University of
South-Eastern Norway

www.usn.no

FMH606 Master's Thesis 2021
Industrial IT and Automation

Manufacturing analysis and data acquisition in advanced machining

Kjell Arne S. Guldbjørnsen

Faculty of Technology, Natural Sciences and Maritime Sciences
Campus Porsgrunn

University of South-Eastern Norway www.usn.no

Course: FMH606 Master's Thesis 2021
Title: *Manufacturing analysis and data acquisition in advanced machining*
Pages: *67*
Keywords: *Condition-based maintenance, Principal component analysis, Mahalanobis distance, Probe calibration, Temperature, Vibration, Runout*

Student: *Kjell Arne S. Guldbjørnsen*
Supervisor: *Håkon Viumdal, Saba Mylvaganam*
External partner: *Håvard Norum, Stefan Köwerich*

Summary:
To cut cost of maintenance, being able to stop the machines at the right time before fault and the possibility to implement zero defect manufacturing is an important part of manufacturing aeroplane engine parts. The purpose of this report is to test a statistical method on different data sources of GKN Aerospace Norway, and Sweden to see what kind of available dataset is best suited to prevent high maintenance cost and identify faults in the machine.

The data sources are based on calibration data of a probe and temperature from a Carnaghi vertical turning lathe machine at GKN Aerospace Norway, and vibration and runout data from the spindle on a GROB milling machine at GKN Aerospace Sweden. Using Principal component analysis and Mahalanobis distance to analyse these datasets will give a better picture of what kind of data to use when implementing condition-based maintenance or optimising fault recognition.

After testing both datasets, the dataset available from the GROB machine made it possible to see when the spindle of the machine slowly started to deteriorate before a fault happened. This was possible since the dataset had a known fault. The calibration dataset shows it is possible to identify deviations from normal calibration, and may make it easier to analyse deviation on later calibration. The method is not implemented into either of the sites, but this report may give more background for further work.

Preface

This report is based on an idea from engineers at GKN Aerospace Norway (GAN). It concerns the use of data collected from the calibration of a cube probe on a Carnaghi machine and vibration data from bearings on a GROB machine at GKN Aerospace Sweden (GAS). GROB is a machine manufacturer from Germany. The main aim is to analyse the data acquired from the machines at GAN and GAS today and use it to make better decisions regarding preventive actions or maintenance of the machines in the future. Today's maintenance is performed with limited analytical background. There are mostly trial and error, tolerances are given by the manufacturer of the probes and spindles, and opinions from experienced engineers when setting the tolerances and rules on changing the probes or spindles. The maintenance on the machines is set to a given interval, also called preventive maintenance. This report utilizes the historical data of the machines to analyse through Principal component analysis and Mahalanobis distance to show the state of the measured equipment. The report is meant for the industry and those who want to improve the maintenance routine to cut the cost of unnecessary expenditure for equipment and machines. It provides an introduction to the methods of analysing data on older systems in the industry. A secondary aim of the project is to see what the bare minimum of dataset or what sensors to analyse to give results for a company, when looking at the health of the machine and its equipment. Due to the COVID-19 epidemic, internal resources in the project were strongly reduced. This affected the data acquisition and background information of the systems at GAN and GAS.

From GAN I would like to thank Håvard Norum as supervisor and Stefan Köwerich for input on the calibration process. From USN I would like to thank Håkon Viumdal as supervisor, Saba Mylvaganam as co-supervisor, and Ru Yan for input on ideas on the analytical part of the report. From GAS I would like to thank Mikael Alm and James Gladh for assistance retrieving data and some background on the vibration and runout data. From IFM Electronic Sweden I would like to thank Ingemar Sjöberg for assistance with more background information on the equipment on the GROB milling machine.

Porsgrunn, 2nd June 2021

Kjell Arne S. Guldbjørnsen

Contents

List of Figures

List of Tables

Nomenclature

Abbreviation	Explanation
1D CNN	1-dimensional convolutional neural network
CMM	Coordinate measuring machine
CNC	Computer numerical control
CBM	Condition-based maintenance
COTS	Commercial off-the-shelf
DL	Deep learning
FCNN	Fully connected neural network
FD	Fault detection
GAN	GKN Aerospace Norway
GAS	GKN Aerospace Sweden
GM	Green Monitor
IPG	In-process gauge
KTT	Kongsberg Terotech
MAAM	Multivariate analysis in advanced machining
MADAAM	Manufacturing analysis and data acquisition in advanced machining
MD	Mahalanobis distance
ML	Machine learning
NN	Neural network
PC	Principal component
PCA	Principal component analysis
PdM	Predictive maintenance
Q-DAS	Qualitative data analysis software
RMP	Radio modular probe
RPM	Revolutions per minute
SVD	Singular value decomposition
USN	University of South-Eastern Norway
VTL	Vertical turning lathe
ZDM	Zero defect manufacturing

1 Introduction

This chapter contains the background for the report, previous work done by GAN, GAS, literature research, and the structure of the report. GKN Aerospace as a global company produces complex parts for the worlds biggest aeroplane engine program, both to the commercial and military market [1]. GAN was an early adapter to tool probing and in-process gauging (IPG) tools in the '80s. Like many other manufacturing industries, most of GAN's machines are from different decades. New machines that are implemented today have hardware (sensors and controlling system) and software (up-to-date operating systems, more flexible data transfer methods) that uses most of the data collected from the probing and sensor data to give a better picture of the state of the machines. The report is based on a project set by GAN concerning the advanced use of probe technology in advanced machining and the wish for further analysis, locate inaccuracies and structuring of data measured in the machines at GAN. A probe is an elaborate switch designed to trigger when in contact with another surface [2]. Knowing the dimensions of the probe, makes it possible to collect the position of the surface, in reference to the zero point of the machine. During the manufacturing of aerospace engine parts, it is crucial that the machines have the same state to be able to achieve zero defect manufacturing (ZDM). Causes to prevent ZDM may be vibrations of the machine, misalignment of a spindle, poorly calibrated equipment used when machining, unwanted stops in the machine, etc. When producing engine parts with tolerances that are 1:1000 of a mm (μm) on the dimensions, many parameters are contributing to deviations when machining. The method used in this report will try to prevent some of these causes to affect the machining of the engine part. Due to the complexity of the machines, these data sources will not be able to allocate all faults or causes for all unwanted stops in the machine, but it will be able to indicate the most common causes, which are misalignment of the spindle and misalignment during machining. The data sources are based on the calibration of probes and temperature data from a Carnaghi vertical turning lathe (VTL) machine at GAN, and vibration and runout data from a GROB milling machine at GAS.

The aim of the report is to be able to identify faults from the data from different machines before irregular manufacturing causes faults in the machine. Another aim is to identify which data source from the machines gives the best result about the state of the equipment measured. This makes it easier when implementing the method on other machines with similar data sources available.

Comments on the project description

The project description is included in Appendix A. Due to some changes during the project, the method is applied to both calibrations of a cube probe and temperature data from a vertical turning lathes Carnaghi machine at GAN, and a dataset from vibration and runout from a GROB milling machine at GAS. The datasets were chosen based on availability, historical data, background information and recommendation from engineers at the sites.

Due to the lack of resources at GAN, the implementation of the method in the machine and automation of the data acquisition was not possible. The method chosen is also more based on condition-based maintenance (CBM) and fault recognition, rather than predictive maintenance (PdM), due to the lack of available data from each of the machines.

1.1 Previous work on data analysis and acquisition

From the 5$^{\text{th}}$ semester of the study program Industrial IT and Automation, testing of the method was used only with cube probe calibration data made available. This was tested with one year of data on the calibration of the cube probe, also with some analysis on temperature measures of the ambient, in the machine and in the cooling liquid tanks. Since the analysis of that report was based on one year of data that had no apparent fault in the machine during that time, no apparent result was accessible [3]. Since this report, Manufacturing analysis and data acquisition in advanced machining (MADAAM) is a continuation of the Multivariate analysis in advanced machining (MAAM) report some of the description are similar. The description of the prevailing system such as the parameters affecting the quality of the part, sensors and methods is still included in this report as there are some changes and more information around the subjects that surfaced during this project. More information on the machine at GAN and Green monitor (GM) is available in the MAAM report. The machine and GM information is not included in this report after an agreement with the main supervisor, since there are no considerable changes.

1.1.1 Data analysis and acquisition at GKN Aerospace Norway

At GAN there have been multiple projects and weekly routines around the collection and analysis of data. One of the outcomes of these projects has been GM. GM is a visualization tool for production and sensor data. Figure 1.1 shows an example of production data shown in GM. The figure shows the number of hours the machine has been in each of the states (active, in cut, maintenance, offline). It also shows the distribution of hours each day in a required duration. The operators are also using it for error registration and shop floor overview. This data is not important to the report itself, but GM as a tool is more discussed concerning implementing the method of the report as a tool for further use at GAN.

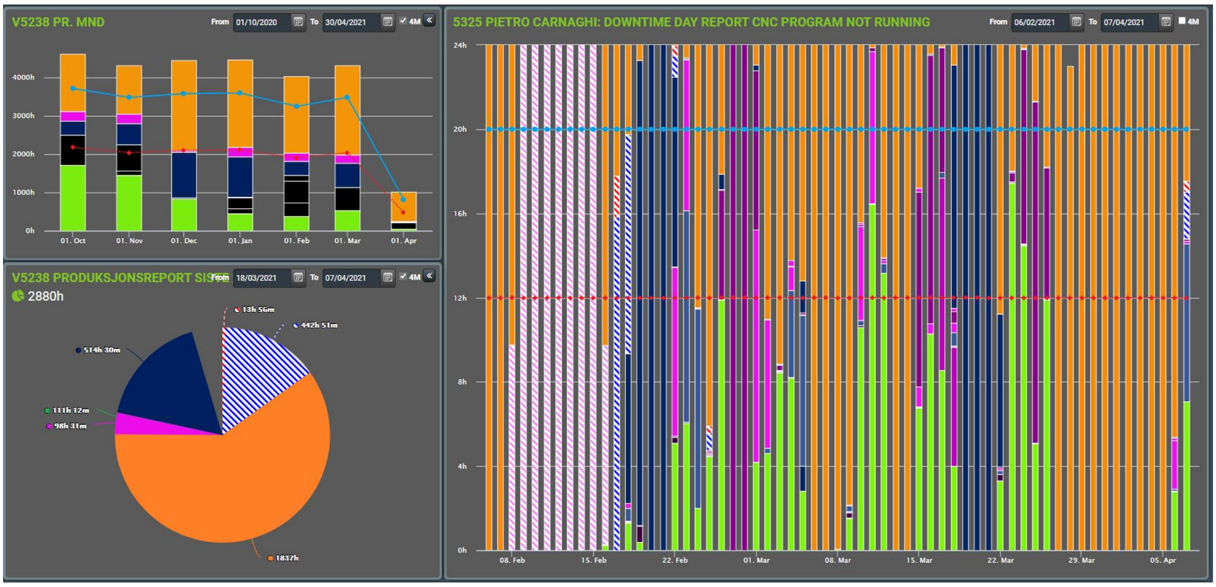

Figure 1.1: Screenshot of available production data in GM at GAN. This data is not important to the report itself.

Weekly analysis of calibration data

The engineer that follows the calibration is collecting the data manually from the machines regularly, usually once a week. A file is saved on the GAN-client (desktop computer) by the part program. All manufacturing machines at GAN have a GAN-client. The way the data is analysed is by looking at the source calibration data and looking at the length of the sides of the cube. After analysing the calibration data, some nominal min-max values are calculated and entered in the calibration process part program. When the sides are greater than the nominal value set during later calibration, the machine stops. The first step of the calibration process is for the operator to make sure that the cube or tool is clean of dirt or cooling liquid. Depending on the severity of the deviation, an engineer is contacted to find a solution to why the deviation happened. The outcome from an earlier project made it possible to change the format of the calibration data to be easier to analyse through Excel. The project also made some of the calibration and temperature data available through qualitative data analysis software (Q-DAS). Q-DAS is a statistical tool that gathers geometrical measurements and machine data. The tool is very restrictive to what kind of analysis to use on the data, the structure of the database is not easily compatible with other analysis software and no intuitive way of exporting the data.

Previous projects on data acquisition

Earlier, there have been projects around business intelligence and software to help visualize and show the data in manufacturing such as order data, manufacturing data and machine availability. These projects have mainly focused on acquiring a good visualization tool and not working on the data input level. This has caused many of the solutions to not being used after the projects are over, due to not getting the results initially required in the beginning. With more resources put in structuring the data at the bottom level, it is more likely that the following projects have a better chance of succeeding.

1.1.2 Data analysis and acquisition at GKN Aerospace Sweden

GAN is the sister company of GAS. GAS uses a similar software as GM called Copilot. Copilot is made in Grafana and is a visualization tool for production and sensor data. An example is shown in Figure 1.2. The figure shows the vibration and runout data from 12.12.19 to 01.05.21 of a GROB machine at GAS. The vibration is measured in acceleration and has the unit mm/s^2. The horizontal yellow and red line is defined tolerances, that stops the machine if exceeded or gives an alarm to the operator of the machine. The runout data is measured in μm. Due to more prioritization and resources on data acquisition at GAS, more data is available.

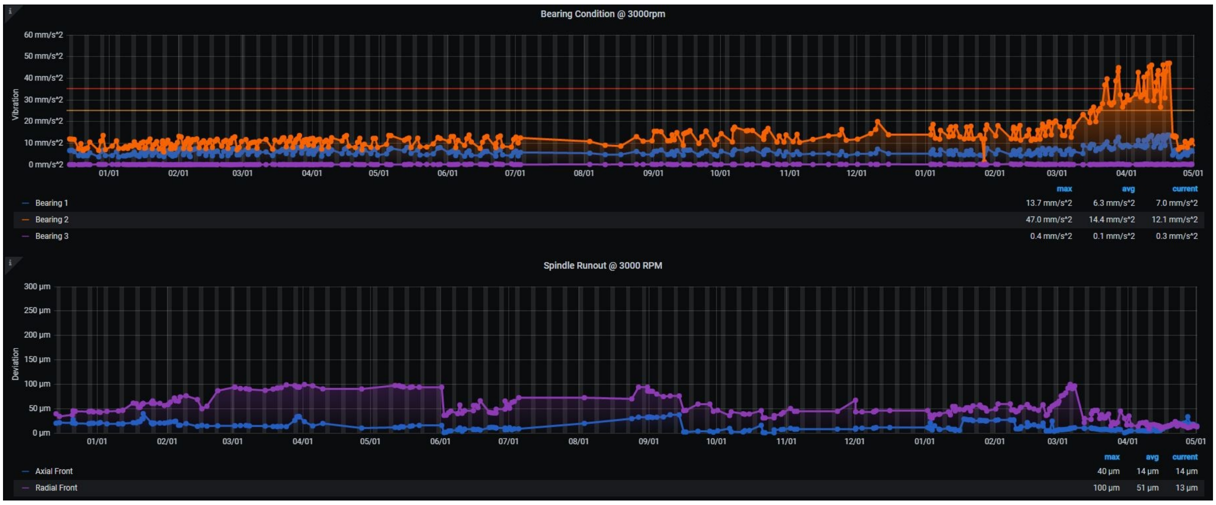

Figure 1.2: Screenshot of source data of vibration, measured in mm/s^2 (top graph), and runout (bottom graph) data, measured in μm from Copilot.

1.1.3 Literature research - Condition-based maintenance and fault detection

As a part of the project, a literature research has been conducted to see the use of the neural network (NN), machine learning (ML), Deep learning (DL), etc on conditional monitoring for the use of condition-based and PdM and fault detection (FD).

In sheet metalworking stamping is one of the most commonly used processes to make products. This study applies a deep 1-dimensional convolutional neural network (1D CNN) along with a fully connected neural network (FCNN) for quality extraction to the classification of wear conditions [4]. Through the study, the vibration measuring showed different trends after some processes. This made it possible to see the rate of deterioration of the stamping tools. These trends were then used to monitor the health of the stamping machine, through the use of 1D CNN. Due to getting 99.8% accuracy when running the 1D

CNN on the vibration data, the vibration data from GAS was implemented in this project.

Railway transportation uses real-time monitoring to be able to secure safety for passengers and cargo. Cloud computing is used on existing train fault diagnosis technology [5]. The fact that cloud computing demands high computing resources and long processing times, contradicts the importance of real-time monitoring on railway transportation. By using edge computing in collaboration with cloud computing real-time monitoring is possible. Edge Intelligence/computing is a distributed computing paradigm that sets the processing resources and datasets closer to the analyser, reducing the response time and network resources [6]. On the shop floor of GAN, multiple sensors on different locations on the machine have different protocols, due to the installation of other sensors years after the machine was implemented. Using a similar solution when gathering data from the shop floor makes it less time consuming when processing the data from all the data sources.

A paper on propulsion system uses data-driven models (DDM) to analyse a large number of historical datasets gathered by on-board systems, without requiring any prior knowledge of the system under analysing. Using DDM to attempt to implement CBM on the propulsion system [7]. If GAN structured the data more and labelled most of the sensor data, it would be possible to use supervised ML techniques to analyse the data. Several of the supervised ML techniques require that the data is labelled. The labelling of the data also gives a good overview of what is available in each of the machines. This makes maintenance, calibration and upgrading of the system around the machine easier.

1.2 Structure of the report

Chapter 2 of the report is the technical background of the prevailing system and the structure of the data of cube calibration and vibration. Chapter 3 is on the description of the data and the validation and groundwork done on the measured data and the background of the methods used. Chapter 4 is on the results of further data analyses. Chapter 5 is the challenges during the project, discussion, conclusion and future work.

2 The existing manufacturing system, processes and maintenance methods

This chapter describes the existing system (type of machine, sensors, probes, equipment on the machine), processes, and methods used today at GAN and GAS. As an example, one machine and its sensors and components are chosen to attempt to make a general solution for the rest of the shop floor at GAN and GAS.

2.1 Prevailing manufacturing system used at GKN Aerospace Norway

The following section describes information on the analysing method used today to be able to indicate faults or deviation from normal manufacturing for the machines, probes and the spindle. The method for calculating the tolerances of the calibration of the cube probe is also discussed. The principle of calibrating a cube probe is to use a master tool. The master tool can have different shapes, but it depends on the type of machine and type of probe. Not all machines have all sides of a cube accessible during calibration. Due to the high precision and accuracy needed when machining aeroplane engine parts, calibration is needed before every process. Figure 2.1 shows most of the parameters affecting the quality of the engine part in some way. In the end, that is the most significant result during manufacturing. All these things must have as little misalignment and deviation as possible to not have an impact on the quality of the engine part after multiple processes. This is important since most of the engine parts at GAN has 25+ operations. Some has as many as 60+ operations. In some of the operations, measurements from earlier machines are used as parameters to implement in the part program of the machine. As an example, this can be the measurement from the coordinate measuring machine (CMM) on how much to machine the part when turning or milling. There has not been done much analysis on these different parameters together but rather done one by one. This is due to the duration of the project, to get a good result from analysing multiples of these parameters, a lot of resources is needed. Table 2.1 shows a more detailed description of what each of the quality parameters is.

Figure 2.1: Parameters affecting the quality of the manufacturing of the different aeroplane engine part.

Table 2.1: Description of the different parameters concerning the quality of an engine part is shown in Figure 2.1.

Parameter	**Equipment type**	**Description**
Machine state	Sensor analysis	Measurements defining the state the machine is in i.e. running, processing, fault, etc
Probe	Touch probe	Measuring the position of the engine part in the machine room
Light probe	Laser probe	Verifying the location of the tool
Cube	Cube probe	Used for calibration of the tool
Operator data	Manual registration	Historical data regarding events on the machine. Operator maintenance, general changes to the machine and sudden stops in the machine
Kongsberg Terotech (KTT) laser	Sensor	Measurement in the machine
Manual measurement	Manual registration	Measurement executed on the engine part with manual measurement tools i.e. calipers, punch, dial gauge, etc
CMM	Measurement machine	Measurements of the engine part in the CMM
IPG	Sensor	Measurements of the engine part in the computer numerical control (CNC) machine
External data	Sensor	Measurements on or around the machine i.e. temperature, accelerometer, level sensor, etc
Artifact	Sensor	Used to evaluate processes in the machine

This report is based on historical data previous to the project start with the cube and external data due to the fact the duration of the project is 5,5 months. To have the background knowledge and resources to analyse all the different parameters more time is necessary.

2.1.1 Probing in the machine

Multiple different sensors are used on the many different engine parts machined at GAN. Some are in the machine itself and others are in a CMM. CMM is only measuring the geometrical shape of the engine parts, making sure the measurements are within the tolerances set by the customer and GAN's internal tolerances. In the different machines, there are used different probes to measure different positions on the engine part or the state of the tools. One of the most used probes in GAN is a Renishaw radio modular probe (RMP) 60 with a LP2 probe (LP is the series of probes, not an abbreviations). RMP60 is used for transmitting the measurements from the LP2 probe. The LP2 probe touches the engine part in specific places to measure the deviation from the specified values, to make sure the fixture and the engine part are in the correct placement inside the machining room. This and the calibration of the tools with the LP2H to make sure that the engine part is being machined as planned. Table 2.2 shows the list of the probes and sensors used in this project to make sure that the tools are calibrated in length- and radius compensation. Each probing is done multiple times during different processes of machining the part. It is part of the part program and will stop the part program if the values are out of the defined tolerances. The part program is a CNC program for one operation done on one specific engine part in that specific machine. Each operation has its own part program linked to it. The temperature is measured with a PT100 element during the calibration process. All temperature measured locations mentioned in Table 2.2. The principle of the PT100 element is to measure the resistance of the platinum element and depending on the type of PT, the resistance correlates to a certain degree.

Table 2.2: List of the chosen sensors and probes on the Carnaghi milling machine.

Name	**Type**	**Description**
Renishaw RMP60	Radio transmitter	Transmitting the measurement from LP2
Renishaw LP2	Measuring probe	Used for calibration of the placement of the engine part
Renishaw LP2H	Measuring probe	Used for calibration of the tools with a cube attachment in this report (cube probe).
Temperature ambient	PT100 element	Measuring the ambient temperature
Temperature machine	PT100 element	Measuring the temperature in the machine
Temperature cooling liquid	PT100 element	Measuring the cooling liquid temperature

Cube calibration

The cube probe is used to calibrate the tools before machining. Figure 2.2 shows how the tool is being calibrated. The lowest box is the cube probe which has a fixed position in the machine. A is the zero point of the machine. The machine zero point is a defined point located in a fixed position at the origin of the machine coordinate system, and it cannot be moved. The machine knows the length of $Zmax$, which is the length from A to the cube. Zm is the length from A to the chuck of the machine where the tool/probe is installed. The interesting thing is the length and radius of the tool. Acquiring Zm makes it possible to calculate $Ztool$ (the length of the tool). After the length of the tool is acquired, the radius of the tool is measured by moving the tool to each side of the cube. The length of the tool can also be manually measured outside the machine, and it depends on the functionality of the machine. The radius and length are then implemented in the part program to compensate for the length and radius to avoid machining too little or too much. The data used in this report is based on the calibration of the cube probe. The description of the calibration of the tool is to get a better understanding of what the cube is used for. The cube probe is calibrated with a tool that has the known length and radius. Also called a master tool, it has different shapes depending on how many sides of the cube probe is calibrated and used in the machine. The principle is the same for the calibration of the probe as it is for the cube. Only when calibrating the cube, a master tool is used.

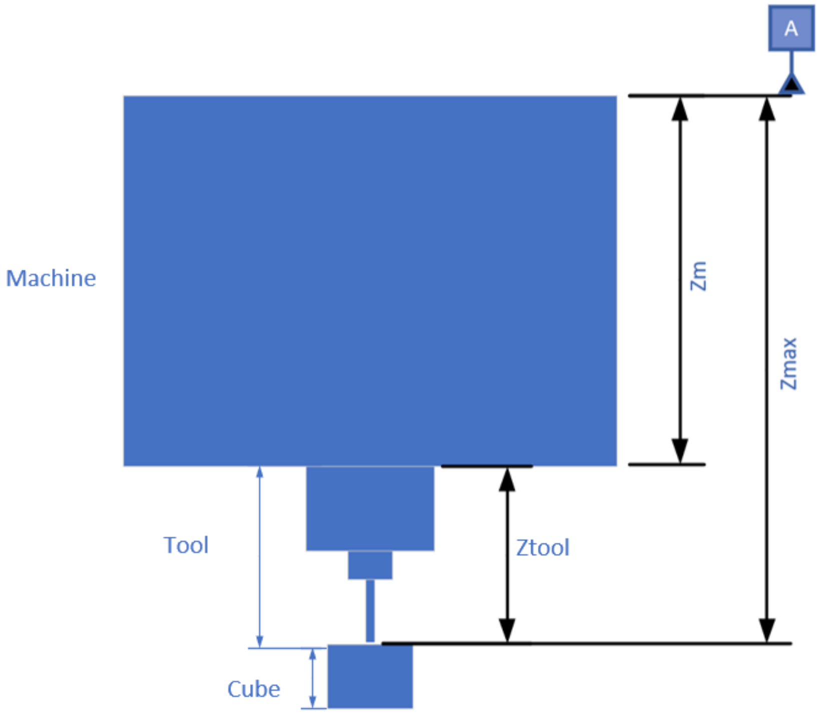

Figure 2.2: Visualization of how the calibration is performed. All the components are included in both the tool and cube calibration.

The cube probe used in the calibration is usually a metal piece connected to a probe that has a lot of geometric shapes. In this case, it is a square. When the tool is calibrated there is a small distance the probe is moved before it triggers, and does the measuring. Figure 2.3 shows the square shape gathered when using all the sides of the cube probe. The notation is described in the next section on the structure of the dataset. The k in the figure indicates which tool/probe is measured. The numbers in the figure are described in Table 2.3. k.2,k.1 indicates the corner of the cube in the top left corner, and the other corners are acquired in the same principle. The corners are used to see if there is any skewness/ellipse form of the tool. This is gathered for the analysis of the calibration data.

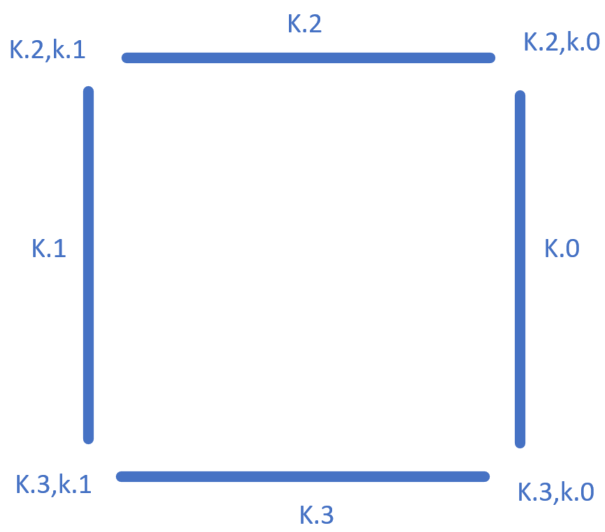

Figure 2.3: Visualization of the different side of the cube. The first code indicates which tool/probe is being measured, and the second indicates the coordinate measured.

2.1.2 Structure of the calibration dataset

The structure of the dataset is decided by the standard from the system used in the machine. As GAN has CNC machines from different decades, there are different systems for the different machines. In the Carnaghi machine, Sinumeric control is used with the control table 840D. Figure 2.4 is a visualization of how the measurements are structured.

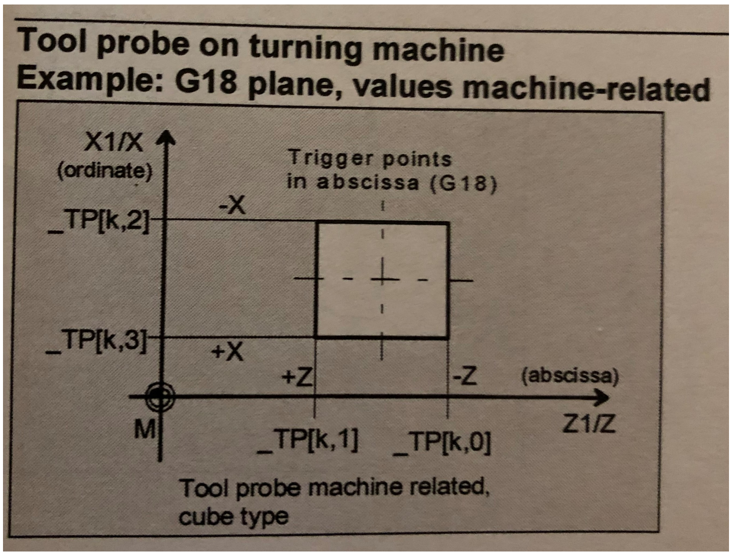

Figure 2.4: Visualization of the calibration values of turning machines - Cube probe. The description is given in Table 2.3.

For reference, the G18 plane mentioned in Figure 2.4 is the reference used in the machine. A different plane is used in different machines. Figure 2.5 shows the orientation of the different planes. *A* is the zero point of the machine. G18 is mostly used in a 3-axis turning machine [8].

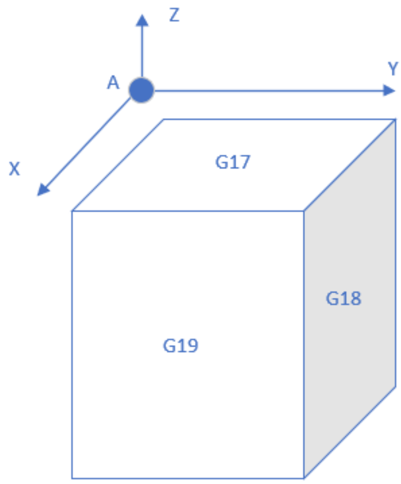

Figure 2.5: Visualization of the plane used in the Carnaghi machine.

The index k indicates a specific tool/probe in the machine. The number is unique for the machine. This means that for one line in the dataset the k is the value for a specific tool/probe for that machine. The number after is in the description of Table 2.3, is for the calibration of tools or probes [9]. _TP[k,4] to _TP[k,9] is not available in the dataset. As it is not measured, the measurements are mentioned due to the assignment of the other ways the probes are calibrated in other machines. The directions Absicca and Ordinate are shown in Figure 2.4.

Table 2.3: Calibration values of milling machines - Cube probe. Measured from the machine zero point.

Code	Description	Direction
_TP[k,0]	Trigger point in minus direction	Absicca (Z1/Z)
_TP[k,1]	Trigger point in plus direction	Absicca (Z1/Z)
_TP[k,2]	Trigger point in minus direction	Ordinate (X1/X)
_TP[k,3]	Trigger point in plus direction	Ordinate (X1/X)
_TP[k,4]	Irrelevant	NA
...		
_TP[k,9]	Irrelevant	NA

2.2 Prevailing manufacturing system used at GKN Aerospace Sweden

For a better understanding of the methods used and the results acquired, the following is a description of part of the manufacturing system in GAS. The machine used is a GROB milling machine and the measurement data is from the bearings and runout data on the spindle in the machine.

2.2.1 Machine - GROB milling

The machine chosen for GAS is a GROB milling machine. This is due to the available data and the possibility to use data related to troubles with the spindle. This way it is possible to base the method of analysing on data with a known fault and not only normal manufacturing. The data used is from late December 2019 until late April 2021. The machine is ideal for complex components where high cutting volumes are required [10]. The spindle is a part of the machine that rotates the tool when milling the product.

Diagnostic electronics for vibration sensors

Measuring of the data is done through the VSE 100 system. According to IFM ”It is a 6 channel diagnostic unit for the evaluation of dynamic signals and analogue inputs. Flexible, detailed monitoring allows the diagnosis of specific machinery faults. Ethernet TCP/IP and Profinet interfaces for connection and integration to higher level systems”. [11] The VSE collects the data every time the part program is performing a warmup cycle of a process. The bearing is rotating at 3000 revolutions per minute (rpm).

2.2.2 Bearings of the spindle

The bearings are a part of the spindle that turns the cutting tool in the machine. The acceleration is measured in mm/s^2. The warmup program is at 3000 rpm which is 50 Hz (Hz = rpm/60). The Grob machine is built for 12000 rpm which gives a frequency of 200 Hz. The warmup program is used to make sure that the state of the machine is acceptable to do the process. Verifying all levels and measures are acceptable for machining. Figure 2.6 shows a simplification of the placement of the bearing rings in the motorized spindle. Bearing 1 is the front bearing, bearing 2 is the bearing preload, and bearing 3 is the rear bearing.

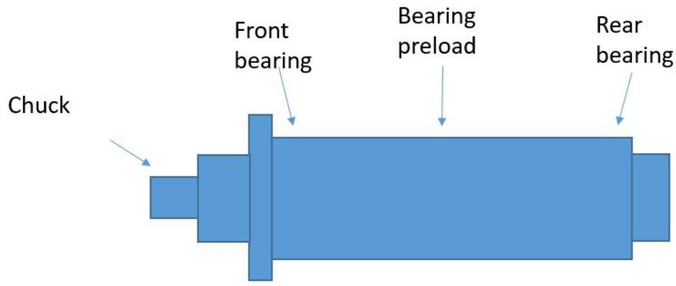

Figure 2.6: Simplification of the placement of the bearings on the spindle of the machine.

2.2.3 Runout of the spindle

The measuring of runout is to control the location of the spindle relative to its axis. If the axis of the spindle is off the axis, the machining may remove more material on the engine part than required or machine uneven on the engine part, resulting in a rework of the engine part or, in the worst case, scrapping. The tolerances set is about a deviation of 70 μm. The tolerances on aeroplane engine parts depending on the operation and type of part are from 1:100 of a mm to 1:1000 of a mm. Figure 2.7 shows on a simplified motorized spindle how the runout affects the spindle. When the spindle has an offset of a few degrees from the main axis it is called axial runout, meaning the axis of rotation is no longer parallel to the main axis. When the spindle is parallel to the main axis, but offset it is called radial runout [12]. The runout is measured with gauges on the spindle at a different location to identify the runout. Radial runout is measured closer to the mounting of the spindle and the axial runout further down the spindle. The difference in the gauges gives the axial runout and the highest value of the two gives the radial runout.

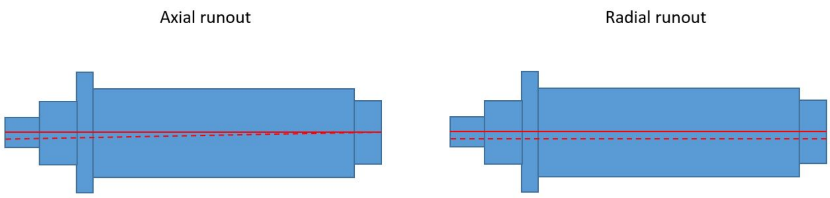

Figure 2.7: Simplification of runout on the spindle. Both axial and radial.

2.3 Different maintenance methods

Maintenance is a crucial part of the manufacturing industry. There are multiple ways to do it, some more costly than others. A factor for all methods is the possibility for a form of unpredictable failure of the equipment. This section describes in short terms the different methods of maintenance in the industry, at the sites of GAN and GAS.

2.3.1 Preventive maintenance

This type of maintenance is periodic and performed in a specified interval. This makes it a high cost compared to the other methods since there is a possibility that the equipment being changed or serviced would have been able to manufacture for longer without any form of interaction. It is also a cost that the machine is not available for manufacturing during maintenance. A positive feature of the preventive method is the cost of repair is low. This is due to a high frequency of maintenance and only repairing unpredictable failures.

2.3.2 Predictive maintenance

Being able to predict the need for maintenance cuts both the cost for repairs and maintenance on the machine and equipment. The need for sensors, formulas and historical data is essential to build a good model for the system. Another essential thing is knowledge of the machine and the process of machining the part. This is the only case if the data is not properly labelled. A crucial part of modelling the predictive system is to make sure that the data used is verified and validated to be true. PdM is mainly modelled with different methods in NN, DL and ML.

2.3.3 Condition-based maintenance

PdM and CBM are relatively similar. Both require sensors, historical data and preliminary knowledge of the system. The difference is that PdM uses the data to predict when the system is in need of maintenance, and CBM uses the data to define tolerances with different methods used on the systems, usually with real-time data. When the parameters of the machine exceed the tolerances, maintenance work is performed.

3 Method to validate the acquired measurement data from the machines

Through testing of the data many different visualizations and analytics software has been used. Mainly to test if the software which is already internally implemented has the flexibility needed to do the method tested in the report. Through this chapter, the pros and cons of the different internal software are described. This chapter shows the way the data has been analysed and preprocessed to use with the method chosen. The analytical method is also described in detail. Figure 3.1 shows the data flow of the calibration data today. The stippled line indicates the wish to get the data automatically to the engineer. This requirement is not part of the project, because of some challenges found during the project regarding the lack of resources and the structuring and formatting of the data.

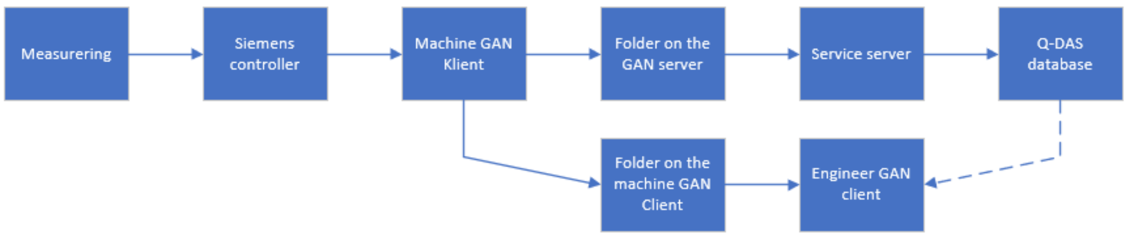

Figure 3.1: Rough data-flow diagram of the data from the measurements from the calibration to the database.

3.1 Analysis of calibration and bearing data

An important part of analysing manufacturing data is to have a sufficient understanding of what the dataset contains. Through meetings with engineers and visual analytical tools, this is attained and described in this section.

3.1.1 Visualisation analysis - Calibration data

The dataset used is sampled between 2020 to 2021 for this analysis. The distribution of each corner is shown in Figure 3.2. The cause of the different clusters for each point is a movement of the cube probe used in the calibration. The change to the coordinates given in the figure is taken into account with a new calibration with the master tool on the new placement of the probe cube. The new coordinates in relation to the machine zero point are added to the part program. In GAN's case, it would be a good routine to check the deviation as part of the maintenance to see if there are some irregularities in the data.

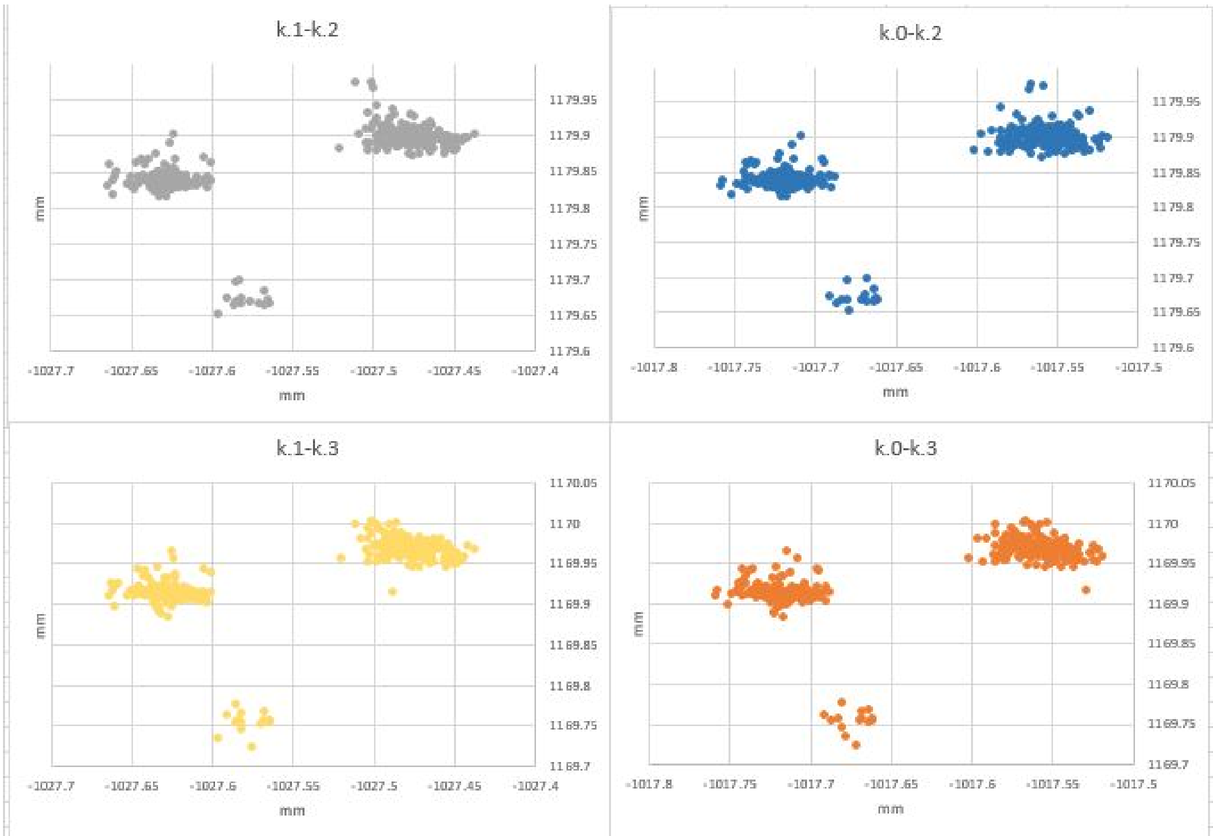

Figure 3.2: Point cloud of the data - Distribution of each corner. The axis is the length in mm from the zero point of the machine.

Figure 3.3 shows the size of the cube and the deviation in comparison to the cube. The point of this analysis was to see if it is possible to look for skewness in the shape made from the points in the dataset. Since all points are made on the two connecting sides, it is not possible to have skewness in the data, only movement of the shape itself. The datapoints on the cube are also measured in one point of the sides. Some deviation is possible in the area of the measurements, but that is due to dirt on the probe or the tool. The shape of the cube is a square, measuring about 1 x 1 cm. The measurements used is the length of the cube subtracted with the trigger distance of the probe.

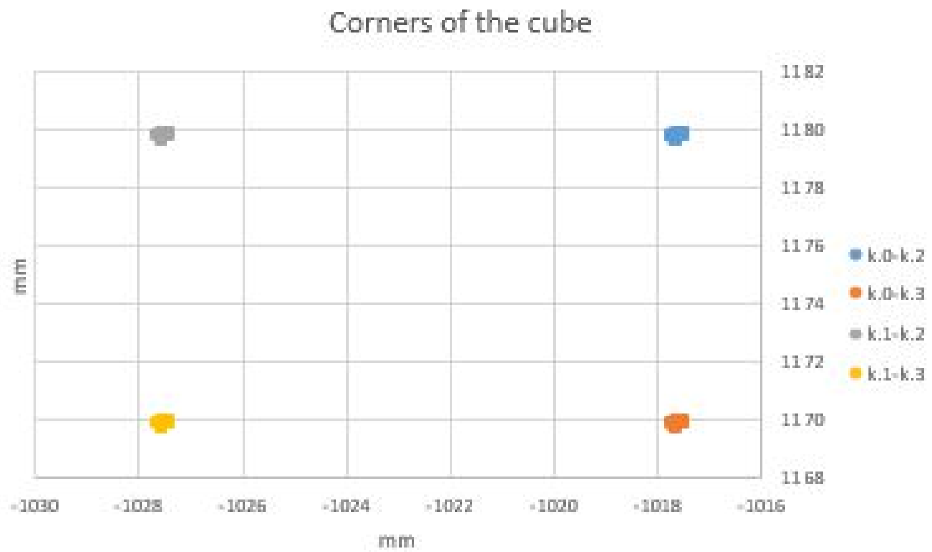

Figure 3.3: Point cloud of the data - Each corner. The axis is the length in mm from the zero point of the machine.

Figure 3.4 shows a graph of source temperature from between 21.12.2020 to 08.06.2021. The temperature is inside the Carnaghi machine (Machine), ambient temperatures of the shop floor mounted on the back of the machine (Ambient), the temperature of the low (KJ_LT) and high(KJ_HT) pressure cooling liquid tank. As the machine is located almost in the centre of the shop floor, it is not impacted if the large overhead doors leading to the outside are opened. The temperatures are added to be sure of this hypnotises. The frequency of acquiring these datasets are the same as the calibration data. It is collected each time the probe is calibrated. This may be multiple times during a process, it depends on the process and the engine part which is machined.

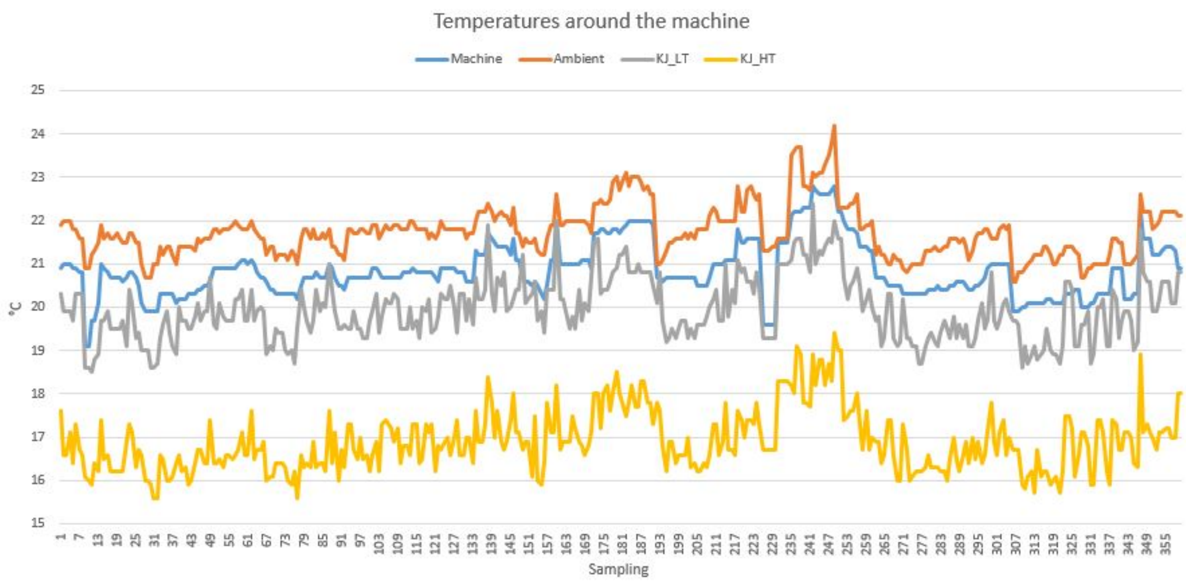

Figure 3.4: Visualization of the temperature in the machine, cooling liquid tank and ambient. Measured around the Carnaghi machine.

After collecting the data from the calibration of the cube shown with the source data in Figure 3.2. Point $k.1$ is subtracted from $k.0$. To give the horizontal length of the cube probe. The same is done with $k.3$ and $k.2$ to get the vertical length. Figure 3.5 shows the distribution of the lengths. Analysing these lengths can show how much the size of the cube probe deviates. Since both the cube and master tool do not change in size, the deviation should be non-existing. The deviation is a combination of the repeatability of the probe and the amount of dirt on the probe. The repeatability is set to 2 μm [13]. The magnitude of the deviation is 0.12 mm, which can make a big impact on the engine part being machined. If the change were to be implemented into the part program, major damage to the part or machine could occur. The cause of the considerable deviation around 14.04.21 (230 in the graph) is not known. Most likely dirt during the calibration, it is not possible to verify as there are no registrations of deviations in the datasets.

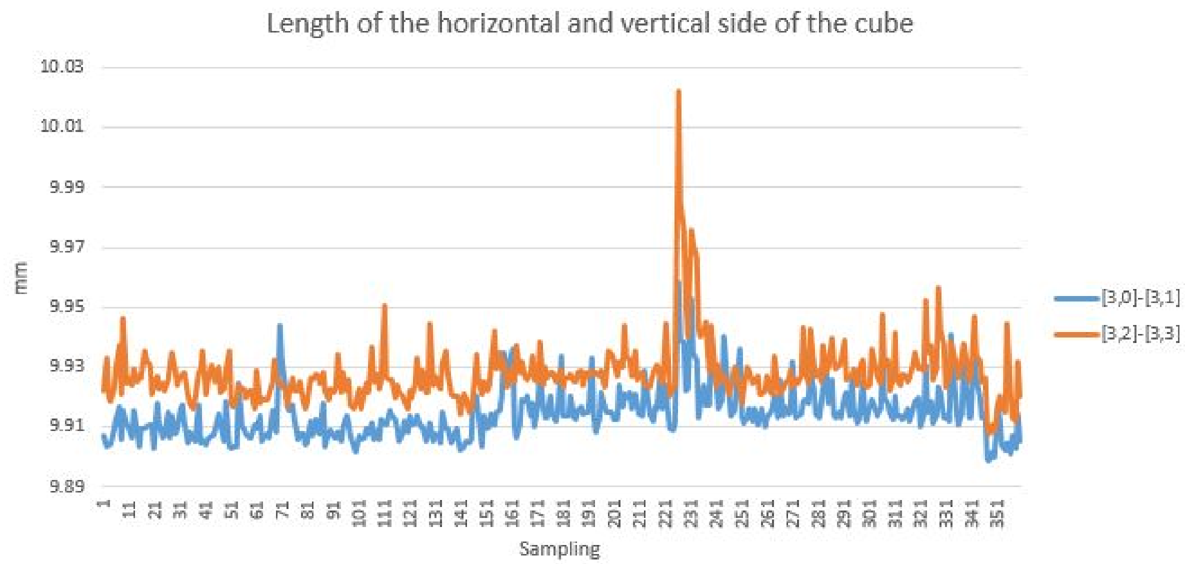

Figure 3.5: Distribution of the length of each side of the cube. Including a fault/unpredicted deviation.

Using the data from Figure 3.4 and Figure 3.5 the correlation is shown in Figure 3.6. The calculated correlation is -0.105 which is a considerably low negative correlation. Pearson's r is used to calculate the correlation. The mathematical formula is in Appendix B. The data with the length of the sides of the cubes are multiplied to give the area of the cube. The data is scaled with Z-score ($z = \frac{x-\bar{x}}{\sigma}$) to see the actual correlation with the degrees and the area of the cube. In the graph, it is showed that to a certain degree, when the temperature rises, the area of the cube decreases, and if the temperature lowers the area rises. This method has also been somewhat used by the engineers at GAN to see if it is correlated with the temperature and the area of the cube. The fact is that the cube itself is measured outside the cube. This is counter-intuitive to what the data indicates, since the area of the cube should be lowered when the temperature is lowered. There has been no result on why this is the case after investigating with an engineer at GAN.

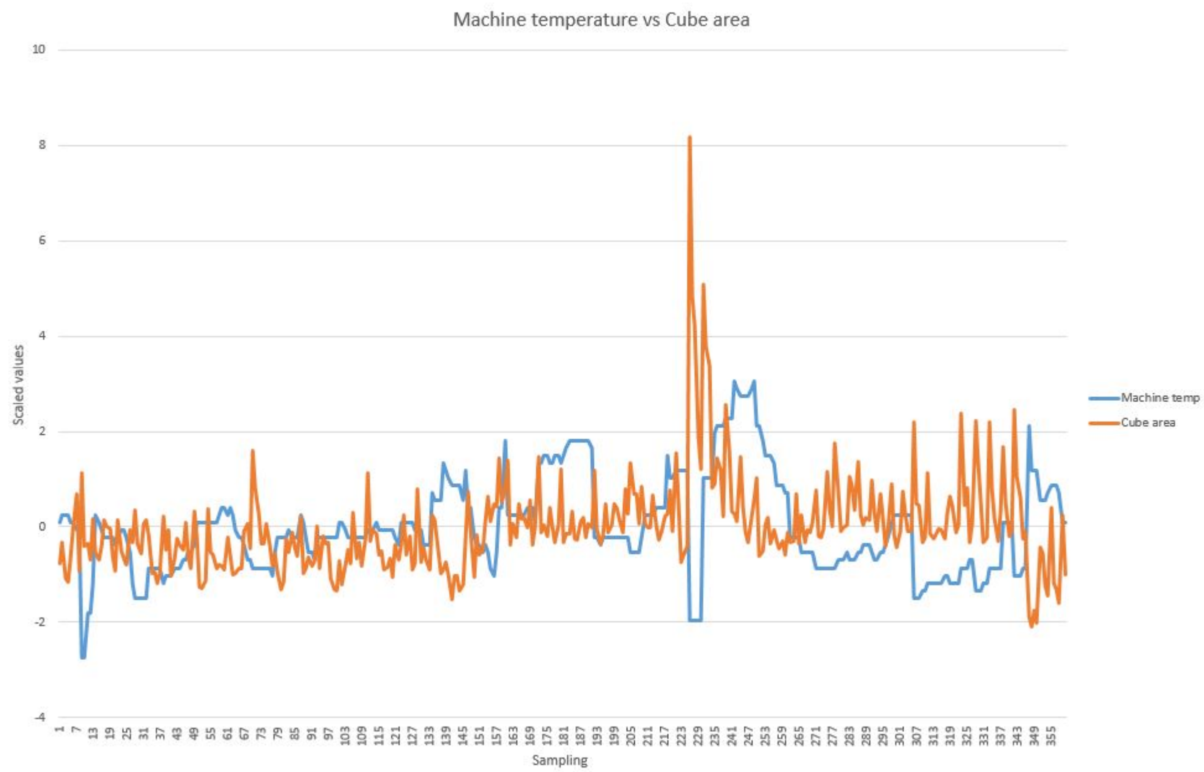

Figure 3.6: Visualization of the correlation between temperature and area of the cube with a dataset with a duration of about 6 months. Including a fault/unpredicted deviation.

Panda data profiling - Cube calibration and temperature

As a part of the source data analysing of the data Pandas data profiling tool was used. The code for the panda data-profiling is included as Appendix C. Figure 3.7 shows Pearson's r. Pearson's r is a measure of linear correlation between two variables in the analysed dataset. The correlation lies between -1 and 1, where -1 indicating perfect negative correlation, 0 indicating no correlation, and 1 indicating perfect positive correlation. For more description of the data-profiling see Appendix D.

The correlation between the length of the sides of the cubes is not surprisingly correlated. The same is the case with all temperatures. With the temperature of the cooling liquid (T_KJ_LT and T_KJ_HT) and one of the lengths [3,0]-[3,1] is mildly correlated as is the length [3,2]-[3,3] and the machine (T_Machine) and ambient (T_Ambient) temperature. This is to such a small degree it is not taken into account for the analysis.

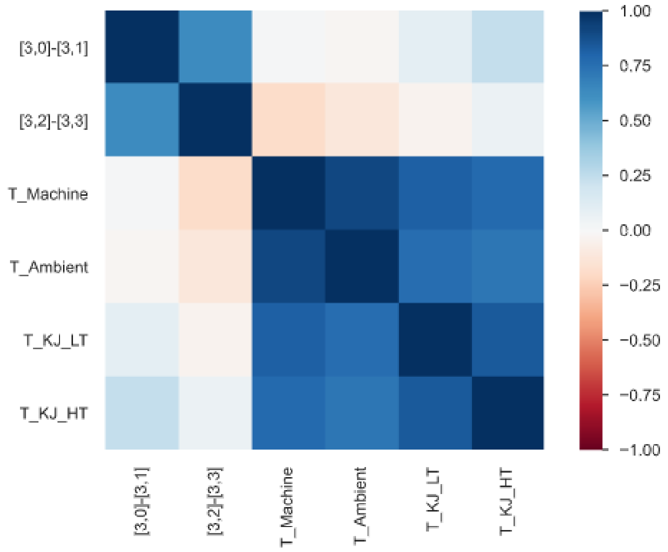

Figure 3.7: Pearson correlation diagram of the calibration and temperature data.

3.1.2 Visualisation analysis - Bearing and runout data

Another test with the method chosen during this project is to use vibration data from the spindle in a GROB machine and the runout data from the spindle to see the runout of the axial front and the radial front. These data sources are gathered during the warmup of the spindle of some different processes.

Copilot visualization tool

These graphs are gathered from the visualization tool called Copilot. Figure 3.8 shows the data collected from late December 2019 until the middle of March 2021. Indicated in the middle of March this year the vibration of the machine started to rise over what is normal manufacturing. Information from the engineers at GAS indicates that the spindle was changed due to runout as viewed in the lower graph of the figure in the same period. The high peaks after is due to trouble when installing the new spindle. As the process of calibrating the probe is more intricate than vibration data, more analysis has been done on the calibration.

Figure 3.8: Source vibration data (upper part of the graph) and runout for the axial and radial front (lower part) with a duration of about 15 months.

Panda data profiling - Vibration and runout

Using the Panda data profiling tool on the vibration and runout data gives a clearer correlation than between the cube calibration and temperature, as shown in Figure 3.9. Bearing 1 and 2 have a higher correlation with the axial and radial front runout data. This is due to the fact that bearing 1 and 2 are mounted on the spindle further from the connection between the machine and spindle. This causes the runout to have a higher impact on the vibration of the bearings than bearing 3 that is mounted near the base of the spindle. As bearing 2 and radial front are more than minimal in relation. The correlation has been calculated to -0.53 which is a moderate negative correlation. This may indicate that the radial gauge is placed near the position of bearing 2. This has not been verified. For more description of the data-profiling see Appendix E.

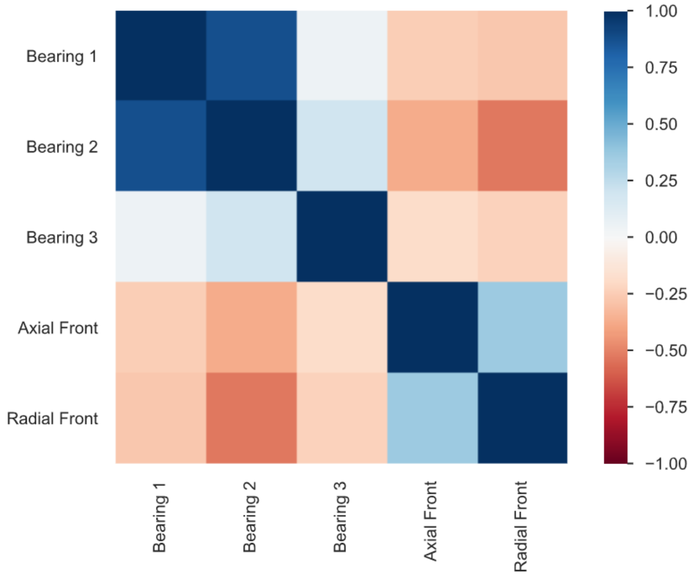

Figure 3.9: Pearson correlation diagram of the vibration and runout data.

3.2 Testing of internal software

This section is on the testing of the software already implemented internally, to see if the software can implement the method used in the report. The software tested is GM, Tableau and to some degree Qlik Sense. Because of some restrictions and difficulties with licenses, Qlik Sense was not tested as thoroughly as the other alternatives.

3.2.1 Green monitor visualization tool

GM is used today to show machine state (is the machine running in cut-off or idle), operation data (How long is there left of an operation, which operation and which engine part is being machined), operator registration for faults or stops in the machines and sensor visualisation (temperature, level of fluid in tanks, vibration and voltage). GM also has the possibility to make its own dashboard and use the data available from some of the databases at GAN. The dashboards are flexible to a certain degree to make customizable graph information tiles with some degrees of freedom. Table 3.1 shows the functions available for the dashboard.Simple mathematical functions are also available. It is possible to use the following applicable functions:

Table 3.1: Useful functions in the visualization tool GM [14].

Function	**Description**	**Unit**
now	Current local time	date/time
TimePerValue	How much time the values of the pointed tag have been meeting the condition in the period between, from and to time. The result time is converted to a decimal number	tagname, condition and date/time
PreviousValue	Returns the previous value of the tag just before the specified time.	Dependent on the tag
MaxForPeriod	Returns the maximal value of the referred tag in the specified period.	Dependent on the tag
MinForPeriod	Returns the minimal value of the referred tag in the specified period.	Dependent on the tag
GoOverLimit	Returns true if tag goes over max limit, otherwise false	True/false
GoOverOrEqualToLimit	Returns true if tag goes over or equal to limit, otherwise false	True/false
GoUnderLimit	Returns true if tag goes under min limit, otherwise false	True/false
GoUnderOrEqualToLimit	Returns true if tag goes under or equal to limit, otherwise false	True/false

When using the data, it is often not possible to get the unit of value that is gathered. This means the user must have the knowledge of the tag to know what is measured. If this had been included when collecting the data it would be more flexible and easier to use by engineers who are not using GM on a daily basis. Since there is some flexibility in the GM tool, it is possible to use the method used in this report, but it requires some changes in formatting, availability of data and more work resources. It is not the best alternative, but possible.

3.2.2 Tableau visualization tool

Tableau is one of the analyzing and visualization tools available to most of the engineers at GAN. During the testing, it came clear that it was hard to use Tableau to show these kinds of data using NN, DL or ML. The software is better utilized on organisation, economy or value stream data and not production or measurement data. The reason the tool was tested is that it is easy for other individuals in the company to access the data from all over the company. It is also easy to implement other sources to the software.

3.3 Development of the analysing method

The data is scaled with a preprocessing method called standard score, also called Z-score. The formula for the standard scaler is shown in Equation 3.1:

$$x_{scaled} = \frac{x - \bar{x}}{\sigma} \tag{3.1}$$

This centres and scales the data variable independently on each feature. This is by calculating the necessary statistics on the sample, to normalize the data. The normalisation is done to contribute to a more equal model fitting when all the different scales of the dataset are the same. This also avoids the dataset having bias from any of the variables. As a method of decomposing the dataset Singular value decomposition (SVD) is used. It takes high dimensional data and distils it into the key correlation of that dataset. The formula for SVD is Equation 3.2 in Appendix B [15]. In this method, SVD is used to compute the Principal component analysis (PCA). SVD is used to decompose the dataset into two principal components (PC). The PC is used in the PCA to find out the dominant direction of variance in the dataset.

The covariance matrix is calculated with Equation 3.2:

$$C(x,y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})^T \tag{3.2}$$

where x and y are two PC, with $\bar{x}$ and $\bar{y}$ being the mean of the two. The covariance is the measure of joint variability of the two PC.

By using the covariance matrix and $\vec{x} = (x_1, x_2, x_3, \ldots, x_n)^T$ and $\vec{y} = (y_1, y_2, y_3, \ldots, y_n)^T$ vectors we get Equation 3.3 [16]:

$$D_M^2(\vec{x}, \vec{y}) = (\vec{x} - \vec{y})^T C^{-1} (\vec{x} - \vec{y}) \tag{3.3}$$

Where:

- D is the Mahalanobis distance (MD)
- $\vec{x}$ is a vector of PC 1
- $\vec{y}$ is a vector of PC 2
- C is the inverse covariance matrix of independent variables

Using the result from Equation 3.3 it is possible to see if the square of MD is following a Chi square distribution if it is assumed that that the variables are following a normal distribution. The distance is given by Equation 3.4.

$$D_M(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T C^{-1} (\vec{x} - \vec{y})} \tag{3.4}$$

Figure 3.10 is a visualisation of the processing of the data done with SVD, PCA and MD. It moves the origin to the centre of the dataset and uses the PC to move the coordinate system to a proper orientation so the dataset is not as distorted as the ellipse indicated. When redrawing the coordinate system, the scatterplot has a circle and using the standard deviation as the unit and makes the horizontal and vertical unit equal. Following the empirical rule (Appendix B - Equation B.2) on the standard deviation as the new unit, it is easy to see if the datapoints are outliers or inside the normal manufacturing. An outlier is treated as an irregular circumstance, and further measures need to be done to make sure the source of the dataset is healthy. The further measures depend on the data source. If there are some irregularities with the spindle as an example. Another warmup can be performed to make sure it is running as normal.

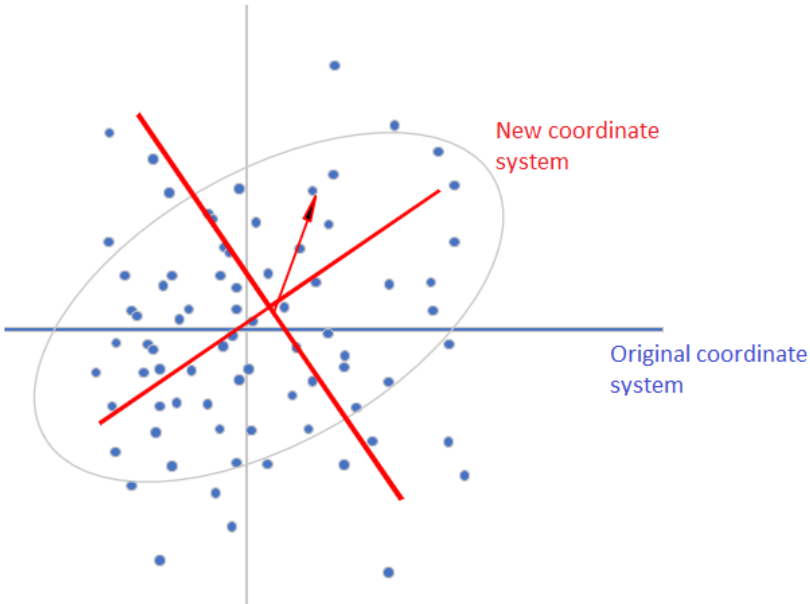

Figure 3.10: Example of the use of MD. The original coordinate system is the distribution of the source dataset. The red coordinate system is the new coordinate system moved to the centre of the datapoints.

The code used is based on a test with FD and CBM on gear bearings degradation. The dataset is collected from NASA, on gear bearing active usage from start until breakage [17]. The datasets chosen in this project are based on showing the availability of data from two sites and show the possible results from the different datasets.

4 Results from the method

Using MD with PCA made it possible to get a more analytical approach to FD, than applied today. CBM is also possible to implement, if real-time data is implemented as input. The code used is added in Appendix F. The method makes it possible to see when the chosen equipment starts to deviate from normal manufacturing. Section 4.1 of this chapter includes the results from the method used on the calibration and temperature data and section 4.2 is on the bearing data with the runout data. Due to the preprocessing of the dataset, the time is changed to sampling in the dataset. This is due to some structural difficulties. This is the case for both the dataset from GAN and GAS.

4.1 Probe calibration and temperature analysis

Figure 4.1 shows the horizontal and vertical length of the cube on the left of the graph, and on the right side is the temperatures of the machine, ambient and cooling liquid for the low- and high-pressure tank. Take notice of the units and the lines through the graph, since there are two units, one line fits the left axis, and the other the right axis. Based on the source dataset, 60% is chosen to be used as the training set of normal manufacturing. That means the last 40% of the dataset is the test set. This is due to the time of the greater deviation in the dataset.

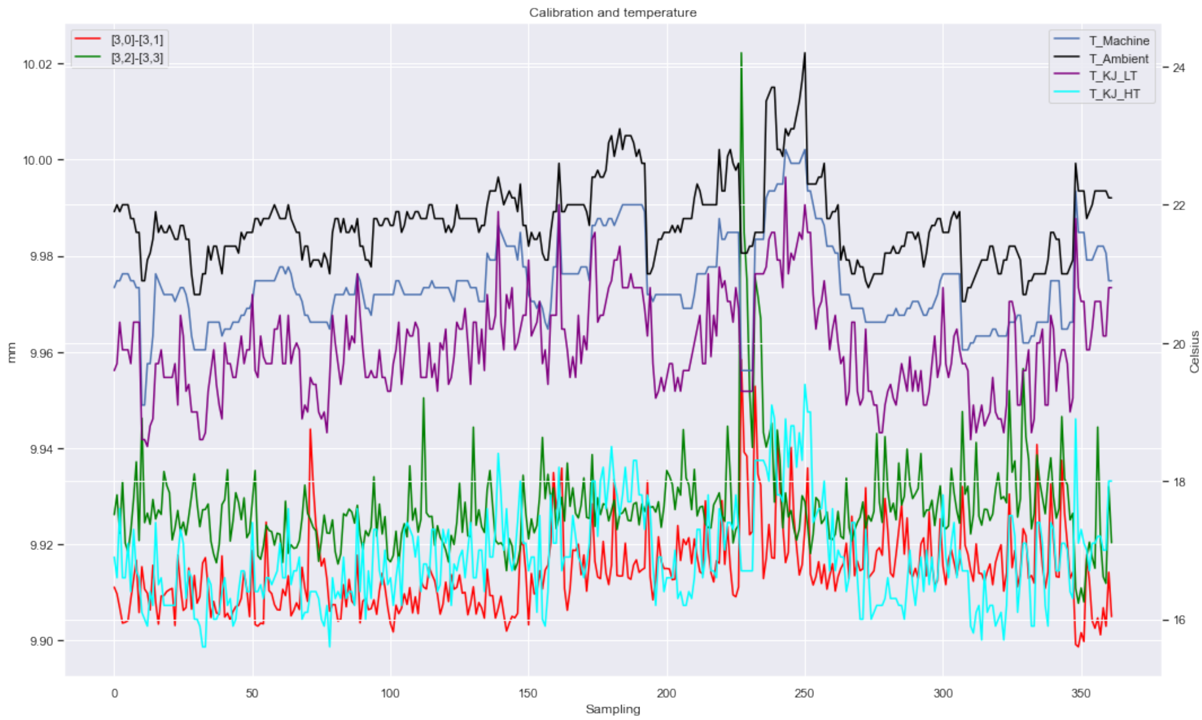

Figure 4.1: Source calibration data from an excel sheet from the machine used by the engineer.

Figure 4.2 shows the square of the MD, and that the data follows a chi-square distribution. This shows that the input variables also are following a normal distribution. The graph is a visualisation tool for the user to validate the quality of the dataset, and if the method is viable for testing.

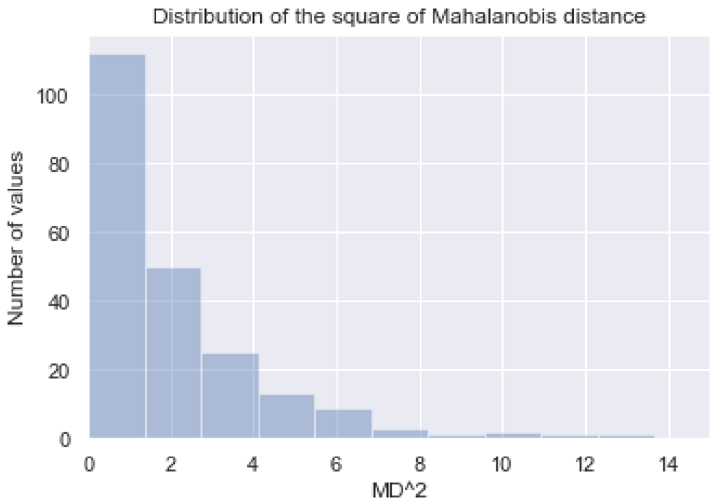

Figure 4.2: Visualization of the square of MD. With the distribution of the values used on the calibration and temperature dataset.

The distribution of MD is shown in Figure 4.3. The graph is used to verify the threshold calculated by the use of the empirical rule on MD. In this case, the threshold is calculated to be 4.599. $3*\boldsymbol{\sigma}$ is used to validate that 99.7% of the datapoints are inside the tolerances. When comparing this to the graph, it correlates with the distribution.

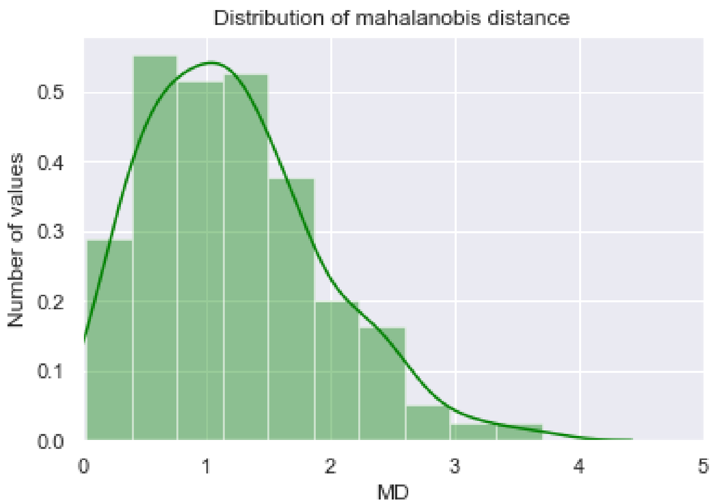

Figure 4.3: Visualization of the distribution of MD. Used on the calibration and temperature dataset.

Figure 4.4 shows the outliers of the dataset used with the calibration and temperature. The outliers are marked with purple circles to show more clearly when the outliers occur. When comparing these outliers with the source data in Figure 4.1, it only finds outliers in the same period as the deviation in the source dataset. This is thought to be because the calibration and temperature data have a low correlation, and that the temperature data only adds noise to the method.

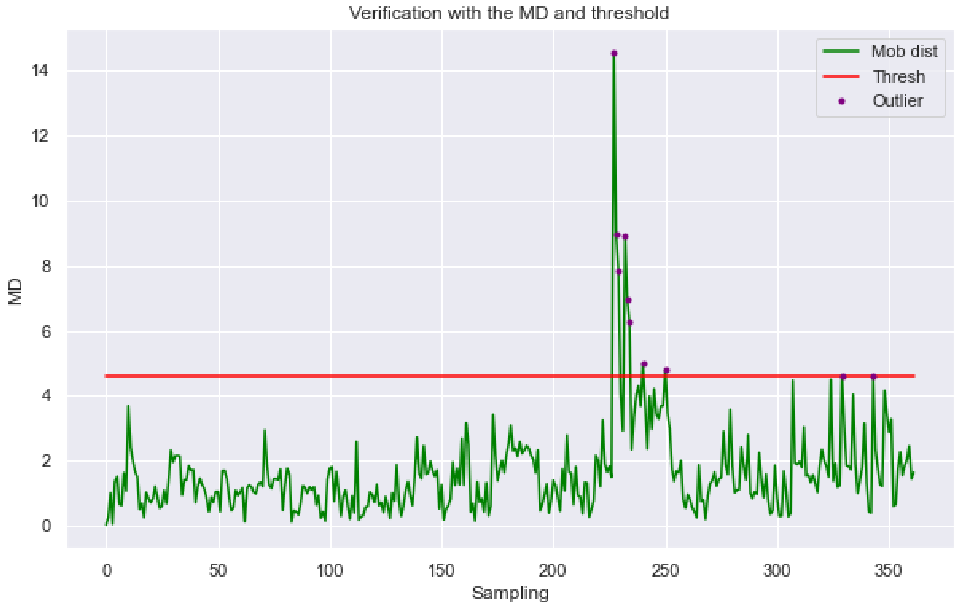

Figure 4.4: MD with the threshold set by the MD calculations on the calibration and temperature dataset.

When applying the same method to only the calibration data, more outliers are visible. The threshold is calculated to be 3. The outliers indicate a greater deviation of trigger length caused by dirt, unknown parameters or poor repeatability of the probe. This indicates that it is not viable to use calibration and temperature data in this method shown in Figure 4.5. Due to the fact that the tolerances set in the method previously used, also acquired about the same amount of outliers. However, the method applied in this report is based on statistical methods used for FD. The method used today is time-consuming and very manual. Comparing this method with FD today would cut the cost of the time spent on analysing.

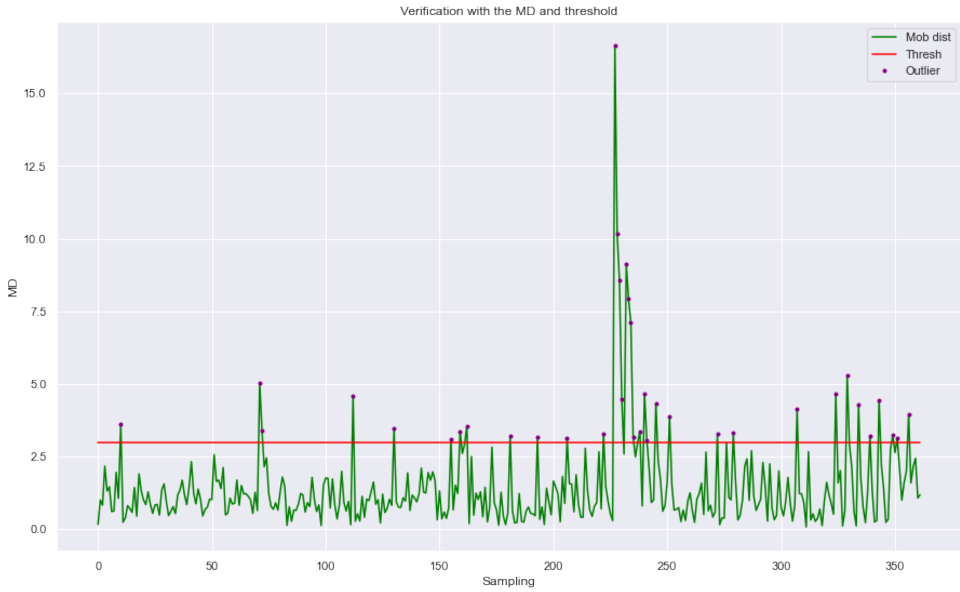

Figure 4.5: MD with the threshold set by the MD calculations on the calibration dataset.

4.2 Bearing and runout analysis

Figure 4.6 shows the runout of the axial- and radial front of the spindle on the right of the graph, measured in μm, and the vibrations of the 3 sets of bearings on the left side measured in mm/s^2. Take notice of the units and the lines through the graph, since there are two units, one line fits the bearing data on left axis, and the other the runout data on the right axis. Based on the source dataset, 60% is chosen to be used as the training set of normal manufacturing. That means the last 40% of the dataset is the test set. This is due to the greater deviation in the dataset. This is mentioned for both datasets since it depends on the dataset.

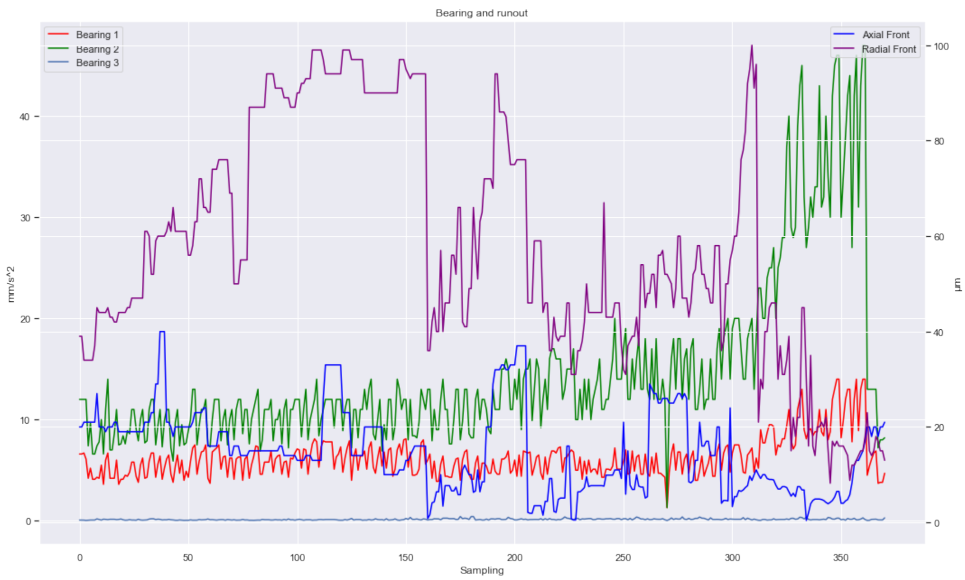

Figure 4.6: Source bearing and runout data from Copilot. The axis to the left is the vibration and to the right is runout data.

Shown in Figure 4.7 is the MD of the bearings and runouts dataset. The threshold is calculated to be 3,752. As bearing 2 and the radial runout have a somewhat negative correlation on the source data, both the bearing and runout is added, after the results of the test for the calibration. When comparing when the MD crossed the threshold the first time, and the time the spindle needed to be changed due to runout, the analysis shows already 12 days before, that the spindle had an irregularity. A longer period before the outlier, it is visible that the manufacturing was more irregular than before, but not substantial. Including a warning tolerance could catch the irregularities sooner and make the engineers more aware of the deterioration. The machine was manufacturing every day, until the change of the spindle. All the values after the change are not taken into account due to installation troubles with the new spindle.

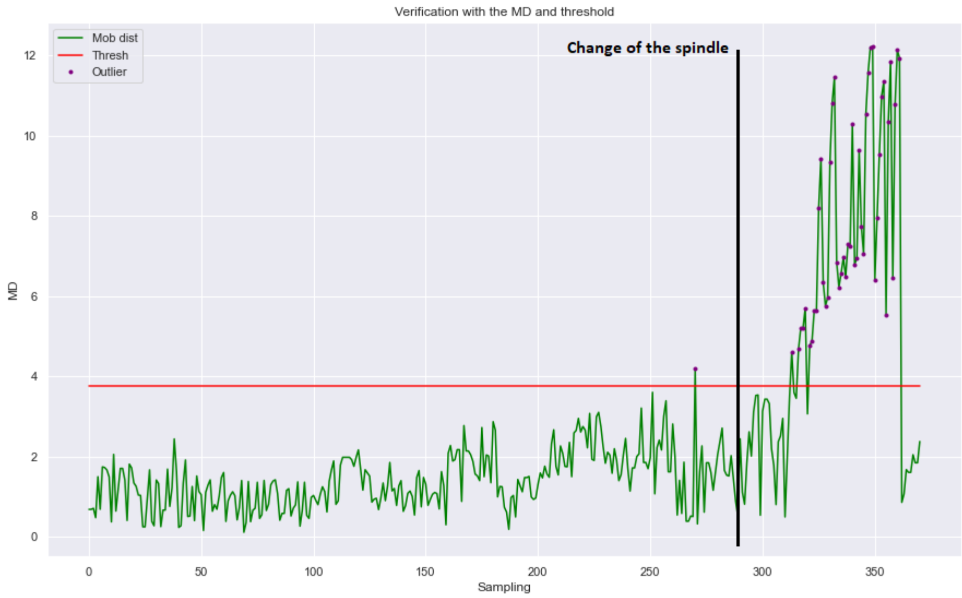

Figure 4.7: MD with the threshold set by the MD calculations on the bearing and runout dataset.

Figure 4.8 shows the same pattern as the calibration data. If there are limited or no correlation between the dataset used together in the method, this method is applicable. However, it is limited and applies too high tolerances, so it is not advised to use in active manufacturing. With the method applied to only the bearings, the graph shows that the bearings are not at normal manufacturing a while before the fault in the spindle. The threshold is calculated to be 3,526, and the first outlier occurred about 9 month before the fault. If the method was implemented on real-time data, precautions could have been made to avoid the change of the spindle. Due to some preprocessing in the dataset with the bearing and runout data, there are more samples in that dataset than bearing alone.

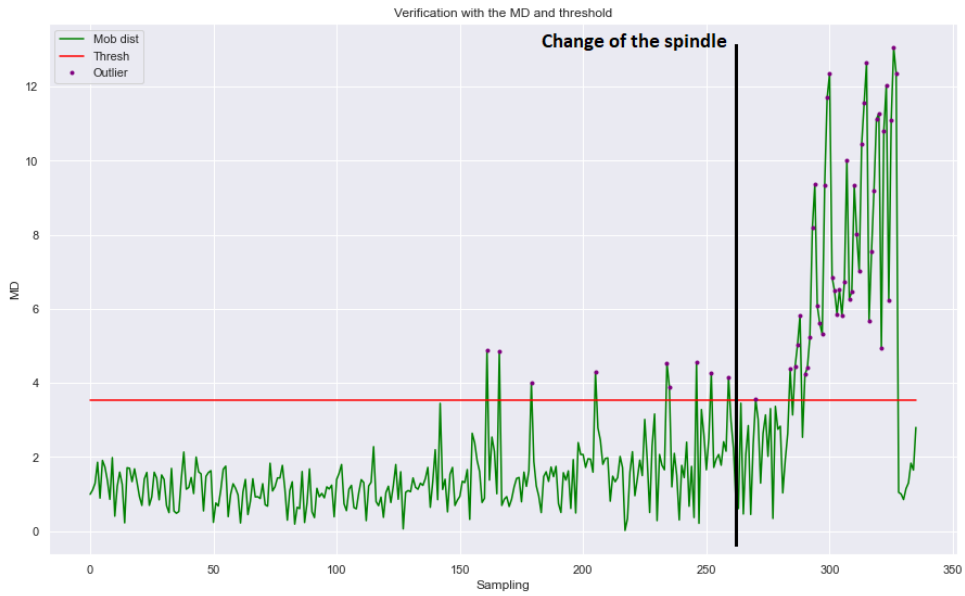

Figure 4.8: MD with the threshold set by the MD calculations on the bearing dataset.

5 Discussion

In this chapter, the challenges during the project, a short description of recommendations of requirements regarding changes to the different software used today, conclusions of the results of the project, and recommendation for further work.

Through the project, it is visible that the data sources used were not viable to analyse together. More data analysis should have been performed earlier in the project to achieve a better result. This is regarding being able to use different data sources together to be able to achieve a fuller view of the state of the machine. The method used in the report was able to identify irregularities before the prevailing method, but this is based on single sources of data. By implementing the method on real-time data with the current visualization tools would make it easier for the engineers to keep track and base the tolerances set on a statistical method, instead of a trial and error method. The method from the report should be able to set the tolerances to be included in the part program. This way it would be able to stop the machine in time, before the fault. Depending on the tool used it may be some computational limits or restrictions on the software used to implement the method. This was not possible to test during the project to any extent, due to lack of reaources.

5.1 Challenges identified during the project

During the project, some challenges occurred. These were both pre-existing challenges and new challenges due to COVID-19. Since GAN is an older company, there has been some employee turnover through the years. This causes many technical systems and solutions to stay without a super user (a person responsible for the system/solution). Due to this, there are different data structures and data flows on the different systems on the shop floor. In addition, there is a lack of documentation describing these structures. As such, there has been some challenges in the project regarding data acquisition, formatting issues and prioritization of resources. Figure 5.1 gives an overview of the challenges.

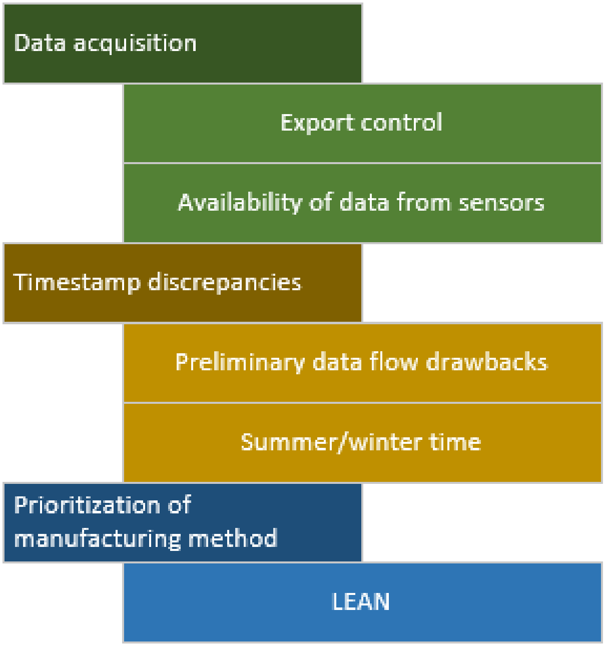

Figure 5.1: Overview of the different challenges that surfaced during the project.

There are also strict IT policies at GKN, especially regarding aeroplane manufacturing of military parts for multiple countries. All the countries handling information of these engine parts have to have a nondisclosure agreement. This also applies to software. The IT department wishes only to have commercial off-the-shelf (COTS) software to make sure that the security is within the requirements of the export control (the security of the data) of the customer. As changes to the processes in manufacturing happen, there are often changes to the way information is handled. With these terms, there is no software developing at GAN. Therefore the priority for this report is the analysing method and exposing challenges when implementing data analysis, and the requirements for implementing new data treatment software.

5.1.1 Data acquisition and export control

Acquiring data from a diverse manufacturing company such as GAN shows that over the year many different systems and data transfer protocols have been used to analyse or acquire the data. This caused some issues with the analysis done for the report. Another difficulty was the lack of availability of an overview of the sensors mounted on the machines. This made it hard to choose the machine that had the most diverse sensors installed. Based on recommendations from the engineers and looking through the databases to find which of the machine had the most available data, the Carnaghi turning machine was chosen for the calibration data. When using Q-DAS to verify the available sensors, many of the listed sensors had no data.

Export control is important for GAN since many of the parts are military and has to follow regulations set by the nations buying those military parts. Sharing production data is a part of the export control. Measures were taken at the beginning of the project to not void GAN's export control policies, when acquiring data both from GAN and GAS. These measures had to be verified with the person responsible for export control at GAN, to make sure that the data was acceptable to analyse for the public. This was also applied to the data from GAS.

5.1.2 Challenges related to data synchronization

Through the analysis and structuring of the data, some challenges showed regarding the timestamp of the dataset used for the calibration. Earlier there has not been any part of the process to change the time or check if the time on the control panel is correct. This made the dataset acquired as a CSV-file having the wrong timestamp when compared to the other databases in GAN. This made it almost impossible to apply it together with data from other sources. The data from Q-DAS also have the chance to get the wrong timestamp in the Q-DAS database, since this data is sent from the control panel to the GAN client, which then is collected from a folder to a server software that quality checks the data. After the quality check it is sent to the Q-DAS database. The server software has a chance to go offline and does not have any means of automatic restart. A locally made program sends a message to the engineer who developed the quality check software. Then it is up to chance that the engineer is at work or not to restart the server application. If the server application is offline for a longer period of time, the new datasets are stopped by the application. When it is restarted, all datasets will get the timestamp the application has when it restarts.

The dataset from GAS also had some challenges related to data synchronization. When exporting the data from Copilot there are not the same number of datapoints and the measurements are different since it is collected from different sources. Using one of the datapoints as the "master data", the other one is a few hours and some minutes offset. Summer time and winter time are taken into account in only one of the sources. In the summer time, it is offset by two hours, and by one hour in the winter time. The data used in this report is structured to not be affected by these discrepancies. This way, the datasets from the bearings are corresponding to those that were sampled on the runout datasets.

5.1.3 Prioritization of manufacturing method

Implementation of the LEAN manufacturing philosophy has been a prioritized activity for several years in GKN. Lean manufacturing is a manufacturing method where descriptive specifications, a good overview of the flow of the product, decrease faults in processing and small improvements are key factors [18]. LEAN is not as much based on using data to help manufacturing, such as finding small or medium improvements on a day-to-day basis. There have been projects before on using data from the machines to acquire the best reliability and uptime of the machines, but too few are still being used in manufacturing. This is because there are too few working actively on using data in manufacturing. Starting this project with little previous work or documentation makes it hard to acquire data. With prioritization and fixed resources on analysing and developing methods, it would cut cost on maintenance and have more predictable and improved manufacturing.

5.2 Recommendations for requirements when implementing new data analysis and acquiring software

One of the aims of the report was to define examples of changes to the prevailing system at GAN to enhance the quality of the method used. This section describes the flexibility needed and limitation of current and new software, and also registration of changes and faults in the machines. A major advantage would be to have a digital overview of the content of the different machines on the shop floor. Then it is easier to know what kind of sensors are available when doing different tests on the machine. This way it would be possible to implement CBM in the company. These are some requirements that should be taken into account on existing machines and when starting a new project, implementing new machines or visualisations/data treatment software in the company.

5.2.1 Flexibility

In a company like GAN, a data acquisition and analysis tool should be available to all engineers working on improving manufacturing. Not necessarily for all to implement new methods, but to have access to contribute with professional opinions and to encourage knowledge sharing. The software should be able to work with many different data sources. It should work on production data (Which operation is next or how long does it machine) as well as sensor data (Temperature, voltage, runout or geometrical data), and be able to analyze the data with user-specified methods. The methods used in the different software's available at GAN today is very limiting.

5.2.2 Limitation

At many work stations at GAN, there are manual measuring, like measuring the radius of a engine part. This also means there is manual registration of the measures. Today, these manual registrations have too much freedom when entering information. This makes it difficult when trying to correlate data with other data sources which collect data automatically. This is due to the quality of the data depends on the operator doing the measurement.

5.2.3 Historical

Registration of cause of fault, operator maintenance, scheduled maintenance or changes to machines and equipment is an important part when doing data analysing. Being able to compare irregularities in the dataset and a description of what has been done to the machine at any time will make it possible to have FD with the cause of the irregularities. Since there are so many parameters concerning the manufacturing of an aeroplane engine part, a valuable result is to be able to distinguish an anomaly from a known fault.

5.3 Conclusion

Regarding the first aim of the report of being able to identify faults/irregularities from the data acquired before the irregular machining causes faults in the machine, it can be stated based on the results from the testing.

From the results given in the testing of the method on calibration and temperature data, it is possible to detect faults or anomalies based on the datasets. Although through the testing it came clear that it is not viable to use the geometrical data from the cube with the temperature in and around the machine. This is due to the fact that the temperature is causing noise to the analysis. When looking at the results from the method used only on the calibration data, a more viable solution was visible. The threshold was calculated to a much more viable level when taking into account the tolerances acceptable to the cube. The calibration data might not be the best use of machine data for accessing the state of the machine, but it is an important part when machining. If there are no validations of the state of the probes in the machine, there would be more rework of the parts, which increases the work hours and cost per engine part.

Similar results can be seen in the results of the bearing and runout of the GROB machine. When applying the method only to the bearing data the irregular manufacturing was indicated much sooner than when implementing the runout data as well. Rather than indicating irregular machining in a single outlier 12 days before the actual changing of the spindle, using only the bearing data the irregular machining was visible for the first time about 9 month ahead of the changing of the spindle. The irregularities are recurring on a regular basis after that, until the change of the spindle. Since this dataset was made available late in the project, it was not possible to compare these irregularities with the quality of the engine parts machined at these periods.

Regarding the second aim of the project of being able to identify which data source from the machines gives the best result about the state of the machine.

When comparing the results of the two analysis, the calibration process and the vibration data of the bearings, the analysis shows that the quality of the result is better when using vibration data. Being able to use this method on real-time data would give a good indication of the state of the spindle in the machine. Since all data is collected during the warmup sequence of the process, it does not give a high frequency of data. This means that this method is only deployable on a machine if it has historical data. GAN and GAS have the possibility to use this method as a CBM instead of the preventive maintenance schedule used today. This gives the machine more uptime and less cost of maintenance.

If the data synchronization challenge was fixed, it would be possible to compare the quality of the engine part with the irregularities in the dataset. This could have caused a rework of the engine parts. This is the case for both the calibration and the bearing dataset.

5.4 Future work

After the results of this report, it is recommended to do some further research into these four topics:

GAN and GAS have multiple different types of machines on the shop floor. This leaves many possible areas for future research. Establishing a complete register of machines with the installed sensors and equipment would make it easier to find fitting sensors for further analysing. If the year and current availability of machining was included, it would also help compare what machines to prioritize when analyzing. It also makes it more clear if some machines could benefit from mounting a new type of sensor.

Implementing the tested method in the machine to verify and validate that it is viable to be used in live testing. These measures are from the non-machining part of the process, so the risk of causing troubles in the cycle is close to none.

Look into comparing the time of irregularity with the engine parts machined to look if the irregularities have affected the quality of the engine part. The irregularities could cause more rework in later operations.

If further analysis should be prioritized in the future. A solution to the time synchronisation challenge must be researched to be sure that all data sources are synchronized to the same time. Being able to have a service/solution to synchronize time on the GAN-clients, control systems, machines and server. This would make it possible to compare data from different sources.

References

[1] MTU Aero Engines, *Engine programs*, Document, Dachauer Straße 665, 80995 Munich, Germany, 2021.

[2] Renishaw. (). '"what is a probe?," accessed mar. 16, 2021,' [Online]. Available: `https://www.renishaw.com/en/what-is-a-probe--32937`.

[3] K. A. S. Guldbjørnsen, *Multivariate analysis in advanced machining (maam)*, Project report, Porsgrunn, Norway, 2020.

[4] C.-Y. Huang and Z. Dzulfikri, 'Stamping monitoring by using an adaptive 1d convolutional neural network.,' *Sensors 2021*, no. 262, p. 21, 2021.

[5] K. Zhang, W. Huang, X. Hou, J. Xu, R. Su and H. Xu, 'A fault diagnosis and visualization method for high-speed train based on edge and cloud collaboration.,' *Appl. Sci.2021*, no. 1251, p. 11, 2021.

[6] Cloudwards. (). '"what is edge computing: The network edge explained," accessed may. 1, 2021,' [Online]. Available: `https://www.cloudwards.net/what-is-edge-computing/`.

[7] F. Cipollinia, L. Onetoa, A. Coraddub, A. Murphyb and D. Anguitaa, *Condition-based maintenance of naval propulsion systems with supervised data analysis.* Paper, Genova, 2018.

[8] GAN, *Grunnl. forståelse av maskin- og styringskonsepter*, Teaching document, Kongsberg, Norway, 2014.

[9] Siemens, *Sinumerik 840D/840Di/810D User's Guide Measuring Cycles*, 10.04, Siemens, Ed. Germany: Siemens, 2004.

[10] Grob, *Grob brochure g-series - 5-axis universal machining centers*, Brochure, Germany, 2020.

[11] ifm, *Ifm vse 100 software manual*, Software manual, Sweden, 2021.

[12] Celera motion. (). '"what are radial and axial runout error?," accessed may. 4, 2021,' [Online]. Available: `https://www.celeramotion.com/applimotion/support/faqs/what-are-radial-and-axial-runout-error/`.

[13] Renishaw, *Lp2 modular probe system for tool setting and workpiece inspection*, Data sheet, 2011.

[14] T. Koleva, D. Mashov, V. Mitev and K. Avdjiev, *Green monitor user manual*, User Manual, 2020.

[15] S. L. Brunton and J. N. Kutz, *Data driven science engineering*, Book, Washington, 2017.

[16] R. Maesschalck, D.Jouan-Rimbaud and D.L.Massart, 'Chemometrics and intelligent laboratory systems,' *Elsevier Science B.V.*, vol. 50, no. 1, pp. 1–18, 2000.

[17] Towards data science. (). '"machine learning for anomaly detection and condition monitoring," accessed feb. 16, 2021,' [Online]. Available: `https://towardsdatascience.com/machine-learning-for-anomaly-detection-and-condition-monitoring-d4614e7de770`.

[18] J. Womack and D. Jones, *Lean thinking : Banish waste and create wealth in your corporation*, Book, New York (NY), 1996.

[19] Corporate finance institute. (). '"what is a correlation?," accessed mar. 20, 2021,' [Online]. Available: `https://corporatefinanceinstitute.com/resources/knowledge/finance/correlation/`.

[20] L. J. Kazmier, *Business statistics : based on Schaum's outline of theory and problems of business statistics*, 3rd ed. New York : McGraw-Hill, 2003.

Appendix A

Master’s thesis 2021 - GKN - MADAAM

Here is included the project description `Master's thesis 2021 - GKN - MADAAM`.

This description is an agreement between USN and the student for what the thesis should contain. Both parties have signed.

University of South-Eastern Norway

FM4017 Project

Title: Multivariate analysis and data acquisition in advanced machining (MADAAM)

USN supervisor: USN campus Porsgrunn, Håkon Viumdal

External partner: GKN Aerospace Norway AS, Håvard Norum

Task background:

GKN Aerospace is a multinational company producing both military and commercial aeroplane parts. GKN Aerospace Norway (GAN) is a part of Kongsberg industry park. GAN mainly produces shafts, turbine cases, exhaust cases, and vanes. Today, during production in the different machines there is some data which is collected manually from tool and part probing, and laser and geometric measurements. The data is collected from SAP, GAN's enterprise resource planning (ERP) system, Q-DAS which contains measurement data from probing and sensors in the machines and process data. The prevailing standard at GAN is to use these measurements to reveal early failures that are related to the tools. The interaction between these data points is also manually analysed with simple correlation. Calibrating of the tools in the machine are mainly performed in three ways; 1) with a probe, 2) with a "cube" probing and 3) with a laser. There are also different processing data from different sensors and geometrical data from coordinate measuring machine (CMM). The machines which use these measurement methods are multi-processing CNC machines.

The main goal of GAN is to further automate the data collection from tool measuring and IPG. This goal of this thesis is to analyse the data acquisition with prediction methods and automate the data acquisition. An important outcome of the automation of the data is to save time of the acquisition and be able to give an example of a predictive analysing system for GAN.

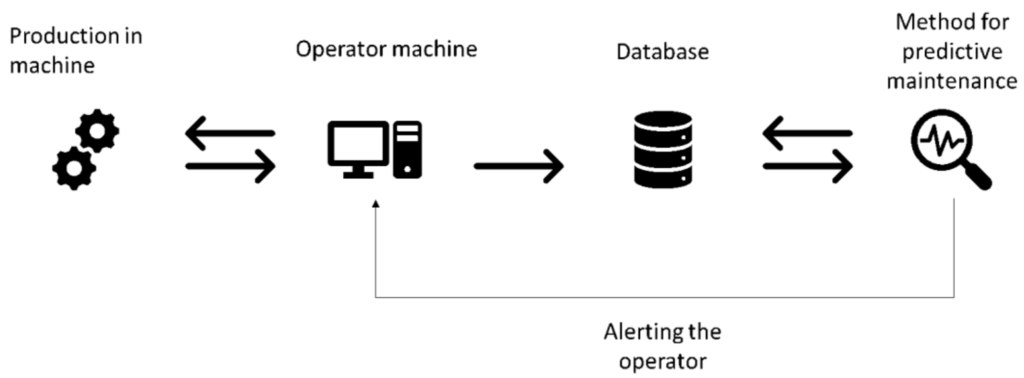

Figure 1 Rough system architecture

Task description:

- Give an overview of the specific process and tools under scrutiny.
- Define the problem related to the prevailing maintenance system.
- Perform a literature survey of similar predictive maintenance system.
- Gather data and perform exploratory data analysis.
- Design a prototype method for predictive maintenance system and data acquisition based on the measurement data and physical insight.
- Test the rough prototype to validate and verify the predictive maintenance system if it can be tested in the existing system at GAN.
- Define examples of changes to the prevailing system at GAN to enhance the quality of the predictive maintenance system.

Student category:

IIA Industry master. Reserved for Kjell Arne Stølen Guldbjørnsen.

The task is suitable for online students (not present at the campus): Reserved

Practical arrangements:

The work will be done at GKN offices. Since the student is already employed at GKN Aerospace Norway a non-disclosure agreement is already signed. The student will have access to necessary measurement data from GKN.

Supervision:
As a general rule, the student is entitled to 15-20 hours of supervision. This includes necessary time for the supervisor to prepare for supervision meetings (reading material to be discussed, etc).

Signatures:

Supervisor signature and date:

Location:

Student(s) signature and date: Kjell Arne S. Guldbjørnsen

Location: Kongsberg 04.02.21

Appendix B

Mathematical formulas

Correlation - Pearson's R [19]

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}} \tag{B.1}$$

Empirical rule [20]

$$1\sigma = 68\%, 2\sigma = 95\%, 3\sigma = 99,7\% (\approx 100\%) \tag{B.2}$$

SVD [15]

$$X = U\Sigma V^T \tag{B.3}$$

Appendix C

Panda data-profiling - Code

Included is the Python code for the Panda data-Profiling tool used in the method chapter. It is used to get a better understanding of the correlation of the datasets used both from GAN and GAS. It will give an overview of the size of the dataset, the distribution of the values and the correlation between the parameters.

```
#import the packages
import pands as pd
import pandas_profiling
import os

data_dir = 'Data'#Change this to the name of the folder for the file of the
    dataset
filename = 'Bearing2021'#Change to the name of the file
# read the file
df = pd.read_csv(os.path.join(data_dir, filename +'.csv'), sep=';')

# run the profile report
profile = df.profile_report(title='Pandas Profiling Report')

# save the report as html file
profile.to_file(output_file="pandas_profiling1.html")
```

Appendix D

Panda data-profiling - Calibration

All of the data-profiling for the calibration dataset is attached, see HTML: Panda_dataprofiling Calibration.html

Appendix E

Panda data-profiling - Bearing/Runout

All of the data-profiling for the Bearing/Runout dataset is attached, see HTML:
Panda_dataprofiling Bearing.html

Appendix F

Python code for the analysis method

The attached file ”MADAAM Python code” includes the code for the project, it is made in Jypiter Notebook in Python 3 code for testing of the method used in the project. It is just used as a testing tool and not possible to implement in any real-time systems. This is due to not having the proper information to implement the necessary protocols in the method.

The testing method collects the data from a csv file, merges multiple files if needed, splits the dataset into training and test sets, scales the data and then pre-processes the data. There are 4 functions in the code. Much of the other code is structuring and notations on the graphs.

The first function is ”cov_matrix(data, verbose=False)” which calculates the coveriance matrix with the dataset imported. verbose = false means that you do not get detailed logging information.

The second function is ”MahalanobisDist(inv_cov_matrix, mean_distr, data, verbose=False)” Which calculates MD. It uses the inverse of the coveriance method and the mean of the distribution of the training set.

The third function is ”MD_threshold(dist, extreme=False, verbose=False)”. Which is calculating tolerances for classifying an anomaly. The empiric rule has been used. The threshold is 3*sigma. Sigma being the standard deviation of MD.

The fourth function is is_pos_def(A). Which check the matrix for positive definite