Heuristic model of packout applied to the 2025 Rockit™ apple fruit set

July 4, 2025

## Introduction

Fruit yield (packout) prediction is an important measurement for Rockit™ apple as it enables the estimation of product available for sale to customers and thus revenue and profitability. Fruit yield is very complex to predict, *a priori*. It is inherently non-linear and is affected by many different factors including: growing conditions, climate, harvest timing, storage and shipping. The key variables may be difficult to understand and measure.

One approach to predicting packout is through a multivariate model which uses modern supervised learning approaches to determine a quantitative outcome for yield (or any other dependent variable of interest) based on a relevant set of predictor variables. Depending on the method selected the model can provide good accuracy of prediction as well as good inference into the underlying drivers of the model. The advantages with this type of apporach is that it incorporates a range of predictor variables which, if properly fitted, will explain the model for a broad range of conditions. The disadvantage of the machine learning approach is the complexity of assembling and configuring the model.

An alternative method is to observe historical data and apply a best fit curve to the data using an appropriate regression method (heuristic approach). The advantage with this apporach is it is simple and relatively easy to configure. The disadvantage is that it generally relies on one (or a small set) of predictor variables and may not explain the underlying variance for all conditions.

Despite the limitation of the heuristic will be used for the following analysis.

## A heuristic approach to modeling packout for Rockit™ apples

### Regular atmosphere (RA)

Based on six years of observation packout (yield), when stored in regular atmosphere, as a function of storage days is given by [Equation 1](#eq-pomod).

Where:

= yield (packout as a %)  
 = amount the packout declines from the starting value to the asymptote  
 = rate of decline of packout  
 = storage days  
 = asymptotic value of packout

The curve can be fitted using non-linear least squares regression using the R base stats programme. Confidence intervals are calculated using boot-strapping and the NLSBoot package in R.

### Controlled atmosphere (CA)

CA fruit behaves quite differently to fruit stored in regular atmosphere (RA). CA rooms are typically small and contain suffient apples for one or two days of packing. Because CA rooms are opened sequentially, there is very little deterioration in primary packout observed over the CA packing period. Historically CA packouts have been modeled as constant (i.e. they do not change as a function of storage days). Typically there is an uplift from the final RA packout to the CA packout.

### Using 2024 packout data as an example

The 2004 packing data set provides a good example of the model described above. The following section uses the 2024 packout data to explain the application of the approach to an actual data set.

|  |
| --- |
| Figure 1: Packout as a function for storage days and storage type for all batches packed at Te Ipu in 2024 |

The individual batches are plotted on [Figure 1](#fig-Model2024), with RA fruit in green and CA fruit in red. The modeled packouts are overlaid as the solid lines and the corresponding confidence intervals shown as grey shaded area around each model line. Note the large overlap between the CA and RA packing in terms of storage days for 2024.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Results for non-linear regression of 2024 regular atmosphere (RA) data   | **Coefficient** | **estimate** | **p.value** | **Significance1** | | --- | --- | --- | --- | | α | 0.205 | 7.47e-23 | \*\*\* | | β | -0.096 | 8.13e-10 | \*\*\* | | θ | 0.655 | 0.00e+00 | \*\*\* | | 1NS = not significant, \* = P < 5%, \*\* = P < 1%, \*\*\* = P < 0.1% | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Results for the linear regression of 2024 controlled atmosphere (CA) data   | **Coefficient** | **estimate** | **p.value** | **Significance1** | | --- | --- | --- | --- | | Intercept | 0.718 | 6.44e-40 | \*\*\* | | Slope | 0.000 | 6.14e-01 | NS | | 1NS = not significant, \* = P < 5%, \*\* = P < 1%, \*\*\* = P < 0.1% | | | | |

In terms of the quality of the model, the summaries are presented in tables [1](#tbl-RARegression) and [2](#tbl-CARegression). For the non-linear regression ([Table 1](#tbl-RARegression)) all the coefficents are significant and the model fit is relatively good. The number of storage days to get to within 1% of θ is 36 days which is a measure of the rate of decline in packout with storage days.

The assumption for the packout behaviour of the CA fruit is that it is constant for the period of packing. [Table 2](#tbl-CARegression) validates this assumption with the linear model showing a zero slope and moreover without any statisical significance (i.e. the hypothesis that the slope is zero cannot be rejected). The CA uplift is generally defined as the difference between the CA packout (assumed as constant) and the final RA packout (taken as θ for the case of 2024). This can be calculated as 6.29% and is close to the long term mean uplift of 4%.

## Analysis of 2025 data

The above approach can be applied to the 2025 data set before packing has been completed, recognising that there will be increased uncertainty in the model where there is no actual data. From a model perspective the season can be divided into three periods:

1. Fruit already packed
2. RA fruit to be packed
3. CA fruit to be packed

### Assumptions for the 2025 model

1. Model coefficients are calculated using existing data. Uncertainty can be quantified using confidence intervals (calculated using bootstrap methods).
2. CA packout will be assumed to be constant and based on the long term mean uplift of 4% (described above). The CA uplift baseline will be taken as the estimated (i.e. expected value) final RA packout from the non-linear model.
3. Input mass (kg) of fruit to be packed will be calculated by the product of the number of bins and the weighted mean bin mass.
4. The future packing programme will be taken from the latest view of S&OP. This is detailed below.
5. The pack dates and packouts are evaluated weekly (based on weekly bins packed) and the storage days are calculated from weighted average harvest dates for each of the storage types (CA & RA).

The three compenents described above are assembled and a final estimated seasonal packout is predicted.

### Determination of the mean bin mass and harvest date

For the calculation of input mass of fruit yet to be packed, the number of bins is multiplied by the mean mass per bin which is determined by dividing the total input mass of fruit packed in 2025 by the number of bins tipped. The gives a mean bin mass of 426 kgs.

Storage days are also required for the model (independent variable) for bins that have yet to be packed. While a future pack date can be specified the storage days also requires a harvest date (storage days is defined as pack date - harvest date). For this a weighted mean harvest date can be determined from the all the bins to be packed. A separate harvest date for CA and RA are calculated and presented in [Table 3](#tbl-WAHarvestDates). Note there are 8 days difference between the two dates which will impact the storage days for each storage type.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 3: Weighted average harvest date for all bins to be packed in 2025 (as of 04 Jul) by storage type   | **Storage type** | **Harvest date** | | --- | --- | | CA | 2025-03-05 | | RA | 2025-03-13 | |

### Status of packing at the time of the report

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4: Bins packed and to be packed by storage type in 2025 at 03 Jul 2025   | **Packing site** | **Bins tipped** | **Input** | **Reject** | **Packout** | | --- | --- | --- | --- | --- | | **tonnes** | **tonnes** | | Sunfruit Limited | 6,508 | 2,767.4 | 388.1 | 85.98% | | Te Ipu Packhouse (RO) | 21,476 | 9,154.8 | 2,073.7 | 77.35% | | **Total** | **27,984** | **11,922.2** | **2,461.8** | **79.35%** | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 5: Field bins cuurently in storage waiting to be processed as of 03 Jul 2025   | **Storage site** | **CA** | **RA** | | --- | --- | --- | | Berl Property Ltd | 5,501 | 41 | | Kirkwood Road | 495 | 14 | | Kiwi crunch (FV) | 140 | 45 | | Sunfruit Limited | 5,192 | 4,145 | | Te Ipu Packhouse (RO) | 12,191 | 7,890 | | **Total** | **23,519** | **12,135** | |

### Packing programme

A packing programme for the rest of the season was suuplied by S&OP and gives the weekly bins tipped by packing facility (i.e. Te Ipu, Sunfruit and Kiwi Crunch). The packing plan is presented as a bar chart in [Figure 2](#fig-PackProgramme1).

### Packing scenarios

Based on fruit quality risk and market demand three different packing scenarios have recently been proposed. A key aim of this analysis is to attempt to model each scenario in terms of overall packout and the timing of the packing programme (i.e. transition from RA to CA and the estimated date of final packing). The three scenarios are:

1. Base ~ all bins are packed (65,178)
2. Base ~ 8,000 - 8,000 RA bins are removed from the Rockit™ apple export channel
3. Base ~ 10,000 - 10,000 RA bins are removed from the Rockit™ apple export channel

The draw down of both RA and CA bins is shown schematically in [Figure 3](#fig-ScenarioPacking) for each of the three scenarios

|  |
| --- |
| Figure 2: Future pack programme by week and packing site for the remainder of the 2025 season |

|  |
| --- |
| Figure 3: Draw down of RA and CA bins for the three packing scenarios. This is based on the packing programme shown in [Figure 2](#fig-PackProgramme1) |

### Calculation of RA to CA transition for each packing scenario

The theoretical transition from RA to CA fruit can be calculated for each scenario using the packing programme given in [Figure 2](#fig-PackProgramme1). These are listed in [Table 6](#tbl-RACATransition) along with the minimum and maximum storage days for each of the different storage conditions.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 6: Estimated CA start and finish weeks and minimum and maximum storage days for RA and CA packing programmes   | **Scenario** | **CA** | | **Storage days** | | | | | --- | --- | --- | --- | --- | --- | --- | | **Week** | | **RA** | | **CA** | | | **Start** | **End** | **Start** | **End** | **Start** | **End** | | Base | 37 | 49 | 110 | 180 | 188 | 272 | | Base-8,000 | 31 | 46 | 110 | 138 | 146 | 251 | | Base-10,000 | 29 | 46 | 110 | 124 | 132 | 251 | |

The gap between the end of the RA storage days and beginning of the CA storage days for each scenario is due to the fact that The weighted mean harvest date for CA fruit is 8 days earlier than for RA ([Table 3](#tbl-WAHarvestDates)). Storage days should not be confused with packing date.

The packout behaviour for the three different scenarios can be visualised and compared in [Figure 4](#fig-ScenarioPackouts) which shows the scatter of the batches already packed and the modeled RA and CA as lines. Note the confidence intervals diverge for the RA storage days without much data.

|  |
| --- |
| Figure 4: Packout forecast as a function of storage days for RA and CA fruit. Three packing scenarios described above are detailed |

## Seasonal packout prediction

Based on the above analysis, a seasonal packout can be estimated by adding the three phases of the model:

1. RA fruit already packed (i.e. actual packout)
2. RA fruit yet to be packed (modeled using the non-linear function)
3. CA fruit yet to be packed

In terms of timing of the various phases of packing these are listed in [Table 6](#tbl-RACATransition). The input and export mass (export mass = input mass - reject mass) are listed in [Table 7](#tbl-inputexportmass) for each of the three phases and packing scenarios. The final estimated seasonal packout is given in [Table 8](#tbl-finalPackout). The seasonal result is calculated by summing the input and export masses for each of the packing phases and the calculating the packout on the aggregated total. As expected the net packout increases as the bins decrease.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 7: Input and Export mass for the three phases of the packing model and the three packing scenarios   | **Scenario** | **Already packed** | | | **RA to go** | | | **CA to go** | | | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **Input** | **Export** | **Packout** | **Input** | **Export** | **Packout** | **Input** | **Export** | **Packout** | | **tonnes** | **tonnes** | **%** | **tonnes** | **tonnes** | **%** | **tonnes** | **tonnes** | **%** | | Base | 11,922.2 | 9,460.4 | 79.4% | 5,169.9 | 3,305.5 | 63.9% | 10,019.9 | 6,647.6 | 66.3% | | Base-8,000 | 11,922.2 | 9,460.4 | 79.4% | 1,761.7 | 1,157.8 | 65.7% | 10,019.9 | 6,874.3 | 68.6% | | Base-10,000 | 11,922.2 | 9,460.4 | 79.4% | 909.6 | 604.1 | 66.4% | 10,019.9 | 6,975.3 | 69.6% | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 8: Final seasonal packout for each of the different scenarios defined above   | **Scenario** | **Seasonal** | | | | --- | --- | --- | --- | | **Input** | **Export** | **Packout** | | **tonnes** | **tonnes** | **%** | | Base | 27,112.1 | 19,413.5 | 71.6% | | Base-8,000 | 23,703.8 | 17,492.6 | 73.8% | | Base-10,000 | 22,851.7 | 17,039.9 | 74.6% | |