert apples). Despite their low caloric content, they contain essential vitamins, minerals, and secondary plant metabolites that help protect against cardiovascular diseases, cancer, and type II diabetes [15–18]. Apple quality is determined by their physical (colour, shape, texture and size) and chemical (flavour, bioactive compounds) traits, attributes which determine their value on the market [10]. The quality of fruit at the end of a long period of storage is partly attributed to harvesting at the optimal maturity [19]. Fruit harvested too early are underdeveloped in flavour, colour and size and are more likely to develop scald and bitter pit storage disorders [1]. Overripe fruit tend to be softer, leading to bruises and rots, and once the autocatalytic ethylene production has started, it greatly reduces the storage life of all apples within its vicinity [1, 11–13]. By harvesting fruit at the optimal maturity, the overall quality can be improved and fruit loss from long-term storage can be minimised [9]. The optimal maturity is when the fruit exhibits attractive qualities but have not started ripening. However, although fairly accurate, the process of determining maturity is time-consuming, labour-intensive and destructive. Because forecasts can only predict within seven days of the harvest date, it becomes challenging for growers to efficiently allocate resources and manpower during the harvest window. Moreover, conventional maturity assessments are only capable of determining the optimum harvest date at the orchard level, this risks

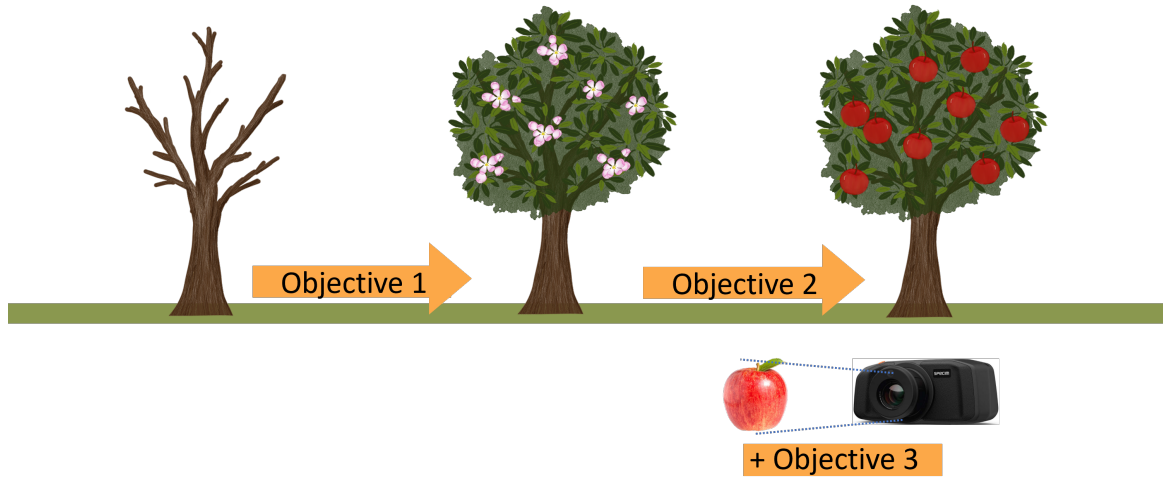


Figure 2.1: Objective points alongside phenological periods from winter to harvest.

including fruit that are either overripe or not fully developed in the final yield. In this thesis, the use of phenology models was explored, along with the evaluation of imaging as a non-destructive maturity assessment method (Figure 2.1). Alternative approaches to current conventional maturity assessment methods were tested: 1) can flowering time be accurately predicted with phenology models, 2) if varying flowering times influence the fruit maturity, and 3) can imaging predict fruit maturity traits?

2.1 Conventional methods for predicting apple maturity

In the following sections, the current processes used to determine apple maturity are discussed. In particular, the biological basis of the mathematical modelling used to determine flowering time and the harvest window and the process of destructive maturity assessments.

2.1.1 Temperature effects on dormancy

Studies since the nineteen hundreds have shown that temperature and day length play major roles in dormancy and dormancy break [20–22]. However, previous studies have shown that members of the Rosaceae family are insensitive to day length, as the growth of apple trees can be induced by 4 weeks of warm temperatures at 21°C in short-day conditions (10 hrs per day) [21]. Cold temperatures are required over a prolonged period to induce dormancy. Dormancy in perennial plants is a survival strategy during unsuitable growing conditions, where most physiological processes are stopped to protect sensitive tissues from environmental stress. Dormancy is characterised into three types: paradormancy, endodormancy, and ecodormancy. Paradormancy occurs during activate growth periods and is regulated by hormonal signals from competing buds that inhibit bud break [23]. In contrast endodormancy and ecodormancy are triggered by environmental changes. Observable signs of the start of endodormancy include leaf fall and growth cessation, usually at the end of autumn. Endodormancy is internally regulated and renders the plant insensitive to warm temperatures. Chilling temperatures alleviate dormancy-related insensitivity, with the required chilling being genetically determined by the specific cultivar [24]. Once chilling requirements are met, endodormancy transitions to ecodormancy, a phase in which growth remains suspended until external conditions become favourable. Dormancy release is followed by budbreak and flowering.

Dormancy is an important part of apple phenology. The consequences of shallow dormancy, when chilling requirements are insufficient, have critical implications for plant growth and development. Previous studies have reported delayed and prolonged flowering periods and greater variation in floral bud development (size, number of spurs, and peduncle length) [20, 25, 26]. The effects of shallow dormancy on floral bud development can increase fruit variation within trees, reducing the size, weight and total yield of fruit. In the worst-case scenario, apple trees in shallow dormancy

could respond to short bursts of warm weather in autumn, causing buds to wake in winter and therefore leading to the damage or loss of flowers and fruit production for the season [24, 27].

2.1.2 Approaches of flowering time models

Apple trees undergo a series of key developmental stages, including dormancy (endodormancy), dormancy release (the start of ecodormancy) and flowering. Various phenology models have been established to represent these phases. Chilling models — Chilling Hours [28, 29], Utah [30] and Dynamic models [31–33] — simulate endodormancy. Whereas, forcing models, such as the Growing Degree Hours and Growing Degree Day models [34, 35] estimate the influence of warm temperatures on ecodormancy until flowering. These key phases of transition are difficult to detect due to the minimal physical changes that accompany them. As such, the application of chilling and forcing models has largely been guided by biological assumptions. The Sequential Model assumes that chill requirements need to be reached before forcing units start accumulating, so these two phases occur one after the other [36]. The Parallel model assumes that chill accumulates at the same time [37]. However, this is not biologically accurate as plants are not responsive to heat during endodormancy; a chill requirement must be met before forcing units are accumulated [24, 38]. Hence, the development of overlapping models. The Alternating Model and Chill overlap models assume that there is an overlapping period when either chilling or forcing units can accumulate, depending on the temperature of the hour and the start date of forcing accumulation is not a fixed date [39, 40]. Notably, the sequential model will accumulate chilling units up to the start of the forcing model. In contrast, the Parallel and Chill Overlap Models assume that chill continues to accumulate until flowering is reached [38]. This distinction is important to the hypothesis that additional chilling exposure may have a beneficial effect by reducing the heat unit threshold to initiate flowering [41–43].

Given the lack of clarity in the biological mechanisms underlying phase transitions and the absence of observable physical changes, the overlap model is considered the most biologically plausible. However, since the chilling and forcing models are two distinct models, a key limitation lies in clarifying how their overlap should be addressed. This led to the development of the PhenoFlex model [44]. The PhenoFlex model has a slope parameter that flexibly integrates both the chilling and forcing models. Thus, depending on the parameter, the models can range from complete overlap, like the Parallel model, to complete independence, like the Sequential model. The PhenoFlex model makes modelling different cultivars more promising as the chilling and heat model parameters can be adjusted for each cultivar.

Parameterisation of model parameters should be conducted when training a phenology model. However, most studies fail to do so. Thus, applying models trained on other species or cultivars on their own datasets without calibration. This practice can lead to inaccurate predictions. When parameterisation is conducted, it is typically done on a single cultivar due to the limited datasets therefore, the best method of parameterisation is unknown.

2.1.3 Temperature effects on fruit development

The growth of the fruit starts from pollination, and it follows a curvilinear pattern, where cell division occurs in the early development, before subsequent cell expansion [45]. The temperatures within the first 50 days following pollination have been linked to the maximum potential fruit size [46] as warmer temperatures affect the rate of cell division [47, 48]. Temperature effects after 40 days after full bloom were found to have a lower impact on soluble sugars, firmness or starch hydrolysis than the initial period [48]. As the rate of fruit development is influenced by temperature, it is possible to predict the optimal harvest date through temperature-based modelling during the growth period.

2.1.4 Approaches of harvest date models

Many models have been presented to determine the optimal harvest date. Some use a set number of days after full bloom (DAFB), while others model the relationship between temperature and the harvest maturity to predict the optimal harvest date for apples [1, 49–51]. However, the key date, full bloom, is often inconsistent between studies. Luton and Hamer [50] consider full bloom when 50% of the flowers are open, 70% for Sugiura et al. [52], and 80% for Blanpied and Silsby [1]. Moreover, by predicting the harvest date with an averaged date, it will not account for the variation in flowering time. Apple trees are in bloom for an average of two weeks in April (in the United Kingdom) [53], therefore, the differences in temperature exposure between the flowers are unaccounted for by these models.

2.1.5 Flowering time effects on fruit maturity

Volz et al. [54] show that fruit from apple flowers that bloom later are smaller and less mature as shown by the greater presence of starch. This also occurs in other species such as tomato [55] and kiwifruit [56, 57]. Böhner and Bangerth [55] show that flowering time has a significant impact on tomato maturity. Normally, there is a 5-day difference in flowering (and pollination) between flowers closer to the stem and the distal flowers. If tomatoes are pollinated on the same date, they ripen at the same time, proving that the variation in flowering time affects tomato maturity. Studies on kiwifruit varieties are inconclusive. Kiwifruit cultivars, Hayward [56] and Zesy002 [57], show that flowers at the end of the cane bloom earlier than those closer to the roots. When picked at a single time point, fruit from the ends were more mature than fruit closer to the root end, suggesting flowering time is positively correlated to fruit maturity. However, the kiwifruit cultivar Hort16A [56] flowers bloomed earlier at the root end, but despite this, the fruit on the tip ends were riper (less firm) despite flowers blooming later. Studies will need to be conducted in apple to determine the

influence of flowering time on fruit maturity variation.

2.1.6 Maturity assessments

Maturity assessments are the current gold standard for maturity assessments. They are the most common way of determining the harvest maturity close to the actual harvest date. A sample of fruit is collected from an orchard and processed to obtain the average starch, firmness and Brix levels. When apples mature, the cell walls start to break down, resulting in softer, more palatable fruit. They also start to convert starch into soluble solids, which decreases the amount of surface starch-iodine staining of apple halves (Figure 2.2) and increases Brix levels. Apples at a starch index of 3 or 4 are the most appropriate for long-term storage [1, 58]. Another common trait is the loss of chlorophyll, as observed by the degreening of the skin. In addition, depending on the cultivar, anthocyanins may be produced, resulting in red pigmentation in the skin.

There are several limitations to these conventional maturity assessments. The most prominent being their reliance on destructive sampling. Each evaluation results in the loss of fruit and involves time-consuming procedures. This highlights a need for a more efficient, non-destructive alternative. Further concerns include the subjective nature of starch iodine staining. Consistency is best achieved when the assessment is performed by a single individual. However, even then, results may vary from day to day. The inherent subjectivity of the process may produce unreliable results. Degreening is also a good way to determine fruit maturity; however, its effectiveness is limited in red-skinned varieties. Red pigments (anthocyanins) may mask the presence of chlorophyll, making it difficult to accurately assess with the human eye. Additionally, anthocyanins are produced in response to light exposure, creating uneven colouring on opposite sides (light exposed vs non-exposed) of the fruit, further complicating visual evaluations.

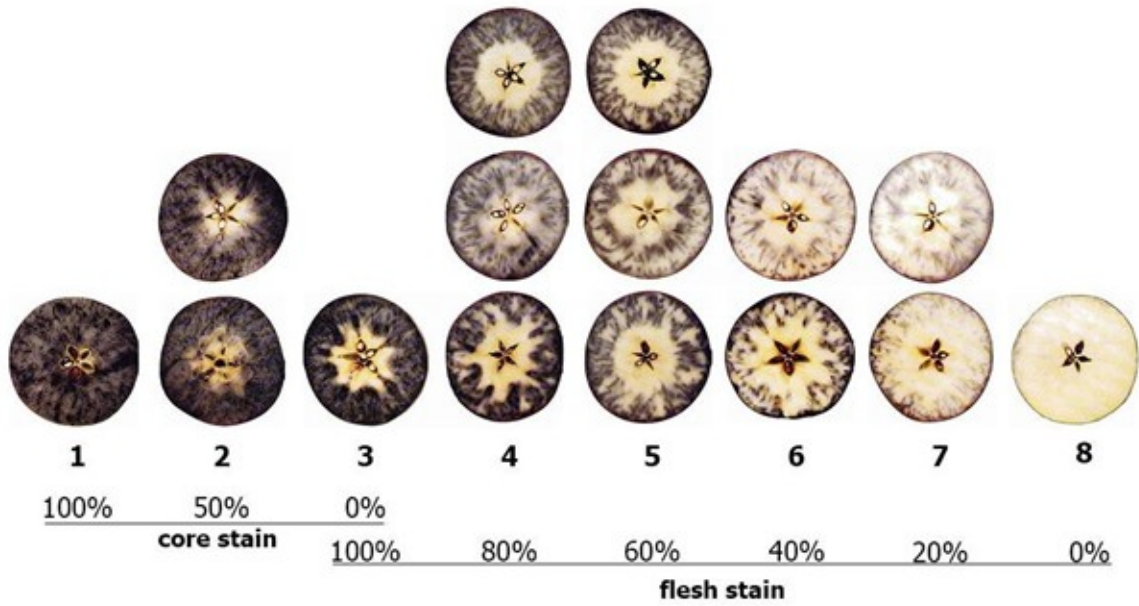


Figure 2.2: Starch iodine indexes in apple [1]. An iodine mixture is used to spray halved apples to determine fruit ripeness from 1 - least ripe to 8 fully ripe. In common practice, the optimal harvest dates for long-term storage are around stages 3 and 4.

2.2 Imaging

Imaging has emerged as a non-destructive method to predict fruit maturity and quality. Other than being a non-destructive testing method, imaging can overcome the limitations of conventional testing methods by obtaining scans faster, being objective in their measurements, being able to detect compounds unseen by the human eye and being able to test the maturity of the fruit of interest rather than using a representative sample. Images are defined as a representation of a surface, with each pixel containing spatial and qualitative information [59]. This makes them powerful tools for analysing complex biological constituents including: dry matter, sugar, water, acids, and minerals [60]. Apple fruit is comprised of 75 - 88% water, which may interfere with soluble sugar detection as they both reflect in the infrared spectra [60–62]. Despite this, imaging techniques are capable of measuring fruit maturity traits, in particular soluble solids and firmness. Apple studies show that

imaging results using detached apples put in a dark chamber with a detector and a light source can be used to class maturity as unripe, ripe, or over-ripe with more than 84-95.5% accuracy [63–67]. These studies validated their classifications using conventional methods.

Imaging data is highly collinear, especially when wavelengths are closely spaced. Multi-variate statistical methods have been used, such as partial least squares regression, multiple linear regression, and a dimension reduction method, such as principal components analysis to address the relationship and remove highly collinear wavelengths in imaging data [65, 68–70]. More recently, deep learning models have been applied to imaging studies to decipher more complex patterns [71, 72]. However, a major challenge with deep learning models lies in their black box nature, which limits the interpretability of model outcomes.

Xu et al. [67] tested the effect of using imaging to predict maturity based on the location the image is taken on the fruit. Capturing images at different places may affect the clarity of the spectra especially when an apple’s surface is curved and this curvature is stronger at the calyx and stalk ends. The spectra intensity is significantly lower when taken from these ends. They show that the best orientation to capture an image is at the equator of the fruit as these best represent the average of all their seven tested positions. Moreover, they captured images from one random position on the fruit to validate their test which further confirmed their results. Even though the position of the fruit doesn’t alter the spectra results, at least two sides of the apple should be tested as the side exposed to light may be firmer than the other sides of the fruit [73–75].

2.3 Summary

This review explored the biological and technological approaches to predicting apple maturity, with a focus on improving harvest timing and fruit quality. Phenology models, which simulate developmental stages, such as dormancy and flowering, are not properly used in practice, often failing to conduct the parameterisation step. The best method of parameterisation has yet to be found, mainly due to the lack of expansive datasets. Additionally, phenology models can offer valuable insights, but their start dates are inconsistent between studies and they are unable to account for within-tree variations in flowering time. Moreover, conventional maturity assessments are destructive and may be subjective. Imaging may be a promising alternative method of fruit maturity assessment due to the non-destructive nature of the process. However, hyperspectral images are highly collinear, requiring more complex modelling techniques such as deep learning models. While deep learning models can capture complex patterns and filter out multi-collinear wavelengths, their black box nature makes them difficult to interpret. External factors such as light quality and fruit orientation can also affect model results. Future studies should focus on identifying the best practice for parameterisation, determining the effect of flowering time variation and evaluating the effectiveness of hyperspectral imaging for apple maturity predictions.

Evaluating the performance of models predicting the flowering times of twenty-six apple cultivars in England

3.1 Introduction

Apple tree dormancy is regulated by temperature [21, 76]. Low temperatures slow plant processes so that young meristems such as leaf and flower buds are protected against harsh environments. There are two types of dormancy that affect the flowering period: endodormancy and ecodormancy. Endodormancy is the state in which trees stop growing in response to chilly conditions. Following this, ecodormancy sets in, signifying that chilling requirements have been met but the plants are awaiting warm temperatures to resume growth. Specifically, the chilling requirement refers to the duration and intensity of chilling temperatures needed to complete endodormancy and likewise, the heat requirement refers to the duration and intensity of warm temperatures needed to overcome ecodormancy. Both the chilling and heat requirements are species-specific [77].

Apples (*Malus domestica* (Suckow) Borkh.) have been cultivated for several thousand years, dating back to 1000 BCE [78] and while only about 100 cultivars are grown commercially, there are over 7500 unique cultivars [79]. Predicting the timing of apple

flowering time is important for variety selection and crop management. Phenology models aim to explain the realities of biology with mathematical representations. Many phenology models have been developed to predict flowering time but there is no consensus to the best phenology model for apple [37, 80–82]. A complete flowering time model comprises of two sub-models: the first models chilling accumulation and the second models heat accumulation. Chilling models estimate completion of endodormancy whereas forcing models represent the hypothetical relationship between warm temperatures and plant development.

Chilling models include the Chilling Hours, Utah and Dynamic models. The Chilling Hours model is a long-established chilling model and takes the cumulative number of hours below a certain temperature threshold [29]. However, it does not consider the negative effects of chill accumulation by high temperatures unlike the Utah model and the Dynamic model. The Utah model calculates chilling accumulation by giving each hour a positive, negative, or no chilling value [30]. As the Utah model results in negative chilling units in tropical and subtropical climates, this model is only appropriate for temperate climates [83]. The Dynamic model is the only known empirical chilling model that utilises bud break experiments to formulate the model, which is why this thesis will focus on this chilling model [24, 41, 84, 85]. The Dynamic model calculates chilling accumulation in two steps. The first step accumulates a pseudo product called precursor to the dormancy-breaking factor (PDBF) which is created in chilling temperatures but destroyed in warm temperatures. Both the formation and destruction follow Arrhenius law, describing the effect of temperature on the rate of chemical reactions in an exponential relationship. The PDBF fluctuating in response to temperature allows the model to adapt to warmer regions, since longer periods of warm temperatures will not result in negative chilling. Additional chilling activates the second step, the irreversible conversion of the PDBF to a chill portion after a critical amount of chilling is reached. These chill portions are stable and cannot be

destroyed by subsequent heat exposure.

The most widely used forcing model is the Growing Degree Hour (GDH) model [34]. The GDH model is the hourly variation of the growing degree days model, which takes the average temperature of the day rather than the hour [34, 86]. A GDH model sums the temperatures suitable for growth (temperatures above a threshold temperature) and each degree is weighted based on its proximity to the optimal growing temperature. The date when the GDH units reach the heat requirement is predicted as the flowering date.

The boundary between endodormancy and ecodormancy is unclear due to a lack of measurable physiological changes. Thus, recent phenology models were developed to integrate chilling and forcing models into a single framework for predictive purposes. To ensure a flexible modelling approach, several forms of transitions from chilling to forcing have been studied, including (1) treating chilling and heat accumulations completely sequentially and separately (chilling must be completed before heat accumulation starts), (2) assuming a complete overlap (chilling and heat accumulation is simultaneous), and (3) assuming a partial overlap between the chilling and heat accumulation phases (heat accumulation can start before chilling is completed but not before the initiation of chilling).

The Parallel model assumes that chilling and forcing units accumulate at the same time [37]. However, it is understood that a chill requirement must be met before forcing units can accrue because any response to heat before chilling satisfaction may lead to flowering in unsuitable conditions [24–26, 38]. In contrast to parallel accumulations of chilling and heat, the Sequential and Unified models both assume independent phases for chilling and heat accumulation [36, 81]. The difference between the Sequential and Unified models is that the Sequential model uses user-chosen chilling and forcing models, whereas the Unified model chooses the chilling and forcing models based on the data. The shape of the adjustable function within the Unified model can follow

the shape for Triangular Chilling, Chilling Days, Sigmoidal and Growing Degree Days models and anything in between for chilling and forcing. These models are commonly used in older combination models. The parameters determine the shape of the model, so it does not prematurely constrain the data to fit a particular model [81]. The StepChill model is a simpler variation of the Unified model. It showed similar predictive performance in a recent comparative study in predicting bud break in major forest tree species compared to the full Unified model despite the use of fewer parameters [44, 87].

Later models were developed with the concept of an overlap existing between the chilling and heat phases and, moreover, they consider a beneficial effect on heat accumulation through additional chilling after chilling requirements have been reached [41–43]. The Chill Overlap model attempts to model an overlap between the chilling and heat models. This model is based on the idea that once endodormancy is satisfied, additional exposure to chilling temperatures will reduce the heat requirement [38, 39, 82]. The PhenoFlex model was developed in 2021 and was used to predict apple and pear flowering time [44]. This model integrates the Dynamic and GDH models as the chilling and forcing models. The difference is that while the Chill Overlap model requires a preset overlap value, the PhenoFlex model allows for no overlap, various degrees of overlap or complete overlap between the chilling and forcing models.

Previously, apple flowering time of different cultivars has been predicted with a range of different approaches. The recent studies have used parameterisation to minimise the root mean square error (RMSE) to get the highest level of predictive accuracy. A typical sequential model combination is the Dynamic and GDH model combination. This combination resulted in fairly good predictions in validation datasets for Boskoop (4.2 days), Cox’s Orange (5.7 days), Golden Delicious (5.12 days) and Jonagold (4.57 days) apples in Belgium [88] but poor predictions for Crisp’s Pink apples in Australia (RMSE of 14.7 days) [82]. A chill overlap model was developed for Crisp’s Pink apples

grown in Australia with a RMSE value of 5.9 days [82]. A StepChill model resulted in a RMSE of 7.68 days for Boskoop apple trees in Germany, but a PhenoFlex model improved these results, reducing the RMSE to 3.82 days [44]. These few studies on a limited number of apple cultivars indicated that the overlapping modelling framework for predicting apple flowering time may be better than a completely sequential model and that apple cultivars may differ significantly in the exact model formulation as well as in parameter values within the same model formulation.

A study in apricots [89] shows that while applying flowering data from many cultivars to a crop does not result in accurate predictions, consolidating cultivars based on their flowering group can improve model accuracy. Apple cultivars are classified into flowering groups (based on similar flowering times), but it would be interesting to see if a naïve clustering method can be used to classify cultivars into groups that are used to generate parameters suitable for the PhenoFlex model.

Previous research has focused on parameterisation for individual cultivars separately, thus in the present study comparative modelling of multiple cultivars based on the PhenoFlex modelling framework is explored. Specifically, the present study focuses on estimation of PhenoFlex model parameters for (1) individual cultivars separately (cultivar-specific), (2) cultivar groups derived from K-means clustering of mean flowering dates and their variation across years (mean flowering time), and (3) a common model fitted to all cultivars (common model). Three approaches were evaluated for their accuracy in predicting flowering dates. The cultivar-specific approach follows conventional protocols when fitting a phenology model. The common model approach is to explore whether a large dataset from combining all different apple cultivars improves model performance. The K-means approach is an intermediate between the cultivar-specific and common model approaches and is used to determine if cultivars can be grouped together using a naïve clustering method based on pattern similarities in their mean flowering dates. The hypothesis is that the cultivar-specific approach

will prove to be the most accurate in predicting flowering dates, followed by K-means grouped and the common approach. Finally, the best of the three approaches is used to compare it to the StepChill model to assess if the PhenoFlex model can achieve comparative results.

3.2 Materials and Methods

3.2.1 Flowering data

The flowering time of twenty-six apple cultivars have been recorded from East Malling, United Kingdom, over the last eighty-five years. The flowering data collected for each cultivar ranged between eighteen to eighty-five years. Some records are shorter than others as some cultivars may have been recorded earlier than others, discontinued or are newer cultivars. Due to the apples being monitored from one site, the environmental variation can be limited within the dataset and directly determine the variation between apple cultivars. The specific years are indicated in Table 3.1 and depicted in Figure 3.1. For each cultivar, their date of the first flower (BBCH 60, according to the BBCH-scale for fruit phenology [90]), will be used for modelling purposes as the first flowers are less affected by environmental conditions other than temperature [38, 82]. The flowering dates are either an average of four trees of the same cultivar or an individual tree, depending on the data availability. The number of trees are recorded in Table 3.1. The cultivars are grown on sixteen different rootstocks, which have not been used as a factor in the analysis as flowering behaviour is assumed to be determined by the scion.

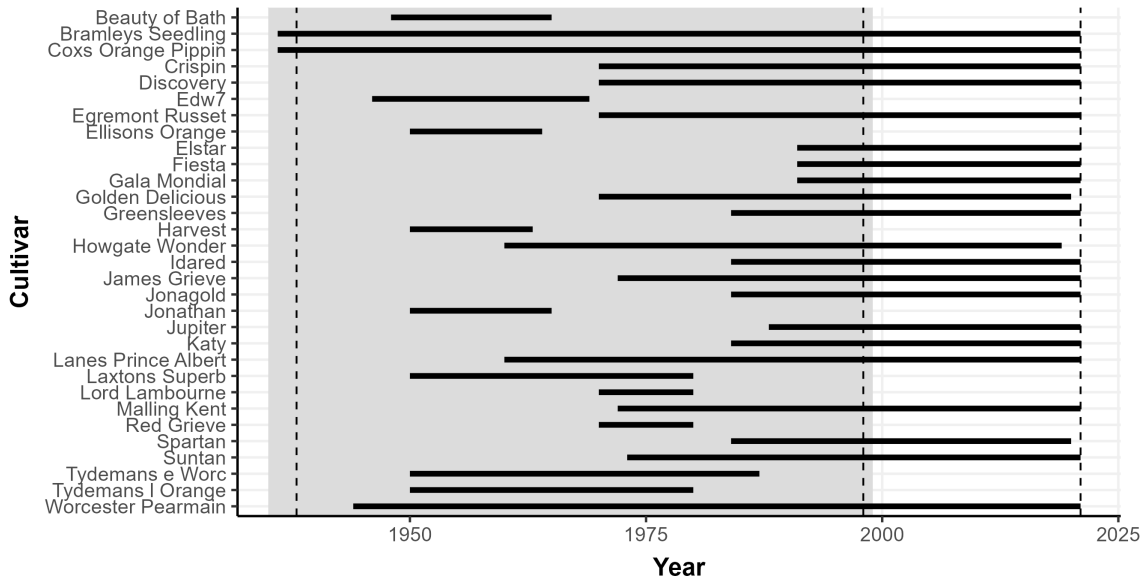
3.2.2 Temperature data

Temperature records were obtained from the East Malling weather station (51.2876°N, 0.4486°E, 33m above mean sea level), an official UK Meteorological Office Station.

Table 3.1: Summary of the range of flowering time data available from East Malling, the total number of years and number of datapoints used to train and test each model for each of the twenty-six cultivars. The table is split by K-means clustering on mean flowering dates and variation across years. The last column represents the standard deviation of flowering.

Cultivar	Starting year	Ending year	Total years	Tree (n)	Training years	Testing years	SD of flowering dates
Group 1							
Beauty of Bath	1948	1965	18	5	13	5	9.89
Crispin	1970	2021	51	5	36	15	9.75
Egremont Russet	1970	2021	52	5	36	16	10.53
Greensleeves	1984	2021	38	3	27	11	9.89
Idared	1984	2021	34	4	24	10	10.27
James Grieve	1972	2021	49	5	34	15	9.35
Jonagold	1984	2021	38	5	27	11	9.12
Group 2							
Edw7	1946	1969	24	13	17	7	7.27
Howgate Wonder	1960	2019	54	3	38	16	7.74
Lanes Prince Albert	1960	2021	62	3	43	19	8.30
Laxton's Superb	1950	1980	25	3	18	7	7.35
Tydemans' Early Worcester	1950	1987	31	5	22	9	7.55
Tydemans' Late Orange	1950	1980	26	5	18	8	6.85
Worcester Pearmain	1944	2021	70	11	49	21	8.16
Group 3							
Bramley's Seedling	1936	2021	81	9	57	24	8.90
Cox's Orange Pippin	1936	2021	85	17	59	26	9.19
Discovery	1970	2021	50	5	35	15	8.84
Elstar	1991	2021	29	1	20	9	7.88
Fiesta	1991	2021	31	2	22	9	8.16
Gala Mondial	1991	2021	30	2	21	9	7.64
Golden Delicious	1970	2020	51	4	36	15	9.17
Jupiter	1988	2021	34	1	24	10	7.95
Katy	1984	2021	37	2	26	11	8.67
Malling Kent	1972	2021	50	5	35	15	8.98
Spartan	1984	2020	36	2	25	11	8.82
Suntan	1973	2021	49	6	34	15	9.52

Figure 3.1: Year range for each of the twenty-six cultivars, shown as black bars. The data highlighted in grey indicate simulated temperatures using latitude and maximum and minimum daily temperatures. The unhighlighted area represents temperature data from real recorded hourly temperatures, supplemented with simulated data where there are missing hourly values. The dashed lines indicate the specific years in which the three days and five hours of missing data occurs.



The orchards were located within 0.75 miles east and 0.31 miles north, 0.21 miles west, 0.18 miles south of the weather station. Fluctuations in hourly temperatures in a day follow typical patterns between maximum and minimum daily temperatures. These patterns can be modelled by a sine function for daytime warming and logarithmic decay for nighttime cooling, respective to a specific geographical latitude. A simulated approach following these patterns was used to generate missing hourly temperature values [44, 91]. Data from 1935 to 1999 are recorded as daily maximum and minimum temperatures so the simulated approach was used to generate hourly temperatures for all hourly observations between 1935 and 1999. These generated hourly temperatures were also used to fill in missing hourly datapoints from 2000 to 2021. Overall, 72.15% of the data was formed by the simulated approach, and 27.84% was of real data values. The remaining 0.011% is due to the inability to simulate hourly temperatures due to

missing daily minimum and maximum temperatures. In total this accounts to three days and five hours of data which is unlikely to significantly affect model predictions. The hourly data from 2000 to 2021 was mostly complete. It consists of 99.8% of real temperature values, 0.24% in simulated data and a negligible amount in missing temperatures (28 hours).

3.2.3 Model formulation

PhenoFlex model

The PhenoFlex model, implemented in the `PhenoFlex_GDHwrapper()` function from the `chillR` package, integrates the framework from the Dynamic model and the GDH model [91]. The PhenoFlex model [44] is fitted with twelve parameters, with the parameters for the chilling requirement (y_c), the heat requirement (z_c) and slope (s_1) linking the Dynamic and GDH models.

The heat accumulated at any point in time (t) is calculated by the PhenoFlex model equation, incorporating the total heat accumulated so far (z) and a portion of the GDH function over the elapsed time and temperature (T) [44]. $Py(y)$ is a function following a sigmoidal pattern which determines the proportion or size (s) of heat that can be accumulated, as a function of the accumulated chill (y) [44]. The inflection point is determined by the critical chilling threshold (y_c) and the slope of the transition is determined by the parameter s_1 . Large values of s_1 indicates lower levels of overlap and vice versa.

Six of the twelve PhenoFlex parameters are associated with the Dynamic model. The hypothetical process to form and destroy the precursor to the dormancy-breaking factor (PDBF) follows Arrhenius law. $E0$ and $E1$ represent the time-independent activation energy, and $A0$ and $A1$ refer to the amplitude of the function. $E0$ and $A0$ contribute to PDBF formation, while $E1$ and $A1$ are involved in PDBF destruction. When x reaches 1, a portion of the PDBF is converted to a stable chilling portion

where it cannot be destroyed by warm temperatures [41]. The pseudo-intermediate (x) is calculated as a function over time, where t is the new time and t_j is the level of x at time j . The portion converted is determined by a sigmoidal function with the inflection point at T_f and slope governed by the slope parameter [41].

Three parameters are associated with the GDH model. The contribution to heat accumulation is dictated by the optimal temperature (T_u), the upper temperature limit (T_c) and the lower temperature limit (T_b) [34]. The difference between optimal and lower temperatures are multiplied with a function which determines the effectiveness of GDH in driving the biological process under consideration [34].

StepChill model

The five-parameter StepChill model was fitted with the `StepChill_Wrapper()` function from the `chillR` package [91]. The first parameter is the chilling threshold (T_c). Any temperature lower than T_c does not contribute to chilling function (CF); temperatures above T_c contributes 1 to the CF (Equation 3.1).

$$CF(T) = \begin{cases} 0, & T \leq T_c \\ 1, & T > T_c \end{cases} \quad (3.1)$$

Chilling hours accumulate until it reaches the chilling requirement (C^*), the second parameter of this model. The heat (forcing) model then begins and is represented by Equation 3.2. This function determines the amount of heat contributed by the temperature (x). In the simplified version of the Unified model, the StepChill function sets parameter a to 0. The third and fourth parameters, b and c , affect whether the sigmoidal curve starts positive and shifts to negative or vice versa and at what temperature this shift occurs. The fifth coefficient F^* represents the heat requirement and budbreak occurs when the heat accumulated reaches F^* [81, 91].

$$Heat(T) = \frac{1}{1 + e^{a(x-c)^2 + b(x-c)}} \quad (3.2)$$

3.2.4 Model optimisation and performance evaluation

The flowering data was split into a training dataset and a test dataset by randomly selecting 70% of the years for each cultivar and leaving the last 30% of the unselected data for the test dataset. This split was done for each cultivar then the split was maintained for subsequent model fitting and comparisons between approaches.

Specific models were fitted to the corresponding training data with a simulated annealing algorithm, wrapped in the `phenologyFitter()` function from the `chillR` package [91]. The simulated annealing mechanism generates model parameters for the model chosen, then aims to reduce the residual sum of squares (RSS) by choosing a new set of model parameters. This process is repeated up to 1000 times or until there are no improvements after 250 iterations. The best fit model was then bootstrapped 99 times with the function `bootstrap.phenologyFit()` in the `chillR` package [80, 91]. The standard errors for the parameters were calculated on the 99 bootstrap values and the original set of fitted parameter values. This process was repeated at least seven times using different starting parameters as the `PhenoFlex` model results can be sensitive to the initial parameters. The reported results were from the run with the smallest residual sum of squares (RSS).

Fitted models were evaluated with Akaike information criterion (AIC) [92], model efficiency (EFF) [93] and RMSE for both the training and test datasets. AIC is a measure of model goodness of fit that considers the number of parameters in the fitted model. The function `AICc` contains a penalty term adjusting for small sample sizes (Equation 3.3).

$$AICc = 2k + n \log\left(\frac{RSS}{n}\right) + \frac{2k^2 + 2k}{n - k - 1} \quad (3.3)$$

The number of parameters is represented by the letter k , the number of samples, n , and the residual sum of squares, RSS . $AICc$ will be used to assess the fitted models and their parameters and determine the model that minimises information loss. $AICc$ values are relative to each other, the smaller the $AICc$ value, the better.

The model efficiency (EFF) compares models that were fitted to the same training dataset. The efficiency is the ratio between the residual sum of squares and the squared sum of the differences between the observed values and the mean (Equation 3.4).

$$EFF = \frac{RSS}{\sum (t_i - \bar{t})^2} \quad (3.4)$$

Root mean squared error ($RMSE$) is a commonly used metric of prediction accuracy. AIC and EFF can only be used to compare between models fitted with the same dataset, while $RMSE$ can be used to compare between models. This is why $RMSE$ is used to evaluate the test dataset.

Due to differences in the amount of flowering data for each cultivar, the Ratio of Performance to InterQuartile distance ($RPIQ$) was used to standardise the prediction errors against the variation of the observed flowering dates.

3.2.5 Comparative modelling

Comparing PhenoFlex models between apple cultivars

Firstly, PhenoFlex models were fitted to individual cultivars, and thus the model was fitted to twenty-six test datasets (one for each cultivar). Next, K-means clustering was applied on mean flowering dates and their variation across years to determine flowering groups. K-means clustering is an unsupervised method which assigns each

observation into a group based on their similarities with other observations in the same group. Principal Component Analysis (PCA) of standardised Z scores were used for interpretability and visualisation. PhenoFlex models were then fitted to each flowering group of cultivars as identified by the K-means clustering. Finally, a single PhenoFlex model was fitted to all twenty-six cultivars. Model performance, particularly for the test datasets, was then evaluated and compared among the three sets of models.

Comparing PhenoFlex and StepChill models using common parameters

In addition to the common PhenoFlex model, a common StepChill model was fitted to the pooled training data of all twenty-six cultivars. The common PhenoFlex and StepChill models were then assessed for the performance.

R version

The analysis was run on R version 4.3.2 (2023-10-31 ucrt). Model fitting and bootstrapping was run on a high throughput computer running R version 4.2.3 (2023-03-15).

3.3 Results

3.3.1 PhenoFlex models fitted to individual cultivars

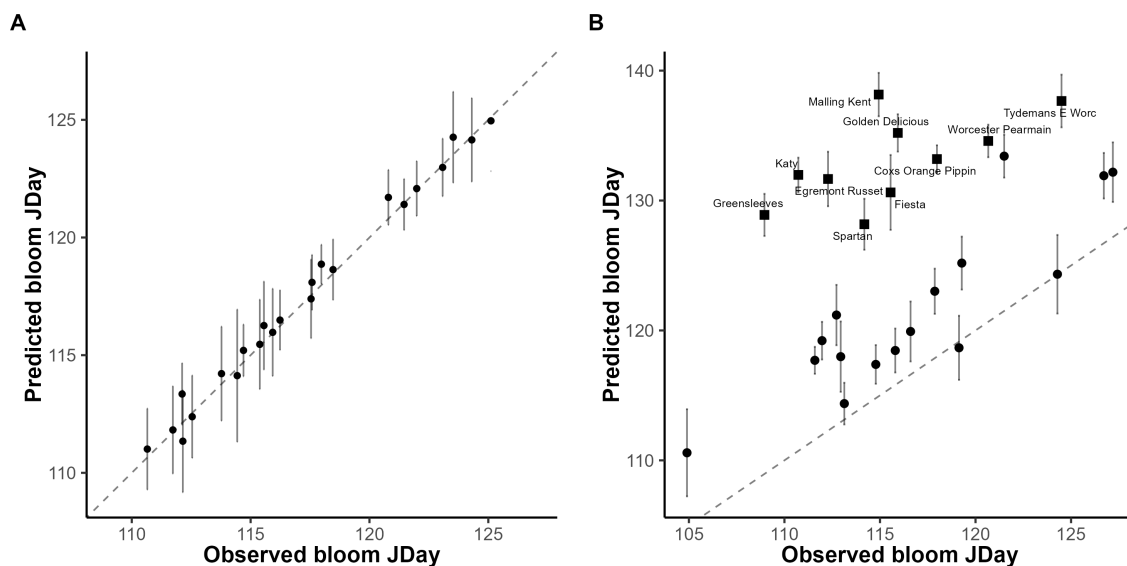
Cultivar-specific parameter estimates for the PhenoFlex model (Table 3.2) were used to predict flowering dates for individual cultivars. The parameters were derived from running the model 10 times with different starting parameters. The parameters were selected from the runs with the lowest RSS for each cultivar. The average RSS was 1194.15 ± 103.69 for the specific model.

Table 3.2: The best fit parameters of the PhenoFlex model for twenty-six apple cultivars, each one optimised by generalised simulated annealing algorithm with 99 bootstraps.

Cultivar	y_c	z_c	s_1	T_u	E_0	E_1	A_0	A_1	T_f	T_c	T_b	slope
Beauty Of Bath	28.66 ± 0.17	300.45 ± 3.15	22.97 ± 0.72	16.65 ± 0.12	3970.6 ± 2.12	$10\,725.61 \pm 5.99$	8721.52 ± 41.25	$5.95 \times 10^{13} \pm 2.13 \times 10^7$	2.03 ± 0.15	30 ± 0.34	3.7 ± 0.04	30.77 ± 0.81
Bramley's Seedling	27.87 ± 0.12	300.59 ± 1.88	0.83 ± 0.71	17.33 ± 0.08	3812.45 ± 5.44	$10\,196.84 \pm 16.56$	$10\,813.98 \pm 68.98$	$5.85 \times 10^{13} \pm 2.23 \times 10^7$	-3.98 ± 0.17	29.88 ± 0.27	2.76 ± 0.02	40.08 ± 0.64
Cox's Orange Pippin	29.86 ± 0.24	280.09 ± 1.47	22.23 ± 0.4	22.06 ± 0.13	3983.1 ± 3.91	$10\,686.31 \pm 11.55$	9782.07 ± 38.61	$5.89 \times 10^{13} \pm 1.67 \times 10^7$	-0.03 ± 0.12	29.44 ± 0.22	2.96 ± 0.02	49.39 ± 0.45
Crispin	20.71 ± 0.2	284.16 ± 1.25	0.66 ± 0.49	15.71 ± 0.08	3891.13 ± 1.84	$10\,298.48 \pm 9.66$	$10\,417.43 \pm 44.17$	$5.93 \times 10^{13} \pm 1.25 \times 10^7$	1.3 ± 0.11	29.69 ± 0.18	3.26 ± 0.02	35.41 ± 0.44
Discovery	25.03 ± 0.13	334.17 ± 1.2	22.82 ± 0.39	15 ± 0.11	3898.5 ± 1.82	$10\,451.35 \pm 4.66$	8542.24 ± 36.85	$5.91 \times 10^{13} \pm 1.52 \times 10^7$	2.51 ± 0.09	25.46 ± 0.18	2.8 ± 0.01	35.46 ± 0.4
Edw 7	23.58 ± 0.11	293.47 ± 1.11	1.4 ± 0.38	16.38 ± 0.1	3828.77 ± 1.84	$10\,244.67 \pm 4.69$	9566.66 ± 30.87	$5.91 \times 10^{13} \pm 1.32 \times 10^7$	4.9 ± 0.08	20.23 ± 0.17	3.38 ± 0.01	50.28 ± 0.37
Egremont Russet	21.3 ± 0.11	370.38 ± 1.07	22.57 ± 0.34	18.62 ± 0.09	3842.75 ± 1.7	$10\,272.34 \pm 4.33$	$10\,503.56 \pm 29.03$	$5.91 \times 10^{13} \pm 1.24 \times 10^7$	3.42 ± 0.08	21.77 ± 0.15	3.21 ± 0.01	49.18 ± 0.34
Elstar	25.32 ± 0.1	310.54 ± 1.11	22.85 ± 0.26	24.57 ± 0.09	3932.29 ± 2.14	$10\,761.05 \pm 7.67$	8403 ± 32.36	$5.94 \times 10^{13} \pm 1.12 \times 10^7$	4.82 ± 0.07	29.9 ± 0.14	3.2 ± 0.01	39.71 ± 0.32
Fiesta	26.26 ± 0.09	287.96 ± 1.02	17.36 ± 0.24	24.78 ± 0.1	3861.87 ± 2.01	$10\,455.76 \pm 6.85$	8887.9 ± 30.3	$5.94 \times 10^{13} \pm 1.04 \times 10^7$	-1 ± 0.07	29.99 ± 0.13	3.18 ± 0.01	55.87 ± 0.31
Gala Mondial	29.18 ± 0.11	325.41 ± 0.85	22.98 ± 0.23	22.6 ± 0.08	3826.05 ± 1.92	$10\,302.38 \pm 5.75$	$10\,376.74 \pm 27.24$	$5.89 \times 10^{13} \pm 9.99 \times 10^6$	0.57 ± 0.07	30 ± 0.12	2.75 ± 0.01	30.13 ± 0.28
Golden Delicious	20.26 ± 0.09	340.27 ± 0.89	22.56 ± 0.21	21.21 ± 0.08	3999.44 ± 1.84	$10\,585.62 \pm 5.97$	9451.45 ± 25.85	$5.94 \times 10^{13} \pm 1.05 \times 10^7$	2.21 ± 0.06	30 ± 0.11	3.13 ± 0.01	32.63 ± 0.28
Greensleeves	29.14 ± 0.1	292.51 ± 0.8	23 ± 0.23	20.56 ± 0.07	3990.64 ± 1.45	$10\,672.78 \pm 4.74$	$10\,407.76 \pm 26.36$	$5.93 \times 10^{13} \pm 9.21 \times 10^6$	1.85 ± 0.06	29.97 ± 0.11	3.43 ± 0.01	30.42 ± 0.26
Howgate Wonder	25.95 ± 0.08	325.05 ± 0.79	4.04 ± 0.24	20.17 ± 0.07	3936.24 ± 1.34	$10\,489.47 \pm 3.69$	9723.83 ± 19.69	$5.91 \times 10^{13} \pm 9.71 \times 10^6$	-0.92 ± 0.06	24.06 ± 0.11	3.5 ± 0.01	32.23 ± 0.25
Idared	22.09 ± 0.08	368.47 ± 0.82	0.47 ± 0.19	19.97 ± 0.07	3868.48 ± 1.62	$10\,317.06 \pm 5.4$	9985.63 ± 22.88	$5.94 \times 10^{13} \pm 9.29 \times 10^6$	0.82 ± 0.05	28.52 ± 0.1	2.71 ± 0.01	31.7 ± 0.24
James Grieve	23.07 ± 0.07	332.38 ± 0.79	22.98 ± 0.19	21.7 ± 0.07	3934.8 ± 1.51	$10\,525.06 \pm 5.04$	8779.79 ± 21.59	$5.94 \times 10^{13} \pm 9.17 \times 10^6$	0.07 ± 0.05	23.1 ± 0.1	2.83 ± 0.01	56.84 ± 0.24
Jonagold	20.36 ± 0.08	330.6 ± 0.69	22.04 ± 0.21	20.92 ± 0.06	3808.8 ± 1.39	$10\,244.36 \pm 4.06$	9746.5 ± 21.69	$5.93 \times 10^{13} \pm 7.78 \times 10^6$	1.83 ± 0.05	29.99 ± 0.1	2.96 ± 0.01	37.5 ± 0.22
Jupiter	28.26 ± 0.08	281.63 ± 0.68	21.98 ± 0.2	21.39 ± 0.06	3880.92 ± 1.33	$10\,414.83 \pm 3.83$	9959.96 ± 21.12	$5.93 \times 10^{13} \pm 7.97 \times 10^6$	0 ± 0.05	22.61 ± 0.1	3.88 ± 0.01	39.43 ± 0.21
Katy	27.35 ± 0.06	303.23 ± 0.72	0.39 ± 0.2	19.98 ± 0.06	3816.23 ± 1.4	$10\,342.18 \pm 4.45$	9022.32 ± 21.24	$5.92 \times 10^{13} \pm 8.33 \times 10^6$	-3.34 ± 0.05	29.83 ± 0.09	3.2 ± 0.01	33.97 ± 0.21
Lanes Prince Albert	28.37 ± 0.07	304.04 ± 0.7	22.59 ± 0.17	23.24 ± 0.06	3869.55 ± 1.31	$10\,539.21 \pm 4.19$	8751.57 ± 19.55	$5.94 \times 10^{13} \pm 7.83 \times 10^6$	1.37 ± 0.05	24.87 ± 0.09	2.72 ± 0.01	43.8 ± 0.21
Laxton's Superb	29.43 ± 0.06	396.47 ± 0.66	22.9 ± 0.19	21.07 ± 0.05	3945.93 ± 1.09	$10\,808.79 \pm 3.57$	8800.54 ± 16.8	$5.91 \times 10^{13} \pm 7.95 \times 10^6$	2.29 ± 0.05	29.97 ± 0.09	2.76 ± 0.01	35.63 ± 0.2
Malling Kent	26.86 ± 0.06	302.5 ± 0.65	1.09 ± 0.16	20.27 ± 0.06	3987 ± 1.23	$10\,850.53 \pm 4.12$	8241.43 ± 18.96	$5.94 \times 10^{13} \pm 7.27 \times 10^6$	0.74 ± 0.04	29.91 ± 0.08	2.84 ± 0.01	30.9 ± 0.2
Spartan	25.91 ± 0.06	319.16 ± 0.64	22.88 ± 0.18	20.12 ± 0.05	3912.47 ± 1.18	$10\,512.29 \pm 3.42$	8733.05 ± 17.47	$5.88 \times 10^{13} \pm 8.58 \times 10^6$	1.57 ± 0.04	29.96 ± 0.08	3.75 ± 0.01	40.08 ± 0.2
Suntan	25.38 ± 0.06	307.09 ± 0.63	22.91 ± 0.16	22.83 ± 0.06	3997.95 ± 1.23	$10\,875.44 \pm 4.18$	9673.65 ± 17.74	$5.94 \times 10^{13} \pm 6.8 \times 10^6$	2.7 ± 0.04	30 ± 0.08	3.83 ± 0.01	30.58 ± 0.19
Tydemans' Early Worcester	27.03 ± 0.06	351.99 ± 0.63	22.97 ± 0.16	20.78 ± 0.05	3963.57 ± 1.2	$10\,676.47 \pm 4.09$	9413.76 ± 17.27	$5.94 \times 10^{13} \pm 6.61 \times 10^6$	2.4 ± 0.04	29.95 ± 0.08	2.89 ± 0.01	38.08 ± 0.19
Tydemans' Late Orange	28.66 ± 0.06	282.34 ± 0.61	23 ± 0.16	20.9 ± 0.05	3990.56 ± 1.15	$10\,785.38 \pm 3.5$	8943.02 ± 16.32	$5.88 \times 10^{13} \pm 7.82 \times 10^6$	-2.04 ± 0.04	29.98 ± 0.08	3.41 ± 0.01	32.16 ± 0.18
Worcester Pearmain	23.66 ± 0.06	309.29 ± 0.6	0.94 ± 0.16	23.16 ± 0.05	3962.44 ± 1.15	$10\,871.61 \pm 4.12$	8593.58 ± 16.48	$5.94 \times 10^{13} \pm 6.22 \times 10^6$	4.12 ± 0.04	25.92 ± 0.07	3.26 ± 0.01	30.93 ± 0.18

The PhenoFlex model fitted well to the individual training datasets of twenty-six cultivars, resulting in an average RMSE of 6.15 ± 0.22 days and an R^2 value of 0.99 (Table 3.3 and Figure 3.2A). A decline in model performance (RMSE 13.8 ± 0.53 days) was observed when the fitted model was used to predict flowering date on the test datasets (Figure 3.2B). The resulting R^2 value was negative (-3.93), indicating a poor model fit. Poor model performance was particularly apparent for ten cultivars: Cox's Orange Pippin, Egremont Russet, Fiesta, Golden Delicious, Greensleeves, Katy, Mallin Kent, Spartan, Tydemans' Early Worcester, and Worcester Pearmain, with RMSEs above 13 days. When these ten cultivars were excluded, the R^2 value of the test data improved to 0.43.

Figure 3.2: A comparison of the mean observed and predicted bloom dates using cultivar-specific parameters on the PhenoFlex model on A) training data and B) test data. The dashed line represents the line of equality or the $y = x$ relationship between the x and y coordinates. The square points in B represent cultivars which have RMSEs greater than 13 days.



Linear regression with forwards and backwards selection were used to determine which of the twelve PhenoFlex parameters are correlated with high RMSEs. There were no significant correlations between high RMSE and any of the twelve parameters.

Of the twenty-six cultivars, seven cultivars – Egremont Russet (5.12 ± 1.26 days),

Table 3.3: Average RMSE and RPIQ for each model parameterisation approach from training and test datasets.

Data	Model	Average RMSE	RPIQ
Test	PhenoFlex - Mean FT group 1	5.46±0.6	1.75
Test	PhenoFlex - Mean FT group 2	4.34±0.47	1.86
Test	PhenoFlex - Mean FT group 3	5.5±0.42	1.55
Test	PhenoFlex - common	5.64±0.30	1.67
Test	PhenoFlex - specific	12.58±0.58	0.75
Test	StepChill - common	32.38±0.46	0.29
Training	PhenoFlex - Mean FT group 1	9.68±0.69	1.06
Training	PhenoFlex - Mean FT group 2	4.98±0.35	1.60
Training	PhenoFlex - Mean FT group 3	8.42±0.42	1.11
Training	PhenoFlex - common	8.62±0.30	1.16
Training	PhenoFlex - specific	6.08±0.22	1.64
Training	StepChill - common	9.6±0.34	1.04

Gala Mondial (4.54 ± 1.58 days), Howgate Wonder (4.49 ± 1.25 days), Jupiter (4.95 ± 1.34 days), Katy (6.42 ± 1.88 days), Lane’s Prince Albert (5.44 ± 1.22 days) and Malling Kent (4.42 ± 1.16 days) – resulted in test data RMSEs smaller than the standard deviation of flowering dates between years. The RPIQ of the cultivar-specific PhenoFlex model was 1.64 for the training data but only 0.75 in the test data (Table 3.3).

3.3.2 PhenoFlex models fitted to groups of cultivars as identified by mean flowering dates and variation across years

The cultivars separated well in the first two dimensions of the PCA scores (Figure 3.3A), with three clusters identified using the silhouette method (Figure 3.3B). Group

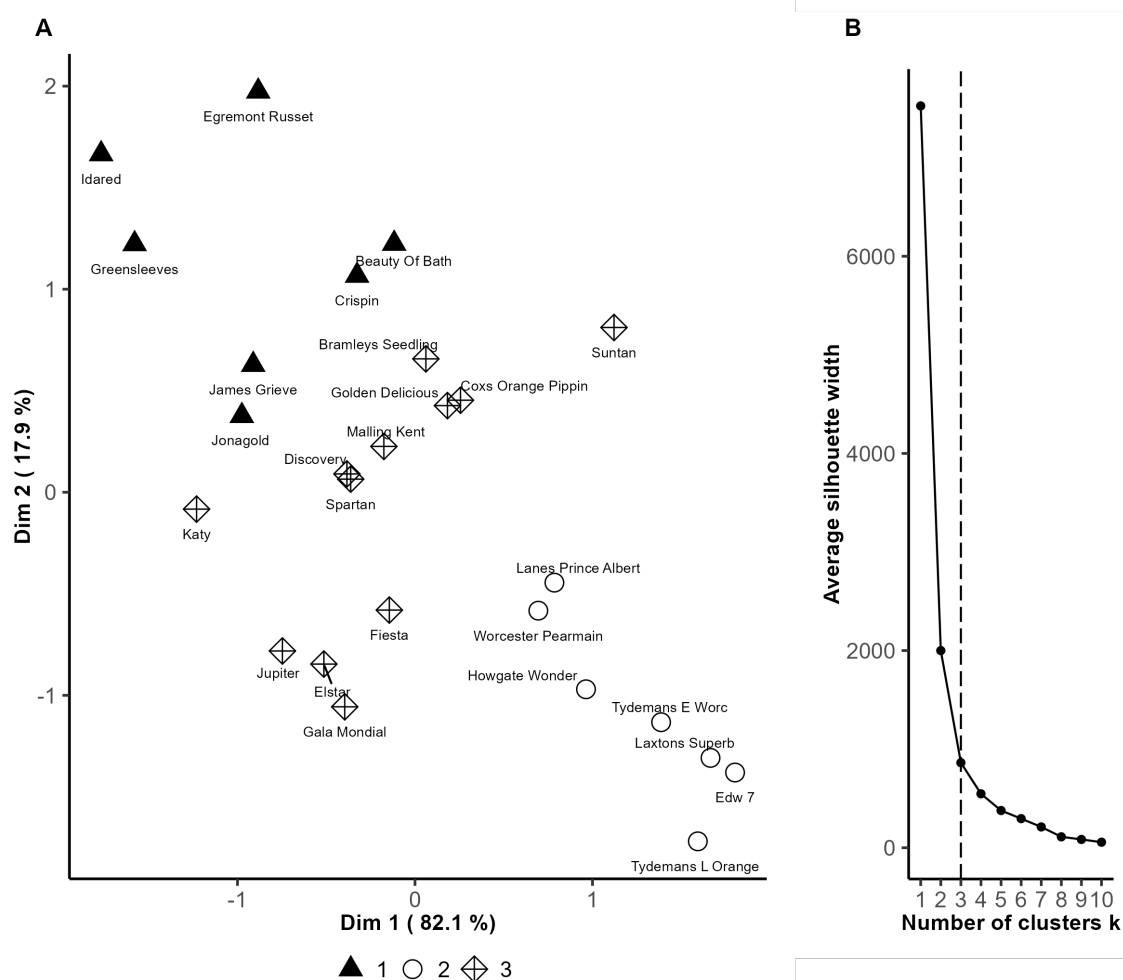
Group	Mean standard deviation of bloom dates	Mean number of training years
Common better	9.89	13.00
Mean Flowering Time better	8.62 ± 0.90	22.50 ± 3.18
Common and Mean Flowering Time similar	8.69 ± 0.09	33.91 ± 1.34
All models similar	8.59 ± 0.08	30.42 ± 0.70

Table 3.4: Analysis of bloom dates and training data across cultivar groups which showed either more accurate predictions with common, specific, or mean flowering time models or whether they are similar. Some only had one cultivar so no standard error was calculated.

one contained 7 cultivars with 280 flowering dates, group 2 contained 7 cultivars with 292 flowering dates and group 3 contained 12 cultivars with 563 flowering dates (Table 3.1). K-means was applied on mean flowering dates and variation across years to divide the cultivars by their flowering behaviours. The variance of flowering patterns in group 1 were the highest at 9.48 days, followed by group 3 at 8.96 days and lastly group 2 at 5.36 days. The cultivars from group 1 contains the least genetic variability with a total of 32 trees, followed by group 2 with 43 trees and group 3 with 56 trees (Table 3.1).

We fitted PhenoFlex models for three groups of cultivars identified via K-means clustering of the means and variations of flowering dates (Figure 3.4A). The model was run 10 times, each with different initial parameters. The average RSS identified from the 10 runs on the mean flowering time groups were 19507.29 ± 201.39 , 5661.96 ± 111.61 and 2911.12 ± 209.45 for groups 1, 2 and 3, respectively. Group 1 consists mostly of cultivars which bloom earlier in the season, group 3 consists of cultivars blooming late in the season, and group 2 contains cultivars which bloom sometime in between. As the groups were split by their mean flowering dates, R^2 values are not that relevant, but were reported to be 0.51, 0.58 and 0.33 on the training data and 0.54, 0.74 and -0.06 for the test data for groups 1, 2 and 3, respectively. Evaluation

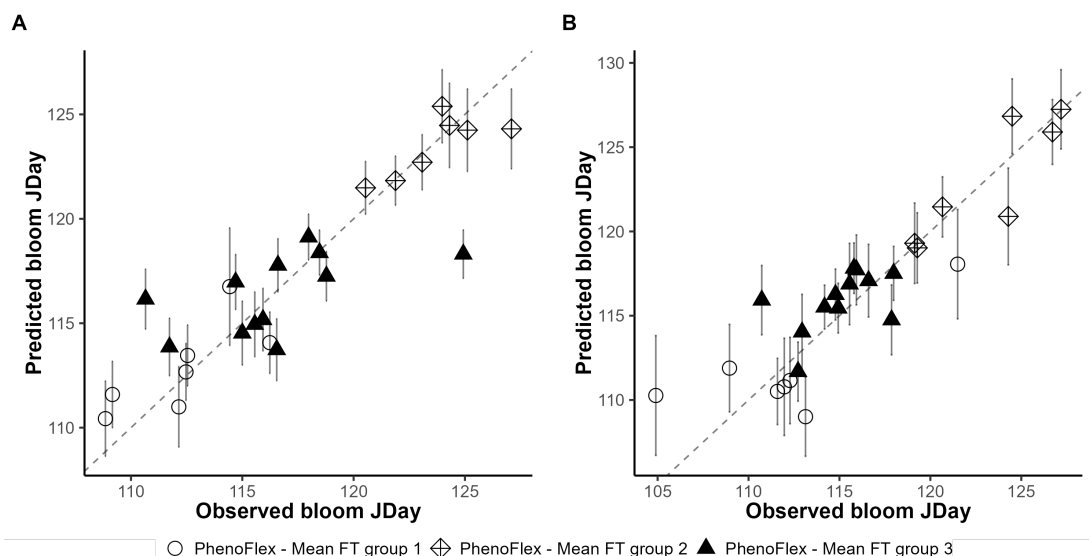
Figure 3.3: K-means clustering presented in a PCA plot for mean flowering time and flowering variation across years for the twenty-six cultivars. B) Silhouette plot indicating three optimal clusters.



against the test datasets showed the best model performance among the three model approaches. The RMSE for groups 1,2 and 3 were 9.68 ± 0.69 , 4.98 ± 0.35 and 8.42 ± 0.43 days on the training data, respectively. The RMSE remains consistent on the test dataset, at 5.46 ± 0.60 , 4.34 ± 0.47 and 5.50 ± 0.42 days for groups 1, 2 and 3, respectively (Figure 3.4). The average RMSE for the three groups were all less than the standard deviation of the interannual flowering dates, so these model parameters identified were significantly better than taking the average flowering date of each cultivar. The mean flowering date approach is a significant improvement on the

predictive accuracy of flowering dates compared to using cultivar-specific parameters. The RPIQ of the mean flowering time groups 1, 2 and 3 were 1.06, 1.60 and 1.11, respectively (Table 3.3). Their RPIQs improved when applied to the test dataset (group1 = 1.75, group 2 = 1.86 and group 3 = 1.55). Overall, the mean flowering date clustered groups performed better than the cultivar-specific model.

Figure 3.4: A comparison of the observed and predicted bloom dates of cultivars grouped by K-means clustering on mean flowering and variation on the PhenoFlex model for A) training data and B) test data. The dashed line represents the line of equality or the $y = x$ relationship between the x and y coordinates.



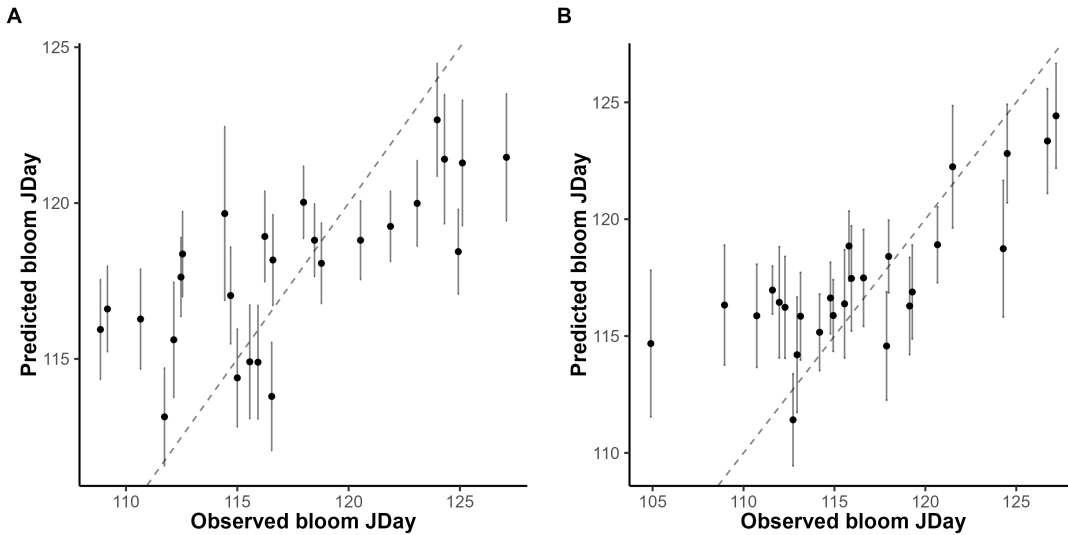
The mean flowering clustered groups performs well but are outperformed by the cultivar-specific model predictions on four cultivars (Beauty of Bath, Edw7, Greensleeves and Katy) on the training data (Figure 3.6A). In the test dataset, the mean flowering clustered group outperformed both the cultivar-specific and common model approaches in all but Beauty of Bath (Figure 3.6B).

3.3.3 A common PhenoFlex model fitted to all cultivars

The common model was run 7 times with various initial parameters. The average RSS identified across the seven runs was 60431.79 ± 785.74 . The standard error is higher than the previous approaches, likely because the model needs to allow for larger errors

to fit a more generalised model with more cultivars. The common model performance on the training data was 8.62 ± 0.31 days, with an R^2 value of 0.44 (Figure 3.5). Its performance on the test data yielded a RMSE of 5.64 ± 0.30 days and R^2 of 0.53, which was better than cultivar-specific model performance. Unlike the specific model, which predicts bloom dates later than the observed bloom date, the common model predicts flowering time around the observed flowering date (Figure 3.5). The RPIQ observed for the common model was 1.16 for the training data and 1.67 for the test data (Table 3.3). This RPIQ is on par with the RPIQ observed using clustered mean flowering dates. The common model produced smaller RMSE than the standard deviation of flowering time for all cultivars, suggesting that model predictions are better than taking the average bloom date for each cultivar.

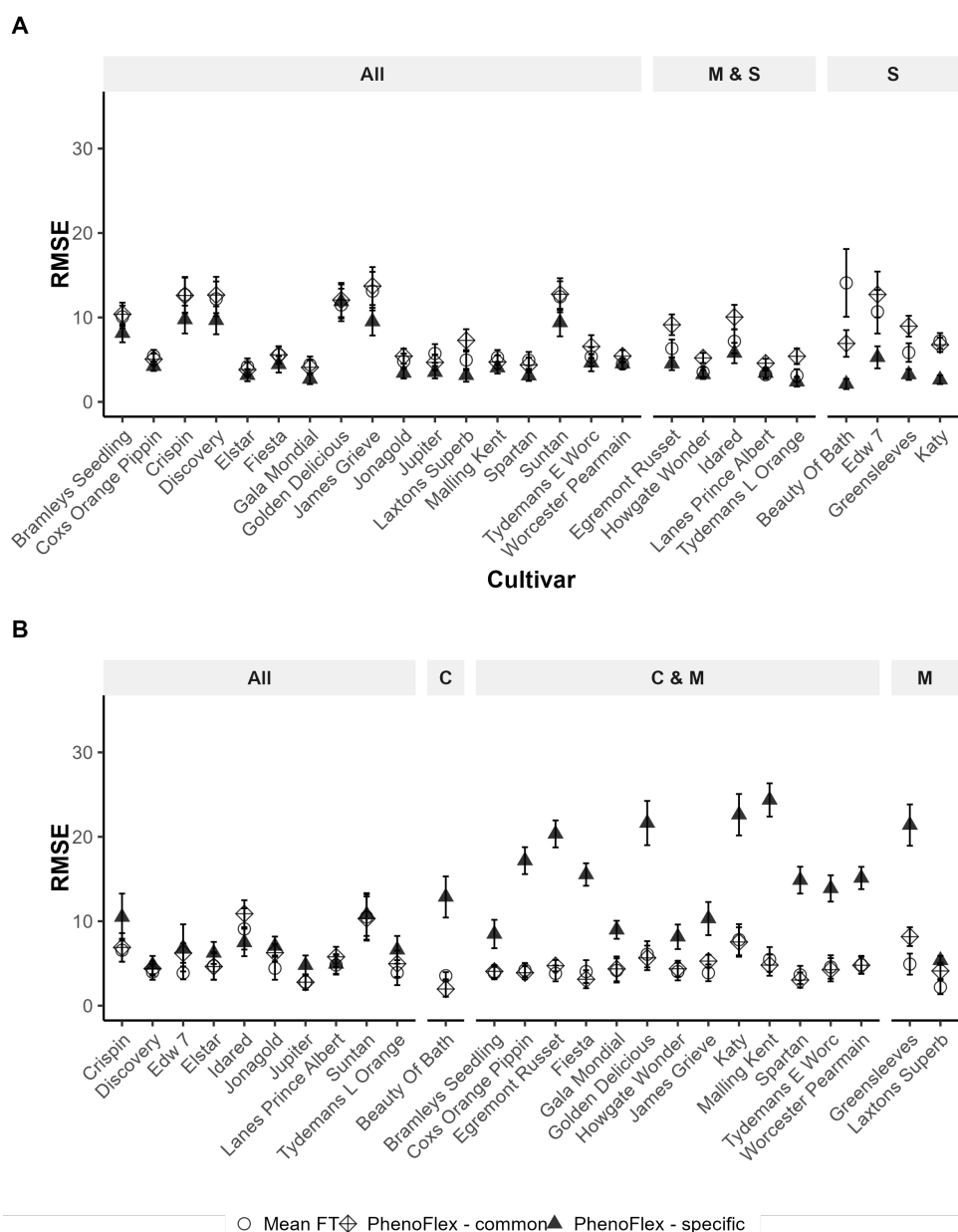
Figure 3.5: A comparison of the observed and predicted bloom dates using common parameters on the PhenoFlex model on A) training data and B) test data. The dashed line represents the line of equality or the $y = x$ relationship between the x and y coordinates.



The cultivar-specific model was able to more accurately predict bloom dates better on the training data (Figure 3.6A), but the common model predictions outperformed the cultivar-specific model predictions or were the same between model approach predictions in the test dataset (Figure 3.6B). The common model approach was

outperformed by the mean flowering cluster approach on Greensleeves and Laxton's Superb (Figure 3.6B).

Figure 3.6: Comparison of RMSE of the cultivar-specific, mean flowering grouped and common models for A) the training data and B) the test data. The graphs are separated by whether the specific (S), grouped by mean flowering date (M) or common (C) models perform better. When two of the approaches perform equally well, two letters are shows (M&S and C&M) or whether there is no difference between approaches (All).



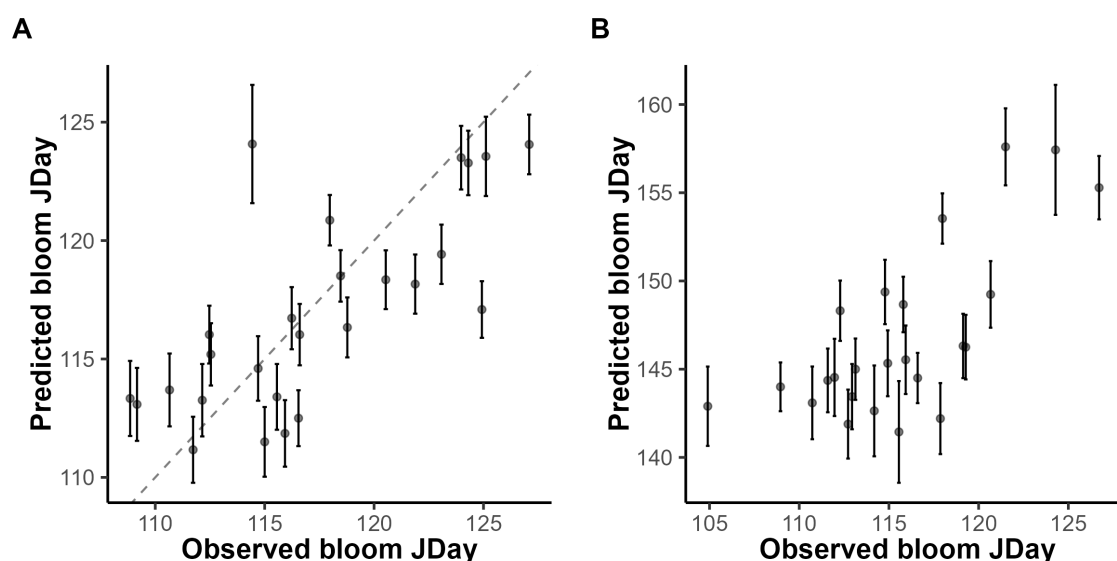
When there are large numbers of data per cultivar (more than 30 years), the cultivar-specific approach does well in predicting bloom dates. When there are around 20 years

of data, the mean bloom date cluster performs well since this method increases the number of datapoints by including more cultivars. When there are only few numbers of years per cultivar (approximately 10 years), but many cultivars are present, it is better to apply the common approach (Table 3.4).

3.3.4 Comparison of the common PhenoFlex and StepChill models

Fitting the training the data with the StepChill model resulted in an average RMSE of 9.60 ± 0.34 days and an R^2 of 0.52 (Figure 3.7A). The common StepChill model did not predict bloom dates well for the test data, resulting in a RMSE of 32.4 ± 0.46 days and R^2 of -30.86. All predicted flowering dates were later than the observed bloom date (Figure 3.7B). Overall, the PhenoFlex model predicted flowering dates 82.59% more accurately than the StepChill model. This agrees with the observed RPIQ values. The common StepChill model produced smaller RPIQ (0.29) on the test dataset compared to the common PhenoFlex model.

Figure 3.7: A comparison of the observed and predicted bloom dates using common parameters on the StepChill model on A) training data and B) test data. The dashed line represents the line of equality or the $y = x$ relationship between the x and y coordinates.



AICc model selection was used to select between possible models which were trained on the same dataset and the Nash-Sutcliffe efficiency of the model was used to quantify the predictive effectiveness of a model, with a good efficiency being close to 1. The common PhenoFlex model with twelve parameters was deemed to be better than the StepChill model with five parameters as PhenoFlex model AICc (3438.53 ± 5.29) was lower than StepChill model AICc (3611.21 ± 4.65), and the PhenoFlex model efficiency (0.27 ± 0.003) was higher than StepChill model efficiency (0.08 ± 0.004).

The common PhenoFlex model was more reliable as it was able to predict bloom dates for all the years, but the common StepChill model was unable to predict 2 datapoints from 1950 on the test dataset.

3.4 Discussion

In this study, a large collection of flowering dates of twenty-six apple cultivars at East Malling, Kent, England was used to fit PhenoFlex models to predict flowering time, and to compare the PhenoFlex and StepChill models that were fitted to the training data of all twenty-six cultivars.

Predictive performance of cultivar-specific approach is the worst of the three modelling approaches, with large RMSEs (13.8 ± 0.53 days). Most predictions were worse than taking the average bloom dates of each cultivar as the predicted bloom date. Flowering predictions for seven of twenty-six cultivars were better than taking the average flowering date. The groups identified by applying K-means on mean flowering dates and their variation were significantly better than using the cultivar-specific model. The model predictions were better than taking the average bloom dates for each cultivar. The common model performed as well as the second approach, with similar RMSEs and RPIQ.

We can speculate that the difference in model performance results from model over-

fitting on training data of smaller datasets, excessive model complexity and noisy training data, with more emphasis placed on the first two factors. Cultivar-specific models tended to do well when the dataset is large (approximately 30 years or more of training data per cultivar), the mean flowering date grouped model did well with at least 20 years of training data per cultivar and the common model does well with even less training data per cultivar. Small datasets can restrict the optimisation functions' performance, increasing the chance that the model parameters converge at a local minimum rather than the global minimum. In the present study, flowering data were only available for a limited number of years for several cultivars and for some, flowering dates were only observed for a limited number of trees, thus limiting the genetic influence on flowering time. Another reason could be the large variability in flowering dates between years for some cultivars or variability in flowering dates between cultivars in the same year, which could make parameter convergence difficult in model optimisation. Problems associated with small datasets are further exacerbated when complex models are fitted.

The common PhenoFlex model performs as well as previous literature since its average RMSE (5.64 ± 0.30 days) is comparable to previous results predicting flowering dates for Boskoop (RMSE = 4.2 days), Cox's Orange Pippin (RMSE = 5.7 days), Golden Delicious (RMSE = 5.12 days), Jonagold apples (RMSE = 4.57 days) [88] and Crispps Pink apples (RMSE = 14.7 days) [82] using the Dynamic and GDH models sequentially. However, previous application of the PhenoFlex model on a single cultivar (Boskoop apples with RMSE = 3.82 days) [44] attained better predictive accuracy than the grouped and common models. As discussed above, the size of available data and number of individual trees are key factors affecting model predictive performance. A potential improvement for the common model would be to retain the CR and HR of the individual models but identify a common set of parameters for the other 10 parameters. In theory, this would be the best common model as previous studies agree

that CR and HR are species specific [77].

Present modelling suggested that cultivars differ in the exact PhenoFlex model parameter estimates, although grouped by mean flowering date and variation of bloom dates and common PhenoFlex model led to more accurate predictions than the cultivar-specific model. Predictions given by both approaches were more or less balanced (namely varying around the observed dates). In contrast, cultivar-specific models predicted flowering later than the observed. Consistent large overpredictions (bias) may suggest model under-fitting. Small datasets may be the cause of poor predictive performance of cultivar-specific models. Under-fitting is more likely due to the fact that the data do not contain sufficient information on the generic feature due to large differences in the generic features between cultivars.

We grouped cultivars together based on their mean flowering date and variation of bloom dates. A previous study attempted a similar modelling approach with several phenology models to predict apricot flowering time, but they did not find good predictive results from modelling at the species level (common model approach) [89]. They did find good results when split into smaller precocity groups for early, intermediate and late flowering, which is similar to what was observed in the current study. The common model was speculated to work well because the data originated from one site, thus reducing the generic variability and difference in response to different conditions.

Drepper et al. [88] identified much larger chilling and heat requirements for Cox's Orange Pippin (CR = 79.6 and HR = 4430), Golden Delicious (CR = 59.84 and HR = 4980), Jonagold (CR = 60.66 and HR = 4980) and Boskoop (CR = 59.16 and HR = 4430). This agrees with the chilling (45) and heat (8500) requirements used by Darbyshire et al., [82] on Boskoop apples. Using a combined chilling and heat model which can adjust for the overlap between the models appears to lower the chill and heat requirements for flowering by 2- to 3-fold. Luedeling et al., [44] identified parameters very similar (CR = 25.4 ± 3.2 and HR = 348 ± 31) to the parameters

identified in the common model (Table 3.5) as well as similar levels of overlap as shown by their $s1$ value of 1 ± 22 . The common model implements a large overlap between the chilling and heat models, as indicated by the smaller $s1$ parameter (0.55 ± 0.08). The present model is more likely to have a greater level of overlap as the $s1$ parameter estimate has a much smaller standard error than in Luedeling et al. [44].

Model	y_c	z_c	$s1$	T_u	$E0$	$E1$
Common	25.36 ± 0.14	286.01 ± 9.34	0.55 ± 0.08	21.22 ± 0.29	$3.82 \times 10^3 \pm 4.33 \times 10^{-1}$	$1.03 \times 10^4 \pm 2.74 \times 10^{-1}$
Group 1	22.6 ± 0.46	321.95 ± 13.09	0.55 ± 7.23	18.61 ± 0.55	$3.82 \times 10^3 \pm 1.22$	$1.02 \times 10^4 \pm 1.49$
Group 2	22.3 ± 0.23	355.14 ± 3.31	22.96 ± 8.04	20.24 ± 0.23	$3.92 \times 10^3 \pm 0.772$	$1.04 \times 10^4 \pm 0.508$
Group 3	28.44 ± 0.18	305.96 ± 4.19	0.76 ± 2.23	19.73 ± 0.15	$3.82 \times 10^3 \pm 0.157$	$1.03 \times 10^4 \pm 0.47$
Model	$A0$	$A1$	T_f	T_c	T_b	slope
Common	$8.51 \times 10^3 \pm 6.70$	$5.91 \times 10^{13} \pm 2.07 \times 10^8$	-3.64 ± 1.73	27.27 ± 3.13	2.85 ± 0.06	46.87 ± 8.45
Group 1	$1.025 \times 10^4 \pm 27.3$	$5.89 \times 10^{13} \pm 2.22 \times 10^8$	0.82 ± 2.18	28.74 ± 3.52	3.53 ± 0.17	40.22 ± 8.78
Group 2	$1.084 \times 10^4 \pm 27.1$	$5.94 \times 10^{13} \pm 2.08 \times 10^8$	2.35 ± 1.69	27.15 ± 3.31	3.18 ± 0.15	52.1 ± 9.48
Group 3	$9.996 \times 10^3 \pm 12.5$	$5.92 \times 10^{13} \pm 2.27 \times 10^8$	-2.15 ± 1.99	29.7 ± 3.55	3.09 ± 0.15	31.01 ± 10.23

Table 3.5: The best fit parameters of the PhenoFlex model for the K-means groups and Common (all cultivars). The standard errors are calculated on 99 bootstraps and the original parameters.

The common PhenoFlex model had much better predictive performance than the StepChill model. Although both models had the similar goodness of fit for the training dataset, the StepChill model had much worse predictive performance, predicting flowering dates much later than observed. Moreover, for 1950, it failed to predict flowering time. The large bias in predictions indicated a model underfitting, suggesting that the model formulation does not capture much of the generic responses of apple flowering development in response to temperature.

In this study, over 70% of hourly temperature data was generated using the recorded daily maximum and minimum temperatures following a sine curve for warming and logarithmic decay for cooling temperatures in the data collected prior to the year 2000 as temperatures were only recorded as daily maximum and minimum values. In doing so, it must be assumed that temperatures more or less follow these trends for warming and cooling. Since the models are parameterised on the available data, their reliability is only as good as the simulated data. In this current study, the effectiveness

of the three approaches in predicting flowering date was compared, and since the hourly temperatures are the same for all approaches, this would unlikely impact the comparisons between models. Nevertheless, it would be informative to repeat this experiment with non-simulated hourly temperatures to understand the impact of using simulated hourly data on model fitting and performance.

Although apple trees are mostly affected by temperature, future applications of this model in different countries may need to consider environmental differences that can affect maturity features. An example is radiance. Comparing the UK, which is more often cloudy, to New Zealand, which has plenty of sun and is affected by low Ozone air from Antarctica, the radiance can vary strongly and affect the predictive results of the models. Even changes in microclimates (shaded vs non-shaded) can affect maturity, let alone major changes in their growing environments. The author strongly recommends parametrising the model using data from their respective countries.

3.5 Conclusion

The present research showed that the PhenoFlex model approach is an improved approach over the StepChill model to predict apple flowering development in relation to temperature, as concluded from higher EFF and lower AICc and RMSE values. Moreover, the common PhenoFlex model had no failed predictions. Contrary to the hypothesis, the common PhenoFlex model and cultivars grouped by mean flowering time was found to result in the best predictive accuracy and highest RPIQ compared to the cultivar-specific approach. Grouping cultivars by similar flowering dates can be used to adjust for low numbers of data in individual cultivars or grouping all data of a species from one region will yield better results than modelling each cultivar independently. Much of the poor model performance may be associated with small data sizes for fitting a complex model such as the PhenoFlex model with 12 parameters. In future, PhenoFlex models should be developed with a much-increased dataset through

merging data from different locations as well as formulate specific CR and HR for each cultivar whilst having common values for the other 10 parameters.

Flowering variation induces apple maturity variation at harvest

4.1 Introduction

It is established that fruit quality after long-term storage has been linked to fruit maturity at the time of harvest [19, 94, 95]. Fruit picked too early are more likely to develop storage disorders such as core rot and internal browning [96]. Moreover, underdeveloped fruit are less attractive to consumers due to poor development in secondary metabolites (flavonoids). The group of secondary metabolites not only adds flavour and colour to fruit, but also have antioxidant, anticancer, antiviral and anti-inflammation properties. On the other hand, although over-ripe fruit contains these beneficial and attractive qualities, they tend to be too soft and easily bruised, leading to rot and making it unsuitable for long-term storage [95, 97]. In addition, over-ripe fruit produces ethylene, which causes an autocatalytic ripening effect on itself and other fruit in the vicinity, causing them to ripen. Picking fruit at the optimal harvest time can thus reduce fruit loss and increase fruit quality.

The optimal time of harvest is difficult to predict because there is no single clear index for apple maturity. Changes in several physical traits can be observed, such as colour, firmness and sweetness, but these traits can be highly variable even among individual fruit on the same tree. These traits can vary year to year, with the extent of variability

depending on cultivars, partially due to the variability in the season's conditions [1, 95]. The industry standard for maturity assessments includes destructive assessments of small samples of fruit, to measure firmness, Starch Pattern Index (SPI), soluble solids and ethylene. Firmness, soluble solids and SPI measurements can be aggregated into a single variable called Strief [98]. However, no single feature can represent the true maturity of apples, and an aggregation of multiple features is more effective [98, 99]. Regardless of the maturity indicator, these assessments only forecast up to a week from the harvest window. Therefore, a clear definition of maturity and earlier indicators for the optimal harvest window are necessary for effective orchard harvest and post-harvest management [50, 100].

A set calendar date after full bloom is not a good predictor of harvest windows, especially with changing climates [99]. Statistical methods can be used to predict harvest windows for specific cultivars. Typically, they are done using growth models based on the accumulation of effective temperatures within a specified temperature limit [51, 101]. Assorted windows of time were tested, as temperature has different effects on fruit depending on the stage of development [48]. Many model variants account for temperatures at different stages of development. Several studies showed that temperatures within 30-days post-bloom improved harvest date predictions [102, 103]. Perry et al. [51] trialled 30-, 40-, 50- and 60-days post-bloom temperatures, but were only able to improve predictions by 1 day for Delicious apples, and no improvement for Golden Delicious apples. Other studies found that a three-month window (June, July and August) improved the prediction accuracy by 3 days [50]. The complete harvest window from full bloom to harvest only improved harvest day predictions by 1 day [104]. Predictive results often varied greatly with cultivars [105]. In most studies, average daily temperatures instead of hourly temperatures were used to predict harvest time. Thus, if the underlying relationship of temperature with fruit development is not linear, the use of hourly models is expected to improve predictions.

The term “full bloom” can mean different stages of flowering: 50% bloom [1], 70% bloom [52] or 80% bloom [50, 106]. Moreover, treating flowering times of all fruit in a single orchard as a single date ignores the variability in the flowering time between trees and between individual flowers on the same tree. Later blooming flowers developed fruit which were smaller and had greater levels of starch in apples [54]. Thus, inaccuracies in predicted harvest time by previous models may be partially attributable to this simplification of flowering time for a given orchard. It can be argued that a better predictive model should be based on individual fruit, from flowering to harvest.

Fruit position within the canopy has been shown to affect maturity of apples [107]. The fruit position is used as a proxy for light exposure, as fruit grown on the outer zones of the trees were more likely to have higher soluble sugars [54, 107, 108] and be firmer [73–75]. Inconsistent results, occasionally showing no change or opposing findings in maturity indices [109], could be due to varying cultivar responses to the environment.

The most commonly used harvest prediction models use only the average flowering date to estimate at a single point in time to harvest. However, the impact of extended flowering time is not considered in these models, nor are other environmental factors. In this study, the aim was to determine the degree of impact on harvest maturity due to flowering time, season (year), region of the canopy and individual trees. Knowing the effect of these factors can improve harvest predictions and improve the overall fruit quality at the picking time. To model temperature effect on fruit development, three growth rate models (linear growing degree hours (linear GDH), non-linear GDH and Thermodynamic) were used to relate temperature to fruit maturity. The relative importance of individual factors (flowering time [i.e., temperature accumulation], season, tree, canopy region) was assessed in the framework of generalised linear modelling (GLM). In our study, flowering time was found to affect fruit maturity at harvest, but the degree of the effect is cultivar dependent.

4.2 Methods

4.2.1 Plants

We focused our research on five apple cultivars: Braeburn, Cox, Fuji, Gala and Golden Delicious. The same 14 trees were monitored over the 2022 and 2023 seasons, except for Cox's Orange Pippin, which only had 5 fruit in 2022. There were three trees for each cultivar, except Fuji for which one tree was severely diseased and thus there were no fruits from this tree reaching maturation. The three trees of the same cultivar were adjacent to each other in an orchard; all five cultivars were situated in adjacent rows, which run north to south, with 2 m spacing between each tree and 2.5 m spacing between each row. All trees were about 12 years old and grafted on M9 rootstocks. The orchard received basic orchard management, but no commercial thinning or pruning was applied. Natural June drop was the only process of fruit thinning. All trees were at most 2.5 m tall and 2 m wide.

4.2.2 Flowering records

Apple flowers grow in clusters. Flowering records were done by tagging clusters of flowers with the date of bloom. The bloom date was noted when the majority (three of five flowers) of the cluster was fully open. Therefore, the flowering dates used in this study were when flowers on positions 2 and 3 were fully open, which usually occurred a day after the king bloom flower opened and a day before the flowers at positions 4 and 5 opened. A total of 1199 flower clusters were tagged between the 14 trees, made up of 12 to 85 clusters per tree (Table 4.1).

4.2.3 Tree canopy zones

A single tree canopy was divided into 7 zones as a proxy for fruit exposure to light: north, south, east, west, upper, inner and lower (Table 4.1). The first 5 zones are

Table 4.1: Summary of fruit and cluster quantities across different trees and zones in 2022 and 2023. Three trees of each cultivar were used in the study, except for Fuji. Collection of fruit occurred in 2022 and 2023 but Cox was only collected in 2023. Cluster represents the total flower clusters collected from each cultivar, year and tree.

Cultivar	Tree	Year	Total fruit	Clusters per tree	Tree Canopy Zones						
					North	South	East	West	Upper	Inner	Lower
Braeburn	BB1	2022	79	50	7	2	8	9	2	13	9
Braeburn	BB1	2023	124	77	3	2	4	5	7	29	27
Braeburn	BB2	2022	63	38	10	2	1	5	6	3	11
Braeburn	BB2	2023	133	78	8	2	8	4	3	28	25
Braeburn	BB3	2022	76	49	5	9	8	7	2	6	12
Braeburn	BB3	2023	72	59	1	2	4	9	11	14	18
Braeburn Total			547	351	34	19	33	39	31	93	102
Cox	CX1	2023	76	58	8	5	10	5	7	17	6
Cox	CX2	2023	37	13	2	0	0	0	0	7	4
Cox	CX3	2023	131	85	3	6	6	5	7	37	21
Cox's Orange Pippin Total			244	156	13	11	16	10	14	61	31
Fuji	FJ2	2022	77	31	2	1	3	0	6	10	9
Fuji	FJ2	2023	56	28	8	2	1	0	1	4	12
Fuji	FJ3	2022	108	40	1	4	6	1	6	7	15
Fuji	FJ3	2023	51	22	1	1	0	1	2	10	7
Fuji Total			292	121	12	8	10	2	15	31	43
Gala	GL1	2022	113	58	8	8	6	7	8	12	9
Gala	GL1	2023	42	21	0	1	2	0	0	10	8
Gala	GL2	2022	102	49	7	6	4	7	3	14	8
Gala	GL2	2023	130	59	4	4	1	0	6	34	10
Gala	GL3	2022	145	78	6	8	6	10	7	16	25
Gala	GL3	2023	65	34	0	0	0	0	2	19	13
Gala Total			597	299	25	27	19	24	26	105	73
Golden Delicious	GD1	2022	93	36	2	1	5	5	9	7	7
Golden Delicious	GD1	2023	80	50	5	6	6	4	3	15	11
Golden Delicious	GD2	2022	113	49	12	6	0	2	3	8	18
Golden Delicious	GD2	2023	82	40	0	2	1	0	1	19	17
Golden Delicious	GD3	2022	196	85	3	6	10	8	12	34	12
Golden Delicious	GD3	2023	23	12	0	0	0	0	2	9	1
Golden Delicious Total			587	272	22	21	22	19	30	92	66

regions on the outer areas of the trees with greater light exposure, whereas fruit from the inner and lower zones were mostly shaded by foliage during fruit development. The upper region consisted of fruit within the upper 25% of the tree. The fruit from the four cardinal directions were picked from the outer edge of the trees. Fruit picked from the inner and lower of the trees were located close to the trunk and within the lower 25% of the trees, respectively.

4.2.4 Temperature records

An official UK Meteorological Office Station situated approximately 465 m from the orchard in East Malling (51.2876°N, 0.4486°E, 33 m above the mean sea level) collected hourly temperatures.

4.2.5 Maturity measurements

In total, 2267 fruit were collected over two years (Table 4.1). The tagged fruit were harvested at the recommended commercial harvest date for specific cultivars in the UK, and tested for starch, soluble solids (Brix) and firmness within 36 hours from picking. They were always picked in the morning over a 3-month period. Firmness was measured by peeling two sides of the apple at 90 degrees in the equatorial region, avoiding obvious bruises, then using a fruit texture analyser (Llyod LRX, UK) fitted with an 11 mm diameter probe to puncture the fruit to a depth of 8 mm. The force at 8 mm was used for the analysis. The soluble solids was measured using an Atago portable benchtop refractometer (palette series, model PR-32 α), using water to calibrate at the start of each sampling day. The starch pattern index (SPI) is the gold standard in determining apple ripeness. An apple was first cut in half at its equator; one of the halves was then immediately stained with potassium iodine mixture (1% w/v iodine and 4% w/v potassium iodide), leaving it to stain for at least 30 minutes; finally the staining pattern was compared to the CTIFL starch conversion

chart for apples to estimate SPI as an index from 1 to 10, and proportion of staining coverage. The 30-minute staining was more practical and allowed for all apples to be evaluated without large breaks between each set. No significant differences in staining pattern occurs within 30 minutes but starch breakdown may occur if left overnight (Per comms). The SPI, firmness and Brix were transformed into a singular unit called the Streif index [98, 110].

$$Streif = Firmness / (Brix * Starch) \quad (4.1)$$

4.2.6 Models describing temperature effects

Three temperature-based models were evaluated for their relationship with fruit maturity (linear GDH, non-linear GDH and Thermodynamic). They were chosen to model the relationship between temperature and maturity. The preliminary analysis showed that the cultivars differed, hence a model was fitted across the two years for each cultivar separately. Hourly temperatures from flowering to harvest were used to calculate the growth units as specified by each model for individual apples. For each model, a simple grid search was used to identify the set of parameters that maximised the Kendall's Tau correlation between SPI and accumulated temperature measurement from flowering to harvest for individual fruit. All parameters were iterated through a range as specified below for each model at a step length of 0.1. Temperatures at East Malling in 2022 and 2023 did not exceed 40 °C, thus higher temperatures was not explored for model parameters.

Linear Growing Degree Hours

The linear GDH, established by Anderson and Seeley [111], assumes a linear relationship of growth with accumulative temperatures above a temperature (base) threshold. It has 3 parameters, T_b , T_c and T_u , representing the base, critical and

optimal temperatures. Temperatures below the base do not count towards GDH units, nor temperatures exceeding the critical threshold. The contribution of each degree increase in temperature from the base linearly increases as temperature increases, up until the optimum temperature. Temperatures between the optimum and critical temperatures accumulate GDH units at the maximum rate. Thus, strictly speaking, this GDH is not linear but two lines joining at T_u .

$$GDH_{linear}(T) = \sum f_{linear_{GDH}}(T) \quad (4.2)$$

$$f_{linear_{GDH}}(T) = \begin{cases} T_u - T_b, & \text{if } T_u \leq T < T_c \\ T - T_b, & \text{if } T_b < T < T_u \\ 0, & \text{if } T \leq T_b \text{ or } T_c \leq T \end{cases} \quad (4.3)$$

In grid search, T_b spanned from 0°C to 10°C, T_u from 15°C to 25°C, and T_c from 30°C to 40°C. The base temperature originally proposed by Richardson et al. [112] was 4.5°C for peach trees. However, a recent study by Tang et al. [113] found that the base temperatures of apple trees may be lower than 4.5°C. The search was extended from 0°C to 10°C to explore the best fitting base temperature. T_u was chosen based on the expected best growth conditions of most living organisms, and finally, the critical temperature was expected to range somewhere between 30 and 40°C.

Non-linear Growing Degree Hours

The second is another well-established model growing degree hour model by Anderson et al. [34]. Opposed to the linear non-linear GDH model, this model assumes a non-linear accumulative relationship of growth with temperature. Each temperature increase from the base causes a non-linear increase in GDH up until the optimum temperature. Temperatures above the optimal gradually decrease in effectiveness in GDH accumulation (Figure 4.3).

$$GDH(T) = \sum ((T_u - T_b) * f_{non-linear_{GDH}}(T)) \quad (4.4)$$

$$f_{non-linear_{GDH}} = \begin{cases} \frac{1}{2} * \left(1 + \cos\left(\pi + \pi * \frac{T-T_b}{T_u-T_b}\right)\right), & \text{if } T_b < T \leq T_u \\ 1 + \cos\left(\frac{\pi}{2} + \frac{\pi}{2} * \frac{T-T_u}{T_c-T_u}\right), & \text{if } T_u < T \leq T_c \\ 0, & \text{otherwise} \end{cases} \quad (4.5)$$

The grid search range for T_b , T_c and T_u was the same as for the linear GDH model.

Thermodynamic model

The Thermodynamic model is a non-linear growth rate model based on the theory of enzyme activity rate variation in response to temperature changes [114, 115]. The parameters for the Thermodynamic model are B , C , TH and ρ . K is the temperature in Kelvin.

$$R(K) = \frac{\frac{\rho K}{298} \exp[B(1 - \frac{298}{K})]}{1 + \exp[C(1 - \frac{TH}{K})]} \quad (4.6)$$

In the grid search, the range of parameters B was 15 to 40°K, C was 5 to 30°K and TH was 290 to 300°K. The last variable in the equation, ρ , is a scaling factor and does not affect the correlation of the estimated growth unit from flowering to maturity with SPI, so it was fixed to 1.

4.2.7 Assessment of the relative importance of experimental factors in fruit maturity

Logistic regression with starch proportion as the response variable was used to determine the effect of flowering time (as approximated by the estimated temperature-based growth unit), year, individual trees, and fruit position within the tree canopy on fruit maturity. In the GLM analysis, a binomial distribution was assumed for the

residual errors. The deviance explained by each experimental factor was calculated by extracting the residual deviance from ANOVA tables calculated using Chi-square (synonymous to likelihood ratio) as the test function. Since this study focused on the temperature effect (flowering time) on fruit maturity, the accumulated growth models estimated by one of the three models was first added in GLM analysis of SPI for each cultivar. Then, year, tree, and canopy region were added sequentially. A nested model approach was used to test for statistical significance of the effect of specific factors on fruit development (SPI).

4.3 Results

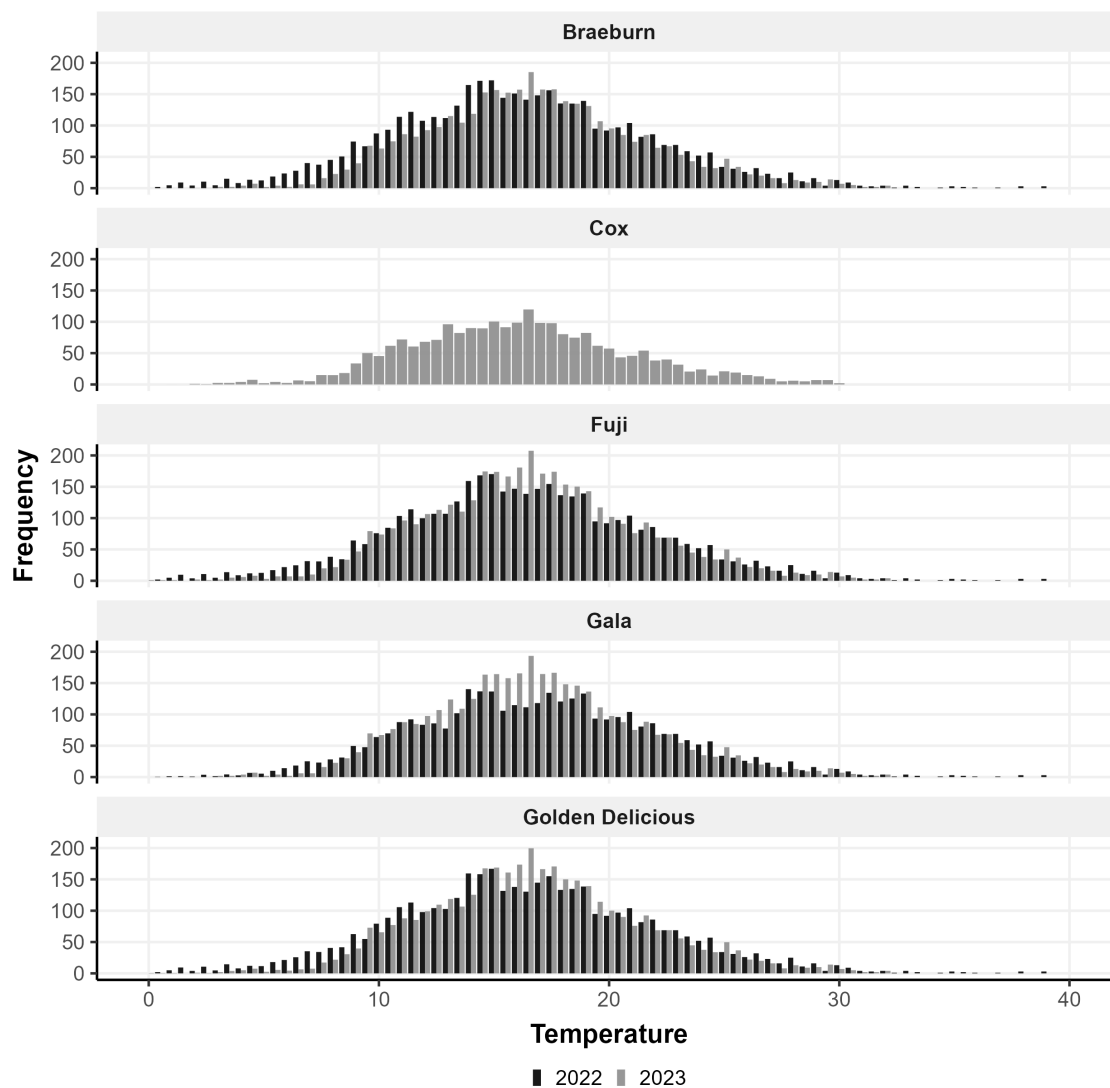
4.3.1 Variability of temperature and effects on flowering

The temperatures in 2022 during fruit development were more extreme than in 2023 and had a slightly lower median temperature (Figure 4.1). The median flowering day in 2022 was earlier than in 2023 by 9.8 days; flowering was spread over 3 weeks in 2022 (Table 4.2). On the other hand, temperatures in 2023 were more often within optimal growing temperatures, so flowering occurred within a shorter time span (~ 11 days), with Cox's Orange Pippin as an exception as there were no fruit available for this cultivar in 2022 (Table 4.2).

Table 4.2: Summary of flowering data from 2022 and 2023 for 5 apple cultivars.

Cultivar	Year	Median flowering Julian Day	Range	Interquartile range
Braeburn	2022	116.0	21	7.00
Braeburn	2023	125.0	10	3.00
Cox	2023	125.0	21	2.00
Fuji	2022	124.5	18	4.00
Fuji	2023	130.0	11	0.00
Gala	2022	120.0	22	8.25
Gala	2023	128.0	12	5.00
Golden Delicious	2022	119.0	19	10.00
Golden Delicious	2023	130.0	12	3.00

Figure 4.1: The frequency of temperatures between 0 and 40°C at 0.5°C increments from the first flowering to harvest for the five apple cultivars in 2022 (black bars) and 2023 (grey bars).



4.3.2 Fruit maturity measurements

Of the five cultivars, Cox is the earliest to mature, whereas Golden Delicious and Fuji matured the latest. Streif can account for several factors in one variable and no single fruit quality parameter has been able to accurately define the maturity of an apple. Therefore, Streif should be a better factor than individual measures for maturity. However, in the present study, Streif segregated the dataset in firmness and soluble solids (Brix) measurements (Figure 4.2). This division of data seen in Braeburn and Gala are not due to differences between year, tree or canopy zones. The division is also seen in Cox indicate no differences between tree or canopy zone. There are no known biological differences between trees, including rootstocks, and since all trees were planted in the same orchard, in adjacent rows, they were exposed to the same environmental conditions and orchard management systems. It is speculated that there might be a latent variable segregating the data. Thus, starch proportion was used as the maturity indicator for the present study. The relationship between Streif and starch proportion appears to follow a logistic function (Figure 4.2), as the degradation of starch starts slowly, then increases rapidly as fruit matures.

4.3.3 Temperature-based fruit development

The correlation between calculated linear GDH and starch proportion had a maximum of 0.54 from Braeburn and a minimum of 0.22 from Fuji (Table 4.3). Golden Delicious and Fuji have the smallest correlation, have nearly identical parameters (Table 4.3) and hence temperature-growth rate relationship (Figure 4.3A). They have the highest optimum and critical temperatures. Gala has the smallest effective temperature range between 9.1 – 30 °C, which may be compensated by the lowest optimal temperature (16.1 °C). Braeburn and Cox have similar temperature-growth rate relationship (Figure 4.2A), but the parameters for Cox are 3 °C lower than Braeburn for all three parameters. For the non-linear GDH model, Braeburn has the highest correlation of 0.54. The least

Figure 4.2: Strief vs soluble solids (Brix), firmness and starch measurements of all apple cultivars collected in 2022 and 2023 from a single orchard. Note: Cox only produced fruit in 2023 and samples were collected from three trees per cultivar, except for Fuji which only had two.

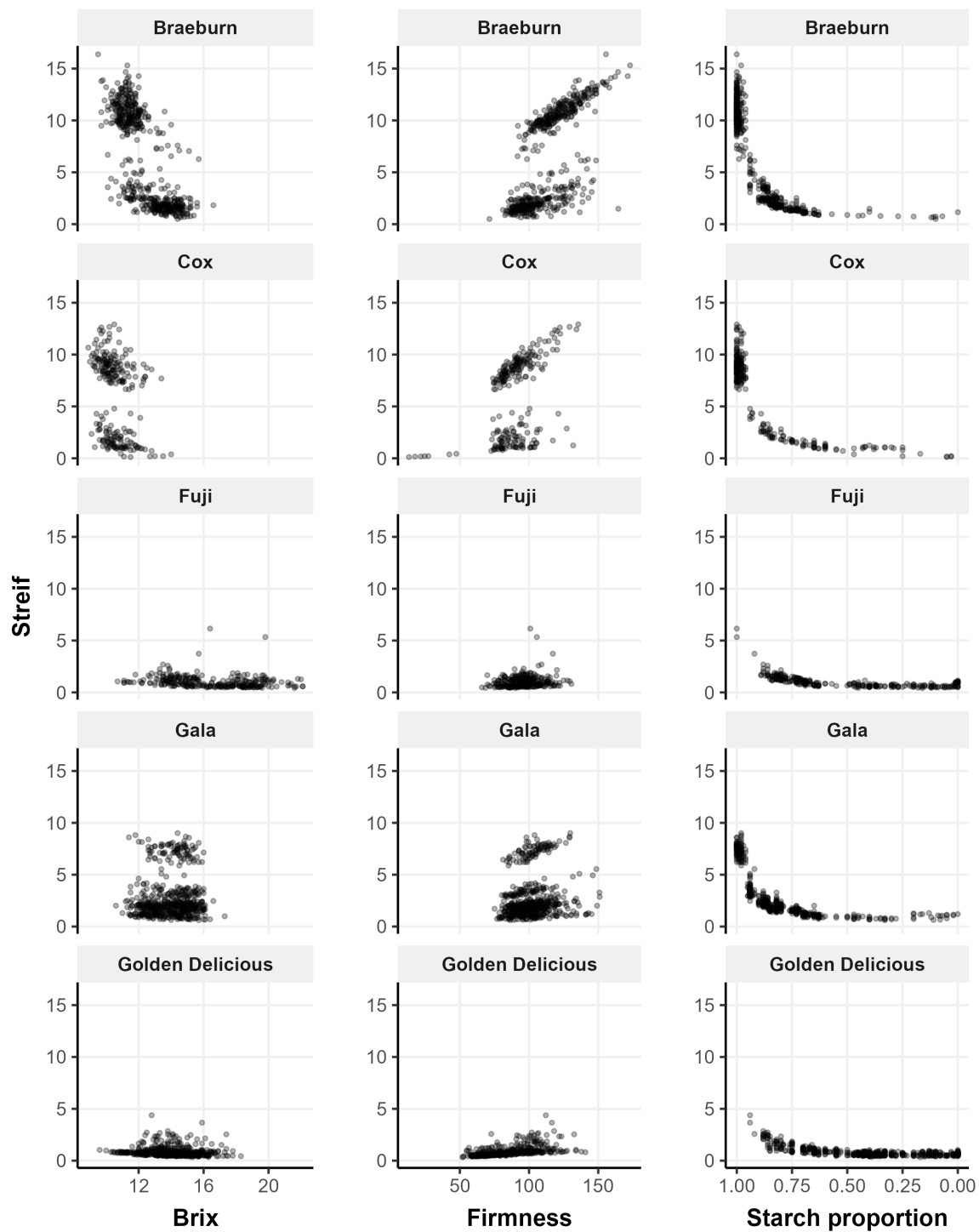


Figure 4.3: Growth rate of apple cultivars for hourly exposure to temperatures between 0 and 40°C using A) linear GDH, B) non-linear GDH and C) Thermodynamic models.

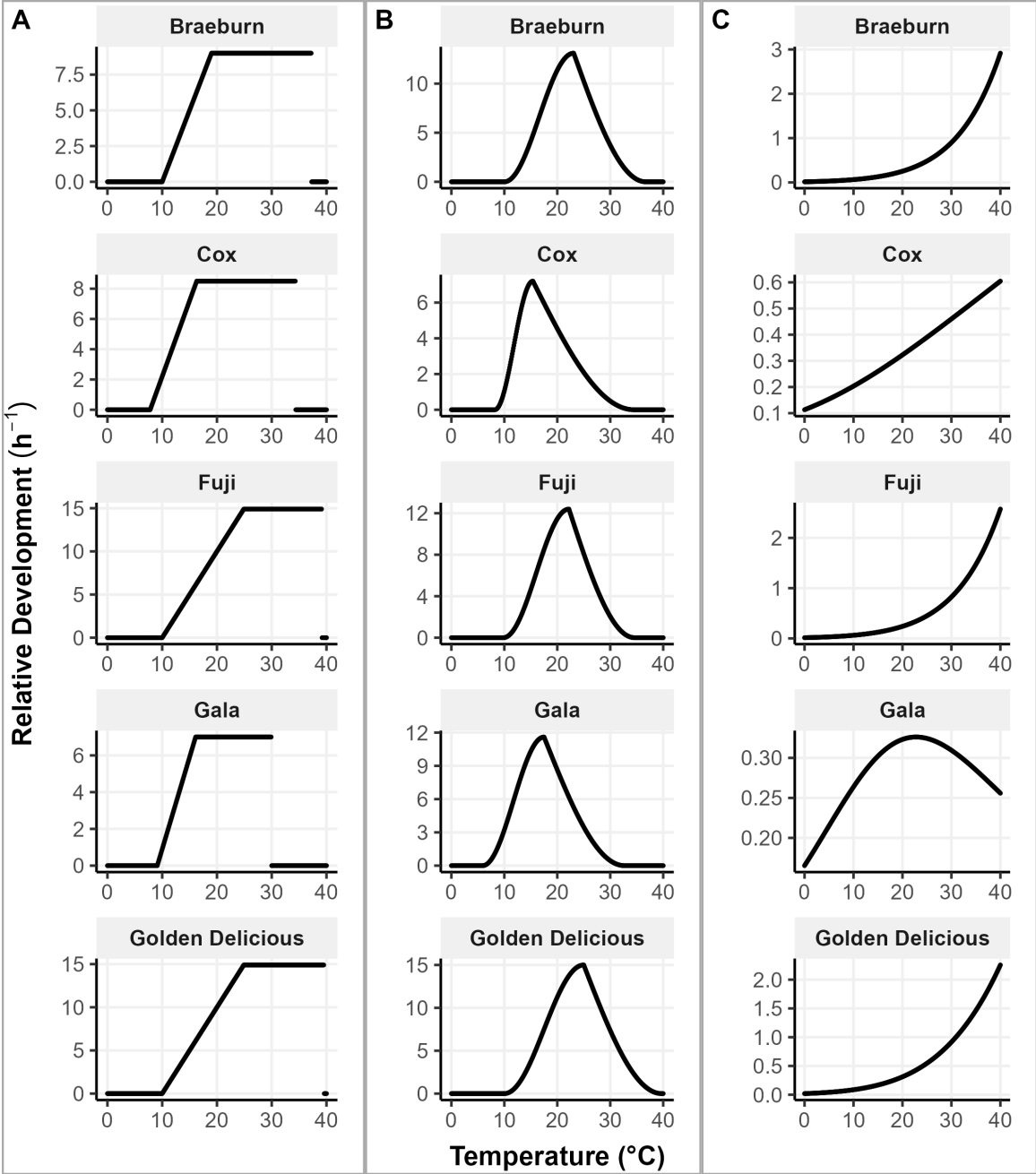


Table 4.3: Linear Growing Degree Hour model parameters estimated by the best correlation (Kendall's Tau) to Starch. Where multiple combinations result in the best correlation, the parameters presented are the closest to the median values. The errors represent the standard deviation of the best correlated parameters.

Cultivar	Best correlation	T_b	T_u	T_c
Braeburn	0.54	10.0 ± 0.00	19.0 ± 0.29	37.3 ± 1.86
Cox	0.41	7.8 ± 0.00	16.3 ± 0.00	34.4 ± 3.28
Fuji	0.22	10.0 ± 0.05	24.9 ± 0.12	39.2 ± 0.21
Gala	0.47	9.1 ± 0.11	16.1 ± 0.27	30.0 ± 0.63
Golden Delicious	0.33	10.0 ± 0.00	24.9 ± 0.06	39.6 ± 0.26

correlated is Fuji, with a weak negative correlation (Table 4.4). Fuji, Golden Delicious and Braeburn have similar temperature-growth rate relationship (Figure 4.3B). Their parameters vary slightly, with Fuji having lower temperature requirements and Golden Delicious being acceptive of high temperatures ($T_c = 39.7^\circ\text{C}$). The critical temperature for Gala is the lowest of the five cultivars. Gala and Cox have low base temperature values (5.9 and 8.2°C , respectively).

Table 4.4: Non-linear Growing Degree Hour model parameters estimated by the best correlation (Kendall's Tau) to Starch. Where multiple combinations result in the best correlation, the parameters presented are the closest to the median values. The errors represent the standard deviation of the best correlated parameters.

Cultivar	Best correlation	T_b	T_u	T_c
Braeburn	0.54	10.0 ± 0.00	23.1 ± 0.28	36.7 ± 2.32
Cox	0.41	8.2 ± 0.28	15.4 ± 0.28	34.5 ± 1.71
Fuji	-0.17	9.8 ± 0.10	22.2 ± 0.49	34.6 ± 2.85
Gala	0.47	5.9 ± 0.53	17.5 ± 0.77	32.7 ± 1.66
Golden Delicious	0.30	10.0 ± 0.00	25.0 ± 0.00	39.7 ± 0.00

The temperature-growth rate relationship as modelled by the Thermodynamic model for Braeburn, Fuji and Golden Delicious are similar (Table 4.5 and Figure 4.3C). They all follow an exponential pattern, suggesting their maximum growth rate has not been reached at 40°C . Similarly, Cox has also not reached its maximum growth rate 40°C ,

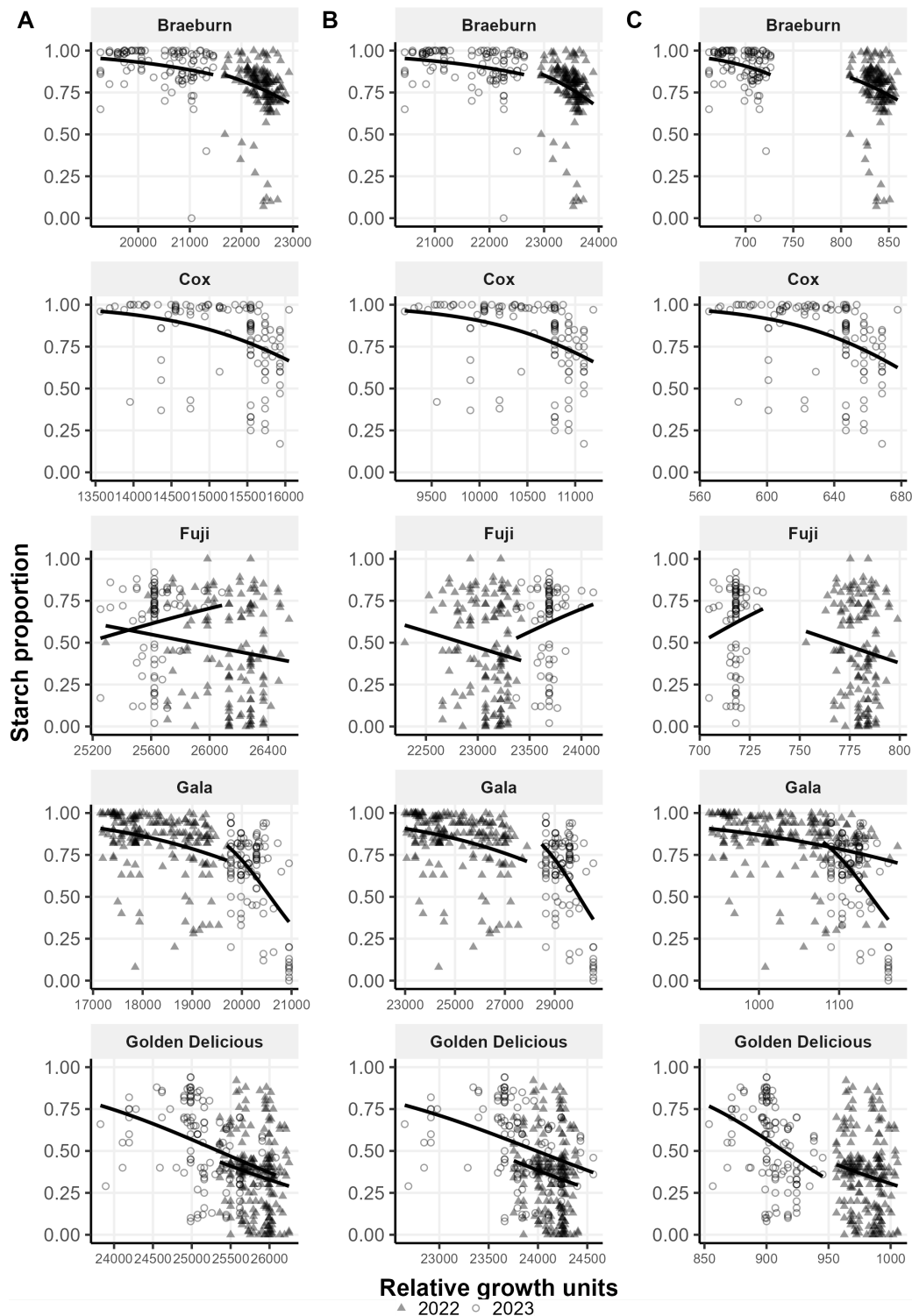
but its growth rate appears to be almost linearly related to temperature. In contrast, the maximum growth rate for Gala is at 22°C. The estimated relative growth rate at 22°C were similar for all cultivars.

Table 4.5: Thermodynamic model parameters estimated by the best correlation (Kendall’s Tau) to Starch. Where multiple combinations result in the best correlation, the parameters presented are the closest to the median values. The errors represent the standard deviation of the best correlated parameters.

Cultivar	Best correlation	B	C	TH
Braeburn	0.53	40.0 ± 0.19	7.9 ± 1.03	293.6 ± 3.34
Cox	0.41	18.8 ± 2.18	17.1 ± 5.06	290.0 ± 0.00
Fuji	0.23	39.8 ± 2.50	9.1 ± 5.30	290.0 ± 0.32
Gala	0.37	16.4 ± 0.52	29.9 ± 0.11	290.7 ± 0.93
Golden Delicious	0.34	40.0 ± 0.00	20.9 ± 0.00	300.0 ± 0.00

As expected, there is a negative correlation between the proportion of starch and accumulated growth unit for all three growth models —linear GDH (Figure 4.4A), non-linear GDH (Figure 4.4B) and Thermodynamic (Figure 4.4C) models. For Braeburn, Cox and Gala, the relationship follows a logistic shape as starch proportion does not change significantly in the early stages of development, but rapidly degrades after accumulation of certain growth units (Figure 4.4). The difference in the trend between years is consistent with the observed relationship of temperature accumulation with maturity: harvesting fruit appeared to be too early in 2023. For Golden Delicious, proportion of starch appears to decrease linearly with increasing accumulated growth units for all three models. The correlation between maturity and relative growth rates for Fuji apples were consistently the lowest (Tables 4.3-4.5), this is reflected by the weak trends observed for Fuji (Figure 4.4). The calculated accumulated growth units were higher in 2023 than in 2022 for linear GDH and Thermodynamic models, but the opposite was true for the non-linear model (Figure 4.4). This change in the accumulated growth units does not occur in the other four cultivars.

Figure 4.4: Proportion of starch (1 – immature and 0 – mature) against relative growth units calculated using A) linear Growing Degree Hours, B) non-linear Growing Degrees Hours and C) Thermodynamic model for each apple cultivar across 2 years. The trend lines show the trends of the values from 2022 and 2023 for each cultivar on each model



4.3.4 Factors contributing to maturity variation

Table 4.6 shows the summary of deviance in the fruit maturity (SPI) attributable to individual factors for individual cultivars. Comparing the deviance explained by the accumulated growth units across the three temperature growth models, the linear GDH model is the most effective for Gala (14.71%), the non-linear GDH model suits Braeburn, explaining 3.37%, and the Thermodynamic model works best for Cox (19.68%) and Golden Delicious (6.76%). However, it should be noted that the deviance attributable to the accumulated growth units was very similar among the three models (Table 4.6). The effect of accumulated growth units is not always statistically significant; only the linear GDH model for Braeburn (2.88%) and Gala (14.71%), non-linear GDH model for Braeburn (3.37%) and Thermodynamic model for Golden Delicious (6.76%) were statistically significant. For Fuji, < 1% of deviance in proportion of starch was explained by accumulated growth units (Table 4.6).

Some of the differences between the two seasons are expected to be accounted for by the accumulated growth units. The year effect did not contribute much to the deviance in proportion of starch for Braeburn or Gala, but it did affect proportion of starch significantly for Fuji with the linear GDH model (2.17%) and Golden Delicious with the non-linear GDH (1.75%) and Thermodynamic models (1.84%).

Differences between individual trees did not significantly affect proportion of starch for Braeburn, Fuji and Gala. In contrast, for Cox's Orange Pippin, tree effects were highly significant for all growth models, contributing 6.71%, 6.75% and 7.55% of deviance in the linear GDH, non-linear GDH and Thermodynamic models, respectively. For Golden Delicious, tree effects were significant for the linear and non-linear GDH models (Table 4.6).

The regions within the canopy contributed to less than 6% of the deviance in the observed proportion of starch, none of which was statistically significant.

Table 4.6: The percentage of the total variation explained by the linear GDH, non-linear GDH and Thermodynamic models for all cultivars. The values represent the percentage explained by each variable. Chi-squared test was used to determine the significance of each variable.

Terms	Linear GDH	Non-linear GDH	Thermodynamic
Braeburn			
Relative growth units	2.88 ***	3.37 ***	2.18
Year	1.07	0.84	0.05
Tree	0.65	0.83	0.46
Region	1.95	1.93	2.10
Cox's Orange Pippin			
Relative growth units	18.44	18.70	19.68
Tree	6.71 ***	6.78 ***	7.55 ***
Region	4.82	4.88	5.61
Fuji			
Relative growth units	0.52	0.61	0.29
Year	2.17 ***	6.40	0.03
Tree	0.58	0.61	0.54
Region	2.90	2.86	3.16
Gala			
Relative growth units	14.71 ***	13.42	12.18
Year	0.64	1.53	4.88
Tree	0.33	0.49	0.66
Region	3.57	3.35	3.06
Golden Delicious			
Relative growth units	6.15	5.86	6.76 ***
Year	0.02	1.75 ***	1.84 ***
Tree	10.86 ***	10.61 ***	11.50
Region	2.09	1.99	2.30
The significance codes denote the p-value thresholds.			
*** $p < 0.001$			
** $p < 0.01$			
* $p < 0.05$			
. $p < 0.1$			

4.4 Discussion

4.4.1 Harvest predictions and flowering effect

Flowering can occur in a quick burst or a longer span of time depending on the temperatures in the season. Harvest prediction models commonly use the average flowering date to predict a single harvest date. The accuracy of calendar date models will decrease as the climate changes. This is because plants can make small changes in their behaviour, such as changing their flowering patterns to avoid harsh temperatures. Only their offspring (with selective breeding) may become more tolerant to extreme temperatures, but commercial apple cultivars are based on clonal propagation. Thus, although calendar date models may have been accurate, but its effectiveness will fade. Moreover, due to only using the average harvest date, it ignores tree-to-tree and within-tree fruit-to-fruit flowering time. The present research assessed whether the within-orchard variation in flowering time can impact fruit maturity at harvest and hence predicted harvest dates. The present research showed that within-orchard flowering time accounted for 2-20% of the variability in fruit maturity (as represented by proportion of starch in fruit), depending on cultivars.

In the present study, different apple cultivars had flowering differences of about 10 days, with the within-cultivar variation of about 10 to 22 days. This difference in within-cultivar flowering time led to differences in accumulated growth units among individual fruit. Since temperature was usually much lower around the flowering time, the impact of this difference is thus expected to depend on the temperature-based growth relationship. The variation due to flowering time was observed to be less than 5% for Braeburn, Fuji and Golden Delicious, but can be much higher in Cox's Orange Pippin (18.4-19.7%) and Gala (12.2-14.7%). Since the difference in flowering time is relatively small even when temperature was usually low in the spring this shows that incorporating fine-resolution flowering time may improve the accuracies in predicting

the harvest window.

An unexpected linear trend was observed only for Golden Delicious. This could be due to a lack of immature Golden Delicious (namely with close to 100% starch), so only the development stage when the proportion of starch is reducing at a linear rate can be observed. Of course, unlike the other cultivars, Golden Delicious may indeed lose starch at a linear rate with accumulated growth units. Fuji consistently resulted in low correlation between starch and maturity. Fuji is deduced to mature differently than the other cultivars, and that Fuji may not depend on temperature as much as other apple cultivars. Another possibility for the lack of temperature relationship for Fuji is the low variability in flowering time, particularly in 2023 with zero interquartile range. The loss of one tree in an already limited number of biological replicates may have compromised the statistical power in our study. Overall, this study should be replicated with a greater number of trees per cultivar to ensure that the results are accurate and reproducible.

4.4.2 Year, tree, region and cultivar effects on maturity variation

Year, tree and canopy region accounted for a small proportion of the observed variability in the proportion of starch after the accumulated growth unit was included in the model. For the year factor, there still appears to be some significant effects, depending on the cultivar. The effect of year on Fuji, with the linear GDH model, and Golden Delicious, with the non-linear GDH and Thermodynamic models were significant, albeit accounting for $< 2\%$ of the deviance. It is assumed that this year-to-year effect could be due to differences in the solar radiation intensity since the temperature effects were accounted for by the accumulated growth units. Further data on multiple years across multiple locations with large differences in solar radiation are required to assess the potential effect of solar radiation on fruit development. A better understanding might

be gained by using a commercial orchard as these apple trees are not commercially thinned each year. The consequences of not thinning fruit include non-uniform fruit, and biennial cropping: less fruit after a season of heavy crop [99].

The effect of canopy region was insignificant for all cultivars. This result is inconsistent with previous findings [73–75, 107]. Most likely, this difference is due to the fact that our trees are small relative to conventional orchard trees; thus, there was no significant difference in shading between fruit in different zones of the same tree. In the present research, the proportion of starch was used as the maturity indicator; but previous studies focused on soluble sugars and firmness. Different maturity indicators may thus also account for the differences in the canopy effects.

It was surprising to observe the significant effects of individual neighboring trees on proportion of starch for Cox and Golden Delicious. As the trees were in proximity and exposed to the same biological and environmental conditions, it is difficult to explain such significant effects. One possible explanation could be that the number of fruit varied greatly among trees (hence possibly more variability in fruit development), which may affect fruit development in the same season and the following season (as these trees were not thinned) [99].

4.4.3 Evaluation of linear GDH, non-linear GDH and Thermodynamic models

In the present study, complicated optimisation algorithms were not used to estimate model parameters. Instead, a simple grid search approach was used to search a set of parameters that maximised the correlation of the estimated accumulated growth units and the observed proportion of starch. This simple approach will not be able to resolve non-converging issues often encountered in fitting complex nonlinear models. This approach was selected for two reasons. Firstly, nearly all parameters are related to minimum, optimal and maximum temperatures for fruit development. For these

parameters, there is a well-defined range based on biological intuition. Secondly, the present study focuses on the relative effect of flowering time (as represented by temperature effects) and other factors on fruit maturity, not the precise parameter values. By ensuring the parameter values maximised the correlation, the maximum effects of flowering time (temperature model) on fruit development within our defined search limits are expected to be captured. In our study, our maximum correlation was 0.54 for linear and non-linear GDH models on Braeburn, this is similar to values from a previous study by Sugiura et al. [52], where their absolute correlation of their linear model was 0.56 in field studies. This gives us confidence that our models are appropriate for our study.

We expected the optimal growth temperatures to be within 15-30°C, similar to values found in [113] and [44], but optimal growth rates have not been determined by the Thermodynamic model, even at 40°C, except for Gala (Figure 4.2C). Optimal temperatures above 40°C are unrealistic, and a detrimental effect was expected on most biological functions when temperatures exceed the realistic optimal temperature range [116]. Although optimal growth temperatures were not identified, since it is likely that temperature fluctuations during the growth season range from 10-30°C, the effective growth for each cultivar modelled by the Thermodynamic model is similar to the other two models. The temperature rates were more biologically sound for the linear and non-linear GDH. However, it should be noted that the linear GDH model is actually non-linear, consisting of two lines that join at the optimal temperature. There is no definitive rationale for preferring one GDH model over the other. The order of maximum relative development rates suggested by the non-linear GDH model follows the order of maturation of cultivars (Figure 4.2); Cox's Orange Pippin matures earliest in the season, followed by Braeburn, Gala, Golden Delicious and Fuji. This makes biological sense, as a faster rate of development suggests a shorter development period is required. Moreover, the non-linear GDH model tends to explain more variation

than the linear GDH model for most cultivars.

Commonly, when using the GDH model, the base temperature is set to 4.5°C, as per the original model [34]. The limits for the base temperature was explored between 0 to 10°C since flowering was observed in the spring when temperatures were usually less than 10°C and growth was not expected below 0°C. In the linear GDH model, the base temperatures were 7.8°C for Cox, 9.1°C for Gala, and 10°C for Braeburn, Fuji and Golden Delicious. The non-linear GDH model had base temperatures of 10°C for Braeburn and Golden Delicious, 8.2°C for Cox, 9.8°C for Fuji and 5.9°C for Gala. Growing degree day models and their variants used similar base temperatures at 10°C [101] or higher [51]. However, papers which fitted model parameters [44, 113] or tested a series of base temperatures [106] found lower base temperatures between 0 to 4°C than observed in our study.

4.4.4 Assumptions and limitations

This project was focused on conventional methods of fruit maturity methods, commonly used by growers. Therefore, the focus was on soluble solids, firmness and starch. The maturity parameter used in this study was the proportion of starch, a standard method for determining fruit maturity. Previous studies indicated that maturity cannot be represented by a single variable. Our findings showed a latent variable present in our Streif measurements (Figure 4.2). Therefore for this study, it must be assumed that starch proportion is an adequate measure of maturity despite it being a highly subjective assessment. Considering that each of the fitted models only explained up to 20% of the variability in fruit maturity, there is still a large amount of unexplained variance. It can be assumed that everything else is due to random variability. Physical assessments of fruit quality are therefore still required closer to the harvest window.

The authors acknowledge that flowering date and pollination date may not be the same. However, for the purpose of this study, it is assumed that flowers are pollinated

when the flowers are fully open. Apple flowers are typically insect-pollinated but pollinators may not always be present (in abundance) in the orchard, so it cannot be guaranteed that the flowers were pollinated on the bloom date. This would introduce a level of error in GDH calculations.

4.4.5 Implementations

Changes in the climate can affect the duration of flowering time, and therefore, the downstream harvest window. Growers can more accurately predict the harvest window, particularly the duration of the harvest window, when accounting for the variation in flowering time. This can reduce the noise for harvest window predictions. Knowing the duration of the harvest window will help orchard management during the harvest season. Our results show there is still a large amount of variation unexplained, therefore it is essential to assess the fruit close to the predicted harvest date, whether by destructive fruit assessments or by non-destructive methods.

4.5 Conclusion

The variation of maturity can be effectively quantified by either the linear or non-linear GDH models. The effect of flowering variation varies depending on the cultivar; the effect is small for Braeburn and Golden Delicious, but large for Cox's Orange Pippin and Gala. No flowering effect was found for Fuji. Considering these variations are induced within a 2-to-3-week period, flowering time is a significant contributor to maturity variation at harvest. Of the factors: year, region and tree specimen, only year and tree were significant for some cultivars but the effects were suspected to be due to non-uniform fruit cropping. Proper orchard management could help resolve differences.

Toward accurate prediction of apple firmness and Brix across countries, seasons and cultivars with hyperspectral imaging

5.1 Introduction

Traditional apple harvest prediction methods are essential for ensuring high-quality apples, but they are time-consuming, labour-intensive and rely on destructive sampling. Imaging has emerged as an efficient, non-destructive method to predict fruit maturity. Spectroscopy captures spectral wavelengths in a localised region, but previous research involving five surface scans per fruit across a large sample of apples did not yield reliable results for predicting maturity (unpublished). Hyperspectral imaging captures spectral data within an area, thus providing spectral and spatial information. Hyperspectral imaging has been successful for maturity trait predictions in apple [68, 70], sweet potato [117], tomato [118], cherry [119] and other fruits. While numerous studies have applied hyperspectral imaging to apples, the dataset in this study is notably broader, encompassing multiple cultivars, growing seasons, and countries.

Visible and near-infrared (VIS/NIR) spectroscopy has been a common area of study in determining fruit maturity qualities due to the obvious green-to-red (or yellow) colour change in the apple development process and the reflective soluble sugars at NIR.

Water is also highly absorptive in NIR [61], and can introduce complexity and noise into the data — particularly relevant given that water makes up approximately 80% of apple content [120]. Due to this noise, assigning wavelengths to specific functional groups (e.g. -OH or -COOH) is ineffective. Most studies identify significant spectral indices — ratios between specific wavelengths — that correlate with the presence and concentration of biological compounds or physiological traits. Some studies observe chlorophyll degradation using the difference in absorption ratio between 670 and 720 nm (IAD) [99, 121–123], and others look at the ratio of anthocyanin (R800/R678) or carotenoids (R550/R700) and chlorophyll [124]. IAD also shows some correlation with firmness [121], soluble solids [121] and starch degradation [122]. However, using IAD to predict maturity parameters may not be consistent year-to-year [122].

Other studies use multivariate statistical methods, which help extract informative wavebands from spectral data. Methods such as Partial Least Squares Regression (PLSR), Principal Component Analysis (PCA), Multiple Linear Regression (MLR) and their variants [68–70] are commonly used to address the high collinearity of spectral wavelengths and to model fruit maturity traits effectively. Wang et al., 2022 [70] found wavelength indices for chlorophyll, carbohydrate and water absorbance. The most significant wavelengths include the normalised difference spectral indices between 767 and 737 nm for soluble solids and 764 and 820 nm for firmness. On the other hand, Li et al., 2020 [69] identified 14 wavelengths in the infrared region (interesting spectra were found between 907 and 1107 nm) to define soluble solids (Brix).

More recently, hyperspectral imaging has been successfully analysed with deep learning models for learning complex patterns [71, 72]. The application of deep learning models to spectral data remains relatively new. When employed, these models often lack interpretability, offering limited insight into the importance of key spectral indices due to their black-box nature. Moreover, only a limited number of studies utilise large-scale dataset that exceeds 300 apples, despite the importance of large sample

sizes when training deep learning models [71, 125].

In the present study, we collected a large apple hyperspectral dataset comprising of 5756 apples, sourced from orchards in both the United Kingdom and New Zealand. Covering 6 cultivars across 3 growing seasons, this dataset provides a robust foundation for predicting key maturity traits in apples. We trained deep learning models with our data and found that: 1) the Vision transformer model performed the best out of four deep learning architectures (2D-Convolutional Neural Network (2D-CNN), 3D-Convolutional Neural Network (3D-CNN), and hybrid CNN-transformer model). 2) The models were only effective for predicting soluble solids and firmness, but not for starch. 3) The wavelengths important for making maturity predictions are in the visible region for both soluble solids (Brix) and firmness (Brix: 400-411, 453-465, 482-488, 538, 600-630, 648-668 and 671-692 nm, and firmness: 400-408, 633-698, 565-600) and one band at 1000 nm for firmness only. 4) The fruit region had no impact on soluble solids (Brix) predictions, but it was a significant factor in firmness predictions. 5) A single side of the fruit is insufficient for making accurate predictions due to increased bias. 6) Models showed that factoring cultivar into the training process improved model predictions. 7) Seasonal data showed specificity, but the model trained on all three seasons improved model performance.

5.2 Material and Methods

5.2.1 Fruit Samples

Fruit of six apple cultivars were collected from Kent, United Kingdom and Hawke's Bay and Nelson, New Zealand in three harvest seasons. Fruit were harvested during the harvest seasons from NZ (February to April) in 2023 and 2024, and from UK (September to October) in 2024. In England, Cox, Braeburn, Fuji, Gala and Golden Delicious were picked from East Malling (51°17'07.0"N 0°27'13.2"E) and additional

Gala, Braeburn, Cox and Jazz were sourced from orchards around Harbledown, UK (51°16'33.5"N 1°2'28.4"E). All cultivars, except Jazz, were picked from orchards at Plant and Food (PFR) Research Hawke's Bay (39°39'37.4"S 176°52'45.9"E). Additional Gala and Fuji were picked from a commercial orchard (39°36'11.5"S 176°49'20.9"E), Golden Delicious was picked from a local grower (39°36'54.2"S 176°50'52.1"E), and Braeburn and Fuji from PFR Nelson (41°06'49.7"S 172°59'04.3"E) in 2023 and 2024. All apples were picked in the morning, imaged, then the firmness, soluble solids and starch of each fruit were measured within 36 hours of harvest at ambient temperature. A total of 5756 apples were harvested (Table 5.1).

Table 5.1: Image and apple count per season split by cultivar. Numbers represent pre-filtering counts.

Cultivar	NZ2023		NZ2024		UK2024	
	Apples	Images	Apples	Images	Apples	Images
Braeburn	47	32	350	236	804	464
Cox	248	180	0	0	119	72
Fuji	811	472	524	360	280	160
Gala	463	356	424	292	850	492
Golden Delicious	132	104	111	76	574	328
Jazz	0	0	0	0	19	12
Total	1701	1144	1409	964	2646	1528

The equipment reported for each maturity feature was used, as follows, in the UK and NZ, respectively. Firmness was measured using a fruit texture analyser (Llyod LRX, UK and GÜSS, South Africa) with an 11 mm diameter probe to a depth of 8 mm. Two regions (approximately perpendicular to each other) along the equatorial region of the fruit were punctured after peeling away the skin. The force at maximum depth was used in this study. Atago refractometers (portable benchtop palette series, model PR-32 α and pocket PAL series, model PAL-1) were used to measure the apple juice collected from the puncture sites made during the firmness process. The refractometers were calibrated at the start of each sampling day using distilled water. Lastly, the

starch percentage was measured by cutting the apples in half horizontally then a potassium iodine solution was applied to one half of the apple by dipping them in a 1% w/v iodine and 4% w/v potassium iodide solution or spraying with 1% w/v potassium iodide and 0.25% w/v iodine. The apples were left to stain for 30 minutes (Per comms) before recording the staining percentage. The firmness readings from both sides of each fruit were averaged to create an average firmness reading.

5.2.2 Obtaining fruit images and initial image processing

Images were taken using two different Specim IQ hyperspectral cameras (Specim Imaging Ltd., Oulu, Finland) using the same settings [126]; one was in New Zealand and the other was in the United Kingdom. The spectral range of the cameras was 400-1000 nm, with 204 spectral bands at 7 nm resolution. Two 750 W tungsten halogen lights (ARRILITE 750 Plus, ARRI, Germany) were each positioned approximately 1 m apart, forming a triangular setup with the photo shooting tent. This setup ensured little shadow was cast on the apples. The tent diffused the light to avoid overexposure caused by direct light on the waxy layer of apples. The camera was set in front of the tent (Figure 5.1). The camera set-up and calibration were followed according to the manufacturer's manual. The reflectance was generated by correcting with white and dark references to reduce background noise. Images were captured following appropriate integration times.

In a dark room, four images of each apple were taken by rotating the apples by 90 degrees on the horizontal axis to attain four equatorial images. These were processed at room temperature before the maturity assessments. Depending on the size of the fruits, each image contained between 2 to 8 apples arranged in two rows. A total of 3636 images were taken (Table 5.1). The image metadata was inputted during the imaging process, ensuring the date, cultivar and fruit numbers were recorded. Some apples may not have images from all four sides due to human errors during the

collection process.

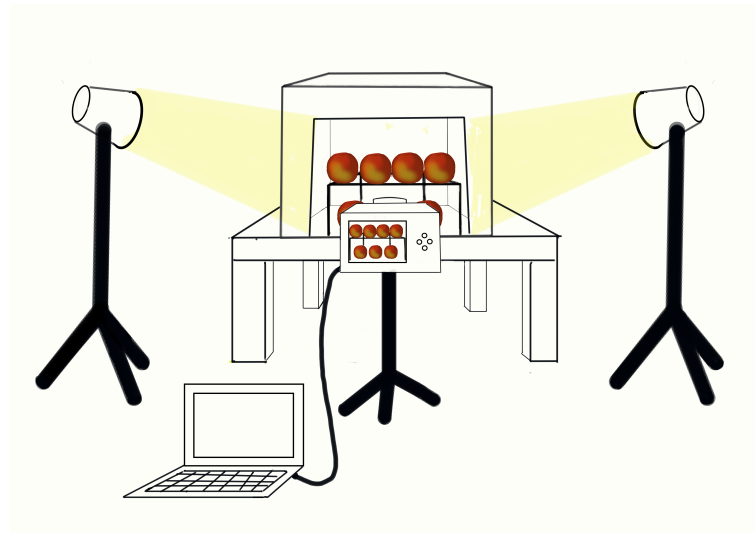


Figure 5.1: Schematic of the hyperspectral imaging equipment setup. Two lights on stands were set up on either side of a camera on a tripod, facing the light diffusion tent containing the apples. The lights and tent formed a triangle with sides of $\sim 1\text{m}$ length.

5.2.3 Image Processing

Since multiple apples were presented in the images, object detection was done using Segment Anything Model 2 (SAM2) [127] on the RGB images generated by the hyperspectral camera. Occasionally, SAM2 would not detect all the apples within the image. To ensure that we retained the correct number of actual apples and those detected by the image detection AI, we counted the number of bounding boxes detected and matched it to the count of fruit noted in the metadata. When the numbers did not match for that image, the image and the maturity data for those apples were removed from the final dataset. 2.75% of the images were filtered out due to mismatches between the detected and actual number of apples in the image. Valid masks were applied onto the hyperspectral images so that any pixel coordinates for any wavelength outside of the segmented area is zeroed.

The apple images within the bounding box coordinates (apple and zeroed background)

were divided into equal parts before locally averaging the pixels, forming either 30-by-30 or 50-by-50 pixel images. This method reduces the size of the tensors and retains spatial information. To further localise just the centre of the images, 5 pixels were removed from each edge of the generated images, thus forming 20-by-20 or 40-by-40 pixels images. Due to the curvature of apples and lower quality of pixel values around the edge of the fruit, removing the edges may improve model accuracy [70]. The spectral data, 204 contiguous bands, made up the third dimension of the image. In addition to the hyperspectral data, each sample was labelled with its corresponding cultivar information using one hot encoding so that it is trainable by machine learning algorithms. The six cultivars in the study were encoded into a binary format and integrated into the last layer of the images, increasing the third dimension from 204 to 210 channels, unless otherwise stated.

The data was first split into their respective cultivars before randomly splitting into training, test and validation datasets at 80, 18 and 2%, respectively. The different training, test and validation datasets for each cultivar were rejoined after the random split to ensure all cultivars were equally represented in all three datasets. The images were further processed to test whether specific regions of the apple were important in model predictions. The images were processed as above, then they were split in half vertically and horizontally before the quadrants were swapped diagonally (Figure 5.2). This caused the outer corners of the apples to be placed in the centre and vice versa.

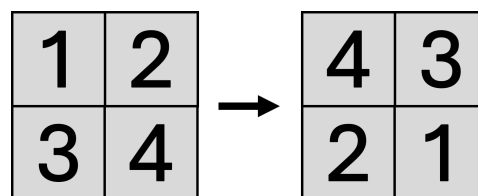


Figure 5.2: Example of a swapped quadrants image. The four corners were swapped diagonally to simulate whether apple regions are important.

5.2.4 Deep Learning Models

The selection of these four model architectures was intended to systematically compare the performance differences of various deep learning strategies in processing hyperspectral data for the values of soluble solids (Brix), firmness, and starch prediction. Specifically, 2D-Convolutional Neural Networks (CNN) excel at extracting local spatial features (Figure 5.3); 3D-CNNs can simultaneously capture joint features in both spatial and spectral domains (Figure 5.4); Hybrid CNN-Transformer models attempt to combine the local receptive field capabilities of CNNs with the global dependency modelling capabilities of Transformers (Figure 5.5); and Vision Transformer (ViT) models rely entirely on self-attention mechanisms for feature learning (Figure 5.6).

All models were implemented on the TensorFlow and Keras frameworks. The process utilised `tf.distribute.MirroredStrategy()` for efficient utilisation of available graphics processing units (GPU). Common training configurations included the use of Adam optimiser and Mean Absolute Error (MAE) as the primary evaluation metric. MAE intuitively reflects the average magnitude of deviation between predicted and true values. During the training process, MAE on the validation set also serves as a key metric for monitoring model performance, making early stopping decisions, and selecting the best model weights (saved via the `ModelCheckpoint` callback). Concurrently, loss function values (such as MSE or Huber loss) during training will also be recorded and analysed to aid in understanding the model's convergence behaviour. The `CSVLogger` recorded a detailed training history.

5.2.5 Hyperparameter Tuning

Bayesian optimisation is a method to optimise model hyperparameters [128]. Bayesian optimisation was used to determine the hyperparameters only for the ViT model, as it demonstrated the strongest baseline performance among the four architectures tested. A combination of the number of layers (1-5), patch size (2-10), projection dimension

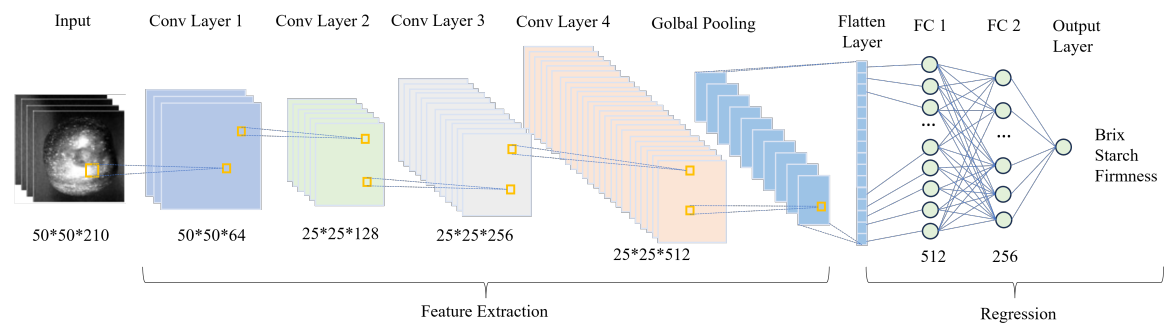


Figure 5.3: 2D-Convolutional Neural Network architecture.

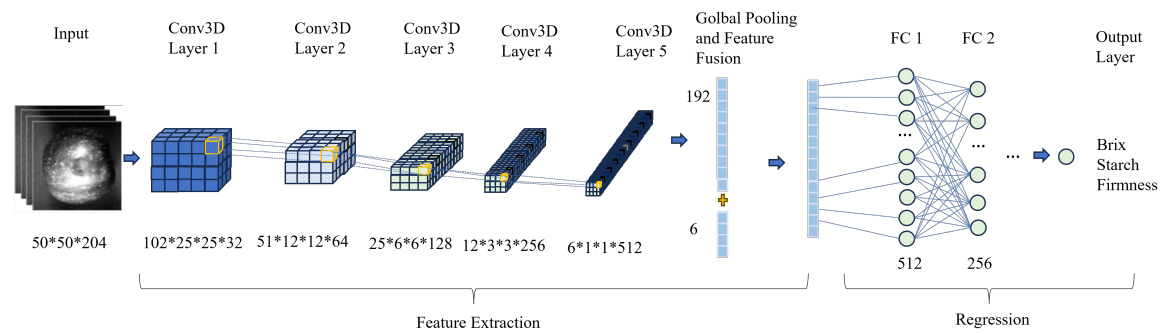


Figure 5.4: 3D-Convolutional Neural Network architecture.

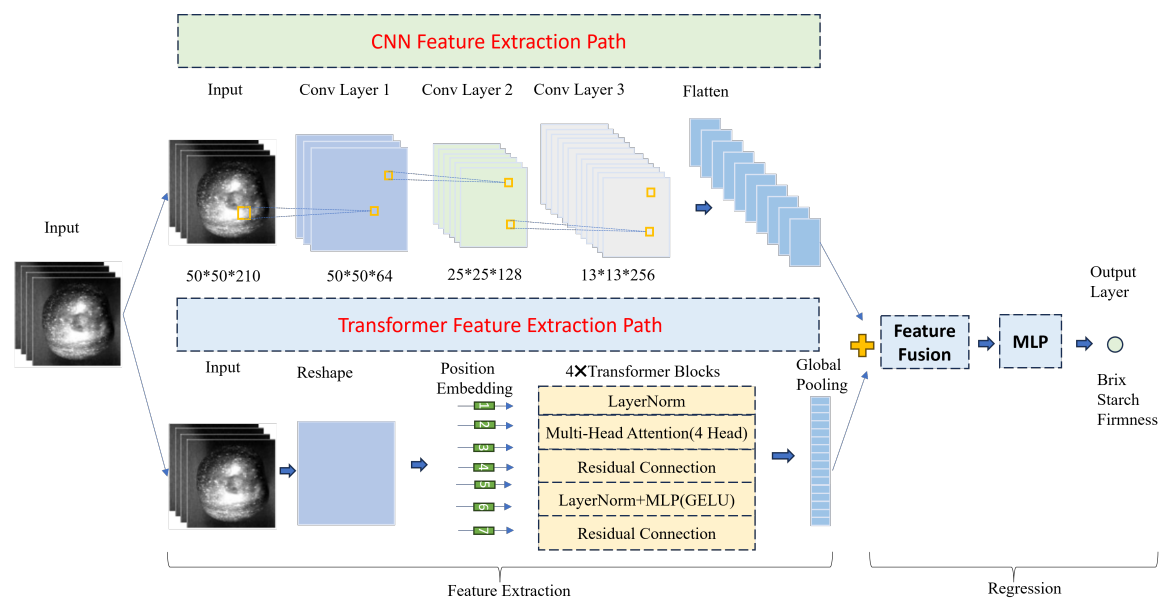


Figure 5.5: Hybrid model architecture, consisting of 2D-CNN and Transformer blocks.

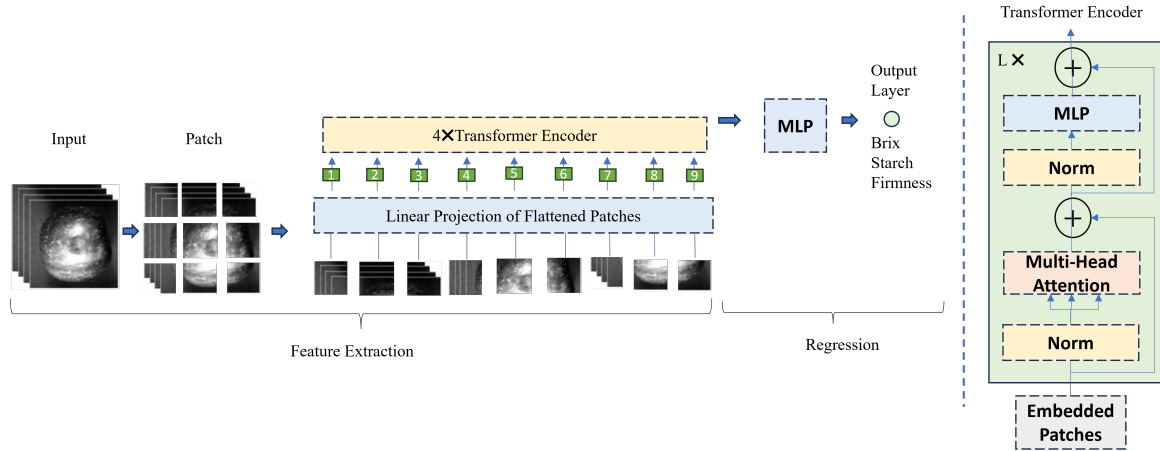


Figure 5.6: Vision Transformer model architecture.

(64-256), number of heads (1-5), multilayer perceptron (MLP) head units (small head: [128, 64], medium head: [256-128] or large head: [256-128-64]) and dropout rates (0.1-0.5) were tested. The larger the head, the more complex patterns can be learnt, but it may overfit and require more time to train. The final model was evaluated as follows.

5.2.6 Model Evaluation and Feature Selection

The three models return outputs for soluble solids (Brix), average firmness and starch percentage, independently. Root Mean Squared Error (RMSE) and the Coefficient of Determination (R^2) were used to determine the performance of the model. Evaluation of important wavebands, apple regions of interest and cultivar was done using Shapley values. Shapley value is based on coalition theory to fairly split a prize pool of money, based on their contribution. The greater their contribution, the greater their winnings. We calculate Shapley values for each feature on our final machine learning models to determine the contribution of each waveband to each maturity feature prediction, as well as the significance of each region of the apple and cultivars.

Floating Point Operations (FLOPs) were used to measure the complexity of models.

The lower the FLOPs value, the lower the computational cost. We used FLOPs to assess the difference between the Bayesian optimised models and the models trained on default parameters.

5.2.7 Hardware and software environment

Models were trained and evaluated with Python 3.10.13, using Keras (v2.15.0), Pytorch (v2.1.0), scikit-learn (v1.4.2), Numpy (v1.26.4), Pandas (v2.2.3), OpenCV (v4.8.1) and Matplotlib (v3.8.2). All ML models were trained on a NVIDIA A100 (80 GB), 503 GB RAM system.

5.2.8 Modelling strategies to assess model performance on maturity features

Several tests were conducted to get the optimised models for soluble solids, firmness and starch (Figure 5.7). We first tested between 2D-CNN, 3D-CNN, Hybrid or ViT models to identify the best performing model type. Secondly, different image sizes (30-by-30 or 50-by-50 pixels) and cropped (20-by-20 or 40-by-40 pixels) images were fed into the best model to determine the best input shape. These models were optimised using Bayesian optimisation functions to tune the hyperparameters of the models for each of the maturity features, independently. On the optimised models, we compared the effects of 1) removing the cultivar encoding, 2) just having one model for all three features and 3) reducing the input to just a single side of the apples (instead of using images from all sides of the fruits). We further tested 4) the effects of segregating the data into their seasons and 5) retaining a percentage of high impact channels and observing the effect on the models. Finally, 6) we tested how important each region of the apple is by splitting apple images into quadrants and swapping quadrants diagonally. All model RMSE and R^2 values were compared between the models.

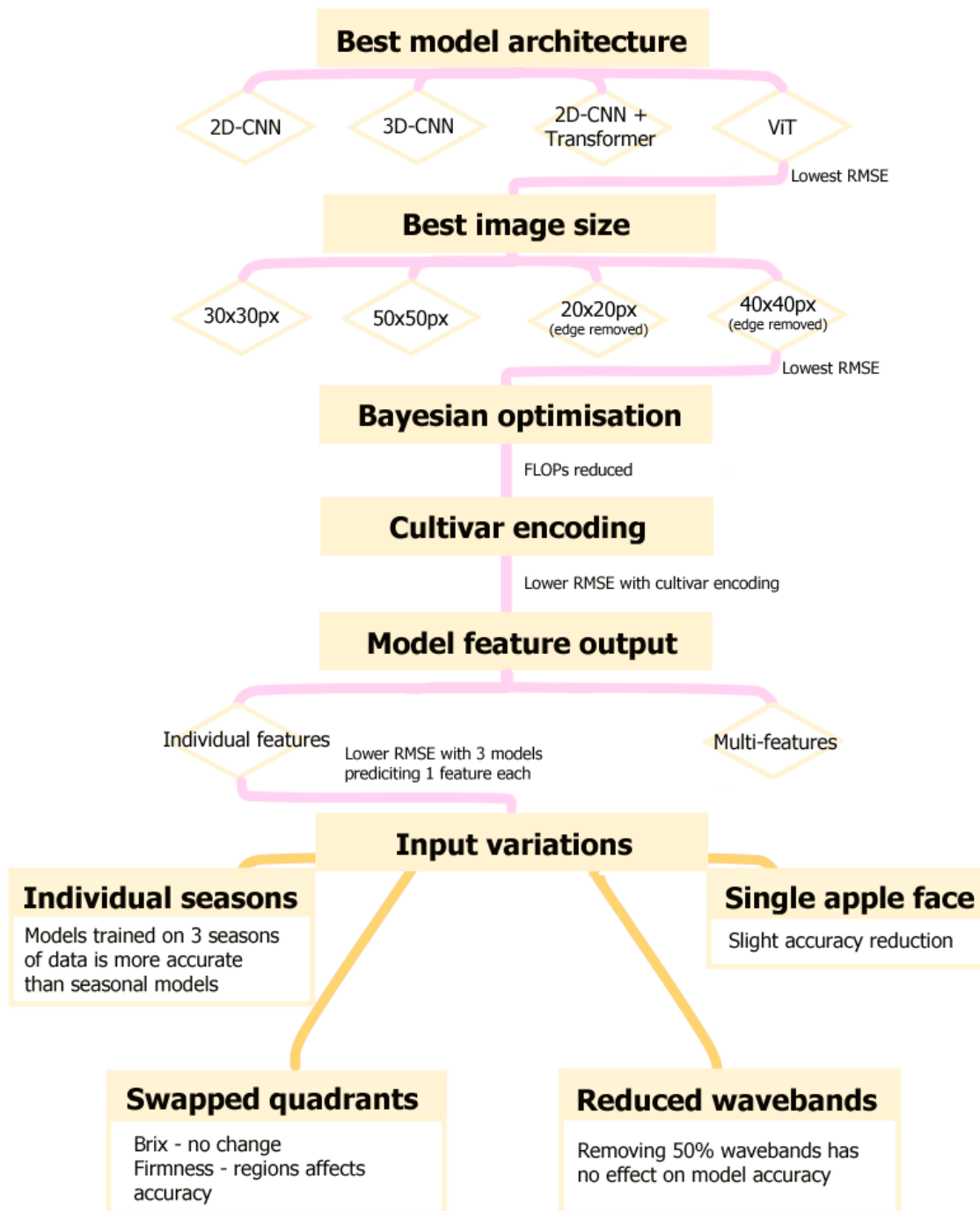


Figure 5.7: Model training and testing workflow. Solid rectangles represent research questions, while diamonds represent the options tested for each query. Pink lines depict the experimental pathway, with annotations along each line representing the rationale behind successive choices. The last four boxes, connected by the yellow lines, show the outcome of the varied inputs.

5.3 Results

5.3.1 Analysis of maturity parameters

Harvests were done over the commercial harvest period for six apple cultivars, and their starch, soluble solids and firmness measurements were recorded (Table 5.2). As expected, soluble solids and firmness are more similar within cultivars than between cultivars. Overall, the fruit with lower firmness was likely to have lower starch levels and vice versa, as seen by the slight positive trend (Figure 5.8).

Firmness is measured twice for each fruit. The overall standard deviation of firmness between sides is 0.46 ± 0.37 kgF, with a minimum of 0 and maximum of 2.63 kgF. The mean standard deviation for each cultivar, ordered from the largest to the smallest, is Jazz (0.56 ± 0.23), Cox (0.54 ± 0.39), Braeburn (0.49 ± 0.42), Fuji (0.48 ± 0.39), Gala (0.44 ± 0.33) and Golden Delicious (0.4 ± 0.35). Jazz was only harvested in the final season and had the least number of samples and therefore shows the least amount of deviance in soluble solids and starch values, but it was ranked the highest in firmness deviance (Table 5.2).

5.3.2 Analysis of Spectral Data

The main trends in spectra are similar between cultivars, namely peaks at 500-700 nm which are coloured pigments such as carotenoids, chlorophyll and anthocyanins; and a large peak at 700 nm which continues onwards to 1000 nm, the limit of the camera. The relative reflectance was distinct for each cultivar, indicating that cultivar types can be inferred through imaging, specifically between 800-900 nm (Figure 5.9). Blue light, between 400-480 nm, is where carotenoids and chlorophyll absorb light. This is consistent with our results as Golden Delicious has the highest reflectance in this region and Braeburn and Jazz had the least reflectance in this range. Golden Delicious shows a high reflectance peak around 530-640 nm, likely due to the low

Table 5.2: Quality distribution table of apples by cultivar from 2023-2024. The columns represent the mean and standard deviation of the mean for firmness, soluble solids(Brix) and starch. Firmness is the average firmness of the two sides of the fruit. The rows represent the six apple cultivars in this study. Each cultivar is calculated for each season separately and with all seasons together.

Cultivar	Season	Firmness (kgF)	Soluble Solids (°Brix)	Starch (%)
Braeburn	NZ2023	8.11 ± 0.63	12.41 ± 1.0	55.00 ± 21.59
	NZ2024	9.25 ± 1.01	11.97 ± 1.04	69.04 ± 22.06
	UK2024	9.72 ± 0.89	10.58 ± 0.82	83.01 ± 15.70
	All	9.51 ± 0.98	11.07 ± 1.13	77.73 ± 19.71
Cox's Orange Pippin	NZ2023	8.36 ± 1.36	13.02 ± 1.49	67.59 ± 22.25
	UK2024	8.06 ± 1.25	10.83 ± 0.95	62.18 ± 15.59
	All	8.26 ± 1.33	12.28 ± 1.69	65.76 ± 20.39
Fuji	NZ2023	7.45 ± 0.82	14.36 ± 1.54	32.64 ± 23.95
	NZ2024	7.39 ± 0.81	14.65 ± 1.41	49.07 ± 23.36
	UK2024	9.01 ± 0.87	12.32 ± 1.12	36.83 ± 18.26
	All	7.71 ± 1.03	14.09 ± 1.66	38.67 ± 23.99
Gala	NZ2023	7.78 ± 1.10	13.15 ± 1.35	39.10 ± 28.28
	NZ2024	7.81 ± 0.84	12.46 ± 1.23	44.36 ± 28.41
	UK2024	8.77 ± 0.81	11.05 ± 0.80	69.81 ± 22.06
	All	8.27 ± 1.02	11.96 ± 1.42	55.27 ± 29.18
Golden Delicious	NZ2023	7.58 ± 0.97	11.92 ± 0.72	84.27 ± 12.41
	NZ2024	7.59 ± 0.61	13.16 ± 0.94	53.06 ± 18.16
	UK2024	7.19 ± 0.68	11.62 ± 1.27	15.15 ± 14.62
	All	7.30 ± 0.75	11.88 ± 1.27	30.66 ± 29.83
Jazz	UK2024	11.35 ± 1.50	11.23 ± 0.97	95.16 ± 3.83

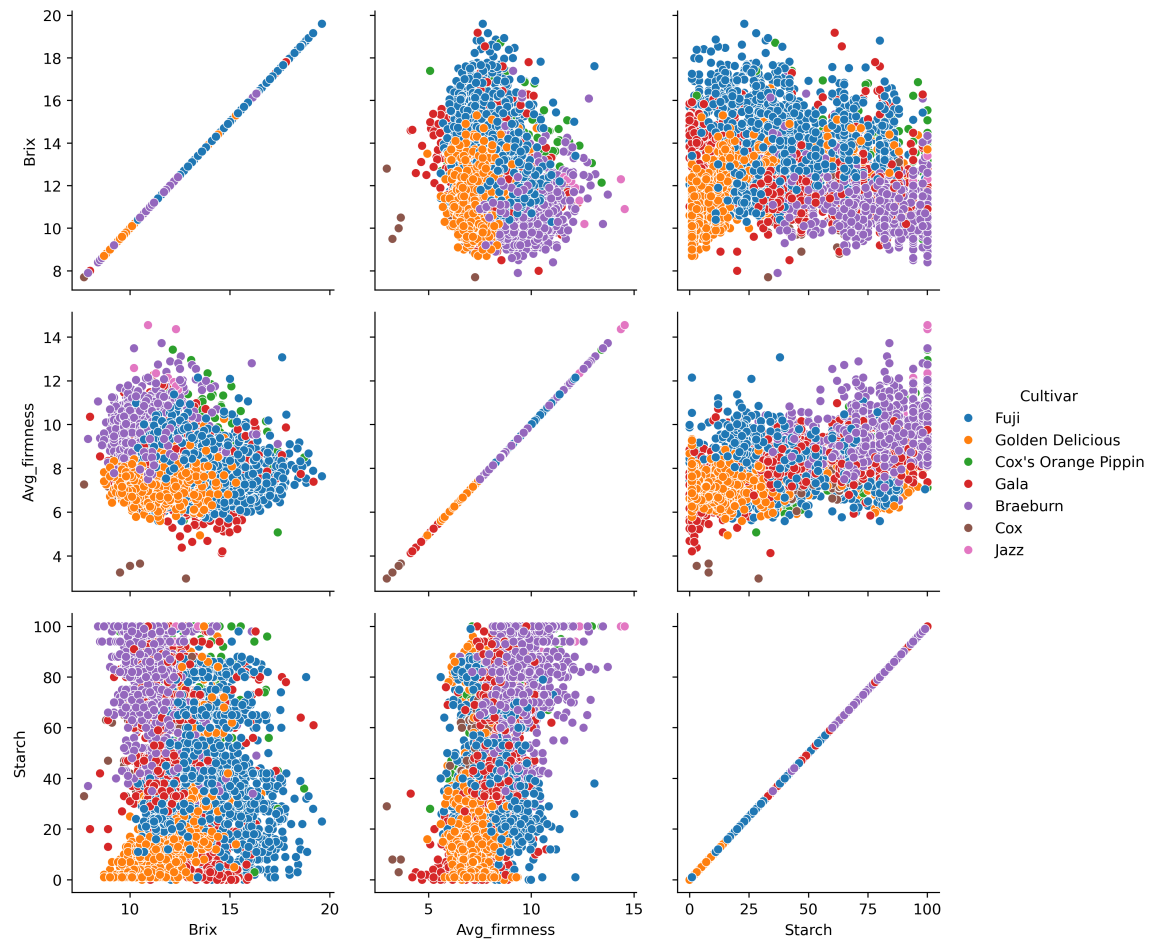


Figure 5.8: Pairwise plot between Brix, firmness and starch, coloured by each cultivar.

levels of carotenoid and anthocyanin present in the skin [124]. Gala likely has the lowest concentration of chlorophyll due to the higher reflectance at 650-670 nm [124]. Braeburn and Jazz contain the highest levels of anthocyanins in the peel, as evident by the low reflectance between 550 and 600 nm. Due to significant differences in wavelengths between cultivars, we encoded the cultivar information in the training process.

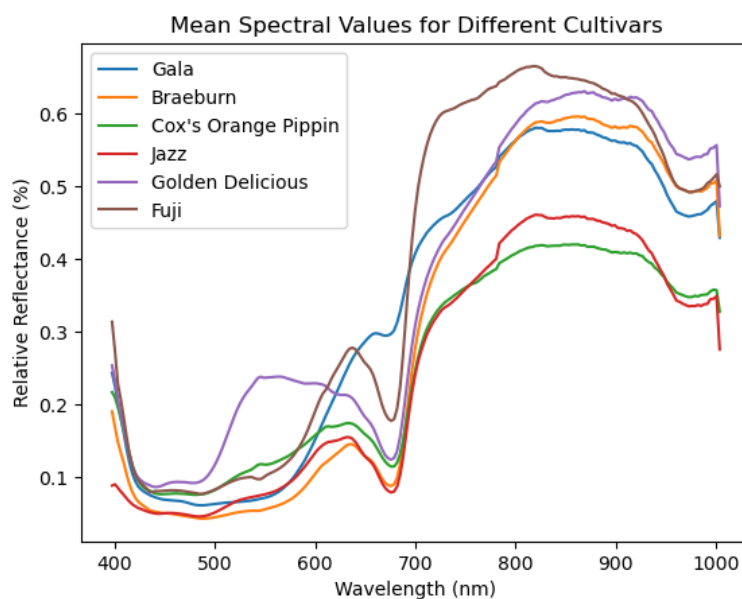


Figure 5.9: Mean reflectance of six apple cultivars between 400-1000 nm.

5.3.3 Model Training Analysis

This study tested which type of model was the best in predicting fruit maturity (Table 5.4; runs 1-4). 2D-CNN, 3D-CNN, Hybrid-Transformer and ViT models were tested on 50-by-50px images. The best performing model was the ViT model, consistently resulting in the lowest RMSE values for Brix, firmness and starch and the highest R^2 values. In comparison, the 2D-CNN, Hybrid and 3D-CNN models were not as accurate as the ViT model. Interestingly, the 3D-CNN model performed better in predicting firmness and starch compared to the Hybrid model, but the Hybrid model did better in predicting soluble solids (Brix). The 2D-model worked well for soluble solids (Brix),

ranking second after the ViT results. Firmness and starch were slightly worse than results from the 3D-CNN model but better than the Hybrid model. The second test determined the best size of the data input Table 5.4; runs 4-7). The images were aggregated into 30-by-30 pixel or 50-by-50 pixel size. The aggregating process would average across more pixels on larger apples such as Braeburn, compared to smaller apples such as Cox’s Orange Pippin due to the differences in the original apple sizes (Table 5.3). We further tested 20-by-20 pixel and 40-by-40 pixel images, which were created by removing 5 pixels from each edge of the original 30-by-30 or 50-by-50 pixel images, respectively. This causes the models to be trained only on the central part of the apple face, as the edges may have reduced signal quality [70]. The images sized 40-by-40 pixels, showed the best results for Brix with the edges removed, despite a slight reduction in model performance in firmness and starch. Between 20-by-20 and 40-by-40 pixels sized apples, soluble solids and firmness had a small improvement in accuracy, but firmness accuracy dropped by a small amount. Therefore, we selected 40-by-40 pixels images to train our subsequent models.

Table 5.3: Mean dimensions of apple cultivars in pixels

Cultivar	Mean Height (px)	Mean Width (px)
Braeburn	130	136
Cox’s Orange Pippin	118	122
Fuji	123	131
Gala	124	132
Golden Delicious	125	132
Jazz	120	119

A Bayesian optimisation function was used to find the optimal hyperparameters of the models (Table 5.5). This changed the number of convolutional blocks (each containing Conv2d, normalisation, max pooling and dropout layers) from 4 blocks to 5 for soluble solids (Brix) and 3 for firmness, and the number of blocks for starch was unchanged. The patch sizes were increased after the optimisation for all features, and the number

Table 5.4: Model summary results for Brix, Firmness, and Starch

Run	Name	Image size	Model	Result	Soluble solids (Brix)		Firmness		Starch	
					RMSE	R^2	RMSE	R^2	RMSE	R^2
1	Model type	50px	2D-CNN	Runs 1–4: Vision transformer model outperformed the other model types	1.14	0.61	0.86	0.53	19.89	0.57
2	Model type	50px	3D-CNN		1.97	-0.16	1.15	0.15	20.46	0.54
3	Model type	50px	Hybrid		1.33	0.47	1.78	-1.03	34.70	-0.31
4	Model type	50px	ViT	Runs 4–7: Results are slightly better with edges removed (either 40 or 20 px images). 40px chosen for subsequent processes	0.94	0.74	0.74	0.65	17.33	0.67
5	Image size	40px	ViT		0.92	0.75	0.76	0.63	17.37	0.67
6	Image size	30px	ViT		0.98	0.71	0.75	0.64	17.79	0.66
7	Image size	20px	ViT		0.95	0.73	0.76	0.63	17.11	0.68
8	Optimised hyperparameters	40px	ViT	Run 5 vs 8: Optimised hyperparameters with Bayesian optimisation function. Results improved or unchanged for all features	0.91	0.75	0.76	0.63	16.79	0.69
9	Without cultivar	40px	ViT	Run 9 vs 8: Embedded cultivar data improved model results for all features	0.97	0.72	0.79	0.60	17.53	0.67
10	All 3 features	40px	ViT	Run 10 vs 8: A model for each feature is more accurate than one model for all three features	1.08	0.65	0.92	0.46	17.13	0.68
11	Season - NZ2023	40px	ViT	Runs 11–13 vs 8: Trained seasons independently but results are not as accurate as all data model	2.11	-0.33	1.01	0.34	28.38	0.12
12	Season - NZ2024	40px	ViT		1.95	-0.14	1.08	0.25	28.87	0.09
13	Season - UK2024	40px	ViT		6.68	-12.37	1.95	-1.44	27.02	0.21
14	Top 20% wave-lengths	40px	ViT	Runs 14–16 vs 8: Retained the specified amount of channels and retrained the models. Retaining 50% of the data was as good as modelling with all channels	0.94	0.73	0.76	0.63	18.49	0.63
15	Top 50% wave-lengths	40px	ViT		0.89	0.76	0.76	0.63	17.31	0.67
16	Top 80% wave-lengths	40px	ViT		0.89	0.76	0.76	0.63	17.07	0.68
17	Swapped corners diagonally	40px	ViT	Run 17 vs 8: Swapped images reduced model accuracy for soluble solids (Brix) and starch but did not affect firmness	1.00	0.70	0.76	0.63	18.75	0.62
18	Single side	40px	ViT	Run 18 vs 8: Trained using one side of the apples	0.96	0.72	0.83	0.55	19.24	0.60

of heads decreased from the default model. The projection dimension increased for Brix but decreased for firmness and starch. The dropout rates also increased for firmness and starch but remained the same for Brix. In terms of MLP heads, Brix and firmness had less layers and similar number of neurons. The MLP head layers and neurons were unchanged for starch. The optimised hyperparameters resulted in models with similar performance for Brix and firmness between runs 5 and 8 (Table 5.4). Starch results showed significant improvement in RMSE (17.37 vs 16.79) and R^2 (0.67 vs 0.69) between run 5 and 8, respectively. Although no large improvements in RMSE or R^2 was observed, all optimised models reduced the number of FLOPs required to make model predictions (Table 5.6). The plot of predicted vs actual values of Brix and firmness show good predictions close to the $x=y$ line, evenly spread across the range with equal deviance on both sides of the line (Figure 5.10). For starch, the points follow a vague positive trend, but it has high deviation, especially around the centre of the plot (between 20 – 80%).

Table 5.5: Bayesian Optimisation results for Brix, firmness and starch models. The default represents the parameters used prior to optimisation.

Model	Transformer Layers	Patch Size	Projection Dim.	Head Number	MLP Head Units	Dropout
Default	4	5	128	8	[256, 128, 64]	0.1
Brix	5	8	256	3	[256, 128]	0.1
Firmness	3	10	64	5	[128, 64]	0.223
Starch	4	9	66	4	[256, 128, 64]	0.137

Table 5.6: Floating Point Operations (FLOPs) calculated for each model before and after Bayesian optimisation

Model	Brix	Firmness	Starch
Pre-optimised model	4.59×10^8	4.59×10^8	4.59×10^8
Optimised model	2.39×10^8	4.48×10^7	3.85×10^7

Encoding cultivar data into the last layer of the training data yielded improved results for all three features (Table 5.4; run 9 vs 8). We also determined if the performance

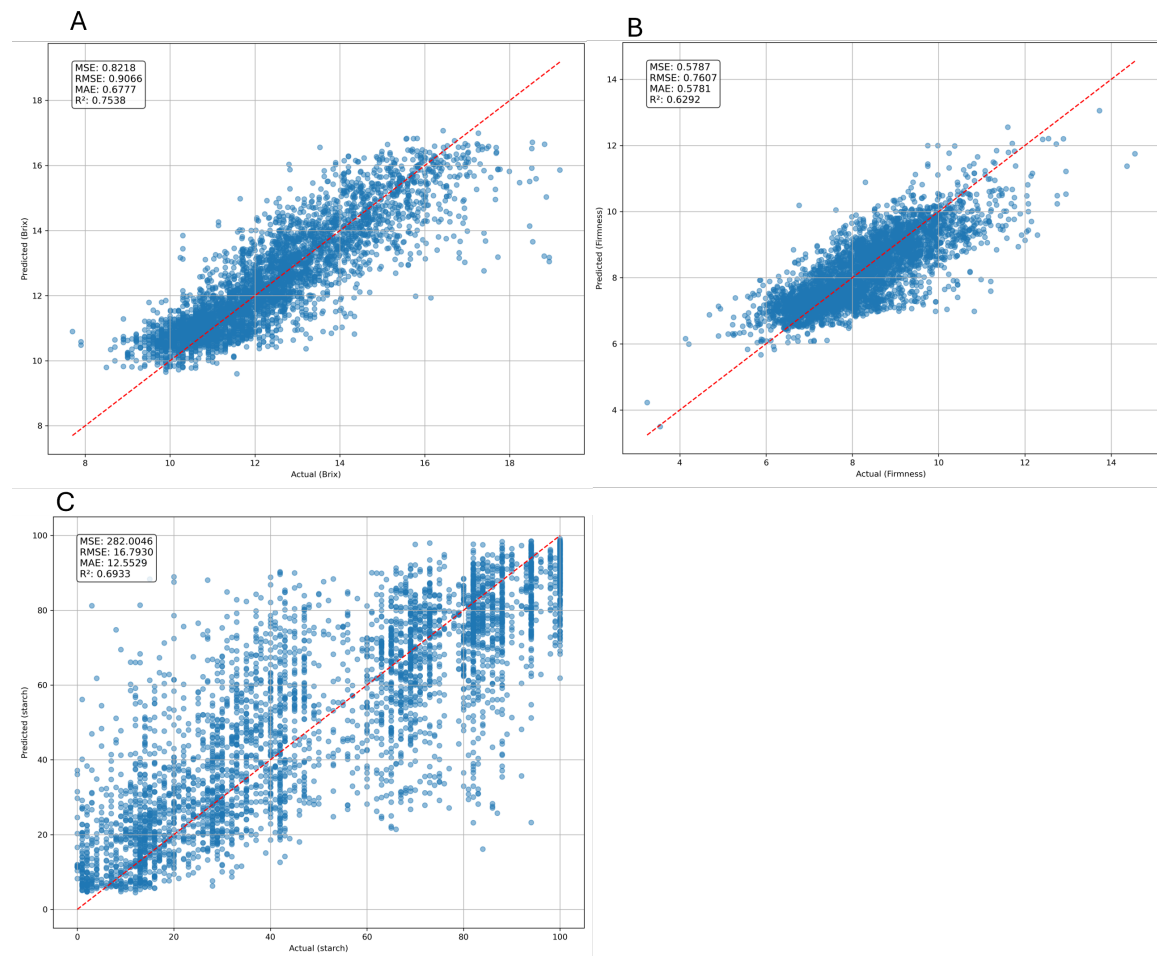


Figure 5.10: Predicted vs Actual values of A) Brix, B) Firmness and C) Starch on the optimised trained models. The red dashed line represents the $x=y$ line.

of a single model, predicting three features (Brix, firmness and starch), would be as accurate as three separate models predicting one feature each. We tested this feature on the 2D-CNN model in preliminary testing (results not shown) and with the ViT model. Model accuracy improved by training three separate models for Brix and firmness compared to a single model (Table 5.4; runs 8 and 10). There was a significant improvement in model performance with Brix and firmness. For Starch, there were no significant differences. Therefore, the subsequent models trained independent features.

5.3.4 Seasonal Comparisons

To see if the variation in model predictions is due to seasonal variation, the models are trained with each season independently (Table 5.4; run 11-13). We cross-validated the seasons with each other, computing each five times on five randomly generated test sample subsets and averaging the RMSE. Applying one season's model on another season, in most cases, will result in worse RMSE for soluble solids and firmness. The exception was the firmness model trained on NZ2024 data and tested on NZ2023 (RMSE = 1.17) data, compared to results on data from the same year (RMSE = 1.28) (Table 5.7). Overall, we observe less predictive accuracy and correlation between predictions and observed results when models are trained using single seasons for soluble solids and firmness. The results for starch always produced high RMSE and did not show seasonal specificity.

5.3.5 Single-sided Analysis

Finally, we tested whether using a single side of the apple would yield accurate results. The model was trained on one of the four sides of the imaging data and its corresponding maturity data. The models showed lower accuracy for all features (Figure 5.11). The tails of Brix and firmness plots start to flatten out, indicating some bias in the models. The shape of the firmness points is less defined and appears to

Table 5.7: Cross-validation of seasonal data on other seasons data. The values are the average RMSE computed across five randomly sampled subsets of the test dataset.

Model		Data		
		NZ2023	NZ2024	UK2024
	NZ2023	Brix		
		1.90	2.20	1.89
		2.18	1.98	2.50
		2.80	2.76	1.47
	NZ2024	Firmness		
		1.16	1.30	1.60
		1.17	1.28	1.62
		1.98	1.92	1.47
	UK2024	Starch		
		36.63	32.06	32.25
		32.80	26.51	38.55
		36.27	36.27	38.73

follow less of a positive trend. Starch remains just as variable as with all four sides of the apple data.

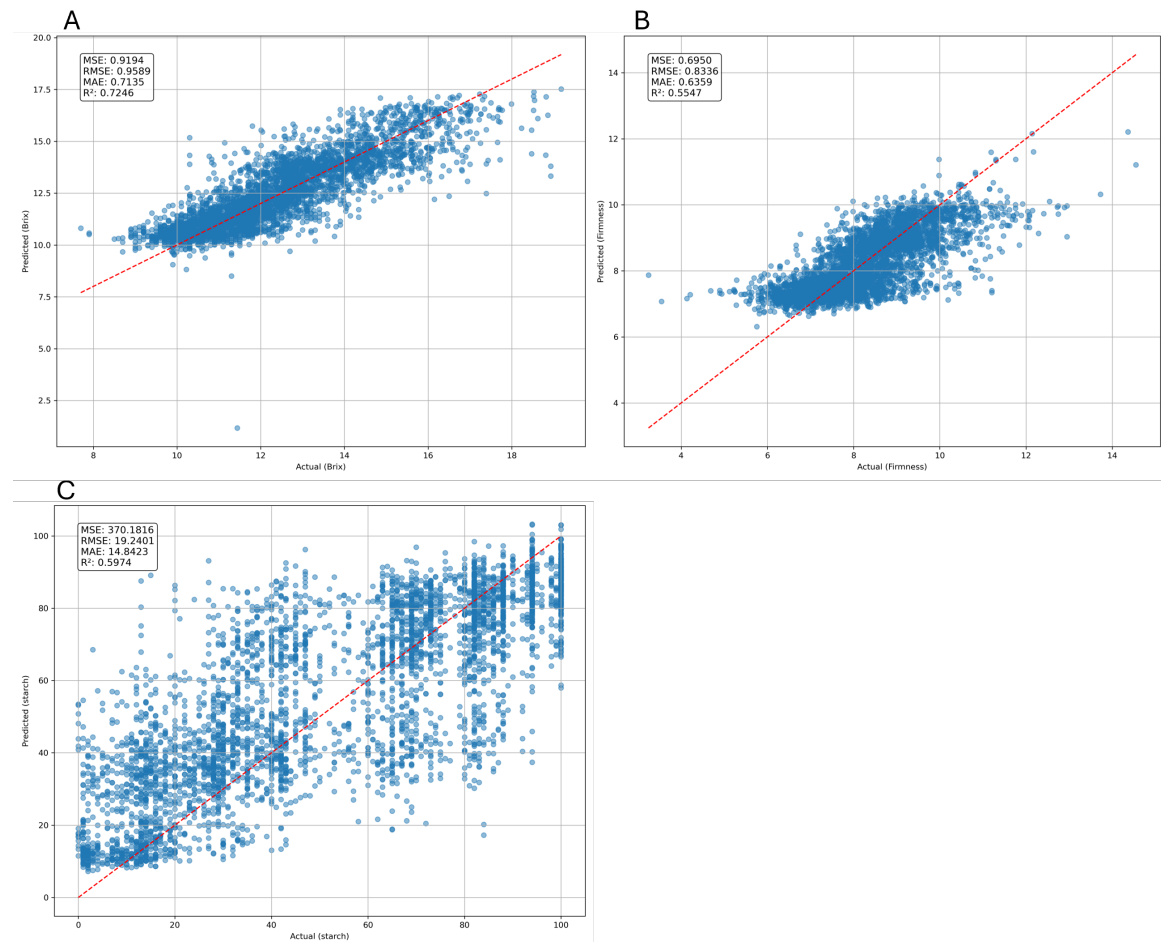


Figure 5.11: Prediction vs actual plot of A) Brix, B) Firmness and C) Starch models on a single side of apple data. The red dashed line represents the $x=y$ line.

5.3.6 Shapley Analysis

Hyperspectral data are highly correlated, making a large portion of the channels redundant. Shapley values indicate that the top 20% of the most important spectral channels were between 400-411, 453-465, 482-488, 538, 600-630, 648-668 and 671-692 nm (peaking at 400, 459, 488, 538, 618, 657 and 683 nm) for Brix (Figure 5.12A); 400-408, 633-698, 565-600 and 1000 nm (peaking at 400, 589, 651, 683, and 1000 nm) for firmness (Figure 5.12B); and 400-411, 547-598, 600-642, 671-692 nm (peaking at

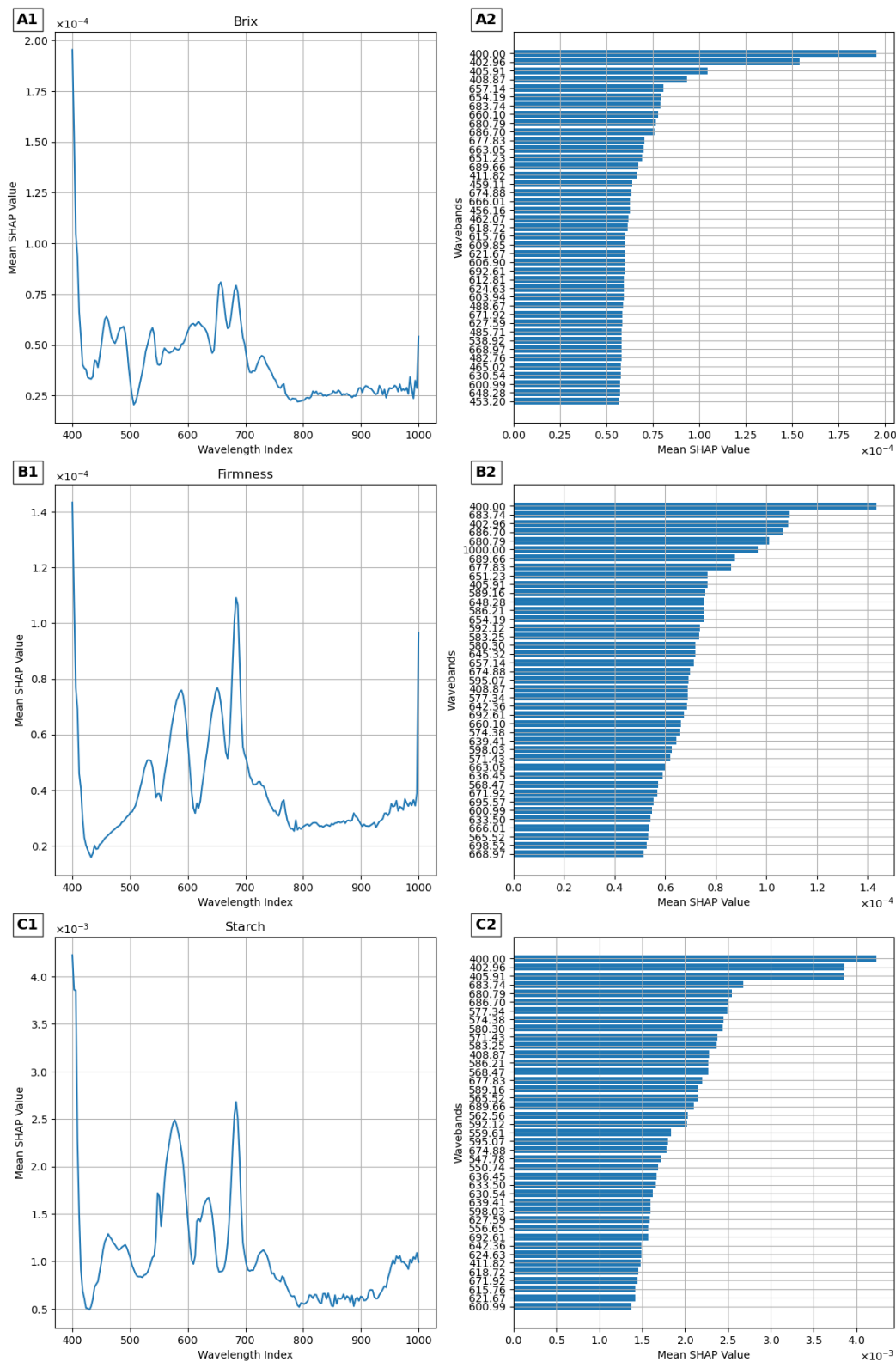


Figure 5.12: Mean spectral importance. 1) Mean Shapley values across wavebands and 2) the top 20% most significant wavebands for the A) Brix, B) firmness and C) starch models.

400, 547, 577, 636 and 683 nm) for starch (Figure 5.12C). The spectra between 800-950 nm had small Shap values, indicating a low level of importance of wavelengths in this region. Using Shap values as the level of importance, we kept 20%, 50% and 80% of the most important channels for each feature and retrained the models with the reduced channels. The results show that reducing the wavelength channels to the top 20% for firmness does not affect model accuracy (RMSE = 0.76 and $R^2 = 0.63$). For soluble solids (Brix), reducing the channel wavelengths to the top 50% slightly improves model accuracy (RMSE = 0.89 and $R^2 = 0.75$). For starch, even keeping the top 80% of channels causes a reduction in model accuracy (RMSE = 17.07 and $R^2 = 0.68$). Removing 50% of the wavelengths does not induce any noticeable bias in model predictions (Figure 5.13).

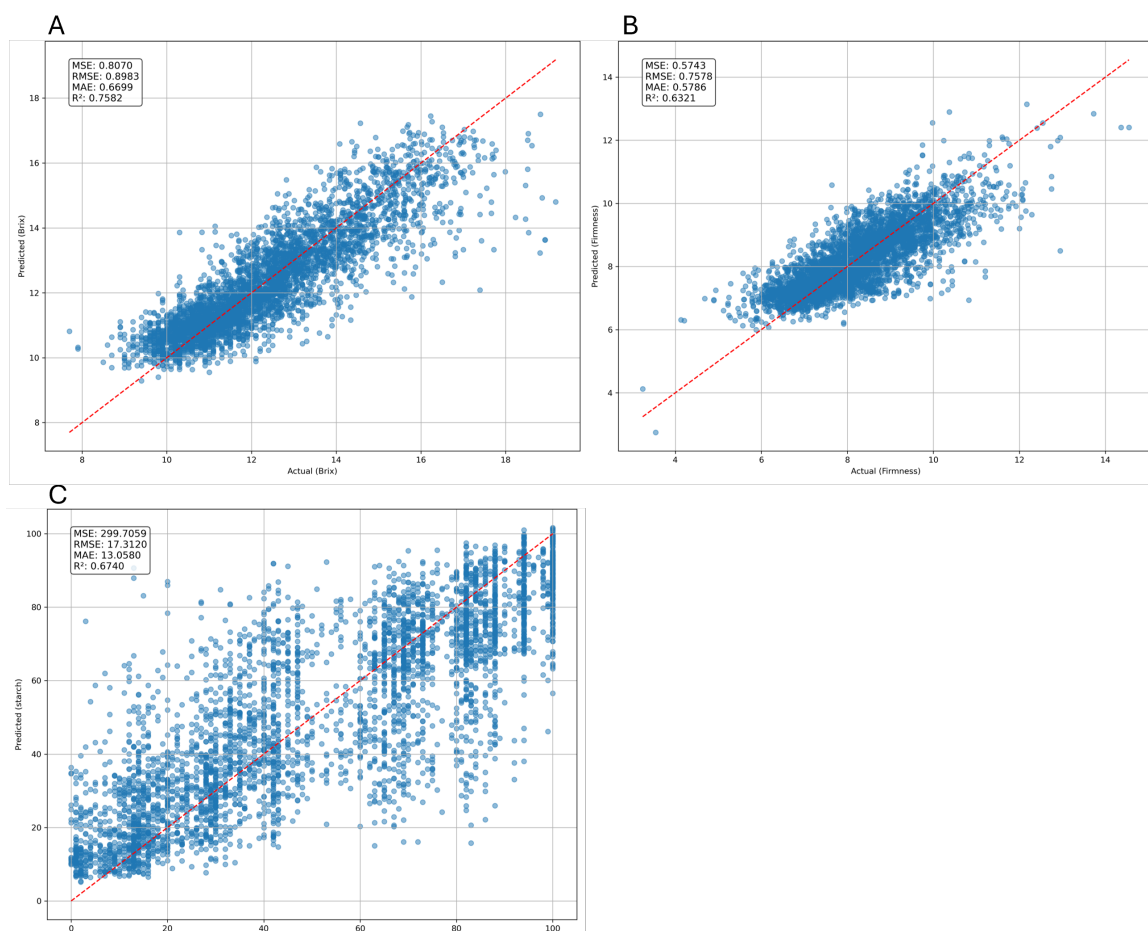


Figure 5.13: Predicted vs Actual plot of A) Brix, B) Firmness and C) Starch models trained with 50% reduced wavelengths. The red dashed line represents the x=y line.

Shapley values were used to evaluate the significance of patches and regions of the apples. The Shapley values indicate that the models put more emphasis on the outer regions of the apple images, in particular the top left corners for Brix and firmness, the bottom right corners for Brix and starch and the top right corner for firmness and starch. Less significance is seen the middle section of the apple images (Figure 5.14).

To test whether this region is significant or whether the models are more attentive to these regions of the data, we trained the model on images with quadrants swapped (Figure 5.2). The Shapley values showed very similar patterning in Brix, indicating that the model tends to look at the top left (first patch into the model) and the bottom right (last patch into the model) (Figure 5.15). The regions are insignificant for Brix. However, the model prediction accuracy dropped for Brix ($\text{RMSE} = 1.00$, $R^2 = 0.7$) compared to the original images ($\text{RMSE} = 0.91$, $R^2 = 0.75$). The pattern of the Shapley values is altered for firmness, the most significant patches are in the bottom row. However, the degree of significance for firmness is very low (maximum Shapley value is 0.001) compared to Brix (0.023) and starch (0.04). For starch, all patches are now significant, with the middle section having slightly less importance than the outer patches. In particular, the top left and bottom right, as before, were important patches, indicating that the models put more attention on the first and last patches. Starch model accuracy dropped slightly between the original images and swapped images (Table 5.4; run 8 vs 17).

5.4 Discussion

In our study, we evaluated the state-of-the-art effective model and tested the effects of wavelengths, regions of interest, cultivar encoding, and seasonality on models predicting soluble solids (Brix), firmness and starch. We found that imaging is an effective method for predicting soluble solids and firmness, but it does not work well for starch. The best model performance is achieved by training Vision Transformer models for each

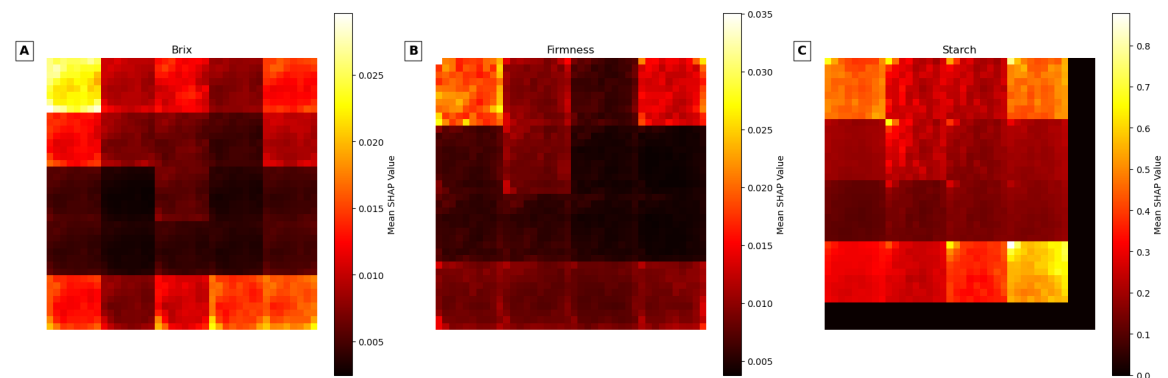


Figure 5.14: Mean spectral importance (Shapley values) per pixel for all apples for A) Brix, B) firmness and C) starch models. The black margin in starch is due to the image size (40-by-40 pixels) being divided into patch sizes of 9-by-9 pixels.

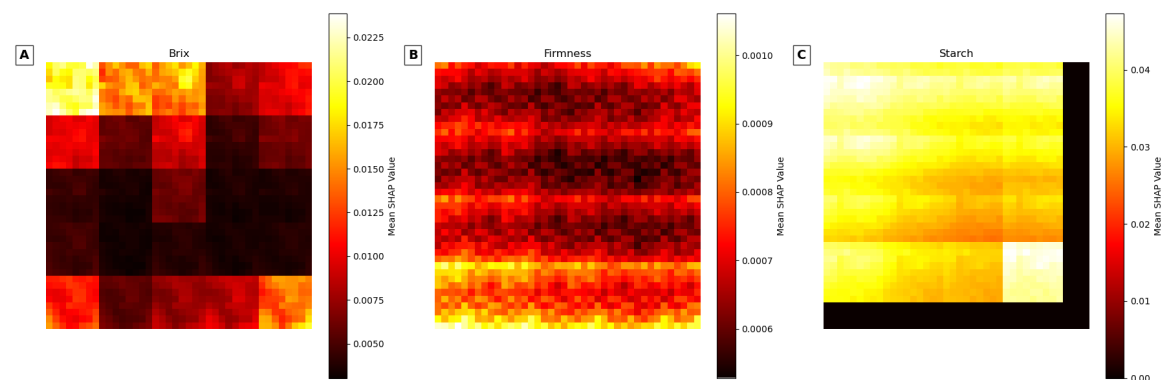


Figure 5.15: Mean spectral importance (Shapley values) per pixel for all apples for A) Brix, B) firmness and C) starch models trained on swapped images. The black margin in starch is due to the image size (40-by-40 pixels) being divided into patch sizes of 9-by-9 pixels.

maturity feature, on datasets from different seasons and geographic locations, training the models on more than one side of the fruit and encoding cultivar information in the training process. Reductions in computational processes can be achieved by reducing the wavelengths to the top 50% of most informative wavelengths.

5.4.1 Cultivar and Seasonal Information

The work in the present study incorporates data from multiple seasons, cultivars and geographical locations, in contrast to most previous studies whose dataset limit their generalisability. We assess the effect this additional information has on our model performance. The broader scope introduces more variability into the models, which may partially explain the comparatively poorer performance metrics compared to more focused studies. However, diversity is essential for building robust and generalisable models. The average spectral wavelengths varied distinctly between cultivars thus, cultivars were encoded in the model training process. Without cultivar encoding, we saw a slight drop in model accuracy for all features. During the preliminary assessment, we found that models trained on individual cultivars yielded inaccurate results and were highly bias (data not presented). Similar findings were observed when training individual seasonal data (Table 5.7). Splitting the training data into smaller sizes typically reduced the model accuracy, a common result in deep learning models. Cross-season prediction tests revealed a degree of season-specificity, with models performing best within their own season. Despite this, the full model (model trained on data from all three seasons) produced the best result, demonstrating the importance of training on diverse datasets to improve consistency and reliability. The authors recommend that data be trained on multi-season and multi-cultivar data for the best model performance.

5.4.2 Model Results

The firmness prediction model performed the best out of the three maturity features tested. We achieved RMSE of 0.76 and R^2 of 0.63 (Table 5.4; run 8). Comparing previous studies from multivariate linear regression (RMSE = 5.76 and R^2 = 0.74 ([68] and RMSE = 1.26 and R^2 = 0.67 [71]), least squares support vector machine (RMSE = 4.26 and R^2 = 0.88) [68] and partial least squares regression (PLSR) (RMSE = 0.99 and R^2 = 0.78) [70], predictions from our deep learning model yielded better RMSE but lower R^2 values. Our results are similar to a study using artificial neural network models (RMSE = 0.722 and R^2 = 0.896) [66]. Their study recorded firmness at a puncture depth of 1 mm, whereas in our study, we used the gold standard of puncture tests at a depth of 8 mm for our firmness measurements. It is interesting to find similar results because the mean light penetration depth within the visible region — where high importance of wavelengths was found (Figure 5.12B) — is 1-2 mm [129]. Our results suggest that HSI with ViT models is an effective method of predicting apple firmness.

Our second-best model was Brix (RMSE = 0.91, R^2 = 0.75). Brix is the measurement of soluble sugars in the juice. Compared to other studies using multivariate linear regression — with reported performance RMSE = 0.6 and R^2 = 0.76 [68], RMSE = 0.412 and R^2 = 0.96 [69] and RMSE = 0.82 and R^2 = 0.76 [71] — as well as least squares support vector machine (RMSE = 0.64 and R^2 = 0.74; [68]), and PLSR, with RMSE = 0.56 and R^2 = 0.79 [130] and RMSE = 0.54 and R^2 = 0.90 [70]), our results using ViT models performed worse. We suspect that soluble solids does not need complex models to make accurate predictions or perhaps it is due to the increased variation as our data is collected from multiple cultivars and seasons.

Starch percentage or the standard starch pattern index is a measurement of starch coverage inside the fruit. Starch degradation starts from the centre of the fruit, near the seeds and therefore, light does not penetrate deep enough for the models to predict

accurate starch predictions [129, 131]. This is evident in our study as predictions closer to unripe (100%) or overripe (0%) were more accurate than predictions in between (Figure 5.11). Thus, starch cannot be accurately predicted by spectral imaging. Starch results will not be further discussed due to the poor results.

5.4.3 Important Channels

Shapley values enable us to decode black box systems and identify which wavelengths are important in the model prediction process. The Shapley values indicate that wavelengths in the visible spectra at ranges 400-411, 453-465, 482-488, 538, 600-630, 648-668 and 671-692 nm were within the top 20% of the most significant wavelengths, with peaks at 400, 459, 488, 538, 618, 657 and 683 nm were important for predicting soluble solids (Brix) values (Figure 5.12A). Similarly, wavelengths between 400-408, 633-698, 565-600 and 1000 nm with peaks at 400, 589, 651, 683, and 1000 nm were within the top 20% of most important wavelengths for firmness (Figure 5.12B). The most important wavelengths identified in our study are similar to those identified by Merzlyak et al. [124] in their Plant Senescence Reflectance Index (PSRI; $(R_{678} - R_{500})/R_{NIR}$ or $(R_{678} - R_{500})/R_{800}$), which are correlated to fruit ripening onset, those used by Pourdarbani et al. [72] (535-560, 835-855 and 950-975 nm) and Zhai et al. [132] (403, 430, 551, 617, and 846 nm) to predict fruit maturity. The detected metabolites are likely chlorophyll, carotenoids and anthocyanins in the visible spectrum and the oxygen-hydrogen bonds abundant in soluble sugars and water in the NIR region [72]. Other studies have also identified several effective spectral ranges for predicting SSC and firmness within the boundaries of important wavelengths identified in our study. For instance, Wang et al. [70] highlighted normal spectral indices around 746/749 nm, while Zhao et al. [133] found the range between 704-805 nm to be effective in sugar content predictions. Similarly, Çetin et al. [71] reported that wavelengths at 505, 511, 704 and 689 nm were useful for SSC prediction. Çetin et al. [71]

attributed ratio combinations between 680, 880, 905 and 940 nm to firmness predictions. Although the wavelengths are not exact, most can also attribute their spectral results to chlorophyll, carotenoids, anthocyanins and O-H bonds for soluble sugars. Despite significant spectra not being exact, they attained similar or better RMSE and R^2 compared to our results. These studies select highly important wavelengths to predict their features. Whereas, in our study, we reduced the number of wavelengths by 50% (retaining 102 channels), proving that most wavelengths between 400-1000 nm are redundant for model predictions. Our process does not guarantee the removal of multicollinear wavelengths, potentially retaining wavelengths with similar information.

5.4.4 Regions of the Fruit

The Shapley values in the original images (Figure 5.14) suggests that edges, and more specifically, the top left and bottom right patches, are the most important regions when making model predictions. This would be consistent with the findings of [134], as the curvature of the apples affects the signal-to-noise quality of the reflectance. We tested the theory by swapping quadrants of the original images (Figure 5.2 and Figure 5.15) and found that the Shapley quadrants do not change the significant regions for Brix. This suggests that the regions of the fruit are unimportant for soluble solids. One reason could be that the soluble sugars of the fruit are homogenous throughout the fruit and therefore, sampling from any region will not make a difference to predictions. Another reason why the top left and bottom right patches may be more significant could be because the model is emphasising on the first and last patch in the model. The reduced accuracy with the swapped quadrants suggests that the quality of the image in the centre part of the apple is not as good as the edges, consistent with the findings from [134].

The original images (Figure 5.14) showed the same trend for firmness; where the centre patches are less important and more importance is placed on the edges. We

anticipated some importance to be placed in the centre as the puncture site when collecting firmness is near the equator of the apple. When the quadrants were swapped (Figure 5.15), the pattern of the Shapley values changed significantly. Each firmness patch was deemed important by Shapley values. This could be due to the uneven firmness across the fruit’s surface, or it might be an artefact of the data collection process — specifically, the side of the fruit used for firmness measurements may not correspond to the side shown in the images. This shows that fruit regions may have some significance for predicting firmness. In the future, if imaging were used for predicting firmness, the side corresponding to the image should be used for firmness measurements.

5.4.5 Single-sided Analysis

In our study, we captured images from all four sides of each apple. We treated each side as an independent apple. Previous studies averaged each side together to create a mean spectral image of the apple [72]. In our study, we treated each side independently because if imaging is applied to an orchard, the imaging system will only capture one side of the fruit. This is our rationale in testing prediction models trained on a single side of the fruit. These models resulted in a reduction in model accuracy and increased bias on the test dataset. The increased error might be due to a decreased amount of training data. This is common in deep learning models. Another reason could be due to a difference in the side of the image and the corresponding side used to measure firmness. We were unable to match up the sides as it was not recorded during the collection process. The variation of firmness between two sides of the fruit deviates by 0.46 ± 0.37 kgF, which may explain the decrease in model accuracy. To improve model results, the same face should be used to measure soluble solids and firmness as the side imaged.

5.4.6 Pre-processing Method

Fruit maturity is largely a factor of fruit size, shape and colour when discerned with the human eye. However, by locally averaging and aggregating pixels together, it removes size and shape information. The models chosen in our study required the same size input. Newer models that allow variable input shapes are available, and it may be worth retraining the models in the current study with different-sized images. In a practical setting within an orchard, retaining a fixed distance between the camera and the apples would be difficult; thus, the current models would achieve better results. There are advantages and disadvantages to either approach. Choosing between any approach would depend on the subject of interest.

5.4.7 Applications

As imaging is a non-destructive process, it enables repeated measurements over time. This allows growers to monitor apple maturity progression over the season. Growers can therefore schedule harvests based on the progression of each fruit. Such temporal tracking could significantly enhance crop management strategies, reduce waste and improve fruit quality at market. However, more research is required to adapt the present model to in-field applications. Environmental factors such as variations in light intensities and angles of the light can influence spectral images. Understanding and compensating for these effects are essential for producing a robust model for in-field maturity tracking.

5.4.8 Limitations

As apples have a curved surface, the light reflection and the distance of the apple edge changes. This results in variable pixel quality, particularly near edges, where shadows or glare may be more pronounced. Such variability can affect the accuracy of model training and introduce bias. In the future, incorporating a curvature-aware

preprocessing method may enhance pixel uniformity and improve model accuracy.

5.5 Conclusion

In summary, the large-scale hyperspectral dataset comprising of 5756 apples from multiple countries, seasons and cultivars have provided a robust foundation for predicting apple maturity features. Its scale and diversity mark a significant step forward in development of reliable, data-driven approaches for non-destructive maturity assessments. The results show that ViT models are able to predict soluble solids ($\text{RMSE} = 0.91$, $R^2 = 0.75$) and firmness ($\text{RMSE} = 0.76$, $R^2 = 0.63$) within reasonable accuracy, demonstrating their potential as effective tools for non-destructive apple maturity assessment across diverse cultivars, seasons and countries. Results from this study suggest that wavelengths around 400-411, 453-465, 482-488, 538, 600-630, 648-668 and 671-692 nm with peaks at 400, 459, 488, 538, 618, 657 and 683 nm are significant for Brix; and 400-408, 633-698, 565-600 and 1000 nm, with peaks at 400, 589, 651, 683, and 1000 nm are significant for firmness predictions with a Vision transformer model. We can reduce the wavelengths used in this study and make soluble solids and firmness predictions as accurately as the full model. It would be of interest to further reduce the number of wavelengths in the study to identify a small subset of key wavelengths. The region of the fruit is not particularly important in making predictions, but a clean signal is important for accuracy. The models show some seasonal specificity so it is important to train models with multi-season and multi-cultivar data for a more generalised model.

Conclusion and Discussion

This study demonstrates the practical potential of the combination of phenology models and hyperspectral imaging with ViT models for non-destructive assessment of apple maturity (soluble solids (Brix) and firmness). By combining both systems, growers can plan harvests more effectively, using long-term forecasting over entire orchards with phenology models and real-time maturity assessment on individual fruit with hyperspectral imaging.

Traditional maturity assessment methods are fundamental in establishing the optimal harvest window. However, assessments are destructive, labour-intensive and time-consuming. In this thesis, we investigated the potential of phenology and hyperspectral imaging as predictive tools for harvest timing. Our findings suggest that phenology models and hyperspectral imaging offer a viable and efficient alternative to conventional methods.

In the first study, we evaluated the effectiveness of phenology models in predicting apple flowering time. We applied the PhenoFlex model to our extensive flowering data, collected across 85 years, on a range of different cultivars from East Malling, UK. The study showed that a common apple model at the species level was better at predicting flowering time for the twenty-six apple cultivars studied than using models trained on individual cultivars. The trained model could predict the flowering date within 5 days of harvest. Similar results can be found with models trained on groups of apple

cultivars clustered by flowering time. The predictions in this approach depended on the internal consistency of each group, and results were within 5-6 days of the harvest window. Both approaches gave predictions comparable to previous flowering time prediction studies, with results ranging between 3-6 days [44, 82, 88]. The results from this study were slightly higher, since we are reporting the average RMSE for all twenty-six cultivars as opposed to individual cultivars. There is some risk in training models on a single cultivar; models trained on individual cultivars can yield polarising (ranging from precise to misleading) results. Our results show that generalised models — models trained on data either aggregated with a large number of cultivars or divided into smaller groups by flowering date are appropriate and more reliable for making flowering time predictions compared to single cultivar approaches.

Since flowering time can be predicted with reasonable accuracy, the harvest date can be forecasted using the average flowering date, a convention commonly used to estimate the harvest date of apples [50, 101]. However, due to the use of the average flowering date in this process, harvest date predictions often overlook the variation in flowering time and its influence on fruit maturity. We recorded the flowering date and fruit maturity from five apple cultivars harvested over two seasons to determine how much flowering time influences the fruit maturity. The growth units were calculated from the flowering date to harvest date for each apple cluster. The results show that up to 20% of maturity variation is explained by flowering time variation, with this effect being more pronounced in early-flowering cultivars than late-flowering cultivars. This suggests that accounting for flowering time variability is crucial for improving the accuracy of harvest predictions, particularly for early-season cultivars. Therefore, phenology models that rely on modelling the climate data alone is insufficient for accurate harvest date predictions. Given the remaining variability in harvest timing predicted by phenology models, there is a need for methods that can assess fruit maturity more accurately, closer to the harvest date.

In this thesis, we tested hyperspectral imaging as a non-destructive method to evaluate fruit maturity. We imaged over 5000 apples over three seasons from both the United Kingdom and New Zealand and applied deep learning models to find patterns between the hyperspectral data and maturity data. Our results demonstrated that deep learning vision transformer models could reliably predict key maturity indicators — soluble sugar content and firmness — with prediction accuracies of 0.91°Brix and 0.79 kgF, respectively. Our results are consistent with previous studies predicting firmness with hyperspectral imaging [68, 71], but our soluble solids (Brix) model underperformed in comparison with previous models [70, 130] suggesting that soluble solids does not require deep learning models for accurate soluble solids predictions. Despite less accurate results, our model can still predict well within reasonable errors.

Overall, phenology models are valuable to ensure that seasonal weather conditions will satisfy the chilling and forcing requirements specific to each cultivar. Without sufficient environmental cues, the apples will not flower and therefore will not fruit. They can also approximate the harvest window for each cultivar. While phenology models provide useful estimates for the harvest window, they fail to capture the full variability induced by differences in flowering time. On the other hand, deep learning models trained on hyperspectral images can accurately determine the levels of soluble solids and firmness at the individual fruit level. However, the deep learning model trained on hyperspectral images may not generalise well beyond the conditions of the current study. Therefore, integrating phenology models with hyperspectral imaging offers a promising alternative to traditional harvest prediction methods. By combining both systems, growers could plan harvests more effectively, using long-term forecasting over entire orchards with phenology models and real-time maturity assessment on individual fruit with hyperspectral imaging.

Applications

Phenology models, such as the PhenoFlex model, can be applied to forecast the harvest window. As temperature data is required as the model inputs, and growers are primarily interested in the coming season's harvest date, historical temperature data can be used to initially simulate temperatures to predict an estimated harvest date, then refine the predictions with actual temperatures experienced by the trees. According to the results from our studies, this will attain an estimated harvest window within ± 7 days of the actual harvest date. Hyperspectral imaging can be applied from the earliest point of the potential harvest day, tracking the development of soluble solids and firmness over time. By monitoring the progress of soluble solids and firmness, the optimal harvest maturity can be determined by observing the changes in soluble solids, which increases when ripe, and firmness, which decreases with maturity, for each fruit. Future applications may extend beyond apples and into other fruits.

6.0.1 Contributions

We show the potential of phenology models and imaging for non-destructive soluble solids and firmness predictions of apple fruit. Generalised multi-cultivar models are more effective for flowering time and maturity quality predictions in both phenology and deep learning ViT models. This allows a simplified approach to model applications since it only requires a single model for different cultivars. We also found a link between the variation of flowering time and flowering time itself, where more maturity variation is induced in early flowering cultivars compared to late flowering cultivars. The models can be applied with cultivar and flowering time information to more accurately determine the harvest dates. Moreover, imaging using the key wavelengths identified in our study — spectral bands within the visible spectrum (400-411, 453-465, 482-488, 538- 565-630, 648-668 and 471-692 nm) and one band (1000 nm) in the near infrared spectrum — can greatly reduce the camera costs, making a more affordable

and non-destructive tool for predicting maturity, thereby making them more widely accessible. Since the data collected in this study spans more than one country, we show a potential for a generalised ViT model to work effectively on apples from different countries.

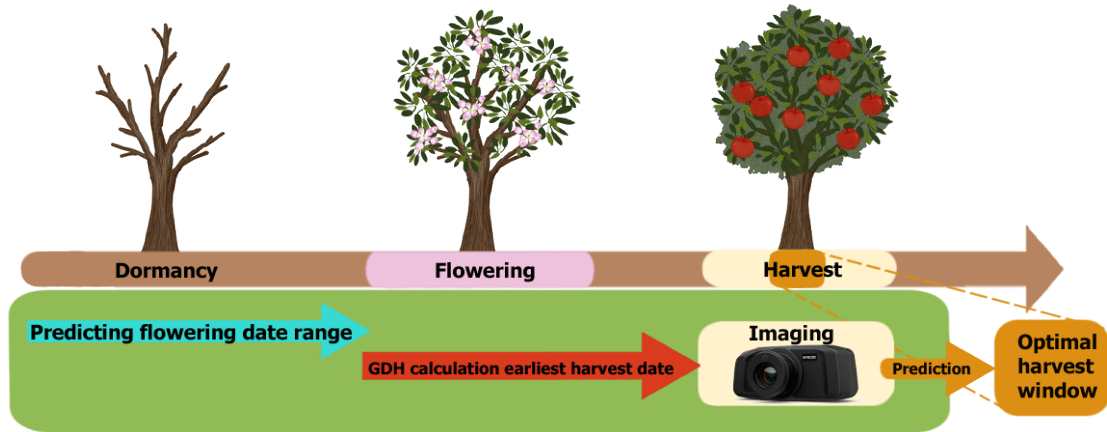


Figure 6.1: Graphical representation of applications. The phenological phases are represented in the top brown bar, and the applications and their applied range are shown below as red and blue bars.

Limitations

Phenology models are dependent on temperature data. When simulated data is used, prediction accuracy can be compromised. This would impact regions where temperature is inconsistent or where there is limited coverage or lack of historical data to simulate with. The increasing unpredictability of the climate due to unusual weather events may introduce more frequent and extreme anomalies. However, previous phenology models will also be vulnerable to unusual weather patterns. Training models with updated temperature data will help maintain model robustness.

Apple varieties are constantly produced through breeding programs. While we present a large range of existing apple cultivars, our results may not generalise well to newer varieties. As a result, our findings may not fully translate to future varieties, limiting the long-term applicability of the results. The models should be continuously trained

and recalibrated with the emergence of new cultivars.

As we captured apple images in a highly controlled environment, the feasibility of hyperspectral imaging will need to be tested in-field with the reduced wavebands identified in our study. The effects of variable lighting and occlusion by leaves or other fruit will affect the accuracy of model results. The distance of the apples from the camera will also affect model results. The apples captured for our study were aligned in a row, with small distances between the camera and each apple. In-field applications would typically scan the entire canopy height for fruit from approximately 1 m away, thus reducing the resolution of the pixels of each apple. Reduced pixel resolution may limit the accuracy and reliability of predictive models. Moreover, the distance of each apple from the camera will introduce inconsistencies, further influencing the model results. These factors will need to be investigated in the future to determine their impact on the model performance in outdoor settings. Further preprocessing of image data will be needed to standardise inputs and mitigate the variability.

References

- [1] G. D. Blanpied and K. J. Silsby, “Predicting harvest date windows for apples,” *Cornell Cooperative Extension*, 1992.
- [2] F. Department for Environment and R. Affairs, “Horticulture statistics - dataset,” DEFRA, Tech. Rep., 2021.
- [3] V. Tijero, F. Girardi, and A. Botton, “Fruit development and primary metabolism in apple,” *Agronomy*, vol. 11, p. 1160, 2021.
- [4] Food and A. Organization, “Top 20 commodities, gross production value in united kingdom of great britain and northern ireland,” https://www.fao.org/faostat/en/#rankings/commodities_by_country, 2023, accessed: 2025-08-30.
- [5] B. Bajzelj and W. Parry, “Food waste in primary production in the uk,” WRAP, Banbury, Tech. Rep., 2019.
- [6] M. Khadiri, H. Boubaker, H. Lahmamsi, M. Taoussi, R. Ezzougari, L. Askarne, A. Farhaoui, E. Ait Barka, and R. Lahlali, “Challenges in apple preservation: Fungicide resistance and emerging biocontrols,” *Physiological and Molecular Plant Pathology*, vol. 129, p. 102205, 2024.
- [7] F. Department for Environment and R. Affairs, “Uk food production to supply ration in 2024,” UK Government, Dashboard, 2024, accessed: 2025-08-31. [Online]. Available: <https://defra-farming-stats.github.io/auk-dashboard/#food-production-to-supply-ratio-by-commodity>
- [8] F. . R. A. Department for Environment, “Marketing standard for apples,” APHA, DEFRA, Tech. Rep., 2024.

-
- [9] L. Argenta, S. de Freitas, J. Mattheis, M. Vieira, and C. Ogoshi, “Characterization and quantification of postharvest losses of apple fruit stored under commercial conditions,” *HortScience*, vol. 56, pp. 608–616, 2021.
- [10] R. Hueppe and K. Zander, “Perfect apples or sustainable production?—consumer perspectives from germany,” *Journal of Consumer Behaviour*, vol. 23, pp. 698–710, 2024.
- [11] M. Ahmadi-Afzadi, I. Tahir, and H. Nybom, “Impact of harvesting time and fruit firmness on the tolerance to fungal storage diseases in an apple germplasm collection,” *Postharvest Biology and Technology*, vol. 82, pp. 51–58, 2013.
- [12] C. Watkins, M. Erkan, J. Nock, K. Iungerman, R. Beaudry, and R. Moran, “Harvest date effects on maturity, quality, and storage disorders of ‘honeycrisp’ apples,” *HortScience*, vol. 40, pp. 164–169, 2005.
- [13] J. Song and F. Bangerth, “The effect of harvest date on aroma compound production from ‘golden delicious’ apple fruit and relationship to respiration and ethylene production,” *Postharvest Biology and Technology*, vol. 8, pp. 259–269, 1996.
- [14] M. Zude, “Comparison of indices and multivariate models to non-destructively predict the fruit chlorophyll by means of visible spectrometry in apple fruit,” *Analytica Chimica Acta*, vol. 481, no. 1, pp. 119–126, 2003.
- [15] N. Bondonno, C. Bondonno, N. Ward, J. Hodgson, and K. Croft, “The cardiovascular health benefits of apples: Whole fruit vs. isolated compounds,” *Trends in Food Science & Technology*, vol. 69, pp. 243–256, 2017.
- [16] J. Boyer and R. Liu, “Apple phytochemicals and their health benefits,” *Nutrition Journal*, vol. 3, pp. 1–15, 2004.

-
- [17] F. Gheyas, S. Blankenship, E. Young, and R. McFeeters, "Dietary fibre composition in developing apple fruits," *Journal of Horticultural Science & Biotechnology*, vol. 73, pp. 631–639, 1998.
- [18] M. Kellerhals, "Introduction to apple (*malus* × *domestica*)," in *Genetics and Genomics of Rosaceae*. Springer, 2009, pp. 73–84.
- [19] M. Knee and S. M. Smith, "Variation in quality of apple fruits stored after harvest on different dates," *Journal of Horticultural Science*, vol. 64, pp. 413–419, 1 1989.
- [20] F. Coville, "The influence of cold in stimulating the growth of plants," *Journal of Agricultural Research*, vol. 20, pp. 151–160, 1920.
- [21] O. Heide and A. Prestrud, "Low temperature, but not photoperiod, controls growth cessation and dormancy induction and release in apple and pear," *Tree Physiology*, vol. 25, pp. 109–114, 2005.
- [22] P. Wareing, "Photoperiodism in woody plants," *Annual Review of Plant Physiology*, vol. 7, pp. 191–214, 1956.
- [23] R. Pertille, I. Citadin, L. Patto, T. Oldoni, S. Scariotto, C. Grigolo, and P.-É. Lauri, "High-chilling requirement apple cultivar has more accentuated acrotony than low-chilling one in mild winter region," *Trees*, vol. 35, pp. 1135–1150, 2021.
- [24] J. Campoy, D. Ruiz, and J. Egea, "Dormancy in temperate fruit trees in a global warming context: a review," *Scientia Horticulturae*, vol. 130, pp. 357–372, 2011.
- [25] A. Erez, "Bud dormancy; phenomenon, problems and solutions in the tropics and subtropics," in *Temperate Fruit Crops in Warm Climates*. Springer, 2000, pp. 17–48.
- [26] J. Petri and G. Leite, "Consequences of insufficient winter chilling on apple tree bud-break," in *VII International Symposium on Temperate Zone Fruits in the*

- Tropics and Subtropics*, vol. 662, 2003, pp. 53–60.
- [27] J. Rodrigo, “Spring frosts in deciduous fruit trees—morphological damage and flower hardiness,” *Scientia Horticulturae*, vol. 85, pp. 155–173, 2000.
- [28] W. Chandler, *Deciduous orchards*. Deciduous orchards, 1942.
- [29] E. Luedeling and P. Brown, “A global analysis of the comparability of winter chill models for fruit and nut trees,” *International Journal of Biometeorology*, vol. 55, pp. 411–421, 2011.
- [30] E. Richardson, “A model for estimating the completion of rest for “redhaven” and “elberta” peach trees,” *HortScience*, vol. 9, pp. 331–332, 1974.
- [31] A. Erez, S. Fishman, G. Linsley-Noakes, and P. Allan, “The dynamic model for rest completion in peach buds,” in *II International Symposium on Computer Modelling in Fruit Research and Orchard Management*, vol. 276, 1989, pp. 165–174.
- [32] S. Fishman, A. Erez, and G. Couvillon, “The temperature dependence of dormancy breaking in plants: mathematical analysis of a two-step model involving a cooperative transition,” *Journal of Theoretical Biology*, vol. 124, pp. 473–483, 1987.
- [33] ———, “The temperature dependence of dormancy breaking in plants: computer simulation of processes studied under controlled temperatures,” *Journal of Theoretical Biology*, vol. 126, pp. 309–321, 1987.
- [34] J. Anderson, E. Richardson, and C. Kesner, “Validation of chill unit and flower bud phenology models for ‘montmorency’ sour cherry,” *I International Symposium on Computer Modelling in Fruit Research and Orchard Management*, vol. 184, pp. 71–78, 1985.

-
- [35] R. Darbyshire, I. Farrera, J. Martinez-Lüscher, G. Leite, V. Mathieu, A. el Yaacoubi, and J.-M. Legave, “A global evaluation of apple flowering phenology models for climate adaptation,” *Agricultural and Forest Meteorology*, vol. 240, pp. 67–77, 2017.
- [36] G. Ashcroft, E. Richardson, and S. Seeley, “A statistical method of determining chill unit and growing degree hour requirements for deciduous fruit trees,” *HortScience*, 1977.
- [37] J. Landsberg, “Apple fruit bud development and growth; analysis and an empirical model,” *Annals of Botany*, vol. 38, pp. 1013–1023, 1974.
- [38] K. Pope, D. da Silva, P. Brown, and T. DeJong, “A biologically based approach to modeling spring phenology in temperate deciduous trees,” *Agricultural and Forest Meteorology*, vol. 198, pp. 15–23, 2014.
- [39] M. Cannell and R. Smith, “Thermal time, chill days and prediction of budburst in picea sitchensis,” *Journal of Applied Ecology*, pp. 951–963, 1983.
- [40] I. Chuine, P. Cour, and D. Rousseau, “Fitting models predicting dates of flowering of temperate-zone trees using simulated annealing,” *Plant, Cell & Environment*, vol. 21, pp. 455–466, 1998.
- [41] A. Erez and G. Couvillon, “Characterization of the influence of moderate temperatures on rest completion in peach,” *Journal of the American Society for Horticultural Science (USA)*, 1987.
- [42] R. Guerriero, S. Indigine, and G. Scalabrelli, “The effect of cyclic and constant temperatures in fulfilling the chilling requirement of two apricot cultivars,” in *VIII International Symposium on Apricot Culture and Decline*, vol. 192, 1985, pp. 41–48.
- [43] A. Naor, M. Flaishman, R. Stern, A. Moshe, and A. Erez, “Temperature effects

- on dormancy completion of vegetative buds in apple,” *Journal of the American Society for Horticultural Science*, vol. 128, pp. 636–641, 2003.
- [44] E. Luedeling, K. Schiffrers, T. Fohrmann, and C. Urbach, “Phenoflex-an integrated model to predict spring phenology in temperate fruit trees,” *Agricultural and Forest Meteorology*, vol. 307, p. 108491, 2021.
- [45] A. Lakso, L. Corelli Grappadelli, J. Barnard, and M. Goffinet, “An expolinear model of the growth pattern of the apple fruit,” *Journal of Horticultural Science*, vol. 70, pp. 389–394, 1995.
- [46] C. J. Stanley, D. S. Tustin, G. B. Lupton, S. McCartney, W. M. Cashmore, and H. N. D. Silva, “Towards understanding the role of temperature in apple fruit growth responses in three geographical regions within new zealand,” *The Journal of Horticultural Science and Biotechnology*, vol. 75, pp. 413–422, 2000.
- [47] C. Atkinson, L. Taylor, and G. Kingswell, “The importance of temperature differences, directly after anthesis, in determining growth and cellular development of malus fruits,” *Journal of Horticultural Science & Biotechnology*, vol. 76, pp. 721–731, 2001.
- [48] I. J. Warrington, T. A. Fulton, E. A. Halligan, and H. N. D. Silva, “Apple fruit growth and maturity are affected by early season temperatures,” *Journal of the American Society for Horticultural Science*, vol. 124, pp. 468–477, 1999.
- [49] G. Baker and R. Brooks, “Climate in relation to deciduous fruit production in california. iii. effect of temperature on number of days from full bloom to harvest of apricot and prune fruits,” in *Proceedings of the American Society for Horticultural Science*, 1944, p. 104.
- [50] M. T. Luton and P. J. C. Hamer, “Predicting the optimum harvest dates for apples using temperature and full-bloom records,” *Journal of horticultural science*, vol. 58, pp. 37–44, 1983.

-
- [51] K. B. Perry, S. M. Blankenship, and C. R. Unrath, "Predicting harvest date of 'delicious' and 'golden delicious' apples using heat unit accumulations," *Agricultural and Forest Meteorology*, vol. 39, pp. 81–88, 1987.
- [52] T. Sugiura, N. Fukuda, T. Tsuchida, M. Sakurai, and H. Sugiura, "Modeling the relationship between apple quality indices and air temperature," *The Horticulture Journal*, vol. 92, no. 4, pp. 424–430, 2023.
- [53] H. Tang, X. Zhai, and X. Xu, "Flowering variation induces apple maturity variation at harvest," *Frontiers in Agronomy*, vol. 7, p. 1545070, 2025.
- [54] R. Volz, J. Palmer, and H. Gibbs, "Within-tree variability in fruit quality and maturity for 'royal gala'apple," in *International Symposium on Quality of Fruit and Vegetables: Influence of Pre-and Post-Harvest Factors and Technology 379*, 1993, pp. 67–74.
- [55] J. Bohner and F. Bangerth, "Effects of fruit set sequence and defoliation on cell number, cell size and hormone levels of tomato fruits (*lycopersicon esculentum* mill.) within a truss," *Plant Growth Regulation*, vol. 7, pp. 141–155, 1988.
- [56] L. Boyd, P. Ramankutty, A. Barnett, T. Dawson, T. Wegrzyn, A. le Guevel, and A. Mowat, "Effect of canopy position on fruit quality in 'hort16a' kiwifruit in new zealand," *Journal of Horticultural Science & Biotechnology*, vol. 83, pp. 791–797, 2008.
- [57] A. Richardson, H. Boldingh, P. Kashuba, G. Knight, and D. Ellingham, "Flowering time determines the weight and composition of *actinidia chinensis* var. *chinensis* 'zesy002' kiwifruit," *Scientia Horticulturae*, vol. 246, pp. 741–748, 2019.
- [58] M. Saltveit Jr and S. Hale, "Determining the maturity of north carolina apples," North Carolina Agricultural Extension Bulletin, Tech. Rep. AG-282, 1982.

- [59] M. Ahmed, A. Villordon, and M. Kamruzzaman, *Hyperspectral imaging*. Elsevier, 2019.
- [60] G. Campeanu, G. Neata, and G. Darjanschi, “Chemical composition of the fruits of several apple cultivars growth as biological crop,” *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*, vol. 37, pp. 161–164, 2009.
- [61] B. Nicolai, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K. Theron, and J. Lammertyn, “Nondestructive measurement of fruit and vegetable quality by means of nir spectroscopy: A review,” *Postharvest Biology and Technology*, vol. 46, pp. 99–118, 2007.
- [62] H. Wang, J. Peng, C. Xie, Y. Bao, and Y. He, “Fruit quality evaluation using spectroscopy technology: a review,” *Sensors*, vol. 15, pp. 11 889–11 927, 2015.
- [63] R. van Beers, B. Aernouts, L. León Gutiérrez, C. Erkinbaev, K. Rutten, A. Schenk, B. Nicolai, and W. Saeys, “Optimal illumination-detection distance and detector size for predicting braeburn apple maturity from vis/nir laser reflectance measurements,” *Food and Bioprocess Technology*, vol. 8, pp. 2123–2136, 2015.
- [64] Z. Guo, M. Wang, A. Agyekum, J. Wu, Q. Chen, M. Zuo, H. El-Seedi, F. Tao, J. Shi, and Q. Ouyang, “Quantitative detection of apple watercore and soluble solids content by near infrared transmittance spectroscopy,” *Journal of Food Engineering*, vol. 279, p. 109955, 2020.
- [65] B. Park, J. Abbott, K.-J. Lee, C. Choi, and K. Choi, “Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of delicious and gala apples,” *Transactions of the ASAE*, vol. 46, p. 1721, 2003.
- [66] A. Peirs, J. Lammertyn, K. Ooms, and B. Nicolai, “Prediction of the optimal picking date of different apple cultivars by means of vis/nir-spectroscopy,”

- Postharvest Biology and Technology*, vol. 21, pp. 189–199, 2001.
- [67] X. Xu, H. Xu, L. Xie, and Y. Ying, “Effect of measurement position on prediction of apple soluble solids content (ssc) by an on-line near-infrared (nir) system,” *Journal of Food Measurement and Characterization*, vol. 13, pp. 506–512, 2019.
- [68] N. Ekramirad, A. Rady, A. Adedeji, and R. Alimardani, “Application of hyperspectral imaging and acoustic emission techniques for apple quality prediction,” *Transactions of the ASABE*, vol. 60, pp. 1391–1401, 2017.
- [69] L. Li, Y. Peng, C. Yang, and Y. Li, “Optical sensing system for detection of the internal and external quality attributes of apples,” *Postharvest Biology and Technology*, vol. 162, p. 111101, 2020.
- [70] F. Wang, C. Zhao, H. Yang, H. Jiang, L. Li, and G. Yang, “Non-destructive and in-site estimation of apple quality and maturity by hyperspectral imaging,” *Computers and Electronics in Agriculture*, vol. 195, p. 106843, 2022.
- [71] N. Çetin, K. Karaman, E. Kavuncuoğlu, B. Yıldırım, and A. Jahanbakhshi, “Using hyperspectral imaging technology and machine learning algorithms for assessing internal quality parameters of apple fruits,” *Chemometrics and Intelligent Laboratory Systems*, vol. 230, p. 104650, 2022.
- [72] R. Pourdarbani, S. Sabzi, D. Kalantari, R. Karimzadeh, E. Ilbeygi, and J. Arribas, “Automatic non-destructive video estimation of maturation levels in fuji apple (*malus malus pumila*) fruit in orchard based on colour (vis) and spectral (nir) data,” *Biosystems Engineering*, vol. 195, pp. 136–151, 2020.
- [73] E. T. Hamadziripi, K. I. Theron, M. Muller, and W. J. Steyn, “Apple compositional and peel color differences resulting from canopy microclimate affect consumer preference for eating quality and appearance,” *HortScience*, vol. 49, pp. 384–392, 2014.

- [74] L. Kalcsits, J. Mattheis, L. Giordani, M. Reid, and K. Mullin, "Fruit canopy positioning affects fruit calcium and potassium concentrations, disorder incidence, and fruit quality for 'honeycrisp' apple," *Canadian Journal of Plant Science*, vol. 99, pp. 761–771, 2019.
- [75] D. Kviklys, J. Viškelis, M. Liaudanskas, V. Janulis, K. Laužikė, G. Samuolienė, N. Uselis, and J. Lanauskas, "Apple fruit growth and quality depend on the position in tree canopy," *Plants*, vol. 11, p. 196, 2022.
- [76] R. Rea and E. Eccel, "Phenological models for blooming of apple in a mountainous region," *International Journal of Biometeorology*, vol. 51, pp. 1–16, 2006. [Online]. Available: <https://doi.org/10.1007/s00484-006-0043-x>
- [77] T. O. Perry, "Dormancy of trees in winter: Photoperiod is only one of the variables which interact to control leaf fall and other dormancy phenomena." *Science*, vol. 171, no. 3966, pp. 29–36, 1971.
- [78] J. Janick, "The origins of fruits, fruit growing, and fruit breeding," *Plant breeding reviews*, vol. 25, no. 1, pp. 255–320, 2005.
- [79] S. Shultz, "Apples," *Journal of Agricultural & Food Information*, vol. 5, no. 4, pp. 77–84, 2003.
- [80] U. Carsten, E. Luedeling, and K. Schiffrs, "Phenoflex," [urlhttps://cran.r-project.org/web/packages/chillR/vignettes/PhenoFlex.html](https://cran.r-project.org/web/packages/chillR/vignettes/PhenoFlex.html), 2022, accessed: 2023-07-08.
- [81] I. Chuine, "A unified model for budburst of trees," *Journal of theoretical biology*, vol. 207, no. 3, pp. 337–347, 2000.
- [82] R. Darbyshire, K. Pope, and I. Goodwin, "An evaluation of the chill overlap model to predict flowering time in apple tree," *Scientia Horticulturae*, vol. 198, pp. 142–149, 2016.

- [83] A. Melke, “The physiology of chilling temperature requirements for dormancy release and bud-break in temperate fruit trees grown at mild winter tropical climate,” *Journal of Plant Studies*, vol. 4, 2015. [Online]. Available: <https://doi.org/10.5539/jps.v4n2p110>
- [84] R. Hauagge and J. Cummins, “Phenotypic variation of length of bud dormancy in apple cultivars and related malus species,” *Journal of the American Society for Horticultural Science*, vol. 116, pp. 100–106, 1991. [Online]. Available: <https://doi.org/10.21273/jashs.116.1.100>
- [85] A. Shaltout and C. Unrath, “Rest completion prediction model for ‘starkrimson delicious’ apples,” *Journal of the American Society for Horticultural Science*, vol. 108, pp. 957–961, 1983. [Online]. Available: <https://doi.org/10.21273/jashs.108.6.957>
- [86] R. Réaumur, “Observations du thermomètre faites pendant l’année mdccxxxv comparées a celles qui ont été faites sous la ligne a l’isle-de-france, a alger et en quelques-unes de nos isles de l’amérique,” *Mémoires de l’Académie Royale des Sciences*, pp. 545–576, 1735.
- [87] D. Asse, C. F. Randin, M. Bonhomme, A. Delestrade, and I. Chuine, “Process-based models outcompete correlative models in projecting spring phenology of trees in a future warmer climate,” *Agricultural and Forest Meteorology*, vol. 285, p. 107931, 2020.
- [88] B. Drepper, A. Gobin, S. Remy, and J. Van Orshoven, “Comparing apple and pear phenology and model performance: what seven decades of observations reveal,” *Agronomy*, vol. 10, p. 73, 2020. [Online]. Available: <https://doi.org/10.3390/agronomy10010073>
- [89] L. Andreini, I. G. de Cortázar-Atauri, I. Chuine, R. Viti, S. Bartolini, D. Ruiz, J. A. Campoy, J. M. Legave, J.-M. Audergon, and P. Bertuzzi, “Understanding

- dormancy release in apricot flower buds (*prunus armeniaca* l.) using several process-based phenological models,” *Agricultural and forest meteorology*, vol. 184, pp. 210–219, 2014.
- [90] U. Meier, H. Graf, H. Hack, M. Heb, W. Kennel, R. Klose, D. Mappes, D. Seipp, R. StauB, and J. Streif, “Phänologische entwicklungsstadien des kernobstes, des steinobstes der johannisbeere und der erdbeere,” *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes*, vol. 46, pp. 141–153, 1994.
- [91] E. Luedeling and E. Fernandez, “chillr: Statistical methods for phenology analysis in temperate fruit trees,” 2022.
- [92] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, 2002.
- [93] J. E. Nash and J. V. Sutcliffe, “River flow forecasting through conceptual models part i—a discussion of principles,” *Journal of hydrology*, vol. 10, no. 3, pp. 282–290, 1970.
- [94] J. R. Magness, H. C. Diehl, and M. H. Haller, *Picking maturity of apples in relation to storage*. US Department of Agriculture, University of Illinois at Urbana-Champaign, 1926.
- [95] M. W. Goncalves, L. C. Argenta, and M. S. D. MARTIN, “Maturity and quality of apple fruit during the harvest period at apple industry,” *Revista Brasileira de Fruticultura*, vol. 39, no. 5, pp. e-825, 2017.
- [96] L. C. Argenta, R. M. Wood, J. P. Mattheis, F. R. Thewes, C. N. Nesi, and D. A. Neuwald, “Factors affecting development of disorders expressed after storage of ‘gala’ apple fruit,” *Postharvest Biology and Technology*, vol. 204, p. 112439, 2023.
- [97] B. G. Wilkinson and R. O. Sharples, “The relation between time of picking and storage disorders in cox’s orange pippin apple fruits,” *Journal*

- of Horticultural Science*, vol. 42, pp. 67–82, 1 1967. [Online]. Available: <https://doi.org/10.1080/00221589.1967.11514194>
- [98] J. Streif, “Optimum harvest date for different apple cultivars in the ’bodensee’ area,” in *COST 94. The postharvest treatment of fruit and vegetables. Determination and prediction of optimum harvest date of apples and pears. Proceedings of a meeting*, A. de Jager, D. Johnson, and E. Höhn, Eds. Lofthus, Norway: COST 94, jun 1994, pp. 15–20, presented at the meeting held on 9–10 June 1994. Published in 1996. Abstract record from CABI Databases, 21 August 1996.
- [99] S. Musacchi and S. Serra, “Apple fruit quality: Overview on pre-harvest factors,” *Scientia Horticulturae*, vol. 234, pp. 409–430, 2018.
- [100] A. Shewa, D. Gobena, and M. Ali, “Review on postharvest quality and handling of apple,” *J. Agric. Sci. Food Technol*, vol. 8, no. 1, pp. 028–032, 2022.
- [101] G. D. Blanpied, “Observations of the ripening and harvest indices at commercial harvest dates of ‘delicious’ apple at its extreme northern latitudes,” *HortScience*, vol. 17, pp. 783–785, 1982.
- [102] G. Blanpied, “The relationship between growing season temperatures, bloom dates and the length of the growing season of red delicious apples in north america.” *Proceedings. American Society for Horticultural Science*, 1964.
- [103] G. D. Blanpied and S. Ben-David, “A new york study of ‘mcintosh’ apple optimum harvest dates1,” *Journal of the American Society for Horticultural Science*, vol. 95, pp. 151–154, 1970.
- [104] F. B. Abeles and G. W. Lightner, “Optimal harvest date equations for west virginia apples,” *HortScience*, vol. 19, pp. 429–430, 1984.
- [105] S. Sapkota, J. Liu, M. T. Islam, P. Ravindran, P. P. Kumar, and S. M. Sherif, “Contrasting bloom dates in two apple cultivars linked to differential levels of phy-

- to hormones and heat requirements during ecodormancy,” *Scientia Horticulturae*, vol. 288, p. 110413, 2021.
- [106] G. Lysiak, “The sum of active temperatures as a method of determining the optimum harvest date of ‘sampion’ and ‘ligol’ apple cultivars,” *Acta Scientiarum Polonorum. Hortorum Cultus*, vol. 11, no. 6, pp. 3–13, 2012.
- [107] T. L. Robinson, E. J. Seeley, and B. H. Barritt, “Effect of light environment and spur age on ‘delicious’ apple fruit size and quality,” *Journal of the American Society for Horticultural Science*, vol. 108, pp. 855–861, 1983.
- [108] P. D. Drogoudi and G. Pantelidis, “Effects of position on canopy and harvest time on fruit physico-chemical and antioxidant properties in different apple cultivars,” *Scientia Horticulturae*, vol. 129, pp. 752–760, 2011.
- [109] M. S. Krishnaprakash, B. Aravindaprasad, C. A. Krishnaprasad, P. Narasimham, S. M. Ananthakrishna, S. Dhanaraj, and V. S. Govindarajan, “Effect of apple position on the tree on maturity and quality,” *Journal of Horticultural Science*, vol. 58, pp. 31–36, 1 1983.
- [110] J. Streif, “optimale erntetermin beim apfel. i. qualitätsentwicklung und reife,” *Horticultural science*, vol. 48, pp. 154–159, 1983.
- [111] J. Anderson and S. Seeley, “Modelling strategy in pomology: development of the utah models,” in *III International Symposium on Computer Modelling in Fruit Research and Orchard Management 313*, 1992, pp. 297–306.
- [112] E. A. Richardson, S. D. Seeley, D. R. Walker, J. Anderson, and G. L. Ashcroft, “Pheno-climatography of spring peach bud development.” *HortScience*, vol. 10, pp. 236–237, 1975.
- [113] H. Tang, X. Zhai, and X. Xu, “Evaluating the performance of models predicting the flowering times of twenty-six apple cultivars in england,” *European Journal*

- of Agronomy*, vol. 160, p. 127319, 2024.
- [114] T. L. Wagner, H.-I. Wu, P. J. H. Sharpe, R. M. Schoolfield, and R. N. Coulson, “Modeling insect development rates: a literature review and application of a biophysical model,” *Annals of the Entomological Society of America*, vol. 77, pp. 208–220, 1984.
- [115] X. Xu, “The effects of constant and fluctuating temperatures on the length of the incubation period of apple powdery mildew (*podosphaera leucotricha*),” *Plant Pathology*, vol. 45, pp. 924–932, 1996.
- [116] M. H. Al-Whaibi, “Plant heat-shock proteins: a mini review,” *Journal of King Saud University-Science*, vol. 23, pp. 139–150, 2011.
- [117] M. Ahmed, A. Villordon, and M. Kamruzzaman, “Hyperspectral imaging and explainable deep-learning for non-destructive quality prediction of sweetpotato,” *Postharvest Biology and Technology*, vol. 222, p. 113379, 2025.
- [118] E. Fass, E. Shlomi, C. Ziv, O. Glickman, and D. Helman, “Machine learning models based on hyperspectral imaging for pre-harvest tomato fruit quality monitoring,” *Computers and Electronics in Agriculture*, vol. 229, p. 109788, 2025.
- [119] Y. Wei, S. Yao, F. Wu, and Q. Yu, “Maturity classification and quality determination of cherry using vnir hyperspectral images and comprehensive chemometrics,” *International Journal of Fruit Science*, vol. 24, pp. 363–379, 2024.
- [120] F. Vesali, M. Gharibkhani, and M. Komarizadeh, “An approach to estimate moisture content of apple with image processing method,” *Australian Journal of Crop Science*, vol. 5, pp. 111–115, 2011.
- [121] J. Nyasordzi, H. Friedman, Z. Schmilovitch, T. Ignat, A. Weksler, I. Rot, and S. Lurie, “Utilizing the iad index to determine internal quality attributes of

- apples at harvest and after storage,” *Postharvest Biology and Technology*, vol. 77, pp. 80–86, 2013.
- [122] J. Sjöstrand, I. Tahir, H. Hovmalm, L. Garkava-Gustavsson, H. Stridh, and M. Olsson, “Comparison between iad and other maturity indices in nine commercially grown apple cultivars,” *Scientia Horticulturae*, vol. 324, p. 112559, 2024.
- [123] V. Ziosi, M. Noferini, G. Fiori, A. Tadiello, L. Trainotti, G. Casadoro, and G. Costa, “A new index based on vis spectroscopy to characterize the progression of ripening in peach fruit,” *Postharvest Biology and Technology*, vol. 49, pp. 319–329, 2008.
- [124] M. Merzlyak, A. Solovchenko, and A. Gitelson, “Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit,” *Postharvest Biology and Technology*, vol. 27, pp. 197–211, 2003.
- [125] H. Zhu, S. Qin, S. Liang, M. Su, P. Wang, and Y. He, “Hyperspectral imaging and machine learning for quality assessment of apples with different bagging types,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, p. 126443, 2025.
- [126] J. Behmann, K. Acebron, D. Emin, S. Bennertz, S. Matsubara, S. Thomas, D. Bohnenkamp, M. Kuska, J. Jussila, and H. Salo, “Specim iq: evaluation of a new, miniaturized handheld hyperspectral camera and its application for plant phenotyping and disease detection,” *Sensors*, vol. 18, p. 441, 2018.
- [127] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, and L. Gustafson, “Sam 2: Segment anything in images and videos,” *arXiv preprint*, vol. arXiv:2408.00714, 2024.
- [128] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, and S.-H. Deng, “Hyperparameter optimization for machine learning models based on bayesian optimization,”

- Journal of Electronic Science and Technology*, vol. 17, pp. 26–40, 2019.
- [129] J. Qin and R. Lu, “Monte carlo simulation for quantification of light transport features in apples,” *Computers and Electronics in Agriculture*, vol. 68, pp. 44–51, 2009.
- [130] S. Fan, Q. Wang, X. Tian, G. Yang, Y. Xia, J. Li, and W. Huang, “Non-destructive evaluation of soluble solids content of apples using a developed portable vis/nir device,” *Biosystems Engineering*, vol. 193, pp. 138–148, 2020.
- [131] J. Lammertyn, A. Peirs, J. De Baerdemaeker, and B. Nicolai, “Light penetration properties of nir radiation in fruit with respect to non-destructive quality assessment,” *Postharvest Biology and Technology*, vol. 18, pp. 121–132, 2000.
- [132] H. Zhai, P. Xie, X. Xie, and S. Sha, “Deep learning-enabled hyperspectral imaging for high-accuracy non-destructive quantification of nutritional components in multi-variety apples,” *Frontiers in Plant Science*, vol. 16, p. 1634785, 2025.
- [133] J. Zhao, S. Vittayapadung, Q. Chen, S. Chaitep, and R. Chuaviroj, “Nondestructive measurement of sugar content of apple using hyperspectral imaging technique,” *Maejo International Journal of Science and Technology*, vol. 3, no. 1, pp. 130–142, 2009.
- [134] X. Tian, J. Li, Q. Wang, S. Fan, W. Huang, and C. Zhao, “A multi-region combined model for non-destructive prediction of soluble solids content in apple, based on brightness grade segmentation of hyperspectral imaging,” *Biosystems Engineering*, vol. 183, pp. 110–120, 2019.