Measuring cutting area performance, based on the length of fries

DEVELOPING A DATA DRIVEN MODEL TO MEASURE CUTTING AREA PERFORMANCE EACH FIVE MINUTES AS SUPPORT FOR CUTTER OPERATORS' DECISION-MAKING

JURRIAN MATHIAS VAN DALEN

BACHELOR OF INDUSTRIAL ENGINEERING & MANAGEMENT

UNIVERSITY OF TWENTE

20TH OF JUNE 2025

Bachelor thesis Industrial Engineering & Management

Author:

J.M. van Dalen (Jurrian)

2999358

BSc Industrial Engineering and Management

University of Twente

University of Twente

Drienerlolaan 5

7522 NB Enschede

Netherlands

McCain Lelystad

Steenstraat 9

8211 AG Lelystad

Netherlands

Supervisors University of Twente

Dr. Dipl.-Ing M. Sharma (Mahak)

Dr. H. Chen (Hao)

Supervisor McCain

H. Feenstra (Harry)

Preface

Dear reader,

In front of you is my report “Measuring cutting area performance, based on the length of fries”. I have conducted this research in collaboration with the Lelystad facility of McCain. This report marks the end of my bachelor’s degree in Industrial Engineering and Management.

At first, I would like to thank McCain for allowing me to perform this research in their production environment. A special thanks to my company supervisor Harry Feenstra. It has been a pleasure to work with you! Your guidance and broad knowledge of the entire process greatly attributed to the quality of my report. I also want to thank my university supervisors Mahak Sharma and Hao Chen for their guidance and feedback. I would like to thank my student buddy Jurgen van der Blom. Our weekly meetings to discuss each others progress and your willingness to provide feedback were much appreciated. Finally, I would like to thank my family and friends for their support.

I hope you will enjoy reading my bachelor thesis!

Jurrian van Dalen

Enschede, June 20th, 2025

Management summary

Introduction

This research is conducted for the Lelystad facility of McCain. McCain is the worlds' largest manufacturer of frozen potato products, including fries. The research aims to increase performance insights at the cutting area based on the length of fries. McCain has troubles with managing length of fries due to the high variation of potato dimensions. This leads to often not meeting quality standards set by the customers. Which consequently leads to increased cost and higher inventories. McCain experienced that existing systems are not able to manage production related decisions based on the length of fries. This research aims improve managing production related decisioning, by developing a data driven decision model. The data for this system was not available before. However new machines have been introduced that allow for data extraction about potato and fries dimensions and lengths respectively, which are important measures to evaluate process performance at the cutting area. The goal of this research is to develop model that is able to measure the cutting area performance based on the length of fries and inform cutting operators, allowing for data driven decision making regarding machine settings.

Methods

First the current physical process was mapped. This was done by performing observations on the production line during production, and with the use of the internal documentation provided by the company, which is based on previous research and experience. Critical points that might cause harm to the fries were specified. The data to be extracted from the data collection devices integrated in the production process was accessed to make informed decisions regarding the possibilities for data evaluation. A basis for the research solution was formed by developing a theoretical data workflow.

An effort is made to improve the potato shape approximation. Previously, the assumption that a potato could be accurately described by a perfect ellipsoid was used (Somsen et al., 2004), (Sari & Gofuku, 2023). Due to the high variety in shapes this was thought not to be of desired accuracy. Therefore, using literature, mathematics, and observations a new approach was introduced.

The data that was collected will form the basis of the research solution. Literature is examined to determine the best approach to conduct a fry extraction from a large sample of potatoes. The modelling steps are visually explained and coded using Python. Validation of the model with different input parameters was performed by comparing results to a real life sample of potatoes that were manually cut into fries. An additional application of the simulation approach was explored to determine the losses occurring from using certain knife configurations.

Results

The effort to enhance the approximation of potato shape yielded variable outcomes; however, for the "King Russet" variety, the results were particularly promising. The new approximation was able to improve the volume approximation based on the acquired exponent value and measured potato dimensions compared to the perfect ellipsoid approximation by 1.3 percentage point. The variety "Fontane" approximation showed worse results, namely -4.4 percentage point. Since the sample size was relatively small, it can not be said with certainty which approach performs better.

The simulation model developed uses the potato dimensions to make a prediction about the fries' distribution. It was found that the models predicted fries' distribution was not sufficiently accurate when compared to real life results. This inaccuracy can be attributed to three identified issues.

1. Insufficiently accurate potato shape approximation
2. Too small sample size to compare results against
3. The used assumption of normally distributed potato dimensions

Reconfiguring the model to measure the losses due to slivers or nubbins, occurring at varying potato widths, under the assumption of a perfect ellipsoid shape approximation. The virtual potatoes are cut into fries. The model categorizes each fry individually into fry, nubbin or sliver, based on its length and thickness. The volume of each fry is determined to compute the total volume in each category. This is performed for 1000 potatoes per width of the potatoes for each knife. It has shown that there is a potential to decrease losses in for certain tube choices. It was found that a potential reduction in volume loss can be attained of between 0.23-0.51% depending on the cut size.

Conclusion

The research has shown that measuring cutting area performance based on the quality identifier - length of fries - is a viable option for the company to explore further. A system has been developed that is able to handle the input parameters given by the data collection devices and by performing a simulation is able to predict the fries' length distribution at the second data collecting device. It was found that the model does not provide statistically valid results. More research must be conducted to ensure valid and reliable results that will support operators' decision-making before integration.

The main findings of the report can be summarised as follows:

- Fry damage mainly occurs at the water jet cutting stage; other parts of the system pose low risk.
- Super-ellipsoid shape approximation showed inconsistent accuracy across potato varieties.
- Monte Carlo simulation is effective for estimating fry extraction and runs fast enough for real-time use, using a Lenovo ThinkPad P15v.
- Model fails to predict fry length distribution accurately due to shape approximation, incorrect normal distribution assumptions, and can not prove statistically that a sufficiently large sample size was used.
- The model can predict fry quantity accurately, the real count deviated 1.8% from the predicted count.
- Knife choice impacts yield; potential loss reduction of 0.23–0.51% depending on tube and knife combination, assuming the potato shape can be approximated using a perfect ellipsoid.

Recommendations

The company is advised to perform more research in the following topic before implementing the system.

1. **Data accuracy**: validate the correctness of data from both the Welliver and Véryx systems.
2. **Consistent data intervals**: standardize Welliver camera data collection intervals to every 5 minutes to support consistent model calculations.
3. **Statistical assumptions**: reassess the assumption of normal distribution in camera data. Each lane likely represents a segment of a truncated normal distribution due to sorting into six categories, not a full normal distribution.
4. **Larger sample size**: conduct the super-ellipsoid shape approximation on a significantly larger sample and test its statistical impact.
5. **Model refinement**: After implementing the above steps, improve the shape model. Explore more advanced shape approximation methods such as:
    a. 3D scanning (Hirtle et al., 2022), investigate if the Welliver cameras' 3D capabilities suffice for these purposes.
    b. Statistical shape modelling (Danielak et al., 2023)
    c. AI-based 2D to 3D prediction (Joyce & Brown, 2022)
6. **User engagement**: Involve operators actively during and after implementation to increase adoption and long-term success.

Some other potential ways to improve the overall fries making process, and that were outside the scope of this research. Therefore, not thoroughly researched however do show potential value are:

1. **Robotic halver settings**: develop a method using Welliver and Véryx data to set correct thresholds for the robotic halvers.
2. **Process monitoring**: add extra length measurements throughout the process to detect fry breakage points and set better baseline parameters.
3. **Tube design optimization**: consider alternative tube designs that allow a wider range of potato widths for better alignment and fry length.

Contents

Glossary

Slivers	-	Cutting residue from the potato, which surface area of the widest part is less than 30% of the cut size.
Nubbins	-	Cutting residue from the potato, which length is less than 50 mm.

1. Introduction

The bachelor thesis will be performed at the Lelystad facility of McCain. The research aims to measure the performance of the cutting area, based on the length of the fries. The company is presented in section 1.1, and the problem identification will be performed in section 1.2. The research design and deliverables will be introduced in section 1.3.

1.1 Company Introduction

McCain is an international company that produces fries and other potato products all over the world. With its forty-nine production facilities, it has a worldwide market share of 25% for fries and is therefore grown into a global leader in the frozen food industry.

McCain has two production facilities in the Netherlands, one in Lelystad and another in Lewedorp. Lelystad is focussing on the production of the traditional fries, whilst Lewedorp is focussing more on the specialty products within the portfolio of McCain.

The production facility in Lelystad is the owner of a fully automated production line, with the intended output of twenty-eight tons of finished product each hour. With around two hundred employees they run the plant 24/7. Once every 2 weeks the production is shut down to clean and perform maintenance on machines. In recent years, McCain has invested considerable time and resources into optimizing production output. Through the implementation of advanced equipment and data analysis techniques, notable improvements have already been achieved. Nevertheless, opportunities for further enhancement still exist.

1.2 Problem Identification

The company sells its products to a wide variety of customers. They trade both B2B as well as B2C. The businesses demand a product that is according to their specifications, including the length of the fries.

1.2.1 Identification of Research Problem

1.2.1.1 Quality Identifier

One of the quality identifiers of the finished product is the length of the fries. The specifications of the product set by the customer - to have a certain distribution of fries' lengths in the finished good - shows the importance of this quality feature. In general, the main concern is a lack of long fries, therefore McCain aims to limit length loss. When the finished good is not within the set specifications of the customer, there is a chance that the product might be returned to McCain, or McCain might need to compensate its customer, by refund or future discounts. Since this can have a major impact on the performance in terms of profitability of a production facility, this quality feature should be actively managed.

Though the improvements made by the company in the last couple of years have increased tonnage output, specific quality features over the production line have not been managed in the last years. The result was a quality management protocol with low adaptability and traceability during production. One section of the production line was upgraded with machines that gather data about the length quality features of the product. With this data the company can measure more often, as well as measuring at more segments of the production line. The more measurements taken at more sections of the production line, the quality features are thought to be better manageable. The company is therefore looking to explore the possibilities of utilizing this information to assist operators to manage product quality. In the month of February 2025, 2,59% of the final product was turned into second quality or even into feed, due to not meeting the quality specifications for the length of the product.

This number is including both product categories, wedges, and fries. The data from February 2024, shows a 9,1% product loss in due to not meeting length specifications and the loss is 1,05% in February 2025 for the fries only. The company currently does not know what the performance of the cutting area is and the length defects originating from that area, increasing difficulties managing the process. Though the product loss of 1.05% seems to be little in comparison to the 9.1% recorded the year before, it is still a product loss of approximately 280 kgs each production hour. Due to the economies of scale of the company, this product loss has profound impact on profitability.

Currently, the length feature of the fries is checked randomly, by collecting a finished bag with fries, thus at the end of the line, and checking the length of each one. Whilst this is a straightforward operation to perform and the found data is accurate, it is performed quite late in the process after an already large batch has been produced. Leading to a process that is hard to manage and needed changes of the settings[1] are not performed in a responsive manner. This is caused by the delay in information transmission between the measurement and the cutting area. The current method makes tracing the origin of the issue much harder.

A lack of insight into this quality feature at the cutting area can have a substantial influence on the length of the finished good. Since wrong settings of cutting machines result in a large quantity of internal defects of the potatoes, with internal breakages or decreased fries length due to misaligning the potatoes. As is shown in Figure 1, a tube with a diameter that is too large (left) allows a potato to turn, consequently reducing the length of the fries. With a tube with a too small diameter (right) the potato does not have enough space to pass through, leading to internal breakages in the potatoes and so in the fries. The tube with the correct diameter (middle) ensures correct alignment, therefore maximizing length. The losses need to be minimized, since it has significant financial consequences. The current way of monitoring the length of the fries is unable to differentiate between each area. Tracing back where length losses occur from are therefore impossible. Thus, cutting area performance can not be determined.

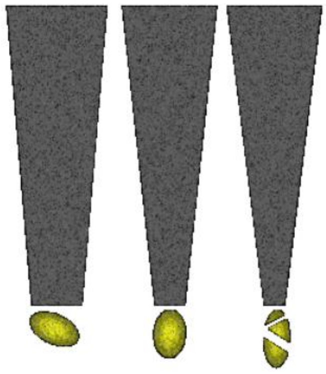


Figure 1: Tube choice

An existing method for verifying the correctness of the settings involves conducting an 80/20 test on the incoming potatoes. The idea is that 80% of the potatoes need to be within the set boundaries, and 20% is allowed to be outside these boundaries. Due to the time it takes to perform these measurements it rarely occurs. This is done at the start of an eight-hour shift or when the product changes. Hence this does not make it a practical possibility to track the product quality effectively and act responsively, leading to operators not knowing if the right machine settings are applied at each time.

[1] *The "settings" that are being referred to in this and later sections, are physical components that must be changed manually and cannot be done by programming. The physical components to change are the knives and tubes, both part of the waterjet cutters. The flow of water needs to be stopped to change these components. This is like changing the batteries of a remote controller, but much bigger.*

1.2.2 Problem Cluster

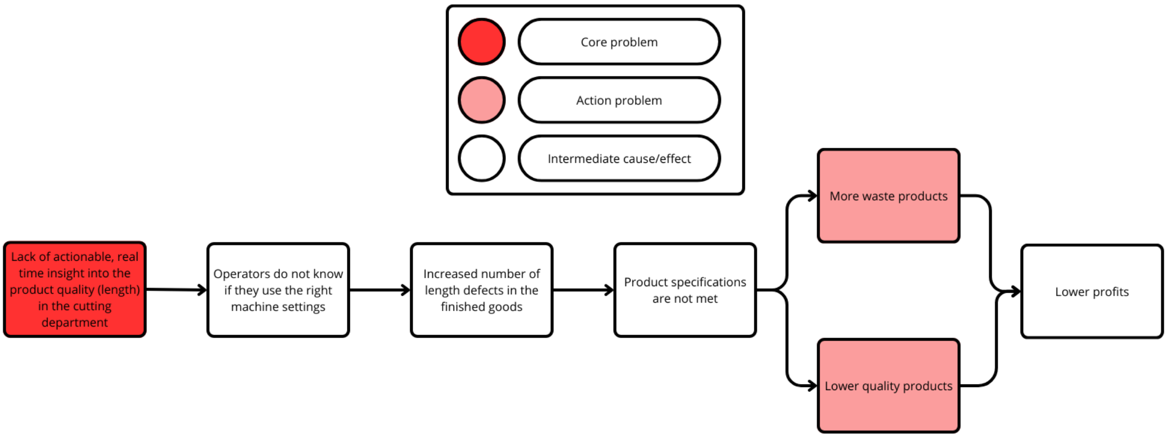


Figure 2: Problem cluster

The company found an issue: one of the main reasons for product loss in the cutting area is that the length of the fries does not meet specifications. Issues in the current process cause inefficiencies and increased costs. The lower-quality products must be reprocessed or repacked. Furthermore, second-quality products are sold at a lower price than high-quality ones, leading to reduced profits. The primary goal is to decrease the number of lower quality products and with that waste products.

A key factor behind these length defects is the number of internal breakages of the potatoes, often due to incorrect machine settings. Operators adjust settings based on experience or irregular 80/20 tests (see section 1.2.1), these manual adjustments are prone to errors that only become apparent later in the production process.

At the core of these issues is the lack of actionable, real-time insight into product quality (length) in the cutting department. This fundamental problem drives inefficiencies throughout the process, resulting in increased waste, lower-quality products, and financial losses. Addressing this core issue will improve production efficiency, product consistency, and overall profitability of the cutting area.

1.2.3 Motivation of Core Problem

Addressing the lack of (close to) real-time, actionable insights into product quality measure length of fries in the cutting department is essential for improving efficiency, reducing waste, and increasing profitability. Currently, incorrect machine settings led to an increased number of length defects, leading to higher reprocessing costs, lower-quality products, and more waste. Without up-to-date quality feedback, operators rely on experience or irregular 80/20 tests, which are not exact enough to confirm the machine settings for a longer period.

Furthermore, reducing reprocessing and repacking work will lower operational costs, saving resources. A streamlined quality control system will improve operators' decision making by providing the necessary data to make informed adjustments.

In a highly competitive market, keeping high product quality is essential for sustaining customer satisfaction and brand reputation. Through waste reduction, the company can achieve long-term profitability, improve sustainability, and increase overall production efficiency.

1.3 Research Design

In the research design, the scope, (sub)research questions and the deliverables are introduced. First the goal and scope will be defined in section 1.3.1. The research question and sub research questions will be introduced in section 1.3.2. The chapter will finish with the introduction of the deliverables in section 1.3.3.

1.3.1 Research Goal and Scope

The goal of this research is to develop a model that is able to measure the cutting area performance based on the length of fries, to enable data-driven decision-making.

This research is solely designed by the bounds of the cutting section of the production line at McCains' Lelystad facility. Implementation of this research and deliverables and its function at other locations and facilities can therefore not be guaranteed. It is, furthermore, bounded by the availability of data provided by the McCain's Lelystad facility.

In the current line the length focus is only on the fries category produced in the factory, the other product category, wedges, is not considered.

The focus is only on the cutting area of the production line. For this reason, the only two machines to be considered are the Welliver cameras and the Véryx optical soring machines. From the Welliver cameras, the mean and standard deviation of the length, width, and height of the potatoes can be found. The Véryx optical sorting machine, can find the length of the fries, and defects due to breakages.

The model that will be used will be found during the literature search on available models in similar use cases. From this literature search the model that is best applicable to McCain's situation will be chosen and used.

1.3.2 Research Question

To offer more insight into the performance of the cutting area for the operator. The main goal of this research will become the following.

How can a data driven model be developed and implemented to support operators' decision making regarding machine settings?

To answer the main research question, some further actions have to be taken. Within the research design the sub research questions are formulated following the methodology of a structured data science methodology (Van Dalen, 2025). These sub questions aim to structure the main research question to provide the needed underlaying knowledge on the main research question to succeed answering correctly.

Chapter II: Situation Analysis

This chapter tries to explore "What does the current process look like?"

The aim of looking at the current process is to increase the understanding of the machines, working manners, standardized procedures, and points of improvement. By finding the needs and wants for the process it will become more clear what variables and constraints in the physical process exist and their impact on the data driven approach. This will be beneficial in applying a solution and the "fit" of the solution can be thought of in advance. Leading to ease of implementation overall. The questions can be answered with the use of the company internal information system, observations, and available machine manufacture information. This will be answered according to the following sub questions.

1. **Sub question - What does the physical configuration look like?**
   A detailed mapping of the process needs to be made, to improve the understanding of the process and its interactions within the system.
2. **Sub question - At which point in the production line are the measurements made?**
   This question must be answered by naming the machines in the mapping of the physical configuration, also understanding how and where this data is collected is essential. Additionally, the methods used by the machines should be introduced to improve the understanding and limitation of the data to be collected.
3. **Sub question - What are the points fries can break at the cutting area?**
   The places where damage can originate from should be determined and sufficiently substantiated, to investigate what issues the new system should be able to address.
4. **Sub question - What are the possibilities of automating data evaluation?**
   The company is looking for a data driven approach to fulfil their quality performance measurements. Therefore, it is necessary to investigate the possibilities on the automation of data evaluating and its performance indicators. This will be done by connecting the available data to a theoretical concept data workflow. This will lay the foundation for the model to be developed.

Chapter III: Theoretical Background

This chapter tries to explore "How can a potato be mathematically interpreted?"

Knowing how to model a potato mathematically will allow calculating and finding the number of fries and the fries length distribution, respectively. The better the mathematical interpretation, the more representative the predicted results will be. Since working with a biological object that varies in shape a perfect mathematical approach using 3 dimensions as measurement inputs might cause issues, as shown in figure 2a in the work of Joyce & Brown, 2022. Determining whether this approach is sufficiently accurate will be done during the validation of the research solution.

Chapter IV: Solution Configuration

This chapter tries to explore "What simulation method is the most accurate in predicting the number of fries extracted from the potatoes?"

Predicting the number, and length distribution from the input fries is hard to find due to all the variations in shape and dimensions. There is a need for a literature review on the best simulation approach to predict the number and length distribution, whilst considering the volatility about shape and dimensions.

And "How can a model be developed that functions and is according to the needs of the company?"

Based on the workflow from the situation analysis, a Python model will be built that is able to consider all the input variables and tries to consider the real operations in sufficient detail, such that the results are valid and reliable.

And "Is the model valid and reliable?"

To make sure that the model represents reality in a correct manner, a statistical test on the output of the model and real life results from a sampled set of fries should be performed. With this the models' validity and reliability will be tested.

Chapter V: Solution Implementation

This chapter tries to explore "How can this model be implemented in the current system for evaluation?"

The back-end implementation of the solution is relevant since it will be using some diverse sources of data, increasing complexity when integrating in the current systems. Another key point is the ease of working with the system by the operators present at the cutting area. Therefore, it is important to investigate how to perform the front-end integration. The model output will be visualized for the operators, such that the information shown is relevant and improves the performance understanding of the cutting area. Also, the expected stakeholders' participation during and after implementation will be discussed.

1.3.3 Deliverables

The company is looking for a solution that informs the operators about the performance of the cutting area based on data. Ultimately supporting the decision-making process of the operators by performing quality (length of the fries) performance measurements. With the use of indicators in the currently used factory systems the operators will be informed when an underperformance occurs and which changes to the settings need to be made.

The company wants a model that can evaluate the cutting area performance by use of data collected by the Welliver cameras and the Véryx in the production process during a short time interval of five minutes (the shortest interval the company is able to extract data from the machines), to create a responsive information system, to be used by the operators. The model must assess whether the cutting area is performing within the parameters – decided later by the company - and identify whether machine settings need to be changed. The model must output relevant performance indicators to the operators, and show the optimal settings.

The KPIs that will be included in the research solution are made in agreement with the company. These KPIs are show actionable insights in the current performance of the cutting area.

The research solution will determine the following three KPIs

1. Discrepancy of fries' length between predicted and actual bucket quantities shown in a percentage per bucket, and an average over all the buckets
2. Categorized fries distribution at the Véryx
3. Knife type to be used at each water jet cutter

The first KPI shows the discrepancy between predicted and actual fries' length distribution. Operators should check the average loss on regular occasion to check whether the length loss is below the acceptable threshold. If the average value is larger than the acceptable threshold, the operators should take action, they should consult the losses per bucket to check which bucket loose most length. This gives an idea which water jet cutter is causing this issue. When the water jet cutter is identified, changes of the settings can be made.

The categorized fries' distribution at the Véryx show the percentage of fries that are within the buckets of <50 mm, 50<x<75 mm, and 75> mm. The specifications of the end product are managed in accordance with these buckets. Every product has its own required categorized fries' distribution. Something that can now also be managed at the cutting area. Since settings can be changed when the required distribution is not met.

The third KPI shows the operators what type of knife to use. The knife choice is depending on the second KPI, when certain specifications are not met it should determine the knife that optimize the expected cutting results. Thus, maximizing output, whilst meeting specifications. Since each water jet cutter handles another average potato width a different type of knife at each cutter might improve returns. Therefore, the operators are notified what type of knife to choose for each water jet cutter.

These KPIs will be visualized in a dashboard for the operators to use. Making the model easy to use in practice.

This will ensure better performance management in the cutting department; due to the actionable insights this model will provide. The increased frequency of measuring will enable the cutting operator to make better evaluated decisions.

2. Situation Analysis

In this chapter the objective is to familiarize the current situation. Becoming aware of the current physical configuration and it limitations and what and where data can be accessed. First current physical machine configuration will be visualized in a process mapping of the cutting area. After that, the machines that will provide the data will be further explained, also diving deeper into the methods of measurement. When the broad picture of the system is sketched, it allows to look further at the details, exploring where potatoes and/or fries might be sensitive to breakages or damages. Considering what might influence the quality, it can be decided how the data should be interpreted and used in a theoretical concept flow.

2.1 Physical Configuration

The cutting area is placed after the potato peeler. The potato peeler removes the peel. However, it can be turned off to produce so called skin-on products. Next, the potatoes are sorted based on their appearance. The machine divides the potatoes into three categories: good, bad, and re-peeling. This ensures the highest recovery of the incoming potatoes. When the potatoes are considered "good", they continue to be treated by a PEF (pulsed electric field) installation. This installation softens the cell tissue of the potatoes making the cutting operation more efficient due to the reduced friction on the knives. The process continues with sorting the potatoes and removing the ones that are too large. The two streams continue separately. The oversized potatoes move to the potato halver system, where they are cut in two equal sized potato halves. The reason being, that the company wants as much fries as possible in the product, whilst still being within length specifications. The defects are determined as a percentage of the count, thus it is better to have more fries, since this will reduce the defect percentage and allow meeting specifications easier. These potatoes are then transported to the conveyor belt with the right sized potatoes. The potatoes arrive at the two Welliver systems, these machines divide the stream of potatoes into six different lanes, based on the height of the potato. Directly behind the Wellivers are the Welliver cameras. This sorting is essential to allocate the potato to the water jet cutter that is best fit given the potato size. The potatoes fall into a tube, directing the potatoes through the water pump, into the tube and through the knives. Potential harm that can occur during this step will be addressed in section 2.3. After the cutting, the fries are decelerated since the water jet cutter brought the potatoes to a significant speed. Then the slivers that occurred are sieved out of the fries. The fries then all fall on one long conveyor belt, directing them to a sieve with larger holes, to filter out the nubbins. The Véryx right after the nubbins sieve, checks the fries for black spots, or other inconsistencies in the product. This machine is also able to measure the length of the fries. In Figure 3 the current configuration is visualized.

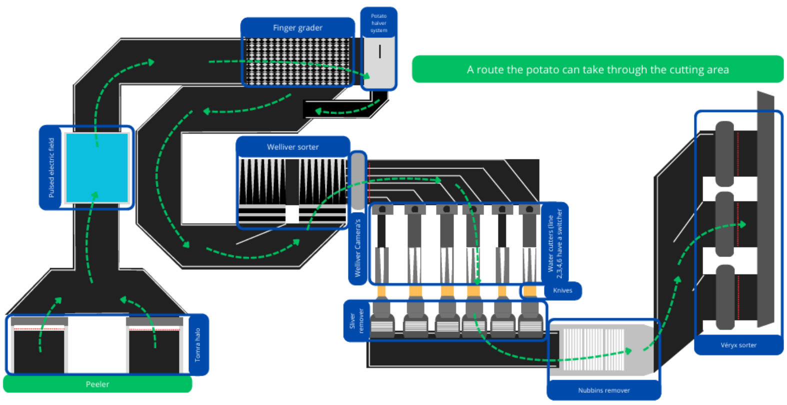


Figure 3: Visual representation of the cutting area at the production facility (top view)

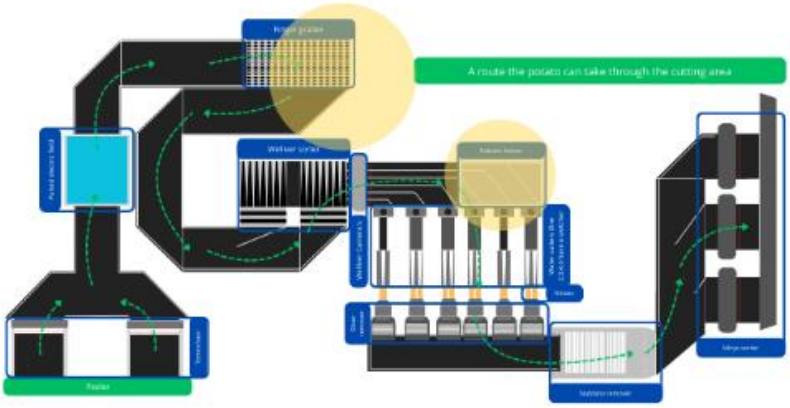


Figure 4: Visual representation of the soon to be process (top view)

In figure 4 the soon "to be" configuration is visualized, the changes compared to the current configuration are highlighted in yellow. The "soon to be process" will be operational after the summer stop. This setup closely mirrors the existing configuration, with the only notable modification being the elimination of the potato halver system positioned after the finger grader, and the integration of a new robotic halver mounted above the cutting lane belt, immediately preceding cutter knives 5 and 6. The size grading will be performed using only whole potatoes, therefore the model needs to be able to consider the halving of potatoes too. The largest potatoes fall onto lane 5 and 6. That is the reason the potatoes are only cut in those two lanes. The robotic halver can make more precise cuts than the current halver. Since, the cutting knives do not make a full cut through the potato, but only 5 or 6 cm wide through the middle. This helps extracting more value from the potatoes, since the fries on the side of the potato will remain potentially longer compared to the current system. The capacity of the robotic halver is significantly larger than the current halver, enabling more potatoes to be halved. Therefore, the number of fries extracted will increase. In figure 5 the current potato halver cut is shown on the left, compared to the new robotic halver cut shown on the right.

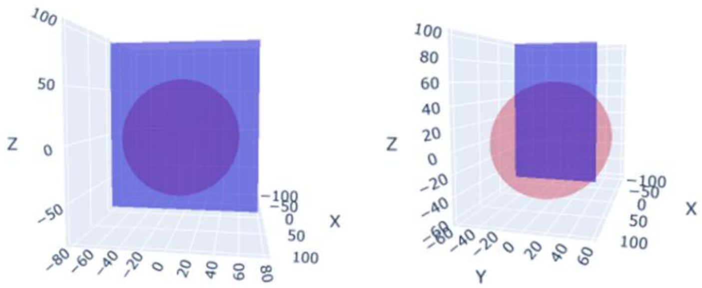


Figure 5: Current potato halver compared to the new robotic halver

2.2 Measuring Devices

The devices that measure the potatoes or the fries are the Welliver cameras and de Véryx optical sorter. The Welliver cameras are positioned in front of the cutters. The Véryx is place after the nubbins sieve at the end of the cutting area.

2.2.1 Welliver Cameras

The six cameras provide a three-dimensional picture of each individual cutting lane belt. The cameras turn off, whilst this 3D picture is sent to a computer, which extracts the data and performs the measurements of the dimensions of each potato. The found dimensions data is sent to the memory of the system. After this operation is finished, the cameras turn on and the routine is repeated.

When five minutes have passed the collected data from the last five minutes is send to the data server, visualized, and ready to be extracted for continued evaluation. There is a need to have the data accessible and updated consistently every 5 minutes, this is currently not the case.

2.2.2 Véryx

The Véryx is an optical sorting machine. In total, three Véryx machines are installed to keep up to suffice capacity needs. This machine evaluates each fry passing and filters out defect fries. Besides the defect analysis, the Véryx is also able to measure each fry' length. This data is initially saved in de memory of the machine and sent to the data server to be extracted every five minutes.

2.3 Damaging Fries

Damaging fries is negatively correlated with the average fries length and positively correlated with the number of fries. For measuring the performance of the cutting area, it is relevant to know what might influence the damage done to the fries. This way the information shown, can better support the operators' decision-making.

2.3.1 Critical Points

The critical points in the production line in between the two measurement points, consists of the cutting operation itself, fall height and sieving. The tube choice and the sharpness of the knives impact quality. However, fall height of the fries on conveyor belts can cause harm. Furthermore, the sieving is done by shaking the sieves in a rapid motion, fries can be thrown around and might break.

From the documentation about the cutting area provide by the company itself. It was found that of these three possibilities the tube choice and the sharpness of the knives are most influential to the length quality of the fries. The other two critical points mentioned are not applicable in this part of the production line. Due to the fries being rather flexible at this stage, thus being able to manage the forces they face. As shown in figure 6.

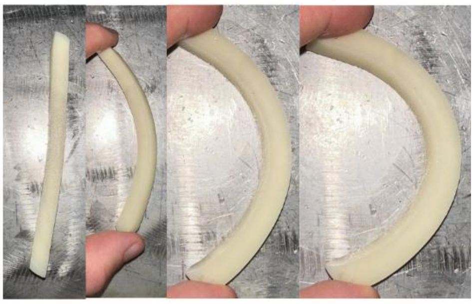


Figure 6: No shattering when bending the fry directly after cutting

The cutting procedure is done with quite some force. About one bar of water pressure is used during cutting. This force together with the friction from the tubes and the knives negatively affects the length of the fries. Since this this is the only operation that influences the fries length with such significance. It shows that providing information about the cutting area performance is a valid identifier to find the correct machine settings.

2.4 Possibilities for Data Evaluation

First it needs to be determined what model could be build with the data at hand. For this a theoretical data workflow will be made. A data workflow shows what data can be accessed and used, what operations to include, and what results are expected. The data workflow can be found in figure 7

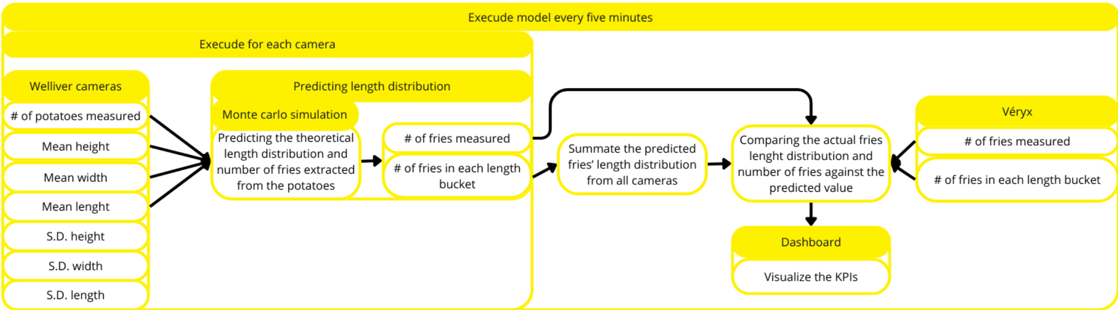


Figure 7: Theoretical data workflow (S.D. stands for standard deviation)

Every five minutes, the model will execute and update its output. The interval is chosen based on the limitations of the Welliver cameras and Véryxs, mentioned in section 2.2.

The Welliver camera system consists of six separate cameras, which all collect their own data. The Welliver cameras will extract seven data points per camera to serve as input for the model. These are the number of potatoes measured and the mean and standard deviation of each largest dimension.

The mean and standard deviation of each dimension will be used to perform a Monte Carlo simulation for many potatoes. This simulation will model a potato, calculate the number of fries extracted and the length of each fry. Performing this simulation for many potatoes, will result in a predicted fries' length distribution, and the predicted number of fries extracted.

This simulation is performed for each Welliver camera. The results from each simulation will be collected and summated.

The Véryx will collect the actual number of fries measured, and the actual fries' length distribution for buckets of 5 mm.

This data from the simulation and the data from the Véryx will be compared and evaluated. The information that can be extracted from this will be converted and visualized into a dashboard.

2.5 Conclusion Situation Analysis

It current configuration of the production process was laid out. Also, the introduction of the new robotic halver was evaluated as a new machine that will be introduced on short notice. The research solution should be able to handle the so called “to be” process, including the robotic halver.

The measuring devices are identified and was found that both the Welliver cameras as well as the Véryx will provide the gathered data regarding dimensions of the potatoes and lengths of the fries respectively, every 5 minutes.

It was found that the only potential critical point within the process line is the cutting operation itself. The friction that occurs between the potatoes and the knives and tubes, and the force that is used to cut the potatoes, negatively influence the length of the fries. Other evaluated points did not show significant signs of harming the fries due to the flexibility of the fries detected after cutting.

The developed theoretical data workflow shows which data will be used as inputs for the model and visualizes the steps that the model will perform. This provides clear directions for model development by laying out each step of the model in a structured and understandable manner.

3. Theoretical Background

In this chapter the aim is to answer the sub research question: "How can a potato be mathematically interpreted?". This question will be answered with the use of existing literature, a mathematical derivation, and is after applied to 100 potatoes to validate the performance.

3.1 Mathematical Approximation of a Potato

To make sure that the computations of the prediction model are accurate. The approximation of shape and volume of potatoes should be considered. A method to approximate the shape and volume of different potato varieties is required. In this research the best approximation for the varieties: King Russet and Fontane will be determined. From figure 8 it can be found that these are the most used varieties in the Lelystad facility, except from the Innovator. However the use of the innovator will decrease in the upcoming years and be replaced by the King Russet, so it was decided to exclude this variety.

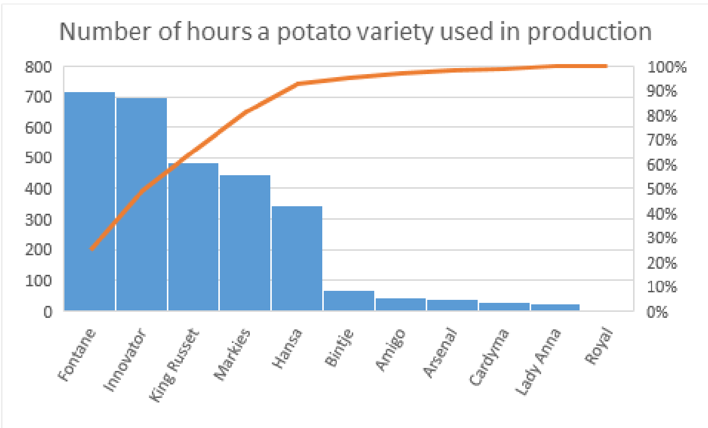


Figure 8: Distribution of varieties, measured in hours of production

3.1.1 Shape and Volume Approximation

According to (Somsen et al., 2004), the potato can be assumed to be like a perfect ellipsoid. This assumption is also used in more recent work of Banagaaya et al., 2014, Yu et al., 2022, Sari & Gofuku, 2023 and Sharma et al., 2025. Though, in the paper of Somsen et al., 2004, they later find that the assumption for a perfect ellipsoid can not be entirely correct, since their results show otherwise. With the use of a regression analysis, they found that the volume of a potato was about 4% higher than the volumetric assumption would suggest using the formula (1) to calculate the volume of a perfect ellipsoid. In the work of Sari & Gofuku, 2023 their perfect ellipsoid fitting to a potato showed to deviate with an Absolute Relative Error (ARE) of 10.9%. Showing an even larger deviation than in the work of Somsen et al., 2004.

$$V = k * L * W * H \quad (1)$$

In a perfect ellipsoid the value for k = $\pi/6$. The value for k, was found to be 4% higher, therefore it is written down as a variable that can be adjusted for the specific potato variety used. In the work of Somsen et al., 2004, only three varieties are discussed. A volume approximation on each potato variety used by the company must therefore be performed to approximation accuracy. L stands for Length, W stands for Width and H stands for Height.

In the work of Banagaaya et al., 2014, the volume approximation is also based on the assumption that a potato can be approximated elliptically. Based on data points, like length, width and height. Like the data provided for this research. They found that the estimated volume, based on the epileptically assumption, resembles the volume approximation by measuring the weight and using the potato density quite accurately. They found that this geometric model can be further improved by finding a better description of the potatoes. Showing that considering the increased volume, determined in “Manufacturing of par-fried French-fries: Part 1” (Somsen et al., 2004) ,is relevant to consider improving these approximations.

The use of an ellipsoidal shape seems to be a justified basis to build upon. During observations in the facility, it was found that the shape of the potatoes is generally wider top and bottom in diameter than a perfect ellipse, as shown in figure 9 and 10.

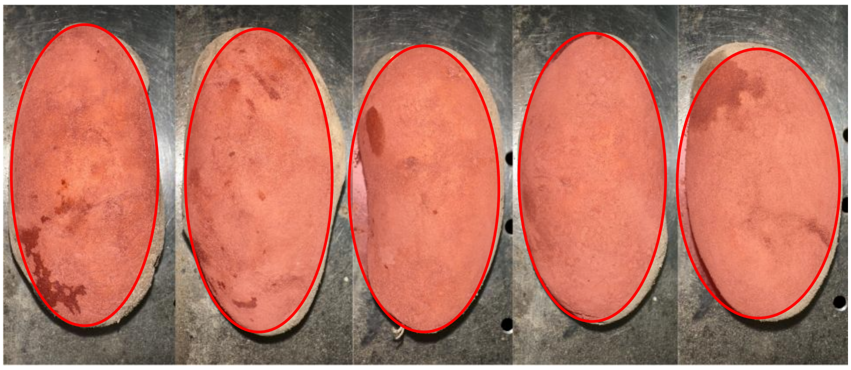


Figure 9: Potato and elliptical estimation (top view)

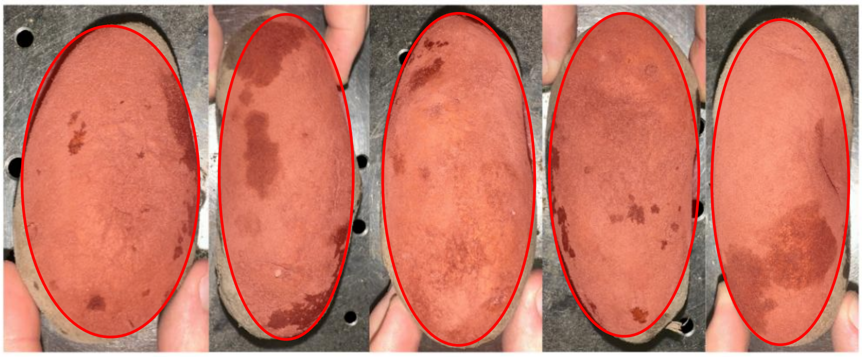


Figure 10: Potato and elliptical estimation (side view)

It might be beneficial to compensate for this fact and improve the approximation of the shape of the potato. The use of a super-elliptical volumetric approximation must be considered. In figure 11, the difference between the two shapes is visualized.

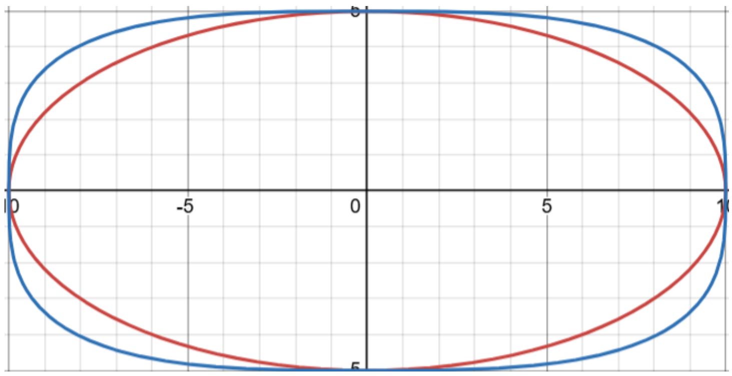


Figure 11: Super-elliptical used compared to ellipse. Red shows the perfect ellipse. The blue line shows the super-elliptical shape.

The formula of the super-ellipse (3) shows quite some resemblance to the formula of the perfect ellipse (2). The difference is the variable power term that is introduced, and the use of absolute value operator, otherwise the super-elliptical shape will not be found for certain values of “m” due to its mathematical behaviour. The “a” and “b” stand for the radius of the ellipse in the “x” or “y” axis respectively. Looking at figure 11, the value of a=10 and b=5 is used as an example.

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1 \quad (2)$$

$$abs\left(\frac{x}{a}\right)^m + abs\left(\frac{y}{b}\right)^m = 1 \quad (3)$$

To create a super-ellipse that is more rectangular of shape, a value larger than 2 should be chosen for “m”. However, it could be possible that a better approximation has a value lower than 2 for “m”. By using this method, it seems likely that a more correct approximation can be found. Therefore, the choice to approximate the exponent “m” is substantiated in order to find the value that approximates the shape best.

3.1.2 Exponent Approximation

The method to perform the exponent approximation is explained in further detail, since this enables the model to be used with multiple varieties in a systematic manner.

To be able to perform the exponent approximation for each potato variety, a standardized and repeatable method needs to be established. How the measurements of potatoes need to be taken, and the computation of the exponent must be decided.

Still one assumption stays, that is that the volumetric object is symmetric in all its planes and is approximated with the use of one value for the exponent. Since the approximation is based on a three-dimensional object, the following equation is considered:

$$abs\left(\frac{x}{a}\right)^m + abs\left(\frac{y}{b}\right)^m + abs\left(\frac{z}{c}\right)^m = 1 \quad (4)$$

The a, b, and c, represent the half of length, width, and height, respectively. Those values are found by measurements 1,2, and 3. The exponent of the equation is computed using measurement 4.

For the first measurements of the potatoes, a random sample of potatoes needs to be selected. For this, the standard procedure is to measure 50 potatoes of each variety. The following measurements needs to be taken:

1. Maximum length
2. Maximum width
3. Maximum height
4. Volume

According to Cooper & Schindler, 2014 validity is defined as “the extent to which a test measures what we actually wish to measure” and reliability as “the accuracy and precision of a measurement procedure”.

To ensure reliability, the measurements will be performed using a calliper to measure the length, width and height of the potato. The volume is measured using an overflow container, of which the water is collected and weighed on a scale with a precision of (±0.2 grams) according to the specifications of the scale. Minor human error can still occur in identifying exact reading from the calliper or ensuring consistent water displacement. However, the use of precise instruments ensures that the results are repeatable and therefore reliable.

Validity is ensured by performing measurements on two different potato varieties of which each a sample of 50 potatoes was taken to ensure external variability, also described in Cooper & Schindler. Splitting the sample into two groups of 30 train samples and 20 test samples will ensure validity, by removing bias in outcome, to compare performance.

The theoretical approach to find the volume of a super-ellipsoid is derived from the work of Jaklič et al., 2000, there the volume is described as:

$$V = 2 * a_1 * a_2 * a_3 * \varepsilon_1 * \varepsilon_2 * \beta\left(\frac{\varepsilon_1}{2} + 1, \varepsilon_1\right) * \beta\left(\frac{\varepsilon_2}{2}, \frac{\varepsilon_2}{2}\right) \quad (5)$$

This formula encloses all the volumes that exists for different exponents for "m" (m influences the rectangularity of the ellipse) in (4). However, this is not the case in this research, thus:

$$\varepsilon_1 = \varepsilon_2 \quad (6)$$

Therefore, ε will be used in the continuation. The relationship between the beta and gamma function is as follows:

$$\beta(Z_1, Z_2) = \frac{\Gamma(Z_1) * \Gamma(Z_2)}{\Gamma(Z_1 + Z_2)} \quad (7)$$

Thus, the original formula for the volume can be rewritten as:

$$V = 2 * a_1 * a_2 * a_3 * \varepsilon^2 * \frac{\Gamma\left(\frac{\varepsilon_1}{2} + 1\right) * \Gamma(\varepsilon_1)}{\Gamma\left(\frac{\varepsilon_1}{2} + 1 + \varepsilon_1\right)} * \frac{\Gamma\left(\frac{\varepsilon_2}{2}\right) * \Gamma\left(\frac{\varepsilon_2}{2}\right)}{\Gamma\left(\frac{\varepsilon_2}{2} + \frac{\varepsilon_2}{2}\right)} \quad (8)$$

From equation (2.15) in the work of Jaklič et al., 2000 it can be determined that:

$$\varepsilon = \frac{2}{m} \quad (9)$$

Resulting in:

$$V = 2 * a_1 * a_2 * a_3 * \left(\frac{2}{m}\right)^2 * \frac{\Gamma\left(\frac{1}{m} + 1\right) * \Gamma\left(\frac{2}{m}\right)}{\Gamma\left(\frac{1}{m} + 1 + \frac{2}{m}\right)} * \frac{\Gamma\left(\frac{1}{m}\right) * \Gamma\left(\frac{1}{m}\right)}{\Gamma\left(\frac{1}{m} + \frac{1}{m}\right)} \quad (10)$$

Using the mathematical rule:

$$\Gamma(n+1) = n\Gamma(n) \quad (11)$$

Gives:

$$V = \frac{8}{m^2} * a_1 * a_2 * a_3 * \frac{\frac{1}{m}}{\frac{3}{m}} * \frac{\Gamma\left(\frac{1}{m}\right) * \left(\Gamma\left(\frac{1}{m}\right)\right)^2}{\Gamma\left(\frac{3}{m}\right)} \quad (12)$$

Simplifies to:

$$V = \frac{8}{3m^2} * a_1 * a_2 * a_3 * \frac{\left(\Gamma\left(\frac{1}{m}\right)\right)^3}{\Gamma\left(\frac{3}{m}\right)} \qquad (13)$$

This formula can now be applied and incorporated to find the exponent "m" based on the volume. The values of a_1, a_2, and a_3 represent the half of length, width and height. This should be a more accurate approximation, than the common assumption made by other researchers that a potato can be described as a perfect ellipsoid.

One example of the calculation from the excel file (Appendix A/B) will be given for the approximation of the exponent:

Length (L) measured	Width (W) measured	Height (H) measured	Volume	Calculated volume	Difference	m
62	61	48	85.8	((8/3*m^2)*(L/2)*(W/2)*(H/2)*((gamma(1/m))^3)/gamma(3/m))*0.001	Abs(volume-calculated volume)	Solver: 1.859

Table 1: Exponent approximation

In the extended computation in the excel file, the difference is used to approximate the exponent. The solver is set to minimize the difference and is only allowed to change the value of "m".

The solver finds one value of "m" for all potatoes. This "m" gives the best approximation of the potato variety.

3.1.3 Results and Comparison

In appendix A and B, the measurements and computations of the varieties King Russet and Fontane respectively can be found. The exponents are approximated and averaged to be used in further calculations. It was found that the exponents differ significantly from each other. The King Russet averages at about 2.137, whilst for the Fontane a value of 1.859 was found.

To put that into perspective, using the measurements of 100, 55, and 50 mm for the length, width and height, respectively. A ARE of 12% was found using these values for "m". This clearly suggest that differences between varieties need to be accounted for.

The data from both varieties was split up in 30 train and 20 test values to remove the training bias. The best value of "m" of the training data was approximated and used to predict the volume of the 20 test potatoes to measure the average percentage difference. For both varieties, the percentage difference was calculated using the average exponents found. For the King Russet an average percentage difference of 7.1% was found, and for Fontane an ARE of 7.1% was found. This result is worse compared to the results of Somsen et al., 2004, where a ARE of 4% was found, however these results are better compared to the ARE found in the work of Sari & Gofuku, 2023.

To find whether this novel approach is indeed better or worse than the perfect ellipsoid assumption made by Somsen et al., 2004 the ARE using a perfect ellipsoid was computed. For the King Russet and Fontane the values of 11.5% and 5.8% respectively were found. This does not clearly show that the new method consistently improves the approximation based on Somsen et al., 2004, however show better results than Sari & Gofuku, 2023. This new method seems more robust and can improve the approximation when potatoes are notably more rectangular than ellipsoidal.

3.2 Conclusion Theoretical Background

It was found that the super-ellipsoid approach is able to adjust the shape op the potato according to the measured volume. The sample that was taken from two varieties showed different results. The King Russet seemed to benefit from this new shape approximation, whilst the Fontane showed worse results than the assumption of a perfect ellipsoid. Therefore, it can not be said with certainty that the new approximation is better or worse than de perfect ellipsoid assumption made by (Somsen et al., 2004).

The King Russet variety is known to be more rectangular shaped compared to other varieties. For varieties that are notably more rectangular shaped than an ellipse, it might be a more robust choice to use a super-ellipsoid approach instead of assuming the shape to be a perfect ellipsoid.

This indicates that, although the method is based on a sound idea, it still requires further improvement and testing on larger sample to confirm its overall accuracy and usefulness. Therefore, it is recommended to increase the sample size of a potato variety and redo the approximation of “m”. This should enable to better predict the length and quantity of the fries in the model.

It was chosen to continue using the new shape approximation in future measurements made in this report, since King Russet will be used as the sampled potato variety later on.

4. Solution Configuration

In section 4.1 a simulation method is found that is able to estimate the predicted number of fries and their lengths. Section 4.2 explains the modelling methodology. In section 4.3 a theoretical approach to determine how two types of cutting knives perform compared to each other. The simulation model was partly reconstructed to measure the volume loss caused by the different cutting knives. The results from this will be used to select the right knife to use in process (necessary information to determine the third KPI, introduced in section 1.3.3).

4.1 Solution Choice

The simulation choice is a key factor in finding the accuracy of the predicted data, as well as the limiting factor for the computational load/speed of the model. In the work of Banagaaya et al., 2014, a similar case was solved through a simulation method. Their decision was to make use of a Monte Carlo simulation approach. Their results seem to suffice the needs of the application in this study. The Monte Carlo simulation allows the input of data distributed according to a probability density function and can predict fries quantities for many potatoes.

Another option that was considered in this research, was the use of a copula. This is a one-dimensional distribution that originates from connecting multivariate input values. Though this method has a much lower computational cost, it is hard to determine the correct values without an extremely large sample size. A copula was used in section 3.1.3, to determine the value of “m”, it was found that the test data shows quite some difference with the train data, for this small sample size. Providing a reason to assume that the use of a copula may not yield sufficient accuracy

In the work of Takahashi et al., 2025, they tried to predict the fruit size of a tomato some time before harvest, using machine learning algorithms. This shows that there is ground to believe that predicting the number and length of fries extracted from a potato is possible. This method is able to adapt to real data well, however requires a large training and test set which requires a large number of manual measurements. This large number of measurements makes this option not feasible to consider in this research.

Danielak et al., 2023 found that the shape of a potato can be classified with the use of 3D cameras. Exact 3D potato data can be used to cut the potato virtually and determine the fries’ length distribution. This is presumably significantly accurate, as is found in Su et al., 2017. They reached a mass prediction, based on the determined volume, of a 4.4% mean percentage error. The data used was depth images of potatoes. Due to the current setup of the system it is not possible to extract the depth images from the cameras, thus this approach is not an option in this research.

Concluding, the use of a Monte Carlo simulation has already been used and proved its suitability in a similar case. The use of a copula was not found in literature for a similar case. Also, the function of a copula in a previous section showed that a copula shows an underperformance between the training and test set. The work of Takahashi et al., 2025 indicates that it is possible to predict fries’ length distribution using a machine learning approach. Their approach is also able to handle the input parameters available by the current system in use by the company. The downside is the fact that a large sample size to be measured by hand is needed to train the model, which is not feasible in the timespan available. The approach to use 3D imagery as input has a high potential to provide accurate results, however the required input data can not be provided by the current system, thus excluding this approach. It was therefore decided that a Monte Carlo simulation was best fit for this research and within limitations of data availability and time constraint.

4.2 Modelling Methodology

The model has four components.

1. Extracting the input data
2. Predicting the theoretical output of fries and their lengths
3. Comparison between the predicted output and the actual output
4. Visualization of the performance parameters

Components one and four are self explanatory since they only consist of copying data from the data provided by the company or show the results of the calculations made in component three. Therefore, only a more detailed explanation will be constructed in this chapter about predicting the theoretical output, and the comparison of the predicted and actual output.

One good thing to mention is that the model is built to handle the “to be” case of the physical configuration, by setting the length to half potatoes to a relevant value so some of the potatoes are cut. When the current physical configuration is used a large value can be used so the simulation does not cut the potatoes in half, since all potatoes are halved before being seen by the Welliver cameras.

4.2.1 Predicting the Theoretical Output

To predict the theoretical output, a Monte Carlo simulation is used. This simulation can, based on the input given from the actual data of the Welliver cameras, simulate potatoes and the number of fries extracted and their lengths for a large quantity of potatoes N. The idea is that by using a large quantity of N, the variance of the potatoes’ length, width and height and variance in the combination of these dimensions are all well represented. Therefore, returning a good representation of reality.

To find the expected fry’ quantity and associated fry’ lengths the simulation performs the following procedure.

Some fixed parameters are set first. These are:

1. N (number of potatoes to simulate)
2. m (exponent of the super ellipsoidal approach (section 3.1.3), based on the potato variety)
3. Cut size (fry dimension in width and height (square fries))
4. Sliver fraction (minimum fraction of cut size a fry needs to be, to be considered a fry)
5. Length to halve potato (value for the length of the potato the robotic halver will cut)
6. Type of knife used (centred or crosshair knife (section 4.3)

After these parameters are set the dimensions of a potato are sampled. This is done by using a normal distribution (which is an assumption that is widely used within the company) with the input parameters being the mean and standard deviation from each of the dimensions provided by the Welliver cameras.

The model assumes that the potato is symmetric in the XY-plane. To reduce the computational complexity of the model. In figure 12 a visual representation can be found of this idea.

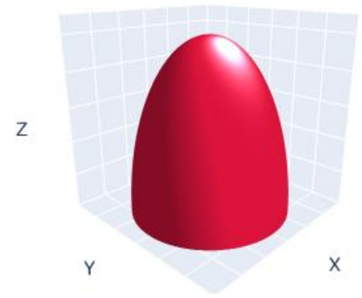


Figure 12: Visual representation of the top part of a modelled potato

From this halved ellipsoid a slice of 1 mm is taken. Shown in figure 13.

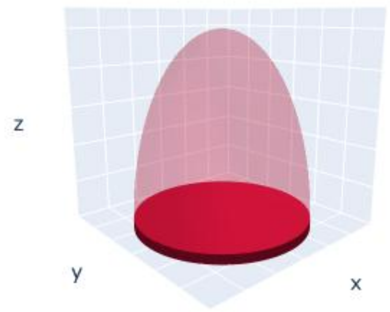


Figure 13: Visual representation of a sliced potato

A grid that represents the cutting blades is added to the slice of the potato. Shown in figure 14.

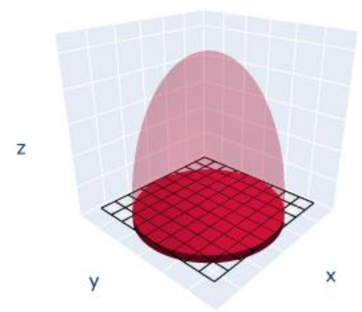


Figure 14: Visual representation of the cutting grid over the potato slice

Then the number of fry parts (grid squares) that are filled by the slice for at least the sliver fraction are marked green, when the grid square is filled less than the sliver fraction it is marked red. In figure 15 a sliver fraction of 0.3 is used as example.

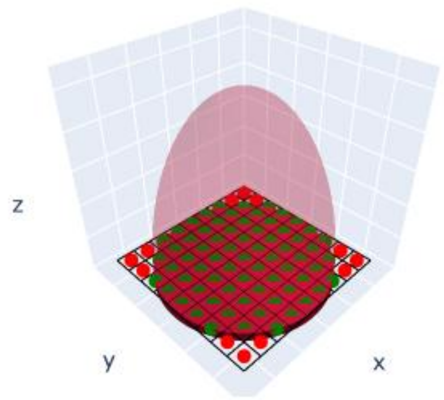


Figure 15: Visual representation of determining the fries inside the slice

The model then takes a count of the number of grid square that have a green marker in figure 15 and turns the grid square to "active". The model continues to a slice with the same thickness, directly above the current slice and performs the same calculation. When a grid was active in slice z-1 and is not filled by any fraction of the slice at z, it will be deactivated and will not be counted any further. In figure 16 a slice higher on the z-axis is shows.

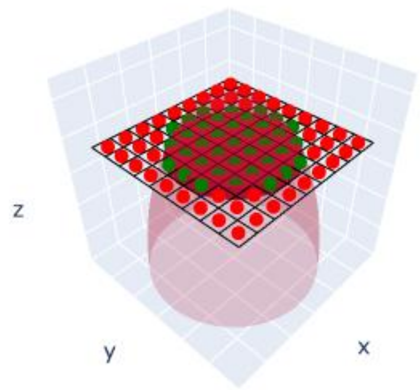


Figure 16: Visual representation of finding the fries inside the slice at a higher point on the z-axis

The count of fries for all slices are now known. The model will now check whether the count decreases when comparing the counts at slice z and z+1. When the count decreases the number where the count decreases with and the value of the previous z will be taken and written in a new array with quantity and half length.

It was decided to measure the length of the fries only at one side of the potato and later multiply the half length with 2 to find the final length of the fries from that potato. One added benefit is that accounting for potatoes that are halved by the robotic halver is easier. In the last case the count of the fries will be multiplied by 2, while the lengths stay the same. This is only done for the part of the potato that is actually cut (see section 2.1)

This computation will be performed for N number of times. The results of all fries' lengths of all potatoes will be visualized to represent the predicted distribution of fries.

An example of a distribution where some of the potatoes are also halved is shown in buckets of 5mm figure 17.

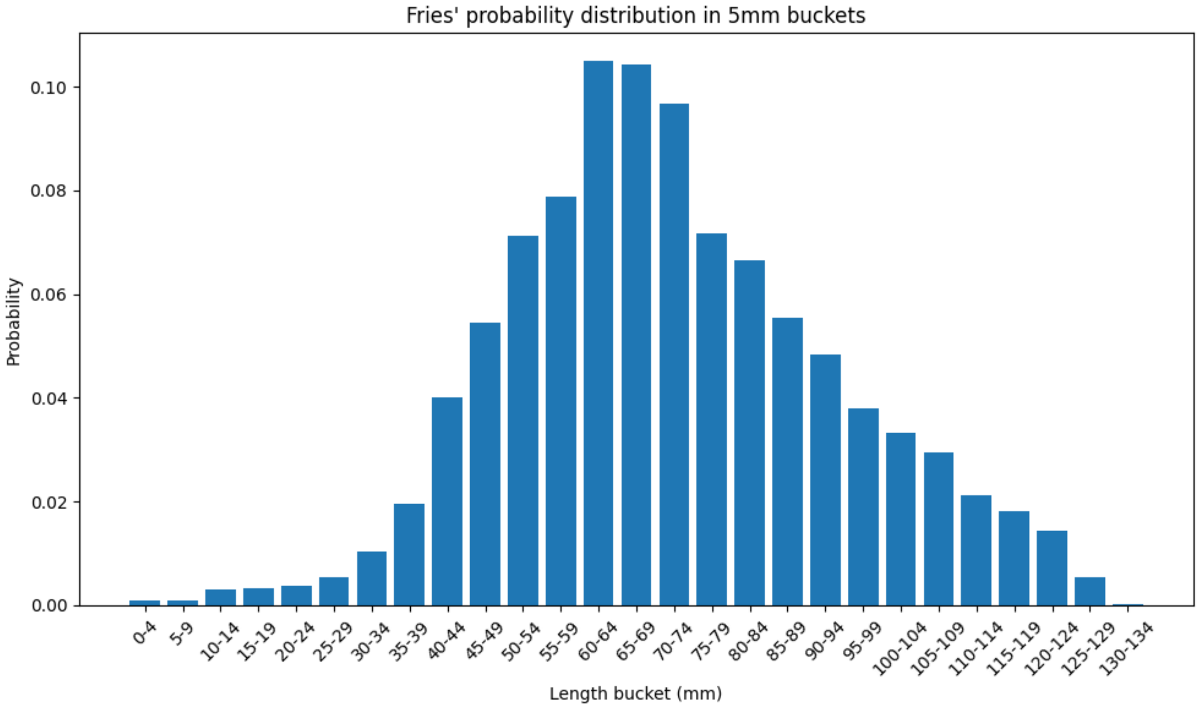


Figure 17: Possible distribution found using the model

4.2.1.1 Simulation Validation on Fries' Distribution

To measure the validity of the simulation model a sample of 50 potatoes is taken. The potatoes are chosen based on their width dimension, these need to be close to each other and within the parameters similar to tubes sizes in the actual configuration (±6 mm). The potato measurements are taken are the same as in section 3.1.2. It was decided to perform the measurement on the potato variety "King Russet" since this variety has shown the best results in the exponent approximation compared to the perfect ellipsoid assumption. Another reason is the larger quantity of fries that can be extracted from each potato on average, due to the larger size, thus providing more data to be used.

The collected data regarding dimensions and volume is analysed in excel to determine the "m-value" and the mean and standard deviation of each dimension of the sample (see appendix C for the sample results, exponent approximation, mean and standard deviation). The potatoes continue on to the cutting knife. This is a special fries knife that can be operated manually, ensuring that slivers and nubbins will correctly get separated from the fries before performing the fries length measurements. All fries will be collected and measured by the length analyser, which is able to determine the length of all fries and categorize them in buckets of 5 mm. This machine is the standard within the company to perform length measurements.

The simulated data distribution is compared to the actual distribution. With the use of the Chi squared test and a confidence interval of 95% it will be determined whether the distributions are similar.

It will be tested against the hypothesis of:

H_0 : the simulated distribution ≠ actual sample distribution

H_1 : the simulated distribution = actual sample distribution

The H_0 must be rejected when the found Chi-squared value is less or equal to the critical value.

When H_0 can be rejected it can be said that: "With 95% certainty, it is found that there is sufficient reason to reject the H_0, and the model's predictions for the number of each length bucket of fries is assumed to be consistent with the observed experimental data.". The model is deemed to be valid.

This validation test is performed twice. Once for the exact potato dimensions and once using a normal distribution and the found mean and standard deviation from the sample.

Additional to the Chi-squared test, a Kolmogorov-Smirnov test was applied to check whether the output data followed the same distribution, based on the shape of the distribution.

4.2.1.1.1 Results of simulation validation based on fries' distribution

The result of this validation process can be found in appendix D and E, those tables show the averages of 50 runs and the calculated Chi-squared value for the exact potato dimensions and the normally distributed dimensions using the found means and standard deviations respectively. The critical value to test the found value against was **28.87**.

For the exact potato dimensions the Chi-squared value of **182.74** was found. Thus, failing to reject the H_0.

For the normally distributed dimensions using the means and standard deviations, the value for the average of 50 runs was found to be **401.27**, this value is much higher than the critical value. Therefore, the H_0 can not be rejected. Thus, the model can not be deemed valid in its current state.

The Kolmogorov-Smirnov test was also applied to 50 simulation runs and the actual output data. It was found that none of the simulated distributions were in accordance with the actual fries' distribution. The simulation output and actual output therefore to not follow the same distribution based on the shape of the distribution. One example of the Kolmogorov-Smirnov test can be found in appendix F.

The results of the fries' distribution as result of the simulation approach is visualized with the actual results in figure 18, using normally distributed dimensions.

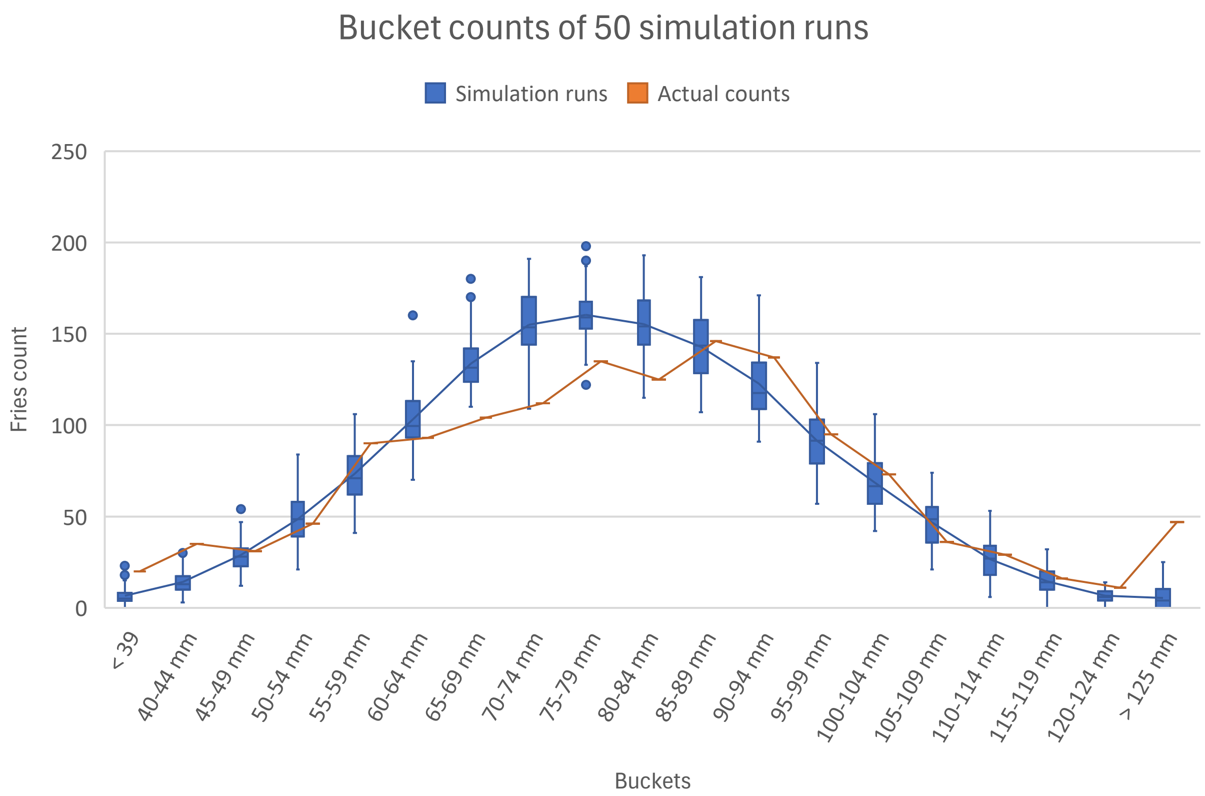


Figure 18: Results simulation VS Actual bucket counts

It can be seen that the validity of the model cannot be proved looking at the two lines according to the Chi-squared test. Though the simulation is not valid, it does show little variation in its output (relative small boxes), arguing that that model is reliable.

Even though the fries' length distribution can not be assumed to be similar, the simulated fry count is within 1.8% over the average counts of the 50 runs.

4.2.1.2 Simulation Validation on Sample Size for Manual Measurements

To determine whether the sample size was sufficiently large. A Chi-squared test on the fries' distribution will be performed using m=2.055 and the exact potato dimensions and the normally distributed mean with standard deviation found from the sample. For the simulated potatoes with the input of a normal distribution with calculated mean and SD it was decided to take the average of 50 simulation runs, to make sure that the Chi-squared test was not performed using an outlier of the general fries' distribution. The option where the exact potato dimensions were used were not simulated 50 times, since all simulation runs use the exact same input data, therefore returning the same results.

The Chi-squared test will be performed for:

H_0 : the distribution using normally distributed inputs ≠ distribution from exact potato dimensions

H_1 : the distribution using normally distributed inputs = distribution from exact potato dimensions

4.2.1.2.1 Results of Simulation Validation on Sample Size

The result of this second validation test can be found in appendix G. The same buckets are used as the previous validation test, therefore the critical Chi-squared value remains the same: **28.87**. The calculated Chi-squared value is **36.29**. This means that the H_0 can not be rejected.

From failing to reject the H_0 it can be concluded that the sample size is statistically not significant to be captured with the use of a normal distribution and the mean and standard deviation.

4.2.1.3 Conclusion

From the first validation test it was found that the model performs well to determine the expected number of fries, however is unable to predict the length of all fries accurately. This can be caused by two things. Either the shape approximation is not sufficiently accurate or the sample size used to determine the input parameters is too small.

To determine whether the shape approximation was sufficiently accurate, both types of inputs were evaluated. Found was that in both cases H_0 can not be rejected. From this, it can be concluded that the shape approximation is insufficiently accurate.

Also a Kolmogorov-Smirnov test was applied to see whether the distributions were similar in shape. This turned out not to be the case, since the observed test statistic was larger than the critical value.

Figure 18, showed that the output of the simulation was reliable. However the validity of the simulation output showed not to be significantly accurate.

The second validation test looked into the fries' distribution results using the exact potato dimensions and the normally distributed sample based on the mean and the standard deviation from the exact dimensions. It was found that the sample size was statistically not sufficiently large to be captured in a normal distribution. Thus, either the sample size was too small, or the assumption of normal distributed potato dimensions is invalid.

Therefore, both sample size and the assumption of normally distributed potato dimensions as well as an insufficient potato approximation are potential causes for invalidating the model.

4.2.2 Data Comparison

The distribution in figure 17 shows an example fries' distribution for a camera.

The model assesses the three KPIs mentioned in section 1.3.3, based on the data gathered from the Véryx and simulation model. These KPIs are then visualized in a dashboard for the operators to use.

The model has to be performed once for each of the cameras. The results are from all cameras are summated to determine all of the final KPI values as a total, since the real-life process does the same. Otherwise, the Véryx data and simulated data can not be compared.

4.3 Optimal Knife Choice

The simulation model introduced in section 4.2 has more applications than to be used as a performance measurement tool only. Another application of the simulation model was found in the best knife selection. For this application there was more to be added to the simulation. Instead of only looking at the quantity and length of the final product it was needed to measure the volume more precisely and make a distinction between different slivers, nubbins and fries.

One important note, the model that showed not to be valid has been used for this application. Due to limitations, it was not possible to improve the shape approximation of the potato further. The assumption that a potato can be described as a perfect ellipsoid was used, since using another shape approximation was not thought to change outcome significantly. Therefore, using the found parameters in real-life have the possibility to show other results than found in this research. Further research on the shape approximation must be performed, after which the best knife choice for each tube can be determined using the model.

4.3.1 Knife Differences

First, the distinction between a centered and a crosshair knife is established. In figure 19 the difference between the two knives is shown.

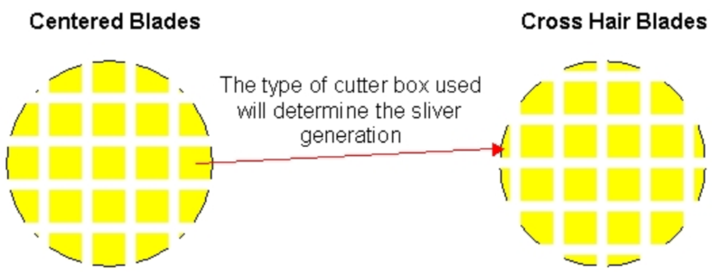


Figure 19: Distinction between a centered and crosshair knife

On the left side the centered knife, the cutting grid of this knife is placed around the centre of the potato. This leads to a different alignment compared to the crosshair knife. The crosshair knife is aligned to go through the centre of the potato in stead of around.

Based on the width of the potato, it is thought that an optimum exists in the knife choice, since the width determines the best alignment with a certain kind of knife. Ultimately decreasing/limiting the losses of nubbins and slivers, leading to an increase in output.

4.3.2 Adaptations to the Model

The model was changed to be able to manage both a centered and a crosshair cut. This was done by offsetting the cutting grid for the centered knife by half a cutting grid in both x and y directions.

To make it possible to record and track the losses during cutting, some additions to the model had to be made. It is needed to track the volume and classify each strip.

Tracking the volume was done by determining the fraction of a grid square of the cutting grid that was filled in with the potato flesh. Shown in figure 20 .

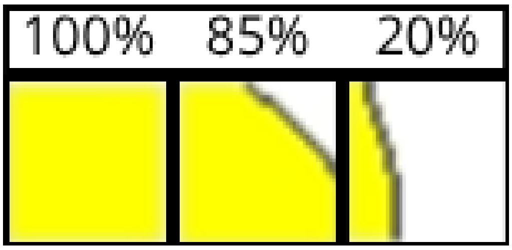


Figure 20: Fraction of cutting grid filled with potato flesh

Each filled fraction was found and multiplied by the grid size to approximate the volume of that sliver, nubbin or fry in that specific layer z. This was done for each layer. All fries that have a fraction of less than in this case a fraction of 0.3 or 30% will be considered a sliver. All fries that remain and are shorter than 50 mm will be considered nubbins. All remaining fries are considered actual fries.

4.3.3 Simulation Runs

The simulation was run using the following input parameters:

1. Length, normally distributed with (mean, Standard Deviation) (100,30)
2. Width, ran for a value between 30 and 108 (based on the tube sizes available)
3. Height, normally distributed (Width, 10)
4. No potato halving
5. The potatoes were assumed to be a perfect ellipsoid
6. Number of potatoes per simulation (once for each width and knife type), 1000
7. Cutting/grid sizes, 7x7mm, 10x10mm, 14x14mm

The simulation runs took between 25 and 56 minutes to finish, running on a Lenovo ThinkPad p15v. The smaller the grid size the longer the computation takes. The simulation was run for each width value between 30 a 108 for both centered and crosshair knives. In total there were 156.000 potatoes simulated per cut-width to draw the results and conclusions from.

4.3.4 Simulation Results

The simulations show the following results in sliver and nubbin (volume) losses as percentage of the whole volume, for 7x7mm, 10x010mm, and 14x14mm respectively in figure 21 . In appendix H the figures are shown in large.

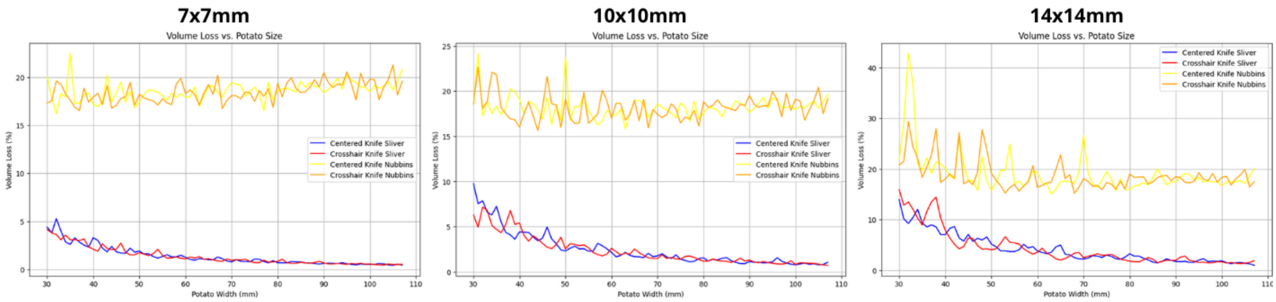


Figure 21: Sliver and nubbin losses for the three cutting sizes compared

The results show that the cutting size influences the total volume losses. The larger the cutting size, the more volume losses occur. This seems logical, using a larger cutting size leads to more fries being categorized as slivers. The nubbin losses also increase when the cut size increases. Especially when the width of the potatoes is small and the cutting size is larger, the nubbin losses increase.

Though the theory (internal documents) classifies nubbins as direct defects and should therefore be filtered out. The specifications of the end product allow a certain percentage of nubbins to be present. Only slivers and small nubbins (length <25 mm) are categorized as length defects. Therefore, it becomes predominantly important to limit the number of slivers created. In figures 22, 23, and 24 the percentage of a reduction of volume losses compared to the other type of knife is shown for each available tube size and interval (±6mm).

The blue bar stands for the centered knives and red the crosshair knives. The legend shows only red, however the transparency of a bar indicates the type of loss.

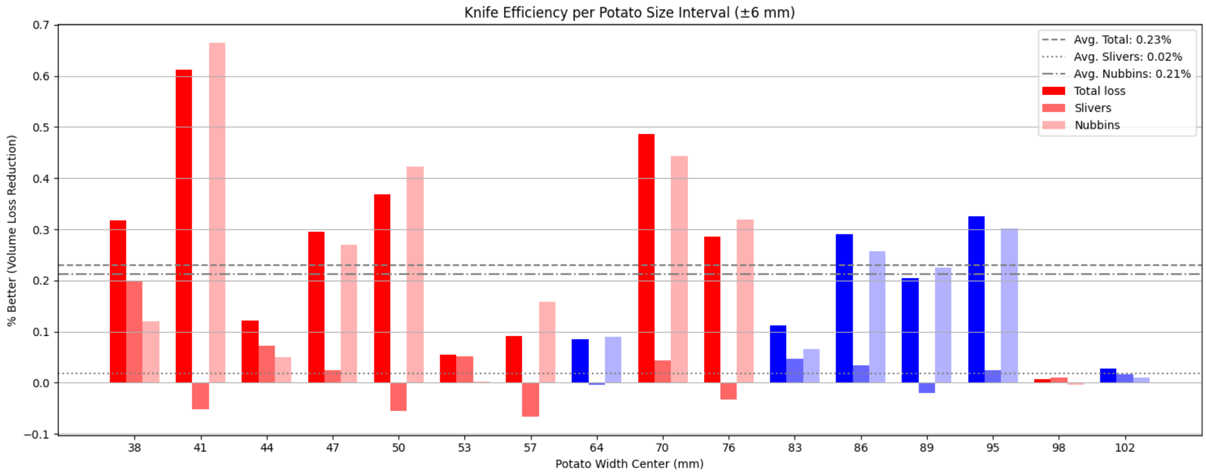


Figure 22: Volume savings 7x7mm

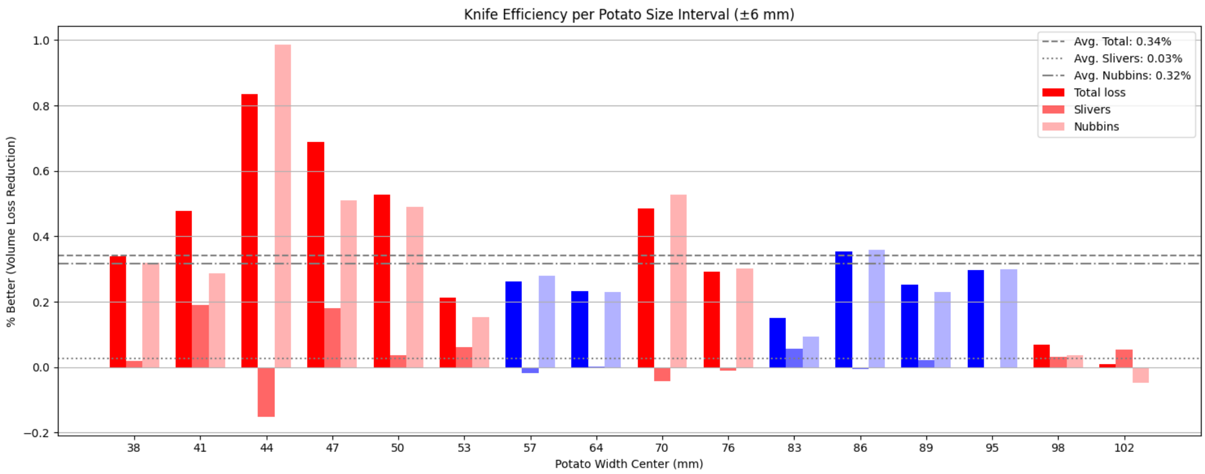


Figure 23: Volume saving 10x10mm

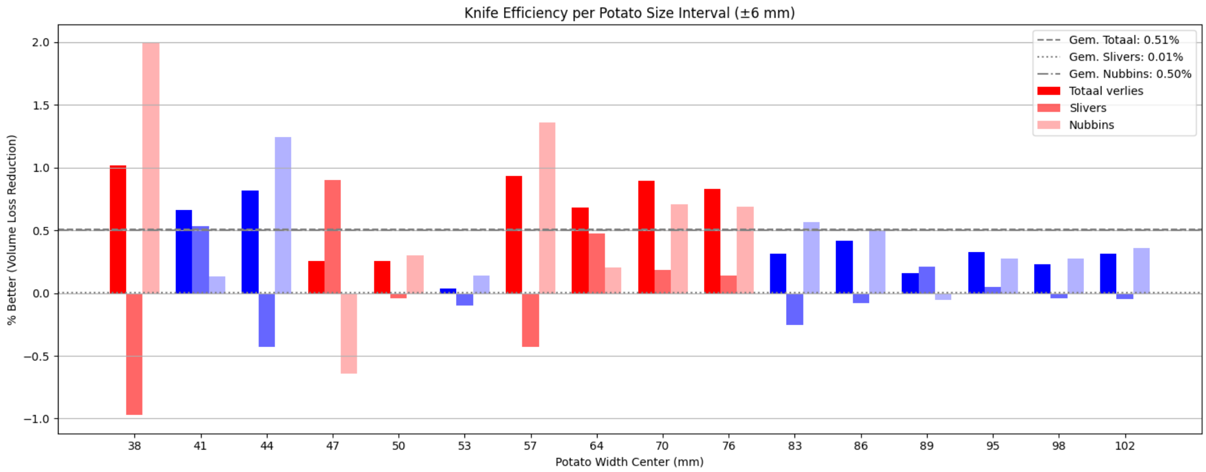


Figure 24: Volume saving 14x14mm

The graphs show for each cutting size and tube widths what knife performs better compared to the other knife configuration. It shows that the crosshair and centered knives each have benefits depending on the tube choice. It was found that for various tube choices the crosshair knives performed better than the centered knives (currently the standard). Switching knife configurations for different tube sizes shows its benefits.

Furthermore, the graphs show some interesting results. First, the knife selection is supported by data and shows that an increased volume can be realised between 0.23% and 0.51%, considering both slivers and nubbins. Another thing that can be learned from this data and knowing that some degree of nubbins is allowed to be present in the end product, is the fact that these results enable the company to steer the cutting area output even further than before by deciding what knife to chose. When according to the Véryx the fries' distribution is within specification, and there is some more room for more nubbins. The company can decide to focus on reducing the slivers, therefore selecting another knife.

For example, the potatoes with a width of 38 and cutting size 14x14mm. The most volume loss reduction can be achieved by using a crosshair knife; however, the graph shows that the centered knife results in almost 1% less sliver losses, and 2% more losses for the nubbins. When 2% additional nubbins will not endanger meeting the specification, it can be decided that focussing on the slivers is more beneficial.

The other way around holds as well, when the company is using the centered knives and length specifications are not met due to too many nubbins present in the end product, it can change the knives to a crosshair knife, reducing the nubbin losses at the cost of increasing the sliver losses.

For most cases it holds that the reduction in sliver and nubbin losses is both positive. For those cases it is not possible to adapt when length specifications are easily met or not. Those cases should always apply the selected knife for that tube size.

4.3.5 Conclusion Knife Choice

When considering only the cutting sizes of 7x7mm, 10x10mm, and 14x14mm, it was found that some improvements can be made with the knife choice, a reduction in volume losses between 0.23% and 0.51%. The results show a reduction of volume and with that mass losses. This will result in a larger output. Another benefit found, was the possibility to adapt the system to the needs at that moment in time. This allows the company to improve on meeting the first time quality or increase its output, based on each individual situation' needs.

4.4 Conclusion Solution Configuration

A Monte Carlo simulation showed the benefits of being able to handle input data using a probability density function and had shown presence in existing literature where it was a suitable option. Therefore, it was chosen to use a Monte Carlo simulation approach in the research solution.

The model methodology visualized and explained the steps that the model should take to determine right extrusion of the fries from the simulated potatoes and determine the fries' distribution.

The model is able to perform a simulation for 1000 potatoes per camera within less than 30 seconds on a Lenovo ThinkPad P15v. Therefore, easily meeting the demand of performing the measurement within 5 minutes.

The current model showed to be invalid within the assumptions. It was found that the results of the fries' length distribution were significantly different between the predicted and actual distributions. Caused by the shape approximation of the potatoes in the simulation model. Which was verified by taking the exact dimensions of the manually measured potatoes and compare the results from the simulation with the actual distribution. Still, the results were significantly different. Thus, it was caused by the insufficiently accurate shape approximation.

Another assumption that had been used was that the manually measured results could be captured using a normal distribution with mean and standard deviation. It was found that these results were also significantly different. Concluding that the sample size was too small to be captured within a normal distribution, or the assumption that the potato dimensions are normally distributed is invalid.

Though the model has been deemed invalid. The simulation to determine the optimal knife to use to limit volume losses during cutting shows results that suggest that output can be increased based on the knife chosen. The differences between knife choices for certain tubes also showed that another method to influence the cutting area output was found. Allowing for more adaptability of the system, with that increasing probability of meeting first time quality more regularly or increase output, based on the situation' needs.

5. Solution Implementation

Before the solution implementation can kick-off, the model should be further developed in accordance with the advice given in the recommendations (section 6.2). In section 5.1 the implementation in the current system will be determined to ensure that the model is used as it is supposed to. Both the backend as well as the frontend of the model will be laid out. In the following section 5.2, the informing and participation of stakeholders is explained.

5.1 Implementation

The company makes use of a general visualization system called "FactoryTalk". This system is connected to all of the machine in the factory. It provides a visual overview of the machine performance and settings. The company wants to enable the insights given by the model to be visualized of FactoryTalk. This system is able to show the KPIs, however this system does not have the computational power to perform the simulation itself. For this a new hardware system must be introduced that has the power to perform the simulations every 5 minutes. Also, the data collected by the Welliver cameras and the Véryx optical sorters, should automatically be available to use. For this there is a connection that must be made between the server that collects and stores the data, and the computation system. The dataflow diagram with all the connections to be made are visualized in figure 25.

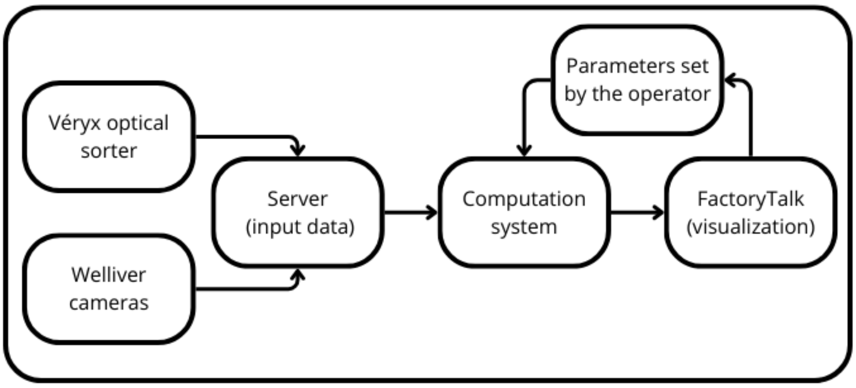


Figure 25: Dataflow diagram visualized

The data from the Welliver cameras, the Véryx and the simulation will be visualized for the operators on FactoryTalk. An example of the dashboard with the relevant KPIs selected by the company can be found in figure 26. The data used is made up and does not represent reality.

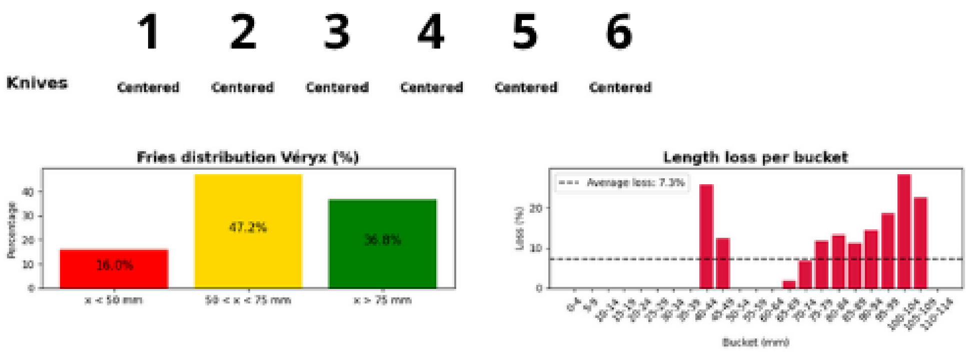


Figure 26: Dashboard with visualizations of the collected data by the Welliver cameras, the Véryx and the simulation

The dashboard shows the following three elements:

1. Optimal knife selection
2. Actual fries' distribution
3. Length loss of fries in 5 mm buckets

Element 1 contains the optimal knife selection for each individual water jet cutter. The knife is selected to pass the first time quality and reduce losses as much as possible. Element 2 shows the actual fries' distribution, operators can use this to determine whether the cutting area is performing in accordance with the set expectations. Element 3 shows the percentile discrepancy between the predicted number of fries and the actual number of fries extracted. When a large discrepancy is found between the two distributions, the average length loss will increase, indicating the operators to act accordingly.

This dashboard will be updated every 5 minutes and use the most current available data.

5.2 Stakeholder Participation

The next step after the implementation of the model, is informing the cutting area operators. Since the model can cause quite some changes compared to their current working environment, it is valuable to create a training course for the operators to attend. The model namely impacts their knife choices, interpreting new available data, and acting according to this new data. The current performance measurements that are done by hand will be replaced. This will lead to structural changes to their operating methods. That is the reason it is advised to introduce the model with a training, where these points are discussed, and where the necessity of the model is explained. This also allows for questions from the operators. According to Baronas & Louis, 1988, user acceptance can be increased by user involvement. Decreasing resistance to change, providing a place for conflict resolution about design issues, and committing users to the system. Also, the work of Hunton & Beeler, 1997 shows that there is a strong positive correlation between human involvement and system performance during introduction as well as gains over a longer period after implementation.

An addition is the evaluation of the model and dashboard after implementation. A survey will be sent out to the operators of the cutting area. This survey can be used as a measure of the implemented system. Allowing discussion on future additions to the system, and if all information is visualized in a way that feels intuitive for the operators. This way it can be determined how valuable the new system is deemed in the eyes of the operators. Since a higher human participation will ensure higher system gains (Hunton & Beeler, 1997). Based on the feedback given, it will be decided whether a discussion session with the operators and the system developers must be organized to better understand the situation. When the situation is understood by both parties, and points for improvement are identified, adjustments can be made. Depending on the significance of the adjustments an optional second discussions sessions/ training session will be organised to reevaluate the system. This cycle will continue until both parties agree on the functionalities of the system.

Concluding that participation of the operators is a much needed element of a succeeding system. Showing the value of a training before implementation and availability of discussion sessions and support (for questions regarding the system) from system development and data analyst during and after implementation.

5.3 Conclusion Solution Implementation

The implementation in section 5.1 shows the back and front-end of a system that is able to measure the performance of the cutting area and inform cutting operators about it. For the back-end implementation, it was mapped which data stream should be built to fulfil the need of the system and evaluate the performance. The front-end looked at the method for visual identifiers that will be shown to the cutter operators.

Stakeholder participation during and after the implementation of the system is beneficial to the later system performance (Hunton & Beeler, 1997). It will be beneficial to introduce the system to the stakeholders by means of a training, to ensure smooth integration into the current working environment. This training should address the following points:

1. Decision making regarding knife choice
2. Interpreting new available data and how to act accordingly
3. Removing the need for manual performance measurements by the operators

After implementation a survey will be send out to collect feedback from the operators. Based on the given feedback, a discussion session will be introduced to understand strengths and weaknesses of the system. System development will make necessary adjustments to the system based on the feedback. When the system has undergo the adjustments an optional second discussions sessions/ training session will be organised to reevaluate the system. This cycle continues until all parties are satisfied.

6. Conclusion, recommendations and limitations

In this chapter the final conclusions, recommendation for the company and limitation of this research will be discussed.

6.1 Conclusion

It was noted that in the current physical configuration and the "to be" configuration there are no critical points for damaging fries due to the flexibility of the fries, except for the cutting operation performed by the water jet cutters. The friction on the potato caused by the tubes and knives, combined with the force on the potato are likely to cause damage to the fries.

The mathematical approximation of a potato using the super-ellipsoid approach has not showed convincing proof of being sufficiently accurate. One sample of 50 potatoes of the King Russet variety showed that the super-ellipsoid approximation improved the shape approximation by 1.3 percent point, whilst the second sample of 50 potatoes of the Fontane variety showed worse results, namely a reduction of 4.4 percent point, using the super-ellipsoid approach compared to the perfect ellipsoid assumption from Somsen et al., 2004.

The model is able to perform a simulation for 2000 potatoes per camera within less than 30 seconds on a Lenovo ThinkPad P15v. Therefore, easily meeting the demand of performing the measurement within 5 minutes.

The current model showed to be invalid within the assumptions. It was found that the results of the fries' length distribution were statistically significantly different between the predicted and actual distributions (see section 4.2.1.1). Caused by the shape approximation of the potatoes in the simulation model. Which was verified by taking the exact dimensions of the manually measured potatoes and compare the results from the simulation with the actual distribution. Still, the results were significantly different. Thus, it was caused by the insufficiently accurate shape approximation.

Another assumption that had been used was that the manually measured results could be captured using a normal distribution with mean and standard deviation. It was found that these results were also significantly different. Concluding that the sample size was too small to be captured within a normal distribution, or the assumption that the potato dimensions are normally distributed is invalid.

However, the quantity estimation was on average within 1.8% of the real count of fries for 50 simulation runs. Thus, the model can accurately predict the fries' quantity.

Statistical model validation could not be proved. The simulation to determine the optimal knife to use to limit volume losses during cutting shows results that suggest that output can be increased based on the knife chosen, assuming a potato can be described by a perfect ellipsoid. The differences between knife choices for certain tubes also showed that another method to influence the cutting area output was found. Allowing for more adaptability of the system, with that increasing probability of meeting first time quality more regularly or increase output, based on the situation' needs. Though it was found that this assumption does not hold looking at the fries' length distribution, it does show its relevance for future research. In the current setting it was found that potential reduction in losses would be between 0.23% and 0.51%. Resulting in significant savings.

6.2 Recommendations

1. The company should make sure that the data that is extracted from both the Welliver cameras as well as the data from the Véryx is correct.
2. Ensure that the interval at which the data is extracted from the Welliver camera is done every 5 minutes, instead of changing interval lengths. To ensure that the model system is able to perform its calculations after consistent intervals.
3. Make changes to the data that is extracted from the Welliver cameras. Instead of assuming that each camera 'sees' normally distributed potato dimensions, check whether this is true. Each lane most likely 'sees' a segment of a normalized truncated normal distribution. Currently the mean and standard deviation is calculated by the cameras based on the potatoes passing underneath each camera. Each lane was therefore evaluated as if the potatoes were normally distributed. However, this is not the case. The whole batch of potatoes is most likely a normalized truncated normal distribution, when this batch is sorted by height into six lanes, each lane becomes a segment of the normalized truncated normal distribution, instead of the potatoes dimensions being normally distributed within a lane.
4. Perform the super-ellipsoidal approximation measurement as described in section 1.3.2 on a much larger sample than 50 potatoes. Determine whether the increased sample size had a positive influence on the shape approximation. Using the statistical test, based on the results from the model, comparing the exact potato dimension input and results using the probability density function as input parameters, as is performed in section 4.2.1.2.
5. After recommendations 1, 2, 3, and 4 have been managed by the company it continues on to improve the model suggested as the research solution. The adaptations to the model based on recommendations 1, 2, 3, and 4 may have effect on the performance of the model. However, most like not of such significance that the model will work as it is supposed to. The shape approximation of the potato is not sufficiently accurate, as was found in section 4.2.1.1. It is recommended to perform more research in that topic. Some research trajectories that can be evaluated and have shown success in the field of biological shape approximations are:
    1. Statistical shape modelling using morphometric measurements (Hirtle et al., 2022)
    2. 3D scanning (Danielak et al., 2023)
    3. 2D imaging to 3D prediction using artificial intelligence (Joyce & Brown, 2022)

    These are three possibilities to consider in future research. The Welliver cameras currently in use do have 3D extraction capabilities. Whether these capabilities are sufficient to use in the application as input for the model needs to be investigated. When a new shape approximation is found that shows sufficient accuracy, and with that validating the model. The model can then be implemented according to section 5.1 in the current system.

6. After implementation, as mentioned in section 5.2, it is important to actively interact with the people working with the system. This will increase system performance and will lead to greater success in the long run. When system performance is found to be valuable by the operators and management of the Lelystad facility, it can be implemented in production lines of the company.
7. Develop a method that is able to determine the right settings for the newly introduced robotic halver. It is suggested that these settings are determined based on the Welliver data, which is able to measure of potatoes above a certain length measurement in lane 5 and 6 as a factor of the total potatoes above a certain length in all lanes. The length distribution found at the Véryx should be used to evaluate whether the length threshold should be changed or whether it is correct, also taking into account that only a percentage (determined by the Welliver cameras) is able to be cut. Since the robotic halvers can only cut potatoes on lines 5 and 6. Therefore

assuming that all potatoes above a certain length will be cut is invalid, since some pass through other lanes.

8. Introducing additional length measurements across the production can help with identifying where breakage of fries occurs. Since the company notices that breakage of fries occurs during production, and in this research it was found that the number of fries that suffer from any damage was relatively small at the cutting area, compared to the number of breakages found during quality inspection. Another benefit additional measurements can bring is enabling to determine what baseline to use for the parameter values at different stages of the process.
9. Consider the introduction of other tubes that allow for a larger variety of potato widths passing through. This will ensure better alignment of the potatoes, therefore extracting more length from the fries. The downside of this is that benefits of knife choice will decrease, however the potential increase in capacity and length should make up for this loss.

6.3 Research Limitations and Discussion

The data found from the manual measurements to determine the “m-value” of the potato variety were split up into a train and test set. This was done to ensure that the “m-value” determined in the train set did not influence the results of the test set, therefore limiting bias in the comparison with the perfect ellipsoid approach and improving internal validity.

The most promising method for estimating the distribution of fry lengths involves accurately modelling the shape of individual potatoes and performing virtual cuts, allowing direct comparison with data from the Véryx system. However, this approach was not feasible due to the unavailability of detailed 3D potato shape data. As an alternative, a simplified approximation of potato shape was developed and used in combination with a Monte Carlo simulation to estimate the fry length distribution. Although this novel approach reduced the complexity of the required input data, its validity could not be statistically proven, limiting its value for implementation and reducing construct validity.

To ensure accurate data during model validation, manual measurements were performed on a sample of 50 potatoes. This approach minimized potential inaccuracies compared to reliance on the Welliver cameras, which introduce inaccuracy due to the bucketed measurement format (±3–5 mm instead of ±1 mm).

The observed potatoes that enter the production facility were found to be generally wider on the ends of the potato, therefore it was investigated whether a super-ellipsoidal approach found better result in approximating a potato. The super-ellipsoid shape approximation used in this research showed improved results for the King Russet variety, for which the model was validated. However, it did not perform well for the Fontane variety, and no convincing proof was found that it improves shape approximation to a sufficient level. Therefore, external validity is limited, and other shape approximation methods should be investigated before the system can be implemented.

Because of issues with the Welliver camera due to incorrect setup and software related issues, leading to data that was inconsistent with reality, it was not possible to find pattern in the data or validate the probability density distribution the potato dimensions shown. This led to assuming that each cutting lane consisted of normally distributed potato dimensions, since this was also assumed by the company. However, it is fairly unlikely to be the case. The result found in section 4.2.1.2.1 showed that this assumption was not in compliance with the used sample. This has to be explored further, since it negatively impacts the results of the simulation. Since manual measurements were taken to test the validity of the model, this Welliver camera issues did not influence the findings of this research

Further research should be conducted to determine the potential for improved data acquisition from the current setup or explore alternative modelling approaches. Additionally, this study indicates that optimizing knife selection for each tube improves production output. However, validation with a more accurate representation of potato shape is essential to confirm the exact percentual improvements for a certain knife choice at a certain tube size.

Bibliography

Banagaaya, N., Budko, N., H. Doorn, van, Khimshiashvili, G., Klooster, R., Lindenbergh, P. J., Verdijck, J., Vermolen, F., & Zwaan, I. N. (2014). *The mathematics of french fries* (pp. 24–33). Technische Universiteit Delft. https://research.tue.nl/en/publications/the-mathematics-of-french-fries

Baronas, A.-M. K., & Louis, M. R. (1988). Restoring a Sense of Control during Implementation: How User Involvement Leads to System Acceptance. *MIS Quarterly*, *12*(1), 111–124. https://doi.org/10.2307/248811

Cooper, D. R., & Schindler, P. S. (2014). *Business research methods*. McGraw-Hill,. http://trove.nla.gov.au/work/16067972?q&sort=holdings+desc&_=1459090492803&versionId=206919510

Danielak, M., Przybył, K., & Koszela, K. (2023). The Need for Machines for the Nondestructive Quality Assessment of Potatoes with the Use of Artificial Intelligence Methods and Imaging Techniques. *Sensors 2023, Vol. 23, Page 1787*, *23*(4), 1787. https://doi.org/10.3390/S23041787

Hirtle, N. O., Stepanuk, J. E. F., Heywood, E. I., Christiansen, F., & Thorne, L. H. (2022). Integrating 3D models with morphometric measurements to improve volumetric estimates in marine mammals. *Methods in Ecology and Evolution*, *13*(11), 2478–2490. https://doi.org/https://doi.org/10.1111/2041-210X.13962

Hunton, J. E., & Beeler, J. D. (1997). Effects of User Participation in Systems Development: A Longitudinal Field Experiment. *MIS Quarterly*, *21*(4), 359–388. https://doi.org/10.2307/249719

Jaklič, A., Leonardis, A., & Solina, F. (2000). Superquadrics and Their Geometric Properties. In A. Jaklič, A. Leonardis, & F. Solina (Eds.), *Segmentation and Recovery of Superquadrics* (pp. 13–39). Springer Netherlands. https://doi.org/10.1007/978-94-015-9456-1_2

Joyce, B. J., & Brown, G. E. (2022). Estimating the volume of biological structures from a single 2D image: considering apparent cross-sectional area as an alternative to the ellipsoid method. *Evolutionary Ecology*, *36*(6), 1061–1076. https://doi.org/10.1007/s10682-022-10211-7

Sari, Y. A., & Gofuku, A. (2023). Measuring food volume from RGB-Depth image with point cloud conversion method using geometrical approach and robust ellipsoid fitting algorithm. *Journal of Food Engineering*, *358*, 111656. https://doi.org/https://doi.org/10.1016/j.jfoodeng.2023.111656

Sharma, A. K., Zotarelli, L., Zare, A., & Sharma, L. K. (2025). Automated potato tuber mass estimation and grading with multiangle 2D images. *Smart Agricultural Technology*, *10*, 100832. https://doi.org/https://doi.org/10.1016/j.atech.2025.100832

Somsen, D., Capelle, A., & Tramper, J. (2004). Manufacturing of par-fried French-fries: Part 1: Production yield as a function of number of tubers per kilogram. *Journal of Food Engineering*, *61*(2), 191–198. https://doi.org/10.1016/S0260-8774(03)00092-X

Su, Q., Kondo, N., Li, M., Sun, H., & Al Riza, D. F. (2017). Potato feature prediction based on machine vision and 3D model rebuilding. *Computers and Electronics in Agriculture*, *137*, 41–51. https://doi.org/https://doi.org/10.1016/j.compag.2017.03.020

Takahashi, M., Kawasaki, Y., Naito, H., Lee, U., & Yoshi, K. (2025). Fruit size prediction of tomato cultivars using machine learning algorithms. *Frontiers in Plant Science*, *16*, 1516255. https://doi.org/10.3389/FPLS.2025.1516255/BIBTEX

Van Dalen, J. (2025). *CREATING A STRUCTURED DATA SCIENCE METHODLOGY Answering a data science research question*.

Yu, Y., Jiang, H., Zhang, X., & Chen, Y. (2022). Identifying Irregular Potatoes Using Hausdorff Distance and Intersection over Union. *Sensors (Basel, Switzerland)*, *22*(15). https://doi.org/10.3390/s22155740

Appendices

Appendix A

In appendix A, the manual dimension and volume measurements of 50 potatoes are presented. The 50 potato samples are split up into two data sets, a training and a test set. The training set consists of 30 samples, this set is used to approximate the exponent by finding the exponent that minimizes the difference between the actual and calculated volume (calculated volume is based on calculation (13) from section 3.1.2). The exponent is then used to determine the volume approximation on the test set to evaluate the volume percentage difference and compare that to the perfect ellipsoid assumption, to determine the approximation's performance.

Training data King Russet

Potato ID King Russet	Length	Width (middle)	Height (middle)	Volume	Calculated volume	Difference	Exponent
1	119	66	59	203.6	258.506	54.906	2.137
2	102	58	55	180.6	181.518	0.918	2.137
3	123	61	54	241.4	226.025	15.375	2.137
4	136	66	60	288	300.443	12.443	2.137
5	152	78	57	377	377.000	0.000	2.137
6	104	57	46	157.2	152.123	5.077	2.137
7	107	52	50	165.2	155.198	10.002	2.137
8	115	59	50	191.2	189.255	1.945	2.137
9	98	58	57	123.2	180.741	57.541	2.137
10	121	66	57	251	253.941	2.941	2.137
11	104	55	50	153.6	159.549	5.949	2.137
12	130	61	50	235.8	221.193	14.607	2.137
13	81	45	43	88.2	87.437	0.763	2.137
14	106	58	49	150.4	168.058	17.658	2.137
15	127	65	53	230.4	244.074	13.674	2.137
16	87	52	48	126.2	121.141	5.059	2.137
17	128	62	47	204.2	208.079	3.879	2.137
18	101	56	51	165.2	160.919	4.281	2.137
19	113	58	49	198.4	179.156	19.244	2.137
20	90	55	47	138	129.787	8.213	2.137
21	102	60	52	173.8	177.535	3.735	2.137
22	136	63	56	277	267.668	9.332	2.137
23	72	47	38	71.8	71.737	0.063	2.137
24	79	50	39	89.8	85.939	3.861	2.137
25	110	56	48	160.4	164.949	4.549	2.137
26	107	50	47	166.4	140.275	26.125	2.137
27	108	49	43	132	126.945	5.055	2.137
28	97	55	46	139.2	136.905	2.295	2.137
29	116	62	47	198.8	188.571	10.229	2.137
30	111	67	57	219.4	236.484	17.084	2.137
						336.801	2.137

Test data King Russet

Potato ID King Russet	Length	Width (middle)	Height (middle)	Volume	Calculated volume	Difference	Exponent
31	102	59	51	173.8	171.219	2.581	2.137
32	128	62	53	245.6	234.642	10.958	2.137
33	84	48	44	110.2	98.970	11.230	2.137
34	86	48	43	99.2	99.023	0.177	2.137
35	71	53	49	91.6	102.863	11.263	2.137
36	110	49	44	113.2	132.303	19.103	2.137
37	120	57	50	189.8	190.789	0.989	2.137
38	133	62	49	241.2	225.407	15.793	2.137
39	86	55	51	141.8	134.574	7.226	2.137
40	101	62	55	208.4	192.134	16.266	2.137
41	121	64	56	240.4	241.926	1.526	2.137
42	73	59	46	118.4	110.525	7.875	2.137
43	132	56	44	183.4	181.444	1.956	2.137
44	117	60	44	193.6	172.313	21.287	2.137
45	119	55	49	201	178.910	22.090	2.137
46	96	51	47	134.4	128.371	6.029	2.137
47	118	59	48	204.2	186.425	17.775	2.137
48	103	50	44	150.2	126.412	23.788	2.137
49	121	53	49	188.6	175.302	13.298	2.137
50	99	58	41	145.6	131.333	14.267	2.137
						225.478	**2.137**

Appendix B

Manual measurements performed on the Fontane variety. Method is similar as was described in the introduction of appendix A.

Training data Fontane

Potato ID Fontane	Length	Width (middle)	Height (middle)	Volume	Calculated volume	Difference	Exponent
1	62	61	48	85.8	88.021	2.2206	1.859
2	82	74	61	190	179.472	10.5279	1.859
3	80	52	55	121.4	110.937	10.4627	1.859
4	73	60	49	100.2	104.062	3.8619	1.859
5	82	61	53	127.8	128.541	0.7409	1.859
6	87	67	47	143.6	132.835	10.7647	1.859
7	79	56	45	100.8	96.527	4.2729	1.859
8	78	64	48	103.4	116.182	12.7816	1.859
9	94	79	55	187.7	198.034	10.3338	1.859
10	74	58	46	94.2	95.728	1.5280	1.859
11	92	63	47	131.4	132.083	0.6833	1.859
12	101	69	52	188.2	175.710	12.4904	1.859
13	75	54	45	84.8	88.367	3.5668	1.859
14	80	65	49	120	123.544	3.5438	1.859
15	61	52	48	74.6	73.824	0.7763	1.859
16	66	54	45	71.2	77.763	6.5628	1.859
17	74	63	47	94.8	106.241	11.4409	1.859
18	69	50	40	54.3	66.911	12.6115	1.859
19	71	62	50	106.4	106.719	0.3190	1.859
20	86	53	38	89.8	83.981	5.8193	1.859
21	69	50	46	70.6	76.948	6.3482	1.859
22	73	59	46	98.8	96.063	2.7374	1.859
23	70	57	52	100.6	100.600	0.0000	1.859
24	72	66	48	111.4	110.596	0.8040	1.859
25	85	66	53	132.6	144.165	11.5652	1.859
26	72	55	51	110.8	97.924	12.8765	1.859
27	77	62	55	135.6	127.311	8.2888	1.859
28	88	73	59	232.8	183.772	49.0281	1.859
29	71	53	46	91	83.929	7.0707	1.859
30	76	63	54	132.2	125.363	6.8370	1.859
						230.8648	1.859

Test data Fontane

Potato ID Fontane	Length	Width (middle)	Height (middle)	Volume	Calculated volume	Difference	Exponent
31	82	64	51	153.8	129.773	24.0266	1.859
32	69	60	51	113	102.375	10.6254	1.859
33	71	53	45	88.4	82.105	6.2952	1.859
34	81	50	44	99	86.403	12.5969	1.859
35	72	50	45	84.6	78.548	6.0517	1.859
36	84	58	50	125.6	118.113	7.4867	1.859
37	75	58	46	105	97.022	7.9783	1.859
38	75	60	49	116	106.913	9.0871	1.859
39	77	57	48	107.4	102.148	5.2523	1.859
40	72	60	44	104.6	92.163	12.4367	1.859
41	78	65	55	140	135.205	4.7951	1.859
42	70	56	50	81.8	95.034	13.2337	1.859
43	83	64	53	143.2	136.507	6.6928	1.859
44	78	59	50	108	111.568	3.5676	1.859
45	65	51	43	68.4	69.115	0.7152	1.859
46	84	62	53	138	133.835	4.1654	1.859
47	81	63	60	149	148.456	0.5438	1.859
48	99	72	56	218	193.543	24.4571	1.859
49	77	59	47	98.8	103.529	4.7290	1.859
50	74	53	43	84.2	81.771	2.4293	1.859
						167.1661	**1.859**

Appendix C

Manual measurements performed on the King Russet variety to determine the exponent that best approximates the actual volume of the potato, and the means and standard deviation of the dimensions of the potato. The dimension measurements and the exponent served as input parameters for the validation process of the simulation.

Potato_ID	Length	Width	Height	Volume	Calculated volume	Difference	Exponent
1	96	57	48	121.8	141.24	19.436	2.055
2	99	60	43	153.2	137.35	15.855	2.055
3	107	62	50	190.4	178.36	12.037	2.055
4	105	67	52	165.8	196.71	30.910	2.055
5	73	60	43	103.8	101.27	2.525	2.055
6	83	63	53	141.6	149.02	7.423	2.055
7	113	64	52	178.6	202.22	23.618	2.055
8	91	57	56	148.6	156.19	7.594	2.055
9	100	62	50	157	166.69	9.694	2.055
10	93	66	56	189.2	184.83	4.369	2.055
11	79	66	50	139.8	140.18	0.384	2.055
12	86	62	58	136.2	166.29	30.094	2.055
13	103	64	52	174.6	184.32	9.723	2.055
14	104	62	54	198.2	187.23	10.969	2.055
15	74	59	54	127	126.78	0.224	2.055
16	109	63	50	203.2	184.63	18.573	2.055
17	75	65	57	141.2	149.42	8.220	2.055
18	129	62	54	180.2	232.24	52.038	2.055
19	100	66	53	164.2	188.10	23.896	2.055
20	88	58	45	122.2	123.50	1.304	2.055
21	116	70	52	231	227.05	3.952	2.055
22	115	69	56	235.4	238.94	3.543	2.055
23	102	68	57	192	212.59	20.590	2.055
24	105	67	49	193.4	185.36	8.039	2.055
25	95	63	52	168	167.35	0.650	2.055
26	91	62	52	134	157.76	23.759	2.055
27	91	67	66	152.2	216.38	64.181	2.055
28	105	61	53	211.2	182.54	28.662	2.055
29	88	62	53	166.6	155.49	11.108	2.055
30	74	57	48	107.2	108.87	1.670	2.055
31	92	63	56	178.8	174.53	4.268	2.055
32	133	64	49	254.8	224.28	30.522	2.055
33	91	61	46	120.6	137.31	16.705	2.055
34	103	65	51	189.8	183.60	6.197	2.055
35	88	67	56	188	177.54	10.457	2.055
36	85	58	52	143	137.85	5.149	2.055
37	90	63	57	183.2	173.79	9.413	2.055
38	105	68	61	234.2	234.20	0.000	2.055
39	78	66	50	108	138.41	30.410	2.055

40	93	60	37	123	111.02	11.982	2.055
41	82	70	55	164	169.76	5.759	2.055
42	64	55	55	105.4	104.10	1.297	2.055
43	96	62	47	145.2	150.42	5.225	2.055
44	77	61	43	110	108.60	1.395	2.055
45	84	56	54	145.8	136.59	9.210	2.055
46	99	66	50	187	175.67	11.326	2.055
47	73	58	44	108.6	100.18	8.424	2.055
48	93	60	47	140	141.02	1.023	2.055
49	81	58	52	96	131.36	35.364	2.055
50	105	58	47	157.8	153.91	3.887	2.055

Mean	94.02	62.6	51.54
SD	14.32	3.82	5.08

Appendix D

The Chi-squared test was performed on the average fries' length distribution of 50 simulations with the use of the normal distribution assumption for the dimensions and the actual fries' length distribution. A Chi-squared value of **401.267** was found, since this value was larger than the critical value of **28.869**, the distributions were found to be different.

Counts per bucket	Average of 50 runs		
Bucket	**Simulation R=2.055 (E)**	**Count Real (O)**	**((O-E)^2)/E**
35-39 mm	6.8	20	25.624
40-44 mm	14.3	35	29.964
45-49 mm	29.4	31	0.087
50-54 mm	49.6	46	0.261
55-59 mm	74.7	90	3.134
60-64 mm	102.6	93	0.898
65-69 mm	133	104	6.323
70-74 mm	154.7	112	11.786
75-79 mm	161.1	135	4.228
80-84 mm	155	125	5.806
85-89 mm	142.4	146	0.091
90-94 mm	122	137	1.844
95-99 mm	91.1	95	0.167
100-104 mm	68.5	73	0.296
105-109 mm	46.3	36	2.291
110-114 mm	26.5	29	0.236
115-119 mm	15.4	16	0.023
120-124 mm	7.1	11	2.142
> 125 mm	5.6	47	306.064
Total counts	**1406.1**	**1381**	

Calculated Chi squared value Sum	401.267
Critical value (0.05,18) (prob, df)	28.869

Appendix E

The Chi-squared test was performed on the average fries' length distribution of 50 simulations with the use of the actual found dimensions from appendix C and the actual fries' length distribution. A Chi-squared value of **182.738** was found, since this value was larger than the critical value of **28.869**, the distributions were found to be different.

Counts per bucket			
Bucket	**Simulation R=2.055 (E)**	**Count Real (O)**	**((O-E)^2)/E**
35-39 mm	8	20	18.000
40-44 mm	12	35	44.083
45-49 mm	20	31	6.050
50-54 mm	57	46	2.123
55-59 mm	72	90	4.500
60-64 mm	99	93	0.364
65-69 mm	138	104	8.377
70-74 mm	149	112	9.188
75-79 mm	161	135	4.199
80-84 mm	152	125	4.796
85-89 mm	149	146	0.060
90-94 mm	107	137	8.411
95-99 mm	92	95	0.098
100-104 mm	90	73	3.211
105-109 mm	43	36	1.140
110-114 mm	26	29	0.346
115-119 mm	9	16	5.444
120-124 mm	7	11	2.286
> 125 mm	16	47	60.063
Total counts	**1407**	**1381**	

Calculated Chi squared value Sum	182.738
Critical value (0.05,18) (prob,df)	28.869

Appendix F

One example of the Kolmogorov-Smirnov test that was performed on 50 simulation output in total is shown. For both the simulation output as well as the actual output a cumulative distribution function (CDF) is made. The differences of the CDFs are taken. The largest difference observed is the test value of these distributions. When the test value exceeds the critical value, the distributions are not similar in shape.

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6
Buckets	Simulation run	ABS(bucket_count/Sum count) + previous_fraction	Actual Count	ABS(bucket_count/Sum count) + previous_fraction	abs(Column 3 - Column 5)
< 39	9	0.006460876	20	0.014482259	0.008021383
40-44 mm	13	0.015793252	35	0.039826213	0.024032961
45-49 mm	23	0.032304379	31	0.062273715	0.029969336
50-54 mm	50	0.068198134	46	0.095582911	0.027384777
55-59 mm	103	0.142139268	90	0.160753077	0.01861381
60-64 mm	117	0.226130653	93	0.228095583	0.00196493
65-69 mm	124	0.315147164	104	0.303403331	0.011743833
70-74 mm	150	0.422828428	112	0.384503983	0.038324445
75-79 mm	164	0.540559943	135	0.482259232	0.05830071
80-84 mm	165	0.659009332	125	0.572773353	0.08623598
85-89 mm	122	0.746590093	146	0.678493845	0.068096248
90-94 mm	105	0.821966978	137	0.777697321	0.044269657
95-99 mm	74	0.875089734	95	0.846488052	0.028601682
100-104 mm	64	0.92103374	73	0.899348298	0.021685442
105-109 mm	51	0.95764537	36	0.925416365	0.032229005
110-114 mm	28	0.977745872	29	0.946415641	0.031330231
115-119 mm	20	0.992103374	16	0.958001448	0.034101926
120-124 mm	9	0.99856425	11	0.965966691	0.032597559
> 125 mm	2	1	47	1	1.11022E-16
Sum	**1393**		**1381**		
MAX KS					**0.08623598**

Critical value	0.032687693

McCain UNIVERSITY OF TWENTE.

Appendix G

The Chi-squared test was performed on the average fries' length distribution of 50 simulations with the use of the actual found dimensions from appendix C and the average fries' length distribution of 50 simulations with the use of the normal distribution assumption for the dimensions. A Chi-squared value of **36.288** was found, since this value was larger than the critical value of **28.869**, the distributions were found to be different. Concluding that the assumption that the input parameters can be described as normally distributed showed not to be the case for this sample.

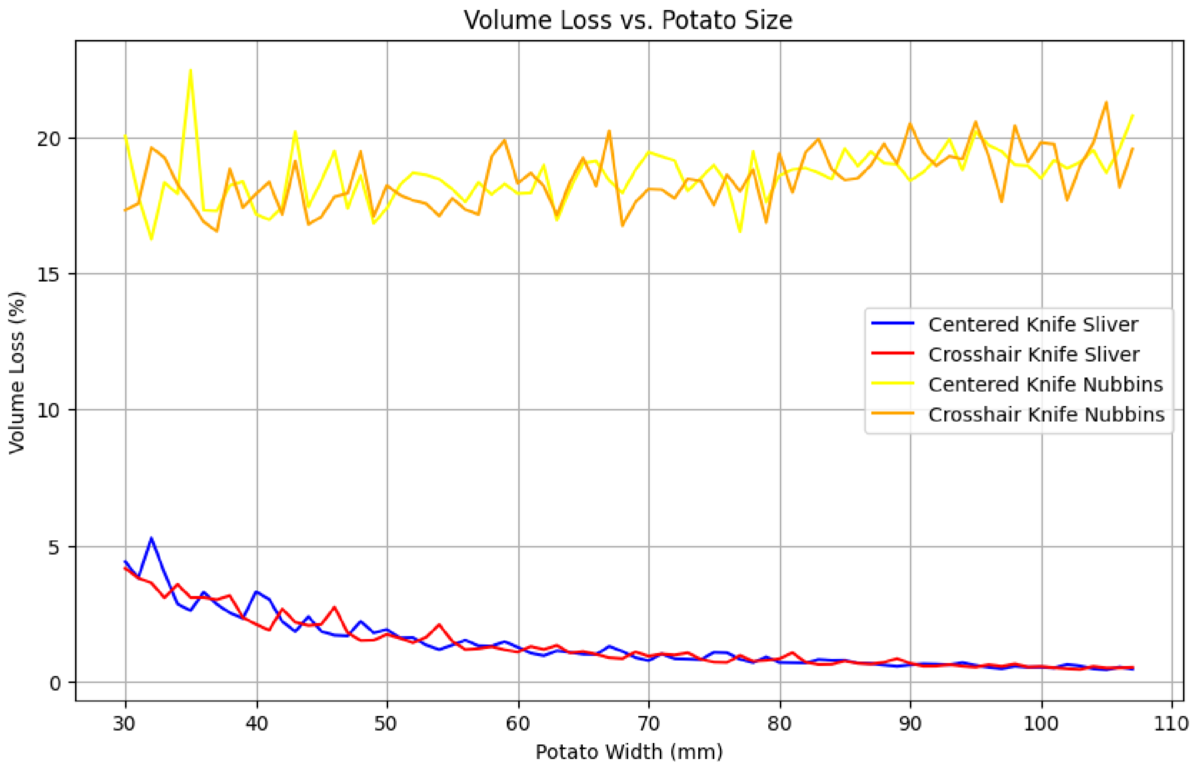
Counts per bucket	**Average of 50 simulation runs**		
Buckets	**Mean and SD (E)**	**Actual potato dimensions (O)**	**((O-E)^2)/E**
<39 mm	6.9	7	0.001449275
40-44 mm	14.3	12	0.36993007
45-49 mm	29.4	20	3.005442177
50-54 mm	49.6	57	1.104032258
55-59 mm	74.7	72	0.097590361
60-64 mm	102.6	99	0.126315789
65-69 mm	133	138	0.187969925
70-74 mm	154.7	149	0.210019392
75-79 mm	161.1	161	6.20732E-05
80-84 mm	155	152	0.058064516
85-89 mm	142.4	149	0.305898876
90-94 mm	122	107	1.844262295
95-99 mm	91.1	92	0.008891328
100-104 mm	68.5	90	6.748175182
105-109 mm	46.3	43	0.235205184
110-114 mm	26.5	26	0.009433962
115-119 mm	15.4	9	2.65974026
120-124 mm	7.1	7	0.001408451
125> mm	5.6	16	19.31428571

Calculated Chi squared value Sum	**36.28817709**
Critical value (0.05,18) (prob,df)	28.86929943

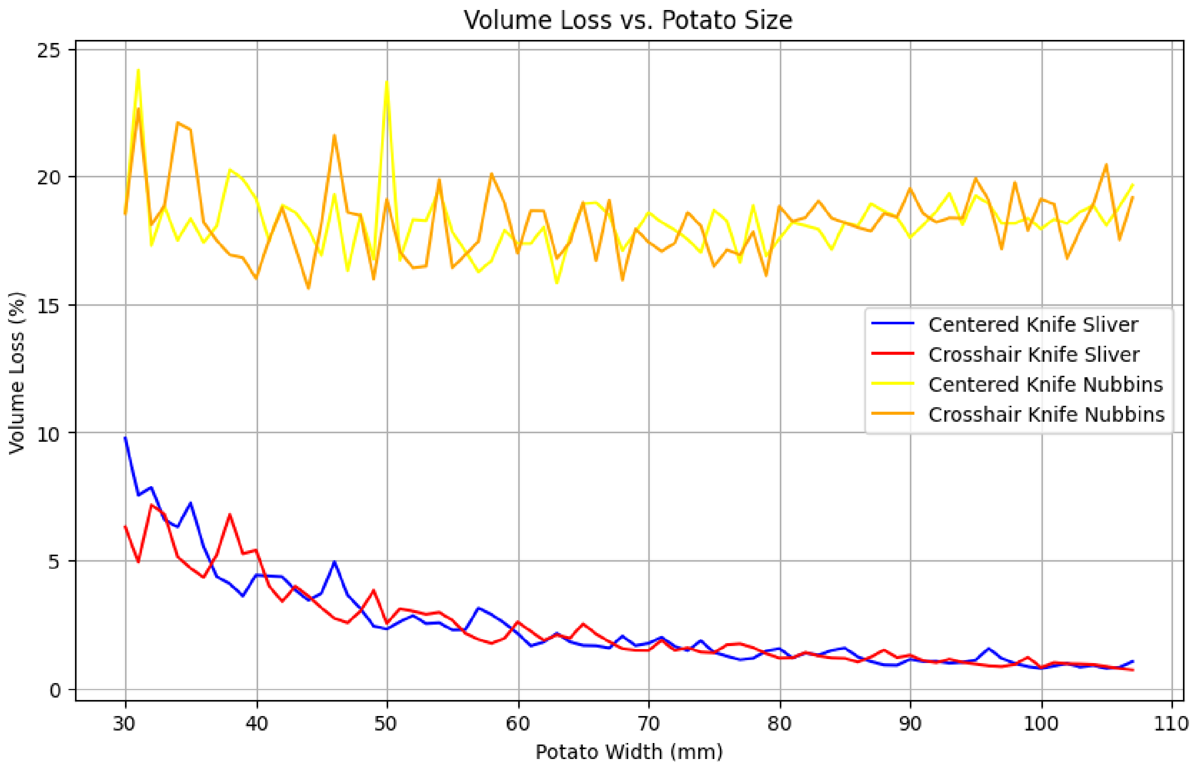
Appendix H

The percentage of volume losses due to slivers and nubbins is shown for three different knife sizes. In general, volume losses increase with larger knife sizes. Notably, nubbin losses tend to rise slightly with increasing potato width for smaller cutting sizes. Sliver losses show a more notable increase for larger knives, which can partly be attributed to the definition of a sliver—any fry with a width of 30% or less of the knife size—resulting in more fragments being classified as slivers when larger knives are used.

7x7mm



10x10mm



14x14mm