Evaluation and Validation of a New Risk Assessment Method

Faculty of Engineering LTH - Division of Production Management
MIOM05 Degree Project in Production Management

Lukas Torbiörnsson & Eddie Berlin

June 10, 2024

Abstract

This Master's thesis evaluates and validates a newly developed risk assessment tool designed for performance guarantees in the sales of rock crushing solutions at Sandvik Rock Processing Solutions, Stationary Crushing & Screening (Stat. C&S). The research aims to compare the new tool against the current expert-based method, assess the relevancy of the variables included, and propose necessary adjustments to improve the accuracy and usability of the tool.

The study was conducted in three phases: a quantitative analysis comparing the new tool's risk scores with those of the current method, qualitative interviews with experts to evaluate the scope and relevance of the variables used, and a final discussion to propose improvements and an implementation plan. Statistical tests revealed significant differences between the risk scores of the two methods, indicating the need for adjustments in the new tool.

Key findings from the investigation highlighted the necessity of adding new variables into the risk assessment tool such as feed material characteristics and machine-specific factors. The current method of calculating individual risk scores by treating variables as independent was generally supported, but the approach for calculating the final risk score requires some refinement to better reflect the system's complexity.

The thesis concludes with recommendations for adjusting the variable scope, refining calculation models, and developing an implementation plan to transition the risk assessment process to the sales areas. These adjustments are important for ensuring the tool's effectiveness and accuracy, improving Stat. C&S risk assessment capabilities, and promoting a more efficient and customer-centered sales process.

Acknowledgements

There are a lot of people who have contributed to the successful completion of this Master's thesis and we would like to extend a thanks to everyone. Firstly, our supervisor at LTH, Danja Sonntag, for her many helpful contributions and continued guidance throughout our Master's thesis project.

Secondly, we would like to thank everyone at Sandvik Stat. C&S and PIE, for welcoming us into the team and answering all of our questions.

We'd like to extend a special thanks to Magnus Franzén and Elizabeth Lee-Norman, our supervisors at PIE, for helping us throughout our thesis with everything from connecting us with experts within rock processing solutions to giving us tips on where to get the best lunch in Svedala.

We are also grateful for all the people at Stat. C&S that have participated in interviews and gave valuable feedback on the risk assessment tool and risks associated with rock processing solutions.

Lastly, we'd like to thank all our colleagues and friends at Lund University for their great company and helpful comments while writing this Master's thesis.

Lund, June 2024

Lukas Torbiörnsson & Eddie Berlin

Contents

Abstract **I**

Acknowledgements **III**

Table of contents **VII**

1 Introduction **1**
1.1 Background 2
1.1.1 Sandvik Group 2
1.1.2 Sandvik Rock Processing Solutions 3
1.1.3 Products 5
1.1.4 Performance Guarantee 5
1.1.5 Risk assessment tool 7
1.2 Problem formulation 10
1.3 Research purpose and questions 11

2 Methodology **13**
2.1 Research Approach 13
2.2 Research Procedure 14
2.2.1 Phase 1: Quantitative Analysis 14
2.2.2 Phase 2: Interviews and Evaluation 16
2.2.3 Phase 3: Discussion and Implementation 18

3 Rock Crushing and Screening **19**
3.1 Rock Crushers 19
3.1.1 CH and CS Crushers 22
3.1.2 CJ Crushers 24
3.1.3 CV Crushers 25
3.1.4 Common Crusher Aspects 26
3.2 Screens 27
3.3 Performance Criteria 32
3.3.1 Capacity 32

3.3.2 Particle Size Distribution 33
3.3.3 Product Shape 35
3.3.4 Oversize and undersize of particles 37

4 Theory **39**
4.1 Multiple linear least-squares regression 39
4.1.1 Terminology 40
4.1.2 Assumptions 40
4.1.3 Parameter estimation using OLS 41
4.2 Multinomial logistic regression 41
4.2.1 Binary multinomial logistic regression 42
4.2.2 Ordered logistic regression 43
4.3 Dummy variable regression 45
4.3.1 Terminology 45
4.4 Statistical inference 46
4.4.1 Wilcoxon signed rank test 47
4.4.2 McNemar's test 48
4.4.3 Coefficient of determination 51
4.4.4 Variance inflation factor 52
4.4.5 Wald's test 53
4.5 Stepwise regression 54

5 Quantitative results **57**
5.1 Data collection 57
5.2 Initial high level analysis 59
5.3 Regression analysis 65
5.3.1 Initial multiple linear regression model 66
5.3.2 Regression of deviation between current and new tool 70

6 Interview Findings **73**
6.1 Scope of variables 74
6.1.1 Crushers 76
6.1.2 Screens 80
6.1.3 Common variables for crushers and screens 83
6.2 Risk Score Calculation Model 85
6.2.1 Individual risk scores 85
6.2.2 Calculating final risk score 88

7 Discussion **91**
7.1 Key recommendations 91

7.1.1 Adjustment of variable scope 91
7.1.2 Adjust the impact on risk for certain variables . . 94
7.1.3 Adjusting probabilities based on test data 96
7.1.4 Alternative way of handling moisture 97
7.1.5 Adjustment of total risk score calculation 99
7.2 Implementation of the Tool 101
7.2.1 EPIS Framework 102
7.2.2 Proposed Implementation Plan 103
7.3 Reliability and Validity 105

8 Summary and Conclusion 109

References 113

A Appendices 117
A.1 Interview Guide 117
A.2 OLS for multiple linear regression 120
A.3 MLE for logistic regression 120
A.4 The stepwise regression procedure 122

1 Introduction

Making informed decisions about the uncertain future is something all people, organizations, and societies have difficulty doing. Macro-level uncertainties include economic and geopolitical uncertainties and the development of public health crises such as the recent COVID-19. Uncertainty is also always apparent on a smaller scale, such as in our daily lives or the day-to-day operations of a company. While uncertainty is closely connected to risk, there is a distinct difference between the two definitions. The terminology uncertainty is used when potential outcomes and the connected probabilities to each outcome are unknown or cannot be reliably estimated (Park and Shapira, 2017). In contrast, the term risk is used in decision-making situations where all potential outcomes and their corresponding probabilities of occurrence are known or can be estimated. A concrete definition of risk is given by Rausand and Haugen (2020), who defines risk as:

"The combined answer to the three questions: (1) What can go wrong? (2) What is the likelihood of that happening? and (3) What are the consequences?"

When assessing a single unfortunate event, where the consequences are known, the remaining task is to answer question number two. The purpose of a risk assessment tool for a known event with known consequences is to give an accurate estimate of the likelihood of the occurrence of that event. In this project, likelihood and probability will be used interchangeably. Referring to the distinction between uncertainty and risk, a successful risk assessment converts uncertainty to a risk, where the likelihood becomes known.

1.1 Background

The purpose of this master thesis project is to evaluate a newly developed risk assessment tool for performance guarantees related to the sales of rock crushing solutions, as well as discuss recommended changes and adjustments to the new tool together with a proposal of how to implement the risk assessment tool within the organization.

This Master's thesis is performed at Sandvik AB for Sandvik Rock Processing Solutions, later referred to as SRP, in Svedala, Skåne. Sandvik Rock Processing Solutions is one of three Business Areas at Sandvik, the other two being Mining and Rock Solutions and Manufacturing and Machining Solutions. The Master's thesis is written at the Stationary Crushing and Screening division of SRP, specifically with the division's Performance and Innovation Excellence department. The former will hereby be referred to as Stat. C&S, and the latter as PIE.

1.1.1 Sandvik Group

Sandvik is a global industrial manufacturing group providing products and solutions for mining and rock excavation, rock processing, and metal cutting. Founded in 1862 in Sandviken, Sweden, by Göran Fredrik Göransson, Sandvik was the first company to commercialize the famous

Bessemer process, the first relatively inexpensive method to mass-produce steel (Sandvik Group, 2020). Today, the company manufactures a wide variety of products and has grown to have around 41,000 employees and sales in 170 countries. In 2023, the company had a total revenue of 126.5 billion SEK with an EBITA margin (earnings before interest, tax, amortization relative to revenue) of 19.4%. SRP is the smallest of the three divisions, having had a 9% share of the revenue in 2023. The two other divisions, Mining and Rock Solutions and Manufacturing and Machining Solutions had 52% and 39% share of the total revenue, respectively (Sandvik Group, 2023).

1.1.2 Sandvik Rock Processing Solutions

SRP is the business area of Sandvik Group that manufactures and sells machines and complete solutions for mineral and rock processing. Founded in 1882 in Svedala, Sweden, the company originally manufactured agricultural machines, but over time started to prioritize machines for rock and mineral processing. In 2001, the Finnish company Metso Corporation aspired to purchase the entire business, which at the time was called Svedala Industri. However, competition regulations enabled Sandvik Group to acquire its production units in Sweden and France. The acquired units, which totaled 900 employees and a yearly revenue of 1.4 billion SEK, were viewed as a natural complement to Sandvik's existing business within construction and infrastructure (Sandvik Group, 2001). Today, SRP consists of 2 946 employees and had a revenue of around 11.5 billion SEK in 2023 (Sandvik Group, 2023). As shown in Figure 1.1, SRP has three divisions, Attachment Tools, Stationary Crushing & Screening, Mobile Crushing & Screening. Stationary Crushing & Screening sells rock crushing systems designed to stay in one place during their entire life cycle. The new risk assessment tool mentioned previously involves only Stationary Rock Crushing solutions, which is why this thesis will solely focus on performance related risk involving Stationary Crushing & Screening's product line.

Customers for these products are mainly found in two industries, infrastructure and mining. The infrastructure industry involves all activities regarding the construction of roads, railways, bridges, and tunnels. In 2022, infrastructure accounted for 61% of SRP's revenues and mining accounted for the remaining 39% (Sandvik Group, 2023). While sales of individual products occur, Stat. C&S's customers usually demand a complete rock-crushing solution where crushers, screens, and feeders are combined.

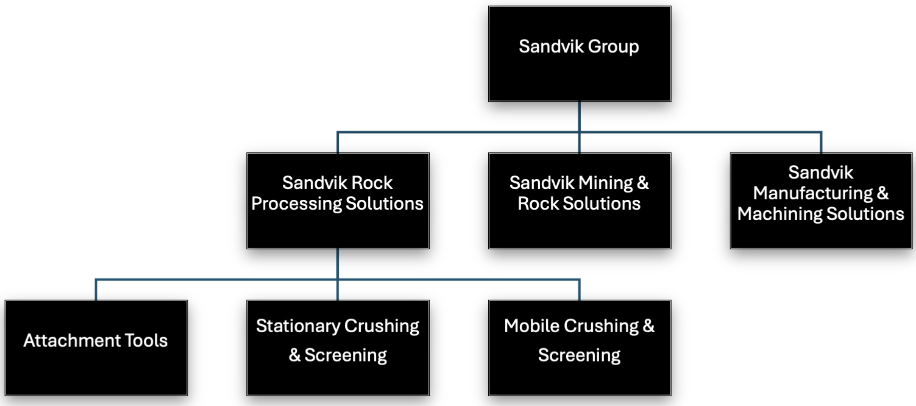

Figure 1.1: An organizational chart over Sandvik Group and Sandvik SRP.

1.1.3 Products

The performance guarantees and risk assessments only evaluate the risk of certain products in the entire product portfolio that Stat. C&S has. These are the rock crushers and screens that are sold relatively frequently and contribute risk to the performance of a system of machines. Crushers is the collective name for machines that disintegrate rocks and minerals into smaller parts and screens separate the rocks into different sizes. Another machine sold with these two is feeders, which ensure that material into the crusher and screens has a regular flow. The feeders are not included in the new risk assessment tool. A more in-depth description of the crushers and screens is available in Chapter 3.

1.1.4 Performance Guarantee

When selling a single piece of equipment or a crushing and screening process, the sales units of Stat. C&S uses an internal digital design tool called PlantDesigner to produce a flowsheet of the system. The flowsheet includes the crushers, screens, and feeders needed, together with information such as output capacity and product specifications, see Figure 1.2 for an example. After inputting the customer's specifications into the design tool and reaching a preliminary agreement, the customer, in some cases, requests a performance guarantee for the solution.

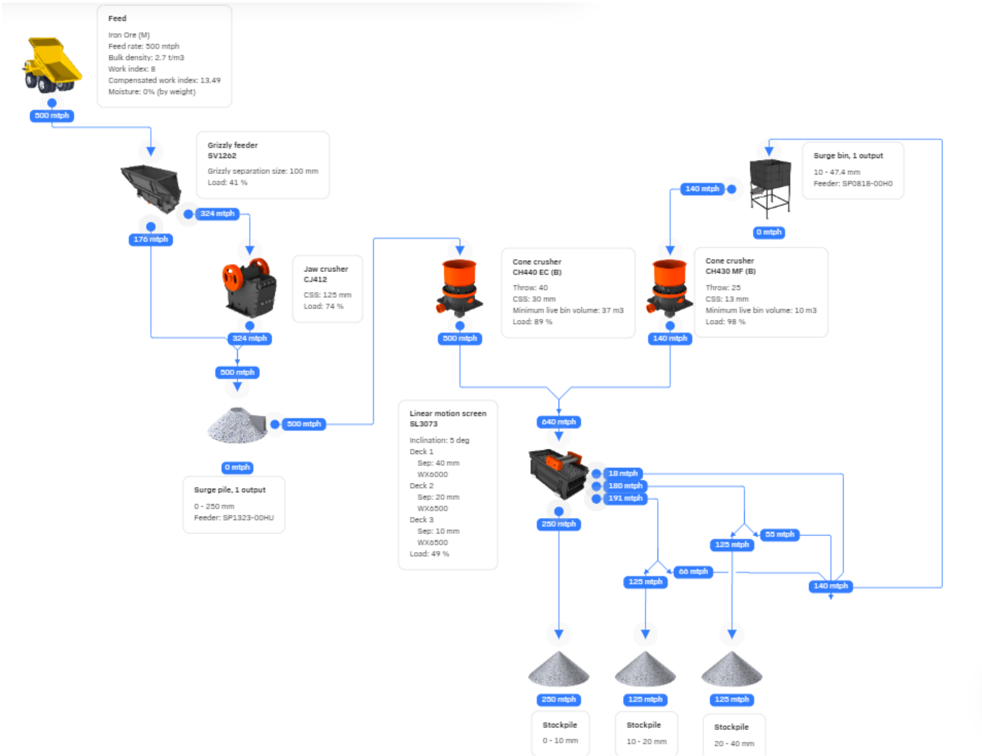

Figure 1.2: An example of a flowsheet from PlantDesigner with flow of material, crushers, and screens included.

The performance guarantee involves minimum performance criteria for the product or system. In this thesis, the term product can refer to one of two things: the machines that Sandvik Stat. C&S sells or the output from a crusher or screen, i.e. crushed or screened rock. The term 'final product' specifically refers to the output from the final crusher or screen in a crushing and screening plant. The most common performance criteria are output capacity, product size grading, particle shape, the maximum degree of over- and undersize of particles, or a combination of these criteria. The criteria will be further explained in Chapter 3. If the above-mentioned criteria are not being met under the test conditions specified in the performance guarantee. Sandvik can be held liable according to the conditions in the performance guarantee. The crushers at Stat. C&S have settings that are allowed to be changed in order to meet the specified performance criteria. For example, if the crushers are not reaching a performance criterion such as a specified capacity, the crusher's original settings can be changed to enhance the capacity of the crusher in order to reach the minimum performance criteria after it is sold. While the fine can be costly for Sandvik, the largest driver for meeting the criteria is avoiding missing out on future engagements or damaging the company's reputation due to a dissatisfied customer.

Before a system is sold, the performance guarantee is created at the respective sales areas, along with a risk assessment which is created centrally at PIE. Different systems have different levels of volatility in terms of performance and subsequently a higher risk of not meeting the performance criteria in the guarantee. The purpose of the risk assessment is to estimate the risk that the system will not perform to the specifications of the performance guarantee. The estimated risk serves as useful information when deciding whether to sell the system along with the performance guarantee or not. If the estimated risk is deemed too high, the sale either does not happen, or further adjustments to the system or performance guarantee have to be done.

1.1.5 Risk assessment tool

This section will explain the current- and new risk assessment tools at Sandvik Stat. C&S. As will be detailed in Chapter 2, the current risk assessment tool will be used as a comparison benchmark for the risk scores given by the new method.

Current risk assessment tool

The current risk assessment tool used at Stat. C&S is built on experience-based inputs for the risk-dependent variables briefly detailed below and is therefore highly dependent on the expertise of the centrally located evaluators at PIE. Currently, the person who is performing the risk assessment grades the estimated risks for a number of different variables divided into separate aspects. Some of these aspects are for instance related to the raw material characteristics, output from the equipment, accuracy, and quality. The grading of the risk-dependent variables gives an overview of the risk, which is taken into account together with an expert's opinion regarding these factors when determining the final risk score. In the current risk assessment tool, the risk score is given as an integer between 1 and 5, 5 meaning a high risk of not meeting the performance criteria, and 1 meaning little or no risk of not meeting it.

New Risk Assessment Tool

Stat. C&S aims to have the entire sales process performed decentrally at the different sales areas. That means that the entire sales process from solution design to the creation of performance guarantees and risk assessments will be performed by the respective sales areas. The reason for this is that Stat. C&S wants the expertise surrounding the entire sales process to be located closer to customers in order to promote a more efficient process and better communication with customers. Stat. C&S has been content with the accuracy of the current risk assessments, but to obtain consistency between sales areas, a new tool is required that is not dependent on the expertise and knowledge of the person who is evaluating the system.

The new tool has been developed but not yet validated and it uses only inputs from Stat. C&S's solution design tool called PlantDesigner. While currently requiring manual inputs for the new tool, the ambition is to automate the risk assessments more in the future.

In the new tool, the user first inserts the crushers and screeners included in the guarantee. The user also inserts the configuration of the machines and the values for the variables that will affect the machine's performance. These numbers are obtained from the flowsheet that is attached to the specified guarantee. The user also inserts the raw material characteristics that are stated in the flowsheet from PlantDesigner and obtained through tests done at Stat. C&S's testing facility in Svedala.

When a value for a variable is inserted in the assessment tool, an individual probability that the machine will meet the performance guarantee is calculated by the method. After all values are inserted for one single machine, these probabilities are then multiplied to provide a final risk score of a certain screen or crusher. In the new tool, this number is presented as a percentage and could be interpreted as the probability that the machine will perform as specified given the values of the variables. After the percentages for all machines are calculated, the tool then calculates the arithmetic mean of the percentages, which is the final risk score.

1.2 Problem formulation

When approaching a formulation of the problem, it is important to understand the risk that the tool aims to estimate. Referring to the definition of risk in the early introduction, what can go wrong is that a sold product or solution that has a performance guarantee does not meet at least one of the performance criteria included in the performance guarantee. An estimate of the likelihood of this happening is the output of the risk assessment method and the consequence of this outcome taking place is that Sandvik can be held liable. To understand the likelihood of not meeting the performance guarantee, an effective approach is to start by describing the unpredictable nature of real-life values for the previously mentioned performance criteria, which are capacity, product grading, shape, and over- and undersize. These will be further explained in Chapter 3.

The values for capacity and product grading are calculated in PlantDesigner based on assumptions of optimal operating conditions. As such, real-life values can sometimes deviate from this. In contrast to capacity and product grading, the values for shape and over- and undersize are not given upfront in PlantDesigner but can be calculated using other internally developed tools. The real-life value of these variables can also deviate from the values calculated in the tools.

The real-life probability density function of the values for the performance criteria is unknown and different for different crushing solution configurations. The relevant configuration settings are aimed to be included in the new risk assessment tool to enable the tool to reflect the risk accurately. To explain with a simple example, a favorably configured crusher might have a nearly deterministic real-life capacity value while a badly configured crusher might have a widely spread normal distribution around the same capacity number. An effective risk assessment tool should observe the included minimum performance criteria and the values for all the relevant configuration settings displayed in PlantDesigner, also referred to as variables from now on. It should also have an appropriate computational model to arrive at a number that accurately reflects the likelihood that the system will meet these criteria. The new risk assessment method is currently in the prototype stage and needs to be verified and evaluated.

1.3 Research purpose and questions

The purpose of the master thesis project is to evaluate a newly developed risk assessment tool for performance guarantees related to the sales of rock crushing solutions, as well as critically discuss changes and adjustments to the new tool. The master thesis project will also propose a plan for the tool's implementation. Following are the questions that the new risk assessment method will be evaluated and verified from:

1. *How does the new tool compare against the current risk assessment done by an expert?*

2. *Does the tool include the right scope of variables affecting the risk?*

3. *Does it have an appropriate model to compute the risk from the variables?*

2 Methodology

This chapter explains the research approach that has been used throughout the thesis. It will also introduce the research procedure, described in Section 2.2 in which the three phases of the master's thesis are introduced. In Section 2.2, the data collection and analysis methodologies used in the master thesis are also described.

2.1 Research Approach

While there is no one-size-fits-all method for research, Denscombe (2010), outlines various approaches tailored to specific contexts, such as case studies or experiments. In this thesis, a mixed-method approach is used, integrating quantitative and qualitative data collection methods to answer the research questions effectively.

The advantage of a mixed-method approach lies in its ability to provide a complete perspective, utilizing the strengths of both quantitative and qualitative data while compensating for their limitations. For this thesis, combining methods can help to validate quantitative findings through qualitative data gathered from interviews, and the other way around. With this approach it's crucial to distinguish between different methods of data collection and use each to validate the other's findings, a process known as triangulation (Denscombe, 2010).

2.2 Research Procedure

This thesis project was performed in three separate phases. The first phase aimed to answer the first research question stated in 1.3, which is to analyze how the new risk assessment method compares against the previous assessments performed by an expert. The second phase aimed to answer the research questions regarding the scope of variables, risk model, and variable weighting. The final phase aimed to critically discuss changes to the method as well as a potential proposal for a final method.

2.2.1 Phase 1: Quantitative Analysis

In this thesis' first phase, the aim was to evaluate how the new risk assessment compared to the current one. Initially, this comparison was intended to be against real-life situations and their outcomes. However, a lack of real-world outcomes (situations where, for example, the performance guarantee was not met) led to a change where the result of the risk assessment tool was compared against the result of a benchmark, which in this thesis is the result from the current risk assessment tool.

Observational techniques were utilized to acquire data related to how well the new risk assessment tool works. Observations mean that the researcher watches what happens in a certain setting, such as using their senses to collect information directly (Denscombe, 2010). The focus of observations in this thesis was on watching how the risk assessment tool was used, collecting data for comparison with the benchmark risk assessment, and identifying knowledge gaps within the tool.

For this, systematic observations were used. Systematic observations mean that the researcher records every observation in the same way or as part of a bigger system of observations. This method is mostly used to gather quantitative data for statistical research. The most important aspects to consider when gathering data in this way are that the produced data is consistent between different observers and that the data is recorded systematically and thoroughly (Denscombe, 2010).

The benchmark in this case was the risk score from the current qualitative risk assessment tool. The results from the current method and the related evaluated systems were identified through secondary data collection in the archive of previous projects at Sandvik Stat. C&S. Secondary data is data collected for purposes that may or may not relate directly to the researcher's current problem. This involves gathering data that have already been recorded and can be used in combination with or to support primary data gathered from other sources (Tashakkori and Teddlie, 2003). This included previous risk assessments using the current method, performance guarantees, and flow sheets of crusher and screener plants from previous projects.

The comparison involved observing and recording the risk scores from the new risk assessment tool, using the same values and setup as in the projects found in the internal database. This was achieved through the following procedure. Firstly, previously performed risk assessments with the current method, along with the related performance guarantee and flowchart from the sold system, were collected from the internal database. Then, using the data from the flow sheet in PlantDesigner, a risk assessment with the new tool was performed. The score from the current risk assessment along with the score from the new method was recorded. Various characteristics of the system in question and the performance guarantee were documented. Additionally, certain characteristics and observations that were identified through this process, such as aspects and factors causing large deviations in risk score and the presence of aspects in the old assessment method that were absent in the new risk assessment tool, were recorded.

After completing the recording and observation, a statistical analysis was conducted to analyze which characteristics had the highest impact on the risk score of the new tool. Also, which of these characteristics affected the score difference between the two methods was analyzed. This analysis aimed to identify different aspects or factors that were not consistent with the benchmark or with interview results. This process, along with the statistical analysis and tools used will be further discussed in Chapter 4.

2.2.2 Phase 2: Interviews and Evaluation

Using insights from Phase 1 combined with interviews, the new method was further evaluated based on the variable scope, risk model, and the weighting of variables.

Interviews are commonly used to gather data for thesis projects, especially to gather qualitative data. There are three main types of interviews: structured, semi-structured, and unstructured. Structured interviews have predetermined questions with limited room for discussion, while unstructured interviews have no predetermined questions and focus on open-ended discussion. Semi-structured interviews fall between these two, allowing some flexibility for discussion based on the predetermined questions (Hurst, 2023).

In this thesis project, semi-structured interviews were conducted to gather data for evaluating the risk assessment tool. This approach was chosen for its balance between providing a consistent set of questions for all interviewees, ensuring a level of reliability and validity, and allowing flexibility for follow-up questions based on individual expertise and experience with rock processing solutions within the organization of Sandvik Stat. C&S.

The interviews were held with employees at Sandvik who had a high level of knowledge about the performance of Stat. C&S's machines and solutions. This primarily included employees at PIE and Sales at Stat. C&S with insight into the scope of variables needed to accurately estimate performance variability and the relative importance of these variables. It was important that the interviews fully capture the experts' knowledge on the matter since available data and publicly available research on the subject of rock crushing were limited.

Insights from Phase 1, such as the inclusion of certain performance criteria leading to an increase in the risk score, were used when formulating the questions in the interview in order to analyze further and find an explanation of what the deviations against the benchmark came from as well as gather knowledge based on the previous observations of the risk assessment tool.

2.2.3 Phase 3: Discussion and Implementation

In Phase 3, the insights from Phase 1 and Phase 2 were used to discuss recommended changes for the new method. This was handled through the process of triangulation, where the results from the comparison with the benchmark and the interviews were used. The focus of this phase was to find evidence of aspects that were problematic within each source and establish a relationship between the qualitative and quantitative data to find key areas in which the new risk assessment method could improve.

Based on the results, an implementation plan for using the new risk assessment tool within Stat. C&S was formulated. This plan outlines key activities required to improve the tool's accuracy and usability and to ensure that the objective of conducting performance guarantees in the sales areas is achieved.

The result of this phase will be further discussed in Chapter 7. In the discussion, aspects were highlighted such as what changes can be made directly to the risk assessment tool, and what changes can be made in the future if the circumstances change. Emphasis of this discussion was held on the feasibility of usage according to Stat. C&S's needs.

3 Rock Crushing and Screening

Multiple aspects impact how well rock crushers and screens perform in real-life scenarios. This chapter aims to provide an overview of some of the performance criteria used at Stat. C&S, including the characteristics of the crushers and screens, and the factors related to the machines and rocks that influence these performance criteria. Special emphasis will be placed on explaining what can be changed in the machining setup if issues arise, as this will affect the level of risk which is higher when there is less flexibility in the system setup.

3.1 Rock Crushers

The product portfolio of Sandvik SRP has many different types of crushers that vary depending on the use case, level of performance required, crushing mechanism, and selection criteria. The new risk assessment tool is developed to assess the risks associated with the CH, CS, CV, and CJ crushers at Stat. C&S (Sandvik Group, 2024).

The main difference between the crushers lies in the crushing mechanism used within each crusher, which in turn influences the properties that affect performance and the associated risks of meeting performance criteria. There are two primary crushing mechanisms for rock crushers: impact crushing and compression crushing. Impact crushing involves, for example, a hammer striking the material passing through the crusher to cause breakage, while compression crushing occurs when, for instance, an iron mantle compresses the rock to induce breakage (Wills and Finch, 2016).

Depending on the size of the rock, applications are commonly categorized as primary, secondary, and tertiary crushing. Primary crushing refers to the crushing of larger rock, usually right after quarrying. Secondary crushing refers to the crushing that occurs after primary crushing is done and naturally involves smaller rock. Fine crushing, sometimes referred to as tertiary crushing, refers to the crushing step after that, and is usually the last step before collecting the final product. Figure 3.1 illustrates the appearance of a CH crusher at Stat. C&S. In Table 3.1 there is a summary of the different crushers included in the risk assessment tool. Feed size refers to the size of rock that is put through the crushers, and max feed size is the maximum size possible to put through the crusher. Capacity refers to the amount of feed material that the crusher can handle per hour.

Figure 3.1: A Sandvik cone crusher of the model CH440. (Sandvik Group, 2024)

Table 3.1: The different types of crushers in the new risk assessment tool scope and their main application, max feed size in mm, and max capacity in metric tons per hour.

Crusher type	**Main application**	**Max feed size (mm)**	**Capacity (MTPH)**
CH and CS crushers	Secondary and Tertiary crushing	18-310	23-2000
CJ crusher	Primary crushing	460-1170	75-1160
CV crusher	Secondary and Tertiary crushing	40-55	37-585

3.1.1 CH and CS Crushers

The CH and CS models at Stat. C&S are cone crushers that rely on crushing between a stationary concave and a rotating mantle within the machine. The steel mantle rotates at high speeds to crush the rocks, moving in a conical pendulum motion that is off-center. As the mantle rotates, it alternately moves closer to and farther away from the concave, creating larger and smaller openings in the crushing chamber where the rocks are crushed. These openings are known as the closed-side setting (CSS) and the open-side setting (OSS). The CSS refers to the distance between the mantle and the concave when the mantle is in its closed position, while the OSS refers to this distance when the mantle is in the open position (Wills and Finch, 2016), see Figure 3.2 for an illustration.

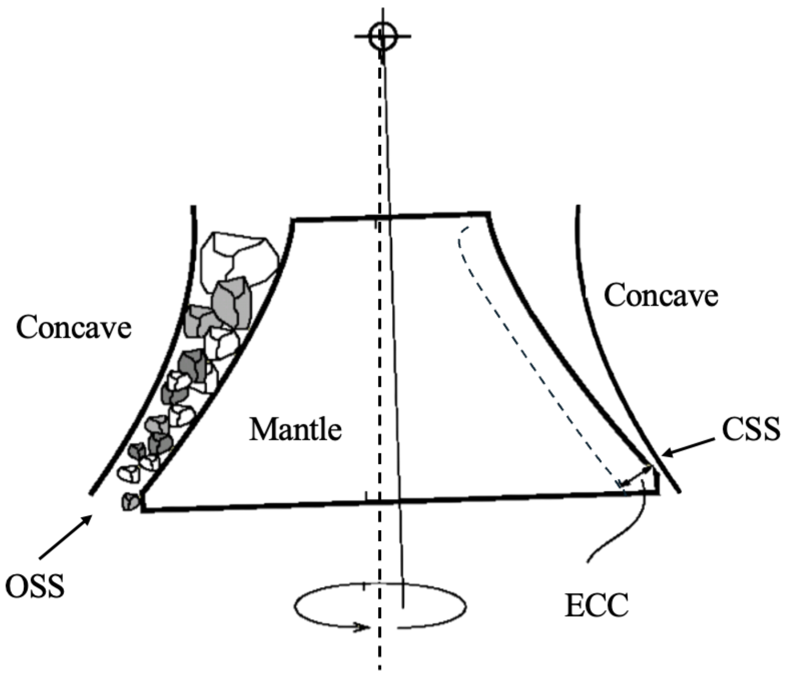

Figure 3.2: An illustration of a cone crusher with OSS, Eccentric Throw (ECC), and CSS (modified from Evertsson, 2015).

When the feed passes through the crusher it will be subjected to several repeated compressions because of the cyclic oscillation of the mantle, and this compression between the mantle and a fixed concave creates a reduction of the size of the feed material. With each compression, the feed material adjusts slightly upwards by being pressed within the chamber of the crusher, and after each compression, the material falls downward due to gravitation. The length of this upward movement defines the number of crushing zones within the crusher (Wills and Finch, 2016).

As can be seen in Table 3.1, the main application of CH and CS crushers is in the secondary or tertiary crushing stages. But in some cases, it can be used in the primary crushing stage, replacing a CJ crusher.

There are two settings on a CH and CS crusher that can be adjusted if the system fails to meet the desired performance criteria. This is the eccentric throw (ECC) and the closed side setting (CSS) of the crusher. The ECC is the difference in length between the CSS and OSS, representing the length of the mantle oscillations within the crushers (see Figure 3.2 for an illustration). The eccentric throw can be altered by adjusting the mantle height, which affects the length of the mantle oscillations (Sandvik, 2024). Similarly, the CSS can be adjusted by changing the horizontal position of the mantle. Due to there being a limit to how much the mantle can move in height and to the sides, there is an upper and lower bound of the value of both CSS and ECC. So, although these settings could be modified, there exists a limit to how much these settings can change before they create issues with the functionality of the machines (Wills and Finch, 2016).

3.1.2 CJ Crushers

The CJ crusher, commonly referred to as the jaw crusher at Stat. C&S, is a compression-based crusher that functions by having one stationary and one moving plate of steel. The breakage is caused by the moving plate moving back and forth toward the fixed plate to compress the material passing through the crusher. Gravity then pushed the material toward an opening at the bottom, with the final size of the material being the same or smaller than the outlet at the bottom. Like the CH- and CS crushers, the jaw crusher has a CSS which is a measurable distance and decides the size of the end product. For the jaw crusher, this distance is the length between the two plates at the bottom outlet when the moving plate is as close as possible to the stationary plate, see Figure 3.3 for an illustration. The main application of a jaw crusher is within primary crushing stages and the jaw crusher is most effective with size reduction of blasted feed material with a top size of at least 150 mm (Wills and Finch, 2016).

The jaw crusher has primarily one setting that can be adjusted if the crusher is not meeting a performance criterion, this is the value of CSS. The value of the CSS can be adjusted by moving the plates toward or further away from each other, but there is a limit to how much this can be done. The jaw crusher thus has a minimum and a maximum CSS that it can achieve (Wills and Finch, 2016).

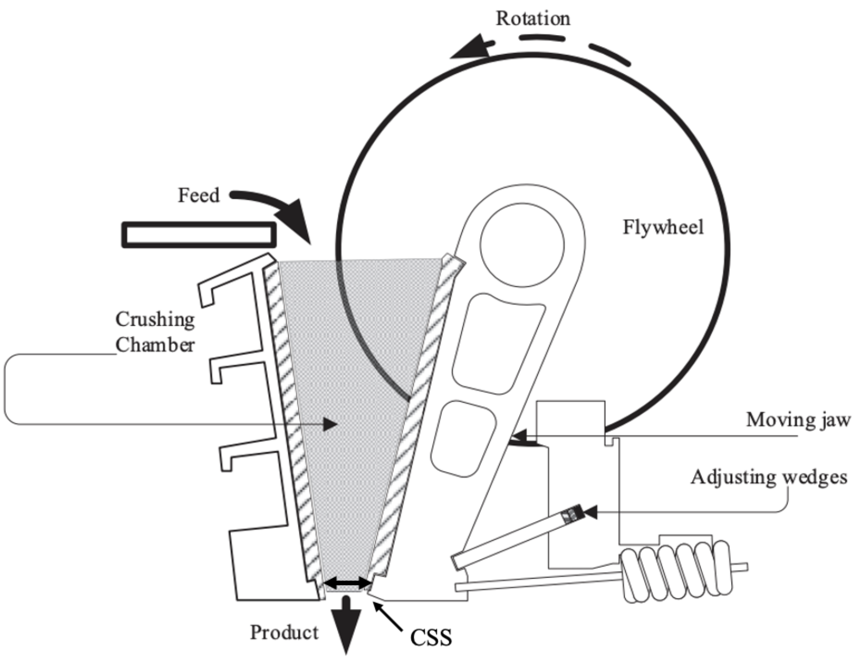

Figure 3.3: Schematic illustration of a jaw crusher by J.Quist, modified to include CSS (Johansson, 2019).

3.1.3 CV Crushers

Sandvik SRP's CV crushers are mainly used in applications where the customer desires a greater shape of the product. The crusher relies on impact crushing to cause breakage. This works by having a high-speed rotor throwing the feed material into a crushing chamber that is lined by the same material. This causes rock-on-rock collisions within the crusher, commonly referred to as interparticle crushing, which aids in providing products with better shape compared to other crushing mechanisms (Wills and Finch, 2016). The CV crusher is most commonly used in tertiary crushing stages.

There is mainly one setting that can be adjusted if the system is not working to the desired specification. This is the Bi-Flow system in the CV crusher. The Bi-Flow works by letting some material flow through the Bi-Flow gates and channel into the crushing chamber, hitting the material coming from the rotor in the opposite direction to create further breakage. The main application of this setting is increasing capacity.

3.1.4 Common Crusher Aspects

One of the most important aspects that customers have to consider when choosing which of these crushers to use is the raw material characteristics and the total reduction ratio of the system. The total reduction ratio of the system is the size ratio between the blasted rock or feed material that goes into the system of crushers and the desired size of the end product. Commonly, the reduction ratio is calculated as:

$$Reduction\,ratio = \frac{F80}{P80} \tag{3.1}$$

where $F80$ is the 80th percentile by size in the feed material and $P80$ is the 80th percentile by size in the end product material (Wills and Finch, 2016).

In some cases, one crusher can handle the entire crushing process but in most cases, multiple crushers are required to get the desired reduction ratio and end product. Thus, all of the crushers must perform well as the input of one crusher can depend on the output of another.

An additional important property to consider is the configuration of the crusher system, e.g. how the connections between the crushers and screens work. Usually, there is a system of conveyors and feeders that transport the rocks between the machines. There are two standard configurations for this: the crusher can either be in a closed- or an open circuit (Wills and Finch, 2016). If the crusher is in an open circuit it means that the rock only goes through the crusher once and then continues on its journey to a secondary- or tertiary crushing stage. This configuration is usually utilized in primary- and secondary crushing stages. If the crusher is in a closed circuit it means that the output product that is larger than a certain size circles back and is fed back into the crusher until it reaches the desired size. This configuration is most common in tertiary crushing stages and when a final product is desired. Both of these configurations will impact how well the crushers and machines will function, but different properties are important depending on the configuration.

3.2 Screens

This section aims to provide an overview of the screen models at Stat. C&S. The selection of which screen type should be used in a crushing plant depends on multiple different factors. Most importantly, the desired screening accuracy together with the capacity of feed material that the screen should be able to handle (Wills and Finch, 2016).

Screens separate rock into different sizes. To do this, screens have multiple decks containing screening media with differently-sized holes for some of the rock to pass through. As an example, a screen for fractioning final products might have two decks for three final products sized 0-4 mm, 4-8 mm, and larger than 8 mm respectively. Depending on the characteristics of the raw material and the output requirements, different materials, hole shapes, and thicknesses for the screening media are used. As an example, for highly abrasive raw materials, a thicker screening panel made of rubber is preferable. If higher accuracy is demanded, steel wire is the better option.

Stat. C&S offers a wide range of different screens to their customers. This thesis will cover the SA screens, SJ screens, and SL screens sold at Stat. C&S (Sandvik Group, 2024). See Figure 3.4 for an image of an SJ screen. While many factors are important when choosing the right screen, three of the most important factors are the maximum feed size, the maximum separation size, and the movement mechanic. Rock smaller than the separation size of a deck will (in optimal conditions) fall to the deck below. Movement mechanic is the movement the screen uses to move the rocks over the screening media. Table 3.2 below summarizes the characteristics of each type of screen covered in this thesis.

Figure 3.4: A Sandvik SJ Screen (Sandvik Group, 2024)

Table 3.2: Table 2: Different types of screens and their main application, max feed size in mm and movement mechanic and max separation size in mm. Scalping refers to the removal of material that is difficult to process.

Type of screen	**Movement mechanic**	**Main application**	**Max feed size (mm)**	**Max separation size top deck (mm)**
SA Screen	Medium and fine screening	Circular motion	150	80
SJ Screen	Medium and fine screening	Circular motion	400	140
SL Screen	Mining duties	Linear motion	300	100

Screens vibrate at high frequencies to promote the stratification of the rocks. In screening processes, the term 'stratification' describes the phenomenon where fine particles move downward through a bed of coarser particles on a vibrating screen bed. Generally, the more successful the stratification of the material is on the screen bed, the better the results will be in terms of screening accuracy and throughput. The promotion of stratification depends on multiple factors, including material density, particle shape, distribution of particle size, and material humidity. The stratification process is most effective when the size of the material is non-homogeneous, for example, when the feed material consists of a mix of large and small materials (Shen and Tong, 2020).

Two main aspects are important to consider when configuring a screen within a crushing- and screening plant. That is the size of the screening area and the separation difference of decks (Wills and Finch, 2016). The screening area decides the screening efficiency, where the capacity of feed material needs to be handled while the bed depth needs to be of the right size. The bed depth refers to the height of the material that is on the screening deck. The bed depth should neither be too large, which would mean that rocks smaller than the holes would not have time for effective stratification and fall through the screening media, or too small, which would mean that the material would vibrate on the screening media without it being pressed downwards by larger rocks above it. The separation difference of decks is important due to the same reasoning, if the separation difference between decks is narrow, e.g the holes are 20 mm on one deck and 18mm on the next, it will be difficult to make sure that the bed depth is consistent and the screening process would be ineffective.

The main difference between the different models included in the risk assessment tool is the type of motion and the inclination of the screen. For example, the SA Model has a circular motion, and a steeper inclination between 15-18 degrees, which means that the capacity is higher, but the material will have less time on the screen and the effective hole size reduces. This means that the screening accuracy is lower. This can however be compensated by having a longer screen and a larger screen area. The SL model uses linear motion and the deck vibrates back and forth with inclination between 0-10 degrees. This means that the material will move slower over the screen providing greater accuracy but at the cost of capacity.

3.3 Performance Criteria

Multiple different performance criteria are used in the performance guarantees given by Stat. C&S. These criteria are in most scenarios specified by a numerical value that the system or product should achieve. This section aims to give an overview of some of the performance criteria utilized and the most important aspects that impact the values of the criteria.

3.3.1 Capacity

In most performance guarantees an output capacity is specified. Capacity in this context refers to the output quantity of crushed rock and is measured in metric tons per hour (MTPH). This performance criterion depends on how well the crusher functions and certain settings available in the crusher. The settings that have the largest impact on how the capacity of a crusher varies are the crusher's CSS and ECC. If the CSS is changed to be larger, e.g. there is a longer distance between the mantle and the concave, the capacity of the crusher will be higher as more material can pass through in a given amount of time. Likewise, a higher value of ECC leads to a greater capacity of the crusher due to more material being able to pass through per second compared to a lower ECC setting (Wills and Finch, 2016). However, in most scenarios, the ECC will likely be the first setting to be adjusted if the specified capacity is not reached. This is due to the CSS impacting other factors such as product size grading and shape, and an adjustment of the CSS to reach a capacity can mean that these other criteria would not be reached.

In a performance guarantee, the capacity is in most scenarios viewed either as the total output capacity or the output capacity of a certain product range. The first case takes the entire crushing and screening plant into account. This means that the output capacity is the same as the capacity of feed material into the system. In the other case where a certain product range is guaranteed, the customer desires an output capacity of rocks within a specific product size, for example, 400 MPTH of rocks between 8-12mm in size. This criterion is referred to as product and capacity.

3.3.2 Particle Size Distribution

In some performance guarantees a particle size distribution (PSD), is given as a performance criterion. The PSD is a description of how the sizes of the material into and out of the crushers and screens vary. It is given as a cumulative percentage of the total weight divided by the total weight over a certain size (Wills and Finch, 2016), see Figure 3.5 for an example of how a PSD curve can look. The PSD curve is calculated in PlantDesigner after each crushing- or screening stage and is also utilized to describe the feed of blasted rock that goes into a system.

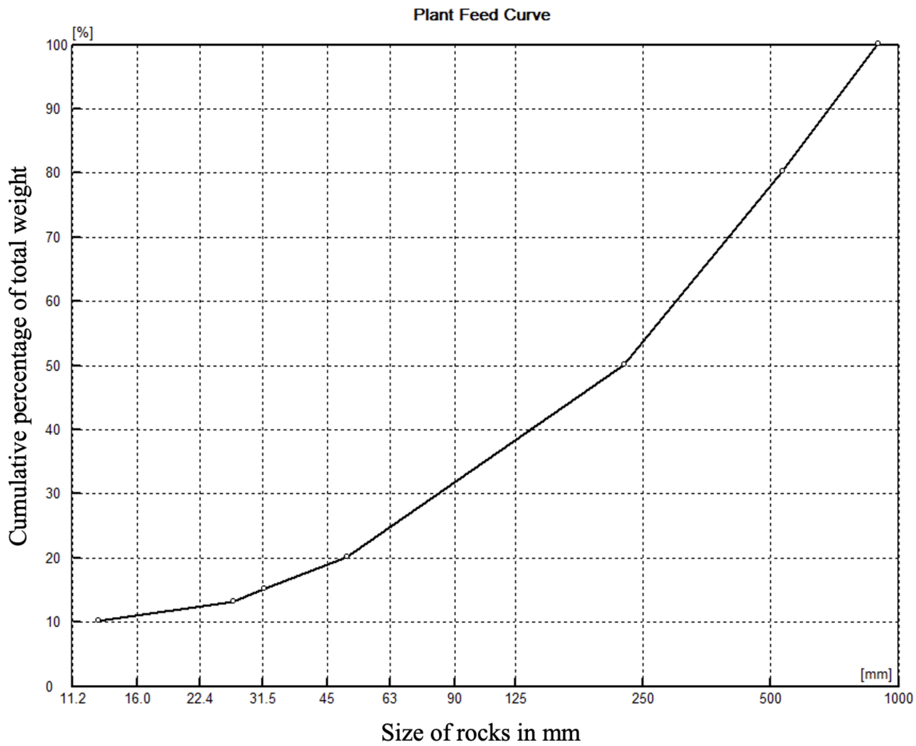

Figure 3.5: An example of a PSD curve, in this case, the curve describes the feed material going into a crushing and screening plant.

Generally, customers want a certain percentage of the feed out of the crushers to be smaller than a predetermined size instead of the entire PSD curve. This is depicted within the industry as PX, where X is indicative of a percentage within the PSD curve and P stands for the product. As an example, a performance criterion could be P40, which means that 40% of the material out of the crusher should be smaller than the size that comes with the criteria. Thus, a P40 of the PSD curve in Figure 3.5 means that 40% of the product should be smaller than about 125mm. In some cases, multiple product ranges could also be included within a single guarantee.

The PSD of the feed is affected by values of multiple characteristics of the crushers. The PSD curve is directly related to the reduction ratio. In the context of PSD as a performance criterion, it is usually preferred to have a higher percentage of material below the specified size after each crushing stage. The calculation of the PSD of the product in a cone crusher is based on two different calculation models. The first of which is a model that describes the flow of material that passes through the crusher and the second model describes the size reduction and breakage process (Evertsson, 2015).

According to Evertsson (2015), three main factors affect the size reduction process for a CH- or CS-crusher. These are the breakage modes, the number of crushing zones, and the compression ratio. These factors in turn depend on eccentric speed, CSS, and the raw material breakage characteristics (Evertsson, 2015). The CSS setting is the only factor that can be adjusted if there are issues with the size of the product. However, there are risks associated with achieving the desired PX or PSD because the CSS setting has a theoretical minimum and maximum value. For example, if the CSS was at its minimum theoretical value and the system still had issues with reaching a certain reduction, there would be a need to switch the chamber of the crusher or the entire machine to correct it and get the desired size of the product.

3.3.3 Product Shape

In some cases, a particular shape is promised with a performance guarantee. Shape, in this case, refers to how similar the dimensions of a rock are to a cube's dimensions, in the sense that the depth, width, and length of the rock are of equal length. The shape is not related to the PSD curve or the actual size of the rock and instead refers to the cubicity of the rock. This can be measured in several different ways depending on where the crusher is sold and the regional norm for shape. The most common measurements are the Flakiness Index (FI) and the Shape Index (SI) according to European standards EN 933-3:2012 and EN 933-4:2008. FI is calculated as

$$FI = \frac{W2}{W1} \cdot 100, \tag{3.2}$$

where $W2$ refers to the total weight of tested particles whose shortest dimension is less than 60% of its mean dimension, and $W1$ refers to the total weight of all particles. SI is calculated as

$$SI = \frac{M2}{M1} \cdot 100, \tag{3.3}$$

where $M2$ is the weight of particles whose length-to-thickness ratio exceeds 3. Length refers to the longest side of a rock and thickness refers to the longest measure perpendicular to the length. $M1$ is the weight of all tested particles.

The shape is dependent on multiple factors. As mentioned in section 3.1.3 the more interparticle crushing there is in a system, the better the shape will be. As such, the CV crusher is generally preferred compared to other crushers if the product shape of the final product is the most important factor in a system. While this is true, the CH and CS crushers can provide a good product shape too but it's harder to predict the expected value of SI or FI out of a system.

3.3.4 Oversize and undersize of particles

The degree of oversize and undersize of particles refers to the allowed percentage by weight of particles within a given size range that is either too large or too small (Wills and Finch, 2016). Oversize indicates that the rocks exceed the specified size in the performance guarantee, while undersize means that the rocks are too small. This criterion is primarily impacted by screens in a plant. If the performance guarantee is not met, the solution would be to adjust the screen configuration or the shape or size of the holes on the screening media.

In most cases, the screen holes are slightly larger compared to the diameter of the rock. For instance, if a customer requests a product size of 15mm, the holes typically have diameters of 17-18mm. As such, there's a possibility that material between 16-18mm will pass through to a lower screening deck, resulting in oversized material. The reason for having slightly larger holes than the product size is to enhance screening efficiency in terms of capacity and to prevent undersized material in the final product.

The presence of undersized material in the product is due to insufficient time for stratification of all materials, causing them to not pass through the holes in the screening media when they should.

See Table 3.3 for an illustrated example where there is a depiction of four different product ranges and an example of what a typical oversize/undersize of these particles could be in the respective ranges.

Table 3.3: An example of allowed oversize and undersize for different size ranges of particles.

Product size (mm)	**Allowed Undersize/Oversize (%)**
0-4	0/20
4-8	15/20
8-12	15/15
12-20	10/15

4 Theory

In this chapter, all theory used in the master thesis are presented. The theory comprises statistics and data analysis, including regression and statistical tests.

4.1 Multiple linear least-squares regression

A linear regression model is a statistical model that estimates the linear relationship between a single response variable and a set of explanatory variables. When the model includes more than one explanatory variable, the model is commonly called a multiple linear regression model. (Fox, 2024). Multiple regression is used in this thesis to predict the effect of certain crushing system characteristics on the risk score for the new tool. Ordinary least squares (OLS) is used to estimate the slope coefficients $\beta_0, \beta_1, \ldots, \beta_k$ in the model presented below in (4.1). Given the assumptions described in Subsection 4.1.2, OLS is the best linear unbiased estimator (Gujarati and Porter, 2009). In short, OLS estimation is an optimization problem where the sum of the squared error terms is minimized with respect to the betas. The estimation process will be detailed in Appendix A.2.

4.1.1 Terminology

The multiple regression equation can be written as:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i, i = 1, \ldots, n, \tag{4.1}$$

where Y_i is the response variable, $x_{i1,} x_{i2}, \ldots, x_{ik}$ are the explanatory variables, $\beta_1, \beta_2, \ldots, \beta_k$ are the slope coefficients for the explanatory variables, β_0 is the intercept, and ε_i is the error term (Fox, 2024). In this master's thesis, Y_i will be related to the risk scores, and x_{ik} will refer to the kth characteristic for crushing solution i, or in other terms the ith collected data row. For convenience, matrix notation will be used, where the equation can be written as:

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{e} \tag{4.2}$$

with:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & & x_{2k} \\ \vdots & & \ddots & \\ 1 & x_{n1} & & x_{nk} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \beta_0 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \mathbf{e} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{4.3}$$

4.1.2 Assumptions

Underlying assumptions for the model concern the error terms in $\mathbf{e}$. The error terms are assumed to be independently normally distributed with constant variance and mean zero, expressed as $\varepsilon_i \sim N(0, \sigma_\epsilon^2)$. Furthermore, explanatory variables x_{ij} are assumed to be fixed or measured without error, as well as independent of ε_i. This means that no component of ε_i is related to the measuring error of x_{ij}. Also, non-linearity is assumed, meaning that there is no linear relationship between explanatory variables (Fox, 2024). An explanation for why this is important is given in subsection 4.4.4.

4.1.3 Parameter estimation using OLS

Ordinary least squares (OLS) is used to estimate the slope coefficients $\beta_0, \beta_1, \ldots, \beta_k$ in the model presented below in (4.1). Given the assumptions described in Subsection 4.1.2, OLS is the best linear unbiased estimator (Gujarati and Porter, 2009). In short, OLS estimation is an optimization problem where the sum of the squared error terms is minimized with respect to the betas. The estimation process will be detailed in Appendix A.2.

4.2 Multinomial logistic regression

In contrast to multiple linear regression, a multinomial logistic regression model approximates the relationship between multiple explanatory variables and a categorical response variable (Fox, 2024). A categorical variable can take on a limited, usually fixed set of values. Logistic regression will be used in this thesis to model the effect of certain variables on the absolute difference in risk scores between the new tool and the current one. The main reason for using logistic regression instead of linear regression for this scenario is that the absolute difference of the scores can take on discrete values 0, 1, 2, 3, 4, and 5. A value of 5 represents the highest difference in risk, and 0 is no difference. This data type is called ordinal data because it implicates a hierarchical order. Ordered logistic regression is commonly used when the response variable is of ordinal type. To arrive at the ordered logistic regression model, the simpler case of logistic regression for a dichotomous response variable will first be explained. A dichotomous variable is a categorical variable that can only take on two possible values, for example, 0 and 1. This type of logistic regression is formally called binary logistic regression. The model is then generalized to apply to ordered response variables. Approximation of the coefficients is done using maximum likelihood estimation, or MLE in short. Because of the dichotomous nature of the response variable for some logistic regression models, MLE is used instead of least squares (Fox, 2024). This is because OLS assumes a linear model and therefore is based on the assumption in 4.1.2 of normally distributed errors. This assumption is not justified theoretically with a dichotomous response variable (Schield, 2017). MLE estimates the slope coefficients that maximize the likelihood of observing the data given that the data follows the logistic regression model. While still an optimization problem, it is distinctly different from OLS which assumes a linear regression model and normally distributed errors and therefore will not provide the same estimation results (Fox, 2024). For interested readers, the estimation method of MLE will be detailed in A.3.

4.2.1 Binary multinomial logistic regression

The purpose of the logistic regression is to classify an outcome variable based on a probability of occurrence. First, let Y_i take on value 1 with probability π_i and 0 with probability $1-\pi_i$. Furthermore, the one-variable logistic distribution function is defined as (Fox, 2024):

$$\Lambda(z) = \frac{1}{1+e^{-z}}, z \in \mathbf{R} \tag{4.4}$$

Now, let us assume π_i to be the probability returned by the logistic distribution function of a linear predictor of several regressors. This is called the linear logistic regression model (Fox, 2024):

$$\begin{aligned} \pi_i &= \Lambda\left(\alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots \beta_k X_{ik}\right) \\ &= \frac{1}{1+e^{-(\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}+\ldots\beta_k X_{ik})}} \end{aligned} \tag{4.5}$$

For interpretability reasons, we want to write the linear predictor on one side of the equation. The inverse linearizing transformation, also known as the log-odds of the model, is given by (Fox, 2024):

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots \beta_k X_{ik} \tag{4.6}$$

4.2.2 Ordered logistic regression

As mentioned a logistic regression model can be formulated to produce an ordered response instead of a dichotomous one. Let ξ_i denote a probability that is a linear function of the included X_is plus a random error ε_i (Fox, 2024):

$$\xi_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_i \tag{4.7}$$

In the case of ordered logistic regression, instead of dividing ξ into two sections like the dichotomous case, one can divide them into m cases, defined by numbers $\alpha_i, i = 1, \ldots, m$. The response variable Y_i is therefore defined as:

$$Y_i = \begin{cases} 1 & \text{if } \xi_i \leq \alpha_1 \\ 2 & \text{if } \alpha_1 < \xi_i \leq \alpha_2 \\ \vdots & \\ m & \text{if } \alpha_m < \xi_i \end{cases} \tag{4.8}$$

Using (4.8), the cumulative distribution function of Y_i can be written as:

$$\begin{gathered} Pr(Y_i \leq k) = Pr(\xi_i \leq \alpha_k) \\ = Pr(\alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_i \ leq \alpha_k) \\ = Pr(\varepsilon_i \leq \alpha_k - \alpha - \beta_1 X_{i1} - \beta_2 X_{i2} - \ldots - \beta_k X_{ik}) \end{gathered} \tag{4.9}$$

Assuming errors ε_i are distributed according to the logistic distribution presented in (4.4), the log-odds of the model can be formulated using (4.9) (Fox, 2024):

$$\begin{gathered} ln \frac{Pr(Y_i > k)}{Pr(Y_i \leq k)} \\ = (\alpha - \alpha_k) + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} \end{gathered} \tag{4.10}$$

for $k = 1, 2, \ldots, m-1$. As is clear from (4.10), slopes are the same for all m cases of Y_i, the only thing that varies is the intercept (Fox, 2024). The fitting of the model is done by the method of maximum likelihood, similar to the binomial case, as first shown by McCullagh (1980).

4.3 Dummy variable regression

Dummy variables are qualitative explanatory variables that take on two or more categories, e.g. they represent either a dichotomous or polytomous factor. A polytomous factor is one that can take on multiple discrete values, e.g. more than two like in the dichotomous case.

4.3.1 Terminology

When including a dichotomous variable, for example whether a performance guarantee includes the criteria PSD or not, a multiple linear regression model can be formulated as (Fox, 2024):

$$Y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \gamma D_i + \epsilon_i, \tag{4.11}$$

where D_i, commonly called the dummy variable regressor, is coded as:

$$D_i = \begin{cases} 1 & \text{when PSD included} \\ 0 & \text{when PSD is not included} \end{cases} \tag{4.12}$$

In the case of a polytomous factor, there are multiple ways of coding the different factors into the regression model. In the case of a dummy variable with three different categories, let's call them category $1, 2$ and 3, they can be represented by introducing two dummy variable regressors to the model, D_1 and D_2 (Fox, 2024):

Category	D_1	D_2
1	1	0
2	0	1
3	0	0

(4.13)

The regression model is now formulated as:

$$Y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \gamma_1 D_1 i + \gamma_2 D_2 i + \varepsilon_i \tag{4.14}$$

It is clear that γ_1 represents the increase in Y between the baseline category 3 and category 1 and that γ_2 represents the increase between category 3 and category 2. It is also clear that for a polytomous dummy variable with n categories, $n-1$ regressors need to be added to the model using this coding technique (Fox, 2024).

4.4 Statistical inference

Statistical inference is the method of analyzing quantitative data and drawing certain conclusions about the characteristics of the population from which the data originates. In the case of this master's thesis, an example of that would be the conclusion that the risk scores from the current and new method are equally distributed. In this section, the statistical tests and calculations that are used for inference in Phase 1 of the research procedure are explained. As with every statistical test, a null hypothesis and an alternative hypothesis are formulated regarding some property of the underlying population from which a sample is collected.

A test statistic with a known distribution is subsequently calculated, and depending on whether this value falls within or outside the rejection region in regards to a certain significance level, the null hypothesis is either rejected or retained.

4.4.1 Wilcoxon signed rank test

Wilcoxon signed rank test is the non-parametric equivalent of the paired samples t-test and should be used when the sample data is severely non-normally distributed. One example of this would be if the data is ordinal (see explanation of ordinal data in subsection 4.2.2). The Wilcoxon signed rank test can for example be used to tell whether the distribution of two paired samples is significantly different or not (Rice, 2007). In the case of this master's thesis, this test will be used to conclude whether the current risk score and the new risk score have equal distributions or not. Since the formats of the risk scores are ordinal, the Wilcoxon test is the right fit for this purpose.

Let $(X_i, Y_i), i = 1, \ldots, n$, be the subcollection of sample data pairs where $X_i - Y_i \neq 0$. Furthermore, let $D_i = X_i - Y_i$ and compute $|D_i|$ and sort the list containing $|D_i|, i = 1, \ldots, n$, in ascending order. After this is complete, ranks $R_i, i = 1, \ldots, n$ are assigned to the list where the smallest has rank 1, the second smallest has rank 2, etc. Now let:

$$T^{+} = \sum_{1 \leq i \leq n, D_i > 0} R_i \tag{4.15}$$

$$T^{-} = \sum_{1 \leq i \leq n, D_i < 0} R_i \tag{4.16}$$

The Wilcoxon Signed Rank test statistic is given by (Rice, 2007):

$$W = \min\left(T^{+}, T^{-}\right) \tag{4.17}$$

If the null hypothesis is that there is no difference in the rank sum, e.g. the two samples have an equal median value, then the following z-value can be computed:

$$z = \frac{W - \mu_W}{\sigma_W} \tag{4.18}$$

Here, μ_W under the above-mentioned null hypothesis is calculated as:

$$\mu_W = \frac{n(n+1)}{4} \tag{4.19}$$

n is the sample size, $n = 177$ in this thesis. The standard error σ_W is calculated as:

$$\sigma_W = \sqrt{\frac{n(n+1)(2n+1) - \sum_{i=1}^{k} \frac{t_i^3 - t_i}{2}}{24}} \tag{4.20}$$

Here, k is the number of tied ranks, e.g. where multiple samples have the same rank. t_i is the number of tied samples for the ith tied rank (DATAtab, 2024). The z-statistic in (4.19) is standard normally distributed under the null hypothesis and is used to calculate a corresponding P-value for the test.

4.4.2 McNemar's test

Similar to the Wilcoxon signed rank test, McNemar's test is also a statistical test suitable for inference paired sample data, specifically dichotomous data. It is commonly used to test whether a before and after effect on the same sample is distinctly different in terms of a dichotomous outcome. In the context of this master's thesis, that would be whether the new risk tool recommends proceeding with sales of certain crushing system (e.g. the risk score is sufficiently low) significantly differently compared to the current tool's risk scores on the same flowsheets (see variable definitions for sales recommendations in (5.6) and (5.5)). For a sample size of N, a 2x2-contingency table can be displayed as (Fagerland et al., 2013):

Table 4.1: An example of a $2x2$ contingency table representing before and after test results. a, b, c, d refer to the frequency of the different outcomes.

	Test 2 positive	Test 2 negative	Row Total
Test 1 positive	a	b	$a+b$
Test 1 negative	c	d	$c+d$
Column Total	$a+c$	$b+d$	N

The null hypothesis would be that the total probabilities for each outcome are the same for both tests:

$$P_a + P_b = P_a + P_c \tag{4.21}$$

$$P_c + P_d = P_b + P_d, \tag{4.22}$$

where P_j denotes the theoretical probability of outcome j in the contingency table, $j = a, b, c, d$. After simplification of the equations, the null and two-sided alternative hypotheses are formulated as (Fagerland et al., 2013):

$$H_0 : P_b = P_c \tag{4.23}$$

$$H_1 : P_b \neq P_c \tag{4.24}$$

The McNemar test statistic is given by:

$$\chi^2 = \frac{(b-c)^2}{b+c} \tag{4.25}$$

With sufficiently large b and c, χ^2 has a chi-squared distribution with one degree of freedom under the null hypothesis. If $b + c < 25$, a variant of the McNemar's test, commonly called the exact McNemar's test or the mid-p McNemar's test, is preferable. In this test, b is compared to an exact binomial distribution with $n = b + c$ and $p = 0.5$. The two-sided exact p-value is calculated as (Fagerland et al., 2013):

$$P_{exact} = 2\sum_{i=b}^{n} \begin{pmatrix} n \\ i \end{pmatrix} 0.5^i (1-0.5)^{n-i} \tag{4.26}$$

To obtain a more conservative result, half of the probability of reaching b under the null hypothesis (the binomial distribution under the null hypothesis) can be subtracted from P_{exact} to calculate what is commonly referred to as the mid P-value (Fagerland et al., 2013):

$$\begin{aligned} P_{mid} &= 2\left(\sum_{i=b}^{n} \begin{pmatrix} n \\ i \end{pmatrix} 0.5^i (1-0.5)^{n-i} - 0.5 \begin{pmatrix} n \\ b \end{pmatrix} 0.5^b (1-0.5)^{n-b}\right) \\ &= P_{exact} - \begin{pmatrix} n \\ b \end{pmatrix} 0.5^b (1-0.5)^{n-b} \end{aligned} \tag{4.27}$$

4.4.3 Coefficient of determination

The coefficient of determination, commonly referred to as R^2, is a measure of the goodness of fit of a linear model to its observed data sample. R^2 is defined as the ratio between the sum of squares (more precisely the sum of the squared error terms) of the fitted model and the sum of squares of the observed data and is therefore a measure of what degree the fitted model explains the variance of the observed data. The mathematical definition is (Fox, 2024):

$$R^2 = \frac{RegSS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} \tag{4.28}$$

$RegSS$ is defined as the sum of squares for the predicted values of the regression, which is calculated by taking the difference between the total sum of squares, TSS, and the sum of squares from the residuals. The total sum of squares, TSS is given by:

$$TSS = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 \tag{4.29}$$

Y_i and $\overline{Y}$ are the observed values and mean value for the dependent variable. The residual sum of squares RSS is given by:

$$RSS = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} \hat{E}_i^2 \tag{4.30}$$

$\hat{Y}_i$ are the fitted values from the regression. Clearly, a value close to 1 is indicative of a well-fitted model to the observed data. Adding more variables to a fitted model will always increase R^2. To prevent an overfitted model, the adjusted R^2, denoted as $\overline{R}^2$ can be used instead (Gujarati and Porter, 2009):

$$\bar{R}^2 = 1 - \frac{(1 - R^2)\,(n-1)}{n-k}, \tag{4.31}$$

where k is the number of explanatory variables in the model and n is the data sample size (number of rows in the sample data).

4.4.4 Variance inflation factor

To analyze the correlation of an explanatory variable, X_j, with other explanatory variables $X_i, i \neq j$, a common measure is the variance inflation factor, or VIF in short. The mathematical definition for the variance inflation factor for X_j, denoted VIF_j, is as follows:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{4.32}$$

R_j^2 is the coefficient of determination (see definition at 4.4.3) from the regression of X_j on all other explanatory variables $X_i, i \neq j$ in the model (Fox, 2024). Let us use X_1 as an example. To calculate R_1^2, first perform ordinary least squares regression, OLS, for the following model (see section 4.1 for an explanation of OLS):

$$X_1 = \alpha_0 + \alpha_2 X_2 + \alpha_3 X_3 + \cdots + \alpha_k X_k + \varepsilon, \tag{4.33}$$

where α_0 is the intercept and ε is the error term. Then, the coefficient of determination is calculated using (4.28) but for the regression model in (4.33) instead of the ordinary linear regression model defined in (4.1). A high value for VIF_j implies a strong correlation between X_j and the other explanatory variables. In the case of perfect multicollinearity, e.g.

when two or more variables have a perfect linear relationship, the VIF is infinite (Fox, 2024). When performing regression modeling in this thesis, an aim is to have a low VIF for the included variables, because the model's explanatory power is higher when the explanatory variables have low correlation.

4.4.5 Wald's test

Wald's test is a common statistical test for an individual coefficient in a logistic regression model to test the null hypothesis $H_0 : \beta_j = \beta_0$. The Wald statistic is calculated as (Fox, 2024):

$$Z_0 = \frac{\hat{\beta}_j - \beta_0}{SE\left(\hat{\beta}_j\right)}, \tag{4.34}$$

where $\hat{\beta}_j$ is the MLE of coefficient β_j and $SE\left(\hat{\beta}_j\right)$ is the asymptotic standard error of $\hat{\beta}_j$.

$SE\left(\hat{\beta}_j\right)$ is commonly estimated using Fisher's information matrix (Fox, 2024):

$$I(\hat{\boldsymbol{\beta}}) = -\frac{\partial}{\partial^2} \ln(L(\hat{\boldsymbol{\beta}})), \tag{4.35}$$

where $L(\hat{\boldsymbol{\beta}})$ is the likelihood function for the parameters in the model (Lixoft, 2024). Inserting the likelihood function from Equation (A.6), we have:

$$\begin{aligned} \mathbf{I}(\hat{\boldsymbol{\beta}}) &= -\frac{\partial}{\partial^2}\left(L\left(\alpha, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k\right)\right) \\ &= \frac{\partial}{\partial^2}\left(\prod_{i=1}^{n} \frac{e^{\left(\alpha+\hat{\beta}_1 X_{i1}+\hat{\beta}_2 X_{i2}+\ldots\hat{\beta}_k X_{ik}\right)^{y_i}}}{1+e^{\alpha+\hat{\beta}_1 X_{i1}+\hat{\beta}_2 X_{i2}+\ldots\hat{\beta}_k X_{ik}}}\right) \end{aligned} \tag{4.36}$$

The inverse matrix is the variance-covariance matrix $\mathbf{C}(\hat{\boldsymbol{\beta}})$

$$\mathbf{C}(\hat{\boldsymbol{\beta}}) = \mathbf{I}(\hat{\boldsymbol{\beta}})^{-1} \tag{4.37}$$

The standard error of the MLE coefficient can now be calculated by taking the square root of the corresponding variance in the variance-covariance matrix (Lixoft, 2024):

$$SE\left(\hat{\beta}_j\right) = \sqrt{\mathbf{C}_{jj}(\hat{\boldsymbol{\beta}})} \tag{4.38}$$

The Wald test statistic follows an asymptotic standard-normal distribution under the null hypothesis. While the Wald's test is usually relatively accurate for large sample sizes, a Type-2 error can sometimes occur under certain circumstances in logistic regression. A type-2 corresponds to wrongfully retaining the null hypothesis. Combining it with the more computation-heavy likelihood-ratio test is therefore preferred (Fox, 2024).

4.5 Stepwise regression

An important aspect of regression modeling is selecting the appropriate variables to include in the models. One common method for variable selection is stepwise regression. This method iteratively compares different models based on predefined criteria to arrive at a satisfactory model (Iain, 2024).

While multiple variable selection criteria can be chosen in the stepwise regression, such as increased R^2 and adjusted R^2, a common criterion is the partial F-test or t-test for an individual slope coefficient. First, a significance level is decided for when a variable gets to enter the model, denoted α_E. Also, a significance level for when to remove a variable is decided, denoted α_R. The model starts off empty with no explanatory variables included. (Iain, 2024) The iterative procedure is explained in Appendix A.4.

5 Quantitative results

As mentioned in Chapter 2, the initial phase of the research procedure is the quantitative comparison between the risk scores from the current risk assessment method and the new method. The quantitative comparison can be separated into two levels. The first level is a high-level comparison between the risk scores of the two methods, where some of the statistical tests described in Chapter 4 are utilized to determine whether the current and new methods provide similar risk scores to a statistically significant degree. If the new method provides a risk score similar to the benchmark, that is a good indicator that the new method is well-performing.

Secondly, an analysis of the variables that affect the performance of the new risk method is conducted. To enable this, values of certain chosen variables were recorded simultaneously as the risk scores were recorded. This has been described in detail in Chapter 2. Using linear and logistic regression models, the impact of certain variables on the deviation of the new model to the current model is assessed. Finally, how well the model fits the data is also analyzed.

5.1 Data collection

The data that was recorded and saved when performing the risk assessment comparison is presented in this section.

Table 5.1: All variables recorded in the quantitative comparison between the current and new risk assessment methods.

Variable Meaning	Notation	Type	Values
New risk rating	NRR	Continuous	$[0, 1]$
Current risk rating	CRR	Discrete	$\{1, 2, 3, 4, 5\}$
Capacity included	CI	Binary	{Yes, No}
Product and capacity included	PCI	Binary	{Yes, No}
PX included	PXI	Binary	{Yes, No}
PSD included	$PSDI$	Binary	{Yes, No}
Over and undersize included	OUI	Binary	{Yes, No}
Shape included	SI	Binary	{Yes, No}
Number of CJ crushers	CJ	Discrete	$\{1, 2, 3, 4\}$
Number of CH crushers	CH	Discrete	$\{1, 2, 3, 4\}$
Number of CS crushers	CS	Discrete	$\{1, 2, 3, 4\}$
Number of CV crushers	CV	Discrete	$\{1, 2, 3, 4\}$
Number of conventional screens	SCR	Discrete	$\{1, 2, 3, 4\}$

The variables presented in Table 5.1 have been chosen in collaboration with people at Stat. C&S. The logic for this is that including these variables in a statistical model would give further insight into the performance of the new tool for different types of assessed rock crushing systems. For example, suppose the new tool has a substantially lower risk than the current risk assessment method for systems where PSD is included in the performance guarantee. That indicates that something should be changed to adjust the impact that PSD has on the risk score of the new tool. A more detailed description of these variables can be found in Chapter 3. In the case of NRR, the lower the value the lower the probability to reach the performance guarantee and the higher the risk. For CRR, the lower the value the lower the score, e.g. a score of 1 would denote a very low risk while a value of 5 represents a very high risk. So the two tools have an opposite scale used to denote the risk.

5.2 Initial high level analysis

As mentioned, the first step in the analysis is to compare the two methods based on the risk score. One problem is that the new risk rating is given on a continuous scale between 0 and 1, while the current risk rating is discrete and can take on numbers between 1 and 5. Another issue is that the two methods have opposing scales, meaning that a 1 for the new risk score is a low risk, and a 5 is a high risk for the current risk score. To compare the two variables in a way that makes sense, the new risk rating is translated to a discrete scale using Stat. C&S's pre-defined translation of the risk from the new format to the current. The reason for doing this instead of translating the current risk score to the new format is that Stat. C&S already has the pre-defined translation from new to current. We define a new variable NAC (New As Current) to represent the new risk rating in the current format:

$$NAC = \begin{cases} 1 & \text{if } 0.8 < NRR \leq 1 \\ 2 & \text{if } 0.6 < NRR \leq 0.8 \\ 3 & \text{if } 0.4 < NRR \leq 0.6 \\ 4 & \text{if } 0.2 < NRR \leq 0.4 \\ 5 & \text{if } 0 \leq NRR \leq 0.2 \end{cases} \tag{5.1}$$

Having defined NAC, one can perform the Wilcoxon signed rank test to evaluate whether there is a substantial difference between NAC and CRR. The Wilcoxon test is used because of the variable type being discrete. Let

$$\Delta_i = NAC_i - CRR_i, \tag{5.2}$$

where i stands for the ith observation, $i = 1, \ldots, 178$. The null hypothesis and the two-sided alternative hypothesis are set as:

$$H_0 : \text{ The median of } \Delta_i \text{ equals } 0 \tag{5.3}$$

$$H_1 : \text{ The median of } \Delta_i \text{ does not equal } 0 \tag{5.4}$$

Running the test Wilcoxon Signed Rank Test in SPSS returns the following output:

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The median of differences between New risk rating as current and Current risk rating equals 0.	Related-Samples Wilcoxon Signed Rank Test	<.001	Reject the null hypothesis.

Hypothesis Test Summary

a. The significance level is .050.

b. Asymptotic significance is displayed.

Figure 5.1: Test summary for the Wilcoxon Signed Rank Test for related samples.

Related-Samples Wilcoxon Signed Rank Test Summary

Total N	177
Test Statistic	1007.000
Standard Error	373.879
Standardized Test Statistic	-7.341
Asymptotic Sig.(2-sided test)	<.001

Figure 5.2: The nominal and standardized values of the Wilcoxon test statistic for the Wilcoxon Signed Rank test along with sample size and standard error.

As can be seen in Figure 5.1 and Figure 5.2 the null hypothesis that the median of Z_i equals 0 is rejected at the significance level of .999 and it can therefore be concluded that the median for the current and new risk scores are different.

Viewing the column charts for the respective risk scores in Figure 5.3, the conclusion can further be made that the distributions of risk scores are different. Viewing Figure 5.3, it is clear that no risk ratings from the current risk assessment tool have a rating of 1 or 5. A possible explanation for this is that the expert performing the risk assessment makes necessary adjustments to the flowchart before giving a risk score of 5, a very high risk. On the other hand, the expert always determines there to be some level of risk of not achieving the performance criteria and therefore never gives out a risk score of 1.

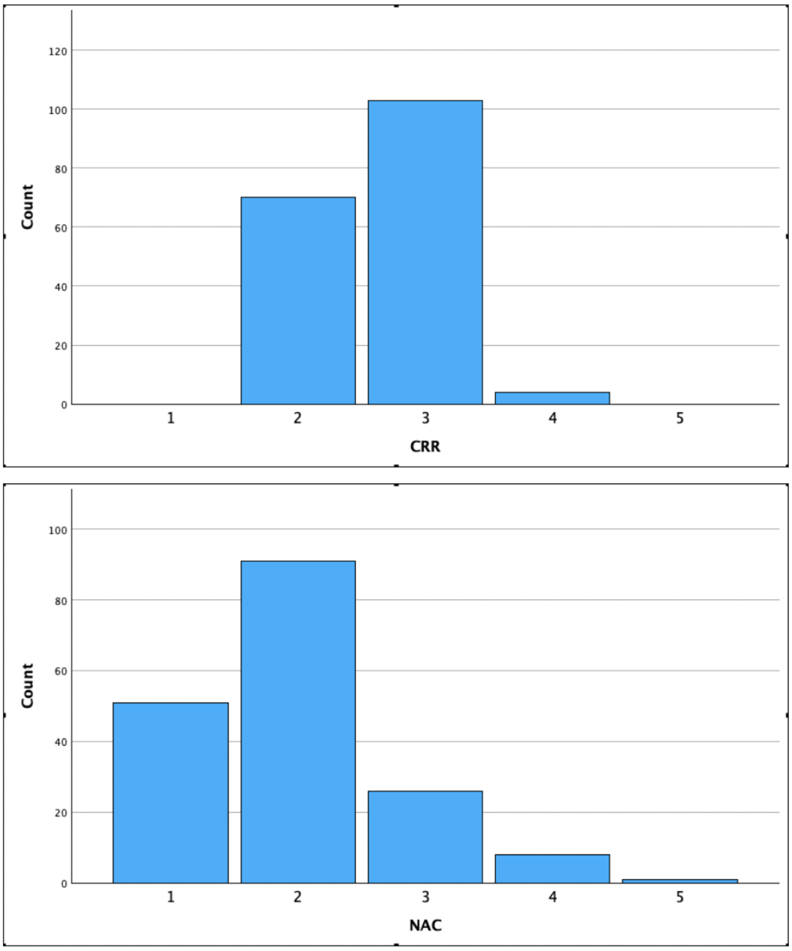

Figure 5.3: The top bar chart is the new risk rating scaled in the current format. The bottom bar chart is the current risk rating. Values are displayed on the x-axis and counts on the y-axis.

One can also look at whether or not the risk assessment tools recommend the system to be sold. Using guidelines for Stat. C&S, two new variables are defined:

$$NRec = \begin{cases} 1 & \text{if } 0.4 \leq NRR \leq 1 \\ 0 & \text{if } 0 \leq NRR < 0.4 \end{cases} \tag{5.5}$$

$$CRec = \begin{cases} 1 & \text{if } 1 \leq CRR \leq 3 \\ 0 & \text{if } 4 \leq NRR \leq 5 \end{cases} \tag{5.6}$$

A value of 1 refers to recommending a sold solution to proceed with implementation, and 0 to not recommend this. Since both $NRec$ and $CRec$ are dichotomous variables based on paired data, an appropriate test to use is McNemar's test. Referring to Chapter 4, Section 4.4.2, the 2 x 2 contingency table for $NRec$ and $CRec$ is given by Table 5.2.

Table 5.2: The 2 x 2 contingency table for $NRec$ and $CRec$.

	$Nrec = 1$	$Nrec = 0$	Row Total
$Crec = 1$	164	9	174
$Crec = 0$	4	0	3
Column Total	168	9	177

Using the notation of n_{ij} for the frequency counts of different outcomes, where i refers to the row number in the table and j to the column number, $i, j = 1, 2$, it is clear that $n_{21} + n_{12} = 13 < 25$, and therefore the exact McNemar's test is used to obtain the p-value. The hypotheses are formulated as:

$$H_0 : P_{12} = P_{21} \tag{5.7}$$

$$H_1 : P_{12} \neq P_{21} \tag{5.8}$$

In accordance to Equation 4.26, we calculate the exact p-value as:

$$
\begin{aligned}
P_{exact} &= 2\sum_{i=9}^{13}\begin{pmatrix}13\\ i\end{pmatrix}0.5^{i}(1-0.5)^{13-i} \\
&= 2\times 0.13342285 \approx 0.267
\end{aligned}
\tag{5.9}
$$

Using Equation 4.27, the mid p-value is now given by:

$$
\begin{aligned}
&= P_{exact} - \begin{pmatrix}13\\ 9\end{pmatrix}0.5^{9}(1-0.5)^{13-9} \\
&\approx 0.267 - 0.087 = 0.18
\end{aligned}
\tag{5.10}
$$

We see that the exact and mid p-values both are higher than 0.05 and therefore we cannot reject the null hypothesis at a significance level of 0.05. Running the test in SPSS provides the same output (the mid p-value is given for Figure 5.4 and the exact p-value for Figure 5.5):

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The distributions of different values across New ready to guarantee and Old ready to guarantee are equally likely.	Related-Samples McNemar Change Test	.180[c]	Retain the null hypothesis.

a. The significance level is ,050.
b. Asymptotic significance is displayed.
c. Exact significance is displayed for this test.

Figure 5.4: Test summary for the performed McNemar's test for related samples.

Related-Samples McNemar Change Test Summary

Total N	177
Test Statistic	1.231[a]
Degree Of Freedom	1
Asymptotic Sig.(2-sided test)	.267
Exact Sig.(2-sided test)	.267

a. The exact p-value is computed based on the binomial distribution because there are 25 or fewer records.

Figure 5.5: Test summary for the performed McNemar's test for related samples.

The figures 5.4 and 5.5 further confirm that the null hypothesis cannot be rejected and we can conclude that the distributions of $CRec$ and $NRec$ are statistically equivalent to a significant level.

5.3 Regression analysis

In this section, the results from multiple statistical regression analyses will be presented. As initially mentioned in Chapter 2, the regression analysis aims to analyze which relevant variables have an impact on the deviation between the risk scores from the new method and the previous method. The first step is to perform a multiple linear regression analysis with the risk score from the new risk method as the response variable. This aims to explain the impact of the included explanatory variables on the risk, which will provide useful insights for triangulation with interviews about which aspects of the risk method bring more risk to the system.

5.3.1 Initial multiple linear regression model

The discrete variables (i.e. the number of different crushers, conventional screens, and CRR) presented in Table 5.1 do not violate any assumptions of the linear model described in (4.1.2). Thus, these variables can be treated as continuous (Fox, 2024). For the binary variables, we define dummy regressors $D_{i,j}, i = 1, \ldots, 177, j = CI, PCI, PXI, PSDI, OUI, SI$ as the following:

$$D_{i,j} = \begin{cases} 1 & \text{if the response for variable } j \text{ is "Yes"} \\ 0 & \text{if the response for variable } j \text{ is "No"} \end{cases} \tag{5.11}$$

We now propose an initial linear regression model as:

$$\begin{gathered} NRR_i = \alpha_i + \beta_{CI} D_{i,CI} + \beta_{PCI} D_{i,PCI} + \beta_{PXI} D_{i,PXI} + \beta_{PSDI} D_{i,PSDI} \\ + \beta_{OUI} D_{i,OUI} + \beta_{SI} D_{i,SI} + \beta_{CJ} CJ_i + \beta_{CH} CH_i + \beta_{CS} CS_i \\ + \beta_{SCR} SCR_i + \varepsilon_i, i = 1, \ldots, 177 \end{gathered} \tag{5.12}$$

Submitting a model containing all the above-mentioned variables, denoted in (5.12), returns Table 5.3 and Table 5.4.

Table 5.3: Model summary for the full linear regression model containing all proposed variables.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	.834	.082		10.127	<.001		
	PCI	-.084	.079	-.168	-1.058	.292	.187	5.361
	PXI	-.011	.040	-.033	-.273	.785	.323	3.094
	PSDI	-.109	.043	-.240	-2.513	.013	.513	1.948
	OUI	-.044	.050	-.102	-.867	.387	.340	2.944
	SI	.043	.037	.092	1.161	.247	.749	1.335
	CJ	-.033	.025	-.103	-1.332	.185	.781	1.281
	CH	.055	.018	.251	3.073	.002	.702	1.425
	CS	.036	.039	.070	.934	.351	.836	1.196
	CV	.067	.048	.103	1.378	.170	.841	1.189
	SCR	-.040	.020	-.215	-2.047	.042	.428	2.338
	CI	-.132	.079	-.264	-1.663	.098	.186	5.364

a. Dependent Variable: NRR

Table 5.4: The initial full model's fitted coefficients and their standard error, t-values, significance levels, and VIF values.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.493[a]	.243	.191	.150

a. Predictors: (Constant), CI, CH, PSDI, OUI, CJ, CS, CV, SI, SCR, PXI, PCI

As is clear from the column labeled "Sig." in Table 5.3, many of the significance levels for the t-tests of the coefficients are larger than 0.05 and are therefore deemed to be non-significant in the model. A high p-value, or significance level as denoted in Table 5.3, indicates strong evidence for the null hypothesis that the slope coefficient is zero. Thus, a high p-value would imply that the slope coefficient is not significantly different from zero. Viewing the VIF values, we can also see some values relatively far from 1, implying collinearity. See subsection 4.4.4 for a definition of VIF. While multiple high VIF-values and high significance values are a clear indication that the model is overfitted, one can also see from Table 5.4 that the R^2 is 0.243. This means that 24.3% of the variance of the dependent variable NRR is explained by the model, which is not a particularly great result, especially for the model containing all variables. However, since the approach of this Master's thesis is to find some interesting connections from the comparisons in an environment where noise and external factors are at play, a model with a relatively low R^2 can still provide useful insights in this context. We will however avoid drawing any insights from this full model because of the significance and attempt to derive a model with fewer variables and better fit instead.

To arrive at a model whose coefficients have a high significance in terms of the t-test, stepwise regression is used (see Chapter 4.5 for a description of stepwise regression). Stepwise regression is useful in this context because it is a simple algorithmic approach to selecting significant variables in the model. SPSS provides the output in Table 5.5 and Table 5.6.

Table 5.5: The four stepwise linear regression models' fitted coefficients and their standard error, t-values, significance levels, and VIF values.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.234[a]	.055	.049	.16261
2	.324[b]	.105	.095	.15870
3	.394[c]	.155	.140	.15464
4	.453[d]	.205	.186	.15044

a. Predictors: (Constant), CI
b. Predictors: (Constant), CI, CH
c. Predictors: (Constant), CI, CH, SCR
d. Predictors: (Constant), CI, CH, SCR, PSDI

Table 5.6: The four stepwise linear regression models' fitted coefficients and their standard error, t-values, significance levels, and VIF values.

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	VIF
1	(Constant)	.812	.035		23.416	<.001		
	CI	-.117	.037	-.234	-3.152	.002	1.000	1.000
2	(Constant)	.763	.037		20.385	<.001		
	CI	-.116	.036	-.232	-3.203	.002	1.000	1.000
	CH	.049	.016	.224	3.088	.002	1.000	1.000
3	(Constant)	.765	.036		20.977	<.001		
	CI	-.100	.036	-.200	-2.799	.006	.979	1.021
	CH	.059	.016	.268	3.716	<.001	.963	1.038
	SCR	-.043	.014	-.231	-3.170	.002	.945	1.058
4	(Constant)	.779	.036		21.800	<.001		
	CI	-.089	.035	-.178	-2.556	.011	.971	1.030
	CH	.056	.015	.253	3.606	<.001	.960	1.042
	SCR	-.050	.013	-.266	-3.714	<.001	.923	1.083
	PSDI	-.103	.032	-.228	-3.252	.001	.963	1.038

a. Dependent Variable: NRR

As presented in Table 5.5, the stepwise regression generated four models, with the final fourth model including the variables $CI, PSDI, CH$, and SCR. As is clear from the p-values for the t-test of the coefficients presented in Table 5.6, they are all below the significance level of 0.05 and are therefore deemed significantly non-zero and of explanatory value. Viewing the VIF values, they are all above 0.9, which indicates low collinearity. For R^2, the value is 0.205 as shown in Table 5.5. While this is not high, the abovementioned viewpoint for the initial full model still applies. As expected, the adjusted R^2, which accounts for the number of variables included in the model, is closer to the adjusted R^2 of the initial full model. The stepwise regression model can be expressed as:

$$NRR_i = 0.779 - 0.089 \times D_{CI} - 0.103 \times D_{PSDI} + 0.056 \times CH - 0.05 \times SCR \tag{5.13}$$

As one can see, we have a negative effect on the risk score from systems where the performance criteria capacity and PSD are included. This is logical since the inclusion of these criteria increases the difficulty of clearing the performance guarantee and therefore increases the risk. The model in (5.13) also reveals that the risk score is increased when including more CH crushers in the system. This may seem like an unusual result at first, as one may believe that a larger system of crushers should provide a higher complexity and therefore more risk of it working as specified in the performance guarantees. There are reasons for this result that will be analyzed further in Chapter 6. The model also reveals a negative relationship between the number of screens in the system and the risk score. A negative effect on the risk score from screens makes sense from a risk perspective for rock crushing systems and is something that will further be confirmed in Chapter 6.

5.3.2 Regression of deviation between current and new tool

In this section, regression models aiming to model the deviation between the current and new risk assessment tools will be proposed and fitted. The aim is to obtain a model that reveals which broad aspects of the new tool lead to deviations in terms of the measured risk. We use the definition of Δ_i in (5.2) and define a variable $ABSDIFF_i$:

$$ABSDIFF_i = |\Delta_i| = |NAC_i - CRR_i| \tag{5.14}$$

Since $ABSDIFF_i$ is an ordered discrete variable with possible values

$0, 1, 2, 3, 4, 5$, we can fit an ordered logistic regression model presented in (4.10). This is a full model containing all variables that are found in 5.1.

Parameter Estimates								
							95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Lower Bound	Upper Bound
Threshold	[ABSDIFF = 0]	-5.945	4.182	2.021	1	.155	-14.142	2.252
	[ABSDIFF = 1]	-3.426	4.172	.674	1	.412	-11.603	4.752
	[ABSDIFF = 2]	.711	4.142	.029	1	.864	-7.406	8.828
Location	[PCI=0]	-.374	1.021	.134	1	.714	-2.376	1.627
	[PCI=1]	0[a]	.	.	0	.	.	.
	[PX=0]	.080	.541	.022	1	.883	-.981	1.141
	[PX=1]	0[a]	.	.	0	.	.	.
	[PSD=0]	.537	.590	.828	1	.363	-.619	1.693
	[PSD=1]	0[a]	.	.	0	.	.	.
	[OUI=0]	-1.216	.736	2.731	1	.098	-2.659	.226
	[OUI=1]	0[a]	.	.	0	.	.	.
	[SI=0]	.555	.528	1.106	1	.293	-.479	1.589
	[SI=1]	0[a]	.	.	0	.	.	.
	[CJ=0]	-.435	1.094	.158	1	.691	-2.578	1.709
	[CJ=1]	-.264	1.085	.059	1	.808	-2.389	1.862
	[CJ=2]	0[a]	.	.	0	.	.	.
	[CH=0]	-2.139	1.324	2.612	1	.106	-4.734	.455
	[CH=1]	-1.611	1.298	1.540	1	.215	-4.156	.933
	[CH=2]	-1.374	1.351	1.035	1	.309	-4.022	1.274
	[CH=3]	0[a]	.	.	0	.	.	.
	[CS=0]	-2.135	2.507	.725	1	.394	-7.048	2.778
	[CS=1]	-2.502	2.589	.934	1	.334	-7.577	2.573
	[CS=2]	0[a]	.	.	0	.	.	.
	[CV=0]	.428	2.041	.044	1	.834	-3.572	4.428
	[CV=1]	.908	2.104	.186	1	.666	-3.216	5.032
	[CV=2]	0[a]	.	.	0	.	.	.
	[SCR=0]	-.274	1.350	.041	1	.839	-2.920	2.372
	[SCR=1]	-1.960	1.231	2.536	1	.111	-4.373	.452
	[SCR=2]	-1.098	1.212	.820	1	.365	-3.473	1.278
	[SCR=3]	-22.905	.000	.	1	.	-22.905	-22.905
	[SCR=4]	0[a]	.	.	0	.	.	.
	[CI=0]	.342	1.041	.108	1	.742	-1.699	2.383
	[CI=1]	0[a]	.	.	0	.	.	.

Figure 5.6: Parameter estimates for the ordinal logistic regression model as well as Wald's statistic values, significance level, and confidence intervals.

As shown in Figure 5.6, no Wald's test for individual slope coefficients β (see Section 4.4.5) reaches a significance level below 0.05. Because of limitations in SPSS, an automatic model selection method such as the stepwise method is not available in the case of ordered logistic regression. Instead, manual trial and error is performed. After a multitude of iterations, beginning with single-variable models, a subset of this model with significant estimates and significance for the dependent variable could not be identified. This result may be due to several factors, including an insufficient sample size for robust regression analysis and the possibility that the chosen variables do not significantly affect the deviation. A further critical discussion on the validity and reliability of the results will be held in Chapter 7.

6 Interview Findings

As mentioned in Chapter 2, the second phase is the qualitative evaluation of the risk assessment tool by conducting interviews. A more in-depth discussion of the interview methodology can be found in Section 2.2.2. This chapter will present the interview findings regarding questions 2 (*Does the tool include the right scope of variables affecting the risk?*) and 3 (*Does it have an appropriate model to compute the risk from the variables?*) presented in Section 1.3. Findings related to question 2 are found in 6.1 and findings related to question 3 are found in 6.2.

The questionnaire that was utilized can be found in Appendix A.1. The questions were constructed to provide a more in-depth discussion of the findings from Phase 1 and discuss the risk assessment tool. In Table 6.1, an introduction to the interviewees is found along with the title and role. The selection was made in collaboration with Stat. C&S and the interviewees work within SRP's organization in different functions.

Table 6.1: A presentation of the interviewees' respective titles and departments of origin within Stat. C&S.

Interviewees	Title	Department
Interviewee 1	Sales Application Engineer	Sales
Interviewee 2	Sales support Manager, Africa & LAM	Crushing Solutions
Interviewee 3	Regional Sales Support Manager	Sales
Interviewee 4	Project- & Lifecycle Manager - Central Europe	Sales
Interviewee 5	Manager Plant Solutions	Sales
Interviewee 6	Process Optimization Expert	PIE
Interviewee 7	Group Interview with Sales Support Managers India Pacific	Crushing Solutions
Interviewee 8	Performance Optimization Expert	PIE

6.1 Scope of variables

In total, the current iteration of the risk assessment tool uses over 60 variables given by PlantDesigner to assess the risk of the system. One goal of conducting the interviews is to determine if the current version of the risk assessment tool uses the correct variables when calculating the risk score or risk probability which will be used interchangeably in this and the subsequent chapter. These terms can refer to one of three things: the final risk score derived from multiple crushers and screens, the output risk score for an individual machine within a crushing and screening plant, or the risk scores of machine-specific variables in the tool (for example, the risk score of a certain value of CSS) that affect the risk score of the individual machines calculated by the tool. The variables in this case affect the risk score from the new risk assessment tool and are not connected to the variables given in Table 5.1. This section is divided into crushers, screens, and common variables that affect both crushers and screens. The latter includes, for example, material characteristics such as the percentage of clay and moisture in the feed material.

In the new risk assessment tool, the performance criteria used in the performance guarantee affect which variables influence the risk scores of the individual machines. For example, a screen would not be affected by the existence of a PX criterion. Likewise, a crusher does not have as large of an impact on an oversize/undersize criterion as a screen. This is due to the former criterion being only affected by the performance of the crusher and the latter criterion only being affected by the performance of the screen. This section will not go through all variables currently used in the risk assessment tool and instead comment on the adjustments mentioned in the interviews.

According to Interviewee 4, the most crucial aspect to consider when assessing the risk of not meeting the performance criteria for a crushing and screening plant is the number of adjustments possible within the plant and the margins available in settings for the machine. Deviations in usage and material can occur in a real-life crushing and screening plant, and if more adjustments can be made, these deviations can more easily be accounted for. Because of this, more possible adjustments for a machine should imply a lower risk score for that machine. However, it is important to consider that fewer adjustments can be made to a screen than to a crusher, which implies a higher risk. Additionally, more factors can go wrong in the screening process, making it a riskier machine for the system (Interviewee 8). This is consistent with the findings in Table 5.6 in Chapter 5, where adding a screen decreases the risk score while adding a crusher increases the risk score due to a crusher having more possible adjustments in comparison with a screen, and therefore contributes with a lower risk overall.

6.1.1 Crushers

This section is divided into the three types of crushers sold at Stat. C&S and will present the interview findings regarding the scope of variables that affect each of these crushers.

CH and CS crushers

For the CH and CS crushers, the ECC and CSS are most likely to be adjusted rather than the layout of the plant or the chamber of the crusher. Due to this, the margin for these settings to their maximum and minimum theoretical values is one of the most critical variables to consider for these types of crushers (Interviewee 4). According to Interviewee 3, this depends on what is specified in the performance guarantee. As the

CSS affects the output PSD curve when changed, Interviewee 3 would rather not change this setting if other adjustments can be made in place, especially by changing the CSS to a higher value and opening up the crusher to reach a higher capacity.

As the risk score currently depends on both the margins to the minimum and maximum value of the CSS, Interviewee 3 believes that the margin between the actual value of the CSS and the maximum CSS is not as important as the margin between the actual ECC setting and the maximum ECC setting. The argument is that CSS in reality is not something that is adjusted to increase capacity unless it is necessary. While opening up the crusher increases capacity, it has the opposite effect on the reduction ratio and PSD, where the reduction ratio is lowered and it will be harder to obtain the desired PX out of the crusher. In the current tool, the margins for these settings have equal importance, an aspect that Interviewee 3 would like to change with the risk assessment tool.

Interviewees 1, 2, and 6 believe that all the variables included in assessing the risk of the CH and CS crushers in the risk assessment tool contribute to the risk in some form. However, the marginal impact that some of the variables have on the risk score can be too high. One example would be the risk of having too little finished material in a CH or CS crusher located within a closed circuit. A smaller percentage of finished material in a closed circuit is reflected by a low risk score in the new risk assessment tool. The risk score of this variable is currently too low, as a smaller percentage of finished material does not necessarily mean the performance will be much lower than anticipated outside of extreme values. The reason for this is that the machines are designed to be able to handle higher percentages of unfinished material (Interviewee 1). Interviewee 2 says that this aspect is more of a "rule of thumb", where it is inefficient and costly to have too much unfinished material in a closed circuit but that it doesn't necessarily mean that the performance criteria wouldn't be met.

None of the interviewees believe that any new variables need to be added to the risk assessment tool for the CH and CS crushers as the currently included risk-dependent variables capture the risk well.

CJ Crushers

The most important aspect that will affect a CJ crusher's performance and ensure its consistency with the values from PlantDesigner is not the settings available for the machine. Rather, it is the ability to ensure that the crusher is fed correctly. Ideally, the feed composition should consist of a mix of small and large rocks to fill the gaps within the crusher (Interviewee 6). However, this aspect is hard to quantify and depends on how the machine is fed. A CJ crusher is almost exclusively utilized in the primary crushing stage and is mostly fed with blasted rocks. For example, an excavator can be used for feeding the crusher, and if that is the case, the performance of the crusher can be dependent on the operator of said excavator. PlantDesigner assumes that the machines are fed optimally in its calculations. Because of this, there can be variations between the simulation and the real-life case. This should be accounted for in the risk assessment tool, which it currently isn't (Interviewee 6).

This could be accounted for through a guideline for how the machine should be fed or by adding another variable. An example of an added variable could be if there are certain feeder arrangements, where some arrangements affect the risk score of the machine (Interviewees 4 and 6).

The second aspect that must be accounted for in the risk assessment tool is the percentage of fines in the feed material (rocks between 0-5 mm in size) (Interviewees 1, 2, 3, and 4). If the percentage of fine material is too high, it can cause the material to compress together and decrease the available capacity of the crusher. In turn, this can also impact the crushing capabilities of the crusher. Therefore, fine material in the feed should increase the risk of not meeting the performance guarantee (Interviewee 3). This can, for example, be accounted for by using the percentage of fine material in the feed into the crusher or state in the tool whether there is an existing screen that separates the fine material from the material to be crushed somewhere before the CJ crusher in the crushing and screening process (Interviewees 2 and 6).

None of the interviewees believe that any of the variables currently included in the new risk assessment tool should be excluded when assessing the risk.

CV crushers

Similar to the case of the CJ crushers, the risk assessment tool should include a variable that describes the percentage of fine material in the feed for a CV crusher. A high percentage of fine material could cause issues within the crusher. As mentioned in 3.1.3, this type of crusher relies on interparticle crushing to break the stone and the collisions between fine material and normal size feed material will not cause as much breakage as if the feed material only consists of rocks of a similar size. Because of this, a large percentage of fine material can cause issues with the size reduction of material and shape (Interviewees 6 and 7).

There should also be an inclusion of a variable that provides the top size of the feed material into the crusher (Interviewee 6) and a variable that describes the flakiness of the feed material. Flaky material in the feed can cause the material to shift and change direction in the air which in turn affects the maximum capacity of the crusher (Interviewee 6).

Similar to the CJ crusher, none of the interviewees considered any of the utilized variables as redundant when assessing the risk.

6.1.2 Screens

For screens, some new aspects should be accounted for in the risk assessment tool to reflect the actual risk of the machines not working as specified by PlantDesigner.

The first aspect is the introduction of more options for the performance criterion degree of oversize/undersize. Currently, three options can be chosen in the risk assessment tool: low, medium, or high accuracy desired. The options should better reflect the actual requirements in the performance guarantees (which are, for example, expressed as 10/15 where 10 means 10% oversize allowed and 15 means 15% undersize allowed). Table 3.3 shows further examples of typical ranges within performance guarantees. In general, finer margins or lower numbers are harder to achieve, and this should be accounted for better in the risk assessment tool (Interviewees 5 and 6).

Interviewee 3 believes that the variable that describes the relationship between the feeder and screen capacity should be omitted from the risk assessment tool. The difference between them is not important, rather, it is important how the screen is fed (Interviewee 3). In a crusher and screening plant, it is rare to have a separate feeder for a screen in the first place. If one exists, the most important thing is that the feed material is correctly distributed across the screen to ensure that the feed depth is consistent in terms of the width across the deck. However, this aspect cannot be quantified as it is outside the scope of the performance guarantee and can only be checked when the plant is running (Interviewees 3 and 5).

When screens separate rocks with similarly small sizes, such as when one deck has a hole size of 4mm, it affects the screening accuracy. This should increase the risk because the performance guarantee is based on PlantDesigner, which does not consider the potential inaccuracy in its simulation. This is reflected in the risk assessment tool for separations under 2mm but should be extended to separation sizes under 5mm (Interviewee 2).

In the new risk assessment tool, if the screen is using wet screening, it increases the risk with the screen. Wet screening is when water is sprayed over the feed material or by adding water to the feed material. This is usually done to increase the capacity and improve the sizing efficiency of the screen (Nelson, 1965). Wet screening should not increase the risk of guaranteeing the screen, as wet screening increases the performance and thus reduces the risk rather than increasing it (Interviewees 1, 2, 3, 5, and 7). However, there are some aspects related to wet screening that are hard to reliably estimate. This is why it can in certain cases be correct to attribute increased risk to systems where wet screening is present (Interview 7).

Another important factor for screening accuracy is the time that material is on the screen. This ensures that the feed material has time for effective stratification and mostly depends on the screen's length. At every time unit, the material has some chance of falling through the holes of the screen, and if the screen is longer, the total chance of this happening is higher as the material has more time for stratification. Currently, this is partly accounted for by having the angle of the screen as a variable, Interviewee 5 would like the addition of the length of the screen as it better reflects the risk of not meeting a prespecified screening accuracy. This does not implicate that the angle of the screen should be omitted as a variable because the angle of the screen impacts other aspects such as the probability that a rock would fall through a screening deck at every vibration (Interviewees 5 and 8).

Interviewee 2 would like the addition of a variable that describes the margin to the maximum capacity of the screens, called carry-over capacity. If there are variations in the amount of material passing the screen and the screen goes over the maximum capacity it can handle, it could cause the screen to not vibrate correctly, reduce the screening accuracy, or make the screen deck break.

Lastly, a variable that describes the amount of flaky material should be added to the risk assessment tool, as flaky material is harder to get through the holes of the screen and thus causes more oversized material in the product. This aspect is partly accounted for by PlantDesigner which uses the percentage of flaky material in its calculations. However, the variable could still be included in the risk assessment tool to provide a more accurate risk assessment (Interviewee 2).

6.1.3 Common variables for crushers and screens

The moisture content of the feed material is a crucial factor affecting the performance of all crushers and screens. Very high moisture levels can cause failure in the entire process (Interview 1). However, high moisture levels do not affect wet screening processes, so the risk that high moisture levels bring to these is eliminated. The amount of moisture in the feed material affects the risk score given by the new risk assessment tool, but only moisture levels up to a given limit are considered. The crushers at Stat. C&S behave predictably at moisture levels within a certain range. There is no significantly increased likelihood of not reaching the performance criteria for feed material with a low moisture level. However, high moisture levels should therefore pose a higher risk to the system in comparison to what it currently does (Interviewee 4). However, aspects such as the amount of moisture in the feed material can be hard to reliably predict, as they can deviate from values given by PlantDesigner in real-life scenarios. This can be due to seasonal aspects or a variety of moisture in the blasted rock material (Interviewee 1). This is especially critical in a case where fine material exists within the feed material, as this makes the process more unpredictable due to the moisture sticking to finer particles in the feed. This causes issues with reaching a specified capacity or PX by impacting the flowability of the material through the crusher (Interviewee 3).

Moisture in the feed material is calculated as a percentage of the entire feed for the crushers and screens. For a screen, the same moisture levels are used for each screen deck. It is said that the "moisture follows the fine material", which implies that a lower separation screen deck will have a higher percentage of moisture in the material passing through the screen deck. As fine material combined with a percentage of moisture in the feed material affects the performance to a greater level, Interviewee 3 means that the risk associated with moisture in the feed material for screens should be calculated per screen deck instead of using the same percentage and impact for every deck of the screen (Interviewee 3).

This is even more important when there is a combination of clay and moisture in the feed material. This affects the flowability of the material and can lead to blockages of the screen holes as the clay adheres to the feed material or the machinery. This issue impacts all types of crushers and screens. Clay is currently considered to have a significant impact on the risk score of individual machines. However, clay alone is not problematic in a crushing and screening plant. Dry clay is easily crushable by the crushers sold at Stat. C&S and does not cause unpredictability in the machines' performance. Therefore, if there is no moisture in the feed material, there is no significant added risk that the system will not perform as specified with high percentages of clay (Interviewees 3, 6, and 7). In the risk assessment tool, the percentage of clay and moisture in the feed material are accounted for separately. Because of this, Interviewee 6 would like to see the inclusion of a combined factor that takes both clay and moisture levels of the feed material into account.

6.2 Risk Score Calculation Model

This section will go through the calculation model of the new risk assessment tool and verify if this model is appropriate when assessing the risk of a crushing and screening plant. Currently, different risk score calculation models are used based on what is being assessed. For a single crusher or screen, the risk scores of the variables are considered independent and multiplied to give the machine's individual risk score. For an entire crushing and screening plant, the final risk score is calculated by taking the arithmetic mean of the individual risk scores of all the crushers and screens in the plant.

One goal of the interviews was to find out if the variables can be treated as independent of each other when assessing the risk or if any important correlations are missing, and if the current method of calculating the final risk score is the best practice to assess the risk of a system of machines. This section will also discuss the models used for calculating variability in performance for the machines, in the cases where ranges of performance are used when calculating the individual risk score of a machine.

6.2.1 Individual risk scores

In the initial risk assessment conducted in Phase 1, it was observed that when variables affecting capacity reached their maximum values, the risk score for achieving the specified capacity in the performance guarantee remained high. This trend extended to other variables where ranges were employed to determine the final risk score. The current risk assessment tool's ranges are derived from testing data from Stat. C&S and other divisions of Sandvik SRP. Within the tool, these ranges are established based on the minimum and maximum values observed during testing. For example, suppose one test showed a maximum capacity of 1100 MTPH for a crusher and another showed 900 MTPH for the same crusher. In that case, the tool takes the difference between these values when assessing the variability of the crushers performance.

This is expressed as a percentage difference between the minimum value from testing and the mean value of all tests in the sample. Using the example provided above, this means the tool would assess a 10% variability in the crusher's performance, given that the mean capacity in the example is 1000 MTPH. So, if all variables used to calculate a predetermined capacity were at their maximum, the risk score reduction is currently only 10%. As such a crusher that would have no margins available for capacity would still return a risk score of 90%, which would imply that the risk of guaranteeing this crusher is next to none.

For the new risk assessment tool, all instances where ranges are used are treated this way. This is incorrect according to Interviewees 3 and 4, if there are no margins and no adjustments possible, the system should be attributed a very high risk of not meeting the performance criteria and a low risk score. Due to how the ranges are used, in about half of the cases the performance criterion would not be met, if the real-life performance falls in the bottom half of the range (Interviewees 3 and 4).

Regarding the method for calculating the individual risk scores of one crusher or screen, an initial hypothesis was formed regarding the current assumption of independence when calculating it. The initial hypothesis was that the different variables should be weighed instead of considered completely independent due to potential correlations between variables and that the addition of new variables to the risk assessment increases the assessed risk of the crushers and screens. One of the interviewees agreed with this hypothesis, as some variables can be viewed as more important than others, and that when adding variables the assessed risk score should not be decreased (Interviewee 6).

In contrast, Interviewee 4 means that all variables in the risk assessment tool have critical values for when it can cause complete failure and disrupt the crushing and screening process, due to this, the risk assessment tool needs the variables to have independence. Interviewee 4 says the current model is correct but could use adjustments for calculating the individual values. Interviewees 1, 2, and 3 agree that the current method used for calculating the individual risk scores is correct when assessing the risk of a machine because the variables can be considered mostly independent.

Regarding potential correlations and relationships between variables, as mentioned in 6.1.3, the real-life performance issues are more impacted when moisture and clay in the feed material are at higher values. Outside of this, when product and capacity are used in the performance guarantee, more adjustments can be made within the system for the CH and CS crushers, as the CSS can be adjusted to a lower value to produce finer products or the ECC can be increased to increase capacity of the crusher. Due to this, the assessed risk should be lower in comparison to when total capacity and PSD are guaranteed separately (Interviewee 1).

6.2.2 Calculating final risk score

The current method of calculating the final risk score for a crushing and screen plant is by taking the arithmetic average of the individual risk scores for the crushers and screens included in the risk assessment. Shape as performance criteria, see 3.3.3 for a description, is treated as an individual risk score like the crushers and screens due to the theoretical shapes being calculated in an external tool. A shape can be guaranteed for separate product ranges, for example, an SI of 20 for products sized 10-14 mm and an SI of 25 for products sized 14-20 mm. If so, each product range is treated as an individual machine when calculating the final risk score.

Taking the arithmetic average of the individual risk scores has pros and cons. When you have more machines included in a guarantee, more adjustments can be made if one machine is not working as specified in the simulation of the system. However, there is also a higher chance of a machine not working in the system due to more machines being included and hence a more complex system with more points of potential errors (Interviewees 3 and 5). As described in the initial part of this Chapter, this seems to be more the case when including more screens. Although this can depend on which system is guaranteed, if only one crusher is responsible for the final product, the risk score of the crusher should be taken into special consideration as it affects the final product and ultimately decides if the performance criteria is met (Interviewee 4). An idea to combat this would be to add a larger weight to the total risk score for the crusher and screens used in secondary or tertiary crushing stages compared to machines in the primary crushing stage (Interviewee 2).

Calculating the shape risk in the current manner may or may not be accurate. It depends on how well the external tool calculates the theoretical value compared to the real-life value. These calculations can then be included in the same risk assessments as the machines and treated as individual machines, or they can be managed separately by using an external sheet that handles the risk associated with the external tool. (Interviewee 6).

7 Discussion

This chapter will discuss the findings, provide key recommendations, and present an implementation plan. If the proposed recommendations are implemented in Stat. C&S, the tool is deemed to be validated in terms of assessing the risk of conducting performance guarantees. The implementation plan will cover two things. Firstly, it will cover how the tool should be implemented and used when conducting performance guarantees. Secondly, it will go over how they should establish best practices for conducting performance guarantees in the sales areas of Stat. C&S.

7.1 Key recommendations

This section will go through the key recommendations derived from the results in Chapter 5 and 6. The recommendations consist of changes to the new risk tool that are both needed for sufficient accuracy and are feasible to implement for Stat. C&S. When implementing any of the proposed recommendations to the risk tool, it is worth noting that some risk values might need to be adjusted to balance the risk score.

7.1.1 Adjustment of variable scope

A key finding from the interviews and analysis was that there are certain gaps in the variable scope of the new risk assessment tool. This concerns screens, CV, and CJ crushers, where there are variables that should be omitted and added to the tool to better reflect the real-life performance-related risks not accounted for by PlantDesigner.

The exact probability values associated with these variables are unclear and need to be further researched. It is therefore important that the variable scope will continue to be developed based on data and findings from real-life situations. As the current scope of variables is based on qualitative data there are still aspects related to the variable scope that may be missing in the new tool. Because of this, the risk assessment tool should not be static in its development after the evaluation and validation, it should be dynamically updated based on emerging risks or operational changes to the machine's usage. It can thus be essential to regularly review and update the variable scope to ensure that it is relevant based on the situation and effective at assessing the risk of not reaching the performance guarantee.

Adding variable for percentage of fine material

As mentioned in 6.1.1, a large percentage of fine material in the feed material of a CJ or CV crusher can affect the performance of the crushing and screening plant. This factor is currently not accounted for in PlantDesigner or the risk assessment tool. As a large percentage of fine material increases the risk with the system, this variable should be included to account for this risk. Adding this variable in PlantDesigner would be relatively straightforward since the percentage of fine material is easily accessed from the PSD curves available in PlantDesigner. However, the risk impact of certain percentages of fine material requires an internal discussion among departments in Stat. C&S.

Adjusting the scope of variables affecting screens

From the results in 6.1.2, it can be observed that there are some potential issues with the scope of variables affecting the screens. Multiple variables should be added to the risk assessment tool, including one that more accurately reflects the oversize/undersize criterion in the performance guarantee, another for carry-over capacity, and a variable for the percentage of flaky material, while the variable that describes the feeder load for screens should be omitted. Additionally, the length of the screen should be included to better describe the time the rocks stay on the screen.

The analysis in 5.3.1 shows that including screens in the risk assessment has a negative slope coefficient in the stepwise linear regression model. Adjusting the variable scope will likely affect this coefficient. Adding new variables to the risk assessment tool always decreases the average risk score due to the calculation model for the individual machines' risk score. A way to counteract this could be to have a more thorough discussion regarding the risk scores of the current variables that affect screens to see whether the current risk scores are consistent with real-life risks. If the risk probabilities are adjusted it can help to increase or maintain the negative slope coefficient marginally when new variables are included instead of lowering it even more.

Adjusting variable scope of CV crushers

For CV crushers two additional variables should be added to the risk assessment tool. This is the top size of the feed and the flakiness of the feed material as these can impact the performance. Using the same argument as earlier, one way to do this would be to adjust the risk probabilities associated with the current scope of variables to counteract the introduction of new variables.

7.1.2 Adjust the impact on risk for certain variables

From the interview results in Chapter 6, it is clear that the impact on the risk score from some included variables should be adjusted to obtain a more accurate risk score. As observed in the Wilcoxon signed rank test conducted in 5.2, the risk scores differ substantially between the current and new risk tools. To improve consistency with the current tool, one approach is to adjust the risk impact of variables that experts have deemed to be inaccurate in relation to real-life performance-related risk. The three main adjustments identified in Chapter 6 are for adjustments on the impact of CSS, finished material, and wet screening.

Adjustment of risk impact from CSS

As mentioned in 6.1.1, the margin between the actual CSS and the maximum CSS should not have the same risk impact as the equivalent margin for the ECC. In this context, impact on risk refers to the marginal difference in the risk score when changing the CSS or ECC one setting larger or smaller. Since there seems to be an agreement on having a linear effect on the risk, (e.g. the same change in risk score per change in setting), one quick adjustment would be to make the marginal changes in risk smaller for the CSS than the ECC. In terms of implementation, this is a quick fix and would not require more than a couple of minutes to actualize.

Adjustment of risk impact from finished material

As stated in the 6.1.1, the risk impact from having a certain share of unfinished material in a closed circuit configuration has been deemed too negative. While it is a sign of crushing inefficiency, it does not necessarily have a large impact in terms of increased risk. There are multiple ways of implementing this adjustment. One possible action is to increase the probability output for this variable. By doing this, the risk is not as negatively affected, while the user still notices the variable and makes sure that the degree of unfinished material is not too low. Another way of accomplishing a similar effect is to remove the variable completely and instead implement this as a pop-up warning in PlantDesigner. This will ensure that the risk tool strictly covers the risks of not meeting the specified performance criteria, under the assumption that the degree of unfinished material bears a negligible amount of that risk.

Adjustment of risk impact from performing wet screening

Currently, performing wet screening returns a risk score of 80%, in contrast to dry screening returning 100%. The common opinion however has been that wet screening, when performed correctly, leads to better screening accuracy than dry screening and should therefore entail lower risk. The counterargument to this, and the reason why it returns an 80% probability, has been that there still is limited knowledge of wet screening and uncertainty around best practices. While still accounting for this uncertainty, the positive aspects of using wet screening could balance the risk score to potentially a 100% output for both screening methods which would mean that this variable could be omitted from the risk assessment tool. As uncertainty around wet screening decreases, the future state would be that dry screening returns the lower number out of the two.

Implementing these changes in the new risk tool would not take long. An additional action to take for Stat. C&S could be to establish an internal training course in wet screening for the salespeople to ensure that everyone applies it most effectively. This training course might need to be updated as research and internal knowledge solidifies.

7.1.3 Adjusting probabilities based on test data

As described in 6.2.1, the probability output from two variables, capacity, and probability to reach CSS, are based on ranges of test data. As Interviewees 3 and 4 state, treating the relative range of test data as the probability output is not correct. To illustrate this, one can use the example where a CH crusher has 100% load, and ECC and CSS on the maximum settings. Furthermore, let us assume the guaranteed capacity is 100 MTPH. The probability of reaching the capacity given in PlantDesigner can then, for example, return 87% as a risk score. This number comes from the range of test data in this example being $100 \pm 100 \times 0.13 = [87, 113]$.

While variability and risk are closely connected, treating probabilities in this way does not align well with theoretical principles. Instead, the probability distribution of the capacity data needs further analysis. If the data appears to follow a normal distribution, it will be symmetrical around the mean (100 in this example). This means that 50% of the time, the outcome will be below the average value shown in PlantDesigner, which equates to not meeting the performance criteria. Therefore, the risk probability should be much lower than the example value of 87%. Assuming the distribution is normal or symmetrical around the mean, the risk probability could be around 50%. If this is not the case, the resulting risk probability could be even lower.

While implementing this change requires some statistical knowledge, it could be relatively straightforward. Furthermore, considering other variables' values are in some cases based on intuition, the data analysis does not need to be particularly rigorous. The use of test data such as in these cases can be considered important in order to reach accurate estimates of probabilities. Therefore, a desired future state is one where a larger share of the included variables' values are based on test data. It needs to be mentioned however that performing test runs and collecting data can be time-consuming, especially to achieve a significantly large sample size.

7.1.4 Alternative way of handling moisture

Moisture is a variable that has been frequently discussed during interviews. While the inclusion of the moisture variable certainly is correct, an alternative way of handling moisture might reflect the associated risk better than the current state. From Chapter 6, three potential corrections have been identified. These are to increase the risk for extreme levels of moisture, to include moisture levels per individual deck in screens, and to include a correlated risk between moisture and clay.

Increase risk for extreme moisture levels

As interviewee 1 mentioned in 6.1.3, higher moisture levels in the feed material can lead to the rock crushing system not working. Since the risk assessment tool only has choice alternatives for moisture levels up to a given percentage, the risk associated with higher levels is not currently captured. If the moisture is at X%, the variable could, for example, return a risk score of 30%. Implementing this would not be particularly time-consuming, serve as an extra caution, and better reflect the risks of operating in very moist conditions. Before implementation, there needs to be a thorough discussion among experts about what probability values to include for this factor.

Inclusion of moisture levels per individual deck

In 6.1.3 it is mentioned that moisture follows the finer material, which leads to lower screen decks having higher moisture levels than decks positioned higher. Currently, in the new tool, there is only one moisture value entered per entire screen. Since higher levels of moisture affect screening performance, entering a moisture value for every deck to accurately calculate the deck bearing the most risk would be advantageous. Moisture values per screen deck have recently been added in PlantDesigner, which allows for straightforward implementation into the risk tool.

Capture variable correlation with moisture

As described in 6.1.3, the interviewee wants a combined variable to capture the risk when higher levels of both clay and moisture are present. When there is a substantial amount of clay in moist conditions, the clay sticks to the rocks and prevents effective crushing and screening. To account for this in the risk tool, a new variable called "Moisture and clay combined" can be created and return a low probability only if moisture and clay have high enough values for stickiness to occur. Regarding which combinations of these values cause stickiness, there is an internal data sheet that provides a description of which values cause stickiness. As such, the implementation of this variable should be relatively easy.

7.1.5 Adjustment of total risk score calculation

In this section, potential adjustments to the total risk score calculation will be highlighted and critically discussed. The highlighted adjustments for the total risk score are a possible change from arithmetic mean to weighted average and an alternative way of including shape risk. It is worth noting that these are not recommendations but rather adjustments derived from Chapter 6 and 5 that are critically discussed.

Including weights for the average calculation

As mentioned in 6.2.2, the new risk assessment tool currently calculates the arithmetic mean to determine the final risk score after calculating the risk scores for individual machines. Additionally, as the number of machines in the system increases, more adjustments can be made to ensure that the system performs as specified in the guarantee. For example, consider a system with two crushers: a primary crusher and a secondary crusher. In comparison to a single crusher, where the output is directly affected by the initial feed, adjustments can be made to the primary crusher in the two-crusher system to improve the feed curve for the secondary crusher. While this logically reduces risk, it also introduces an additional point of potential error. Specifically, if the primary crusher in the two-crusher system cannot achieve the guaranteed capacity and the input to the secondary crusher is below the required criteria, the entire system slows down.

This matter of an increased amount of machines leading to a higher risk might be more correct in the case of the number of screens than the number of crushers. This is because screens are viewed as more risky and have fewer available adjustments as seen in 6.1. In the case of crushers, the decrease in risk from an increase in possible adjustments may outweigh the increase in risk from increased complexity and points of error. This may not be the case for screens.

The results in 5.13 show that there is some correlation between the amount of crushers in a system and an increase in risk score, and the opposite effect for screens. Assuming that the description above is true, this is an indication that taking an average for the final risk score is a representative measure of the real-life risk. However, the weighted average that Interviewee 2 proposes in 6.1, might reflect the risk more precisely than an arithmetic average where every machine has equal weight. Deciding on what weights to use, however, might be difficult since it depends on the configuration of the system and is likely to lack data support.

Alternative way of including shape risk

Another point covered in 6.2.2 was the current inclusion of individual shape products in the average of the new tool. One initial change that can be made is to calculate one single shape related risk. This would prevent inflation of the risk score if many low risk shape products are included. Instead, a total shape risk can be calculated by taking the average or the minimum of the separate shape products. It can then be included in the average for the risk score as one individual component instead of multiple.

As proposed in 6.2.2, one can also manage shape completely separately. While this may result in a more accurate total risk score, it leads to the current risk tool not representing the risk of the solution not meeting the performance guarantee in its entirety, since shape is one of the performance criteria.

7.2 Implementation of the Tool

The original purpose of developing the new risk assessment tool was to ultimately enable the writing and approval of the entire process of performance guarantees in the sales organization of Stat. C&S. To facilitate this, the tool first needed to be evaluated and validated, which is one of the purposes of this thesis project. The tool is deemed to be valid if the proposed changes above are accounted for. The second purpose of this thesis project was the effective implementation of the tool within the organization. This section will go through an initial implementation plan that ensures that the tool can be utilized effectively and does not compromise the high quality of the currently conducted risk assessments and performance guarantees.

When implementing a service or tool within an organization, an implementation framework should ideally be used either before or throughout the implementation effort (Moullin et al., 2020). Recommendations from Moullin et al. (2020) for approaching an implementation include selecting the appropriate framework, determining the factors that matter for implementation, deciding on an effective implementation strategy, specifying the outcome or goal for the implementation, and writing a proposal that should serve as a guideline for the implementation process.

7.2.1 EPIS Framework

A framework that can be used to examine and plan the implementation proposal is the Exploration, Preparation, Implementation, and Sustainment (EPIS) framework. This framework utilizes four well-defined phases that can aid in planning an implementation effort. The EPIS framework is most commonly used in healthcare practices but can be modified to fit into other research fields as well (Moullin et al., 2019).

The exploration phase involves engaging key stakeholders to gather input and support and exploring the potential fit of the risk assessment tool within the organization (Moullin et al., 2019). In the context of the risk assessment tool, this is the phase the tool is currently in. During this phase, questions related to how the tool should be implemented and used within Stat. C&S are gathered from stakeholders within Stat. C&S.

Interviewee 2 suggests that the tool in its current state could serve as a "sanity check," highlighting critical risk aspects when conducting performance guarantees. However, it needs to be continually updated to adjust the final risk scores to be more precise in its risk assessment. The performance guarantees and subsequent risk assessments conducted in the sales organization should also be written by a process expert designated for that purpose by the organization (Interviewee 2). One issue with this is that it can be time-consuming and impact the effectiveness of the employee's regular responsibilities.

For the preparation phase, the primary objective is to identify potential barriers and facilitators of the implementation. It is especially important to deliver training and technical assistance to employees (Moullin et al., 2019). There is currently a guide that explains all variables used in the new tool (Interviewee 8), but as the tool evolves and changes based on the recommendations in 7.1, this guide needs to be updated to enable easy use of the tool. The person conducting the risk assessment must understand why there is a risk with certain variables. Interviewee 4 mentions that it is important for the person conducting the risk assessments to understand the tool's limitations and maintain a continued dialogue with the customers to ensure that the machines are used most effectively.

The implementation phase involves putting the tool into practice and conducting performance guarantees and risk assessments with its assistance. The final phase, the sustainment phase, involves maintaining and institutionalizing the tool (Moullin et al., 2019). This includes continually updating the tool based on the variable scope and the underlying data obtained through testing. A good practice would also be to establish a system where instances of unmet performance guarantees are recorded and systematically analyzed to ensure the tool has optimal accuracy.

7.2.2 Proposed Implementation Plan

A good implementation plan should follow the four phases specified above. This section will go through each phase and propose key activities that can be made within each phase based on the observations and findings obtained throughout this thesis project.

In the exploration phase, the tool should be implemented in a pilot version to gather initial feedback and assess its effectiveness. The tool can, for example, initially be implemented in one sales area of Stat. C&S for this purpose. The focus of this pilot implementation would be to collect more data on how the tool performs, focusing on the accuracy and usability of the tool.

In the preparation phase, training materials for how to use the tool should be created, including guides that explain all variables used and construct a guideline that describes best practices for conducting performance guarantees. In addition, training sessions for employees to ensure they understand how to use the tool and conduct performance guarantees and risk assessments effectively can be arranged in this phase. The existing guide should also be updated based on further feedback and adjustments that have been made to the tool to ensure the existing guide remains relevant and accurate. A support system can also be established, to assist sales areas with questions when conducting performance guarantees and handling high-risk risk assessments.

During the implementation phase, the tool should be rolled out across the entire organization. Performance guarantees should then be conducted entirely by the sales areas of Stat. C&S with the assistance of the tool. It's important to continuously monitor progress to address any issues that may arise with the use of the tool. Additionally, ongoing support should be provided to the sales areas through employees experienced in conducting performance guarantees or from employees with extensive knowledge of risks associated with rock processing solutions if deemed necessary.

When the implementation reaches the sustainment phase, the tool should be regularly updated based on feedback and new data, ensuring that it evolves to meet changes within the norms of conducting performance guarantees. Additionally, it's important to normalize the use of the tool in Stat. C&S's standard operating procedure for conducting performance guarantees. Finally, implementing a practice of recording instances where performance guarantees are not met and systematically analyzing these cases will help improve the tool's accuracy and effectiveness.

7.3 Reliability and Validity

Before moving on to the summary and conclusion of this Master's thesis, it is important to make a brief comment on the reliability and validity of this report. Reliability refers to the consistency of a measure, meaning the extent to which similar results can be obtained by following the same steps of an investigation. Validity refers to the extent to which the concept that a researcher aims to investigate is actually measured (Heale and Twycross, 2015).

One aspect that affects the reliability of this thesis is the collection of data for the risk assessments. As described in 2.2.1, the collection of data for the quantitative analysis was performed systematically. However, some instructions in the new risk assessment tool could be left for interpretation. An available user manual for the tool prevented some of this, but there is still a possibility that two people doing the same risk assessment could get different risk scores. Furthermore, there is a risk of human error when manually entering large amounts of data such as in this case.

The conducted interviews followed a semi-structured format and touched on different areas of the interview guide found in Appendix A.1. As such, the interviews explored different topics based on the interviewees' perceptions of the most critical risks affecting a rock-crushing and screening plant. Therefore, if a similar interview guide were to be used for the same investigation, it is uncertain whether identical results would be obtained.

The interviewees were selected in collaboration with Stat. C&S and have extensive experience, both in using PlantDesigner and in conducting performance guarantees. However, the interviewees have varying roles and viewpoints which must be considered when analyzing the results. It is also not certain that a similar result would be obtained if different interviewee candidates were selected for this Master's thesis.

We found that drawing strong conclusions from the quantitative analysis was somewhat challenging due to limitations in the available data. With the different formats of the new and current risk scores, and especially the fact that no score of 1 or 5 of the current tool was present, comparison and further analysis of the deviation were difficult. Also, if more time was available, a larger set of variables could be recorded which could enable a model with higher explanatory power to be derived. On top of that, we believe the analysis would be stronger with more than the 177 collected data rows in this thesis. However, the time-intensive nature of the collection method and the availability of previously performed risk assessments unfortunately constrained the amount of data that could be collected.

8 Summary and Conclusion

The primary objective of this Master's thesis was to evaluate and validate a newly developed risk assessment tool for performance guarantees related to the sales of rock crushing solutions at Sandvik Rock Processing Solutions - Stationary Crushing and Screening (Stat. C&S). The research and data collection was conducted through quantitative statistical analysis and interviews, which culminated in establishing key recommendations and a proposed implementation plan.

The new risk assessment tool was compared against the current expert-based method. Statistical tests, including the Wilcoxon Signed Rank Test and McNemar's Test, revealed significant differences between the risk scores of the current and new risk assessment tools. This indicates that while the new tool provides a systematic approach, it requires adjustments to align more closely with the expert risk assessments. Also, linear and logistic regression was performed to analyze the risk score response from certain characteristics of the crushing systems.

The new risk assessment tool had a mostly correct variable scope but interviews with experts highlighted the need for small adjustments to this. For example, the inclusion of variables such as the percentage of fine material in the feed for CJ and CV crushers, and the carry-over capacity for screens, were deemed necessary.

The current model of computing individual risk scores by treating variables as independent was generally supported. However, the model for computing the final risk score by averaging individual scores was questioned in the interviews. It was suggested that a weighted approach might better reflect the risk of the system.

From this, several key recommendations could be established. Firstly, the addition of new variables into the risk assessment tool such as feed material characteristics and machine-specific factors, and the risk impact of several variables. Secondly, an adjustment of the calculation method when measuring variability in capacity and considering a weighted approach for calculating the final risk score to account for the criticality of different machines in the system. An implementation plan for transitioning the risk assessment process to the sales areas was also proposed. The plan includes training sales engineers on the new tool, establishing best practices, and continuously updating the tool based on feedback and real-life performance data.

The new risk assessment tool shows promise in providing a systematic and consistent approach to evaluating performance guarantees for rock-crushing solutions. However, to ensure its effectiveness and accuracy, several adjustments were deemed necessary based on the recommendations. By refining the variable scope and calculation models, and implementing a robust training and feedback mechanism, Stat. C&S can improve its risk assessment capabilities and more accurately reflect the risk associated with meeting specified performance criteria. The successful implementation of these recommendations and the risk assessment tool will not only improve the accuracy of risk assessments but also promote a more efficient and customer-centered sales process.

Bibliography

DATAtab (2024). *Wilcoxon Signed-Rank Test.* URL: https://datatab.net/tutorial/wilcoxon-test (visited on 05/22/2024).

Denscombe, Martyn (2010). *The Good Research Guide.* 4th. Open University Press. ISBN: 978-0-335-24139-2.

EN 933-3 (n.d.). *EN 933-3:2012 - Tests for geometrical properties of aggregates - Part 3: Determination of particle shape - Flakiness index.* URL: https://standards.iteh.ai/catalog/standards/cen/0ccc19d3-9861-4d02-b943-5f166e7cf6d6/en-933-3-2012 (visited on 02/20/2024).

EN 933-4 (n.d.). *EN 933-4:2008 - Tests for geometrical properties of aggregates - Part 4: Determination of particle shape - Shape index.* URL: https://standards.iteh.ai/catalog/standards/cen/86ed46c0-f6fe-4f0b-9381-f4309c159b0f/en-933-4-2008 (visited on 02/20/2024).

Evertsson, Magnus (May 2015). *Cone Crusher Performance.* Tech. rep. DOI: 10.13140/RG.2.1.2212.7526.

Fagerland, Morten W., Stian Lydersen, and Petter Laake (July 13, 2013). "The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional". In: *BMC Medical Research Methodology* 13.1, p. 91. DOI: 10.1186/1471-2288-13-91.

Fox, John (Apr. 28, 2024). *Applied Regression Analysis and Generalized Linear Models.* 3rd ed. London: SAGE Publications, Inc. 790 pp. URL:

https://us.sagepub.com/en-us/nam/applied-regression-analysis-and-generalized-linear-models/book237254 (visited on 05/02/2024).

Gujarati, Damodar N. and Dawn C. Porter (2009). *Basic econometrics*. 5. ed. The McGraw-Hill series Economics. Boston, Mass.: McGraw-Hill Irwin. 922 pp. ISBN: 978-0-07-337577-9.

Heale, Roberta and Alison Twycross (2015). “Validity and reliability in quantitative studies”. In: *Evidence-Based Nursing* 18.3, pp. 66–67. DOI: 10.1136/eb-2015-102129.

Hurst, Allison (Mar. 2023). *Introduction to Qualitative Research Methods*. Oregon State University. URL: https://open.oregonstate.education/qualresearchmethods/ (visited on 05/17/2024).

Iain, Pardoe (2024). *10.2 - Stepwise Regression | STAT 501*. Stat 501 - The Pennsylvania State University. URL: https://online.stat.psu.edu/stat501/lesson/10/10.2 (visited on 05/06/2024).

Johansson, Marcus (Dec. 2019). “Efficient Modeling and Control of Crushing Processes in Minerals Processing”. PhD thesis. DOI: 10.13140/RG.2.2.21960.67843.

Lixoft (2024). *Standard error using the Fisher Information Matrix*. Monolix. URL: https://monolix.lixoft.com/tasks/standard-error-using-the-fisher-information-matrix/ (visited on 04/11/2024).

McCullagh, Peter (1980). “Regression Models for Ordinal Data”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 42.2, pp. 109–127. DOI: 10.1111/j.2517-6161.1980.tb01109.x.

Moullin, Joanna C., Kelsey S. Dickson, Nicole A. Stadnick, Bianca Albers, Per Nilsen, Sarabeth Broder-Fingert, Barbara Mukasa, and Gregory A. Aarons (Apr. 2020). "Ten recommendations for using implementation frameworks in research and practice". In: *Implementation Science Communications* 1.1, p. 42. DOI: 10.1186/s43058-020-00023-7.

Moullin, Joanna C., Kelsey S. Dickson, Nicole A. Stadnick, Borsika Rabin, and Gregory A. Aarons (Jan. 2019). "Systematic review of the Exploration, Preparation, Implementation, Sustainment (EPIS) framework". In: *Implementation Science* 14.1, p. 1. DOI: 10.1186/s13012-018-0842-6.

Nelson, Archibald (1965). *A Dictionary of Mining*. New York: Philosophical Library Inc., pp. 523–523.

Park, K. Francis and Zur Shapira (2017). *Risk and Uncertainty*. Palgrave Macmillan UK, pp. 1–7. DOI: 10.1057/978-1-349-94848-2_250-1.

Rausand, Marvin and Stein Haugen (2020). *The Words of Risk Analysis*. John Wiley & Sons, Ltd, pp. 15–57. DOI: 10.1002/9781119377351.ch2.

Rice, John A. (2007). *Mathematical statistics and data analysis*. 3rd ed. Duxbury advanced series. Thomson/Brooks/Cole. 1 p. ISBN: 978-0-534-39942-9.

Sandvik (2024). *What is the eccentric throw?* URL: https://www.rockprocessing.sandvik/en/the-knowledge-hub/crushing-chambers/what-is-the-eccentric-throw/ (visited on 04/02/2024).

Sandvik Group (2001). *Sandvik har träffat överenskommelse om att köpa del av Svedala*. URL: https://www.home.sandvik/se/nyheter-ochmedia/nyheter/2001/06/sandvik-har-traffat-overenskommelse-om-att-kopa-del-av-svedala/ (visited on 04/23/2024).

Sandvik Group (2020). *History Special "Innovation" by Meet Sandvik*. en. URL: https://podcasters.spotify.com/pod/show/meetsandvik/embed/episodes/22-History-Special-Innovation-eo3uj0 (visited on 04/23/2024).

– (2023). "Sandvik Annual Report 2023". In: URL: https://www.annualreport.sandvik/en/2023/_assets/downloads/entire-en-svk-ar23.pdf.

– (2024). *ROCK PROCESSING EQUIPMENT AND SOLUTIONS*. URL: https://www.rockprocessing.sandvik/en/products/ (visited on 04/23/2024).

Schield, Milo (July 2017). "Teaching Logistic Regression using Ordinary Least Squares in Excel". In: *ASA Proceedings of the Section on Statistical Education*.

Shen, Guolang and Xin Tong (May 2020). "Particle stratification of a vibrating screen with translation-swing composite motion". In: *Journal of Vibroengineering* 22.3, pp. 498–508. DOI: 10.21595/jve.2019.20683.

Tashakkori, Abbas and Charles Teddlie (2003). *Handbook of Mixed Methods in Social & Behavioral Research*. SAGE. ISBN: 978-0-7619-2073-1.

Wills, Barry A. and James A. Finch (Jan. 2016). *Wills' Mineral Processing Technology (Eighth Edition)*. Butterworth-Heinemann. ISBN: 978-0-08-097053-0.

A Appendices

A.1 Interview Guide

Crushers

Scope of variables

- In your opinion, what are the most critical variables that affect how a Sandvik crusher functions?

- Going through the variables in the new risk method:

  - What variables are you missing?

  - What variables are redundant?

Weighting of variables and computation of risk

- After explanation: What do you think about how the weighting and computation of variables are handled in the new method?

- How should it be handled?

- What crusher variables are absolutely critical for the performance of the machine?

  - For capacity

  - For product grading

  - For shape (SI/FI index)

- Right now the risk assessment weighs all variables equally, given the variables presented above, how would you rank these in importance? (low-medium-high weight)

- Artificial weights (compare them against each other if necessary)

- "100 Euro-test"

- Right now variables values are multiplied, what do you think about this way of computing risks?

Screens

Scope of variables

- In your opinion, what are the most critical variables that affect how a Sandvik screen performs?

- Going through the variables in the new risk method:

  - What variables are you missing?

  - What variables are redundant?

Weighting of variables and computation of risk

- What do you think about how the weighting and computation of variables are handled in the new method?

- How should it be handled?

- What screen variables are critical for the performance of the machine?

- Right now the risk assessment weighs all variables equally, given the variables presented above, how would you rank these in importance? (low-medium-high weight)

- Artificial weights (compare them against each other if necessary)
- "100 Euro-testet"

Overall

- What do you think about calculating the average of the risk scores for the individual machines (and shapes) for the final score?
- Should shape be handled differently?
- Are there some important correlations between variables that should be taken into consideration?
  - What are those?
- What are your thoughts on potentially including over- and undersize in the method?
- Explanation of data handling possibly needed: In the instances when there are test data available for the risk variables, what are your thoughts on how the data is handled?
- Right now very limited use of raw material characteristics is used in the risk calculations (only Work Index for CV crushers), what other risks should take into consideration raw material characteristics?
- Aside from the points discussed above: Do you see any other issues of using the new method?
- How do you think this new risk assessment method should be implemented and used in the Sandvik organization?

A.2 OLS for multiple linear regression

In this section, the OLS approximation will be derived. Starting with the equation in 4.2, to find the least squares coefficients we want to find $\mathbf{b}$ that minimizes the residual sum of squares. (Fox, 2024). Expressed as a function $\mathbf{F}(\mathbf{b})$, we have:

$$\begin{aligned} \mathbf{F}(\mathbf{b}) &= \sum_{i=1}^{n} \varepsilon_i^2 = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) \\ &= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} - \mathbf{X}'\mathbf{b}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \\ &= \mathbf{y}'\mathbf{y} - (2\mathbf{y}'\mathbf{X})\,\mathbf{b} + \mathbf{b}'\,(\mathbf{X}'\mathbf{X})\,\mathbf{b} \end{aligned} \tag{A.1}$$

Vector derivation with regards to $\mathbf{b}$ gives:

$$\frac{\delta \mathbf{F}(\mathbf{b})}{\delta \mathbf{b}} = \mathbf{0} - 2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b} \tag{A.2}$$

After setting the derivative to zero we get:

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}' \tag{A.3}$$

For nonsingular $\mathbf{X}\mathbf{X}'$, a unique solution for $\mathbf{b}$ is then found by:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y} \tag{A.4}$$

It can easily be shown that the solution minimizes $\mathbf{F}(\mathbf{b})$ (Fox, 2024).

A.3 MLE for logistic regression

In this section, maximum-likelihood estimation, MLE, for binary and ordered logistic regression will be derived. Let response variable Y_i take on values 0 and 1 with probabilities π_i and $1 - \pi_i$ respectively. The probability distribution for Y_i can be written as (Fox, 2024):

$$p\,(y_i) = \Pr\,(Y_i = y_i) = \pi_i^{y_i}\,(1 - \pi_i)^{1-y_i} \tag{A.5}$$

Since observations are independent, the joint probability for the full scope of n observations can be written as:

$$\begin{aligned} p\left(y_1, y_2 \ldots, y_n\right) &= p\left(y_1\right) p\left(y_2\right) \ldots p\left(y_n\right) \\ &= \prod_{i=1}^{n} p\left(y_i\right) = \prod_{i=1}^{n} \pi_i^{y_i}\left(1-\pi_i\right)^{1-y_i} \\ &= \prod_{i=1}^{n}\left(\frac{\pi_i}{1-\pi_i}\right)^{y_i}\left(1-\pi_i\right) \end{aligned} \tag{A.6}$$

Using expression

$$\frac{\pi_i}{1-\pi_i} = e^{\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}+\ldots\beta_k X_{ik}} \tag{A.7}$$

and after some rewriting

$$1-\pi_i = \frac{1}{1+e^{\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}+\ldots\beta_k X_{ik}}} \tag{A.8}$$

Inserting these into (A.6) we get the following expression (Fox, 2024):

$$p\left(y_1, y_2 \ldots, y_n\right) = \prod_{i=1}^{n} \frac{e^{\left(\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}+\ldots\beta_k X_{ik}\right)^{y_i}}}{1+e^{\alpha+\beta_1 X_{i1}+\beta_2 X_{i2}+\ldots\beta_k X_{ik}}} \tag{A.9}$$

By treating the observed data as fixed instead of a variable, this function can be seen as the likelihood function $L\left(\alpha, \beta_1, \beta_2, \ldots, \beta_k\right) = L(\boldsymbol{\beta})$ for the linear logistic model. Maximizing the function for the values of $\alpha, \beta_1, \beta_2, \ldots, \beta_k$ returns the maximum-likelihood estimates, which we can notate as $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ or $\hat{\boldsymbol{\beta}}$ in vector format (Fox, 2024).

For convenience, let $\mathbf{x}_i' \equiv \left(1, X_{i1}, X_{i2}, \ldots, X_{ik}\right)$ be the row of the observed design matrix $\mathbf{X}$, defined in 4.2. Referring to Equation A.9, the probability of independent observations of Y given the values of $\mathbf{X}$ is:

$$p\left(y_1, y_2 \ldots, y_n \mid \mathbf{X}\right) = L(\boldsymbol{\beta}) = \prod_{i=1}^{n} e^{\mathbf{x}_i \beta y_i}\left(1+e^{\mathbf{x}_{\mathrm{i}}' \boldsymbol{\beta}}\right)^{-1} \tag{A.10}$$

Taking the natural logarithm on both sides returns the log-likelihood (Fox, 2024):

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} Y_i \mathbf{x_i'}\boldsymbol{\beta} - \sum_{i=1}^{n} \ln\left(1 + e^{\mathbf{x_i'}\boldsymbol{\beta}}\right) \tag{A.11}$$

Taking the partial derivatives with respect to β and setting them to maximize the likelihood yields the following equations (note that maximizing the log-likelihood yields the same result as maximizing the likelihood):

$$\sum_{i=1}^{n} \left[\frac{1}{1 + e^{-\boldsymbol{x}_i'\hat{\boldsymbol{\beta}}}}\right] \boldsymbol{x}_i = \sum_{i=1}^{n} Y_i \boldsymbol{x}_i \tag{A.12}$$

The solution $\hat{\boldsymbol{\beta}}$ is the vector of maximum likelihood estimates (Fox, 2024).

A.4 The stepwise regression procedure

Step 1: Fit each of the one-predictor models, i.e. y regressed on x_1, y regressed on x_2 etc. Of the variables $x_i, i = 1...n$, that have a P-value from the t-test that is lower than α_E, the variable with the lowest P-value is entered in the model. If the no model with a P-value lower than α_E is found, the procedure ends.

Step 2: In this process, suppose x_1 is initially included in the model, then every two-predictor model containing x_1 is fitted and compared. For instance, y is regressed on x_1 and x_2, then on x_1 and x_3, and so forth, up to x_1 and x_n. The variables $x_i, i = 2, ..., n$ whose P-value is lower than α_E are compared and the model including the second predictor with the lowest P-value is chosen. If no P-value is lower than α_E, the procedure stops and the one-predictor model is selected as the final model. If for example x_2 was selected as the best predictor, the procedure now goes back and tests if x_1 is significant in the chosen two predictor model. A t-test is therefore performed for the null hypothesis $\beta_1 = 0$. If the P-value is greater than α_R, remove x_1 from the model. Otherwise, keep the two predictor models.

Step 3: Repeat Step 2 but now compare all three predictor models including x_1 and x_2. When a third predictor is selected, test x_1 and x_2 individually for significance and remove if not significant. This iteration continues until no more predictors are found with a P-value lower than α_E.