CHALMERS
UNIVERSITY OF TECHNOLOGY

Analysis of Production Disturbances with the Use of Big Data

Identifying Quality Issues Through a Case Study at an Electronics Manufacturer

Master's thesis in Production Engineering

MADELEINE ARNASON
MATILDA IDOFFSSON

DEPARTMENT OF INDUSTRIAL AND MATERIALS SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020
www.chalmers.se

Master's thesis 2020

Analysis of Production Disturbances with the Use of Big Data

Identifying Quality Issues Through a Case Study at an Electronics Manufacturer

MADELEINE ARNASON & MATILDA IDOFFSSON

Department of Industrial and Materials Science
Division of Production Systems
Chalmers University of Technology
Gothenburg, Sweden 2020

Analysis of Production Disturbances with the Use of Big Data
Identifying Quality Issues Through a Case Study at an Electronics Manufacturer
MADELEINE ARNASON
MATILDA IDOFFSSON

Supervisor: Adriana Ito, Department of Industrial and Materials Science
Examiner: Anders Skoogh, Department of Industrial and Materials Science

Master's Thesis 2020
Department of Industrial and Materials Science
Division of Production Systems
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Illustration of data use and analysis in production

Gothenburg, Sweden 2020

Analysis of Production Disturbances with the Use of Big Data
Identifying Quality Issues Through a Case Study at an Electronics Manufacturer
MADELEINE ARNASON & MATILDA IDOFFSSON
Department of Industrial and Materials Science
Chalmers University of Technology

Abstract

In Industry 4.0 and Quality 4.0 one of the key enablers is big data and the analysis of it. The interest in Industry 4.0 and the possibilities that comes with it seems to be limitless. However, the implementation comes with big challenges and there is a big knowledge gap when it comes to implementation. Big data analysis can be one of the key enablers to reduce production disturbances and to gain a more controlled and predictable manufacturing process. Therefore, this thesis has investigated how data can be analyzed and what has to be considered when analyzing production data. A case study has been performed at a company in Sweden that has been gathering data for a while but not yet started the analysis. Through current state mapping of the production and the pre-gathered data, insights of the manufacturing process have been obtained. Production tests were performed where a gap in the data was uncovered in order to investigate whether that could affect the production outcome. Through those investigations, this thesis could conclude that it is of very high interest and importance to have a closed loop of data to enable complete and reliable analysis.

Keywords: Industry 4.0, Quality 4.0, Production Prediction, Production Disturbances, Big Data Analysis, Quality Improvement, Manufacturing Processes

Acknowledgements

This thesis would not have been possible without the help of a lot of people. Therefore we want to thank our supervisors at Chalmers and at the company, Adriana, Kristoffer and Rakesh. This thesis would not have been possible without your valuable knowledge and guidance.

It would also have been impossible to conduct this thesis without the help from everyone who helped us at the production site, a special thanks to Robert and Stefan who has done everything and more to help us even if you might not have the time.

And to all of you lovely operators in the production who devoted your time to help us conduct the experiments when we could not be present, your work is more than appreciated.

Madeleine Arnason & Matilda Idoffsson
Gothenburg, June 2020

Contents

List of Figures

List of Tables

List of Abbreviations

AA	Active Assembly.
DMAIC	Define, Measure, Analyze, Improve, Control.
DR	Dry Room.
ECA	Environmentally Controlled Area.
FIFO	First In First Out.
FT	Function Test.
IoT	Internet of things.
LCL	Lower Control Limit.
OEE	Overall Equipment Efficiency.
P/F	Pass or Fail.
PD	Production Disturbance.
PL3	Production Line 3.
PL4	Production Line 4.
QM	Quality Management.
SQL	Structured Query Language.
UCL	Upper Control Limit.

1 Introduction

One of the biggest issues in today's manufacturing industries is the low Overall Equipment Efficiency (OEE) [1]. However, it is not always evident what the cause for low efficiency is.

1.1 Background

OEE is often used in order to find where in the production the biggest issues exist. However, studies show that companies are using it in an inconsistent way and define the different inputs for OEE and what a production disturbance is, in different ways [1].

To handle the production disturbances and to find them, data and big data is often collected to enable analyzes of production. The use of big data is growing in multiple areas, not only in production. In production, there has been an increasing trend over the last years towards more research within big data [3]. High quality and stable input data without noise is of great importance to be able to produce a reliable prediction of the production. The use of big data has been shown to enable up to 90% prediction accuracy of the production stability if the input is of high quality [4]. With the correct use and interpretation of big data, OEE can be improved and the disturbances can be lowered [5]. However, understanding what data that should be collected and how to interpret it is not as investigated as the fact that data should be used [5]. There is no universal model that fits all productions, what data to collect and how it should be used needs to be adapted to each specific production system. This is a not an easy task and is one of the main challenges when it comes to big data in manufacturing [6]. The use and analysis of big data in production is one of the key aspects to Industry 4.0 [6]. With the strive towards Industry 4.0 and use of big data, hidden patterns and unknown correlations can be found. Patterns can then enable a higher understanding of the production outcome. Within Industry 4.0 a higher quality of the end product can be achieved through a more thorough analysis [7].

A Swedish company specified in manufacturing electronic products develops and produces the products refereed to as product A and product B. The production of product B is well known and stable but the technique behind product A is fairly new and the process suffers more from product quality issues. This ends up in scrap due to not being able to identify the causes of the errors in early processes. Products are therefore kept in the production and more value is added along the way before the errors are found in the quality check at the end of the production [8]. The production processes collect a lot of data and in the current situation most data is only stored in a database without any further analysis. The scrap rate has become a big problem and it ends up in a lot of money, work and time going to waste. The case company has started to look into the collected data and to investigate how it can be used. Data collecting started about one year ago and only analysis limited to one or two parameters have been done. The product is increasing in demand and is predicted to increase more over the coming years which means that the case company is expanding fast and are in need of addressing the quality issues now.

The current scrap rate is tangible and close to its upper limits and in the case of expansion, the scrap must become more controlled to be justified.

1.2 Problem Description

The case company currently suffers from unpredictable losses in production. The product is very complex to produce and very sensitive to environmental factors which make it harder to find the exact cause of the errors in the production. They are expanding the production and will increase the number of production lines in the near future. The number of errors in the production will increase if prediction of errors is not performed and prevented.

1.3 Aim

The aim of this thesis is to conduct a case study in order to understand how big data can be used to predict and prevent possible production disturbances and from that knowledge enable control of the production.

Another aim is to produce a method or procedure that the case company can follow for similar issues in the future. Both for interpreting parameters in the current production and how to perform data analysis. The procedures are supposed to be as general as possible to be adaptable to other productions using big data.

1.4 Limitations

The case company has multiple production sites producing different products around the world, only one site and one product will be analyzed in this thesis. Historical and existing data collected during the last year from the current production of the chosen product will be used. In case there is a lack of data, nominal data will be used or tests will be performed if possible.

The thesis focuses on finding parameters for possible error prediction in the production. There will be no implementation of the findings due to lack of time, only recommendations on what can be improved and possible further analysis will be made.

Investigated parameters will be used to either find reasons to lower OEE or possibly dismiss preconceptions the case company might have of what the issue is in the production. No implementation or measuring of the OEE will be made or analyzed to find if OEE is actually lowered.

1.5 Extraordinary Circumstances

During the thesis work, the Corona pandemic erupted and affected the production and parameter investigation. The production was not able to run at full speed and the staff was partly on furlough leave.

1.6 Research Questions

In this thesis the following four research questions will be answered. RQ1 and RQ2 are of scientific nature while RQ3 and RQ4 are more relevant for the case company.

RQ1. Is it possible to find an appropriate procedure for the case company to follow in future projects to apply in similar situations?

RQ2. For which reasons is it important to have a closed loop of data for data analysis of production performance?

RQ3. Is it possible to find reasons of the high scrap rate through analysis of production disturbances?

RQ4. Can parameters be found in production data to predict the production outcome?

1.7 Outline of the Report

This report consists of 7 main sections and appendices in the end. This first section gives a broader view of what the problem of investigation for this thesis is and the background of the solutions that are sought. In the second chapter, the theory needed to understand the production processes and the theories used for building the case are presented. The methods used in this report and thesis work can be found in section 3. There, information regarding the plans and applications of the theory can be found and how they are interpreted. All results can then be found in the fourth chapter, where the current state, the results of data collection and data analysis are presented. The discussion is presented in the fifth section where methods, results and theory will be constructively discussed. In the fifth section the conclusion will be presented that contains the most important and fundamental findings of this thesis. Lastly, there is a section regarding recommendations of future work to be done, this a recommendation from the authors to the case company and for similar studies regarding big data as a tool for production prediction.

2 Theoretical Framework

This section will present theory connected to this thesis, the theory is relevant to understand the thesis' methodology and to strengthen findings made in the thesis.

2.1 Manufacturing Process

The choice of manufacturing process is depending on what type product that is to be manufactured, cost, design, work organization, company culture, material, quality and etc.[9, 10] A manufacturing process consists of multiple smaller processes that together becomes a production system. It can also consist of different flows, such as line, process, product or functional flow [11]. For this case study, the product is a small electronic product and the processes are therefore chosen to fit small products in big volume and with a high demand on quality.

2.2 Production Disturbances

The concept of Production Disturbance (PD) may vary depending on what the focus is, but it is something that affects the production or the planned production time in a negative way. A disturbance can affect the work environment as well as the quality of the products [10]. What is defined as a PD can only be defined from what the purpose is with the process. In figure 2.1, 21 different parameters has been classified as PD in a survey performed in a study 2016 [1]. A disturbance must not end up in a negative result but it is something that has occurred out of control [12]. In this study, disturbances used are: *Adjustments* and *Scrap or quality problems of the product*. Disturbances also occur out of control and affects the production outcome negatively.

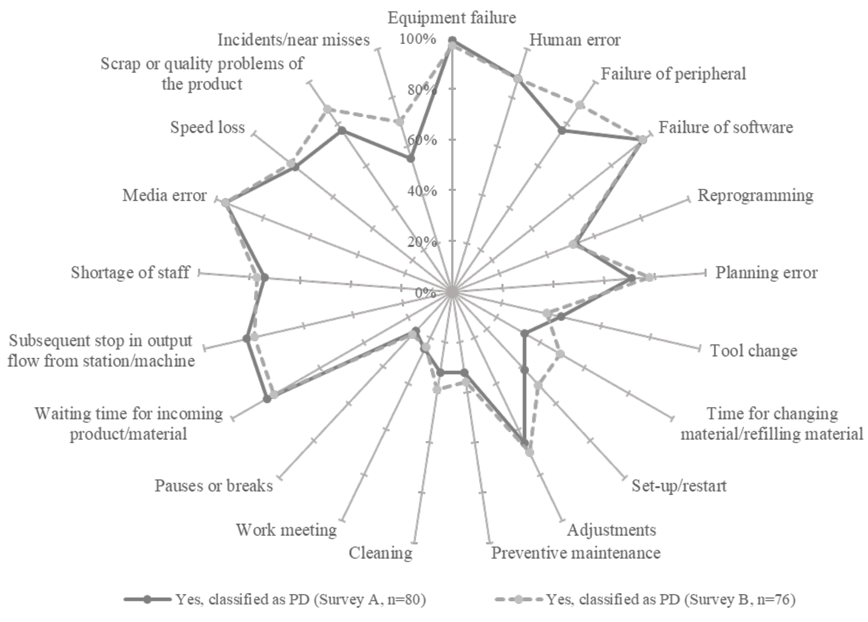


Figure 2.1: Classification of production disturbances, reconstructed from [1]

Disturbances can be categorized into three sections which are downtime, quality and speed losses. Downtime can then also be sub categorized into planned and unplanned disturbances as seen in figure 2.2 [2, 13]. To handle a PD, actions must be taken to correct, prevent and eliminate the disturbance to ensure that the production can be stabilized directly and continuously in the future [10]. What action to do to correct, prevent or eliminate the disturbance depends on supplier, machine, etc. [2]. Since the definition of a PD differs from company to company, something that is defined as a disturbance for one company might not be for another. When it comes to quality in PDs, rejection of products is included as can be seen in table 2.1. Rejection of products can be a result of adjustments of the machines which can affect the production in either positive or negative ways, therefore rejection can be considered a PD [2]. In this study the quality losses will be investigated with the use of big data to enable a prediction of the disturbances that could decrease the level of losses due to improved quality.

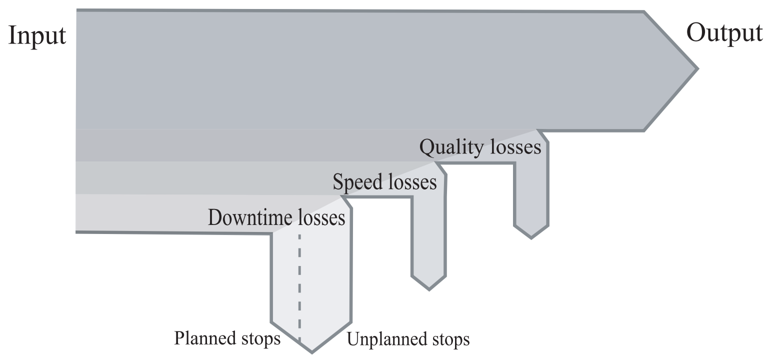


Figure 2.2: PD categories according to [2]

PDs must be prevented to an as great extent as possible to avoid decrease of output. When a potential PD is identified it must be taken into consideration and be prevented before it turns into a disturbance [10]. Manufacturing of electronic products are becoming increasingly complex due to smaller components and a demand of better performance [14]. With a dynamic and a more complex production system the disturbances are harder to handle and the available methods to prevent them are few [10].

Table 2.1: Different PDs and their categorization [2]

Downtime losses		**Speed losses**	**Quality losses**
Planned	Unplanned		
Set-up Maintenance	Damage Repair Lack of material Blocking	Idle running Speed or cycle losses Ramp-up	Rejection of products Adjustments

2.3 Assembly Process in Electronics Manufacturing

The tolerances are very narrow in electronics manufacturing which means that the manufacturing needs to be very precise to meet the high demands from costumers. To ensure this, different types of manufacturing methods are used by different companies and the main types can be divided into active and passive assembly[15]. The choice of assembly method also impacts the choice of fastener method.

2.3.1 Fastener in Assembly and its Effects

There are multiple types of fasteners used in electronics manufacturing and the two main ones for attaching components to each other are adhesive and screwing [16].

2.3.1.1 Screw

Screwing components to each other is fast and gives the same result every time but gives no room for errors and can result in dust on the components that can effect the performance [16]. The method is generally not effected by the environment at the assembly process such as humidity or temperature and is therefore a stable method. Screwing components together disables the possibility to move the components in X and Y direction as well as roll, pitch, yaw angles [16].

2.3.1.2 Adhesive

Adhesive gives a freedom in movement when aligning components relative each other but gives a higher variance of spread and shrinkage of the adhesive [16]. Depending on the adhesive used it can be affected by environmental factors such as temperature and humidity. Some adhesives are very sensitive to moist and can lose performance if the environment is not controlled enough [17].

The adhesive must be able to cure quickly to avoid movement when the assembly has been performed. Other important factors for the adhesive can be electrical properties, high strength and flexibility. In most cases the adhesive is not supposed to be conductive to avoid short circuit [17]. One of the factors that must be taken in consideration is the shrinkage level that occurs when the adhesive hardens. The adhesive must have the right properties to not spread uncontrollably or that leads to an uncontrollable production.

Adhesive provides the possibility to move in multiple directions and angles which enables more degrees of freedom as can be seen in figure 2.3. The component can be adjusted by 3-6 axis depending on were the best quality is achieved [15].

2.3.2 Assembly Process Available in Industry

The assembly method used depends a lot on the product being produced, its quality demands, cost and lead time. Assembly can be performed with an active process that measures continuously to meet requirements or a passive method that performs its task independent of the best performance.

2.3.2.1 Active Assembly (AA)

In Active Assembly (AA) a dynamic assembly process is performed where the

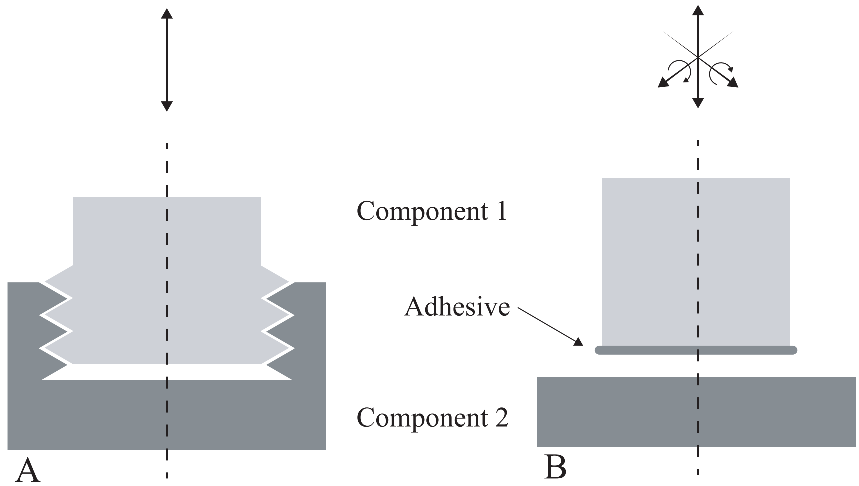


Figure 2.3: Different fastening methods for electronic components, and their degrees of freedom, A. shows screwing method and B. shows the use of adhesive to attach the components.

parts to be put together are mounted with dynamic parameters adjusted to the specific parts and aligned to meet the best possible performance[15]. The AA can also compensate for some of the product variance of the components and thereby produce a higher quality for the end product [18]. In AA the components can be adjusted in 3-6 axes depending on were the highest resolution is achieved [15]. It is most common to use either adhesive or screwing to attach the components [16]. With the use of screwing the components can only be adjusted in the Z-axis. With the use of adhesive the components are free to move relative each other in multiple axes as illustrated in figure 2.3. Depending on the machine used, the degrees of freedom can be set between 3 and 6 axes [16].

AA consists of a *closed loop* of data that feeds the results back to machine continuously to change the parameters and in that way find the best possible performance as can be seen in figure 2.4. The closed loop of data is key in AA, the alignment adjustments of the assembly are done relative to the feedback from measurements performed during assembly [15]. With AA the component that provides the output data must be connected to the machine performing the measurement, which is often the same machine that performs the assembly [15]. In AA the adhesive is applied before the alignment starts since the process continuously reads the data and finds the best position [15].

2.3.2.2 Passive Assembly

Passive assembly in comparison to active does not consist of a closed loop of data, as seen in figure 2.4. In passive assembly the components are mounted together with a pre-defined set of parameters which gives no room for variation in supplier material. The components can be placed with the use of a fixture or with the help of visual marks [15]. In passive assembly the components are often first measured in a *pre-alignment* and when the best position is found the adhesive is applied [15].

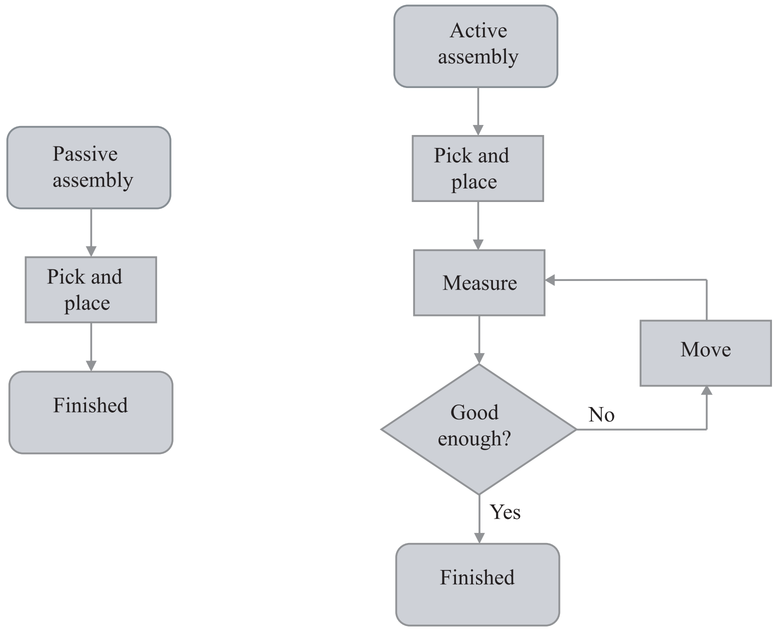


Figure 2.4: Simplified flowchart showing the process in Passive Assembly to the left and Active Assembly to the right

2.3.3 Assembly Method and Fastener at the Case Company

At the company where the case study is made, the choice of assembly process is an AA process with adhesive as fastener. The case company has multiple resources that performs the AA and the exact process in each resource varies depending on the manufacturer of the machine. The differences lies within what component moves relatively to what which results in a difference in the alignment path but ultimately can achieve the same result. In the assembly process, an adhesive that is sensitive to humidity is used and therefore the process is performed in an Environmentally Controlled Area (ECA) where the humidity and temperature are controlled and kept stable.

2.4 Big Data

There is no specific definition of *big data*, big data often refers to data in big volumes [19]. One problem with big data is the differences of what people defines as big or small data [19]. Big data is sometimes defined using a combination of five categories: Volume, Variety, Velocity, Value and Veracity [19, 20]. Data can be produced from different platforms [20]. In this production system, data produced from the machines' activities will be analyzed. Data generated as log-files from a machine can be characterized as *Internet of things (IoT)* [20].

2.4.1 Big Data for Production Prediction

The use of big data in production and for prediction of the production outcome has been a growing area over the last few years. Looking into how often articles have been written on the subject of production in combination with big data there has been a clear increase over the years [3]. There has been a significant growth in the area of manufacturing and analysis. Data in production can for example be used to detect disturbances in machines or to detect reasons for product errors [3]. Big data has been identified as one of the key enablers for quality and process monitoring [3]. This is important since being able to monitor quality through analysis of data, a higher stability can be reached and the errors can be prevented. However, it is hard to know what data should be used and how to implement the results of big data analysis [20].

2.4.2 Approach to Analyzing and Interpreting Big Data

Projects can often follow a plan-do-check-act cycle and it can be interpreted into five phases: Define, Measure, Analyze, Improve, Control (DMAIC) [21]. Once the full DMAIC cycle has been processed it can start over and attack a new problem. The idea is to enable analysis without the use of high statistical expertise and to continuously improve [21]. A closed loop of knowledge and information enables easier monitoring and better approach to the problems, see figure 2.5 for a visualization of the closed loop.

Define-phase

The first phase to be performed is the Define-phase, here the problem at the current state is defined to enable the knowledge throughout the process. In this phase the scope should be defined and a prioritization should be made from cause-and-effect relations [21].

Measure-phase

In the second phase the current state should be measured and good behaviors should be identified. This, to enable the possibility to facilitate the findings of bad behaviors [21]. For example, it can consist of finding the mean and standard deviation at a good time of performance.

Analyze-phase

For the analysis phase, the main goal is to find root-causes and to analyze data gathered in the measuring phase [21]. Finding patterns, parameters that are correlated or comparing mean and standard deviation are examples of analyzing methods that will be used.

Improve-phase

In the fourth phase, the results from the analysis are to be tested and solutions to the identified problem confirmed. Depending on the type of production, solutions can either be tested directly in the running production or applied to a simulation first [21]. When the desired outcome is reached the phase is finished.

Control-phase

With a finished project and a solved problem, the ability to reproduce the process should be documented [21]. Therefore the control-phase consists of enabling a continuation of the process for upcoming projects and to enable monitoring for preventive work [21].

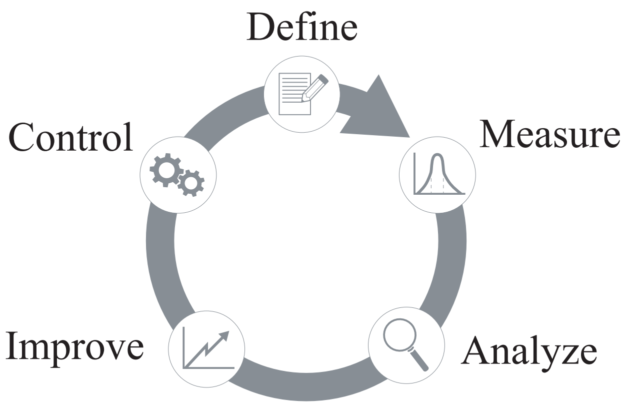


Figure 2.5: The basic idea for the approach of the problem

2.4.3 Analysis and Gathering of Production Data

For analysis of data and efficiency of the production, it is important that every machine records data that can be used. It is then of great importance to know what can be read from the data to understand what can be improved [22]. The data can be used to predict probable failures and to do so, data logs must be relevant and data from sensors, machines and historical maintenance data must be analyzed. To be able to collect all data it is very important to have an environment constructed for data gathering and to save historical data from processes [23]. Path analysis can be used to find correlations between critical events and data captured through the process, and also to identify machines that contribute to errors [19]. A group of machines can together affect the faults and it is therefore important to investigate the machines together as a line and not only individually. The products are affected both from downstream processes as well as upstream and can result in a more complex analysis [23].

2.5 Quality Management (QM)

For every organization, some sort of Quality Management (QM) is in use. It is chosen to fit the company and their work, processes and products. There are different definitions of Quality Management (QM), Demming defines it as where quality focuses on achieving a stable process [24]. When a product's variations are within limits the process is stable and considered predictable [24]. Demming also says that the management must be involved at a system level to reach high quality. The responsibility to ensure good quality does not lay within the inspection department or the inspectors, it must be driven by managers and be continuous throughout the whole process [24].

2.5.1 Quality in Industry 4.0 and Big Data

QM and quality work has changed a lot over the years. From the beginning the focus was to have more quantity, more material and more services [25]. The development has evolved from only hardware to focus on either hardware or software and is now developing into a combination where both have to collaborate [25]. With Industry 4.0 one of the main goals is to find new improved ways to increase the quality of products [26]. Through the use of big data and analysis, manufacturing and product quality can be improved and errors can be prevented [20, 26]. The development of Artificial Intelligence (AI) is moving forwards and to reach Quality 4.0 within Industry 4.0, the understanding and use of big data must be developed through the whole value chain. Interconnection between all types of machines is one of the core features in Industry 4.0 [26]. This means that the process and resources must communicate with each other, not only through a centralized system [25]. With this, a data-driven system for decisions can be developed and the processes can then be controlled according to different circumstances predicted by data [27]. Today, data is often accessed and analyzed after the production of a product is completed. With the use of a connected production there is a possibility to access quality data in real-time which enables changes to be implemented directly [25].

2.5.2 Possible Quality Defects

The product consists of three components where component 2 is attached with adhesive to component 1. Adhesive is then applied to the base (component 1 and 2) and then Component 3 is attached with the use of AA as described in section 2.3.3. Some quality defects that might occur in the manufacturing processes of the investigated electronic product are presented below. The defects are potential defects caused by the assembly process in the ECA.

2.5.2.1 Error in Alignment

One defect that might occur is in the alignment of component 3 when attached to the base. The center-line of component 3 must be aligned to the center of the base to ensure the best performance. See figure 2.6 for an example of how the alignment error might look like. Component 3 is tilted and incorrectly aligned relative the reference line of the base which can affect the performance.

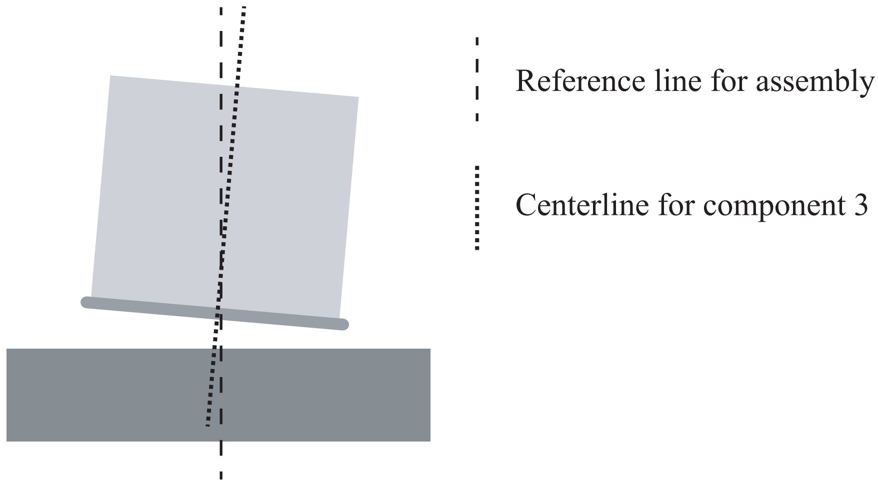


Figure 2.6: Alignment of component 3 on to component 1 with a tilt compared to the center line.

2.5.2.2 Amount of Adhesive

The total height of the product is an essential parameter for the product's result and performance. The performance is negatively affected by both higher and lower height of the product. Components of the product have a very small variance of height but the total product height is affected by the amount and shrinkage level of adhesive when it hardens, as can be seen in figure 2.7. The assembly of component 3 has a preconfigured offset to compensate for the average shrinkage level. Since it is preconfigured, variations of shrinkage can not be compensated for.

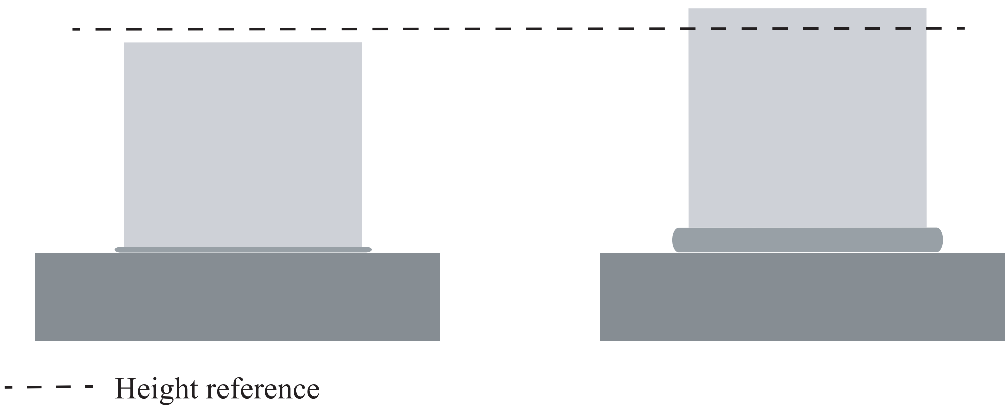


Figure 2.7: Effects of how to much or too little adhesive affects the height of the total product, too little adhesive to the left and too much to the right.

2.5.2.3 Spread of Adhesive

When assembling component 3 to the base, adhesive is applied to component 3. In the assembly process, component 3 is then moved around to find the best result as described in section 2.3.2.1. The adhesive might then be spread over a bigger area and affect component 2 as can be seen in figure 2.8, sub-figure B. When applying the adhesive it must also be evenly spread on component 3 to avoid gaps of the adhesive, as seen in figure 2.8, sub-figure C.

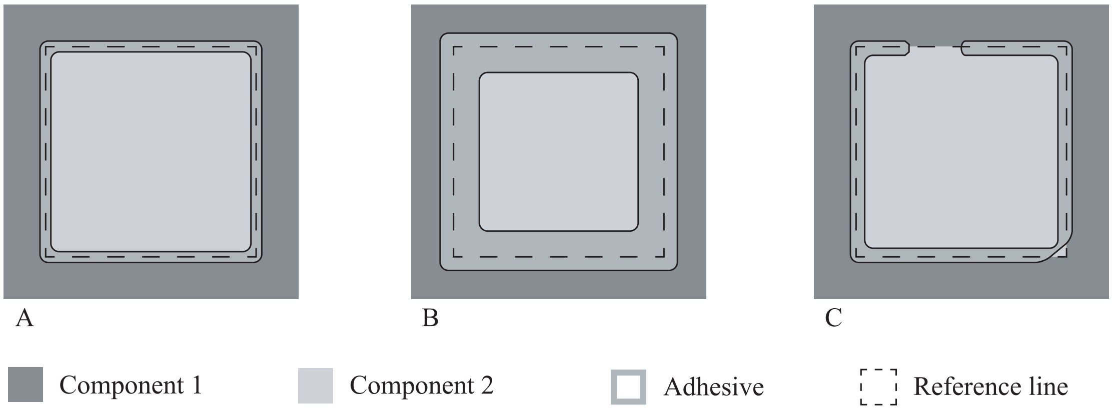


Figure 2.8: Adhesive spread. A shows perfect spread, B shows too much spread of adhesive and C shows gaps of missing adhesive

2.5.2.4 Adhesive Strength

As mentioned in section 2.3.3, the adhesive used is sensitive to moisture. If the adhesive is exposed to a higher level of humidity than allowed, the adhesive is affected and the strength of the adhesion is lowered. The case company has done an investigation of what happens when the used adhesive is subjected to different amounts of moisture. Figure 2.9 shows an illustration of the results where the adhesive that has been exposed to higher moisture develops cracks, which results in a lower durability and strength.

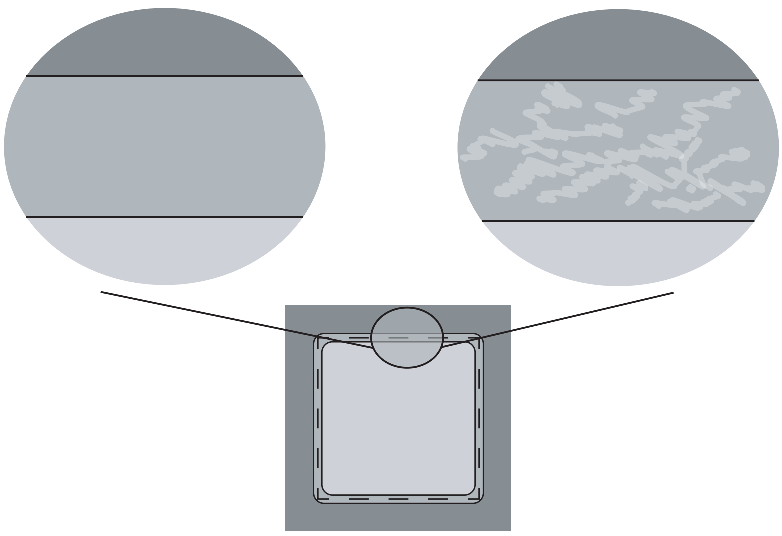


Figure 2.9: Non affected adhesive to the left and adhesive affected by moisture on the right

2.6 Statistics

For the use of the DMAIC cycle, simple statistical tools are supposed to be used in-house and to facilitate the implementation of the analysis [21]. The statistical tools used in this thesis are presented in the coming sections. Examples of used statistical tools that are distribution tests, descriptive statistics and scatterplots in both 2D and 3D.

2.6.1 Descriptive Statistics

Descriptive statistics such as mean, variance and standard deviation can be used to describe the apperance and spread of data and to summarize data [28]. To get good use of descriptive statistics it is important to have a large enough dataset and to present data that is comparable to each other [29].

Mean ($\overline{x}$)
The mean value is often used to describe the central tendency of data and the average value of the data [29]. The mean is defined with equation (2.1).

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.1}$$

where: $\sum_{i=1}^{n} x_i$ = is the sum of all values
n = is the number of values

Variance (σ^2)
The variance is a measurement of the difference between each value in the set of data to the mean [29]. Variance is defined with equation (2.2).

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{2.2}$$

Standard Deviation (σ)
Standard deviation is simply defined by taking the square root of the variance. It is defined in equation (2.3) and describes the average deviation from the mean and can be used to describe the spread of values [29].

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{2.3}$$

Standard Error/Mean Deviation ($\sigma_{\overline{x}}$)
To describe the accuracy of the mean the standard error can be used as seen in equation (2.4). A lower value describes a more accurate mean for the dataset [29].

$$\sigma_{\overline{x}} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}| \tag{2.4}$$

2.6.2 Correlation

To describe if variables correlate to each other the measurement correlation is used [30]. The correlation equation can be found in equation (2.5), for all cases $-1 < r < 1$. If $r = 1$ the variables have an absolute positive relationship and $r = -1$ has an absolute negative relationship. The closer r is to 0 means less association between the different variables [30]. The correlation can be seen as a graph where a more line shaped plot indicates a higher correlation, see figure 2.10 for example of highly correlated data. The equation results in one r_{xy} for each correlated parameters, these can then be summarized and presented as table such as table 2.2.

$$r_{xy} = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2\sum(y-\overline{y})^2}} \tag{2.5}$$

where: x, y = two different independent sets of variables
$\overline{x}, \overline{y}$ = mean of the two variables x and y

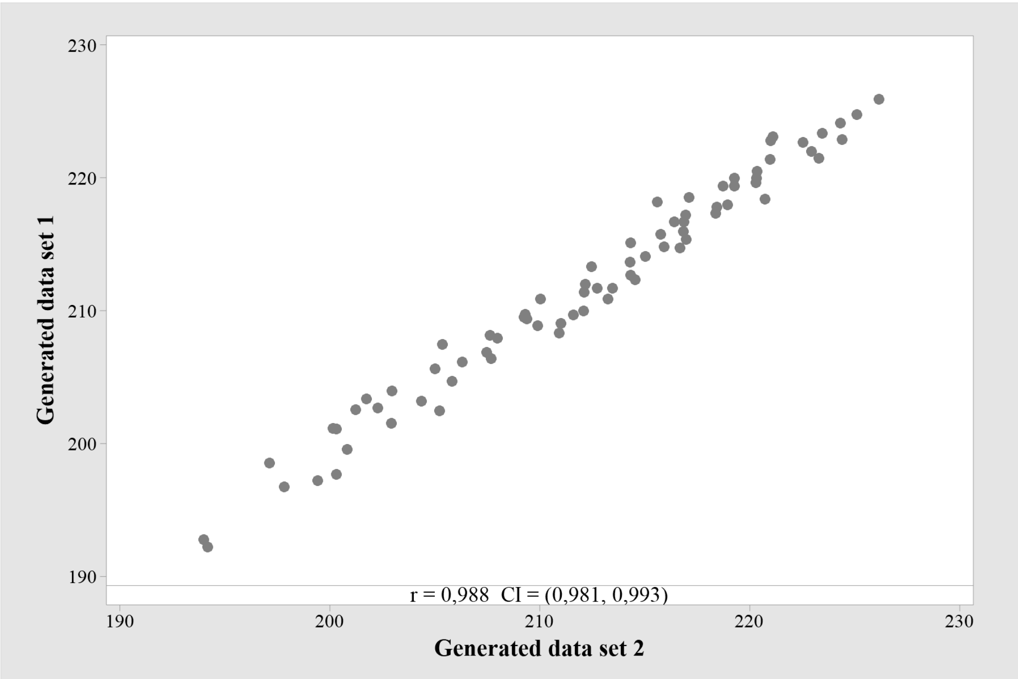


Figure 2.10: Correlation plot generated from correlated data with $r_{12} = 0.988$ in Minitab

Table 2.2: Example of a correlation matrix from Excel with Analysis ToolPak

	Parameter 1	Parameter 2	Parameter 3
Parameter 1	$r_{11} = 1$		
Parameter 2	r_{12}	$r_{22} = 1$	
Parameter 3	r_{13}	r_{23}	$r_{33} = 1$

2.6.3 Normality Test with Anderson-Darling

To see if the distribution of the data is normal a normality test can be used, one of the most common tests is the Anderson-Darling. The Anderson-Darling test is based on an empirical distribution function and is known to perform very well [31]. The Anderson-Darling is a goodness-of-fit test and is used to determine whether a dataset can be considered to be fitted according to the given specified distribution, in this case a normal distribution [32]. To get a visual result of a normality test a normal quantile plot can be used [33]. If the data plotted in a normal quantile plot tends to a linear pattern the data is normal distributed as can be seen in the example in figure 2.11.

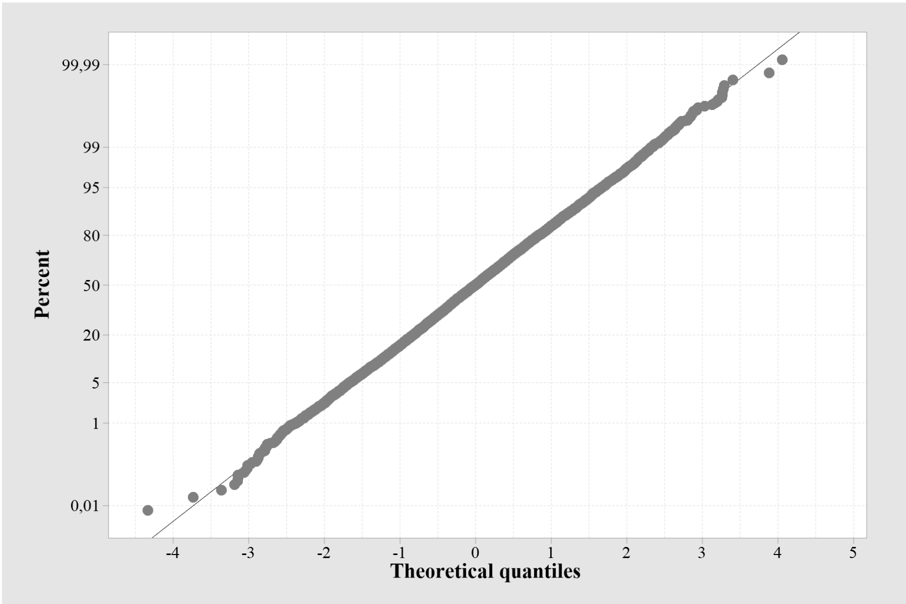


Figure 2.11: Example of a normal quantile plot of random normal data generated in Minitab

2.6.4 Scatter Plots, 2D and 3D

A scatter plot or scatter diagram shows the relation between two parameters, or three if a 3D plot is constructed. Additionally to the parameters, numerical or ordinal variables can be used for the point size and a nominal parameters can be used for color definition [34]. In a scatter plot, each dot represents one data point formed by the intersection of the parameters [35].

2.6.5 Line Charts

Line charts can be used to show the frequency of a discrete variable. The data is presented in a graph with two axes and all data points are connected with a line [36]. Line charts are efficient to show changes between two variables.

2.6.6 Relative Change

The change between two values can be described with a percentage growth factor according to equation (2.6) [37]. To get the relative change, p is solved from equation (2.6) and then is equation (2.7) gained.

$$q_1 = q_0\left(1 + \frac{p}{100}\right) \tag{2.6}$$

where: q_0 = Initial value
q_1 = Final value
p = Growth factor

$$p = \frac{q_1 - q_0}{q_0} \cdot 100 \tag{2.7}$$

2.7 Case Study

When conducting a study that does not require any control over behaviors and that aims to answer questions on *how* and *why*, a case study is a appropriate strategy to use [38]. When wanting to answer the question on *why* something happens a case study can be designed to answer that in contrast to an experiment that can answer questions like *how much* and *what*. Within a case study, other strategies can be used, such as surveys and experiments. A case study uses both direct observations and systematic interviewing as added sources to form a historical study [38]. The case study also consists of a variety of possible shreds of evidence to use, such as documents, observations and interviews. This makes the case study broader than the typical historical study [38]. A case study can use existing and new information or data to find connections to issues. Case studies can be a very good method when studying examples like quality information or a product [39]. When choosing parameters to investigate in a case study it is more flexible than in a rational study. In a case study, the parameters can be chosen to fit the situation and not vice versa. The ability to draw conclusions in a case study is depending on how well the parameters have been chosen and investigated [40]. When using too many parameters it can be hard to draw conclusions due to the complex level that might turn into something uninteresting and that is not understandable [40].

3 Methodology

The thesis was conducted as a case study at one of the case company's production sites in Sweden. Case study was an appropriate method to use because of the uncontrollable behavior of the system and the focus of the current production. According to Yin [38] a case study focuses on answering *Why* and *How* of a problem which was what this project wanted to answer. The study was mainly conducted with two main focuses, case study and documentation of the approach of finding reasons for errors. This, to be able to answer and investigate the issues regarding problems in the production.

Case study

One of the aims is to produce a method or procedure that the case company can apply to other production disturbances. The case study consisted of data collection in the form of interviews, surveys and data extraction from the case company's database. For interviews, a protocol was used to categorize answers and to assure the questions answered are comparable between different interviewees. Interviewees were chosen to reflect different levels of management as well as process experts in order to assure a variation of interest. The collected data from interviews and the database were summarized, analyzed and evaluated in four phases as seen in figure 3.1. The phases represent the DMAIC phases Define, Measure and Analyze. Here, the analysis phase was divided in to phase 3, analysis, and phase 4, evaluation.

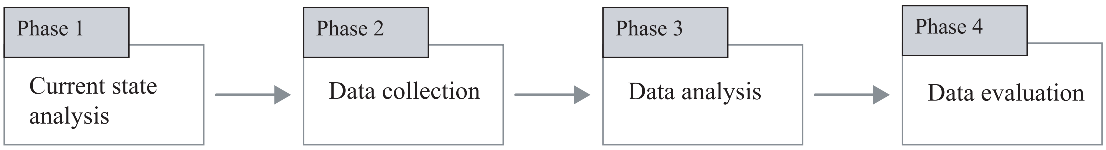


Figure 3.1: Illustration of the research phases

3.1 Phase 1: Current State Analysis

In order to understand how the production worked and to find possible correlations between factors, a current state mapping was done. This was a way to accurately identify what production disturbances have the greatest impact on the production as a whole. It acted as a guide to understand where to start and give a broader view of where the issues were and how different issues could be related. The current state analysis consisted mainly of go-and-see actions and to talk to relevant process owners and experts in order to fully understand the investigated production. For the go-and-see actions, the production facility was visited and the machines were observed while production was running. The documentation of the current state was made through photographs, videos and taking notes. In the current state, the existing database was also shown to the authors and some parameters were explained to give a broader view on what parameters could be expected to be found. This was also to understand how data was stored and what type of data that could be stored and eventually used for analysis. A Master's thesis was conducted 2019 in the production and performed a very thorough current state mapping which was used to gain knowledge of the production and the issues identified one year ago [8].

3.2 Phase 2: Data Collection

Multiple data collection methods were used and combined to gather the best possible prerequisites to the oncoming analysis. It started with interviews to get to know the problems in the production and continued with a survey. Then two production tests were performed and lastly data was extracted from the database.

3.2.1 Production Mapping Through Meetings

Interviews were performed with different people to get their point of view on where the issues lie. The interviews were adapted to the specific person's special area of knowledge and a specific questionnaire was not possible to follow. However, five standardized questions were used for the purpose of the comparability. The questions asked can be seen in table 3.1.

Table 3.1: Questions asked on interviews with different people

Number for reference	Question
1	What do you think can affect the production outcome?
2	What data or parameters can be extracted from your area?
3	According to you, what would be interesting to find?
4	What obstacles do you see in your area that can be hard to analyze with data?
5	Who else do you think we should talk to in regards to the subject of the thesis?

3.2.2 Parameter Classification Through Survey

Based on the interviews performed, a survey was created in order to gather opinion on parameters. Parameters were chosen both from the interviews and from thoughts of the authors of this thesis to enable a different perspective and opinions that the case company employees might not have reflected on before. The questionnaire sent out can be seen in appendix A. The first question gathered the information on the replicants professional title in order to ensure that different focus areas were covered. This, also to investigate conflicts of interests and enable correlation between work area and parameters of interest. They were then asked to grade the given parameters in how they effect the production outcome from *not affecting at all* to *definitely affecting* or the option *I don't know*. All parameters are seen in table 3.2. If any parameters were missed in the questionnaire a free text question was added in the end to enable the replicants to add additional thoughts.

Table 3.2: All parameters that were asked to be graded in survey

Number	Question
1	Humidity in ECA
2	Temperature in ECA
3	Product variation from supplier
4	Operators in ECA
5	Component 2 placement on component 1
6	Effect on component 3 from pre-treatment in assembly
7	Fixture in assembly
8	Adhesive placement
9	Variation of movement in assembly (adhesive spread)
10	Time from ECA to hardening 2
11	Placement in hardening 2
12	Time in hardening
13	Temperature drop when hardening 2 is open
14	Time from hardening 2 to Function Test (FT)
15	Total production time

3.2.3 Existing Data Gathering

Through the parameters with the highest rate from the survey, parameters were chosen to be extracted from the database if possible. Those parameters that were not available in the database were gathered through a test in production, see more in section 3.2.4. Parameters available were requested from a production engineer to be extracted from a database using Structured Query Language (SQL). All data stretched over the time period 2019-04-01 to 2020-04-06 and was extracted for each product produced during the time, both failed and passed products. Data was extracted in multiple csv-files depending on the production line and the placement of storage in the database.

3.2.4 Data Collection Through Production Tests

A lack of data was found in the production at a hardening process. To close the loop of data in the investigated process, the two following tests were planned and executed. The data for humidity and temperature was collected with two digital sensors over the time period of 21 days. For measurement data from the products of investigation the Function Test (FT) stations were used and the data was then extracted from the database.

3.2.4.1 Test 1: Time and Placement in Hardening 2

For the purpose of investigating whether the placement of the trays of products in the hardening process affects the production outcome or not, test 1 was performed. The test also included the possible cycle time differences of the hardening 2 process. 24 trays of products, produced on the same line in the ECA and during the same shift was used for the test. The cycle time in the hardening process is 6 ± 1 hours.

The units were divided into three time intervals 5, 6 and 7 hours. They were also divided into 8 different positions in the resource, one for each corner of the cubic shaped machine. The 8 positions were divided into 3 different time parameters meaning that $8 \cdot 3 = 24$ trays were used. The placements in the process can be seen in figure 3.2. Each tray was marked with the position number and time parameter, for example T1-4 5h, where T1 stands for Test 1, 4 for position 4 and 5h for the time 5 hours.

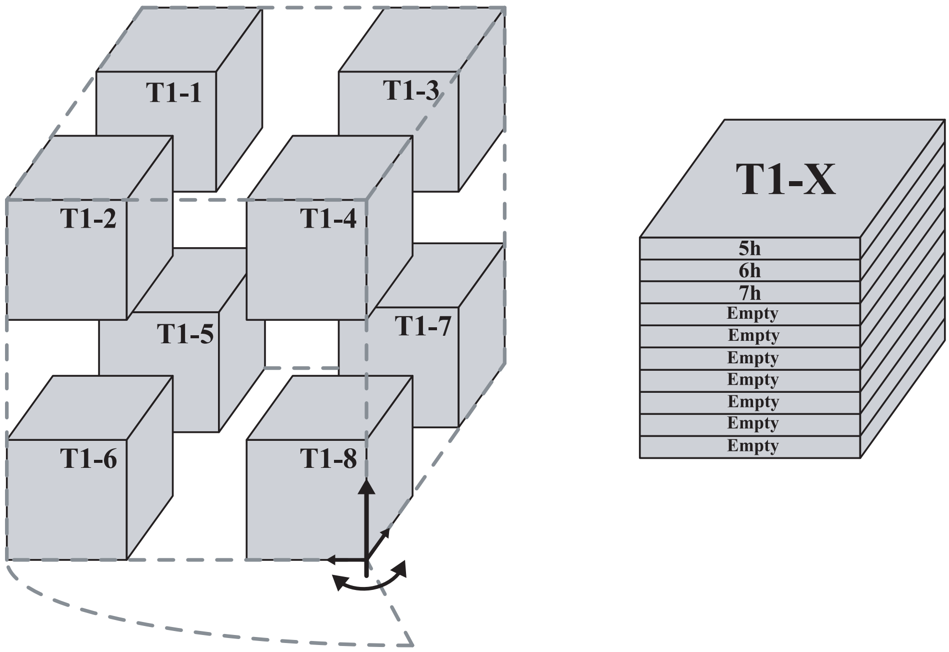


Figure 3.2: Test 1: Placement of trays in hardening 2 to the left and to the right each position's division of timestamped trays

After the hardening process the trays were put to cool down to reach room temperature before being able to be tested in FT. When reaching room temperature they were placed in a climate cabinet until the FT-resource was available. The climate cabinet ensured a stable humidity and temperature and was used to prevent environmental effects on the products.

All trays within the same time variable was tested in the same ascending order starting with T1-1 and ending with T1-8. The timestamp of T1-1 and T1-8 at FT was noted to be able to extract the data for the products from the database based on the time the products were tested.

3.2.4.2 Test 2: Humidity in First In First Out (FIFO) Buffer

The second test performed used 10 trays of products, all produced at the same line in ECA and on the same shift. Every second tray was marked with First In First Out (FIFO) and the others with Dry Room (DR) to have as even produced products as possible and to assure that there was no significant difference in temperature. All trays were placed in the center position of the hardening 2 process, as seen in figure 3.3. All trays were placed in the resource at the same time and removed after 6 hours. After they had cooled down to room temperature they were tested in FT and then placed either in the FIFO or the DR corresponding to their markings. After 7, 14 and 21 days all trays were tested in FT again. Similar to Test 1 the start and stop times were noted to enable data extraction from the database.

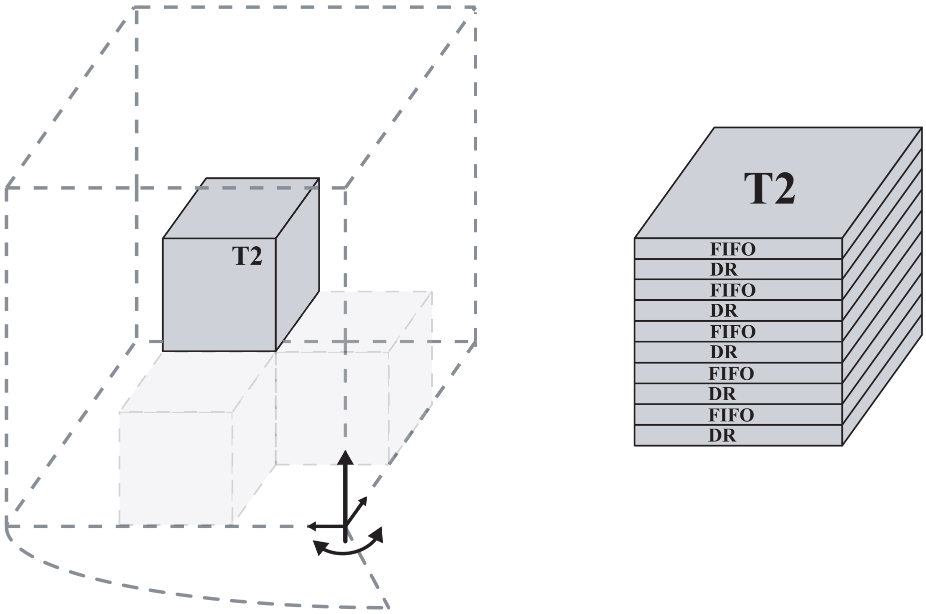


Figure 3.3: Test 2: Placement of trays in hardening 2 to the left and to the right each tray's placement in the position T2

3.2.4.3 Actions Taken to Simulate Normal Production

In a full running production the door is opened to insert new trays or to remove finished ones during the process time. Due to lower production then normal (see section 1.5 for more details), no new trays were finished to go into the hardening process during the test time, therefore it was simulated. To simulate a full speed running production the door of the resource was opened once every 90 minutes for approximately 2 minutes which approximately corresponds to a normal production.

3.3 Phase 3: Data Analysis

Analysis of data from the database was made in the statistical software Minitab, version 19. The first step in the analysis was to merge all data parameters into one file to make them comparable. For the products that had been tested in FT multiple times the last test result was used. For analysis over time for the products that had been tested multiple times a data-set that was not merged was used. The data collected from the production test was also analyzed with Excel.

3.3.1 Descriptive Statistics for Comparison of Data Sets

To find the mean, median, standard deviation, maximum and minimum both Excel and Minitab was used. Minitab can be used for bigger larger amounts of data than Excel and they were used respective to the amount of data that needed to be analyzed. Excel can handle up to approximately 1 million rows [41] while Minitab can handle 10 million rows [42].

Descriptive Statistics in Excel
To find the descriptive statistics for the two tests performed in production, Excel was used. The amount of data for Test 2 was about 900 rows and 10 columns making Excel easy to use for that amount of data. The data was sorted with a Pivot table, to enable sorting of different parameters and test runs. For descriptive analysis the add-in *Analysis ToolPak* was used which generated the following parameters (parameters marked with bold font were used for further analysis).

- **Mean**
- Standard Error
- Median
- Mode
- **Standard Deviation**
- Sample Variance
- Kurtosis
- Skewness
- Range
- **Minimum**
- **Maximum**
- Sum
- **Count**

The parameters mean, standard deviation, maximum and minimum were used to compare different data-sets to each other to find differences in spread. Test 2 used multiple test runs and two different buffer areas and was therefore compared between runs and buffers. To determine the change between each run the relative change was used. They were then compared to each buffer to find differences. For test 2, the time in the buffer was also analyzed by comparing the changes over time in order to determine if the results changed. The mean and standard deviation was compared to find differences of the spread and the distribution between runs. When errors occurred in FT, the parameter was given the value -9999 so maximum and minimum was used to filter them out to have an accurate mean value.

Descriptive Statistics in Minitab
For data extracted from the database, Minitab was used for performing descriptive statistics analysis. In Minitab the function *Display Descriptive Statistics* was used. Minitab then generated the following parameters (bold font parameters were used for analysis).

- **N (sample size)**
- **Mean**
- Standard Error of the mean
- **Standard Deviation**
- **Minimum**
- **Maximum**
- Q1
- Median
- Q3

The parameters were used to identify extreme values and to determine how the data was positioned and spread. Sample size was used to determine the amount of parameters and was used for comparison between different categorizing parameters.

3.3.2 Distribution Tests of Data from Database and Tests

To understand how the data was distributed, a normality test was performed in Minitab with the statistical function *Normality test*. The test was performed on the FT measurement values.

3.3.3 Scatterplots, in 2D and 3D to Analyze Parameters

Scatterplots was the main analysis tool to compare and find patterns between parameters and how the outcome in the FT was affected. Scatterplots were created in Minitab with two or three parameters plotted against each other and grouped with either assembly Pass or Fail (P/F) or FT P/F. All parameters plotted against each other and their grouping parameters can be seen in Appendix B.

3.3.4 Line Plots for Data Analysis

Line plots were mainly used to show differences over time for the buffer test. The relative change in FT values were plotted in a line graph, showing one line for each buffer and one line showing the total change of all tested units over time.

3.3.5 Correlation Between Test Runs in Production

To compare and find correlation between the different test runs in Test 2, a correlation matrix was made with Excel add-in Analysis ToolPak. The ToolPak generated a matrix of how the parameters in the columns related to the parameters in the rows. To generate a plot presentation of the correlation, Minitab was used. Minitab also generated the same matrix as Excel did.

3.4 Phase 4: Data Evaluation

To evaluate the analysis result, different methods were used depending on the type of analysis performed. Parameters were compared against each other and plots were visually inspected.

3.4.1 Comparison of Descriptive Statistics

Differences found in mean and standard deviation between different runs indicated a change depending on the time from hardening to FT when evaluating the tests. When evaluating the data from database, statistics were evaluated to find different patterns over time. This was evaluated further as dividing different assembly stations and FT resources to find patterns or differences depending on the prerequisites.

3.4.2 Interpreting Scatter- and Line plots

To evaluate the scatter plots, spread and correlation between the parameters was evaluated. The categorizing parameter indicated if there was a pattern between the parameters on X- and Y-axes. If only failed units were present in one area of the plot that indicated a correlation between the parameters and could imply a possible parameter that has a negative effect on the outcome of the production. When plotting with FT value the Upper Control Limit (UCL) and Lower Control Limit (LCL) were used to identify what values that could be disregarded.

3.4.3 Evaluation of FT Value Change Over Time

The change in the FT values was compared to each other using the relative change between runs. This was to gain comparable results independent on the actual value from the test, which could differ depending on the unit of test. To see the total change over time of the relative change, a line plot was used to visualize the percentage decrease. Test run 4 was excluded from the relative change investigation due to that the FT was mistakenly performed in another resource, measuring values differently.

3.4.4 Interpreting the Correlation Values

The correlation was mainly investigated form run to run in Test 2 in succeeding order. All tests were not performed in the same FT resource leading to differences in values. The correlation matrix was then evaluated to see if the values were related to each other even though different stations were used. The table of correlation values was evaluated as a first step to see if it inclined a correlation.

3.5 Construction of Guidelines for Investigation of Production Data

The amount of available big data was too much to be fully analyzed in this thesis and every day more data is stored. Therefore, guidelines for the case company to follow in similar projects were created along with this thesis. The instructions summarize the above mentioned methods used and were made to enable a step-by-step guide that the case company can follow when using analyzes to predict production outcome. The guidelines were based upon the theory and the methods to enable a higher level of understanding and a quicker start-up for anyone who would start a data analysis on the same topic as this thesis.

The guidelines consist partly of questions to be asked to enable a more open mindset and to give a broader view of what can affect the production outcome and what does not. For the production personnel it can be hard to see beyond what is already known. Therefore, questions that the authors used to find the investigated parameters for the thesis were used for the guidelines to enable an open mind.

The guidelines were written in English and as a short guide, to enable a faster start for anyone wanting to duplicate the work methods. For better use of the statistical software Minitab, a summarized guide of the main functions used was also included in the guidelines.

4 Results

This section presents the results of the thesis in corresponding order to the thesis outline. The sections is divided according to the phases and methods.

4.1 Current State

In the current state, all available information was investigated to find what data that was available. A go-and-see tour in the ECA was also performed to enable a higher understanding of the processes and figure out what parameters can affect the products.

The production processes at the case company consists of a combination of line and process flow. The first part of the manufacturing consists of four parallel line flows that evolves into a process flow with parallel resources as can be seen in figure 4.1.

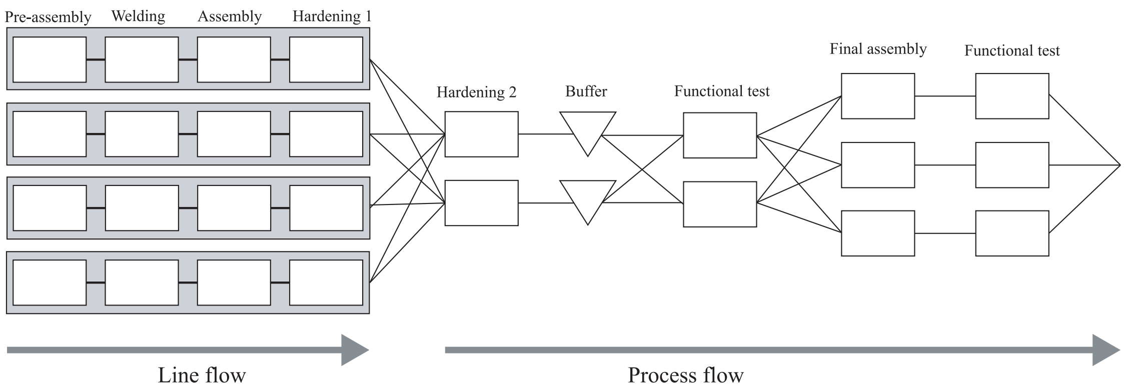


Figure 4.1: Production flow at the case company and the division of line and process flow

Each product is being processed according to figure 4.2, this thesis investigates up until the first functional test seen in figure 4.1. The final assembly is not part of this thesis since most scrap due to quality issues happens before the final assembly.. In an ECA the processes order is as follows: pre-assembly, welding, hardening and assembly. Due to the different production lines having different set-up and work processes, the values from parameters from the different assembly processes are not directly comparable. After leaving the ECA the product goes to a second hardening process for 6 ± 1 hour. When leaving the second hardening process the products are put into a FIFO buffer pending FT. If a product is failed in FT it can go back for re-testing for up to 18 days. For a more detailed description of the processes see Appendix C.

4.2 Data Collection

Data collection has been performed with multiple methods and the results of each method will be presented in the upcoming sections.

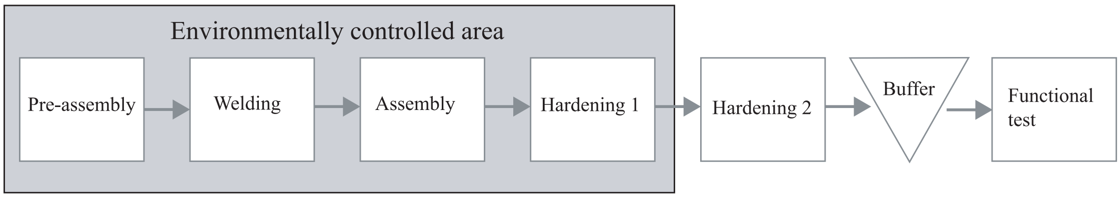


Figure 4.2: Process flow that each product goes through

4.2.1 Interviews and Meetings

Interviews were performed to gather information with the purpose to investigate the possible errors in the production. Several people were interviewed to gather different views on the possible errors. The answers were used to create the survey seen in Appendix A.

4.2.2 Survey

The survey had a response rate of 41% and the consulted respondents positions were the following: *Production engineer*, *Global process owner*, *System test engineer*, *Process engineer for manufacturing processes* and *Area manager*. All respondents answered that *time from hardening 2 to FT* will *Definitely have effect* on the production outcome. The second most chosen parameter was *Time in hardening 2*, followed by *Variation in movement in assembly*. The 5 highest rated parameters according to the consulted are presented in figure 4.3. All answers from the survey can be seen in Appendix D.

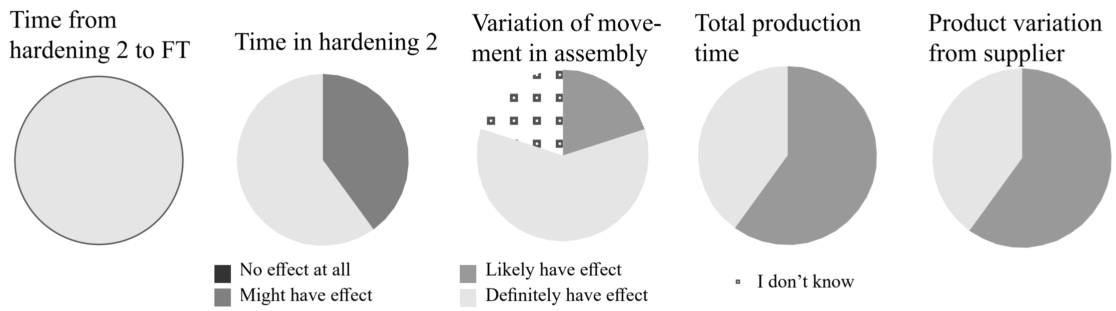


Figure 4.3: 5 highest rated parameters according to survey answers

Answers to the question *Are there any other parameters you think can affect the function test?*, can be seen in table 4.1. Any answers regarding production out of the scope has been excluded as well as answers containing confidential information.

4.2.3 Cause and Effect Diagram Before Data Analysis

The answers of the survey resulted in parameters interesting to investigate and were summarized in a fish-bone diagram that can be seen in figure 4.4. The parameters were divided into five categories according to different processes and areas of the production.

Table 4.1: Answers given to question 3 in the survey

Number for reference	**Are there any other parameters you think can affect the functional test?**
1	Different stations (functional testing stations) might not measure the same values
2	Fixture handling in Testing Equipment
3	Possible is variation in the adhesive from supplier
4	Capability of the active assembly equipment to accurately move component 3 to the desired pre-compensation location
5	The adhesive can continue to shrink over time and temperature

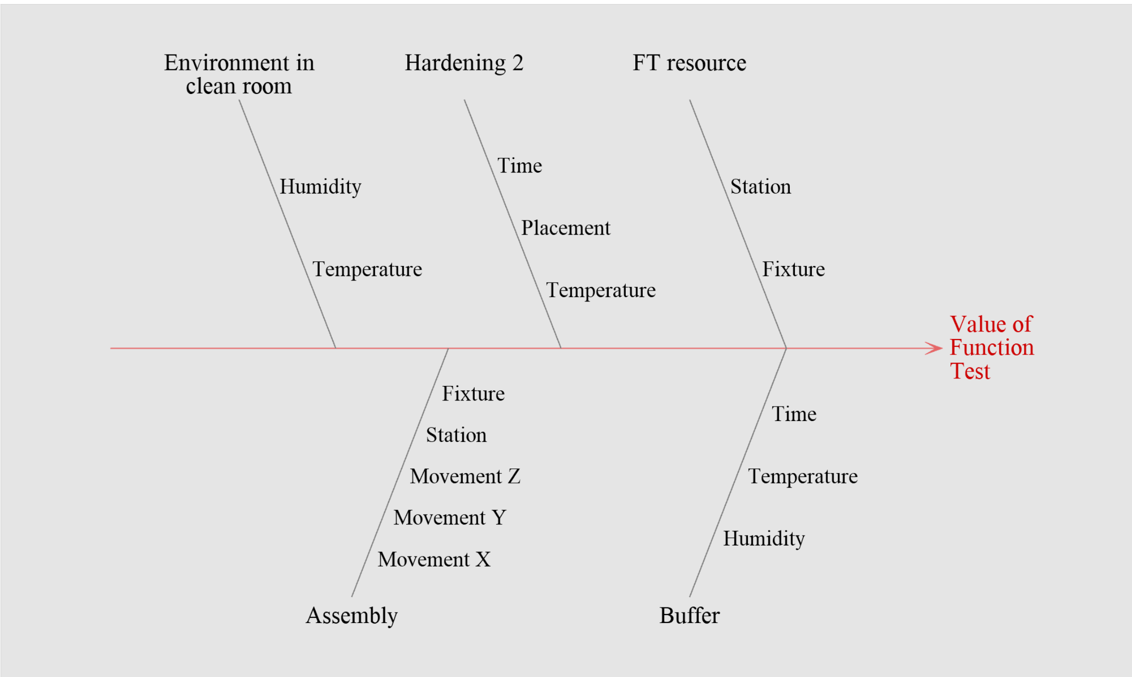


Figure 4.4: Fish-bone diagram of parameters that might affect the production outcome.

4.2.4 Gathering of Existing Data

Through the current state analysis and the interviews, interesting parameters were chosen, those available in the existing data were extracted. Approximately one year of data was extracted for the analysis. The highest rated parameters from the survey were chosen to have a higher priority in the data extraction. All parameters extracted from the database can be seen in Appendix E with their corresponding format and if they were used for further analysis. Data was collected for each Product-ID which generated 2 million rows and was stored in a .txt file.

4.3 Tests in Production

In the following section the results from the two production tests will be presented.

4.3.1 Test 1, Hardening Test

The hardening test had two areas of investigation, time and placement, they are presented separately below.

Placement Investigation
The temperature in the hardening process dropped approximately 9% when open for 2 minutes, the temperature drop was then regained in 1 minute. For the units that were in the process for 5 hours, the spread of FT values depending on the position are presented in figure 4.5. The units from the even numbered positions were placed closed to the door and as seen in 4.2 the standard deviation does not differ significantly. Units that was hardened for 7 hours can be seen in figure 4.6 and there the standard deviation of for positions differ. The even numbered positions have a standard deviation of $7,93$ while the odd numbered have $6,03$ as standard deviation. It was not possible to extract any results from the 6 hour test (see section 5.6.1).

Table 4.2: Descriptive statistics for 5 and 7 hours in the hardening 2 process depending on placement

		5H		**7H**	
Placement	Position number	Mean	StDev	Mean	StDev
Furthest from the door	1, 3, 5, 7	70,91	2,69	80,50	6,03
Closest to the door	2, 4, 6, 8	70,84	2,23	78,83	7,93
All placements	1-8	70,87	2,46	79,66	6,98

Time Investigation
The differences of the spread between the 5 and 7 hours in hardening can be seen in figure 4.7 with their corresponding positions. The 7 hour units generally had a higher FT value than the 5 hour units. This can also be seen in the differences of the mean values in table 4.2, where the mean value of 5 hour units was $70,87$ compared to $79,66$ for the 7 hour units.

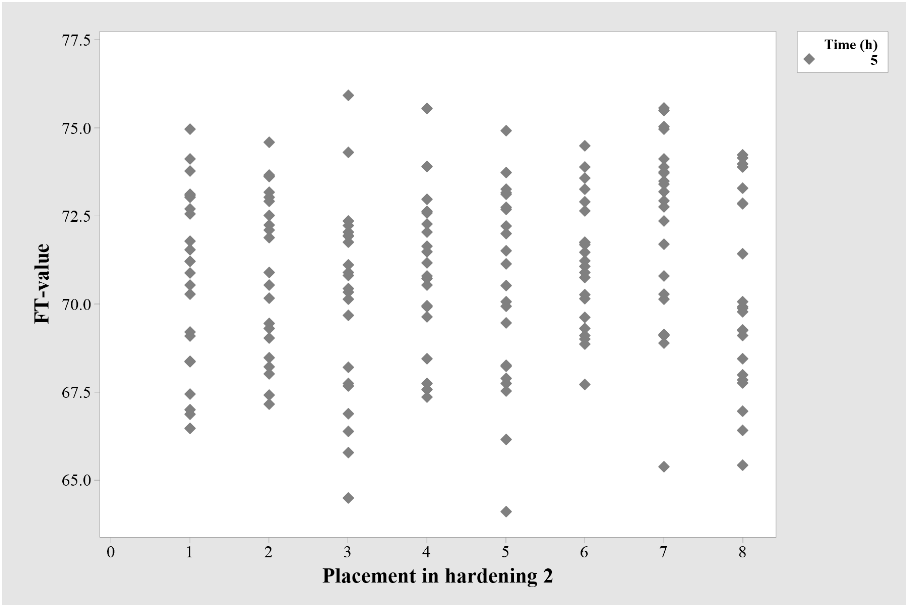


Figure 4.5: FT value depending on the placement in the hardening 2 for the units hardened for 5 hours

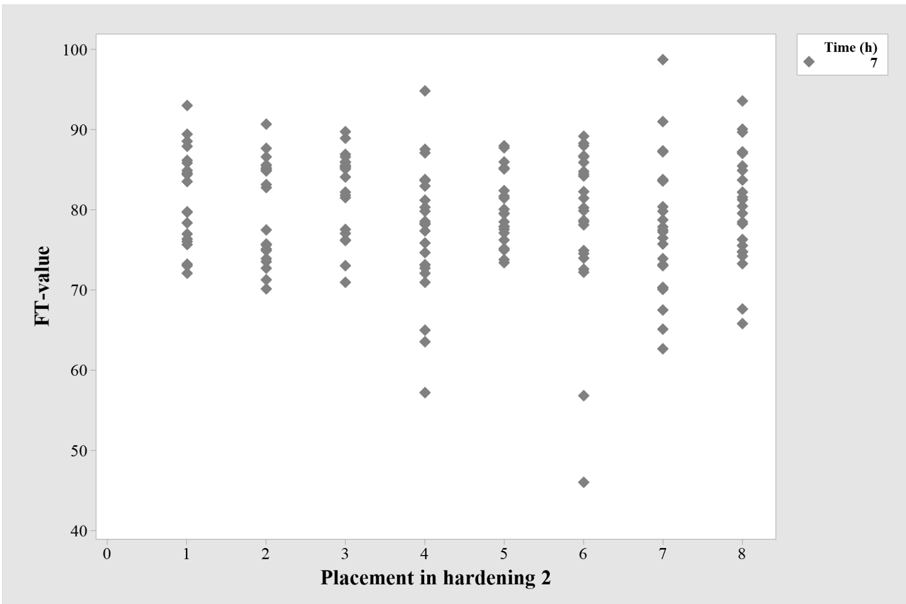


Figure 4.6: FT value depending on the placement in the hardening 2 for the units hardened for 7 hours

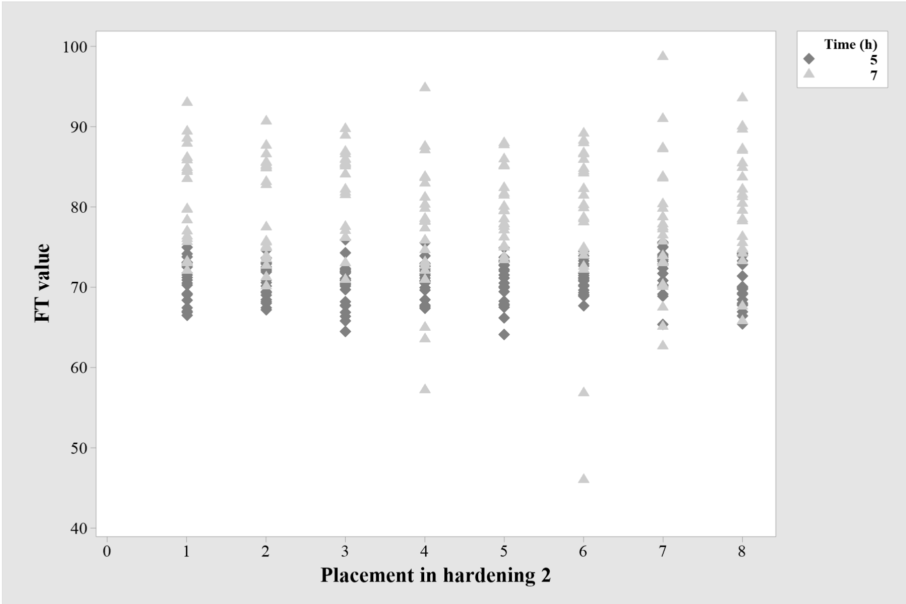


Figure 4.7: All units' placement in the hardening 2 process and their corresponding FT value

4.3.2 Test 2, Buffer Test

Firstly, results for each individual buffer are presented and then the combination of them are presented. When presenting the combination of the buffers, results of time aspects are shown. The differences between the buffers were humidity and temperature, which can be seen numerically in table 4.3 and graphically in appendix G.

Table 4.3: Humidity and temperature data for DR and FIFO buffer

	Buffer	Mean	StDev
Temperature [°C]	DR	28,9	1,2
	FIFO	19,2	0,5
Humidity [%rH]	DR	6,8	2,3
	FIFO	44,4	3,4

DR Buffer

The correlation between the four test runs of the units stored in the DR can be seen in figure 4.5. To see the graphic presentation of the correlation, see appendix F. The correlation between run 2 and 3 was the highest with 0.973 and the lowest was between 1 and 4 which was 0.596. The average FT value from test run 1, 2 and 3 for the units in the DR buffer was $70,2$ and the standard deviation was $3,5$. All values for the descriptive statistics can be seen in table 4.4.

FIFO Buffer

As for the DR buffer the highest correlation was between run 2 and 3, here with a

Table 4.4: Descriptive statistics for each test run and buffer

		Test run				
	Buffer	1	2	3	4*	Total of 1-3*
Mean	DR	72,2	69,5	69,0	73,8	70,2
	FIFO	71,6	69,9	68,9	80,1	70,1
Standard Deviation	DR	2,3	3,5	3,7	9,0	3,5
	FIFO	2,4	2,9	3,2	5,9	3,1
Minimum	DR	63,9	59,2	57,1	39,3	57,1
	FIFO	64,3	60,7	59,8	65,8	59,8
Maximum	DR	76,1	75,4	75,3	95,6	76,1
	FIFO	75,6	75,1	75,1	94,4	75,6
Count	DR	115	120	113	101	348
	FIFO	116	118	117	103	351

*Test run 4 was performed in a different FT resource

Table 4.5: Correlation values between test runs for units in DR buffer

	Test run 1	Test run 2	Test run 3	Test run 4
Test run 1	1			
Test run 2	0,807	1		
Test run 3	0,815	0,973	1	
Test run 4	0,596	0,896	0,914	1

value on 0.980 and the lowest between 1 and 4 with the value 0.772. All correlation values between the runs are seen in table 4.6, correlation plots are seen in appendix F. In the FIFO buffer the measured FT values can be seen in table 4.4. The average value from the three first test runs was $70,1$ with a standard deviation of $3,1$.

Table 4.6: Correlation values between test runs for units in FIFO buffer

	Test run 1	Test run 2	Test run 3	Test run 4
Test run 1	1			
Test run 2	0,892	1		
Test run 3	0,855	0,980	1	
Test run 4	0,772	0,930	0,949	1

Differences in Values Between Buffers

The difference between the buffers' mean values was $0,1$ and the difference in standard deviation was $0,4$. The difference also means that the FIFO buffer's mean corresponds to $99,9\%$ of the DR buffer's mean value and the standard deviation corresponds to 88% of the DR.

Buffer Tests Over Time

The average change of FT value, depending on buffer, is presented in figure 4.8. The difference between test run 1 and test run 2 was in average a decrease of 3% . Test run 3 had a further decrease of 1% from the previous test.

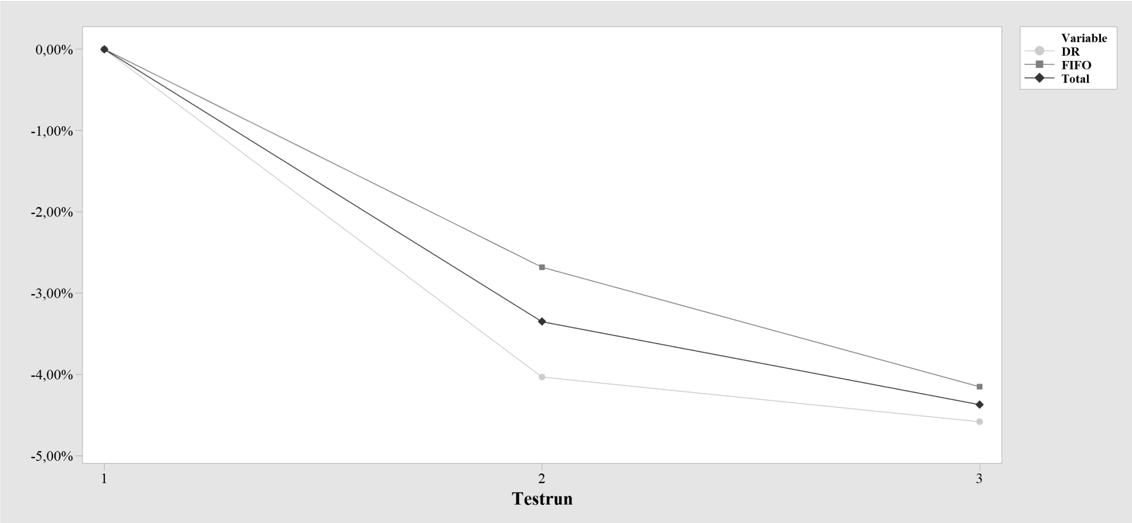


Figure 4.8: Relative change over time of FT value in the buffers and the total average change

4.4 Data Analysis and Evaluation

Presentation of the parameters investigated will be in the order of the production flow, starting from raw material and ending up in the finished component. Firstly, the summarized descriptive statistics will be presented for the FT values.

4.4.1 Distribution test of FT value

In figure 4.9 the result of a Anderson-Darling normality test can be seen for the FT value. From the plot and with a p-value $< 0,005$ it can be seen that the data is not normally distributed. Normality tests and plots for Align X and Align Y can be seen in Appendix H which shows that they are also not normally distributed.

4.4.2 General Descriptive Statistics

Data for the whole year was analyzed and divided in to assembly level and FT level.

Assembly Level
The different production lines are set up in different ways, so it was important to investigate the differences for failed units coming from the different production lines. Table 4.7 shows that there was a slight difference between the production lines, PL4 had approximately 2% more units failing in assembly and almost 3% more in FT.

FT level
Descriptive statistics for FT values, depending on the FT resource, are presented in table 4.8. The percentage of units, passed, failed and total for each FT resource

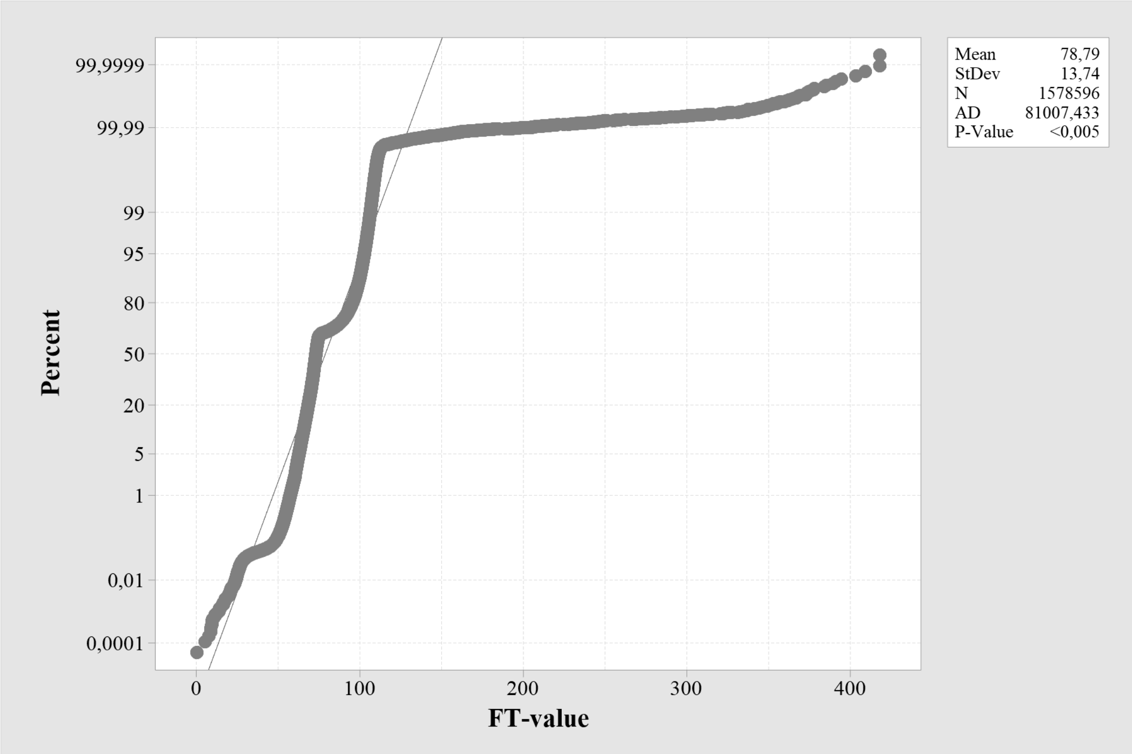


Figure 4.9: Normality test plotted for FT-value

can also be seen in the table. It was clear that most failed units came from resource 1, 65% of the failed units. In total 6% of the units failed during the time period April 2019 to April 2020. During the year the case company has made some improvements and changes, therefore has data for the two first months of 2020 also been summarized and can be seen in table 4.9. During 2020 fewer units have been tested in FT resource 1, only 16% of the total units compared to 21% of the units for the full year of data.

Table 4.7: Percentage of failed units in Assembly and FT depending on Production Line

Failed Units	PL3	PL4
Assembly	6.04%	8.37%
FT	4.86%	7.64%

4.4.3 Material Variation

In order to be able to investigate how the different parameters might affect the end result, there also needed to be an investigation of how the incoming material affected the different steps of the production.

Assembly Level

As presented in section 4.1, data from the different assembly lines are not directly

Table 4.8: FT-value statistics for each resource and total during one year

FT resource	FT status	N	% of N Total	Mean	StDev	Min	Max	% of all FT resources
1	F	58405	17,9%	61,1	5,5	0	231,0	65%
	P	267863	82,1%	66,2	3,0	60,0	82,1	18%
	Total	326268		65,3	4,1	0	231,0	21%
2	F	30816	5,0%	81,4	25,2	4,9	418,1	34%
	P	584222	95,0%	94,4	7,9	60,0	149,0	39%
	Total	615038		93,9	9,5	4,9	418,1	39%
3	F	1159	0,2%	63,9	11,1	8,5	112,7	1%
	P	643877	99,8%	71,4	2,4	53,2	81,8	43%
	Total	645036		71,4	2,5	8,5	112,7	41%
All FT resources	F	90380	6%	66,8	16,9	0,0	418,1	
	P	1495962	94%	79,5	13,2	53,2	149,0	
	Total	1586342		78,8	13,7	0,0	418,1	

Table 4.9: FT-value statistics for each FT resource during the 2 first months of 2020

FT resource	FT-status	N	% of N total	Mean	StDev	Min	Max	% of all FT resources
1	F	8012	17,6%	60,8	6,5	15,0	202,4	71,2%
	P	37502	82,4%	66,6	3,8	60,0	82,1	14%
	Total	45514		65,6	5,0	15,0	202,4	16%
2	F	3152	3,8%	73,9	28,8	4,9	311,5	28%
	P	80655	96,2%	88,4	8,5	60,1	148,8	29%
	Total	83807		87,8	10,4	4,9	311,5	29%
3	F	94	0,1%	70,7	11,7	32,8	112,7	0,8%
	P	156992	99,9%	71,6	2,7	54,0	78,9	57%
	Total	157086		71,6	2,8	32,8	112,7	55%
All FT resources	F	11258	3,9%	64,6	17,3	4,9	311,5	
	P	275149	96,1%	75,9	9,8	54,0	148,8	
	Total	286407		75,4	10,4	4,9	311,5	

comparable and were therefore plotted separately against the batches to show how the incoming material affected the assembly process. Figure 4.10 shows an example of how the different alignment values in the different camera lines differ depending on the batch. There were some clear patterns that indicated a correlation between batch and a larger range of the values. See figures I.1-I.3 in appendix I for further examples.

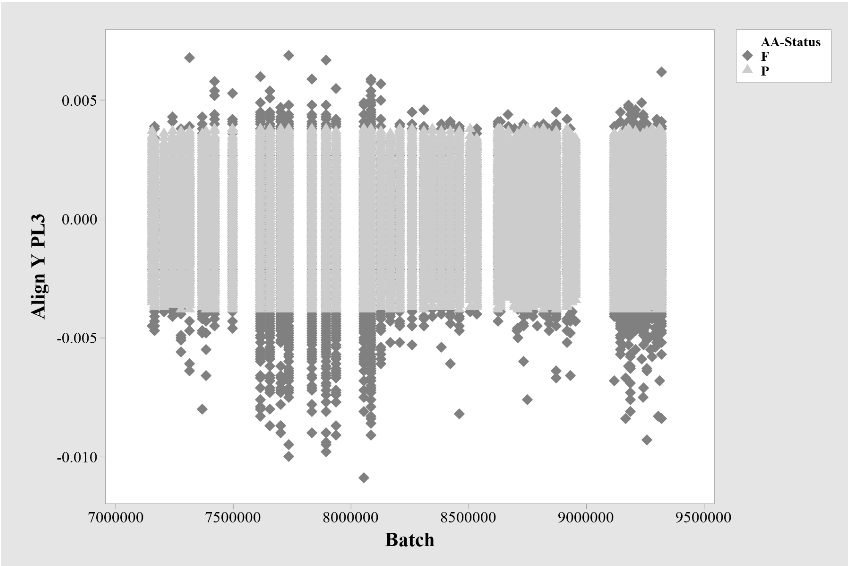


Figure 4.10: Batch with align value in Y-direction and grouped by P/F in PL3 assembly

FT Level
In order to evaluate if the material variation also had an affect on the final component, the FT value was plotted against batch number and can be seen in figure 4.11. There was a clear difference in the spread depending on batch but not in assembly line.

4.4.4 Temperature and Humidity in ECA

For the analysis of temperature and humidity in ECA the data was plotted against both time of assembly as well as FT value which can be seen in the following two sections.

Temperature
To see if the temperature affected the assembly results, temperature (Y-axis) was plotted over time (X-axis) grouped with the P/F results as can be seen in figure 4.14. When comparing the range of P/F, the graphs were more or less identical and there was no indication of temperature being the cause for units to fail.

Further, to see if the temperature affected the end-result of the finished component, temperature was plotted against FT value and grouped by FT P/F as can be seen in figure 4.13. When disregarding the values outside the allowed FT limits

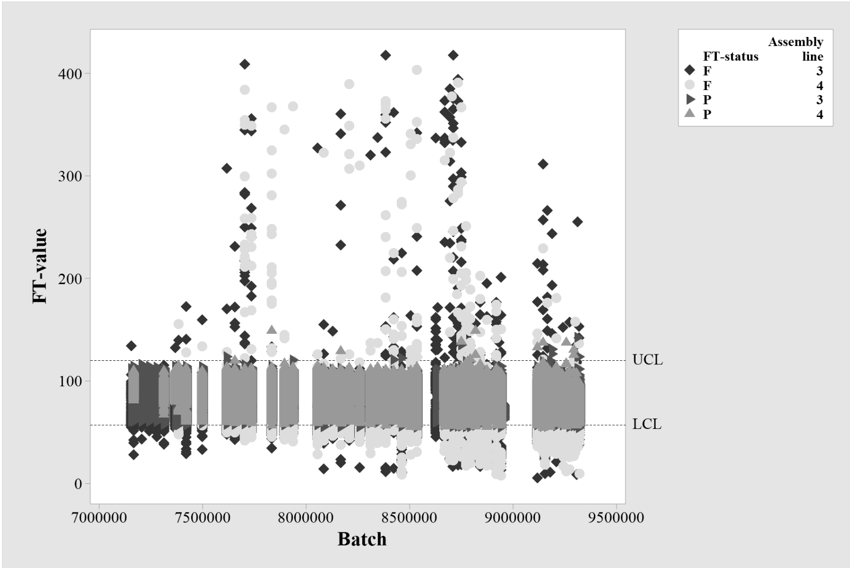


Figure 4.11: FT value against batch grouped by FT status P/F and assembly line

(LCL, UCL), P/F range was more or less evenly spread over the different temperature levels. No connection could be made between failed units and temperature.

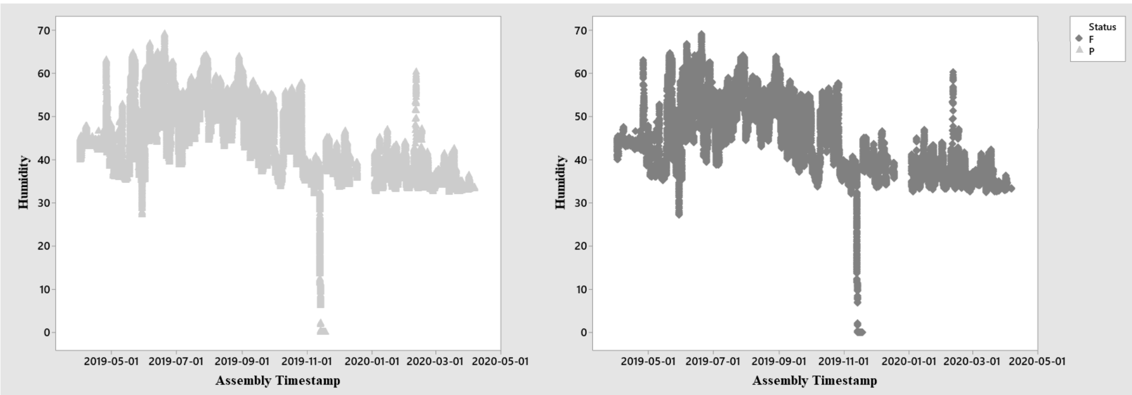


Figure 4.12: Humidity and time of assembly in ECA with passed units on the left and failed units to the right

Humidity

As with temperature, humidity (Y-axis) was plotted over time (X-axis) and grouped by assembly P/F as can be seen in figure 4.12. The range of P/F was more or less identical with temperature therefore, no connection could be made with humidity and failed units.

To investigate if the ECA humidity had any effect on the final result of the component, it was also plotted against the final FT value. Figure 4.15 shows humidity on the X-axis and FT value on the Y-axis and when disregarding the values outside LCL and UCL the results were spread evenly and no connection could be made between the ECA humidity and failed units in FT.

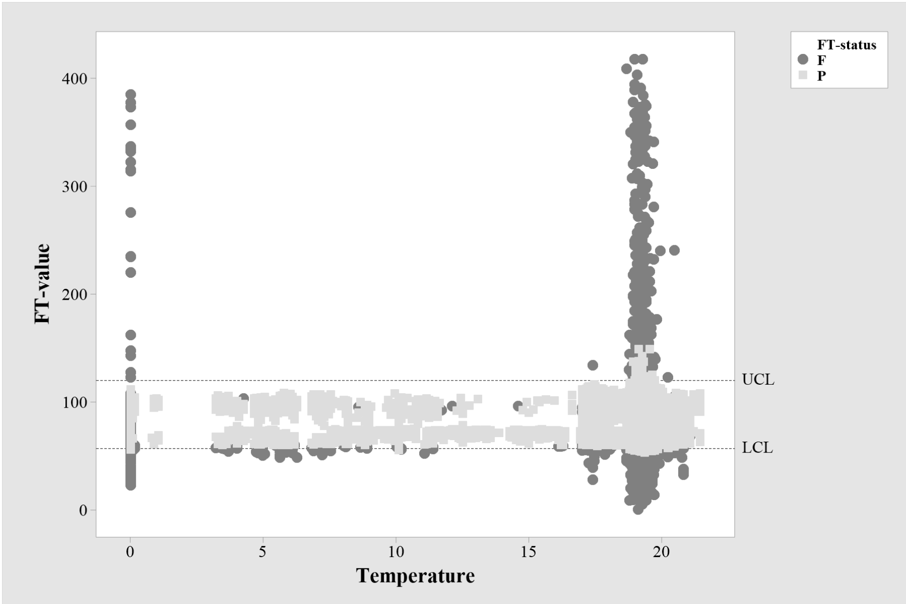


Figure 4.13: ECA Temperature with FT value grouped by FT P/F

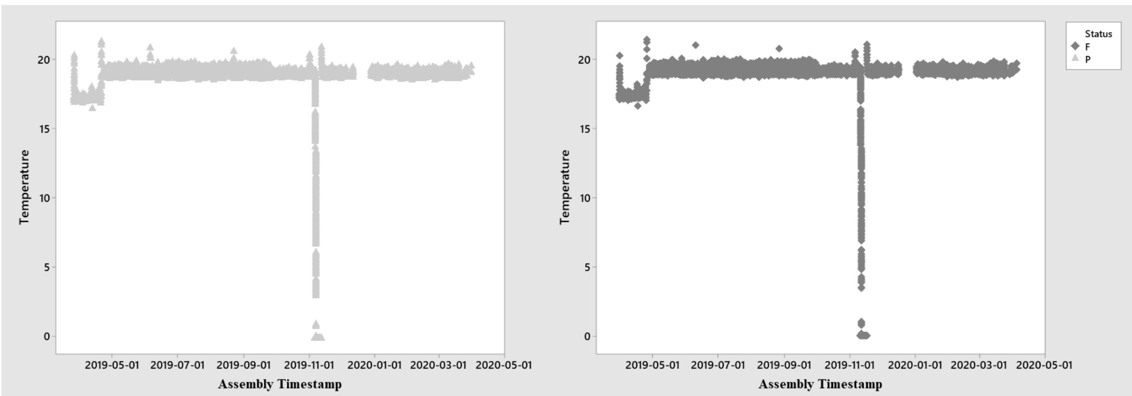


Figure 4.14: Temperature and time of assembly in ECA with passed units on the left and failed units to the right

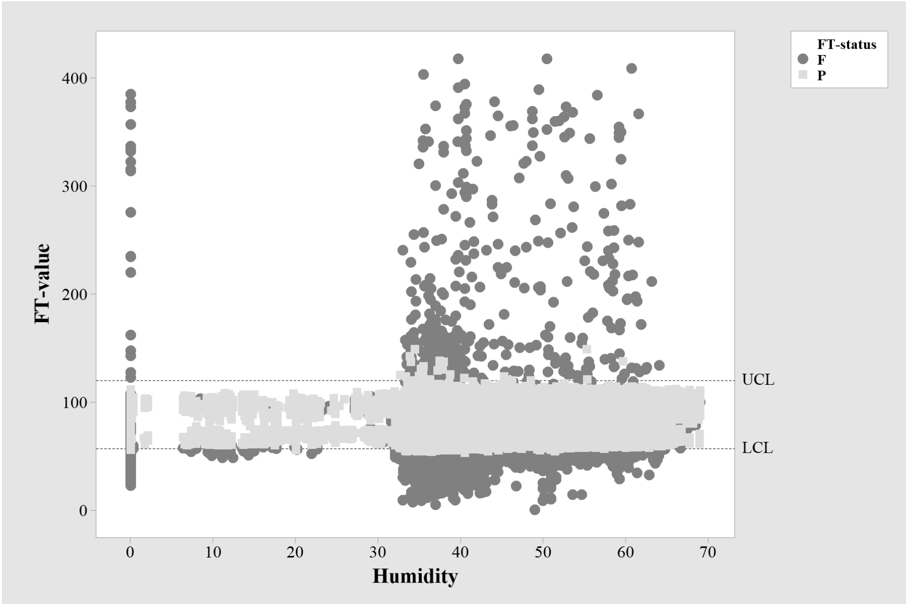


Figure 4.15: ECA Humidity with FT value grouped with FT P/F

4.4.5 Fixture

As mentioned in section 4.1, the assembly is performed in four different production lines. The component investigated was primarily produced in Production Line 3 (PL3) and Production Line 4 (PL4). Therefore, only fixtures from Production Line 3 (PL3) and Production Line 4 (PL4) were investigated. Fixtures were compared both directly in the assembly process against the align values as well as in FT level to see if there was any effect on the final component.

Assembly Level

Each production line has two assembly machines, the first two numbers in every fixture code represents production line and machine. Figure 4.16 shows an example of how the different fixtures in PL3 affects the align values during assembly. There was a very clear difference between the two machines in PL3 where four out of five fixtures had a much larger range in the align values than in the first machine.

When it comes to PL4, the pattern was not as clear as in PL3, there was rather a difference between fixtures instead of between machines, this can be seen in figure 4.17. When looking at the total range of the align values in the different machines, (see table 4.10) the range difference for PL4 was only 15% but differed with 36% in PL3. Same pattern could be seen for Align X in PL3 and Align Y in PL4 when disregarding outliers, graphs showing this can be found in appendix J.

In order to see not only what the range of alignment values were for the different fixtures and production lines, percentages of failed units were plotted in a bar chart. The time periods chosen for this was April 2019 to April 2020 as well as January 2020 to February 2020. This, to see if any changes made to the machines during the year had any changes in the effect of the different fixtures. As can be seen in figure 4.18, the fixture that had the highest fail percentage for the entire year (fixture 4219), had no data for 2020 since it was decommissioned. However, some of the

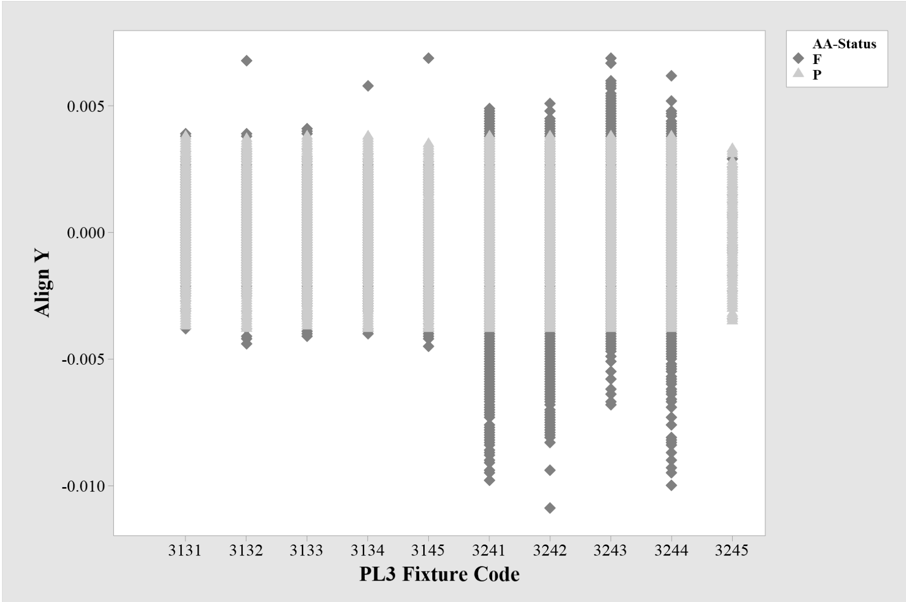


Figure 4.16: Fixture with align value in Y-direction and grouped by P/F in PL3 assembly

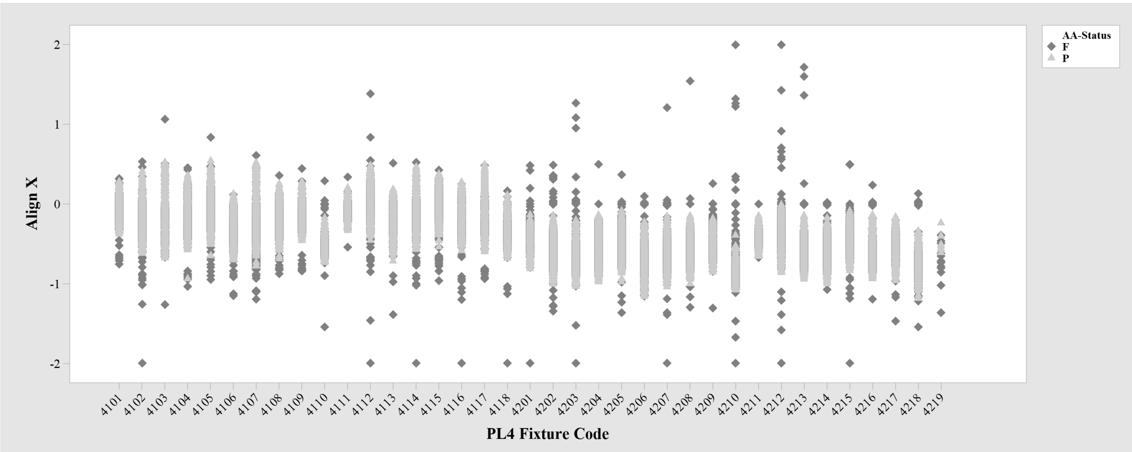


Figure 4.17: Fixture with align value in X-direction and grouped by P/F in PL4 assembly

Table 4.10: Alignment range for the different machines in PL3 and PL4

Production Line	Machine	Range Align Value
3	1	0.011
	2	0.018
4	1	3.384
	2	4.000

fixtures with high fail percentage, as for example fixture 3245, 4110, 4118, 4210 and 4218 were still in use.

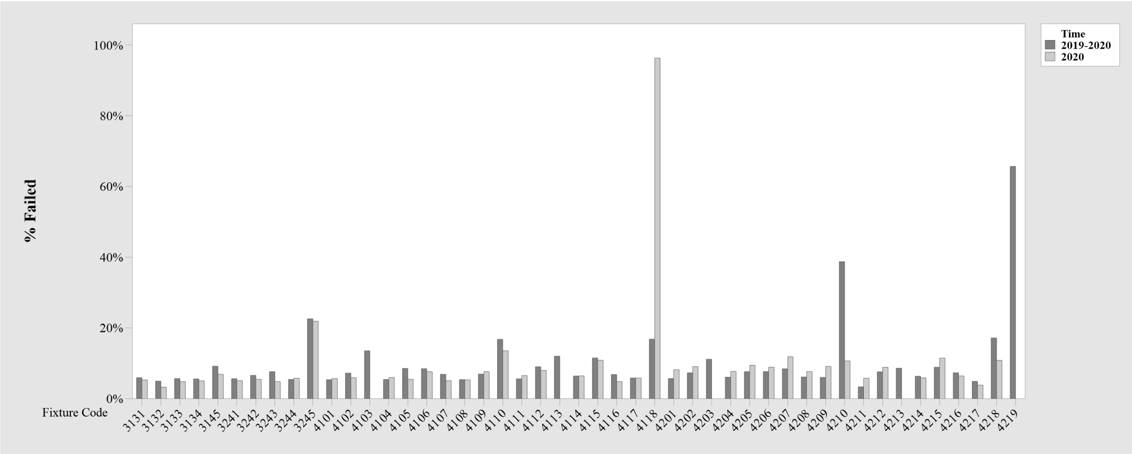


Figure 4.18: Percentage of failed units over time where 100% is all the units that has been assembled for the given time period

FT Level

Out of the units that passed the assembly process, investigations were made to evaluate if the fixtures in the assembly also affect the final FT value of the components. To see this, FT value was plotted against the fixture codes which can be seen in figure 4.19 and 4.20. No clear pattern could be seen here, those fixtures that were used the least (3145, 3245 and 4219) had the least spread FT values, so no connection could be established between assembly fixtures and FT values.

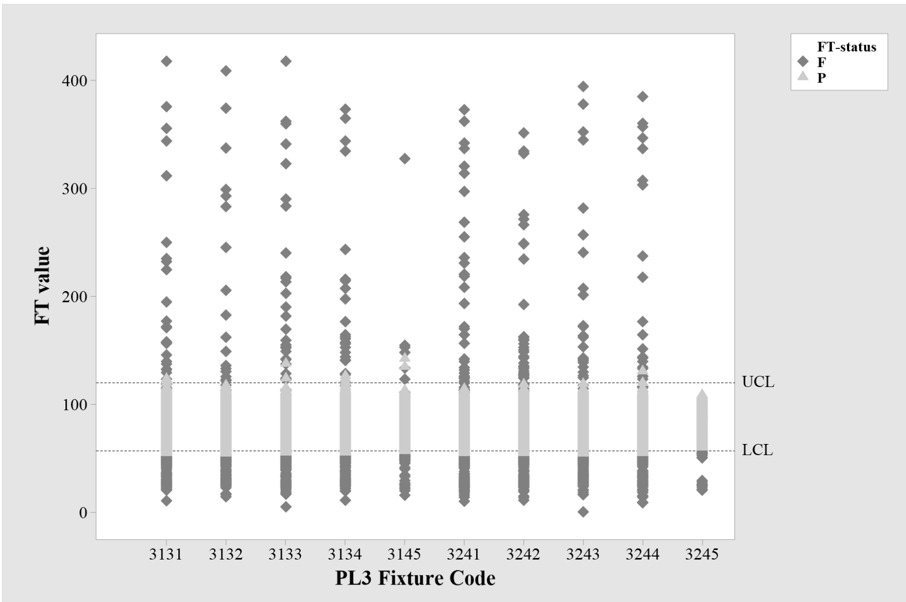


Figure 4.19: Fixture with FT value and grouped by FT-status P/F for PL3

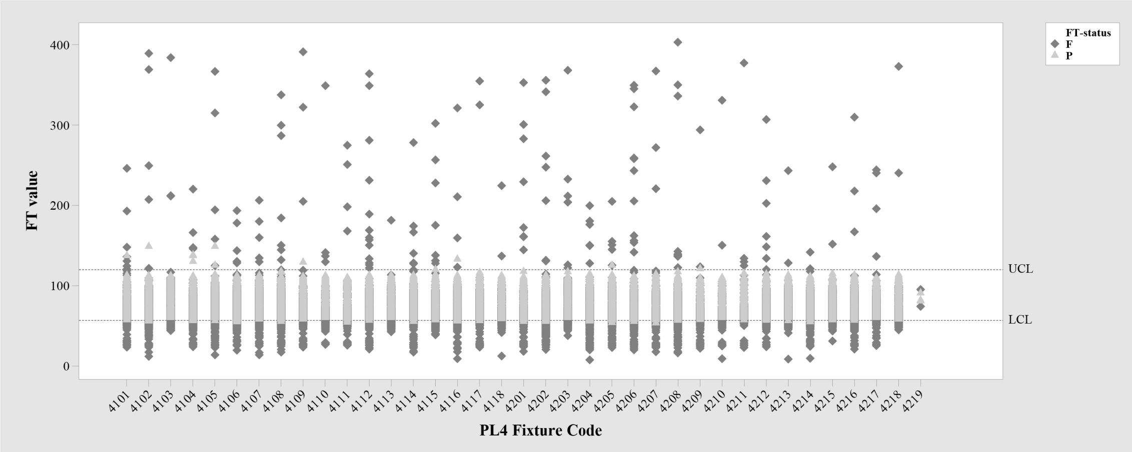


Figure 4.20: Fixture with FT value and grouped by FT-status P/F for PL4

4.4.6 Variation of Movement in Assembly

In previous sections, the alignment values have been used to show how other parameters affects the assembly process. Here, the effect of the alignment values on assembly P/F and the FT value of the final component is presented.

Assembly Level

The alignment process in the different production lines were set up differently. PL4 had a preconfigured path it followed and PL3 had a free path searching process which can be seen by the difference in the P/F pattern in figure 4.21 and 4.22. Those figures also show that units with large alignment values failed in the assembly process.

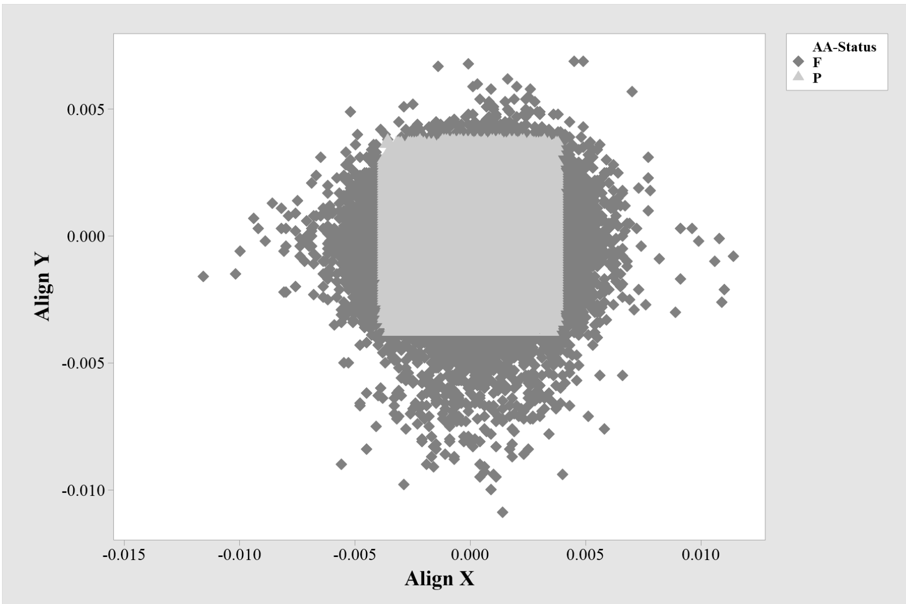


Figure 4.21: PL3 Align X vs Align Y grouped by Assembly status P/F

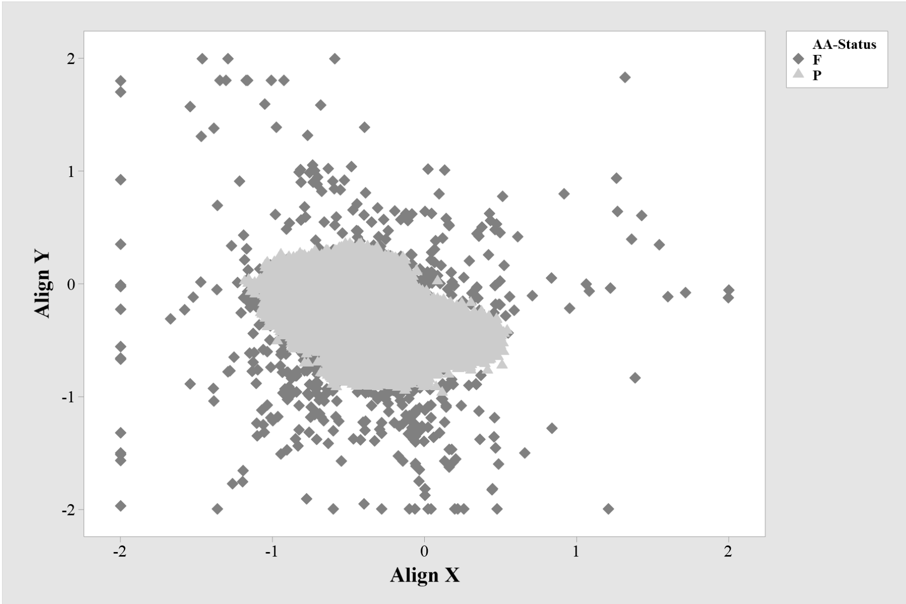


Figure 4.22: PL4 Align X vs Align Y grouped by Assembly status P/F

FT level

In order to see how the variation of alignment values affected the final component, Align X and Align Y were plotted in a 3D scatterplot against FT value for the units that passed in the assembly process. As mentioned in section 4.1, PL3 and PL4 are not directly comparable so they were plotted in different graphs.

As can be seen in figure 4.23 the failed units were spread evenly along with the passed units. The higher FT values that had failed did not follow any specific pattern and therefore no connection could be made to variation of movement in assembly.

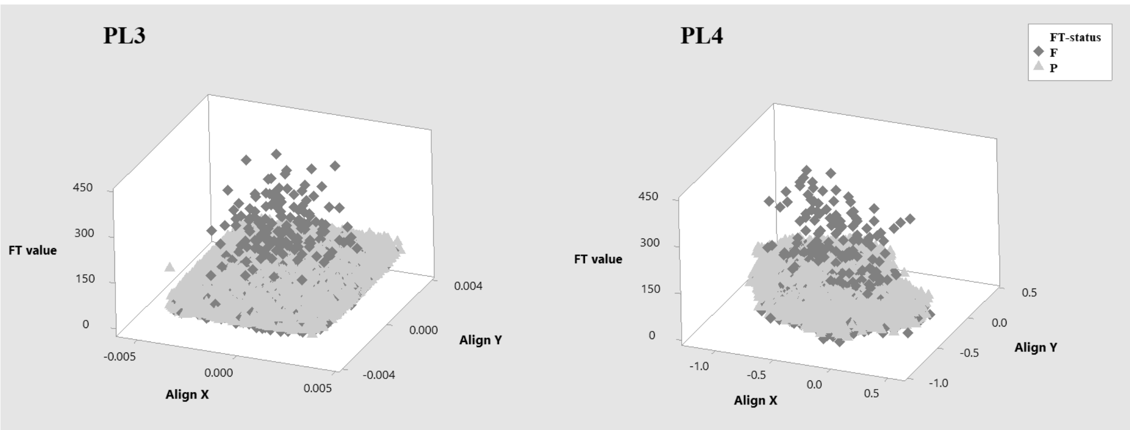


Figure 4.23: Align value in X-direction and Y-direction with FT value grouped by FT P/F for PL3 and PL4

4.4.7 Time from Assembly to FT

In order to see if variation of time from assembly to FT had any affect on the final component, FT value was plotted against *days from assembly to FT* as can be seen in figure 4.24. When disregarding values that fell outside LCL and UCL, no pattern could be seen between *time from assembly to FT* and FT value since P/F was spread evenly across the graph.

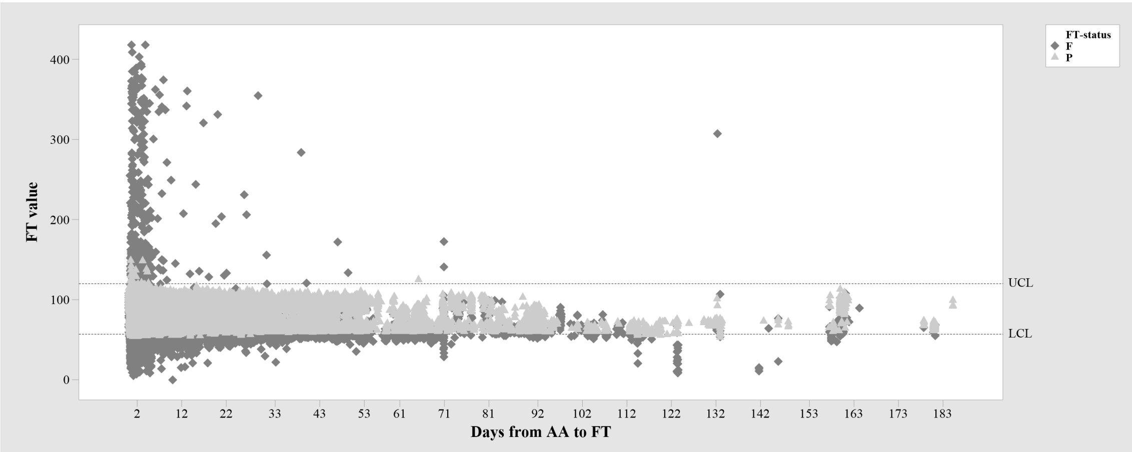


Figure 4.24: Days from assembly to FT with FT value grouped with FT P/F

4.4.8 Function Test Resource

There are three FT resources as mentioned in section 4.1. In order to see if there was any effect on the FT value for the final component depending on which resource the components were tested in, the FT value was plotted against FT resource. FT2 had the largest range and FT3 had the smallest range as can be seen in figure 4.25.

The same pattern could be seen when looking at the range value in table 4.11. When looking at how many units that failed in each resource, there was a clear difference with FT3 having the lowest percentage of failed components (see table 4.11). There was a clear difference in the performance of the different FT resources.

Table 4.11: FT resource range and percentage failed units

Resource	Range	Failed units
FT1	231	17.9%
FT2	413	5.0%
FT3	104	0.2%

4.5 Cause and Effect Diagram after Data Analysis

A fish-bone diagram created from the results of the data analysis and evaluation is presented in figure 4.26. Parameters concluded to not have any effect in analysis has been excluded in the diagram. Parameters that were not investigated enough are present in the figure but marked with an asterisk (*).

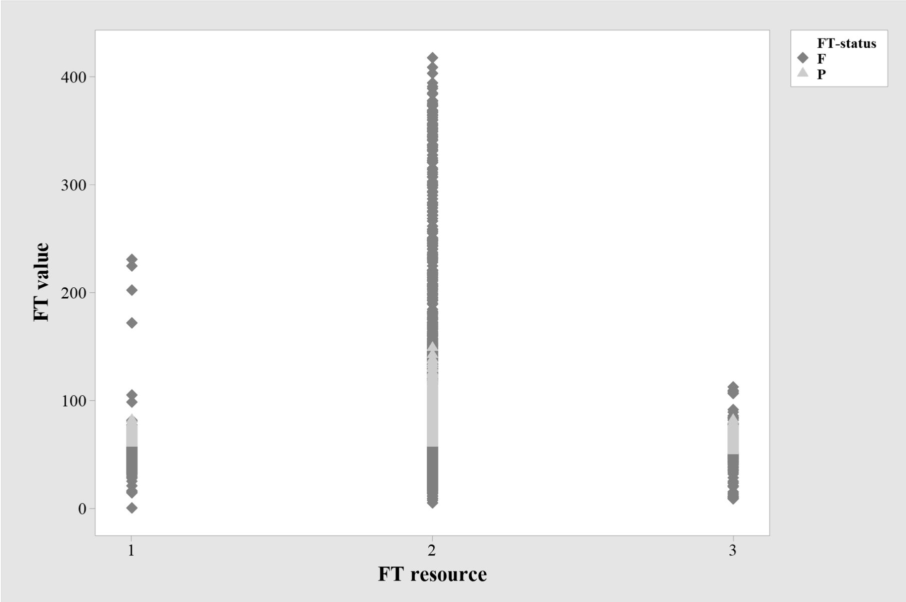


Figure 4.25: FT value in different FT resources

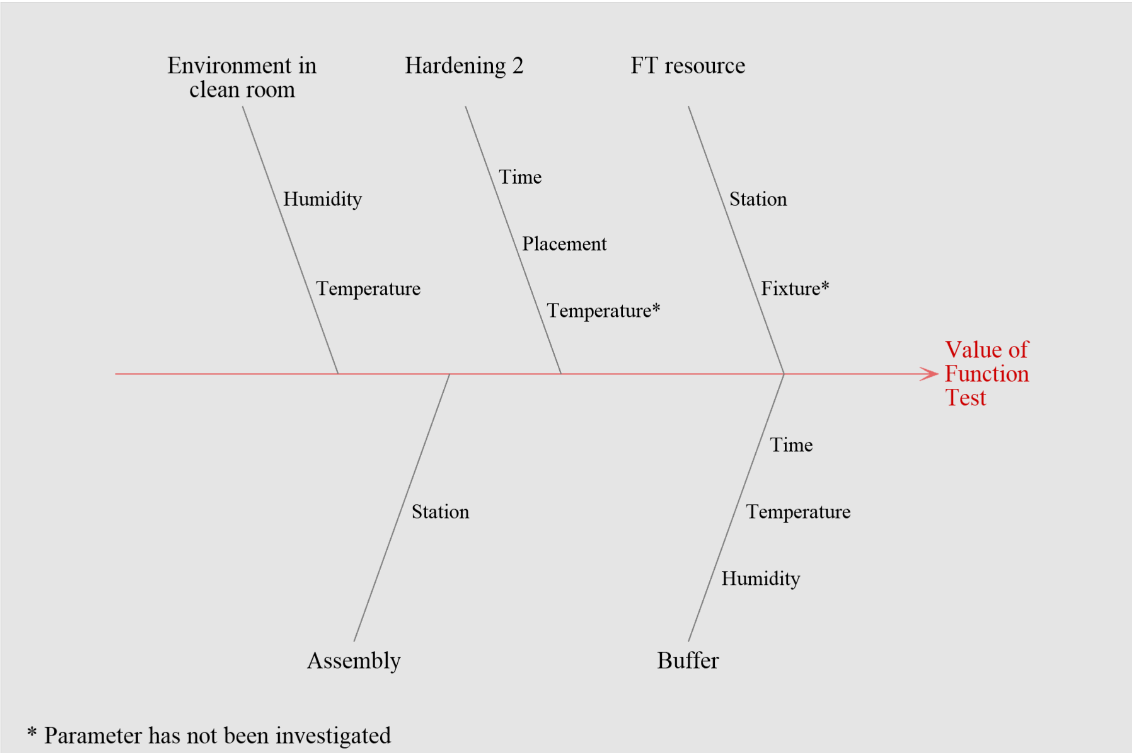


Figure 4.26: Fish-bone diagram on parameters that does affect the production outcome

5 Discussion

In this section, methods, results and interpretation of results will be discussed. The effect of the extraordinary circumstances will also be reviewed on how they interfered with the production tests and analysis of existing data. Lastly, the research questions will be discussed, whether they have been answered through the study or not.

5.1 Data Collection and Correlated Methods

Firstly, the choice of methods will be discussed for the data collection performed through interviews and survey. Secondly, the method for data collection from the database will be discussed. Data collected in the production tests will be presented together with its results in section 5.2. The data collection was performed according to the Define-phase in DMAIC described in section 2.4.2 to enable understanding of the current state. The phase enabled knowledge throughout the processes and defined cause-and-effect relations in a very good way, just as E. van Dijk described [21]. Through the different methods for data collection a triangulation could be made and different stakeholders were taken in consideration.

5.1.1 Survey and Interviews

To gain more useful and comparable information, more than five standardized questions would have been preferable. Even though it would have been hard to gain full comparability due to the conflict of interest, answers would have become more comparable. The interviews gave a broad view on how the production works and what parameters the product is affected by, which greatly helped the current state mapping. Through the interviews, a well fitted survey was able to be sent out and it broadened the view of possible problems for the respondents. Through the survey it was possible to both find parameters and prioritize them depending on the interest from the case company. In the survey more answers would have been preferable, only 41% of the respondents answered the survey after 3 reminders were sent out. This was a lower rate than expected but the responses were well spread among the different areas of interest which increased the credibility of the answers.

5.1.2 Data Extraction from the Database

For gaining knowledge of what parameters existed and to know how they were named in the database, multiple people had to be consulted. The risk of making a mistake was high due to the complexity of the stored data. In order to extract data from the database, a process engineer had to be consulted to extract the specific parameters. In the database, multiple parameters could have the same name which led to confusion. Different production lines could also have different names for the same parameter. This made the extraction harder than expected. Due to the inconsistent naming convention, there is a risk of investigating incorrect data. The consulted person who extracted the parameters was well familiar with SQL and how it works but not the specific parameters and their information. When parameters were chosen, the extraction was easily conducted and the files were well sorted.

When it comes to the use of big data [3] mentions that when more and more data started to be collected, existing analytical models could no longer be used in the same way. The company in this case study had not started using the data to its' full potential yet. Analysis made so far with the collected data was mostly to for example adjust some pre-set parameters in a process. However, no full scale analysis on how different parameters affect the product further then in the process that owned the parameter had been done. It became apparent that the lack of an overall knowledge of how the product was affected was not possible since there until recently had not been anyone department or responsible for understanding how data from different processes correlate in the end. This had lead to a lack of communication between the departments and overlapping issues had not been addressed.

5.2 Tests Performed in Production and the Results

For both test performed in the production, some results indicate that the production outcome can be effected by the the process that has a gap of data. They also show that some parameters does not affect the outcome in a negative way.

5.2.1 Time and Placement Test

This test had two different aspects of testing, time in the hardening process and the placement in the resource.

Time in Hardening 2 Process
There were clear differences in the mean values of the two different time parameters. For the units hardened during 5 hours the mean was 70.91 and for the 7 hour units the mean was 80.50. Both time parameters were within the LCL and UCL, no differences could be seen in P/F. The *Time in Hardening 2* was the second most guessed parameter to affect the production outcome in the survey. From the results of the test the conclusions can be drawn that the time does affect the FT value.

Placement in Hardening 2 Resource
When looking at the placement of the 5 hour units, no correlation can be seen between the different positions. No significant difference was seen either in the statistics or the plots. On the other hand when looking at the units of 7 hours there is clear difference between the units in the front and in the back. The standard deviation differs between 6.03 in the back and 7.93 in the front. This, can also be seen in the spread of values in the plot. For the units hardened for 7 hours, the door was opened more times than for the 5 hour units due to removing the previously finished units. This, means that they have been exposed to more temperature drops than the 5 hour units. The units furthest from the door was not as exposed to drops in temperature as the units in the front which can explain the big differences in spread of values. The test was only performed with 1 tray consisting of 24 units in each position per time parameter. In order to be able to draw definitive conclusions a more extensive test must be performed. However, this initial test indicates that the position does affect the FT value.

5.2.2 Choice of Buffer and Re-Test Time

In the test with different buffers, the goal was to look into how the environmental parameters at the current buffer area could affect the units while waiting or FT. The second goal was to see if the current re-testing time of 18 days is correctly defined or if the time frame should be changed.

Choice of Buffer
Both temperature and humidity differs a lot between the two buffers, as expected in this test. The idea was to see if the humidity could affect the results and outcome due to the moisture sensitive adhesive. As seen in the results, the two different buffers does not differ remarkably. The difference in the mean value was insignificant and the standard deviation barely differed. This means that no significant differences of the buffers could be seen. The humidity and temperature of the current buffer can therefore be concluded to not affect the FT value significantly in either negative nor positive way.

Re-Test Time
The case company has a policy to re-test failed units for up to 18 days after the first FT to see if the units will pass. However, this does not seem to be supported by the results of this thesis. The results showed that after 7 days the FT value decreases with 3% but after that it only decreased with 1% during the next 7 days which is not a significant change. The fourth test, 21 days, was mistakenly performed in a different FT resource and was therefore not comparable with the other test runs. With a correlation of 0.98 respectively 0.97 for the two buffers between run 2 and 3, the changes can be concluded to be very small and the re-test time of 18 days is not necessary. If any waiting time should be applied for re-test, 7 days is enough.

5.3 Data Analysis

The amount of data extracted was very large and due to the spread and non-normal distributed data, statistical analysis became hard and time consuming. Therefore, the use of scatterplots and visual representation of the data became a suitable method to identify relations between parameters. Due to the varying LCL and UCL levels in the different resources, scatterplots could also serve as an useful help for comparison between parameters. The historical data that was used for the analysis, enabled path analysis and made it possible to identify machines both individually and together that contribute to errors [19, 23].

5.3.1 Important Parameters Found in the Survey

From the survey there were five parameters that most of the respondents thought would affect the FT result of the units.
Those were:

1. Time from hardening 2 to FT
2. Time in hardening 2
3. Variation of movement in assembly

4. Total production time
5. Product variation from supplier

Parameter 1 and 3 were found to not have any noticeable effect on the FT value. The results in sections 4.4.6 and 4.4.7 support this claim. Parameter 3 does however have a very strong effect on what units pass or fail in the assembly process.

5.3.2 Important Parameters Found Through Analysis of Extracted Data

In section 4, the parameters were presented in process order. It was also interesting to investigate whether they had any effect on the assembly P/F, which also was presented in order to understand the process more profoundly.

Interestingly, there were some parameters that from the survey and interviews were expected to have an effect on the FT value but only had a direct effect on the assembly process. The only parameters that seemed to have an effect in both assembly and FT were Material Variation and Assembly Line.

Parameters that indicates an effect on the assembly process are:

- Material Variation
- Assembly
  - Fixture
  - Line/Machine
  - Movement X
  - Movement Y

As mentioned in section 5.2, the tests in the production show that time in buffer as well as time and placement in hardening 2 seem to have an effect on the FT value. However, even though these parameters seem to have an effect, most of the units that fail depend on which FT resource they are tested in.

Parameters that indicates an effect on the FT value are:

- Material Variation
- Assembly Station
- Time in Buffer
- FT Resource
- Hardening 2
  - Time
  - Placement

5.3.3 Possible Worst Case Scenario of Extreme Values

When combining the worst case scenarios for the different parameters it is interesting to think about how the results from the production tests can affect the outcome. The tests indicate that the FT value decreases approximately 3% over a time period of 7 days. They also indicate that the position and time in the hardening process affects the FT value. If a unit has a low value in itself and is then placed close to the door for only 5 hours in the hardening 2 process, the FT value will be lower than with the opposite prerequisites. If the unit also is tested directly after hardening 2, there is a great risk of the value decreasing even further. Since the units are tested after the final assembly, units with decreased FT value can still be found. However, more value has been added along the way. There has been no results indicating that the value would increase or be close to the UCL which is why the maximum value will not be further discussed. The only high FT values are from units tested in FT2. FT2 has a known mechanical problem meaning that no further analysis of maximum extreme values could be performed.

5.4 Guidelines for Companies for Similar Studies

PDs are not constant so it is important that the process is iterative and in order to have continuous improvement of the production. The guidelines can be followed in further studies both at the case company where the thesis was conducted or at any other similar project. A flow chart representing these guidelines can be found in appendix K

1. What PD is your focus?
2. Map which parameters that can affect this disturbance. Make sure to also consider parameters that data is not currently collected for.
3. Consult people with different areas of interest (Production managers, Operators, Product owners, Area managers, Process owners etc.)
4. Are the parameters available or are production tests required? Make sure to not disregard parameters due to preconceptions that they are not affecting the PD
5. Create a fish-bone diagram to indicate what and how the parameters affect the PD.
6. Make a plan for which parameters that are relevant to compare to each other. If there are several steps in the process, make sure to investigate how parameters affect all steps, if applicable, for example see section 4.4.5.
7. Extract data from the database and perform production tests if required.
8. Sort and process data, Minitab is a good tool for this since it provides an easy way to merge and gather data from different sources in to one project in order to facilitate analysis and evaluation of data. Minitab can also handle more data than for example Excel.
9. Analyze data according to your plan. Scatterplots and normality test are two good ways to understand the behavior or trends of the data.
10. Evaluate what the plots indicate and if necessary, make a plan for further tests and studies to understand the root cause of the PD (go back to step 9).
11. Be sure to present the findings in a way that enables actions to be taken in order to correct, prevent or eliminate the PD. This means that results might need to be presented in different ways to be relevant for all stakeholders.
12. Re-evaluate the PD after implementations are made to see if it can be further improved.

And lastly, don't forget to attack the next PD.

5.5 Recommended Continuous Work on the Study

To be able to integrate the information gathered through this study, further work needs to be done. Parameters that partly can be the cause of the high scrap rate has been found. Implementation of changes has not been performed to justify the effect of the found parameters. Without further tests it can not be concluded for sure that the parameters affect the outcome. More data should be analyzed and the available parameters should be mapped to ensure the actual value of them. Suggestions on further work and recommendations can be seen in section 6.

5.6 The Social Situation's Impact on the Thesis Work

During the spring of 2020 the Corona pandemic erupted and affected the social situation and the conditions of performing this thesis. The two tests performed in the production was affected by the furlough and were delayed or affected due to the circumstances.

5.6.1 Effects on the Hardening Test

The hardening test was only affected slightly. The units were supposed to be tested in FT as soon as they had cooled down but due to personnel shortage the testing was delayed and were tested on different days. The 6 hour units could not be analyzed due to data gathering issues. Timestamps for start and stop were noted when testing but when extracting the data with corresponding timestamps, enough units did not correspond to the noted time interval. Thus, the 6 hour units had to be excluded from analysis.

5.6.2 Effects on the Buffer Test

The buffer test had to be restarted due to personnel shortage. The authors could not be present at the re-test due to travel restrictions. Because of the inability to be present at the test, this part was not supervised by the authors and the test plan can not be confirmed have been followed perfectly. The test was restarted one week after the initial start and then performed according to plan as far as the authors are aware. All test runs were supposed to be performed in FT 3 but due to production disturbances the 4th test was performed in FT2. The resource that was used in the 4th test has a different measuring method and the results were therefore not comparable.

5.6.3 Effects on the Data Gathering and Analysis

It was very hard to get in touch with people during the furlough which affected the possibility to get answers on specific parameters' impact and how parameters were defined. The database extraction lacked unit information for the parameters which became very hard to get due to the furlough and lack on people available to answer.

5.7 Application of Theories in this Case Study

In section 2, theories to understand the framework of this thesis were presented. Theories regarding quality, big data and PD was presented along with statistical framework. In this section, the applications of these theories for the specific case company will be discussed and how they can be connected to the study's findings.

The definition of PDs used in this case study was quality affecting disturbances according to PD categorization [2, 13]. This study showed that quality issues as a PD can include *Rejection of products* as well as *Adjustments in machines*. Products were found to be rejected in multiple different steps and some of the errors could be avoided by adjustments in the machines or in the procedures. This concludes also that quality is an important parameter when describing a PD. This study clearly showed that depending on the resource used for function test the outcome was very different and can therefore be concluded to be a PD which is supported by what Ingemansson described in his 2004 article about PD reduction [2].

The case company had a lot of stored, non-analyzed data when this study began. Through the analysis, possible quality improvements were found to decrease PDs and to increase product quality. This concludes that big data can be used for identification of errors in both processes and product quality. The data was however poorly structured and the amount of available parameters was excessive which made it hard to know what data that should be used. This is in line with what Gordon et.al [20] also has shown and presented. The fact that quality can be improved with big data has also been shown to be possible in this study which Kuo and Kusiak also presented in their article about big data in production research [3].

For this specific case company's production site it has been shown to indicate that big data can be used for prediction of errors in order to decrease PDs and improve quality. The theories presented has been concluded to cohere with this study's findings. The theories also indicate that the issues found at this case company are common issues among other manufacturing companies. Through larger studies with multiple cases, a generalization could be achieved to conclude the results presented. The approach for analysis was based on DMAIC as well as simple statistical tools and can therefore be concluded to be applicable to other manufacturing companies for investigation of prediction with big data in further studies.

5.8 Research Questions

Each research question will be discussed in the following sections as well as the results in combination with the theory presented earlier.

5.8.1 RQ1 - Appropriate Procedure for Future Projects

An appropriate procedure was found in the form of questions that the authors asked in order to conduct the thesis. The questions could be generalized and made specific for data analysis of production disturbances instead of being specified to the case company. The questions were constructed to enable an open mind instead of asking directly about the current known state. The initial questions focuses on understanding the disturbance. The step for consulting people promotes an open mind and to make sure different views are taken in consideration, not only the investigating person's interests. Later, are steps related to this thesis in methods and analysis of data. When reaching any conclusions in data, it is important to remember to present the findings according to stakeholder's interest. If the findings are not presented with the stakeholder's interest in mind, projects can for example be hard to get economic backing or to get colleagues motivated for implementing changes.

5.8.2 RQ2 - Closed Loop of Data

The production is well covered with data except for the hardening 2 process and according to the consulted people in the interviews the general opinion seemed to be that it was inconsequential. The results of the survey showed that the placement in the hardening 2 process was not thought of as an impacting parameter. The same goes for the temperature drop when the door is opened. This thesis has shown that both placement and temperature might affect the test values which can indicate that the gap of data affects the outcome. To be be able to fully analyze and improve the production with the DMAIC process and to have a closed loop of knowledge and information the data also have to be a closed loop [21]. As seen in the literature presented in the theory section, one of the core features for Industry 4.0 is the interconnection between machines [26]. The interconnection enables a closed loop of data and through the gathered data, higher quality of products can be achieved [26]. In the gap at the investigated production there is no controlled information and this gap has lead to misconceptions on what happens there. When there is a gap of data it is not possible to perform full path analysis and the results can therefore be misleading in were the true problems lie.

5.8.3 RQ3 - Reasons of High Scrap Rate in PD

In this thesis the production disturbances were defined as quality affecting disturbances that occurs out of control and affects the production in a negative way. In the study the high scrap rate became very clear through high percentage of failed units presented in the results. To be able to correct, prevent and eliminate those quality disturbances the scrap rate must be decreased to a more acceptable level [10]. In this case company, rejected products becomes a very high rate in the scrapped products and therefore they are seen as a major disturbance. Major issues in the machines have been identified as a reason for the rejected units, especially in the FT resources the differences are clear. Settings or prerequisites in the resources differs and they also generate very different levels of scrap. Reasons for the high scrap rate as production disturbance can therefore be concluded to exist in machines' settings, both in assembly and testing.

5.8.4 RQ4 - Parameters to Predict Production Outcome

Data can be used to increase quality through monitoring the processes but in order to be able to increase the quality, parameters must be identified and controlled. The results indicates that multiple useful parameters can be found both in data gathered from the machines as well as data gathered from the production tests when investigating the gap of data. Parameters have been found in multiple areas of the production, from material variance to differences between the test stations. There lies a lot of potential in increasing the quality through monitoring and controlling the process with big data [24, 25]. Some parameters have been shown to not affect the outcome greatly and can be excluded as having any substantial effect but when combining the effect of multiple parameters they become a potential problem. This, however, has not yet been investigated. All production data is not analyzed and more parameters can be found but the thesis has shown that there are parameters that affects the production outcome negatively. The parameters must be further analyzed but the production can be partly predicted with the amount of data analyzed at this moment. To be able to predict, prevent and decrease the scrap rate, implementation also has to be made in the production. The results and the theory both showed that through big data and analysis production outcome can be predicted and PDs can be decreased. Parameters that indicates product errors have been found. With implementation and further evaluation on how to improve the production a higher quality can be reached. This is in line with what Kuo and Kusiak [3] also has presented.

6 Suggestions for Future Work

The study is not complete and more actions need to be taken. The database contains more parameters to be analyzed as well as parameters that might not have any analyzing purposes. There must also be further analysis on some parameters and a closed loop of data would be preferable since the tests indicates changes of the FT value. The recommendations on future work or future implementations are:

- Close the loop of data
- Analyze more parameters, all analyzed parameters in the study are seen in Appendix B
- Further path analysis for combinations of resources
- Compare FT value in AA to FT value in FT resource to understand if any resource needs to be re-calibrated and to investigate how the FT value changes from Assembly to Hardening
- Re-structuring of the database is needed and renaming of parameters to follow a naming convention will facilitate further analysis.
- The production tests were performed in a small scale and gave some indications of the effect on the production but needs further testing in a bigger scale and with more controlled parameters to confirm or deny the findings
- An AI to control time in oven depending on FT-value measured in AA since production tests show that the longer units are in the oven the higher value they get. Also, FT-value shrinks slightly during the first week after hardening

7 Conclusion

The aim of this thesis was to investigate whether big data could be used for analysis to predict and prevent production disturbances. This was investigated through a case study consisting of data collection, data analysis and tests in production in order to find answers to four research questions. The work was divided into 4 phases namely current state analysis, data collection, data analysis and data evaluation. In the current state analysis, triangulation was made through interviews, go-and-see tour and a survey that resulted in a number of parameters interesting to the issue at hand. A gap of data was found in the production that disabled the ability to perform a full path analysis. Therefore, two tests were performed in the production to investigate whether production disturbances could be traced to the processes where there currently was no data stored. Due to the social circumstances during the spring of 2020 the tests were not performed perfectly as planned. Nevertheless, the results showed that there is a need of closing the gap of data. The results also showed that for prevention of production disturbances and improvements in quality, data analysis has a significant importance.

A second aim was to find and construct a method or procedure for further investigation of data analysis in production prediction. The guidelines that were constructed were built on the methods and the conductance of this thesis' methods and results. From the findings and the issues that the authors faced along the way, knowledge was gained and then summarized into guidelines in how to analyze a production system using big data. The idea was to have guidelines adaptable to other productions and were therefore constructed using questions to enable an open mindset.

In the below list, extracts from the procedure are presented. These are especially useful when analyzing production disturbances in most manufacturing companies.

- A closed loop of data is required to enable analysis of a whole production system and to predict the outcome in a reliable way.
- Consider different conflicts of interest from stakeholders when investigating what production disturbance that should be the focus of the analysis.
- Analysis of all possible production paths can show differences between parallel resources and identify issues in the production system.
- Parameters must not always relate directly from to production resources, environmental factors might also affect the production performance.
- A DMAIC process for continuous improvements and data analysis is an appropriate procedure to keep searching for solutions to prevent PDs.
- A lot of data is gathered within Industry 4.0, make sure to keep it sorted and well defined.
  - Investigate if all stored parameters are actually useful, a lot of unnecessary data takes a lot of storage and may contribute to a messy database the is hard to navigate.
  - Not all useful parameters are collected and stored, this also has to be investigated in order to fully enable a thorough analysis of production data.

References

[1] J. Bokrantz, A. Skoogh, T. Ylipää, and J. Stahre, "Handling of production disturbances in the manufacturing industry," *Journal of Manufacturing Technology Management*, vol. 27, no. 8, pp. 1054–1057, 2016.

[2] A. Ingemansson and G. Bolmsjö, "Improved efficiency with production disturbance reduction in manufacturing systems based on discrete-event simulation." *Journal of Manufacturing Technology Management*, vol. 15, no. 3, p. 267, 2004. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=edb&AN=20177376&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[3] Y.-H. Kuo and A. Kusiak, "From data to big data in production research: the past and future trends," *International Journal of Production Research*, vol. 57, no. 15-16, pp. 4828–4853, 2019.

[4] C. Noshi, M. Eissa, R. Abdalla, and J. Schubert, "An intelligent data driven approach for production prediction," *Offshore Technology Conference*, 01 2019.

[5] M. Abuali, "Managing big data improves factory processes with oee, other key metrics." *MANUFACTURING ENGINEERING*, vol. 153, no. 1, pp. 32 – 38, 2014. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=edswsc&AN=000338901200009&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[6] K. Kumar, D. Zindani, and J. P. Davim, *Industry 4.0. [electronic resource] : Developments towards the Fourth Industrial Revolution.*, ser. Manufacturing and Surface Engineering. Springer Singapore, 2019. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=cat07472a&AN=clec.SPRINGERLINK9789811381652&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[7] A. Gilchrist, *Introducing Industry 4.0.* Berkeley, CA: Apress, 2016, pp. 195–215. [Online]. Available: https://doi.org/10.1007/978-1-4842-2047-4_13

[8] A. Lindbäck and P. Johansson, "Improve production flow and productivity by increasing predictability and control at an electronics manufacturer," MSc Thesis, Department of Industrial & Material Sciences, Chalmers University of Technology, Gothenburg, Sweden, 2019. [Online]. Available: https://hdl.handle.net/20.500.12380/256705

[9] U. Singh and M. Dwivedi, *Manufacturing Process.* New Delhi: New Age International Ltd, 2009. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=cat07472a&AN=clec.EBC442128&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[10] T. Ylipää, *High-reliability manufacturing systems.*, ser. Report / Department of Human Factors Engineering, Chalmers University of Technology, Göteborg,

Sweden: 9. Chalmers Library Print Collection, 2000. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=cat07470a&AN=clc.345e253046a44eaea5e24ad83020dc88&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[11] M. Bellgran and K. Säfsten, *Production development : design and operation of production systems.* Springer, 2010. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=cat07470a&AN=clc.b877ebe7944441b093dd65220aeb7c05&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[12] J. B. Matson and D. C. McFarlane, "Assessing the responsiveness of existing production operations." *International Journal of Operations & Production Management*, vol. 19, no. 8, p. 765, 1999. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=edb&AN=2334179&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[13] J. Ericsson, *Störningsanalys av tillverkningssystem: ett viktigt verktyg inom lean production.* Dept. of Production and Materials Engineering., Lund University, 1997. [Online]. Available: https://books.google.se/books?id=IrQJNAAACAAJ

[14] F. E. Sahin, "Lens design for active alignment of mobile phone cameras." *Optical Engineering*, vol. 56, no. 6, p. 1, 2017. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=edb&AN=124470512&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[15] E. Frid, "Optical assembly – active vs passive," https://www.beckermus.com/single-post/2018/04/04/Optical-Assembly-%E2%80%93-Active-VS-Passive, 2018, recieved: 2020-04-07.

[16] A. Sag, "Active alignment will be critical for future optical systems manufacturing," 2017.

[17] J. Gould, "Adhesives in electronics," *Technologywatch*, 2004. [Online]. Available: https://www.lboro.ac.uk/microsites/mechman/research/ipm-ktn/pdf/Technology_review/adhesives-in-electronics.pdf

[18] T. Turner, M. Lachapelle, and R. Kirschner, "Active alignment techniques improve lens centering." *PHOTONICS SPECTRA*, vol. 48, no. 12, pp. 42 – 44, n.d. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=edswsc&AN=000346625400027&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[19] H. Mohanty, P. Bhuyan, and D. Chenthati, *Big Data. [electronic resource] : A Primer.*, ser. Studies in Big Data: 11. Springer India, 2015. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=cat07472a&AN=clec.SPRINGERLINK9788132224945&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[20] BCS and Associates and Staff and K. Gordon, *Big Data : Opportunities and Challenges.* BCS Learning & Development Limited, 2014. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=cat07472a&AN=clec.EBC1650370&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[21] E. Van Dijk, “Smart data collection is required for continuously improving your production process,” *Quality Magazine*, 2018.

[22] B. Marr, *Big Data : Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance.* John Wiley & Sons, Incorporated, 2015. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=cat07472a&AN=clec.EBC1895824&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[23] H. S. Sim, “Big data analysis methodology for smart manufacturing systems.” *INTERNATIONAL JOURNAL OF PRECISION ENGINEERING AND MANUFACTURING*, vol. 20, no. 6, pp. 973 – 982, 2019. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=edswsc&AN=000469387800009&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[24] C. N. Madu, *Handbook of Total Quality Management. [electronic resource].* Springer US, 1998. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=cat07472a&AN=clec.SPRINGERLINK9781461552819&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[25] J. S. Lim, *Quality management in engineering : a scientific and systematic approach.* Taylor & Francis, a CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa, plc, 2020. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=cat07472a&AN=clec.ybp16165498&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[26] H. Ahuett-Garza and T. Kurfess, “A brief discussion on the trends of habilitating technologies for industry 4.0 and smart manufacturing.” *Manufacturing Letters*, vol. 15, no. Industry 4.0 and Smart Manufacturing, pp. 60–63, 2018. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=edselc&AN=edselc.2-52.0-85042167777&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[27] S. Sader, I. Husti, and M. Daróczi, “Industry 4.0 as a key enabler toward successful implementation of total quality management practices,” *Periodica Polytechnica Social and Management Sciences*, vol. 27, no. 2, pp. 131–140, (2019). [Online]. Available: https://pp.bme.hu/so/article/view/12675

[28] D. Byrne, “How can I use descriptive statistics?” *Project Planner*, 2017, Accessed: 29 April 2020.

[29] R. R. Wilcox, *Basic Statistics : Understanding Conventional Methods*

and Modern Insights. Oxford University Press USA - OSO, 2009. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=cat07472a&AN=clec.EBC453655&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[30] S. Dowdy, S. Wearden, and D. Chilko, *Statistics for research*, 3rd ed. Hoboken, New Jersey: John Wiley & Sons, 2004.

[31] H. C. Thode, *Normality Tests.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 999–1000. [Online]. Available: https://doi.org/10.1007/978-3-642-04898-2_423

[32] T. W. Anderson, *Anderson–Darling Tests of Goodness-of-Fit.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 52–54. [Online]. Available: https://doi.org/10.1007/978-3-642-04898-2_118

[33] E. Seier, *Normality Tests: Power Comparison.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1000–1003. [Online]. Available: https://doi.org/10.1007/978-3-642-04898-2_421

[34] T. Rahlf, *Scatter Plots.* Cham: Springer International Publishing, 2019, pp. 303–326. [Online]. Available: https://doi.org/10.1007/978-3-030-28444-2_9

[35] W. Vogt, "Dictionary of statistics & methodology," Thousand Oaks, California, May 2005. [Online]. Available: https://methods.sagepub.com/reference/dictionary-of-statistics-methodology

[36] Y. Dodge, *Line Chart.* New York, NY: Springer New York, 2008, pp. 316–316. [Online]. Available: https://doi.org/10.1007/978-0-387-32833-1_234

[37] L. Rade and B. Westergren, *Mathematics Handbook. [electronic resource] : for Science and Engineering.* Springer Berlin Heidelberg, 1999. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=cat07472a&AN=clec.SPRINGERLINK9783662035566&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

[38] R. Yin, *Case study research: design and methods*, 2nd ed., ser. Applied social research methods series. Sage Publications, 1994.

[39] V. Chris, T. Nikos, and F. Mark, "Case research in operations management," *International Journal of Operations & Production Management*, vol. 22, no. 2, pp. 195–219, Jan 2002. [Online]. Available: https://doi.org/10.1108/01443570210414329

[40] J. Meredith, "Building operations management theory through case and field research." *Journal of Operations Management*, vol. 16, no. 4, pp. 441 – 454, 1998. [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=buh&AN=11480541&site=eds-live&scope=site&custid=s3911979&authtype=sso&group=main&profile=eds

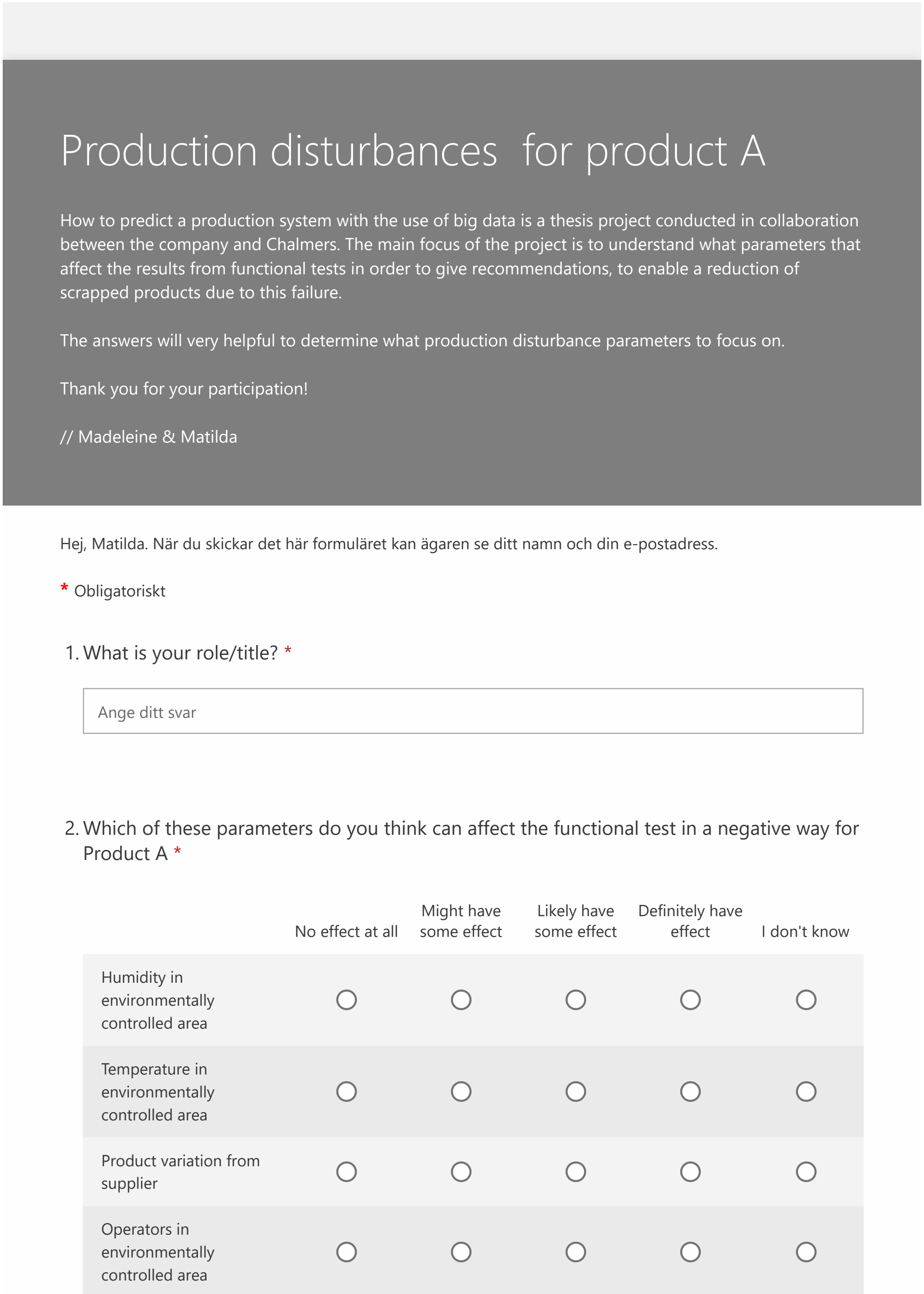
[41] Microsoft. Excel specifications and limits. Accessed on: 2020-

05-05. [Online]. Available: https://support.office.com/en-us/article/excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3

[42] Minitab 18 Support. Worksheets. Accessed on: 2020-05-05. [Online]. Available: https://support.minitab.com/en-us/minitab/18/help-and-how-to/minitab-environment/interface/interface-components/worksheets/

A Survey Questionnaire

The survey was sent out to various people, asking them to answer what parameters that might affect the production outcome, the questionnaire can be seen in figure A.1 and A.2



Production disturbances for product A

How to predict a production system with the use of big data is a thesis project conducted in collaboration between the company and Chalmers. The main focus of the project is to understand what parameters that affect the results from functional tests in order to give recommendations, to enable a reduction of scrapped products due to this failure.

The answers will very helpful to determine what production disturbance parameters to focus on.

Thank you for your participation!

// Madeleine & Matilda

Hej, Matilda. När du skickar det här formuläret kan ägaren se ditt namn och din e-postadress.

* Obligatoriskt

1. What is your role/title? *

Ange ditt svar

2. Which of these parameters do you think can affect the functional test in a negative way for Product A *

	No effect at all	Might have some effect	Likely have some effect	Definitely have effect	I don't know
Humidity in environmentally controlled area	○	○	○	○	○
Temperature in environmentally controlled area	○	○	○	○	○
Product variation from supplier	○	○	○	○	○
Operators in environmentally controlled area	○	○	○	○	○

Figure A.1: Survey sent out about what parameters that might affect the production outcome, page 1

	No effect at all	Might have some effect	Likely have some effect	Definitely have effect	I don't know
Component 2 placement on component 1	○	○	○	○	○
Effect on component 3 from pre-treatment in assembly	○	○	○	○	○
Fixture in assembly	○	○	○	○	○
Adhesive placement	○	○	○	○	○
Variation of movement in assembly (adhesive spread)	○	○	○	○	○
Time from environmentally controlled area to hardening 2	○	○	○	○	○
Placement in hardening 2	○	○	○	○	○
Time in hardening	○	○	○	○	○
Temperature drop when hardening 2 is open	○	○	○	○	○
Time from hardening 2 to functional test	○	○	○	○	○
Total production time	○	○	○	○	○

3. Are there any other parameters you think can affect the results from the functional test? *

Ange ditt svar

Skicka

Figure A.2: Survey sent out about what parameters that might affect the production outcome, page 2

B Parameters Plotted and Grouping Used

In table B.1, all parameters that were plotted against each other in 2D scatterplots are presented. The parameters plotted in 3D scatterplots are presented in table B.2.

Table B.1: Parameters plotted against each other and their grouping

Parameter 1	**Parameter 2**	**Grouping**
Timestamp	Humidity in ECA	AA-status
Timestamp	Temperature in ECA	AA-status
FT-value	Humidity in ECA	FT-status
FT-value	Temperature in ECA	FT-status
Batch	Movement X PL3	AA-status
Batch	Movement X PL4	AA-status
Batch	Movement Y PL3	AA-status
Batch	Movement Y PL4	AA-status
Movement X in PL3	Movement Y in PL3	AA-status
Movement X in PL4	Movement Y in PL4	AA-status
PL3 Fixture code	Movement X in PL3	AA-status
PL4 Fixture code	Movement X in PL4	AA-status
PL3 Fixture code	Movement Y in PL3	AA-status
PL4 Fixture code	Movement Y in PL4	AA-status
Batch	FT-value	FT-status & Assembly line
PL3 Fixture code	FT-value	FT-status
PL4 Fixture code	FT-value	FT-status
FT resource	FT-value	FT-status
Days from AA to FT	FT-value	FT-status
Fixture code	% Failed units	Year of production

Table B.2: Parameters plotted against each other in 3D-plots and their grouping

Parameter 1	**Parameter 2**	**Parameter 3**	**Grouping**
Movement X in PL3	Movement Y in PL3	FT-value	FT-status
Movement X in PL4	Movement Y in PL4	FT-value	FT-status

C Manufacturing Process in Detail at the Company

The production processes for the product will be further described in this section. Each process will be briefly described in the corresponding order of the production flow.

ECA

In the ECA the temperature and the humidity is continuously measured and controlled to ensure a stable environment. This, because of the high sensitivity of the components before being assembled.

Pre-Assembly

In this process the first components are assembled through attaching a component with an adhesive to the board. The process is performed automatically in a machine without operator involvement. An automatic quality check is performed and in case of any errors an operator will be notified and make a manual quality check of the product.

Welding

The second process in the production line is welding. Here, the two components are automatically welded together and wires are attached to enable electrical communication between the components and to the computer that will communicate with the finished product. Even after the welding process a quality check is performed to ensure the product quality and if an error occurs an operator is called on to manually inspect the quality.

Assembly

In the assembly process one more component is attached through an AA, see more in section 2.3.2 about the assembly method. The AA is an automatic process that applies adhesive and places the components in the correct position.

Hardening 1

The first hardening process is a quick hardening to lock the components in position to make sure the components stay in place and are safe to move without affecting the result of the assembly process. Hardening 1 is the last process in the ECA. After this process the product leaves the area and goes into to the ordinary production area.

Hardening 2

Hardening 2 is performed at a different location at the production site and the products are moved from the ECA to the hardening process with the help of a Kanban train operated by a person. The product is placed on a tray in the hardening resource and the process time is 6 ± 1 hours.

Buffer

Products are placed in a FIFO buffer and remain there until called from the upcoming functional test. The time the product stays in the buffer is not controlled or measured.

Functional test

When the product has finished all the steps in the first half of the production it goes to FT. Here the products are tested to ensure the product quality in multiple parameters. The test is performed in four different parts to ensure the quality and the performance. Only one of the four tests is applicable to the PD considered in this thesis, so it is only values from that test that has been analyzed. The different FT resources can have different LCL and UCL over time so when plotting the limits, the most used limits will be seen. If a product has failed in FT the product can be re-tested during up to 18 days.

D Survey Answers

All answers collected from the survey are presented in figure D.1.

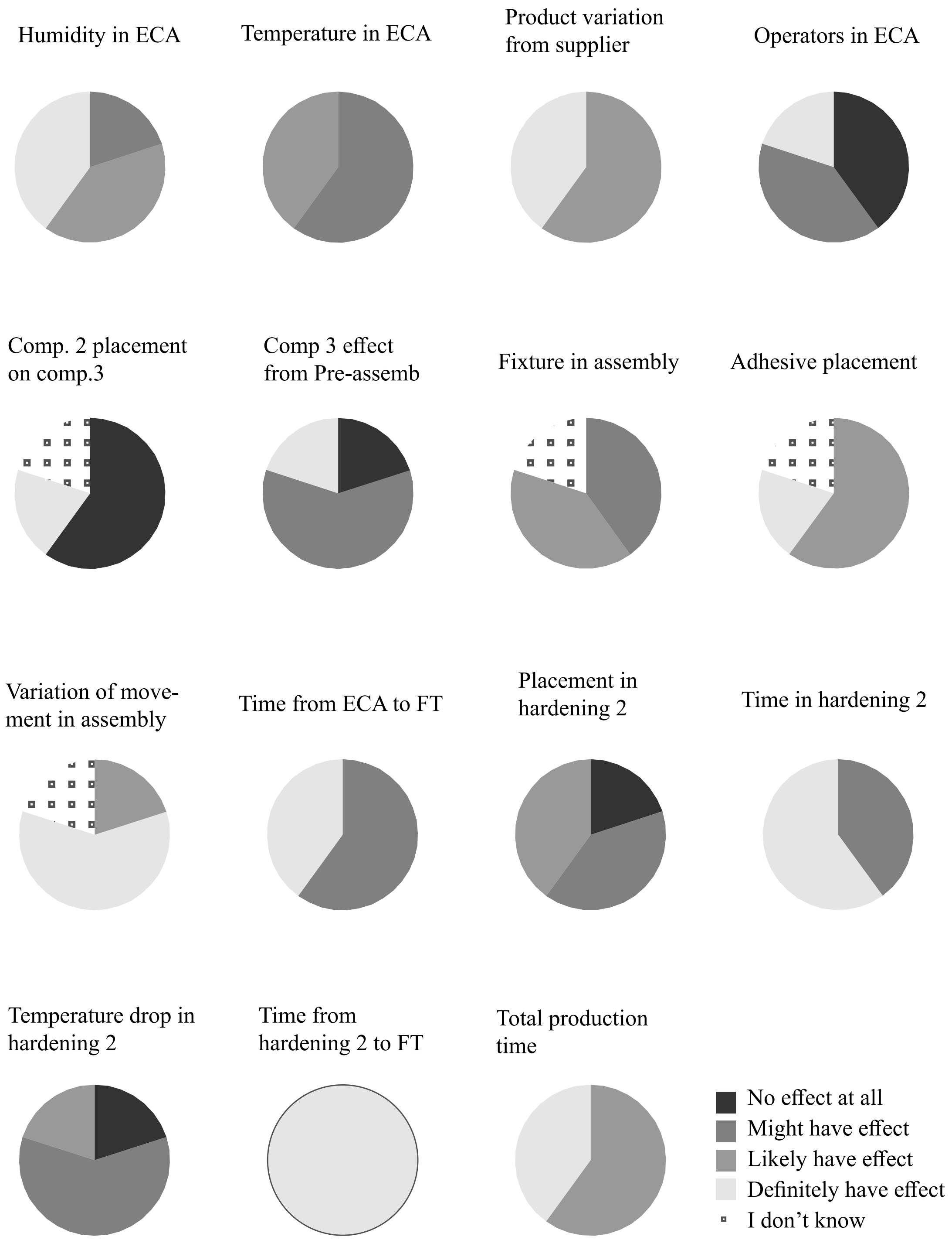


Figure D.1: Pie chart summarizing answers from the survey performed

E All Parameters Extracted from the Database

All parameters extracted from the database can be seen in table E.1.

Table E.1: All parameters that were extracted from the database, their format and if they were used in analysis

Parameter	**Format**	**Used**
Product ID	Text	Yes
AA-TimeStamp	Date	Yes
Production line	Text	Yes
AA-P/F	Text	Yes
Align_X	Number	Yes
Align_Y	Number	Yes
Align_Z	Number	Yes
Roll	Number	No
Pitch	Number	No
Yaw	Number	No
Fixture code	Number	Yes
Material Batch	Text	Yes
FT-Timestamp	Date	Yes
FT resource	Number	Yes
FT-Value	Number	Yes
FT-P/F	Text	Yes
Timestamp Temp/Humidity	Date	Yes
Humidity	Number	Yes
Temperature	Number	Yes

F Correlation Matrices for Buffer Test

The correlation matrix from the production test of the different buffers can be seen in figure F.1 for the DR and figure F.2 for the FIFO buffer.

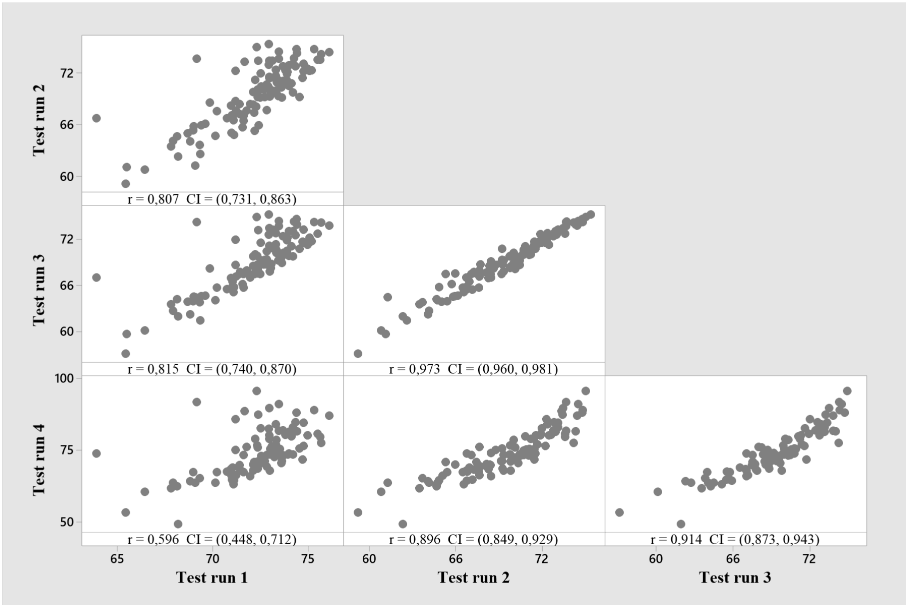


Figure F.1: Correlation matrix between test runs of units in the DR buffer

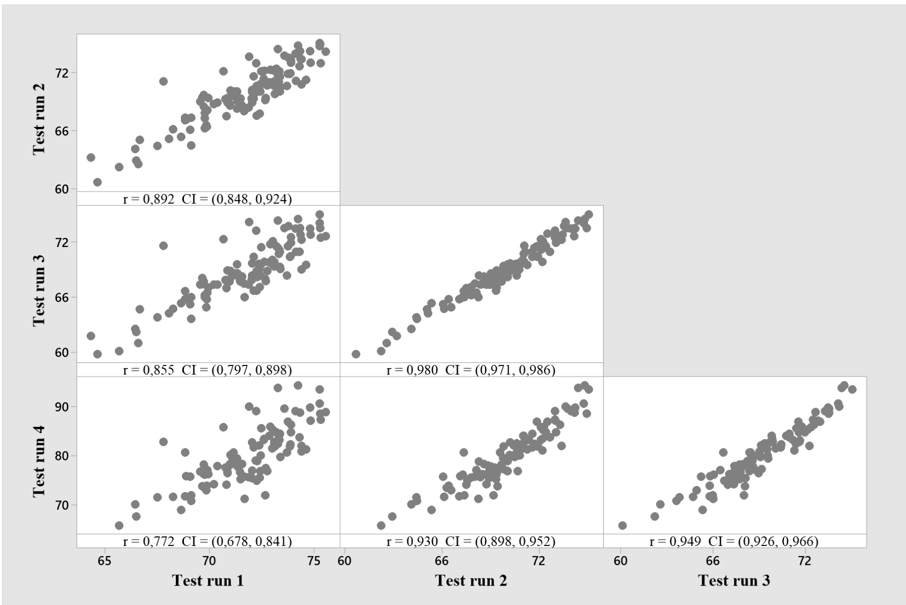


Figure F.2: Correlation matrix between test runs of units stored in FIFO buffer

G Humidity and Temperature Plots for Buffer Test

Humidity and temperature in the different buffer areas are presented in this Appendix, humidity data is seen in figure G.1 and temperature in figure G.2.

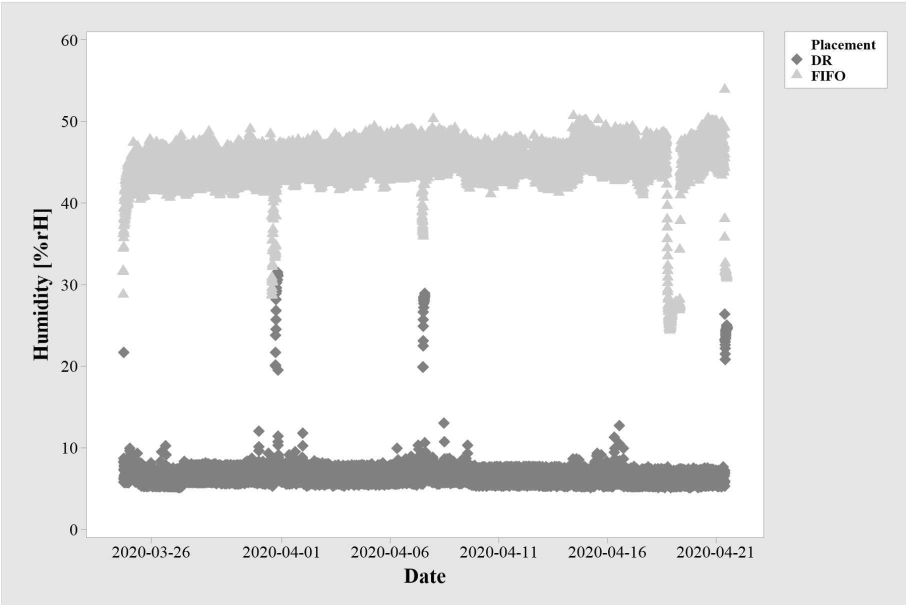


Figure G.1: Humidity in DR and FIFO

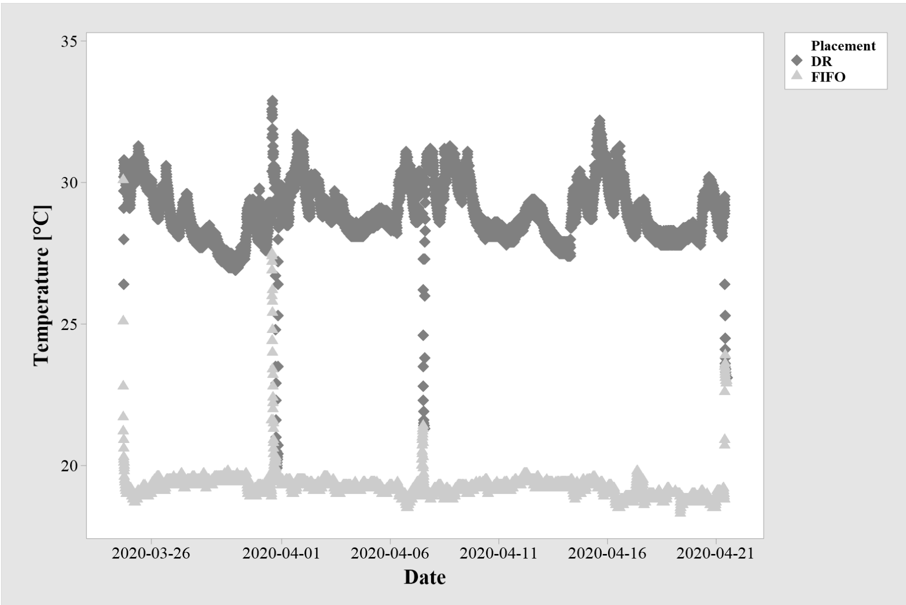


Figure G.2: Temperature in DR and FIFO

H Normality test for Align X and Align Y

Normality tests and their plots are presented in figure H.1 for Align X and figure H.2 for Align Y.

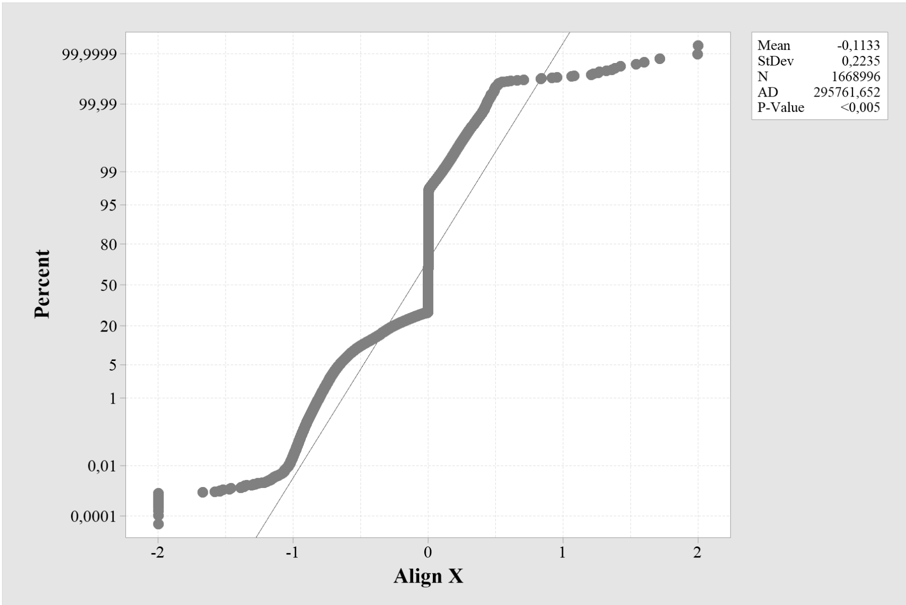


Figure H.1: Normality test for Align X

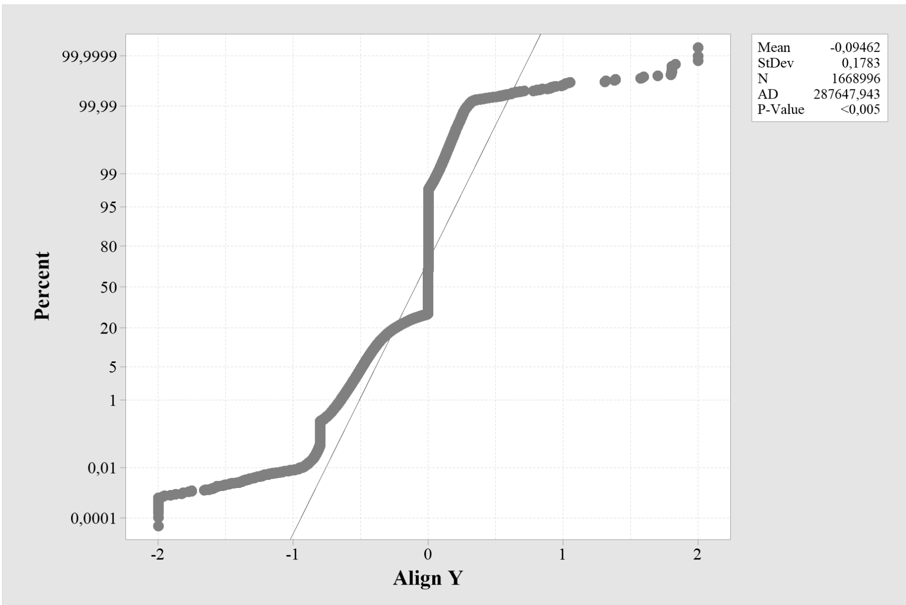


Figure H.2: Normality test for Align Y

I Material Variation in Assembly Level

In this appendix the material variation can be seen plotted against movements in X and Y directions. In figure I.1 movement in X in PL3 can be seen. For PL4 movement in X can be seen in figure I.2 and Y movement in figure I.3.

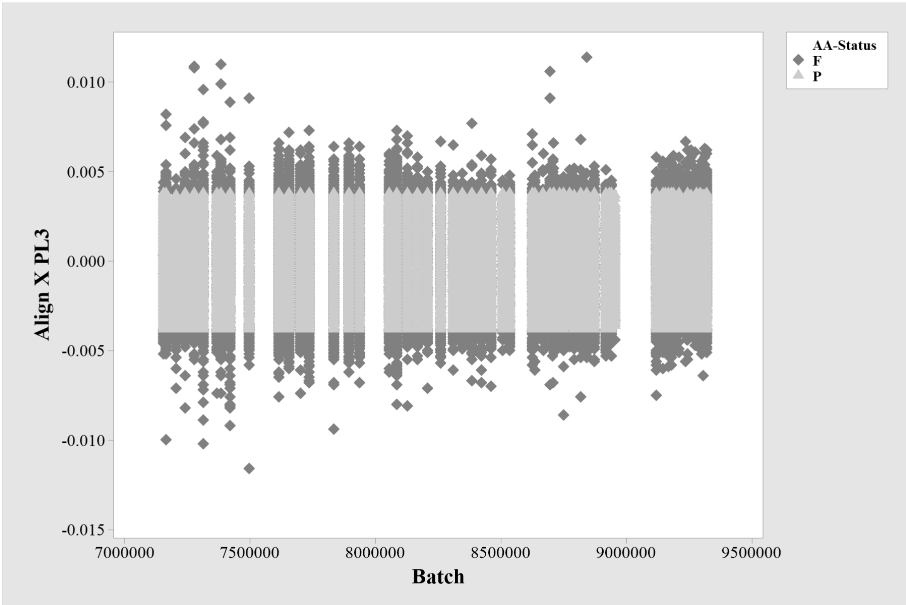


Figure I.1: Batch with align value in X-direction and grouped by P/F in PL3 assembly

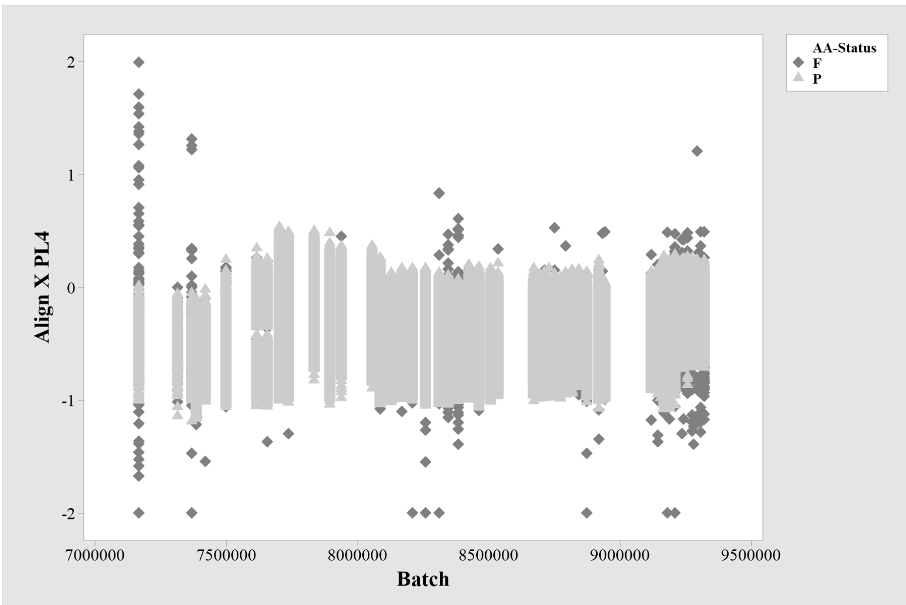


Figure I.2: Batch with align value in X-direction and grouped by P/F in PL4 assembly

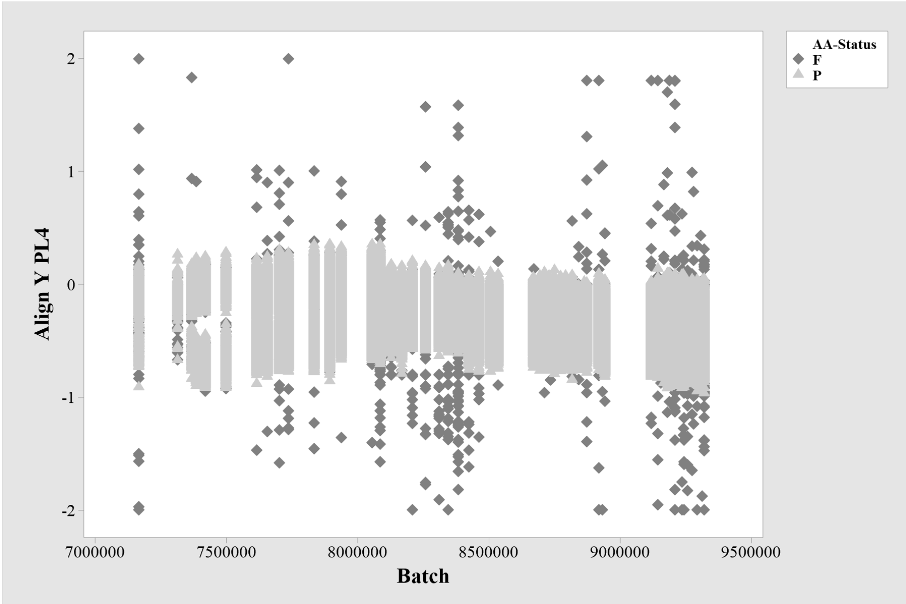


Figure I.3: Batch with align value in Y-direction and grouped by P/F in PL4 assembly

J Production Line Fixtures

The production line 3 and 4 plotted against the movement in X respectively Y direction. PL3 is seen in figure J.1 and PL4 in figure J.2.

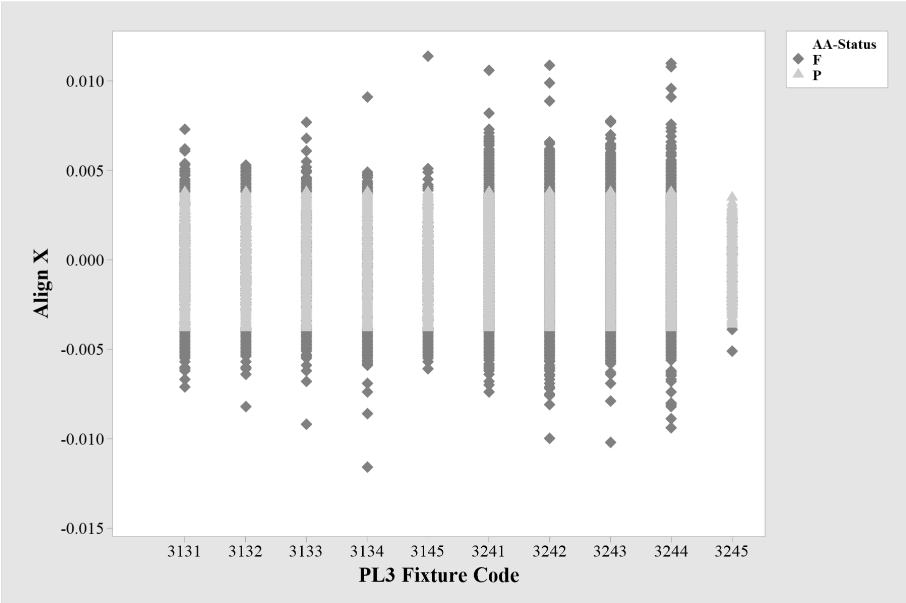


Figure J.1: Fixture with align value in X-direction and grouped by P/F in PL3 assembly

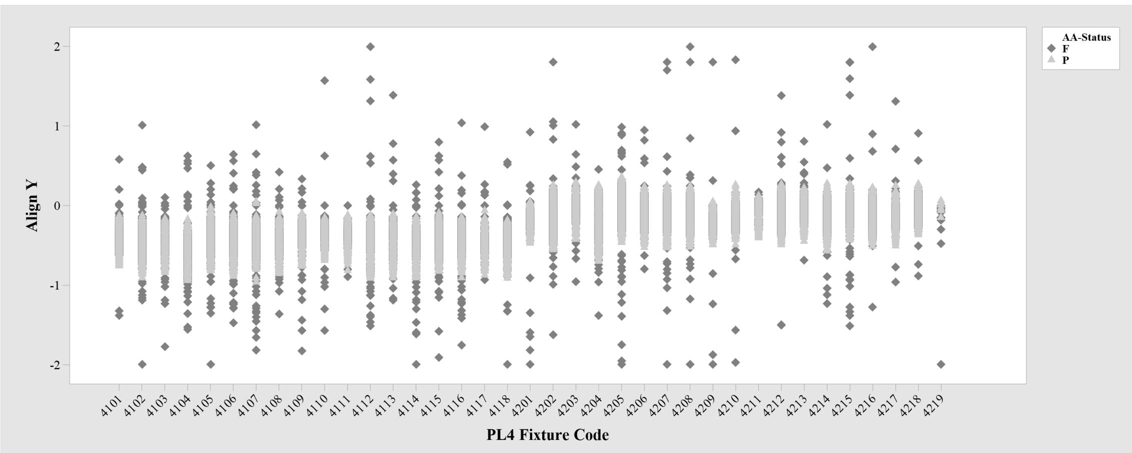


Figure J.2: Fixture with align value in Y-direction and grouped by P/F in PL4 assembly

K Guidelines

Flowchart of the general guidelines to help the case company and other manufacturing companies deal with PDs using big data.

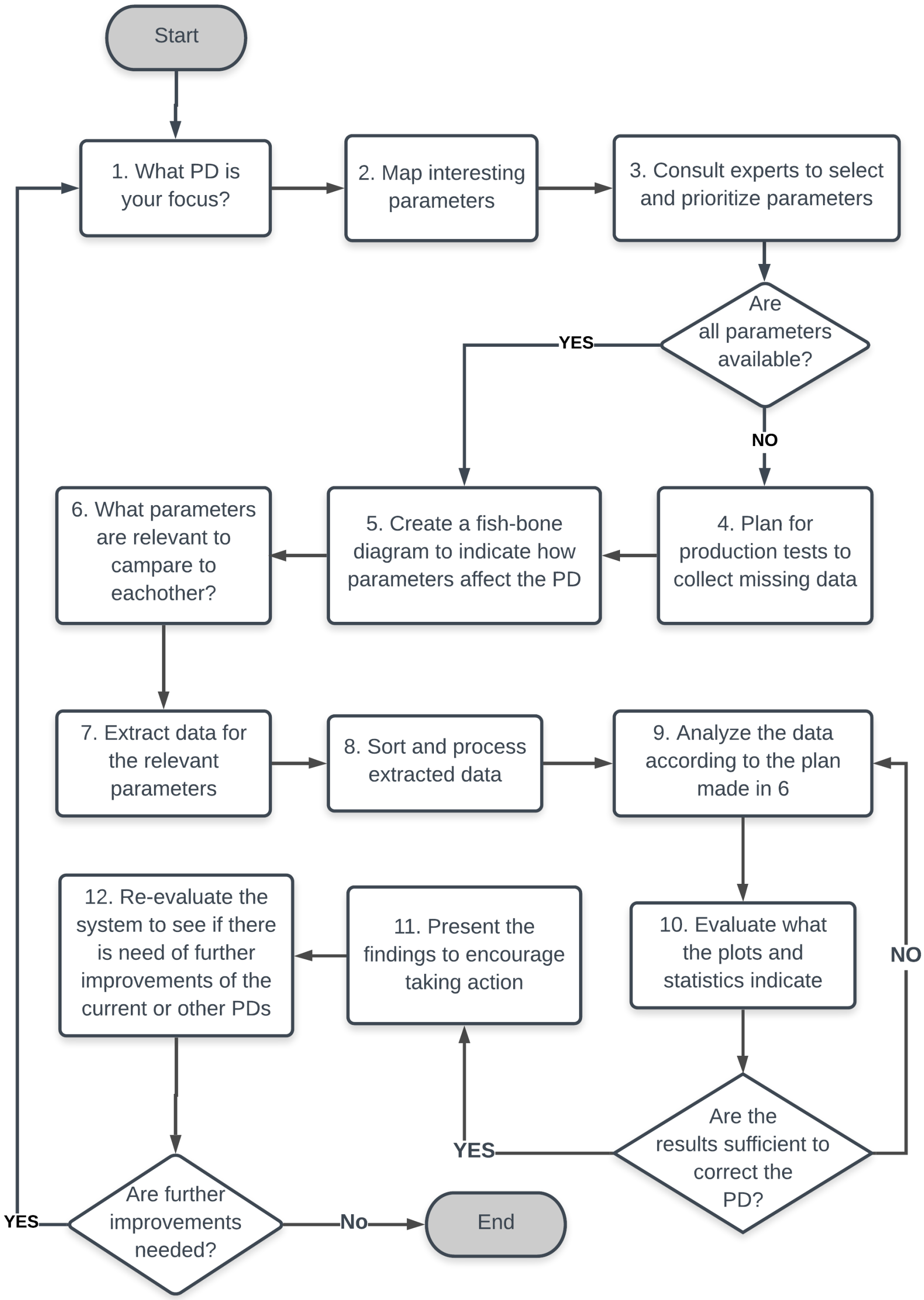


Figure K.1: Flowchart for general guidelines when analyzing PDs in manufacturing processes