Final report

Algorithm hosting platform use and proof of concept

Project code: P.PSH.2201

Prepared by: Shawn McGrath (1), Dave Swain (2), Will Swain (2), Ashad Kabir (1), Arafat Hossain (1) and Lihong Zheng (1)

(1) Charles Sturt University; (2) Terracipher Pty Ltd

Date published: 26 June 2023

PUBLISHED BY
Meat & Livestock Australia Limited
PO Box 1961
NORTH SYDNEY NSW 2059

This is an MLA Donor Company funded project.

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

Abstract

This project demonstrated how a combination of technologies can facilitate user-controlled data sharing and assigned data ownership to produce insights using real-world data from red meat supply chain partners. The aim of the project was to deliver a working example of how an integrated digital data platform that connects producer's data directly with data analytics researchers, can lead to better adoption of research and development outcomes. The project showcased to producers the potential for automated data flows to feed directly through to research algorithms, to help them make data driven decisions.

Six producers representing seven different Property Identification Codes (PICs) engaged with the project, signing up through AgriTrakka to share their "real-world" data with data analysts from Charles Sturt University (CSU). The dataset used for analysis includes 3,231 unique cattle RFID tags from five properties and 45,313 individual weight records, with 2,197 cattle accessing an automated paddock weighing system during the data collection period.

The 'AgriTrakka' product now provides project participants with the ability to access and direct automated in-paddock weigh data, plus uploaded spreadsheet data from yard weights through to the project data storage. Algorithms developed within the project are now hosted on the Shaipup algorithm hosting platform and are available for producers to access, while still undergoing continued refinement and improvement. Reports are being prepared for the producers who contributed to the project, providing them with feedback and insight on their data using the algorithms developed.

1 Executive summary

Background

The increased use of agricultural technology on farms is facilitating increased data collection on individual properties; however, the full potential to utilise this data to inform management decisions is not yet being realised. This project used the example of cattle live weight data collected on multiple farms to demonstrate how a combination of technologies can facilitate user-controlled data sharing and assigned data ownership to produce data driven insights between red meat supply chain partners. The project showcased to producers the potential for automated data flows to feed directly through to research algorithms to help producers make data driven decisions. The specific algorithms for use were identified as part of this project.

Objectives

The aim of the project was to deliver a working example of how an integrated digital data platform can lead to better adoption of research and development outcomes. The project showcased to producers the potential for automated data flows to feed directly through to research algorithms to help producers make data driven decisions. Six producers signed up to AgriTrakka and turned on data flows for an automated in-paddock data collection source (Optiweigh). The combined data set incorporated data from five separate properties for analysis. The data was used by researchers to develop three algorithms with a fourth algorithm for prediction of sale dates or weights based on cleaned data sets currently under development and expected to be delivered soon. Two producers have indicated their interest in further involvement with the project and research team. This demonstrates that the approach modelled within this project can assist producers to answer specific questions or overcome the problems they are having with aligning weights from in-paddock and in-yard data. There is an opportunity to continue to refine algorithms for the growth prediction tool that will be more universally relevant as more data, including across multiple properties and years, is made available to the project team.

Methodology

The team operated on a fast iterative approach with a weekly team meeting cadence. There were several phases to the project including:

- Implementing and testing the cloud computing capabilities and data pipeline to progress algorithm development;

- Exploring in paddock weighing and opportunities to increase accuracy and prediction for cattle weights. This was focussed predominantly on the data set from the CSU farm; and

- Producer engagement. This included recruitment of producers, getting producers to sign on to AgriTrakka and direct data to the data analyst, and development and refinement of algorithms hosted on Shaipup; development of the model for a market readiness tool; and designing an auto-populated report to provide direct feedback to producers on their own data.

Results/key findings

Six producers representing seven different Property Identification Codes (PICs) engaged with the project to sign up through AgriTrakka and share their data with data analysts from Charles Sturt University (CSU). The dataset used for analysis includes a total of 3,231 unique cattle RFID tags from 5 properties and a total of 45,313 individual weight records, with 2,197 cattle using the automated paddock weighing systems throughout the data collection period.

The 'AgriTrakka' product now provides project participants with the ability to access automated in-paddock data weighing plus uploaded spreadsheet data from yard weights and direct these weights through to the project data storage warehouse. Algorithms developed within the project are now hosted on Shaipup (as a component of AgriTrakka) and are undergoing continued refinement. Reports are being prepared for the producers who contributed to the project, providing them feedback and insight on their data using the algorithms developed. A model for market prediction tool based on linear growth rates has been developed and it is expected this base model will be delivered to Shaipup soon. It has been identified that more data will be required across more years to refine this model for non-linear cattle live weight gain for individual farms.

Benefits to industry

Within 12 months the project developed and demonstrated a data sharing infrastructure and algorithm hosting platform which is now available for industry. The infrastructure demonstrated how it is possible to improve product development and research outcomes through aggregation of data from a combination of agricultural technologies. Algorithms for data standardisation and data cleaning are now available on the Shaipup website. This project has improved the understanding of limitations of in-paddock and in-yard weighing and deployed algorithms which can improve the confidence producers have in the data that their on-farm live weight measuring devices have generated.

Future research and recommendations

The technology explored in this project is relatively new and producers are still exploring how accurate the data they are receiving is and how they can apply it within their production system. One large producer who had identified a large discrepancy between in-paddock and in-yard weights is seeking on-going engagement with the project team to improve their understanding of the accuracy of data. We believe that the approach to the problem used in this project, which considers population dynamics, can assist with cleaning data and helping producers make best use of the technology for tracking cattle growth. The project identified the need for mechanisms to incentivise producers to utilise the tools and algorithms developed in this project. The development of an industry benchmarking tool is one concept we are seeking to explore in a future project.

Table of contents

2 Background

A previous project "Trakka: Making Data Flow", which concluded in January 2023, developed and tested data-sharing infrastructure (AgriTrakka) to technology readiness level (TRL) 8. This infrastructure allowed producers to share their data with service providers, with the ability to control the data flow by easily turning on and off the dataflows. Next steps identified for the infrastructure were refining the dashboard and to move towards user adoption, including integration of data publishers and data subscribers. A new algorithm hosting platform, Shaipup, was in development. A use case was selected that combined: newly available technologies on the Charles Sturt University (CSU) Global Digital Farm; the data analytics and algorithm development skills within the project team; and the infrastructure now available in AgriTrakka and Shaipup. The current project therefore sought to leverage off other emerging technologies to provide a targeted use case scenario by integrating a series of algorithms that support livestock performance data analyses, and assess the quality of the data, aiming to deliver a livestock market readiness prediction tool.

Use case

The increased use of technology on farms is facilitating increased data collection on individual properties. Many cattle producers now have permanent scales installed in their yards. These "in-yard" systems collect "static" weights, that is, cattle stand on the scales as they move through a race and a single weight is recorded for the event. There is usually a long period between weighing events with gaps of weeks or months between each weighing event. More recently there has been development of "in-paddock" weighing systems such as Walk over Weighing (WOW | Tru Test Datamars (farmingmadebetter.com) and Optiweigh (In-paddock livestock weighing | Optiweigh | Armidale). These systems allow liveweight data to be captured in the paddock without needing to yard the cattle and allows for more frequent live weight measurements. The increased frequency of measurement increases the amount of data being generated on farms that have invested in this technology.

Despite generating large amounts of data, the full potential to utilise this data to inform management decisions is not yet being realised. These data have intrinsic variability both between animals in a group and between measurements over time. Therefore, producers need to be able to assess the accuracy of the data to make informed decisions. Furthermore, for farmers to share data means that they lose control/ownership of it, and it is difficult to share. Most data generated tends to remain within the digital infrastructure of the ag tech service provider through which it was generated. By sharing data with researchers, farmers can contribute to the development of useful algorithms that can overcome some of the above problems such as assessing data accuracy. There is also potential for farmers to derive value from data generated on their farm by sharing with external parties such as researchers and other service providers to contribute to the development of new algorithms and technologies.

3 Objectives

The project aimed to utilise data generated on farms from in-paddock (a series of weight records collected automatically over a short interval and then averaged) and in-yard weighing systems to demonstrate how producer-controlled sharing of data can provide tangible outcomes for individual farms and the wider industry. The two key objectives were:

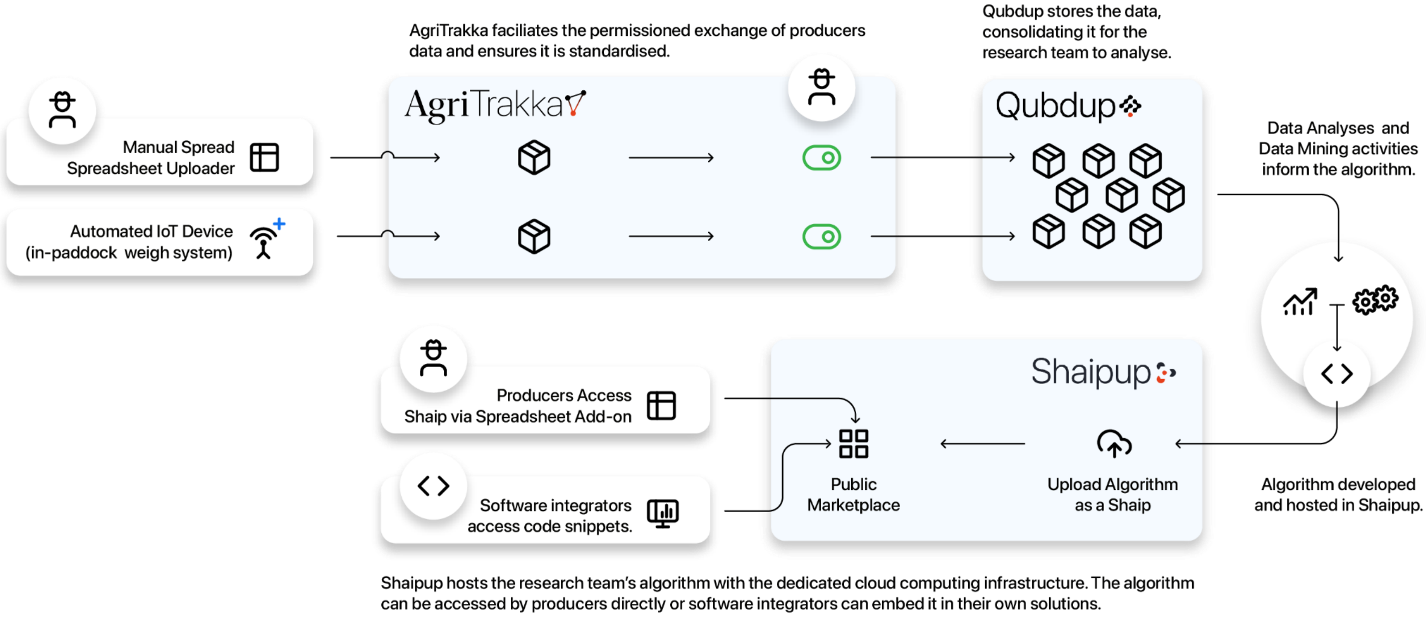
- To demonstrate the viability and user acceptance of using algorithms and data to make predictions of market readiness using automated performance data recording.
- At least six producers to be involved in sharing data and using algorithms to predict market readiness.

4 Methodology

4.1 Developing cloud computing infrastructure

The cloud computing infrastructure is integrated within TerraCipher's software services. These services use cloud computing to deliver data sharing through AgriTrakka, algorithm hosting through Shaipup and data storage through Qubdup. Each of these services operate independently from one another, however, can now be accessed through a single sign on (SSO). This allows simpler access across the infrastructure and provides users the opportunity to integrate their data across the three core services. Figure 1 summarises how the infrastructure facilitates the permissioned exchange of producers' data from their in-paddock weigh systems and spreadsheet software through to the research project team. This allowed the team to analyse the data and develop a set of decision support algorithms that were made available to producers and software developers via the Shaipup marketplace.

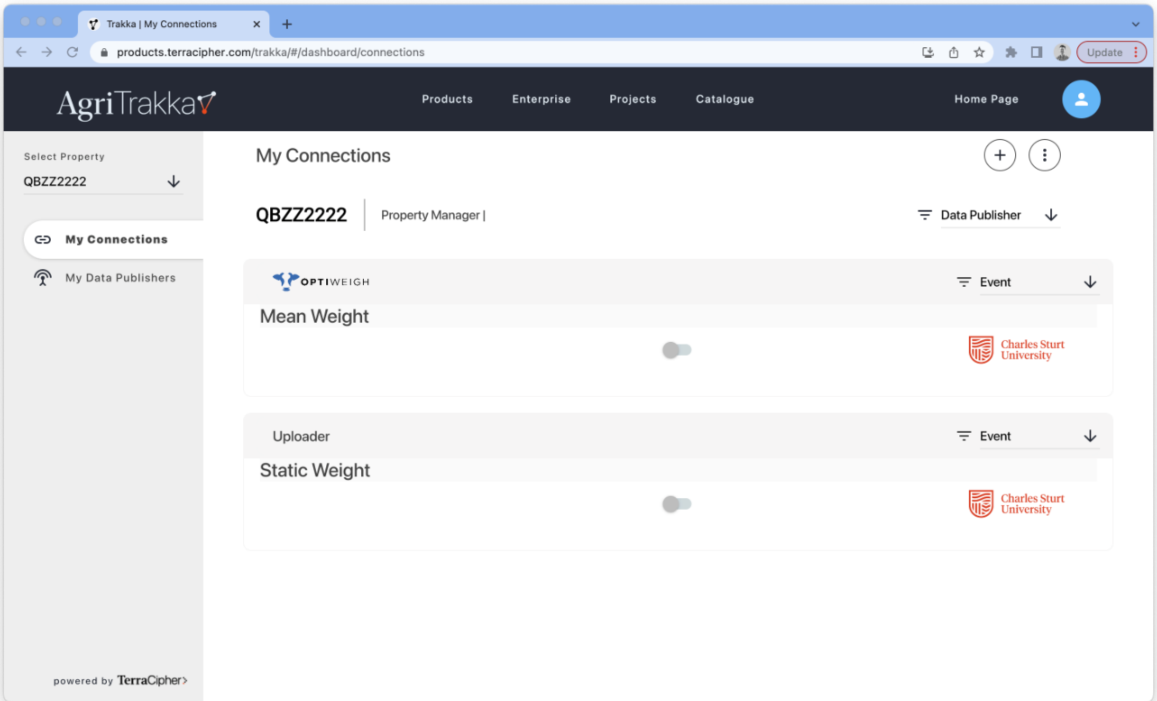
Figure 1 – An overview of TerraCipher's products showing they interact to facilitate data exchange for analysis and algorithm hosting for research translation and adoption.



4.1.1 AgriTrakka data sharing

AgriTrakka provides project participants with the ability to access and direct automated in-paddock weigh data, plus uploaded spreadsheet in-yard weight data through to the project data storage warehouse. Figure 2 provides an example of the AgriTrakka dashboard demonstrating the two published data flows ("Mean Weight" and "Static Weight") and the project endpoint.

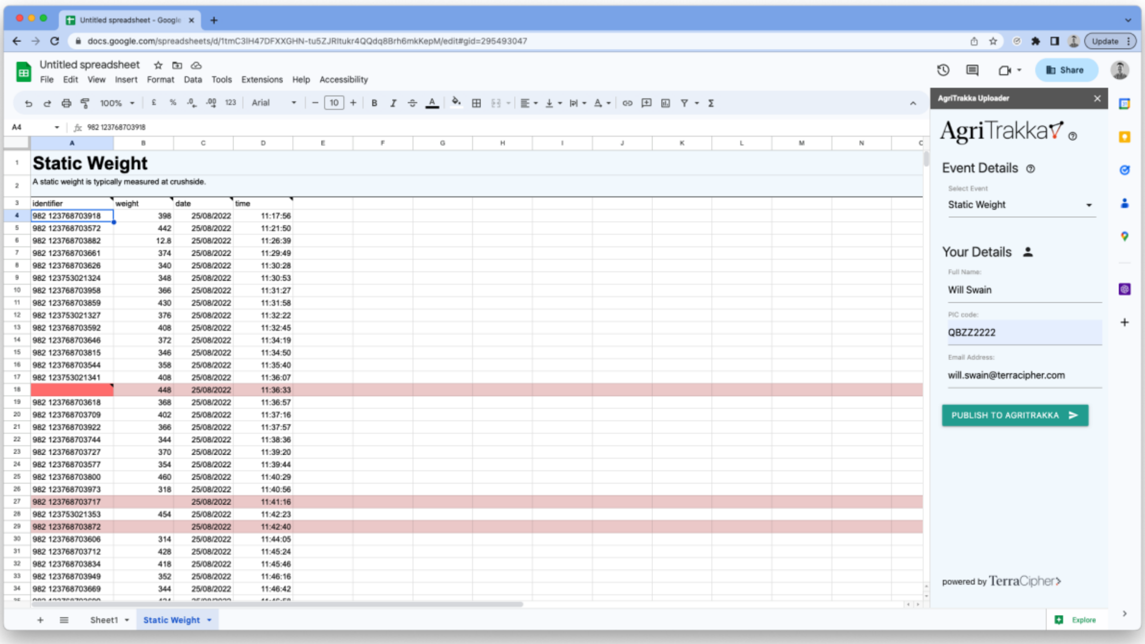
Figure 2 – AgriTrakka dashboard showing Optiweigh -> CSU connection option for Mean Weight events.



To facilitate easier uploading of spreadsheet data, the project team developed a web-based add-on for the Google Sheet platform (Fig. 3). This system integrates with AgriTrakka and provides users with the ability to upload data directly from their spreadsheet software. The application also ensures easy connection to AgriTrakka and provides a data check and formatting function to meet the Integrity Systems Company (ISC) animal data schema. The team has made this system freely available via the Google Workplace Marketplace.

At this stage users will need to use a free google email address to access the Uploader. While this doesn't pose a significant barrier to access, for the next phase TerraCipher is developing a web interface to allow farmers to access the algorithm directly through their Shaipup dashboard. A Microsoft Excel extension is also being considered for future developments.

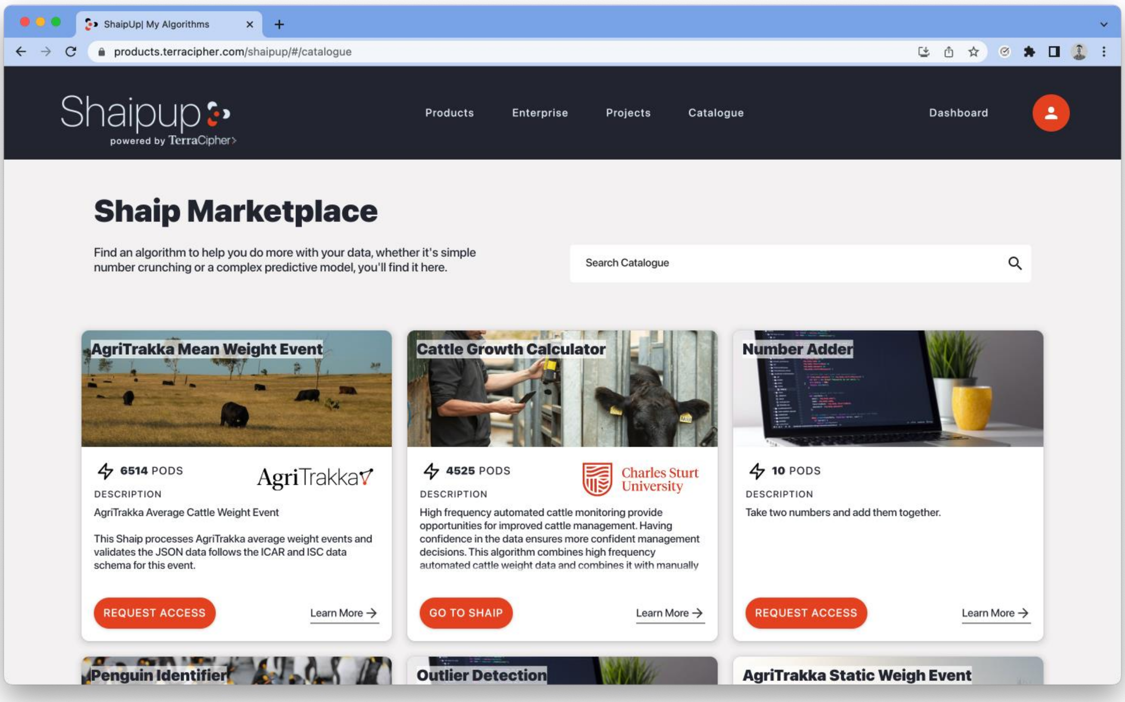
Figure 3 – AgriTrakka web-based Google Sheets spreadsheet uploader Add-on highlighting erroneous Static Weight data that doesn't align to the data standards.



4.1.2 Shaipup algorithm hosting

The Shaipup platform enables users to process their data using algorithms, models or calculators (Shaips) that are publicly available. Algorithms developed by TerraCipher, Charles Sturt University or other algorithm developers can be requested through the public marketplace in Shaipup. Once the request is approved, the user can pass data in and receive results. The project has implemented two initial algorithms that provide users the ability to ensure their data meets the ISC animal data schema standards for both yard weights and in paddock mean weights (a series of weight records collected automatically over a short interval and then averaged). Figure 4 shows the marketplace dashboard with the two AgriTrakka data standardisation algorithms available for users, along with one of CSU's cattle weights algorithms.

Figure 4 – Data standardisation algorithms available for users on the project. Algorithms (Shaips) are available through Shaipup on the Shaip Marketplace screen.

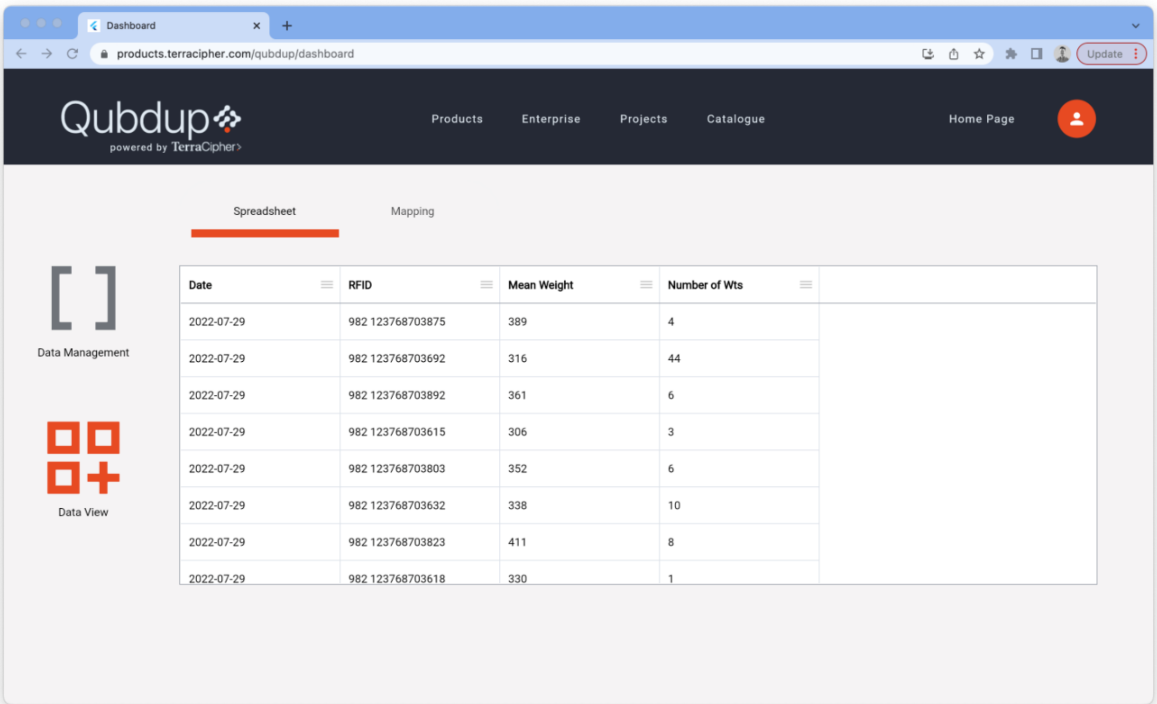


Once access is approved, the user can then access these algorithms, for example via their dashboard. The producers are able to utilise the algorithms, for example to clean their data, but are not able to alter the algorithms.

4.1.3 Qubdup data warehouse

The Qubdup data warehouse provides a repository for the project data, as well as allowing producers to access their own consolidated and standardised data. The dashboard integrates within the broader TerraCipher applications and allows producers and researchers to filter and download the data. Figure 5 shows the user dashboard with example data from the CSU farm.

Figure 5 – The Qubdup dashboard demonstrating the consolidated data.



4.2 Recruiting producers

The project team engaged with six producers representing seven separate properties. Six producers signed up to AgriTrakka and five contributed data within the time bounds of the project; this represents data from six properties. All producers had an Optiweigh system and agreed to upload static yard weights. To increase the data analytics opportunities the AgriTrakka infrastructure was able to access historic data collected prior to the producers signing up to AgriTrakka.

A data sharing agreement was put in place. This was reviewed by the CSU legal team and designed so that there was an initial summary that producers would see when turning on their data. This included a link to the expanded terms and conditions. The short summary for the sign-up page stated:

- The Data Sharing Agreement aims to ensure data is only shared with suitably approved project researchers (the Project Team).

- Producers retain ownership of their raw data and grant rights to the Project Team to access and use the data.

· Confidentiality of personal information will be maintained at all times.

· You may end your participation in the Project research and the Data Sharing Agreement by written notice to Charles Sturt University by contacting Project Lead Dr. Shawn McGrath shmcgrath@csu.edu.au. Upon receipt of written notice of your withdrawal Charles Sturt University will immediately halt the commencement of new project activities using your Raw data. Any activities already commenced which use your Raw data will continue, but your Raw data will be deleted once any existing activities utilising the data have been finalised.

· Raw data will be stored on a secure server at Charles Sturt University during the project.

· Raw data will be deleted from the server at the conclusion of the project (29 November 2023) unless advised by the Data Originator, in which case, data will be stored on a secure server for up to 7 years.

4.3 Data analysis and algorithm development

4.3.1 Preliminary Data Collection

The project started data collection from the Charles Sturt University (CSU) Global Digital Farm (GDF), which allowed the data pipelines to be tested as well as providing initial data for analyses. Eight months of Optiweigh data from a mob of weaner steers (n=230) on the CSU farm was made available to the data analyst. This included continuous collection of data available from the in-paddock weight unit including approximately 5,000 weight events collected between weaning in February 2022 and sale of steers in Spring 2022. In addition, there were four events where the cattle were weighed in the yards (static weights): 6 February 2022 (weaning); 10 May 2022 (prior to commencement of crop grazing); 1 July 2022 (when steers were vaccinated with Bovilis MH + IBR [Coopers Animal Health]); and 25 August 2022 (post-crop grazing/pre-sale weight).

The system provided daily monitoring of any CSU farm Optiweigh mean weight events. Static weights collected in the yards were uploaded by the project team using the Google Sheets platform developed for this purpose.

4.3.2 Data Analysis and Exploring Potential Algorithms

The data standardisation facilitated by AgriTrakka ensured data from different devices and properties could be integrated and easily analysed. The data analysis was initially run on local computers using both Python and R programming languages using a consolidated data set. The code development was managed through GitHub to ensure team members from different organisations could access the latest version and any developments and changes had a version number.

Before any data analysis was started the initial code checked the data, ensuring data types were consistent (e.g. dates, numbers or strings) and added in factors that could be used to subset and group the data for subsequent analyses e.g. time periods, properties or weighing system / weight event types.

The initial data analyses focussed on exploring relationships between “static” yard weights and in paddock weighing devices using the data from the Optiweigh most closely aligned to the week that the yard weights were collected. These analyses allowed the project team to evaluate the accuracy of the in-paddock weighing devices and determine the best method for identifying and removing outliers. Analysis of the static yard weights included power analysis, cluster analysis, outlier detection and regression analysis. The power analysis used the static weight data to determine the minimum number of cattle required to derive a representative group average. These results could then be applied to the in-paddock weight data. The cluster analyses and outlier detection was used to identify outliers in the static weights so they could be removed from subsequent analysis. Regression analysis was used with the static weight data to determine growth rate predictions between weight periods. Erroneous static weights for individual animals were identified by comparing individual cattle growth rates with the overall average.

The final analyses considered the value of the data for a ‘strategic market readiness’ algorithm. Initial work was completed using regression based on the in-paddock weight data to predict future cattle weights. In addition, the data including 3231 unique animals collected by collaborating producers during the period Fri 05 Jun 2020 to Sat 20 May 2023, was used to evaluate the regression and identify how the in-paddock weight data could be used to inform when conditions resulted in change points (where growth rates either increased or decreased) or to track overall growth against expected growth and identify points where management intervention was required to ensure cattle met market expectations.

4.4 Producer reports

The initial data analysis and reporting from the CSU farm was further developed to deliver a customised report based on a standard template (see Appendix 1) for each producer contributing data to the project. These reports provide the producer with an overview of their data. The standardised format will provide the opportunity to further develop benchmarking statistics and allow producers to evaluate how they can derive better value from their in-paddock weighing.

5 Results

5.4 Producer engagement and product acceptance

5.4.1 Producer participation

The project recruited six producers who signed up to AgriTrakka and contributed data to the project. Several challenges were identified through the sign-up phase of the project:

- While all producers were interested, there were differing levels of engagement. Highly engaged producers tended to contact the project team to ensure sign-up happened quickly and overcome any problems. Less engaged producers, while very happy to provide the data, were less likely to follow up to overcome any problems with sign-up and upload of data. This affected the pace of data flow.
- The data sharing agreement was accepted readily by most producers; however, corporate producers required more oversight from their own legal departments, which delayed data being made available to the project team.
- The short duration of the project stated on the original Data Sharing Agreement (30 May 2023) was also an impediment, and further consideration of wording of data sharing agreements into the future is required. The expiry date on the Data Sharing Agreement was extended to 29 November 2023 to match the contract between MLA Donor Company and CSU.

The engagement with producers did allow the project team to identify where the likely problems might be in sign-up and uploading of data, and fixes were implemented to overcome these. For less engaged producers, any impediment through the signup process was enough for them to disengage and not continue through the process. There were three points of friction which hindered most producers during sign up including:

1. Password strength validation, unique username and username format validation error messages made the sign-up process cumbersome. While these are important for cyber security reasons, it frustrated users and caused less engaged producers to give up on the process altogether. Clearer messaging was implemented to better guide users through the sign-up process.
2. During the PIC validation process during the setup of Trakka, there was often a misalignment between the NLIS account and PIC details that the producer had on record compared against the details on record in ISC's PIC registra. This was usually due to producers having multiple accounts registered across ISC products, with only one of the accounts being registered to the PIC. This was typically resolved through a trial and error approach between the different accounts.
3. Verification emails were sometimes identified as SPAM, or in some cases were never received, preventing the user from accessing their account. While a simple fix, this required manual verification from TerraCipher to resolve.

Due to the small number of participants, the project team engaged in video calls to guide the participants through the sign-up process, allowing the project team to immediately rectify any issues encountered. This also allowed the project team to receive feedback directly from participants.

The automated reports (see Appendix for two examples) provides a tool that will allow producers that share their data to get direct feedback. The initial report provides high level descriptive statistics, however, the framework enables more refined feedback to producers in the future and demonstrates the value for them sharing their data and accessing any of the algorithms.

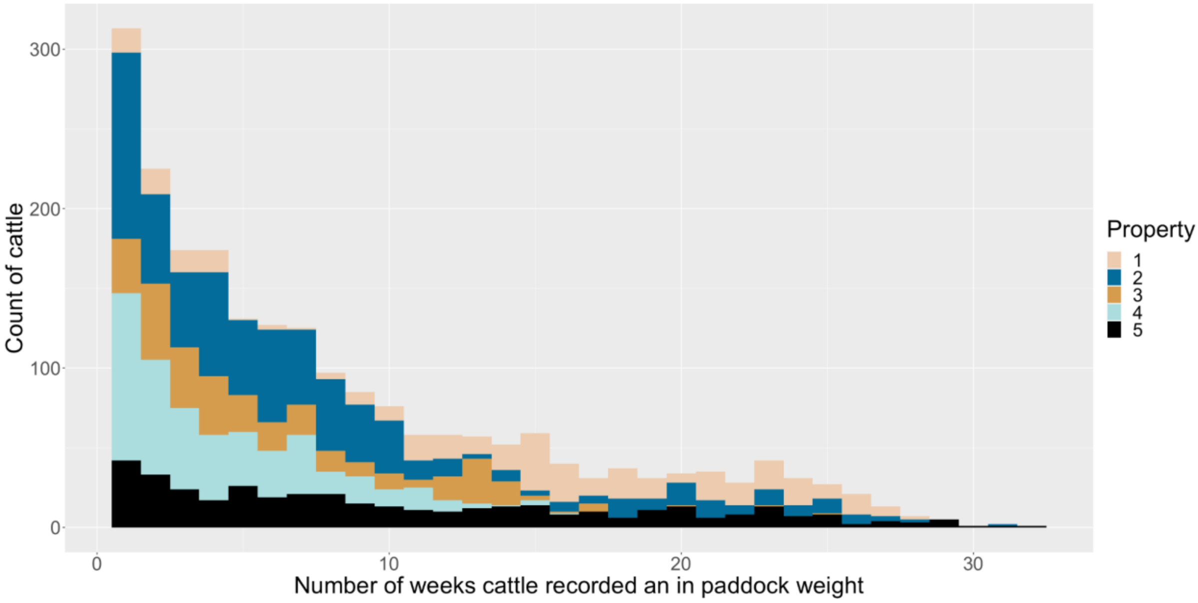
5.4.2 Overview of producer data collected in the project

The dataset includes a total of 3231 unique cattle RFID tags from 5 producers. A sixth producer had signed up but had not provided authentication to access the in-paddock weighing data by the cut-off date for data access for inclusion in this report. The data covers the time period from Fri 05 Jun 2020 to Sat 20 May 2023. The data includes single weights from cattle in the yards, also referred to as static weights, and automated weights from cattle in the paddock. These automated weights are averaged across multiple weights each day and input as a mean daily weight for each animal. The data set has a total of 45,313 individual weight records.

The initial analyses grouped weight data into weeks of the year providing more data for comparative analyses.

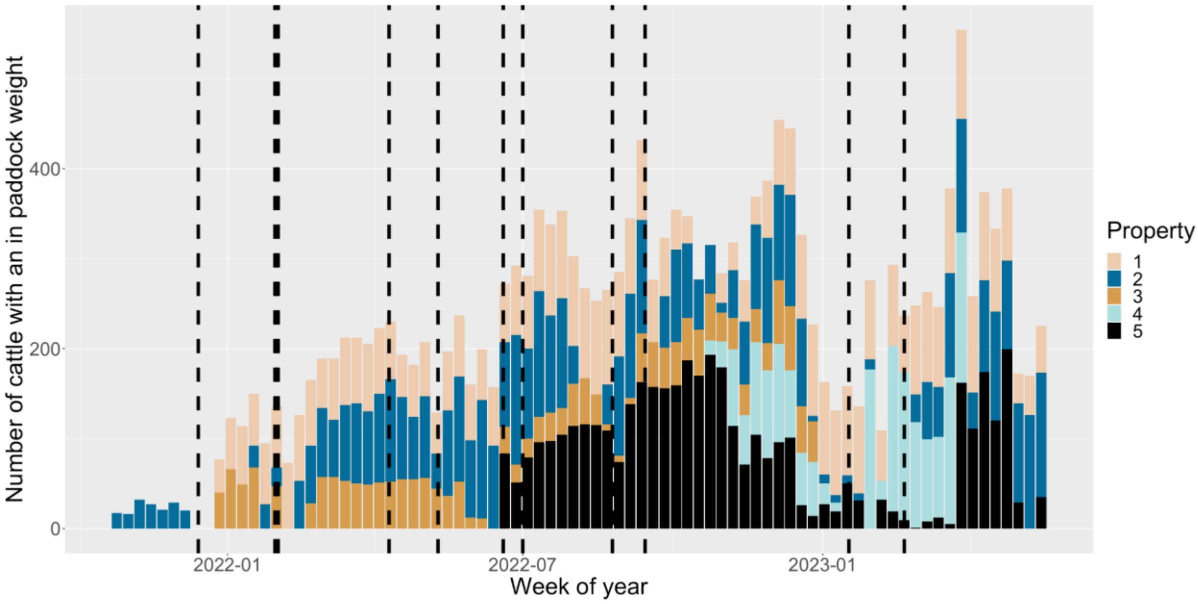
The automated paddock weights rely on cattle accessing an in-paddock weighing device. Not all cattle access this device and those cattle that do will have variable numbers of visits on any given day. The mean paddock weight includes additional data on the number of visits. The overall summary of activity of cattle that have mean weights shows there are 2,197 cattle that have used the system throughout the data collection period. The mean number of weeks that cattle recorded at least one weight during any given week is 8 and the maximum number for at least one animal weights is 32 weeks (see Fig. 6).

Figure 6 – Number of times individual cattle accessed Optiweigh



The overview of the weekly history and numbers of cattle that visited the in paddock weighing system and the timing of the static yard weights is shown in Figure 7. The black hatched lines on the graph show the weeks with a static weighing event.

Figure 7 – Weekly history of cattle weights, hatched black lines represent weeks where static yard weights were collected



5.5 Implementing and testing the cloud computing capabilities

The project developed and delivered both the dashboard and backend data management systems. These were initially tested using the CSU research farm data. There have been over 5,000 individual cattle events uploaded including 3,500 automated cattle weigh events.

The project has tested all stages of the user dashboard within the project team and with the producers who have signed up to the project, and these are all working. Interactions with producers allowed refinements to be made and bugs identified and rectified.

Given the goal of the project was to provide a pipeline of data that can be used to develop new algorithms that can be used by producers, the project team also tested the capability of the system to deliver some preliminary data analyses.

5.5.1 Data pipeline to progress algorithm development

The CSU initial data analysis was used to explore whether the proposed cloud computing is able to provide a data pipeline that can be used to derive insight. Critical to the data pipeline is capturing and standardising all data inputs. The Qubdup data storage retains all data from AgriTrakka. The raw JSON format data is retained with additional fields assigned to allow rapid queries that allow easy sorting. Customised queries are developed to provide appropriate data fields that can be used for data analyses. In addition, the query data is also checked to ensure there are no duplicates and that only data from the approved data supplier is available.

Comma separated value (csv) data format was identified as the preferred format to conduct the analyses. A scheduled daily query updates the csv file, and this is then accessed through a cloud data storage. The data is also passed to the Qubdup dashboard where it is available for display and download. Preliminary analyses from the CSU farm demonstrated the data pipeline and showed how useful it can be for data analyses.

5.6 Exploring in paddock weighing and opportunities to increase accuracy and prediction for cattle weights

5.6.1 Exploration of data using CSU farm data

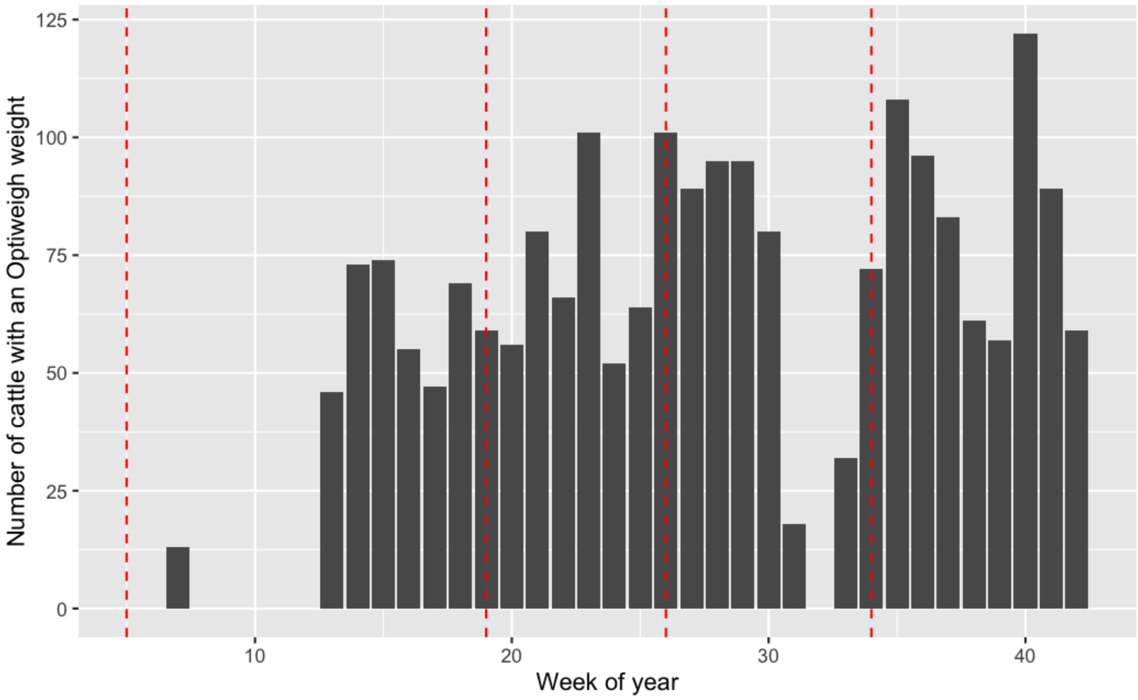
The CSU farm data were collected between February and October 2022 including Optiweigh automated weight data and four static weight events where cattle were manually weighed in a set of yards. The Optiweigh data were recovered using the Optiweigh API which provides single mean daily weight for any individual animal that accessed the Optiweigh front foot weigher. This required producers to contact Optiweigh giving permission for the Project team to access the data. In addition to the mean weight, the API also provides data on the number of individual weights that were used to determine the mean weight.

All data were processed through the TerraCipher AgriTrakka platform where both the static weights and the mean weights were standardised following the ISC animal data schema. The data standardisation ensured consistency in the data formatting and enabled the data to be stored in a Qubdup database facilitating rapid queries and formatting.

Preliminary data analyses were completed using the R programming language. The aim was to better understand the data and determine any patterns or opportunities to: (1) improve data quality by more standardised data filtering; and (2) to explore opportunities for using the Optiweigh data in a predictive algorithm.

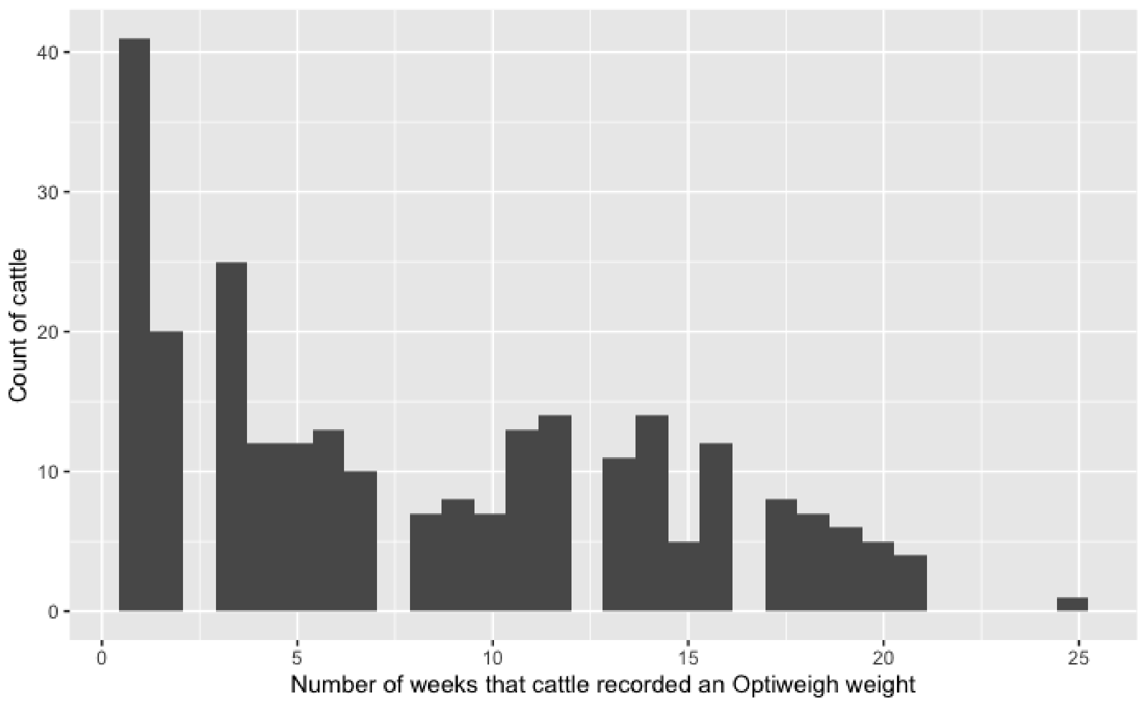
The dataset included a total of 444 unique cattle RFID tags. The initial analyses grouped weight data into weeks of the year providing more data for comparative analyses. Figure 8 shows the number of cattle using the Optiweigh unit in each week and shows that there was only one week where more than 50% of the cattle used the unit at least once.

Figure 8 – Cattle accessing in paddock weigher based on week of the year (2022). Red hatched lines indicate the weeks with static yard weights.



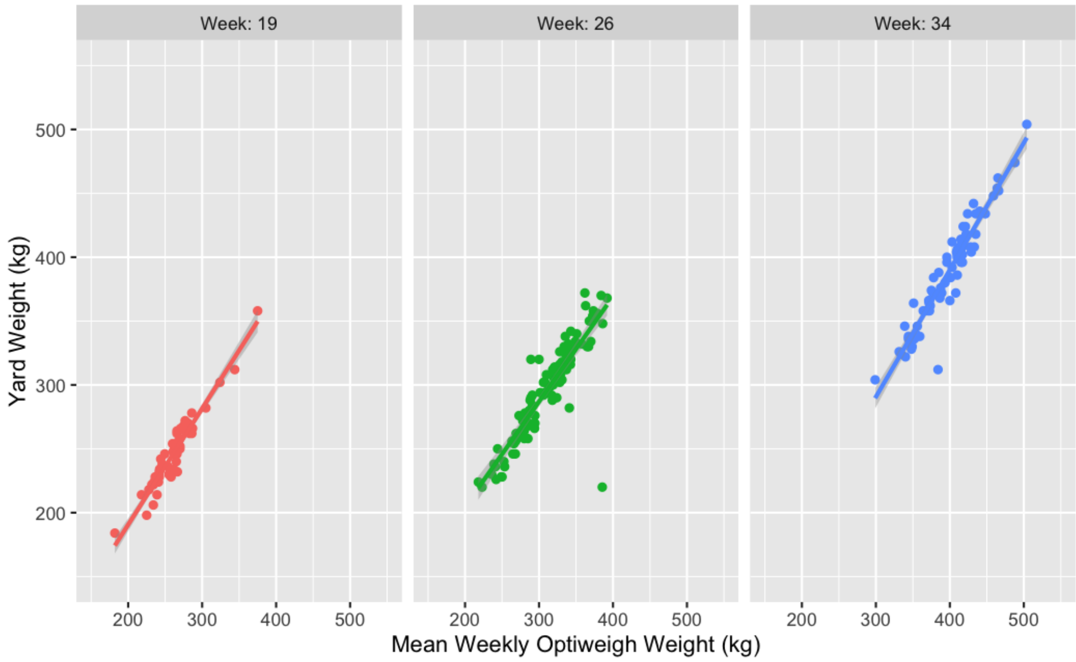
The overall summary of activity of cattle using the Optiweigh system shows 255 cattle that have used the system throughout the data collection period. The mean number of weeks that cattle accessed the unit is 7, with one animal having their data collected over 25 weeks. The histogram in Figure 9 shows there were many cattle that have one week's worth of data.

Figure 9 – Distribution of total weeks that cattle used the in-paddock weighing



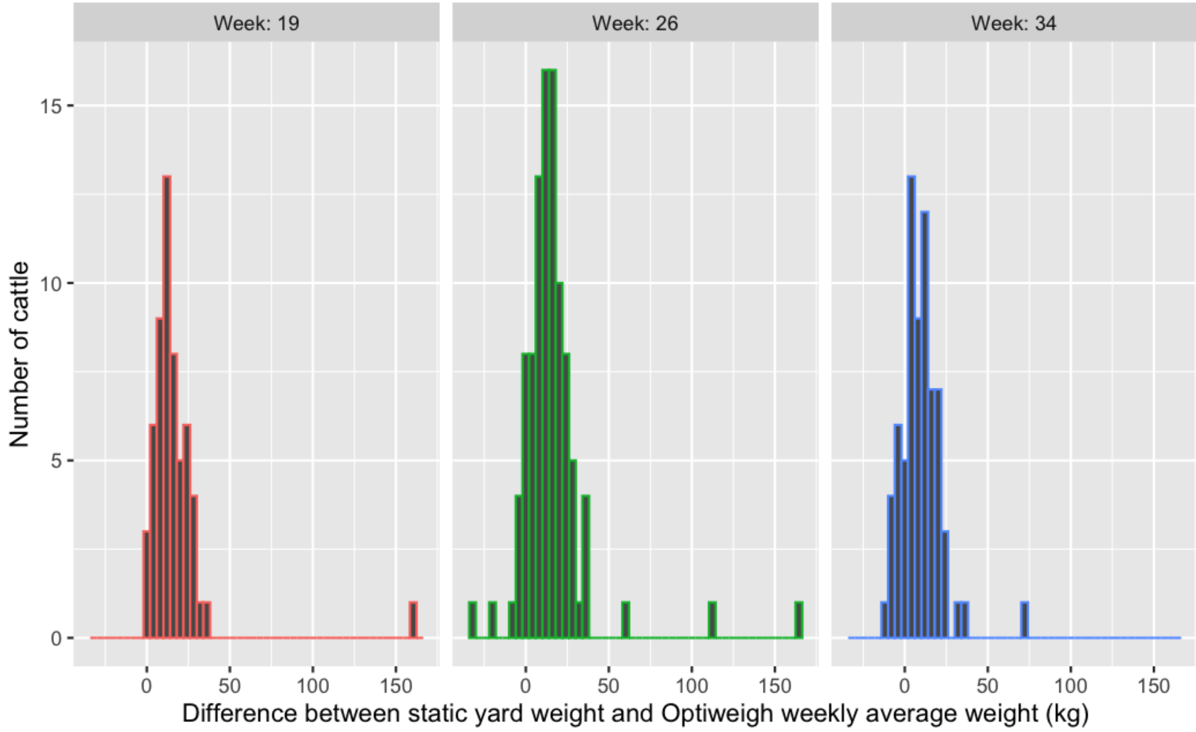
Preliminary data analyses compared the in-paddock weigh data with the individual weeks where static weights were recorded. For each static weighing event all cattle with weekly in-paddock weight records were identified. Figure 10 provides a scatter plot static weight against weekly paddock weights for each of the recording periods.

Figure 10 – Mean weekly weights from in-paddock weighting plotted against static yards weights with a line of best fit indicating the correlation



The difference between static and mean weekly in-paddock weights were calculated and the frequency plot of these differences is shown in Figure 11.

Figure 11 – Frequency plots of the differences between static yards weights and paddock weights.



Overall, there was good agreement between the in-paddock weigher weekly average weights and the static yard weights. It should be noted that the individual daily paddock weights were averaged for each week of the year. There was no data filtering based on additional information regarding the number of weights that were used for the daily average and this averaging method may be improved by putting minimum thresholds for the number of times an animal visits within a week. There are cases where cattle may only have one daily weight. It is likely more data points for an average weight will increase the accuracy. It was identified that an algorithm that removes potential outliers from the paddock weighing data should reduce differences between the static weights and paddock weights, and this was developed in the next phase (below).

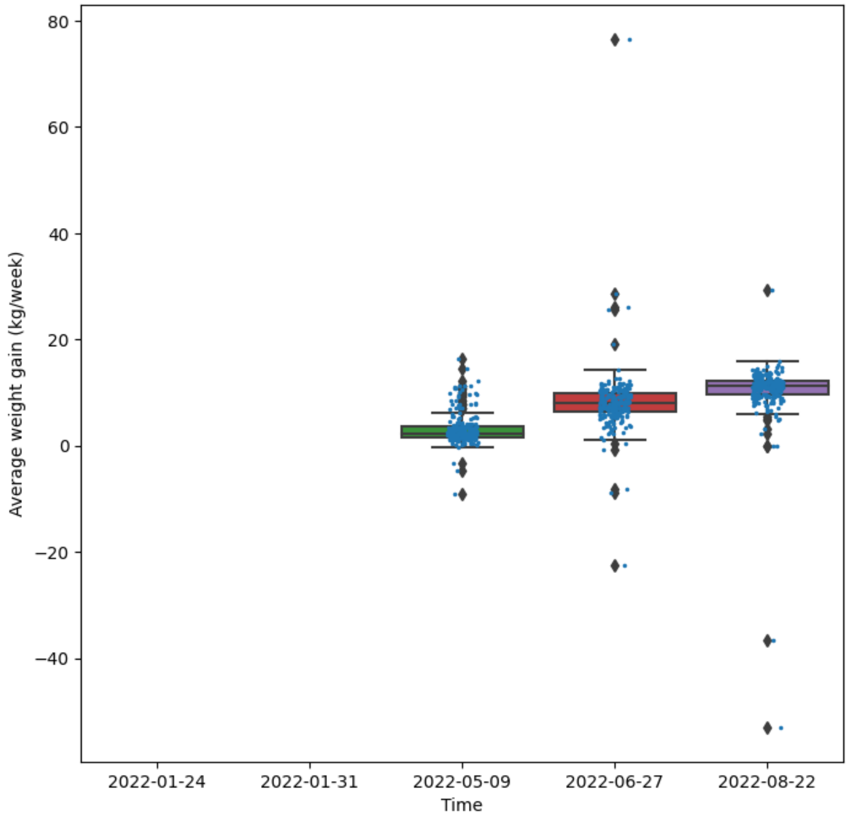
These data analyses provided an opportunity to test the value of the data and explore how the cloud infrastructure and workflow can deliver to the core project objective linked to enabling the delivery of algorithms that can help cattle producers.

5.6.2 Algorithm development and deployment on Shaipup

Using robust population statistical methods to remove potential outliers was identified as a critical goal. The first algorithm hosted by Shaipup enables producers to check their static yard weight data and determine not only outliers but also sub-groups of cattle in the larger mob.

It was identified early in the project that the static weight data that was being used to clean the in-paddock weight data and establish population dynamics also contained errors. The first step was therefore to develop an algorithm for outlier detection and removal from the static weight data, prior to using the static weight data to clean the Optiweigh data. Figure 12 provides an example from the CSU farm data of weekly weight gains calculated between the static weight events.

Figure 12 – Boxplot of population dynamics of calculated average weekly cattle weight gains calculated at three yard weighing events at CSU farm. Blue dots represent data points.



One animal had a calculated weekly growth rate between 9 May and 27 June of almost 80 kg per week. This is a potential outlier based on the statistical measures such as median, first quartile, third quartile of weight gain displayed in the boxplot and also considering that an 80 kg weekly weight gain is biologically impossible. Therefore, our approach identifies a set of possible outliers in the first step using the interquartile range (IQR) criterion, and a statistical test (Rosner's test) was performed to confirm the identified data point is an outlier.

A model for the initial Market Predictor was developed that built on the improved understanding of population dynamics gained in this project. This model is based on a linear growth rate path from a point in time to the projected or desired sale date and is derived by regression. However, the producer datasets show that growth is non-linear. It is expected as more data becomes available across multiple years a sale date/sale weight projection based on non-linear growth rate will be able to be developed, which highlights the strength of the project model. The prediction tool uses the cleaned and standardised weight and growth rate data to provide a firm basis for prediction of future weights. There is the potential to provide several predictions, based on current growth rates and growth rates input by the producer, which could be based on poor, intermediate and good projections of seasonal conditions (and hence low, moderate and high growth rates). An example of this is shown on the Producer Reports in the Appendix.

The static weight data provides opportunities to apply standard growth curves to individual cattle and predict when they will be market ready. In a future iteration the in-paddock weights could be used to check cattle weights to determine progress against the target weights, and periodic assessment of the in-paddock weights could be used to look for change points that reflect animals moving ahead or behind the growth target. From this, early warning could be linked to strategic management intervention strategies. Deriving subgroups based on the static yard weights and determining how many in-paddock weights are required to represent a subgroup weight profile will increase confidence, reduce costs (e.g. mustering cattle to weigh) and allow strategic use of inputs (e.g. supplements).

The algorithms targeted and/or developed in this project and descriptions of their application are displayed in Table 1.

Table 1. Summary of project algorithms currently hosted on Shaipup or under development

Algorithm	Status	Description
AgriTrakka Static Weigh Event	Deployed on ShaipUp	This algorithm receives and validates AgriTrakka static weigh events and enables these events to be passed through to the AgriTrakka data sharing dashboard
AgriTrakka Mean Weight Event	Deployed on ShaipUp	This algorithm processes AgriTrakka average weight events and validates the JSON data follows the ICAR and ISC data schema for this event
Refined Cattle Weights	Deployed on ShaipUp	High frequency automated cattle monitoring provide opportunities for improved cattle management. Having confidence in the data ensures more confident management decisions. This algorithm combines high frequency automated cattle weight data and combines it with manually recorded cattle weight data to identify and remove potentially erroneous data. The algorithm returns statistics on the before and after data and returns the updated data.
Outlier Detection	Deployed on Shaipup	Outlier detection algorithm identifies any possible outlier(s) in the weight data based on mean growth of the remaining cattle.
Market Predictor	Under Development - currently generated for producer reports (see Appendix 1)	Algorithm uses known current weight and growth rate data and producer inputted target weights or growth rates to estimate sale date or sale weight on specified date.

Four algorithms are now deployed on Shaipup. The Market Readiness Tool has been incorporated into the producer reports (Appendix 1) and is continuing development for deployment on Shaipup. It is likely refinement of all algorithms will continue as more data becomes available.

5.6.3 Producer reports

The final component of the project was developing a standard report to provide feedback to collaborating producers on their own datasets. The producer reports (see Appendix for two examples) provided a standard description of each participating producer's data. The report is auto-populated from their own data and describes the patterns in their data.

In both examples provided, the baseline group average data assessments were improved by:

1) identifying outlier data points in the static yard weights; and
2) ensuring that weekly average weights from the paddock included at least four individual daily weights.

Although the improvement in the mean mob weight was modest in both examples, the population distribution of weight data was tightened by applying the two above criteria. While group average weights might be relatively accurate it is the overall distribution that impacts on the ability to predict an individual animal's weight. Even a relatively small variance in the group averages can result in large discrepancies when predicting individual cattle weights.

The background information from the producer reports has the potential to create benchmarking and identify when the in-paddock weighing system is working well and when it could be improved. The benchmarking data related to cattle use of the in-paddock weighing can be further used to monitor how successful changes to management are in delivering better quality data.

6 Discussion

A key outcome from the project was being able to develop domain data analytics capacity at CSU. A data analyst was employed on a short-term casual contract on the project from February until early May 2023 to do data analytics and algorithm development. Domain learning takes a significant amount of time and it is important to allocate sufficient time to this. In addition the use of short term projects makes it difficult to retain skills developed within projects. Consideration of how to develop and retain appropriate skills into the future to facilitate future projects is warranted.

6.4 Learnings from producer engagement

Some producers were excited and highly engaged and wanted to know how they can get the most out of the new technology they are deploying in their operations. Others, while interested in the technology and future applications, appear to be reasonably comfortable with what the technology can currently deliver. This second group may be less motivated to engage deeply in projects, although they are willing to share data as long as it is easy to do.

Having producers agree to be part of projects such as this is easy; motivating producers to directly contribute, such as undergo the sign-up process, is harder. So while producers often said they would sign-up, this did not always eventuate. The project has developed a report for producers to provide feedback from the project on their own data and the broader project outputs (see Appendix) and it is anticipated that seeing the algorithms applied to their data will provide additional motivation for producers, and the earlier this can be done, the better. We are also considering other uses of the data through aggregation at industry level such as benchmarking that may also provide motivation for producers to interact with these tools.

The process needs to be made as easy as possible for producers and missing the first opportunity, for example due to a flow error, can result in producers not following through with sign-up or the final approvals required for data flow. In this project we did make it clear that this was an early stage concept and there would likely be some problems identified as the sign-up process progressed.

6.5 Algorithm development

The short time frame of the project has meant that algorithm development is still preliminary, despite some of these algorithms now being deployed. Some producers have only signed up to the project very late, and more extensive testing of the algorithms on these and other datasets will be required and will likely result in further refinement of the algorithms.

Some producers engaged with this project have specific questions to be answered and are wanting on-going engagement with the project team. We are considering funding opportunities to support on-going activities.

The project aimed to look at the commercial opportunities. A possible structure would be a producer pay model, where producers pay for using the Agritrakka infrastructure (TerraCipher), producers pay to access the algorithms on Shaipup (The algorithm developer e.g. CSU and TerraCipher + producers who contribute to algorithm development), and the host (TerraCipher) charges the algorithm owner for hosting on Shaipup.

7 Conclusion

7.4 Key findings

Producers are still exploring the application of new agricultural technology and it will take time for them to understand the limitations and opportunities. Our project has identified that the mean in-paddock and in-yards cattle weights on some properties are reasonably similar although there is large variation for individual cattle weights (see Appendix 1, Fig. 6 and 7) which can be improved by applying data cleaning and threshold criteria such as the minimum number of times an animal is weighed by the in-paddock system during the week. On other properties the in-paddock and in-yard weights are not well aligned. Some producers from the latter group are seeking further support to understand the population dynamics of their herd and how they can combine the different technologies. We are investigating how to fund this additional analysis.

Producers are generally willing to share data with researchers; however, the mechanisms for doing this need to be easy. Notwithstanding some early teething issues, the systems developed in this project do facilitate an easy mechanism for sharing of data; however, the project team believes that providing motivation such as early reporting on their own datasets will encourage uptake and usage of the algorithms. The producers who were most active in the project in signing up and utilising the systems were those who are highly interested in the ag technology and its potential or those that have identified possible concerns with the accuracy of their data and are seeking support from researchers to better understand their data.

A key benefit of using a combination of technologies is being able to easily identify outliers and possible erroneous data from in-paddock and in-yard cattle weighing systems, which is foundational for being able to confidently make decisions based on the data that is collected. The need to understand population dynamics within a herd is important when using in-paddock weighing that collects only a sample of weights within the population. We have identified that using the combination of in-paddock and in-yard weight collection systems provides a way to incorporate these population dynamics within the analysis which is particularly important when making predictions.

Understanding the accuracy and relationship between the static yard weights and the paddock mean weekly weights provides the opportunity to determine seasonal growth paths and use these data to estimate market access options. Static yard weights provide a starting reference point for the whole herd; these data determine the population distribution and can be used to derive growth scenarios. The mean weekly paddock weights are referenced against the static weights to ensure sufficient representative cattle are used to derive a population or sub-population mean weight estimate. The higher frequency in paddock weighing is used to identify change points and define triggers for management intervention.

7.5 Benefits to industry

The project has developed and demonstrated data sharing infrastructure that is now available for industry. Algorithms for data standardisation and data cleaning for cattle live weights collected using in-paddock and in-yard weighing systems, are now available on the Shaipup website. The project also demonstrated how utilising this infrastructure can improve product development and research outcomes, through aggregation of data from a combination of agricultural technologies. The project has improved the understanding of limitations of in-paddock and in-yard weighing and deployed algorithms which can improve the confidence producers have in the data generated on their farm.

8 Future research and recommendations

The technology explored in this project is relatively new and producers are still exploring how accurate the data they are receiving is and how they can apply it in their production system. Two producers have identified they are keen to continue working with the research team, in particular one large producer who has identified a large discrepancy between in-paddock and in-yard weights. We believe that our approach to the problem, that considers population dynamics, can assist with cleaning data and helping producers make best use of the technology for tracking cattle growth.

Additional funding will be sought to allow continued data analytics and algorithm development and refinement. We have identified the need for clear reporting to encourage more producers to utilise the tools and algorithms developed in this project. The development of an industry benchmarking tool is one concept we are seeking to explore in a future project. Charles Sturt University is continuing to support Shaipup and we aim to continue to partner to work on driving increased value for producers.

9 Appendix 1 Examples of Producer Reports

Algorithms to Refine Cattle Weights
Using
Autonomous In Paddock Weighing Devices
Property Report for PIC XXXXXX1

CSU & MLA-MDC

2022-10-21

10 Background

All data were processed through the TerraCipher AgriTrakka platform where both the static weights and the mean weights were standardised following the Integrity Systems data schema. Data standardisation ensures consistency for data comparisons, data storage and data insight via Shaipup algorithms.

Preliminary data analyses aims to better understand the data by determining patterns or opportunities to improve data quality through standardised data filtering. Eventually combined data is used to identify underlying data analytics delivered through standardised algorithms increasing the opportunities to derive more meaningful insight from the data.

11 Data Overview

The dataset includes a total of 747 unique cattle RFID tags from property identification code (PIC) xxxxxxx1. The data covers the time period from Fri 29 Oct 2021 to Sat 20 May 2023. The data includes single weights from cattle in the yards, also referred to as static weights, and automated weights from cattle in the paddock. These automated weights are averaged across multiple weights each day and input as a mean daily weight for each animal.

The initial analyses grouped weight data into weeks of the year providing more data for comparative analyses.

The automated paddock weights rely on cattle accessing an in paddock weighing device, not all cattle access this device and those cattle that do will have variable numbers of visits on any given day. The mean paddock weight includes additional data on the number of visits. The overall summary of activity of cattle that have mean weights shows there are 686 cattle that have used the system throughout the data collection period. The mean number of weeks that cattle recorded at least one weight during any given week is 7 and the maximum number for at least one animal weights is 31 weeks (see fig. 1).

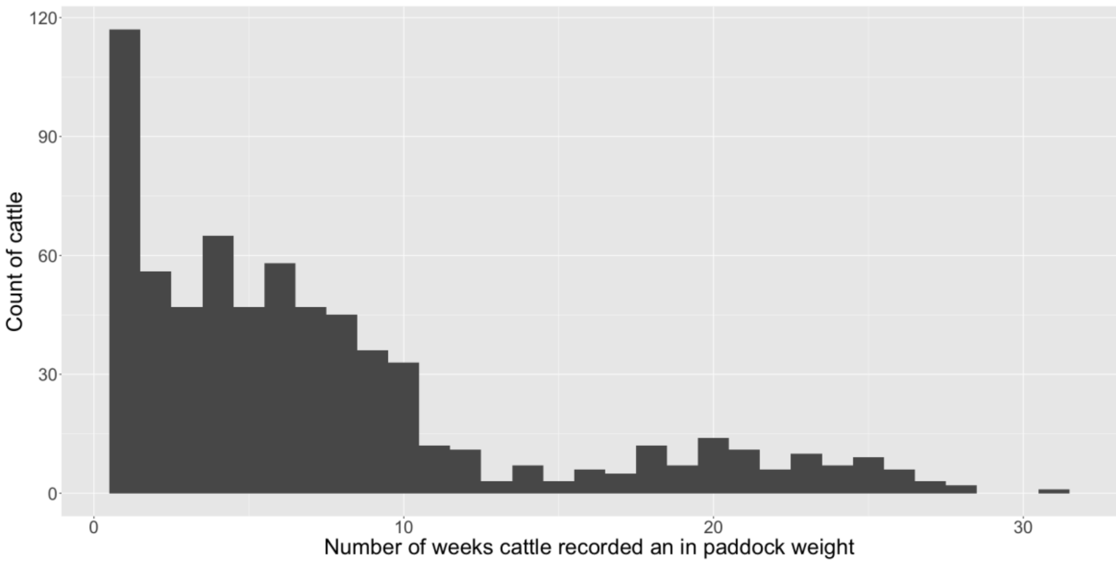
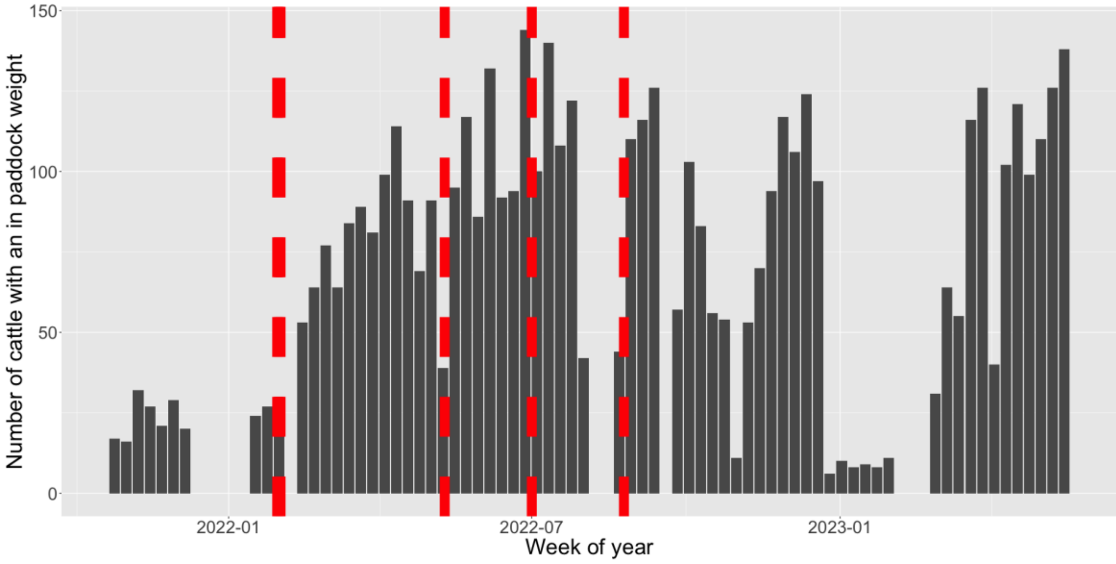


Figure 1: Number of times individual cattle accessed Optiweigh

The overview of the weekly history and numbers of cattle that visit the in paddock weighing system and the timing of the static yard weights is shown in fig.2. The red lines on the graph show the weeks with a static weighing event.



#

Figure 2: Weekly history of cattle weights, red lines represent the weeks that have static yard weights

12 Validation of in paddock weighing systems

To better understand the in paddock weight estimates we compare data from the individual weeks where static yards weights were recorded. For each static weighing event all cattle that also have an in paddock weight record were identified. By plotting all of the static weights against the corresponding weekly average mean weights it provides an estimate of the accuracy of the in paddock weighing (see fig. 3).

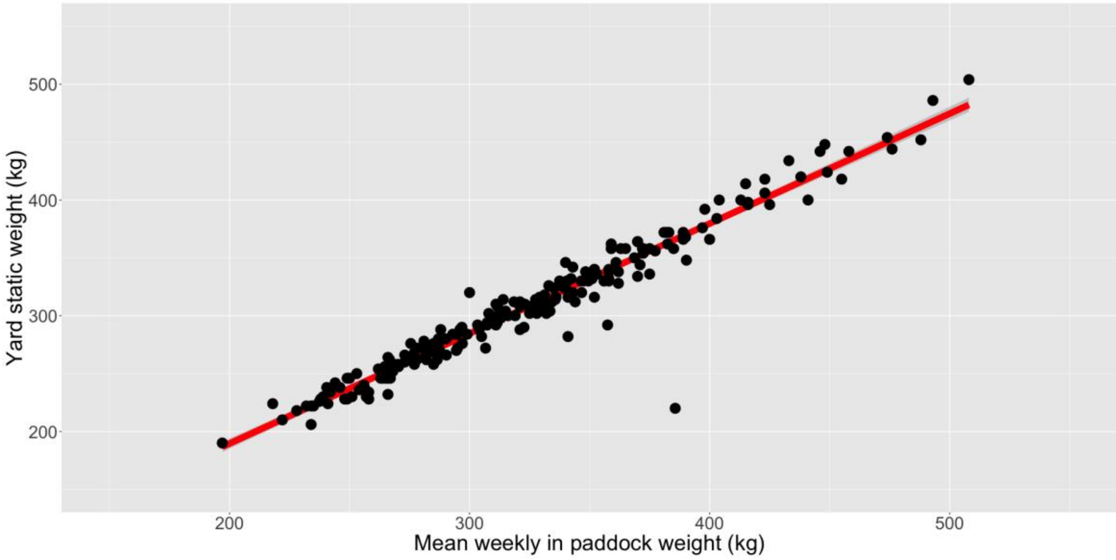


Figure 3: Validation of paddock weights using yard static weights

Individual weeks of static data combined with paddock weights provide the growth trend over time (see fig. 4).

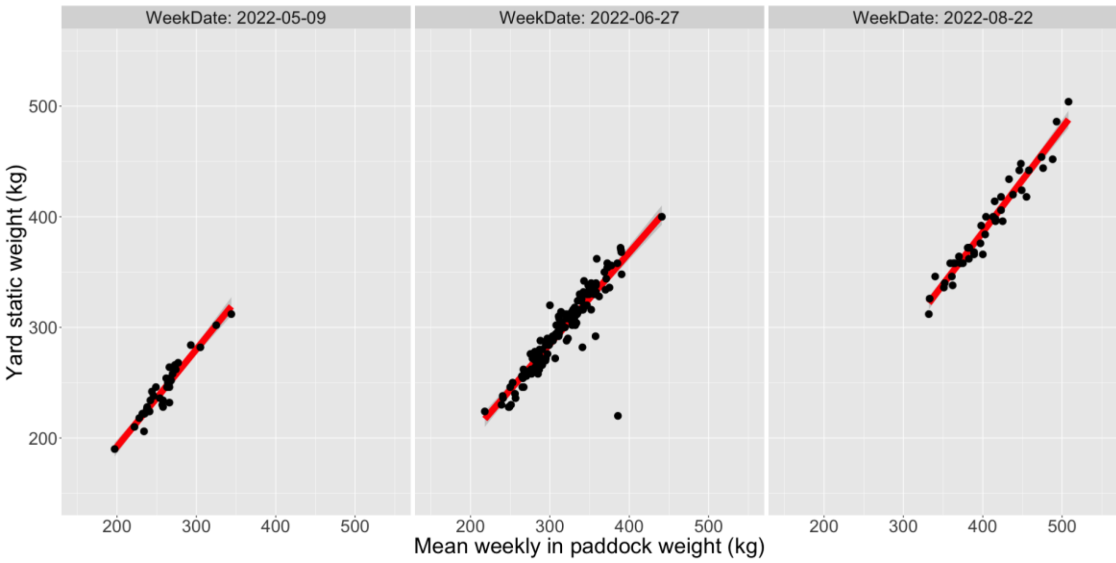


Figure 4: Separate weekly validation of paddock weights using yard static weights

A frequency plot is used to plot the overall distribution of the difference between the mean weekly paddock weight and in the in yard static weights (see fig. 5). The overall mean difference between the two weighing systems is 17.74kg.

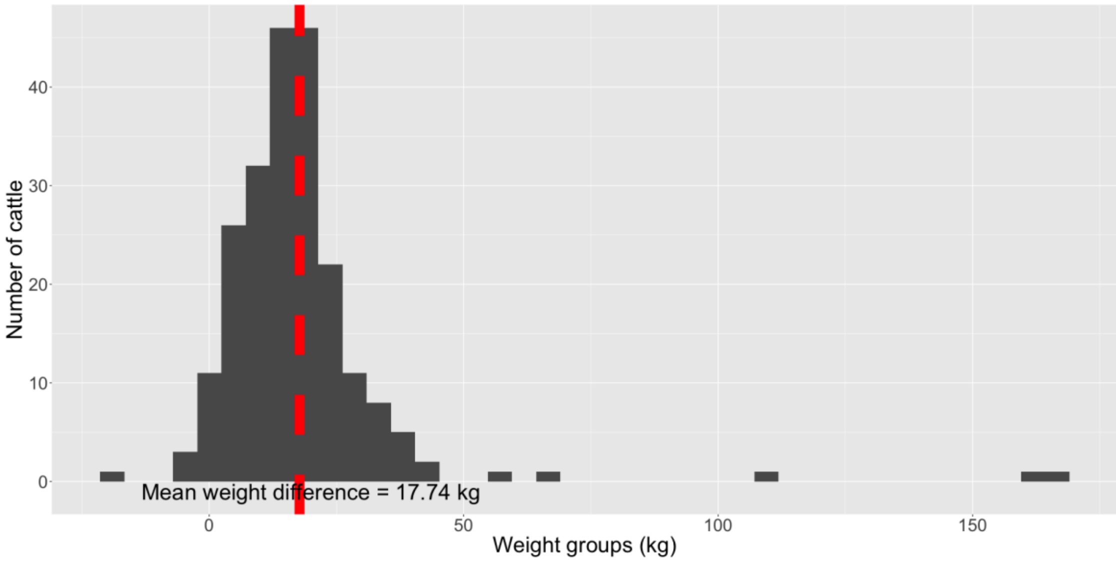


Figure 5: Frequency of weight difference between paddock weekly average and static weights collected in the same week

A density plot represents the frequency units as density adding up to one. A normal distribution using the mean and standard deviation of the weight difference is used to generate a density curve (see fig. 6). If all static weights were perfectly matched to the paddock weekly mean weights all cattle would have a mean difference of 0 and be aligned as a single column with the count of cattle representing the height of the bar. The density curve is used to demonstrate how removing outliers increases the accuracy of paddock weighing data.

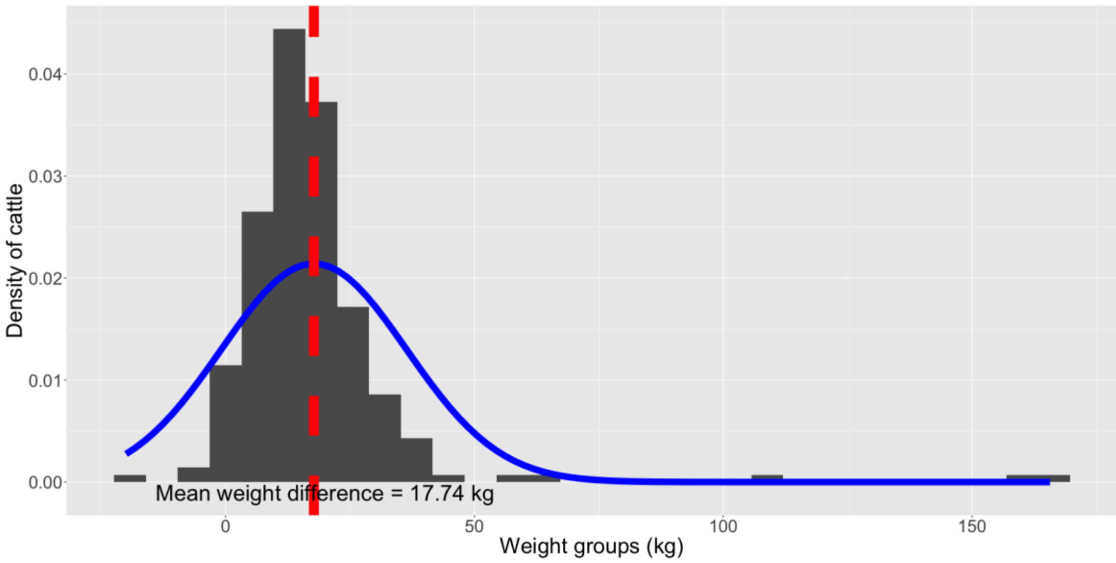


Figure 6: Density plot weight difference between paddock weekly average and static weights collected in the same week

The static yard weights can have inaccuracies, these data are refined by removing all values that fall one standard deviation outside the mean weight difference. The initial algorithm removes all weekly mean paddock weights that have four or less individual weights. The revised density curve (see fig. 7) shows the new mean weight difference is 16.24kg.

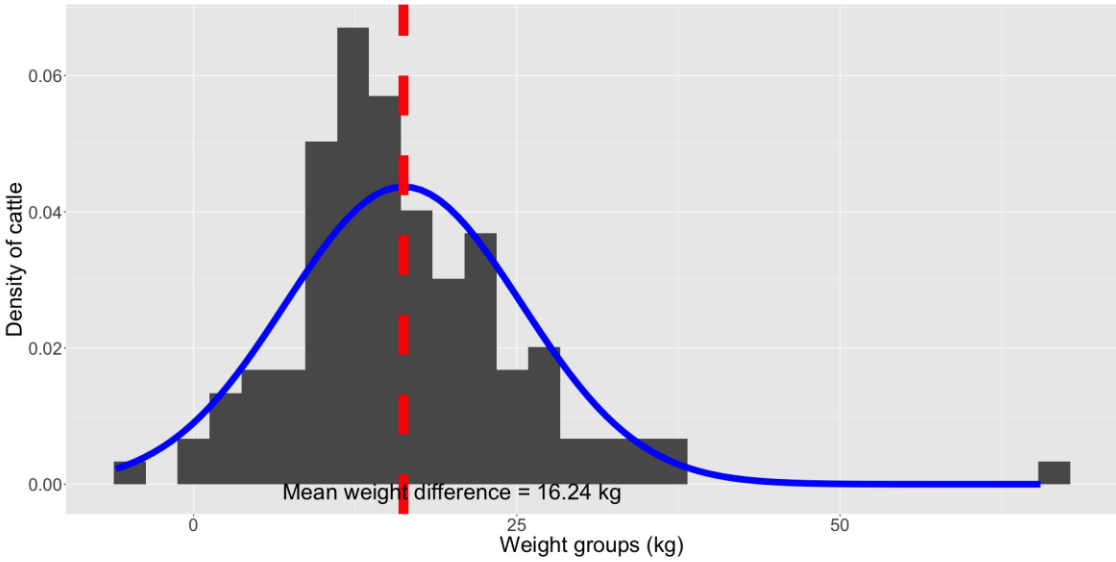


Figure 7: Density plot of weight difference using corrected data between refined paddock weekly average and static weights with outliers removed but all collected in the same week

13 Using static weights to target market opportunities

Understanding the accuracy and relationship between the static yard weights and the paddock mean weekly weights provides the opportunity to determine seasonal growth paths and use these data to estimate market access options. Static yard weight provides a starting reference point for the whole herd, these data determine the population distribution and can be used to derive growth scenarios. The mean weekly paddock weights are referenced against the static weights to ensure sufficient representative cattle are used to derive a population or sub-population mean weight estimate. The higher frequency in paddock weighing is used to identify change points and define triggers for management intervention.

By taking static weights and applying growth rates to the population it is possible to determine potential cattle that are market ready through the growing season. The target weight and growth rates can be adjusted, for these data the target weight is set at 450kg. The static weights are used to provide a profile of weights (see fig. 8) for weights collected on Mon 31 Jan 2022. Different growth rate options are applied and show how high, medium and low growth rates can impact market readiness (see fig. 9).

Using the static weights and market predictor the in-paddock weighing can be used to track progress against market goals.

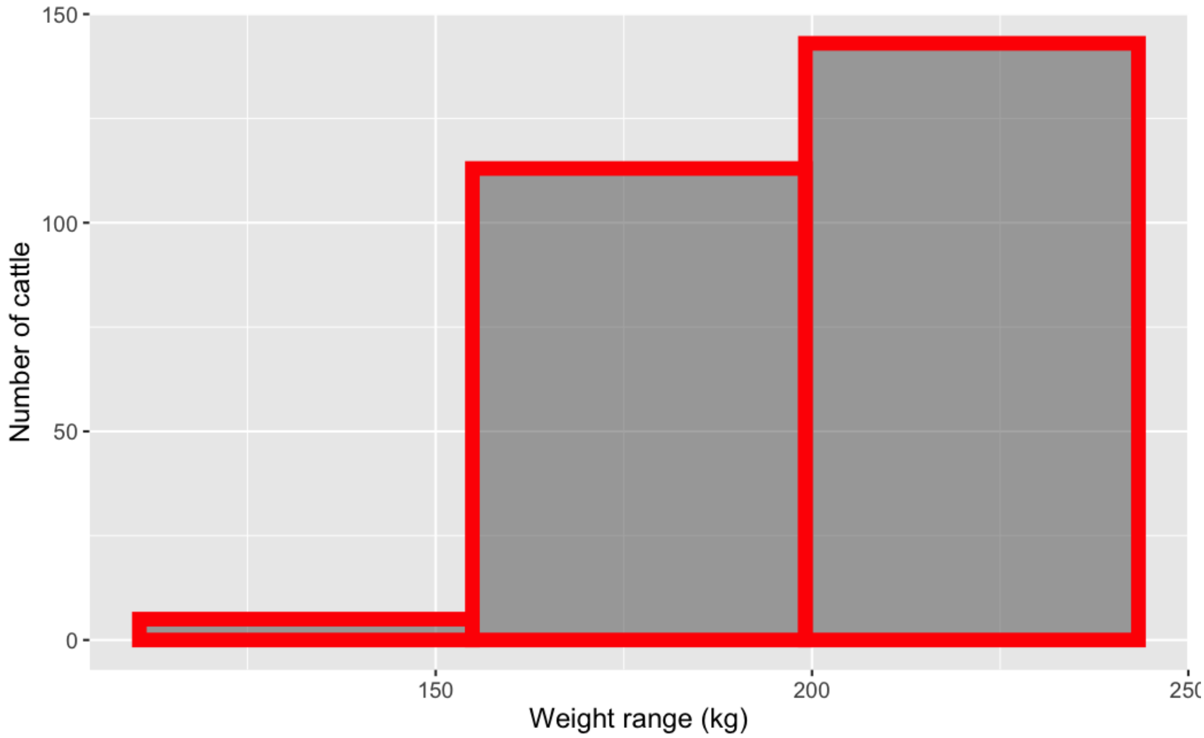


Figure 8: Frequency plot of cattle weights put into high, medium and low weight classes from static weight data collected on Mon 31 Jan 2022

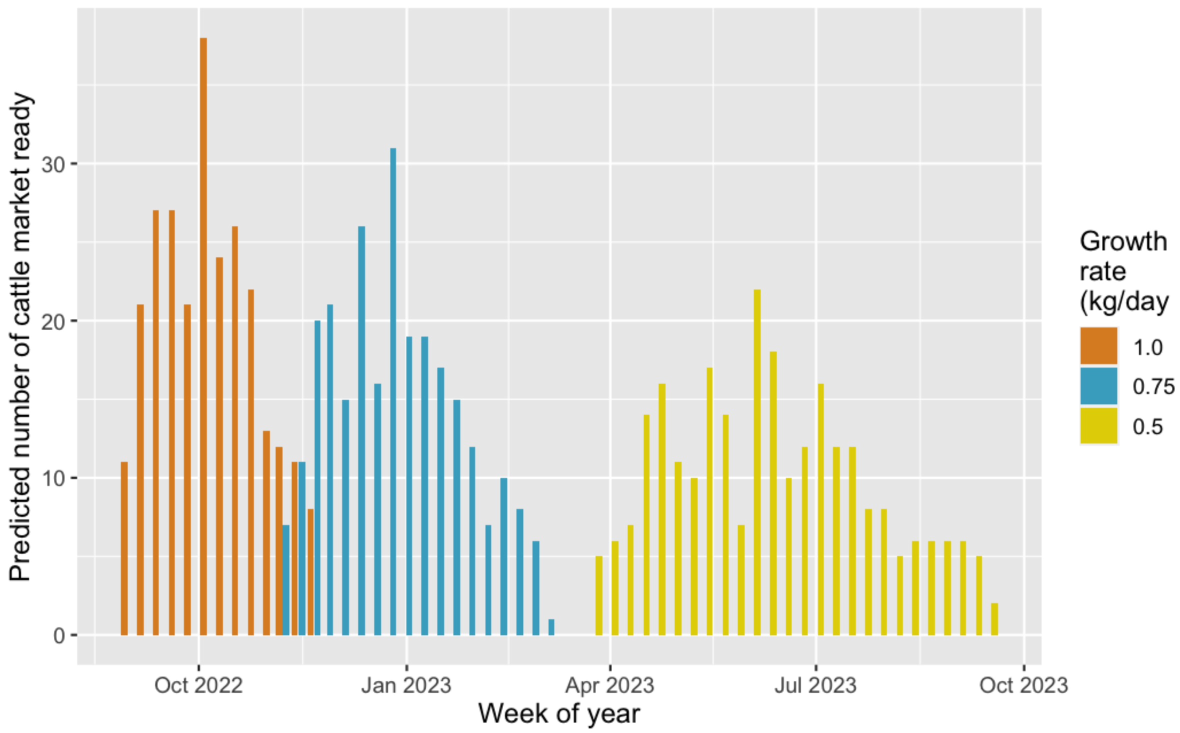


Figure 9: Predicted market readiness (target weight is 450 kg) using static weights collected on Mon 31 Jan 2022 and applying three growth rates assuming linear growth throughout the period

Algorithms to Refine Cattle Weights Using Autonomous In Paddock Weighing Devices

Property Report for PIC XXXXXX2

CSU & MLA-MDC

2022-10-21

14 Background

All data were processed through the TerraCipher AgriTrakka platform where both the static weights and the mean weights were standardised following the Integrity Systems data schema. Data standardisation ensures consistency for data comparisons, data storage and data insight via Shaipup algorithms.

Preliminary data analyses aims to better understand the data by determining patterns or opportunities to improve data quality through standardised data filtering. Eventually combined data is used to identify underlying data analytics delivered through standardised algorithms increasing the opportunities to derive more meaningful insight from the data.

15 Data Overview

The dataset includes a total of 468 unique cattle RFID tags from property identification code (PIC) XXXXX2. The data covers the time period from Sun 02 Jan 2022 to Sat 20 May 2023. The data includes single weights from cattle in the yards, also referred to as static weights, and automated weights from cattle in the paddock. These automated weights are averaged across multiple weights each day and input as a mean daily weight for each animal.

The initial analyses grouped weight data into weeks of the year providing more data for comparative analyses.

The automated paddock weights rely on cattle accessing an in paddock weighing device, not all cattle access this device and those cattle that do will have variable numbers of visits on any given day. The mean paddock weight includes additional data on the number of visits. The overall summary of activity of cattle that have mean weights shows there are 349 cattle that have used the system throughout the data collection period. The mean number of weeks that cattle recorded at least one weight during any given week is 15 and the maximum number for at least one animal weights is 28 weeks (see fig. 1).

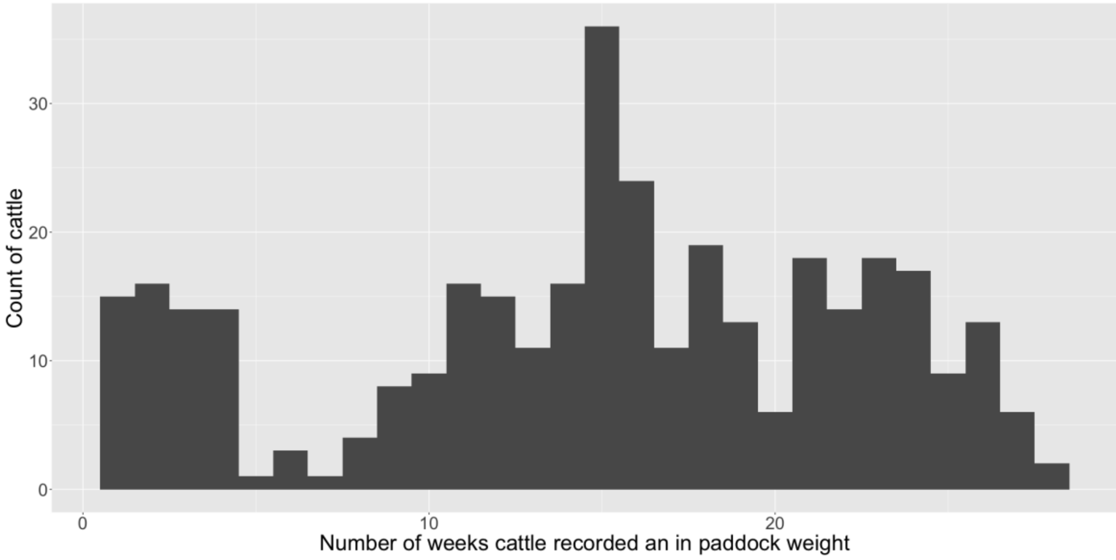


Figure 1: Number of times individual cattle accessed Optiweigh

The overview of the weekly history and numbers of cattle that visit the in paddock weighing system and the timing of the static yard weights is shown in fig.2. The red lines on the graph show the weeks with a static weighing event.

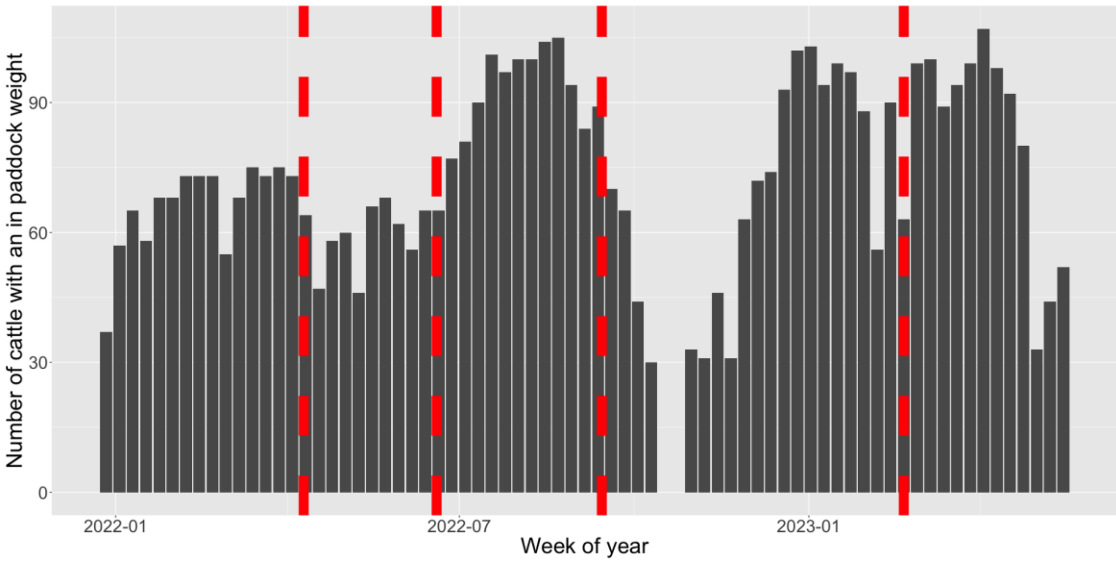


Figure 2: Weekly history of cattle weights, red lines represent the weeks that have static yard weights

16 Validation of in paddock weighing systems

To better understand the in paddock weight estimates we compare data from the individual weeks where static yards weights were recorded. For each static weighing event all cattle that also have an in paddock weight record were identified. By plotting all of the static weights against the corresponding weekly average mean weights it provides an estimate of the accuracy of the in paddock weighing (see fig. 3).

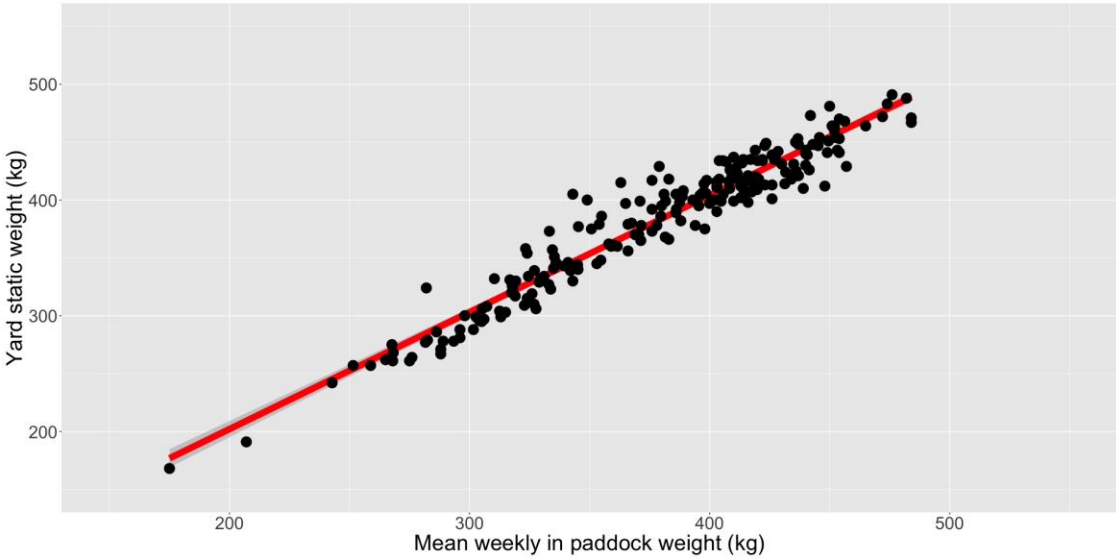


Figure 3: Validation of paddock weights using yard static weights

Individual weeks of static data combined with paddock weights provide the growth trend over time (see fig. 4).

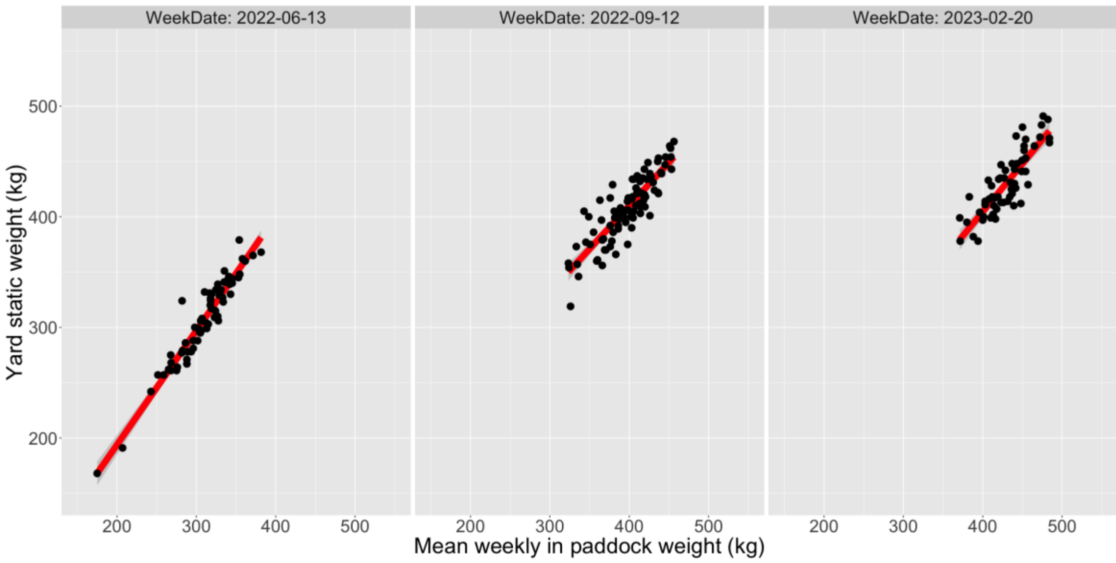


Figure 4: Separate weekly validation of paddock weights using yard static weights

A frequency plot is used to plot the overall distribution of the difference between the mean weekly paddock weight and in the in yard static weights (see fig. 5). The overall mean difference between the two weighing systems is -3.85kg.

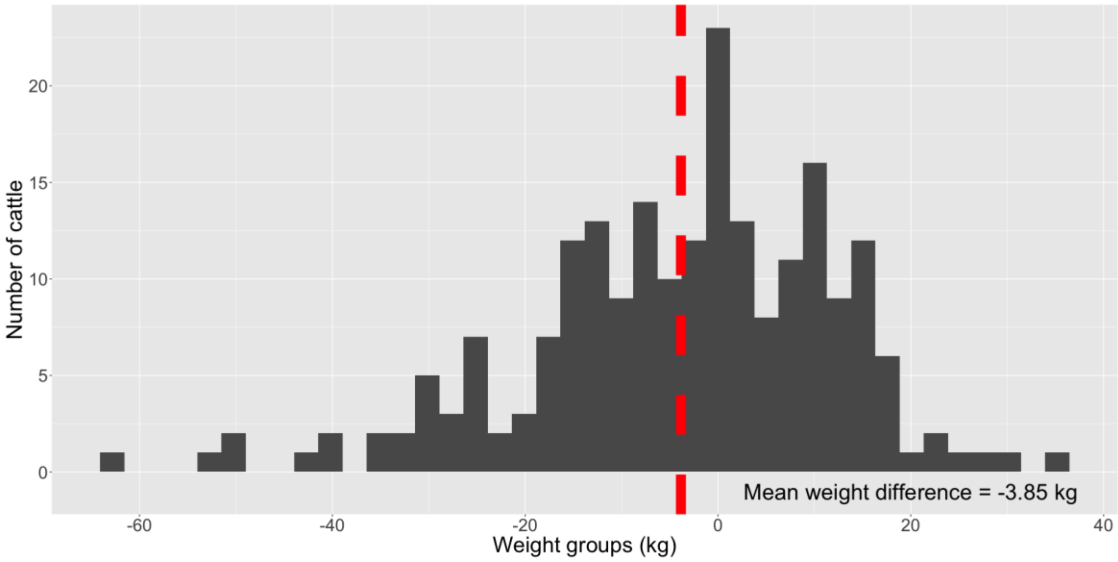


Figure 5: Frequency of weight difference between paddock weekly average and static weights collected in the same week

A density plot represents the frequency units as density adding up to one. A normal distribution using the mean and standard deviation of the weight difference is used to generate a density curve (see fig. 6). If all static weights were perfectly matched to the paddock weekly mean weights all cattle would have a mean difference of 0 and be aligned as a single column with the count of cattle representing the height of the bar. The density curve is used to demonstrate how removing outliers increases the accuracy of paddock weighing data.

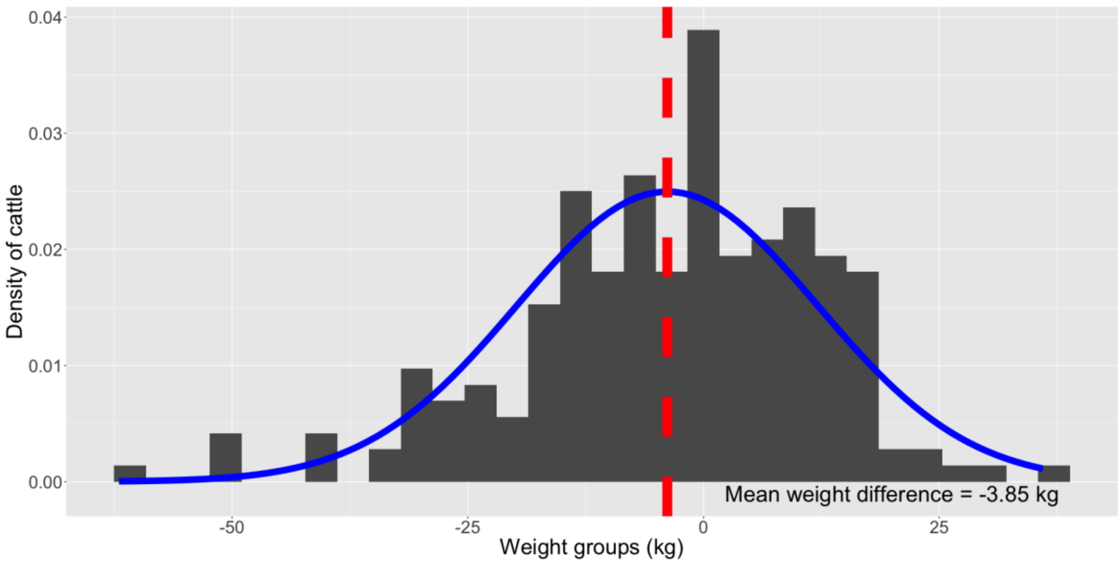


Figure 6: Density plot weight difference between paddock weekly average and static weights collected in the same week

The static yard weights can have inaccuracies, these data are refined by removing all values that fall one standard deviation outside the mean weight difference. The initial algorithm removes all weekly mean paddock weights that have four or less individual weights. The revised density curve (see fig. 7) shows the new mean weight difference is -2.28kg.

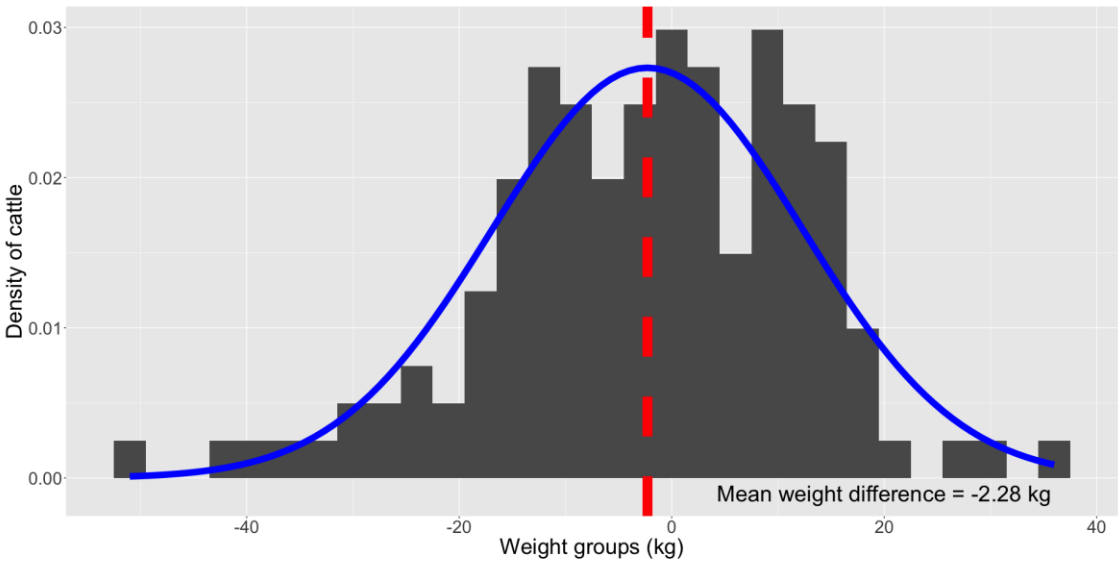


Figure 7: Density plot of weight difference using corrected data between refined paddock weekly average and static weights with outliers removed but all collected in the same week

17 Using static weights to target market opportunities

Understanding the accuracy and relationship between the static yard weights and the paddock mean weekly weights provides the opportunity to determine seasonal growth paths and use these data to estimate market access options. Static yard weight provides a starting reference point for the whole herd, these data determine the population distribution and can be used to derive growth scenarios. The mean weekly paddock weights are referenced against the static weights to ensure sufficient representative cattle are used to derive a population or sub-population mean weight estimate. The higher frequency in paddock weighing is used to identify change points and define triggers for management intervention.

By taking static weights and applying growth rates to the population it is possible to determine potential cattle that are market ready through the growing season. The target weight and growth rates can be adjusted, for these data the target weight is set at 450kg. The static weights are used to provide a profile of weights (see fig. 8) for weights collected on Mon 13 Jun 2022. Different growth rate options are applied and show how high, medium and low growth rates can impact market readiness (see fig. 9).

Using the static weights and market predictor the in-paddock weighing can be used to track progress against market goals.

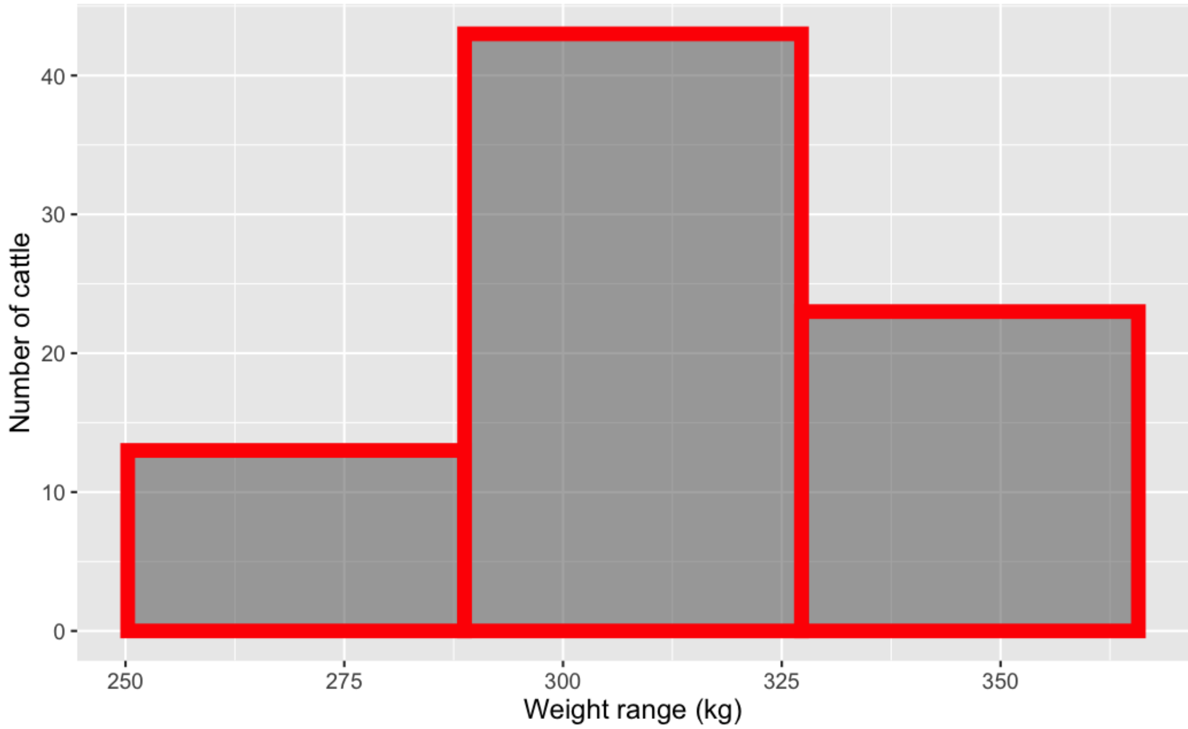


Figure 8: Frequency plot of cattle weights put into high, medium and low weight classes from static weight data collected on Mon 13 Jun 2022

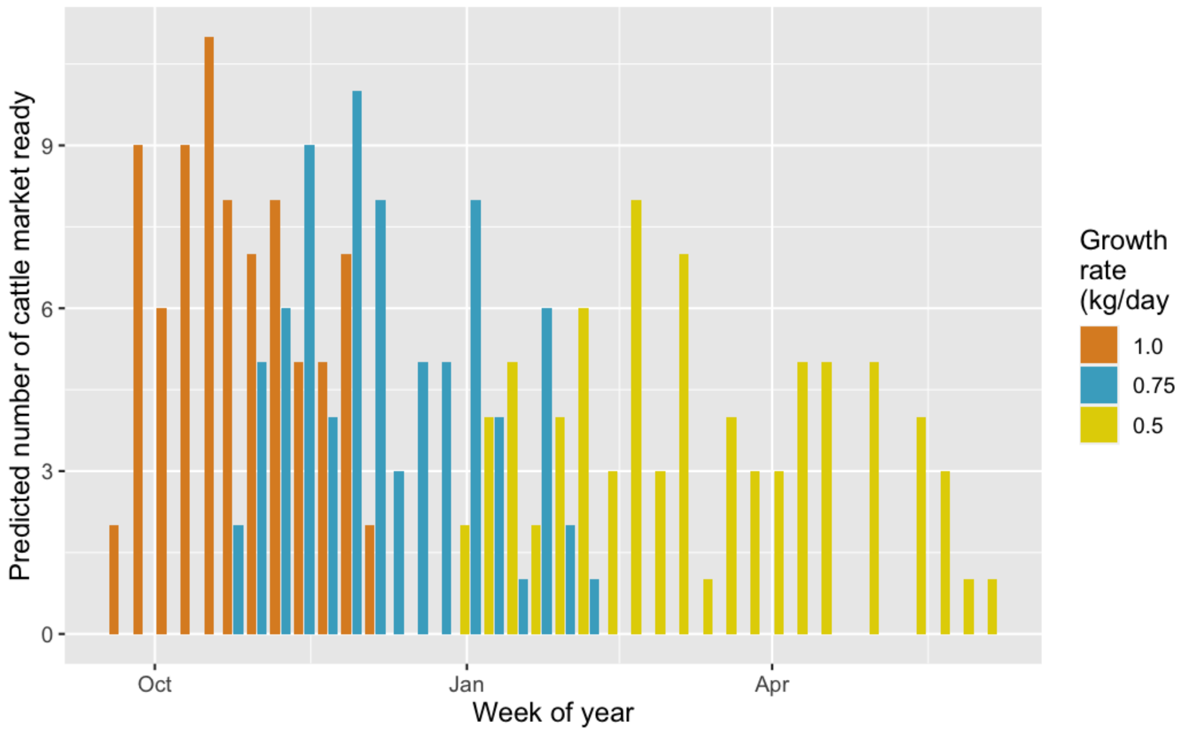


Figure 9: Predicted market readiness (target weight is 450 kg) using static weights collected on Mon 13 Jun 2022 and applying three growth rates assuming linear growth throughout the period