Kiwifruit Yield Prediction

Delivering predictive models of fruit yield count and fruit sizing for kiwifruit crops

A Food Agility CRC project

By Dr Evan Webster, Dr Jianlong Zhou, Dr Zhidong Li,
Dr Merlin Zhang, Dr Tuo Mao, Jichao Kan & Ashley Rootsey

FINAL REPORT

July 2020 - October 2022

[ISBN 978-0-6454652-3-5]

Project Name: Kiwifruit Yield Prediction
Project No: FA049

Citation: Dr Evan Webster, Dr Jianlong Zhou, Dr Zhidong Li, Dr Merlin Zhang, Dr Tuo Mao, Jichao Kan, and Ashley Rootsey (2023), Kiwifruit Yield Prediction, Food Agility CRC, Sydney.
Cover: Kiwifruits sliced lengthwise (Credit: Canva)

IMPORTANT INFORMATION

Author contact details
Name: Ashley Rootsey
Phone: 0418 943 895
Email: projects@foodagility.com

Food Agility CRC contact details
University of Technology Sydney, Broadway
Ultimo NSW 2007

+61 2 8001 6119
hello@foodagility.com
www.foodagility.com

Electronically published by Food Agility CRC at www.foodagility.com in [February 2023].

The Food Agility CRC is funded under the Australian Government Cooperative Research Centres Program.

Australian Government
Department of Industry, Science and Resources
AusIndustry
Cooperative Research Centres Program

ACKNOWLEDGEMENTS

The project would like to acknowledge the ongoing support of the team from Zespri International in the delivery and rollout of the project, namely Peter McHannigan and Mark Graham, who provided initiative and leadership in commencing the partnership.

The project would also like to acknowledge the support of Anothony Pangborn, of Agriculture Risk Management Ltd for his support in coordinating NZ-based kiwifruit packhouses and other industry stakeholders.

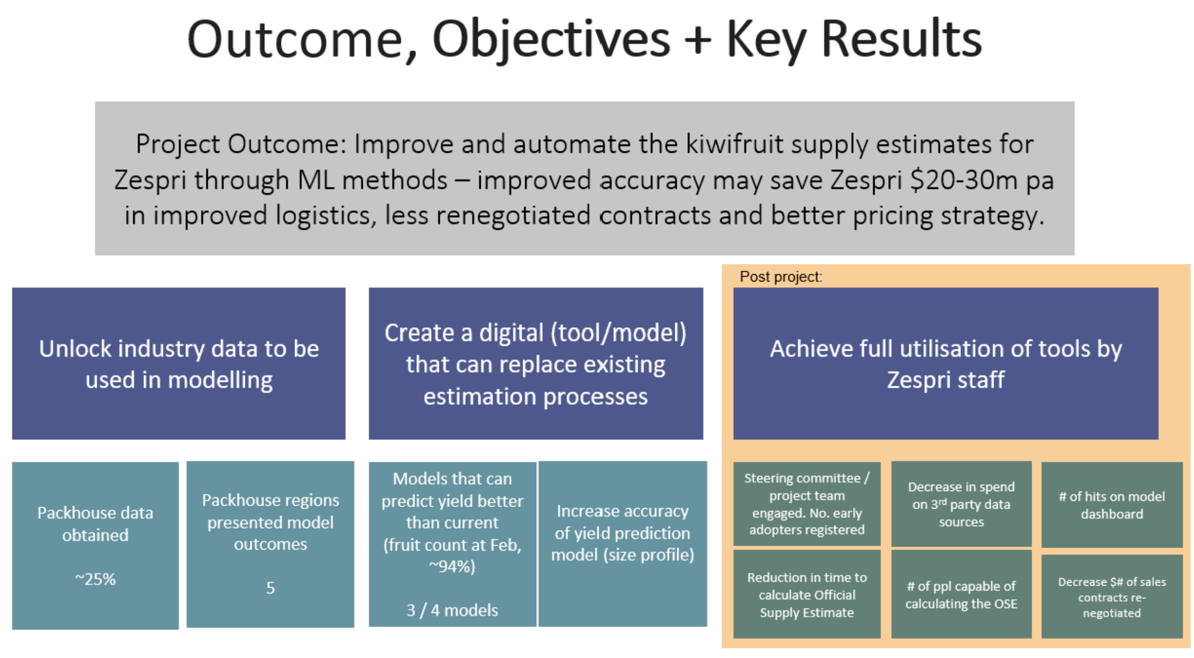
Table of Contents

PROJECT DESCRIPTION

The project aims to accurately predict the national kiwifruit supply of New Zealand each growing season, through producing machine learning models that incorporate both total quantity of fruit (measured in # of trays) and fruit size (number of kiwifruit in a tray).

Prediction models are primarily based on a collation of publicly available datasets and were delivered to Zespri by UTS in partnership with The Yield Technology Solutions.

The project achieved accurate POC models which will enable Zespri to improve its forward contracting of fruit supply, better optimise supply logistics and reduce cost of yield forecasting activities. These models are to be fully integrated into The Yield's Digital Playbook solution to further enhance accuracy and operational performance.



PROJECT PARTNERS

The Yield Technology Solutions is a leading Australian agricultural technology company who uses real time data and AI to power technology that solves challenges at farm level and throughout food supply chains. Their customers are typically commercial growers and large enterprises in irrigated and specialty crops.

The **University of Technology Sydney (UTS)** Data Science Institute (DSI), led by Distinguished Professor Fang Chen. This institute specialises in the delivery of world-class Artificial Intelligence/Machine Learning projects across a variety of industries.

Zespri International Limited is the world's largest marketer of kiwi fruit, selling kiwi fruit into more than 60 countries and managing 30% of the global volume. They aim to create a worldwide, year-round supply of kiwi fruit to the growing market. To do this, they've formed partnerships with 2,700 growers; having almost quadrupled this number since 2000.

Food Agility CRC were a leading investor and facilitator of this collaboration.

EXECUTIVE SUMMARY

Forward estimation of harvest yields and supply has always been an inherently risky and largely uncertain exercise, caused by the extreme susceptibility of agricultural crops to seasonal variability and uncontrollable elements such as weather and climate. Despite this, the need for effective and accurate forward planning in agricultural supply chains is critical for agribusinesses to be globally competitive and sustainable.

Zespri faces particularly pertinent needs for accurate fruit supply estimates for the purposes of supply chain resource allocation, market access planning, forward contracting and price setting. We also recognise that there are key stages throughout the growing season where supply estimates are needed to help conduct these core business functions.

The Yield has a proven record in delivering advanced AI-driven climate and weather services to horticultural growers and has demonstrated its ability to use these services for the purposes of harvest yield predictions at site, regional and national levels in key adjacent horticultural crops such as wine grapes and berries. This project works to test the hypotheses that similar supply forecasting services may be feasible in kiwifruit at national and regional scales to the level of accuracy desired by Zespri, and that these predictions can be achieved nationally with little-to-no direct on-farm input.

The results clearly demonstrate the viability of machine learning and modern data science to derive the OSE and its size profile using historical Zespri submit data and NIWA weather data alone. The estimates from the machine learning models satisfied the total tray count target (<6% error) for 3 of the 5 years tested, among variety and growing types. Averaged over 5 harvest seasons 2017 – 2021, the total tray count error was 6.4% for GACK, 8.5% for GAOB, 5.1% for HWCK, and 5.7% for HWOB. With regard to Zespri's size profiling target (<1.5% error per size category), the average % error was 2.5% for GACK, 2.3% for GAOB, 2.4% for HWCK, and 2.5% for HWOB, over the 2017 – 2021 harvest seasons. When averaged over the last 3 harvest seasons (2019 – 2021), the % error was 2.0% for GACK, 1.8% for GAOB, 0.7% for HWCK, and 2.2% for HWOB.

Given the research to date, there is clear opportunity for additional improvement, within a commercial data science capacity, to develop a robust and reliable solution to forecast the overall count and size profile of the OSE within the target values desired by Zespri.

INDUSTRY IMPACT METRICS

GACK: 95.2%
GAOB: 92.9%
HWCK: 94.9%
HWOB: 94.4%

Accuracy of kiwi count prediction model

GACK: 79.7%
GAOB: 77.0%
HWCK: 84.0%
HWOB: 75.7%

Accuracy of the total tray size prediction model

25%

NZ packhouses (by volume) agreed to share data with project

$20-30 million

Estimated operational savings for Zespri by adoption of accurate yield prediction models

END-USER PROFILE

Crop Supply Manager, Zespri

Historically Zespri's Crop Supply Manager has relied on ad-hoc information and subjective observations made throughout the growing season to deliver an Official Supply Estimate for Zespri sales and marketing teams.

The Crop Supply Manager provides a national estimate of kiwifruit supply in key times of the growing season which is used for forward contracting kiwifruit export sales and logistics.

They are typically the only person in Zespri responsible for delivering these estimates.

OBJECTIVES

This research project aims to estimate the official supply count and size profile of kiwifruit trays, per variety and growing type, using state of the art machine learning forecast models. The models seek to exploit gridded weather data (NIWA and IBM), to derive relevant weather model inputs, determined through initial workshops with Zespri stakeholders and industry experts. Forecasts of the OSE are sought per tray size category for each variety and growing type. While the models are required to estimate the total OSE, forecast values are also desired per KPIN to determine the source of the fruit geographically.

Forecasted tray counts per size category and KPIN have been generated for the 2021 and 2022 harvest seasons, calculated in February of each harvest year. In addition to this the predictive power of the models has been assessed against the 2017, 2018, 2019, 2020, and 2021 harvest seasons to determine the comparative performance of the models against Zespri's own OSE forecasts. The results of these comparisons are shown and summarised below.

The current research models act as a proof-of-concept (PoC) demonstration of the capabilities of machine learning methodologies, driven by gridded weather data and historical harvest data. This research PoC was undertaken by The Yield's research partners, The Food Agility Cooperative Research Centre (FACRC) and The University of Technology Sydney (UTS). Going forward, The Yield seeks to use this PoC as a foundation with which it can build commercial machine learning solutions to provide accurate, robust, and repeatable forecasts of the OSE in the lead up and during harvest. This will provide Zespri with a repeatable machine-driven reference point with which to refine their OSE, enabling them to mitigate many of the challenges caused by seasonal variability of supply, improve their price realisation capabilities, and thereby deliver greater commercial returns both internally and to the wider New Zealand kiwifruit industry.

Significance

Complex interactions between the environment and plants cause costly uncertainty and variability in yield, fruit quality, sizing, and maturity in vine crops such as kiwifruit. Uncertainty in kiwifruit yield and fruit sizing has previously been estimated by Zespri to cost $20 – 30 million per year. Since around 95% of Zespri's revenues are returned to growers, the impact of mitigating these losses could have significant positive impact for the entire industry.

METHODOLOGY

Data

This project fundamentally uses historical and past season orchard-level yield data, combined with national and regional weather data for fruit supply predictions. The yield data includes number of trays in each size group, after all fruit rejections. The weather data is from NIWA, the public weather service in New Zealand. Additional data inputs such as smart monitoring field observations data, and rejection rate data from packhouses were also tested for utility in this project.

Methodology

Official supply estimates are produced through the application of data engineering and machine learning modelling, enabling estimates to be generated per KPIN of tray counts per size group and for each growing method and variety.

The development of tailored engineered weather features, required for the machine learning models, were conducted through consultation workshops with key Zespri stakeholders and project teams. These Zespri workshops and the information collected from them were critical to the development of important weather-related harvest features that influence size distributions and tray counts throughout the growing season (Fig. 1). In addition to the engineered weather features, Zespri submit data from the past 10 years was used to derive relevant harvest features that characterised the typical growing patterns per variety and growing type at and orchard level.

Building from a combined set of engineered weather and harvest features (Fig. 2), key input feature correlations were analysed to ascertain their impact on harvest level outcomes at an orchard level. Through this process, key feature combinations were identified per variety and growing type. From these feature sets, different machine learning models were exhaustively trialled over the past 5 harvest seasons to determine the optimum feature set, model type, and parameters that would provide the best OSE forecast per growing type and variety (Fig. 2). Using this optimum feature set, past season harvest data and weather data were used to train a machine learning model (Fig. 2) which, was then used with up-to-date weather data for the current season to generate estimates of trays for each KPIN and size category (Fig. 2). Aggregating these estimated values provided the equivalent of Zespri's OSE by count and size profile.

Harvest Season	June	July	August	September	October	November	December	January	February	March	April	May	Harvest Outcome	
	Dormancy		Bud Burst / Bud Break		Flowering & Pollination		Fruit Set	Fruit Growth		Early Harvest	Harvest / Leaf Fall			
2010	Winter Chill, Warm Spells, PYR, GDD		Frost Events, High Temperatures, High Winds, or High Rainfall		Rain, Wind, Low or High Temperatures, Frost Events		Hail, Wind, Wet Conditions	Drought, Rain, Dry Conditions, and High Winds		Rain, Low Temperatures, Dry Conditions, Light Levels	Frost Risk, Low Temperatures, Hail, and Rain		Known	Training Data
2011	Winter Chill, Warm Spells, PYR, GDD		Frost Events, High Temperatures, High Winds, or High Rainfall		Rain, Wind, Low or High Temperatures, Frost Events		Hail, Wind, Wet Conditions	Drought, Rain, Dry Conditions, and High Winds		Rain, Low Temperatures, Dry Conditions, Light Levels	Frost Risk, Low Temperatures, Hail, and Rain		Known	
...	Winter Chill, Warm Spells, PYR, GDD		Frost Events, High Temperatures, High Winds, or High Rainfall		Rain, Wind, Low or High Temperatures, Frost Events		Hail, Wind, Wet Conditions	Drought, Rain, Dry Conditions, and High Winds		Rain, Low Temperatures, Dry Conditions, Light Levels	Frost Risk, Low Temperatures, Hail, and Rain		Known	
2022	Winter Chill, Warm Spells, PYR, GDD		Frost Events, High Temperatures, High Winds, or High Rainfall		Rain, Wind, Low or High Temperatures, Frost Events		Hail, Wind, Wet Conditions	Drought, Rain, Dry Conditions, and High Winds		Rain, Low Temperatures, Dry Conditions, Light Levels	Frost Risk, Low Temperatures, Hail, and Rain		Known	
2023	Winter Chill, Warm Spells, PYR, GDD		Frost Events, High Temperatures, High Winds, or High Rainfall		Rain, Wind, Low or High Temperatures, Frost Events (Known through weather predictions only)		Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	To be Estimated	Forecast Data

Figure 1: Summary of the way that relevant weather features were engineered and used within the training and prediction datasets throughout the harvest season.

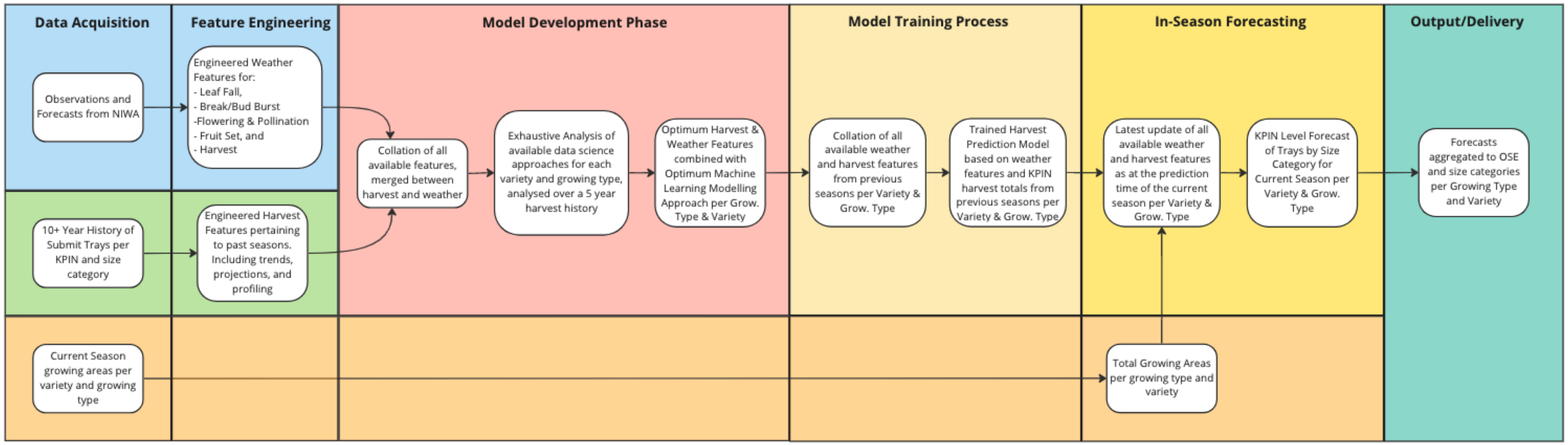


Figure 2: Overview of the methodology used within this research project, ranging from data acquisition, feature engineering for weather and harvest data, model selection and development, model training, in-season forecasting, and delivery of predictions.

Terminology:

Count prediction refers to the total number of delivered trays of kiwifruit. Size categories delineate the number of kiwifruit within each tray, therefore a larger size group number indicates smaller sized fruits. Tray quantities are only available after individual packhouses pack the fruit. The prediction for count is to predict total equivalent trays, and the prediction for size is to predict the distribution of size proportions/percentages among the size categories (sum=100%). The ultimate output from the research modelling defines the number of trays for each size category for each variety and growing type.

Two kiwifruit varieties and two growing methods are analysed in this project. The two varieties are Gold kiwifruit and Green (Hayward) kiwifruit which are denoted using GA and HW respectively. Two growing methods, organic and conventional, are represented by OB and CK respectively.

The prediction of tray-based values, including count and size groups with count, is evaluated by accuracy. The accuracy is calculated by 1-error which is the difference between total predicted trays and total actual trays divided by total actual trays. The accuracy is usually less than 1 (or 100%) where 1 means the perfect prediction.

Weighted mean average percentage accuracy (WMAPA) (Eq. 1) is the accuracy aggregation used to assess the performance of tray counts across all size categories. The calculation weights size categories more heavily if they contain higher numbers of actual trays while size categories with smaller numbers of trays are weighted less heavily, such as size 16 and 42 trays. This is to remove the data uncertainty bias from very small size groups.

$$WMAPA = 100.\frac{\sum_{i=0}^{n}(Act_i - ABS|For_i - Act_i|)}{\sum_{i=0}^{n} Act_i} \quad \text{(Eq. 1)}$$

Rationale:

The project tested three prediction components for total tray count, size (distributions, proportions of each size group), and total trays per size category. The purposes of testing these three components are to: 1) find the best prediction methodology; 2) see if count and size can relate complimentarily in combined predictions; and 3) investigate how auxiliary datasets, such as Smart Monitoring Data and Packhouse Data could be integrated for improved prediction accuracy.

Technical challenges:

- The approach taken in this research project to forecast upcoming tray counts at harvest was sensitive to orchard level details, such as planting area, on-farm management, etc. However, this information was largely unavailable in this project due to Zespri's role in downstream marketing rather than on-farm production. We overcame this by using location information and historical yield information to distinguish each orchard.

- NIWA weather observations and forecasts for this research project were sourced from weather stations that are at varying proximities to the orchards, creating uncertainty as to how representative the weather forecasts and observations were to growing conditions at each orchard. This risk was mitigated by using advanced spatial interpolation techniques like kriging to estimate the growing conditions at individual orchards.
- Short time series data is used for the prediction, where the observed harvest data is typically available historically for 5 years for gold and 8 years for green kiwifruit varieties. This relatively short period of data in a machine learning context may sometimes cause over-fitting of the models. To mitigate this impact, methods were used to prevent the overfitting of data within the models.
- In-season farm management data is not available as an input for the models. Key management decisions such as irrigation and nutrient application will impact the site level outcomes of kiwifruit growth, however, data reflecting these practices is not available in the project. We investigated the possibility of using Smart Monitoring observation data to gain some useful insight into on-farm crop growth during the season.
- The overall amount of Packhouse data obtained for this project was less than 25% of the total available in New Zealand and therefore its use within the project was limited. Furthermore, the data is represented by total fruit volumes, requiring transformation to equivalent trays to be directly compatible with Zespri submit data and the objectives of the project.

RESULTS

The research project results are divided into two main categories:

1. Estimates of total tray counts per harvest season, summarised by variety and growing type; and
2. Estimates of size profiles per harvest season, summarised by variety and growing type.

Both categories consist of estimates generated from machine learning models trained using historical weather data and Zespri submit data extending back to the 2010 harvest season. The results below demonstrate the viability of the trained models in relation to forecast performance for the 2017, 2018, 2019, 2020, and 2021 harvest seasons. Only the harvest and weather data from past seasons is used when training the models (Fig. 1). Machine Learning models were trained at different points through the season, from November through to March, producing tray counts per variety and growing type and each size category at a KPIN/orchard level. The reconciled outcomes, aggregated according to the categories above, are evaluated for models trained in February for the past 5 seasons (2017, 2018, 2019, 2020, and 2021). Predictions for the 2022 season have been generated and have been shared previously with Zespri but are yet to be fully reconciled.

When aggregated to a total tray count, irrespective of size, the results varied by variety and growing type. Gold Conventional estimates were within the 94% - 106% threshold for 2018, 2019, and 2020 but were above the threshold in 2017 and below the threshold in 2021 (Fig. 3). Gold Organic estimates were within the threshold in 2017 and 2018 but were below the threshold in 2019, 2020, and 2021 (Fig. 3). Green Conventional estimates were within the thresholds for 2018, 2019, and 2020 but overpredicted the total trays in 2017 and under predicted the total trays in 2021 (Fig. 3). Green Organic estimates were within the threshold in 2017, 2019, and 2020 but under predicted the total trays in 2018 and 2021 (Fig. 3). Significant rainfall in the 2017 harvest season was identified as a contributor to the over prediction of estimates in Gold and Green Conventional varieties while, high harvest productivity and crop densities were not adequately captured for the 2021 harvest season leading to the under-prediction present across all varieties and growing types (Fig. 4). The unique 2017 and 2021 growing conditions and harvest outcomes are now part of the machine learning training engine and will better represent these impacts in future harvest seasons.

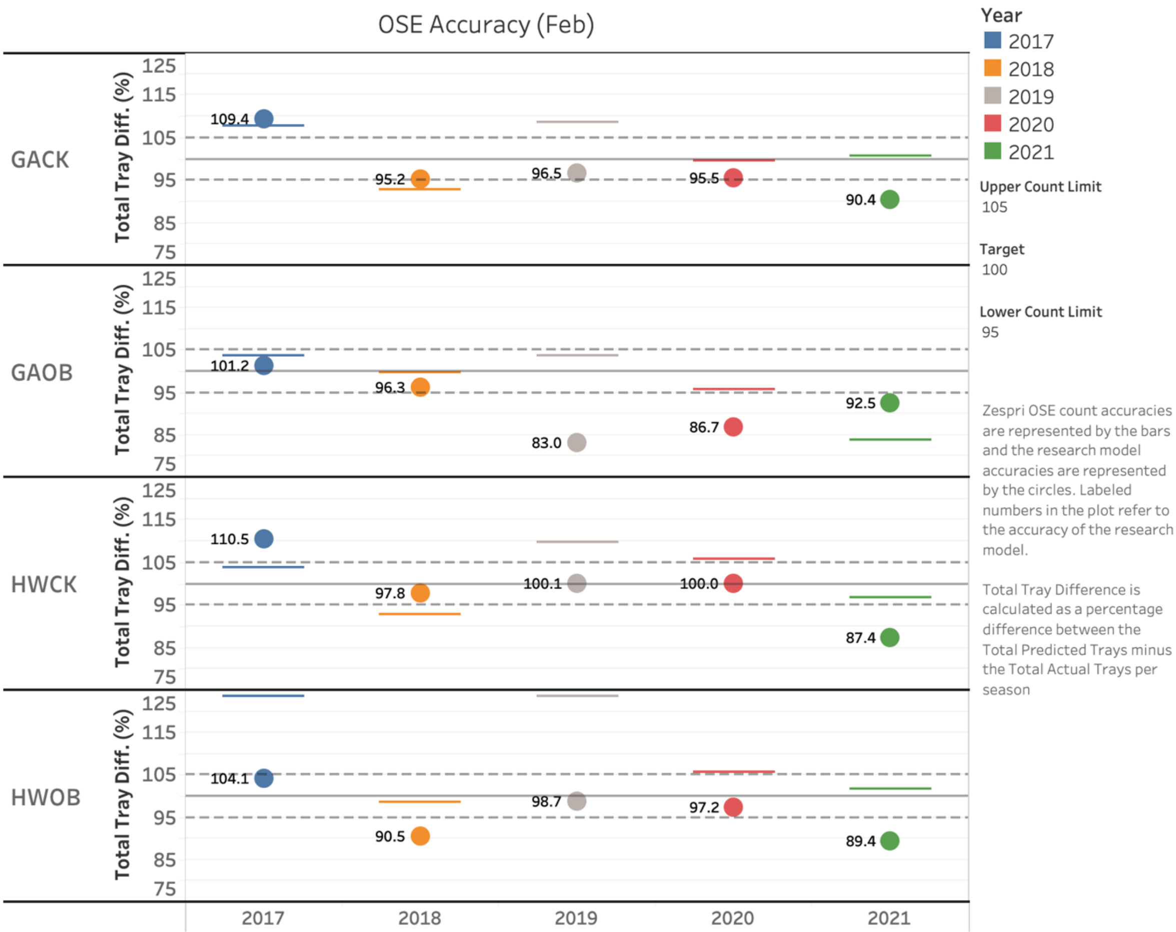


Figure 3: A chart summarising the accuracy of Official Supply Estimates forecasted by the UTS research model (Circles) and Zespri (bars) for 2017 – 2021 harvest seasons. Estimates are referenced to Feb of each harvest year.

Crop density values per variety and growing type over the 2017 – 2021 harvest seasons reveal distinct differences between Green and Gold varieties, with Gold crop densities shown to increase over the time period and Green crop densities shown to be variable but stable over time (Fig. 4). Over predictions and under predictions generally reflect the variations shown for the total tray counts (Fig. 3 & Fig. 4).

Figure 4: A summary of the relative density estimates as of February for both the UTS research model (Blue) and Zespri's OSE estimate (Green). The official supply estimate density is indicated to demonstrate the over or under predication of fruit densities (Yellow). Summaries for 2017 through 2021 are shown.

For size distribution estimates, predictive performance varied by harvest year, variety, and growing type (Fig. 5 & Fig. 6), with size performance, on average, seen to increase from year to year as more training data became available (Fig. 7 & Fig. 8). Size profile estimates generally matched the relative size profile of the submit data, with stronger deviations seen in 2017 and 2018 for both Green and Gold varieties, where tray count estimates favoured smaller fruit than what ended up being harvested (Fig. 5 & Fig. 6). Higher rainfall volumes in 2017 are thought to have been the driver of these discrepancies.

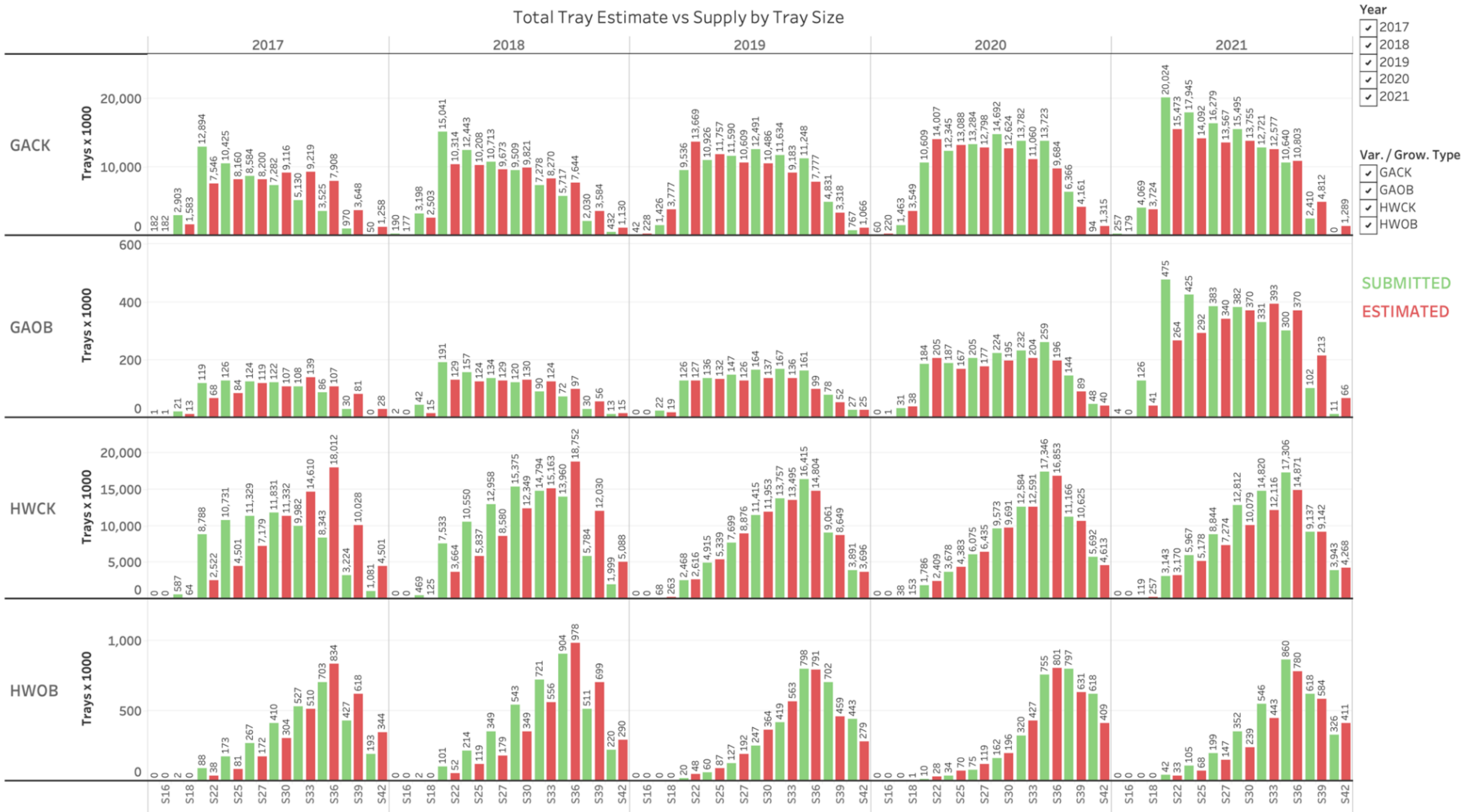


Figure 5: A breakdown of size predictions and submit data actuals per growing type, variety, and harvest season. Submit data is coloured in Green and estimated sizes are coloured in Red. Forecasted estimates of tray counts per size category are valid as at February of each harvest year.

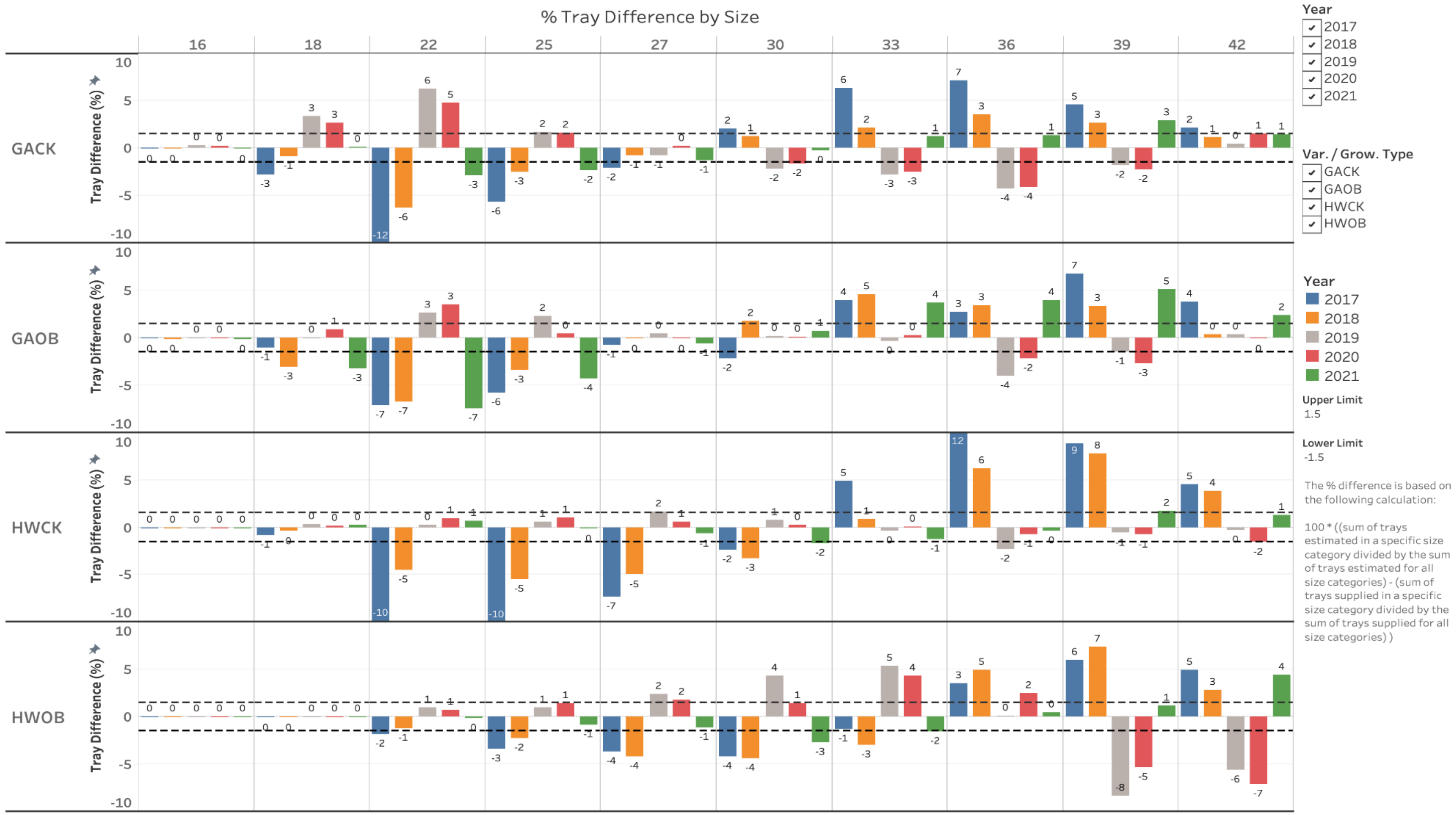


Figure 6: A breakdown of % differences between total trays estimated and submitted, separated per size category, growing type, variety, and harvest season. Indicators for the desired 1.5% error threshold are represented by black dotted lines above and below the target value of zero % difference. Occasions where the coloured bars are within these dashed lines indicates that the size prediction is satisfactory according to Zespri's requirements.

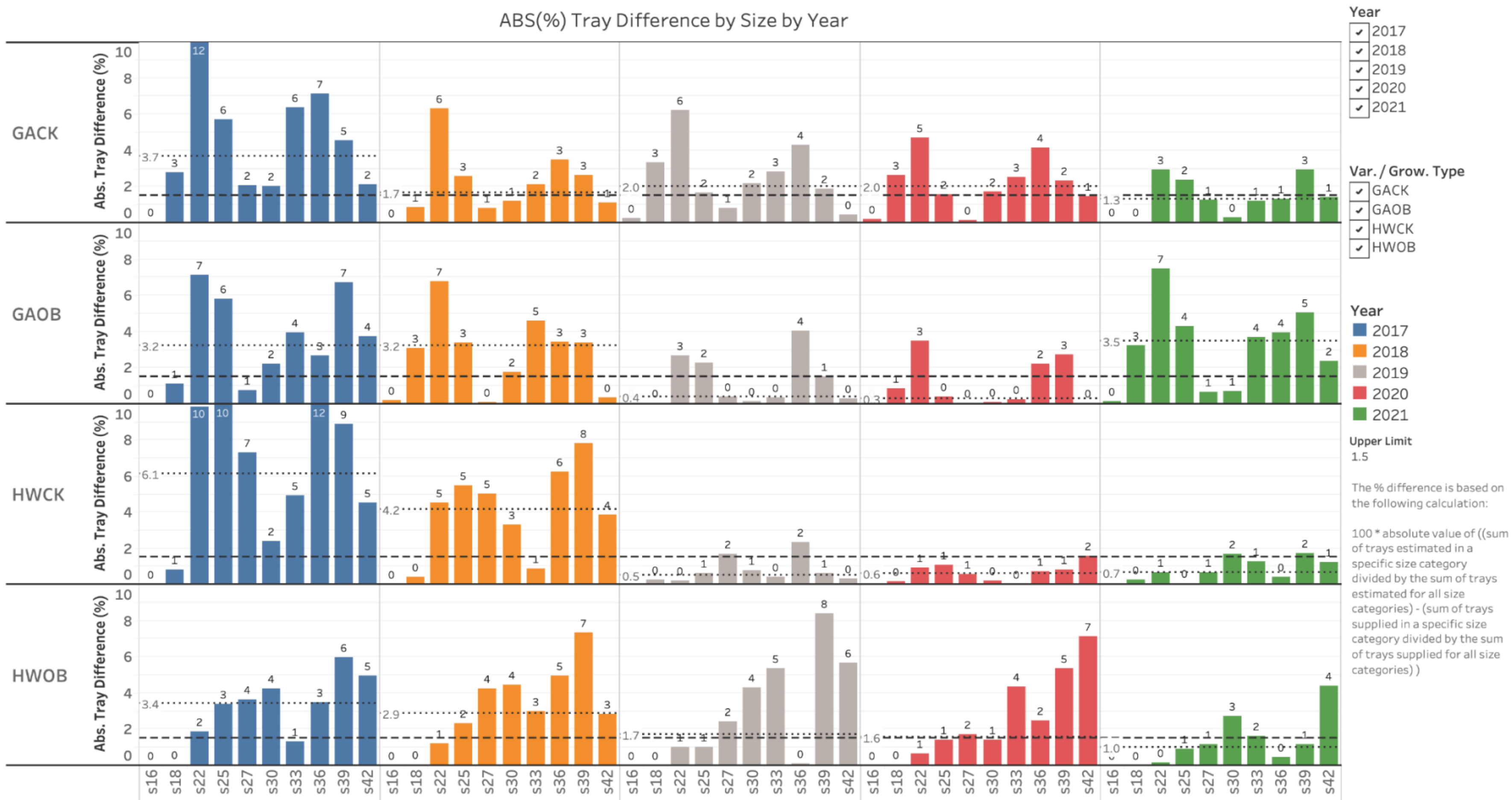


Figure 7: A breakdown of absolute % differences per size category, variety, growing type, and harvest season. The target threshold of 1.5% is represented by the black dashed lines. Median % error values per harvest season, variety, and growing type are represented by grey dotted lines and absolute median % values are displayed in each pane. Occasions where the coloured bars are within these dashed lines indicates that the size prediction is satisfactory according to Zespri's requirements.

Across harvest seasons, the median % size difference for Green Conventional fruit was within the required target threshold for 2019, 2020, and 2021 (Fig. 7). For Green Organic fruit, the median % size difference was within the target threshold for 2021 only but was very close to the target threshold in 2019 and 2020 (Fig. 7). For Gold Conventional fruit, the median % size difference was within the target threshold for 2021 but was only just outside the target threshold for 2018, 2019, and 2020. Gold Organic size estimates were clearly within the target threshold in 2019 and 2020 but outside of the threshold in 2017, 2018, and 2021 (Fig. 7). Overall, the machine learning models developed within the research project demonstrate the clear potential of this technology to achieve the required accuracy thresholds for total trays estimated by size category for each variety and growing type.

Figure 8: Weighted accuracy, averaged across all tray sizes and summarised by harvest year, variety, and growing type. The dotted lines represent the average accuracy over the 2017, 2018, 2019, 2020, and 2021 harvest seasons.

Weighted accuracies (Eq. 1) over the 2017 – 2021 harvest seasons ranged from 35.5% to 93.8% across harvest seasons, varieties, and growing types, with an overall average weighted accuracy of 74.3% (2017 – 2021) and 81.1% for the last 3 harvest seasons (2019 – 2021) (Fig. 8). Overall, when assessing the performance as a weighted accuracy across size categories, the 4 growing type and variety combinations were all very similar (Fig. 8). The models employed to make size and count predictions were complex and clearly benefit from additional training data when assessed by the total trays estimated per size category. With regards to weighted accuracy, while there was no reference target available from Zespri, it remains a useful metric for review.

CONCLUSIONS AND RECOMMENDATIONS

This research project has successfully demonstrated the capacity for weather-based machine learning approaches to forecast an estimate of kiwifruit supply (count and size) within the desired accuracy targets set by Zespri. This is an encouraging result for the proof-of-concept research project and has also been a valuable experience in uncovering opportunities for further refinements and improvements to the modelling methodology and the datasets used to achieve these outcomes.

Key recommendations that may further improve the outcomes of this work following the conclusion of this project include:

1. A shift from NIWA weather data to a more comprehensive weather service and climate forecaster such as IBM’s The Weather Company will be useful in strengthening the models’ ability to adapt to intra-seasonal climatic changes across key production regions. Increasingly accurate weather data and medium- and long-term predictions of climate throughout the growing season will help to improve accuracy and reliability of delivered national supply estimates. More granular weather zoning relative to NIWA data may also enable improved regional and kpin-level prediction accuracies.
2. The project has tested the ability to integrate complimentary datasets such as smart monitoring data into the machine learning models with positive indications of accuracy uplift, however to-date the models have only received limited in-field data, predominantly pertaining to fruit quality. With access to historically collected in-field observations, data such as flower counts and fruit density, we would be able to add

this information complementarily to the weather metrics and test its impact on kpin and national supply estimates.

3. The project worked with some select packhouse data and planned methodologies to derive predictions of fruit rejection rate, which would unlock additional insight for the predictions in estimating both export and non-export quality fruit. There was deemed insufficient data so far to complete this testing thoroughly, so additional work in partnering and data sharing with packhouses may be useful in achieving an accurate prediction of rejected fruit.
4. Additional machine learning methodologies are available that may likely increase the national count estimate and size profile performance of the models. Several important learnings from The Yield's commercial work with Wine and Berry customers are likely to be beneficial to both national count and size estimations.
5. This project was limited in the data sources that it tested, other data sources that drive variability in kiwifruit harvests can be integrated to improve the overall performance of count and size estimates. Where data sources are not available historically, they can be integrated through functional or biophysical methods as an additional post processing pipeline step.

NEXT STEPS

Following the conclusion of this project The Yield Technology Solutions successfully facilitated an ongoing commercial and research partnership with Zespri International, in a deal which promises to further extend the innovative delivery of national and regional kiwifruit supply forecasts.

This marks a significant success story for all partners involved in the collaboration and adds a further major landmark in the continued global expansion of Australian agtech services.

In February 2025, Yamaha Motors announced the launch of Yamaha Agriculture, completing the acquisition of The Yield Technology Solutions and Robotics Plus as part of this process.

Kiwifruit Yield Prediction

By Dr Evan Webster, Dr Jianlong Zhou, Dr Zhidong Li, Dr Merlin Zhang, Jichao Kan & Ashley Rootsey

January 2023

Project No. FA049

hello@foodagility.com
@FoodAgility
foodagility.com