Quality Data Management in the Next Industrial Revolution

A Study of Prerequisites for Industry 4.0 at GKN Aerospace Sweden

Robert Erkki
Philip Johnsson

Industrial and Management Engineering, master's level
2018

Luleå University of Technology
Department of Business Administration, Technology and Social Sciences
ACKNOWLEDGEMENTS

First and foremost, we would like to thank our supervisor Sören Knuts for his undaunted enthusiasm and curiosity. We hope that you find the final report an interesting read. Moreover, it is necessary to extend our thanks to everyone else working at GKN Aerospace Trollhättan, especially Alexander, Hans-Olof and Kasper. Finally, an acknowledgement to our respective families and Erik Lovén as well as opponents at Luleå University of Technology for seeing this thesis through.

Luleå, June 2018

Robert Erkki

Philip Johnsson
ABSTRACT

The so-called Industry 4.0 is by its agitators commonly denoted as the fourth industrial revolution and promises to turn the manufacturing sector on its head. However, everything that glimmers is not gold and in the backwash of hefty consultant fees questions arises: What are the drivers behind Industry 4.0? Which barriers exists? How does one prepare its manufacturing procedures in anticipation of the (if ever) coming era? What is the internet of things and what file sizes’ is characterised as big data?

To answer these questions, this thesis aims to resolve the ambiguity surrounding the definitions of Industry 4.0, as well as clarify the fuzziness of a data-driven manufacturing approach. Ergo, the comprehensive usage of data, including collection and storage, quality control, and analysis. In order to do so, this thesis was carried out as a case study at GKN Aerospace Sweden (GAS). Through interviews and observations, as well as a literature review of the subject, the thesis examined different process’ data-driven needs from a quality management perspective.

The findings of this thesis show that the collection of quality data at GAS is mainly concerned with explicitly stated customer requirements. As such, the data available for the examined processes is proven inadequate for multivariate analytics. The transition towards a data-driven state of manufacturing involves a five-stage process wherein data collection through sensors is seen as a key enabler for multivariate analytics and a deepened process knowledge. Together, these efforts form the prerequisites for Industry 4.0.

In order to effectively start transition towards Industry 4.0, near-time recommendations for GAS includes: capture all data, with emphasize on process data; improve the accessibility of data; and ultimately taking advantage of advanced analytics. Collectively, these undertakings pave the way for the actual improvements of Industry 4.0, such as digital twins, machine cognition, and process self-optimization. Finally, due to the delimitations of the case study, the findings are but generalized for companies with similar characteristics, i.e. complex processes with low volumes.
<table>
<thead>
<tr>
<th><strong>ABBREVIATIONS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cyber-Physical System (CPS)</strong></td>
</tr>
<tr>
<td><strong>GKN Aerospace Sweden (GAS)</strong></td>
</tr>
<tr>
<td><strong>Internet of Things (IoT)</strong></td>
</tr>
<tr>
<td><strong>Machine Learning (ML)</strong></td>
</tr>
<tr>
<td><strong>Partial Least Square (PLS)</strong></td>
</tr>
<tr>
<td><strong>Principal Component Analysis (PCA)</strong></td>
</tr>
<tr>
<td><strong>Quality Data Management</strong></td>
</tr>
<tr>
<td><strong>Statistical Process Control (SPC)</strong></td>
</tr>
<tr>
<td><strong>Zettabyte</strong></td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

1 INTRODUCTION .................................................................................................................. 1  
1.1 Problem Discussion ....................................................................................................... 2  
1.2 Aim ................................................................................................................................ 3  
1.3 Research Scope and Delimitations ................................................................................ 4  
1.4 Thesis Disposition ......................................................................................................... 4  
2 METHOD ............................................................................................................................... 6  
2.1 Research Purpose .......................................................................................................... 6  
2.2 Research Approach ....................................................................................................... 6  
2.3 Research Strategy ......................................................................................................... 6  
2.4 Techniques for Data Collection ..................................................................................... 7  
2.5 Data Analysis ................................................................................................................ 9  
2.6 Research Credibility ..................................................................................................... 9  
3 THEORETICAL FRAME OF REFERENCE ........................................................................ 11  
3.1 Industry 4.0 ................................................................................................................... 11  
3.2 Data Quality .................................................................................................................. 15  
3.3 Statistical Process Control ............................................................................................ 18  
3.4 Multivariate Analyse Techniques .................................................................................. 21  
3.5 Frame of Reference Conceptualised ............................................................................. 22  
4 EMPIRICAL FINDINGS ...................................................................................................... 24  
4.1 A Paradigm Shift in Quality Control ............................................................................. 24  
4.2 Quality Data Management ............................................................................................ 25  
4.3 Planning Ahead: Industry 4.0 ....................................................................................... 27  
4.4 Internal Case Studies of Quality Data Management .................................................... 28  
4.5 Case Example I: Electrolux Home Care & SDA ......................................................... 33  
4.6 Case Example II: GKN Aerospace Norway AS .......................................................... 34  
5 ANALYSIS .......................................................................................................................... 36  
5.1 Assessment of Quality Data Management at GAS ..................................................... 36  
5.2 Industry 4.0 and Internal Process’ Applicability ........................................................ 39  
5.3 A Multivariate Attempt on Process A & B .................................................................. 41  
6 CONCLUSIONS AND RECOMMENDATIONS ............................................................ 43  
6.1 Recommendations ....................................................................................................... 44  
7 DISCUSSION ...................................................................................................................... 46  
8 REFERENCES ....................................................................................................................... 48
APPENDICES

APPENDIX A: INTERVIEW GUIDE SYSTEM OWNER.................................................(2 pages)
APPENDIX B: INTERVIEW GUIDE METHOD OWNER...........................................(2 pages)
APPENDIX C: INTERVIEW GUIDE INDUSTRY 4.0-VISION......................................(1 page)
APPENDIX D: RESPONDENTS KNOWLEDGE LEVEL...........................................(1 page)
APPENDIX E: DATASET AND JMP OUTPUTS FOR PROCESS A.................................(2 pages)
APPENDIX F: DATASET AND JMP OUTPUT FOR PROCESS B................................(1 page)

TABLE OF FIGURES

Figure 1.1 – Interconnection between thesis disposition and research areas...........................5
Figure 3.1 – The 5C architecture for CPS-implementation ......................................................15
Figure 3.2 – The knowledge discovery process ........................................................................17
Figure 3.3 – A standard univariate control chart for monitoring a process ............................18
Figure 3.4 – An illustration of a multivariate scenario wherein a hidden outlier is found .........20
Figure 3.5 – The thesis’ theoretical frame of reference conceptualised ....................................23
Figure 4.1 – An illustration over the historical approach to quality control at GAS ....................24
Figure 4.2 – The current approach to quality control at GAS ...................................................25
Figure 4.3 – An illustration over the general quality data management at GAS .......................26
Figure 4.4 – A compilation of data visualised in the analysis tool ..........................................26
Figure 4.5 – Industry 4.0 visualised on a system level ............................................................27
Figure 4.6 – An illustrative overview of Process A .................................................................28
Figure 4.7 – Overview of data storage in Process A ...............................................................29
Figure 4.8 – An overview of Process B’s operations ...............................................................31
Figure 4.9 – A schematic overview of Process C .................................................................32
Figure 4.10 – Part-fit optimisation illustrated .......................................................................35
Figure 5.1 – Respondents view of Industry 4.0 ....................................................................40
Figure 6.1 – A proposal for a future state of quality data management at GAS .........................45
1 INTRODUCTION

From the dawn of time mankind has had an everlasting thirst for increased effectiveness and technological advancements. By harnessing the power of elements such as water and steam, the first industrial revolution took off at the end of the 18th century. Soon additional innovations such as electricity and programmable logic controllers followed which enhanced automation further. Today, mankind stands on the brink of yet another disruptive revolution.

The idea of man and machine, living in harmony and symbiosis, have long been seen as an idea stemming from science-fiction. However, as technology continue to advance further in the 21st century – cyber-physical integration might become a reality. The magnitude of cyber-physical systems (CPS) is signified by Rajkumar, Lee, Sha, and Stankovic (2010):

“CPS are physical and engineered systems whose operations are monitored, coordinated, controlled and integrated by a computing and communication core. This intimate coupling between the cyber and physical will be manifested from the nano-world to large-scale wide-area systems of systems.”

CPS combines deeply embedded computation and communication to interact with physical processes as to add new capabilities beyond the original physical systems (Wang, Törngren, & Onori, 2015). Albeit a premature technology, some argue (e.g. Lee, 2008; Rajkumar et al., 2010) that CPS has the potential to surpass the entirety of the 20th-century IT-revolution. According to Rajkumar et al. (2010) recent technological advancements can be held responsible for the future promise of CPS: the proliferation of sensors and computers with low-cost and increased capability, as well as the revolution of wireless communication and abundance of internet bandwidth.

The phenomena of ubiquitous computing networks form the Internet of Things (IoT) and is anticipated to unroll the fourth stage of industrialization. IoT was originally coined in 1999 to describe wireless communication through the usage of integrated sensors and computers (Wang et al., 2015). With assistance of Radio Frequency Identification (RFID) and sensors, ordinary “things” or objects forms embedded and invisibly networks around us (Gubbi, Buyya, Marusis, & Palaniswami, 2013). From a manufacturing perspective the benefit is twofold. Allowing field devices to communicate and interact with centralised controllers could synchronise production, while simultaneously enabling real-time responses through decentralised analytics and decision making (Boston Consulting Group, 2015).

With the advancement of sensors and development of IoT, generation of large-scale data is imminent. A white paper released by the International Data Corporation (IDC, 2017) forecasts the annual data growth in 2025 to 163 zettabytes. A number that is tenfold to the 16.1 zettabytes of data generated in 2016. This is what is commonly called as big data and the overwhelming majority of data will be driven by IoT-devices generating data in real-time (ibid.). The accessibility to all this information opens up for greatly enhanced understanding of processes and makes a solid foundation for decision making. Therefore, big data is applicable in a wide range of businesses, the manufacturing domain being one of them.

Altogether these technological developments arise as some of the key drivers of the political project denoted Industry 4.0. Initially a German project to secure its global leadership within manufacturing (Kagermann, Wahlster, & Helbig, 2013), the idea has since gained immense traction and spread globally. At its core, Industry 4.0 emphasis the integration of traditional manufacturing systems and CPS to achieve three objectives (ibid.):
1. Horizontal integration through value networks,
2. End-to-end digital integration of engineering across the entire value chain,
3. Vertical integration and networked manufacturing systems.

From a practical standpoint the Industry 4.0-concept might be perceived as just a futuristic project. However, in a large-scale global survey of the aerospace, defence and security industry conducted by PwC (2016) the participants responded that they are currently investing in Industry 4.0-projects. Within five years’ time, 76 percent of the respondents predict that their company will have reached an advanced level of digitisation. The study points out that the leading actors have gone from Industry 4.0-hype into making real investments in the area and expect to invest approximately 5 percent of their annual revenues into Industry 4.0-related projects the coming five years. The investment into these projects is estimated to increase productivity and lead to cost reductions, ultimately increasing the revenue created. Moreover, Boston Consulting Group (2015) estimates that Industry 4.0 can bring reduction in conversion costs between 20 – 30 percent in the area of component manufacturing.

1.1 Problem Discussion

A consequence of the transition to the digital era is the overwhelming volume of data and how to make sense of it (Lynch, 2008). Participants in a study conducted by PwC (2016) expressed their distress. Only one in six of the surveyed companies stated that they had advanced data analytics in place today. However, 82 percent believed that data analytics will be of great importance in five years. Researchers seem to share the same view regarding the increasing importance of data analytics in manufacturing. According to Wuest, Weimer, Irgens, and Thoben (2016) the main reason being the ongoing trend of rising complexity in the manufacturing domain. This can be seen in both the production line as well as the product characteristics (Wiendahl & Scholtesiek, 1994). In addition, the changing environment of customers’ demands creates further uncertainties (Monostori, 2003). Altogether this creates new challenges in designing traceable processes with the ability to frequently adapt to changes.

As shown in the background, the Industry 4.0-concept promises increased productivity and cost reductions. While many new concepts are introduced, such as CPS and big data, they come without a detailed explanation on how to achieve them. In the fear of missing out, many organisations are starting to invest in these projects. But when there is no clear path towards the goal it is unclear whether they are getting their money’s worth. Many concepts are getting immensely popularized and might not be applicable for all. Organisations needs to be able to sort through these concepts and understand them in relation to their specific processes.

Moreover, these concepts show promising application within quality management as measurement data becomes ubiquitous. Albeit no explicit definition exists, one possible approach towards Industry 4.0 is to view it as an intersection between traditional quality management and computer science. Within this merge of areas, it is necessary to review and understand the different users’ needs in order to translate it to a working concept. As such, this thesis’ problematizes this inquiry as the subject of quality data management. Referring to the comprehensive usage of data, including collection and storage, quality control, and analysis in the era of Industry 4.0.

1.1.1 GKN Aerospace Sweden

In this study a closer look at processes at GKN Aerospace Sweden (GAS) has been made in relation to Industry 4.0. GAS is a world leading manufacturer of aerospace components which can be found
in both commercial aircrafts and military fighter jets. However, many of their production lines are not characterized by either data driven processes or advanced data analytics. Today some processes rely on intangible know-how acquired through experience and without this knowledge the process performance can be endangered. Furthermore, new environmental guidelines can force the usage of different materials and chemicals which even the experienced personnel do not understand the impact of. The solution today is in some cases to make qualified guesses to see what works. A maturing idea at the company is to work towards data-driven production and predictive analysis.

In the case of GAS, a data driven production is further complicated by low volume production and demand for high tolerances. With volumes pending from 10 to 500 parts per product per year comes problems when applying statistical tools to a production line with GAS’ characteristic. Moreover, GAS has an established machine park in place with a varied lifespan. Therefore, the acquisition of a brand-new machine park in the nearest future is not an economic viable solution in order to evolve into a CPS-factory immediately. As such, the modernization of the production must consider the applicability of old equipment.

The product requirements come both from the external customer and the internal design department to ensure the products functionality. Flight safety standards require GAS to record some product data during the products lifespan. As such, it has naturally become a standard for the company to focus their quality assurance program towards measuring critical characteristics on the product. On the other hand, little effort is invested in storing other types of data. Data that can be useful in many types of analysis.

1.2 Aim

The aim of this master’s thesis is to investigate and analyse potential applicability of digitalisation trends within manufacturing. Hereafter denoted under its umbrella term Industry 4.0; the query at hand serve to examine and determine – from a quality perspective – useful tools and improvements brought by recent advancements. To fulfil its aim, the thesis seeks to explore four research areas of varied nature:

1. **Within a manufacturing context, clarify the data-driven perspective of Industry 4.0.**

   Industry 4.0 is merely an umbrella term for untapped potential within manufacturing brought by modern technological advancements. As such, a lot of ambiguity surrounds the term and this thesis aims to clarify in general what manufacturing could expect from this trend.

2. **Identify methods for collection and multivariate analysis of immense data volumes.**

   Recent advancements and the coming of Industry 4.0 is to a large extent depending upon digitalisation. Seeing how data is fundamental for deterministic measurements of quality, finding effective methods for collection and analysis of industrial big data is of pressing concern.

3. **Examine internal processes to understand the needs and applicability of a data-driven Industry 4.0-perspective.**

   To make use of the exploratory nature of the previously research areas, the study aims to apply the knowledge internally at GAS on a selection of processes. The idea being that it is necessary to review and describe the current situation in order to prepare suitable processes for improvements along the lines of an Industry 4.0-scenario.
4. **Propose recommendations for GAS’ quality data management with respect to Industry 4.0.**

Finally, research area 4 serve to conclude the thesis’ findings and determine potential applicability of Industry 4.0 at GAS. As such, conclusions should be based upon the thesis’ theoretical framework explored in research area 1 and 2 together with the descriptive analysis of research area 3.

### 1.3 Research Scope and Delimitations

This thesis is limited to only study the data-driven approach of Industry 4.0, specifically within a manufacturing context, from a perspective of quality management. As such, a lot of potential areas of implementation as well as other possible technologies found within the concept (e.g. Additive Manufacturing, Augmented Reality, Cybersecurity, Cloud Computing etc.) were disregarded. Thus, the theories found within the thesis is a selection of all available approaches to Industry 4.0. Lastly, the thesis is a mix of exploratory and descriptive phases and the results of the thesis is to an extent based on the current conditions shown at GAS and a sample of case examples. Therefore, theoretical generalisations and practical recommendations addresses similar industries of GAS’ characterisations.

### 1.4 ThesisDisposition

Following the introduction of the thesis, the dispositions continues with Chapter 2 Method, wherein the outline of the used methodology is described and justified. Upon its basis the thesis starts of its task of aim fulfilment by an investigation of academic literature, forming Chapter 3 Theoretical Frame of Reference. Furthermore, the continuation of the thesis divides into both external and internal empirical findings in Chapter 4. Altogether forming the basis for Chapter 5 Analysis and ultimately leading to the aim fulfilment of Chapter 6 Conclusions and Recommendations. Lastly, a discussion of the thesis implications and contributions, including a critical review of its credibility, is carried out in Chapter 7. Figure 1.1 illustrates how the thesis disposition correlates with its respective research area.
Figure 1.1 – Interconnection between thesis disposition and research areas
2 METHOD

2.1 Research Purpose

The classification of a chosen research purpose is usually defined by the conductors’ initial knowledge of the problem area (Patel & Davidson, 2007, p. 12). Saunders et al. (2007, pp. 133-134) considers exploratory studies to be suitable when the precise nature of a problem is unknown to the researchers. As this thesis’ research areas aim to inquire both the understanding of a concept, and its applicability in a specific context, a twofold research purpose was chosen. Seeing how the study initially sought to increase the understanding of Industry 4.0 and its related tools, the first two areas of interest came to be answered through an exploratory purpose.

To further deepen the study, the knowledge acquired throughout the exploratory phase was to be converted into practice with the latter RA’s. In order to do so, a descriptive approach was used. According to Robson (2002, p. 59), a descriptive research aims to “[…] portray an accurate profile of persons, events or situations”. Therefore, the necessity of a clear picture of the phenomena, which the researchers hopes to collect data on, is stressed by Saunders et al. (2007, p. 134). The authors thereby conclude that a descriptive study may be used as an extension to a previously conducted exploratory research.

2.2 Research Approach

This thesis was conducted by utilisation of both an inductive and deductive approaches. In practice, this meant that a combination of theories and best practices was studied in parallel to empirical findings and hypothetical tests. Adopting abduction as research reasoning allowed for an alternation between the two traditional approaches in a satisfactory manner. Moreover, the befallen choice of abduction allowed the thesis aims to be approached from two directions. In its initial and exploratory phase, the majority of data was collected through observations and semi-structured interviews. Thus, classified as a qualitative approach (Patel & Davidson, 2003, p. 119). However, as stated by Saunders et al. (2007, pp. 145-146), a mixed method research enables the usage of both qualitative and quantitative techniques and analysis, which came to be the case for this thesis. To aid the purpose of process qualification in the latter aims, elements of quantitative data was analysed. As such, the research was conducted through a sequential approach in order to triangulate the most important issues (Saunders et al., 2007, pp. 146-147).

2.3 Research Strategy

As the thesis purpose is to examine and determine the data-driven maturity of processes within manufacturing, a case study was deemed suitable as research strategy. Mainly due to the magnitude of the subject, delimitations were deemed necessary in order to conduct the thesis. The strength of a case study is the ability to review a specific process and thereafter extrapolate it to describe a larger context (Ejvegård, 2009, p. 35). According to Morris and Wood (as cited in Saunders et al., 2007, p. 139), the strategy of case study will be favourable if its goal is to gain a rich understanding of how a process enacts within a given context. Therefore, the strategy is most often used within exploratory research (Saunders et al., 2007, p. 139). In addition, data collection techniques may be various and used in combination. Usage of triangulation and revision of multiple sources of data are therefore common practice (ibid). Out of the numerous strategies (e.g. experiment, survey, case study, action research, grounded theory, ethnography, & archival research) listed by Saunders et al. (2007, p. 135), case study was the strategy that foremost aligned with the overall purpose.
2.4 Techniques for Data Collection

During the thesis two main methods for gathering of information were used: primary data was collected for the thesis’ purpose while secondary data includes data previously gathered by others.

**Primary Data: Interviews**

To gain understanding of the studied processes’ quality data management, interviews were predominantly used to collect data. In the spirit of an exploratory study, these interviews were mainly conducted with a non-standardised semi-structured format (Saunders et al., 2007, p. 313). In total, 16 interviews were held and the templates used during the interviews can be seen in Appendix A, B, and C respectively. In addition, throughout the thesis several unstructured interviews were carried out in an informal fashion. Classified as non-directive by Saunders et al. (2007, p. 312), these interviews gave the interviewee the opportunity to talk freely in relation to the topic. Its primary contribution to the thesis was to help frame the problem area and clarify uncertain observations. Due to the limited number of people with insight in the studied process, structured interviews or questionnaires were deemed inefficient.

As recommended by the literature (Patel & Davidson, 2003, p. 83; Saunders et al., 2007, p. 312; Ejvegård, 2009, pp. 51-52) the semi-structured interviews were, when approved by the interviewee, audio-recorded. Albeit subsequent transcription was time consuming; recording enabled the interviewer to uphold a consistent focus on the matter of questioning and exploration of the topic at hand. Otherwise if the interviewee denied the usage of a recording device, the work of interviewing and memorandum was divided equally between the authors. Moreover, all semi-structured interviews started off with a clarification of its purpose and contribution to the thesis. All interviewees were offered the possibility to read the transcription and point out eventual errors (Ejvegård, 2009, p. 52). In addition, unless otherwise explicitly stated by the interviewee, all candidates’ identities were kept confidential for the sake of general ethical etiquette (Saunders et al., 2007, p. 187; Ejvegård, 2009, p. 53).

Concerning the sampling of interviewees, the method of choice was aligned with the thesis approach and strategy, and therefore non-probabilistic. In contrast to probability sampling, non-probabilistic entail that generalisation of the population is possible, but not on statistical grounds (Saunders et al., 2007, p. 207). As such, the sampling selection was not chosen at random but upon judgement in order to select cases with the highest likelihood to answer the research question (Saunders et al., 2007, p. 230). To ensure a certain degree of representativeness, the judgemental sampling was done with assistance of the thesis’ supervisor and reference group at GAS. The sampling characteristics deviated over time, and the initial exploratory phase involved self-selection, whereas the later stages focused on critical case sampling (Saunders et al., 2007, p. 232). A compiled list of the interviewees questioned during the thesis is summarized in Table 2.1.
Primary Data: Observations

In order to fulfill the thesis aim of internal comparison between processes, as well as examination of quality data management, observations made a significant contribution. Guided by exploratory purposes, observation served the thesis in respect to both gathering of information, as well as to complement information collected through other means (Patel & Davidson, 2003, pp. 87-88). According to Saunders et al. (2007, p. 282) could collection through observation be characterised as either participant or structured. Since the objective of the data collection was to understand the processes in terms of “why?” rather than “how often?” participant observations were chosen (Saunders et al., 2007, p. 293).

Furthermore, by participating in the observed processes a deeper understanding could be developed (Ejvegård, 2009, p. 76). Likewise to the sampling of interviewees, the observed processes were closely tied to the thesis aims. The processes studied internally were of varied age, with the train-of-thought being to gain accessible benchmarking, and therefore facilitate the identification of the processes maturity from a data-driven perspective.

Secondary Data: Literature Review and Internal Documentation

To complement the thesis’ foundation of information secondary data from multiple sources has been examined. This includes both quantitative and qualitative data classified as documentary data (Saunders et al., 2007, p. 248). First and foremost, the theoretical frame of references was based upon qualitative data found within academic journals and methodology books. The academic journals are primarily collected from the databases Scopus and Google Scholar. The keywords used, either stand-alone or in combination, is presented in Table 2.2.

Table 2.1 – Compilation of interviewees and their respective titles (*WtP = Walk the Process)

<table>
<thead>
<tr>
<th>Interview Guide: System (Appendix A)</th>
<th>Interview Guide: Method (Appendix B)</th>
<th>Industry 4.0-vision (Appendix C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Engineer</td>
<td>Materials and Process Engineer</td>
<td>Data Analyst</td>
</tr>
<tr>
<td>Engineering Support</td>
<td>Manager Production (+WtP*)</td>
<td>Industrial Engineer</td>
</tr>
<tr>
<td>IT-consultant</td>
<td>Process Engineer</td>
<td>Manager Engineering</td>
</tr>
<tr>
<td>System and Support Engineer</td>
<td>Process Engineer (+WtP)</td>
<td>Manager Quality</td>
</tr>
<tr>
<td></td>
<td>Process Operator (+WtP)</td>
<td>Process Engineer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quality Director (External)</td>
</tr>
</tbody>
</table>

Table 2.2 – The literature reviews components and its respective major keywords

<table>
<thead>
<tr>
<th>Section</th>
<th>Theoretical Subject</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>Data Quality</td>
<td>“Data quality”, “Big data”, “Information quality”, “Knowledge discovery process”, “Data mining”, “Key characteristics”, “Product development”</td>
</tr>
<tr>
<td>3.3</td>
<td>Statistical Process Control</td>
<td>“Multivariate process control”</td>
</tr>
<tr>
<td>3.4</td>
<td>Multivariate Analyse Techniques</td>
<td>“Principal component analysis”, “Machine learning”, “Manufacturing”, “Neural networks”</td>
</tr>
</tbody>
</table>

The search results were sorted on highest citations and selected based on relevance to the study. Many research areas connecting to the subject of Industry 4.0 was at the time not yet explicitly
established within academia. Therefore, conference papers were featured to an extent in the frame of reference associated to Industry 4.0.

Furthermore, in addition to GKN’s internal documentation, secondary data sets containing historical process records were also analysed. In some cases, these records stretched over several years and/or contained sensitive information. Therefore, the data sets were processed by authorized personnel and thus compiled in advance. Due to the sensitive nature of these documents only a snippet is presented in Appendix E & F respectively. The ramifications of these operations are further discussed in Chapter 7.

2.5 Data Analysis

Throughout the case study, a duality of analytical methods was used. This strategy is referred to as the “usage of both qualitative and quantitative data” by Yin (2009, pp. 132-133) and is one of four general strategies proposed. The befallen choice of a dual strategy was made clear as part of the thesis’ purpose was to determine the internal data quality management from a selection of processes. As such, substantial amounts of quantitative data were examined and deemed critical in testing one of the thesis’ key propositions (Yin, 2009, p. 133). For this operation a statistical software programme called JMP 13 was used.

The qualitative data gathered arise from primary data collected through interviews. Its main contribution to the thesis was to build an understanding of the examined processes as well as to give insight in the systems used internally. In addition, there was occasions were quantification of qualitative data was deemed desirable (e.g. Figure 5.1 & Appendix D). This approach is described as limited but a useful supplement to the principal means of analysing qualitative data (Saunders et al., 2007, p. 505).

2.6 Research Credibility

As a research design is meant to represent a logical set of statements (Yin, 2009, p. 40) its credibility and findings must be discussed and its quality judged (Saunders et al., 2007, p. 149) In order to develop a reliable research design certain proactive measures should be thought-out beforehand to reduce the possibility of getting the wrong answer to one’s research (ibid.) Therefore, the need of adequate emphasized attention towards the design’s reliability and validity is urged by Saunders et al. (2007, p. 149)

2.6.1 Reliability

Proper research methodology demand that if a later investigator would to repeat the same procedures as described she should arrive at the same conclusions, i.e. reliability (Yin, 2009, p. 45). As recommended by the literature (e.g. Yin, 2009, p. 45; David & Sutton, 2016, p. 220) notorious attention to documentation of the research’s procedures were carried out in order to increase the thesis’ reliability. However, as highlighted by David and Sutton (2016, p. 220) it is often practically impossible to use the test-retest-method to verify the reliability of interviewee’s answers. As such, to ensure the proper construction of the thesis’ questionnaires, pilot studies were conducted to test their correctness and logic. Inevitable, the usage of interviews as source of evidence runs the risk of response bias or reflexivity, i.e. the interviewee gives what the interviewer wants to hear (Yin, 2009, p. 102).
2.6.2 Validity

To assess empirical social research Yin (2009, p. 40) recommend evaluation of three sets of validity: construct, internal, and external. In turn, the different sets are further explained by Kidder and Judd (as cited in Yin, 2009, p. 40):

- **Construct**: identifying correct operational measures for the concepts being studied
- **Internal**: seeking to establish a causal relationship, whereby certain conditions are believed to lead to other conditions, as distinguished from spurious relationships
- **External**: defining the domain to which a study’s findings can be generalised

As discussed by Yin (2009, p. 40), internal validity is of concern when conducting explanatory or causal studies which is not the case of this research design. Therefore, proactive measures were only regarded towards construct and external validity.

A critical criticism towards case studies is the subjective judgement used to collect data (Yin, 2009, p. 41). To circumvent this and ensure that the thesis’ research corresponded with its purpose studied, the sampling of interviewees was key to assure that the findings were focused on the area of interest. Non-probabilistic sampling coupled with judgemental sampling aided by the thesis’ supervisor and reference group at GAS consolidated the correctness and generalisation as far as possible. Moreover, tactics used to further strengthen the thesis’ construct validity included the usage of multiple sources of evidence and to have the report drafts recurrent reviewed by key informants.

Regarding external validity the research design of the thesis made sure to incorporate both internal benchmarks between different processes as well as external case examples. As elaborated by David and Sutton (2016, p. 221) external validity refers to generalisability of the research findings for a wider population as initially studied. Albeit external elements were integrated into the examined topic, the befallen organisations were restricted by accessibility and might not be entirely representative for the manufacturing sector as a whole.
3 THEORETICAL FRAME OF REFERENCE

3.1 Industry 4.0

*Today we stand on the cusp of a fourth industrial revolution; one which promises to marry the worlds of production and network connectivity in an 'Internet of Things' which makes 'Industrie 4.0' a reality. 'Smart production' becomes the norm in a world where intelligent ICT-based machines, systems and networks are capable of independently exchanging and responding to information to manage industrial production processes. – GTAI (2014)*

As can be seen in the quote above, Industry 4.0 is surrounded by somewhat mysterious and disruptive innovation. However, plenty of enterprises (e.g. Bosch, Festo, SAP, TRUMPF, WITTEINSTEIN, etc.), institutions (e.g. BMBF, acatech, DFKI, Fraunhofer-Gesellschaft, etc.) and researchers believe that the increased development of Industry 4.0 could radically transform the manufacturing sector as we know it (GTAI, 2014). Staying true to the nature of innovation, the project is surrounded by a lot of fuzziness and ambiguity. Each project initiative takes on its own name, and the more commonly ones are referred to as: Industrial Internet, Advanced Manufacturing, and Smart Factory. However, this thesis aim to look beyond political initiatives and instead study the underlying frameworks forming Industry 4.0.

3.1.1 Internet of Things

The term Internet of Things (IoT) is used to describe a state of connectivity for things otherwise perceived as ordinary analogical. Through the integration of wireless communication abilities with sensors and computing, uniquely identified things provides data without human interaction (Wang et al., 2015). As IoT abide a network of uniquely addressed objects based upon standard communication, Atzori, Iera, and Morabito (2010) sees the paradigm shift towards IoT ultimately as a result of three different visions: things (sensors), internet (middleware), and semantic (knowledge). Furthermore, Gubbi et al. (2013), in conjunction with Atzori et al. (2010), believes that the usefulness of IoT can only be truly unleashed in an application whereas the three orientations intersect.

The ingenuity of IoT is that these applications could be found in many different domains. Whereas legacy systems, which have been designed for specific purposes with limited flexibility, the initiative of IoT demands application and platforms which can capture, communicate, store, access, and share data from the physical world (Barnaghi, Wang, Henson, & Taylor, 2012). A primary result of collection and processing of data through interconnected devices are a heightened situation awareness – thus enabling machines and human users to make more intelligent decisions (Barnaghi et al., 2012). As such, the potential areas of application are near endless: smart homes, health industry, transport, logistics, and environmental monitoring (Kranenburg et al., 2011). Surely there is no doubt that IoT will find its way into the sector of manufacturing as well (Barnaghi et al., 2012; Gubbi et al., 2013; Lee, Lapira, Bagheri, & Kao, 2013).

According to Gubbi et al. (2013) the technology of Radio Frequency Identification (RFID) presented a major breakthrough in the embedded communication paradigm. Acting as electronic barcodes, RFID-tags allows for a significant improvement on item-level traceability. Smart products are able to communicate via wireless networks, thus they known their own identity, production history, specifications, and documentation (Sadeghi, Wachsmann, & Waidner, 2015).

With the coming of smart sensors the collection of data has become an uncomplicated but overwhelming exercise. According to Lee et al. (2013) it is imperative for manufacturers to integrate
advanced computing and cyber-physical systems (CPS) to be able to reap the possibilities of a big data environment. However, the arising challenge of providing the right data, for the right purpose, at the right time remains (Lee et al., 2013). With abundance of multimodal data collected through sensors and devices (e.g. temperature, vibrations, & sound), the diversity and ubiquity makes the task of processing and interpreting big data complicated (Barnahagi et al., 2012).

### 3.1.2 Big Data

As mentioned in the introduction the data created in 2016 was 16.1 zettabyte (IDC, 2017). In comparison, an earlier study by IDC (2011) stated that the volume created in 2010 was 1 zettabyte. The enormous increase in data opens up for businesses to base their decisions on data to a greater extent, ultimately making better decisions if analysed effectively (Biswas & Sen, 2017). This exponential increase in data is estimated to continue and in 2025, IDC (2017) predicts the data creation in the world to 163 zettabytes. This tremendous amount of data is what is commonly referred to as big data. However, big data refers to more than just the volume of data. Laney (2001) defined the characteristics for large data sets as: volume, velocity and variety, altogether creating the 3V’s model. Volume is the mass of the data handled, velocity referring to the speed data is handled, and variety as in the different data formats creating various incompatible types (e.g. text, video, structured data, etc.).

Even though Laney’s model was not initially created to describe big data, it has since been used and gained popularity. However, as to be suspected with such a buzzword, there exists a lot of ambiguity surrounding its definition. A decade after Laney, IDC (2011) presented their definition of big data as following:

> “Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis.”

In this definition, high emphasize is put on value, thus adding a fourth V. Moreover, recently additional V’s have been used to describe the availability and trustworthiness of data, resulting in additions such as, Veracity, Verification or Validation (Babiceanu & Seker, 2016). However, the key feature of big data is that it accounts for more than just a high volume of data, as the name can be interpreted as. A more common formulation of big data is the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time (Chen & Mao, 2014).

**Industrial big data**

All of these different definitions are aimed towards applications in environments changing from healthcare to social media. When talking about big data in a manufacturing context a distinct segregation can be made. The researchers of the above-mentioned definitions focus on human-generated or human-related data instead of machine-generated data (Lee, Kao & Yang, 2014). The machine-generate data is sometimes called industrial big data. However, the above definition with the three V’s and the additional V’s are still viable dimensions to industrial big data (Babiceanu & Seker, 2016).

Industrial big data is forecasted to be an underlying key enabler of the manufacturing domain’s adaption of CPS (Lee, Lapira, Bagheri & Kao, 2013). To make this possible, industrial big data uses advanced techniques, such as predictive analytics and machine learning, to enable timely and accurate insights leading to better decision making (Shin, Woo & Rachuri, 2014). Old data analytical
methods are still applicable when analyzing big data sets. For example, correlation, regression and statistical analysis can help draw further understanding from the dataset (Chen, Mao & Liu, 2014).

A typical knowledge discovery process from a data project can follow these steps in order: recording, cleaning, analysis, visualization and finally decision making (Chen and Zhang, 2014). Dutta and Bose (2015) suggests that an industrial big data project can follow a similar design. Namely, they suggest the projects start with a strategic groundwork in form of research, formation of cross functional teams and a project roadmap. This is followed by a data analytics part where data is collected, analysed, modelled and visualized. Finally, all comes together in the implementation phase, where insight is generated from the data and integrated with IT-systems.

3.1.3 Machine Health Monitoring
Production lines with constant optimal performance is rarely seen in real factories, instead machines are prone to degradation over time. According to Jardine, Lin and Banjevic (2006) a common strategy to combat this degradation is by time-based preventative maintenance, which maintain the machine on periodic intervals regardless of health status. Modern machines require high reliability and precision and the costs of maintaining these periodically are expensive. In a way to lower these costs, maintenance can more optimally be performed right before the machine starts losing performance. This can be done by monitoring the fundamental input readings, i.e. sensors, and adapt the maintenance accordingly, called condition based maintenance (ibid.). To further increase the accuracy of maintenance predictions the remaining useful life of machines can be estimated through statistical models (Si, Wang, Hu & Zhou, 2011).

The increasing use of connected sensors on machines, through IoT devices, is expanding the input readings and therefore strengthening the capabilities of machine health monitoring. Lee, Kao, and Yang (2014) argues that the machine health monitoring through IoT will reduce costs by minimizing machine downtime and the ability to make more reliable prognostics will support supply chain management and guarantee machine performance. Furthermore, extensive monitoring of the machines also acts as a key enabler for concepts such as cyber-physical systems and digital twins.

3.1.4 Cyber-Physical Systems
Most producing companies demand quick market introduction for their products, and if successful, an easy scalability of production. As such, the time-to-volume and time-to-market becomes considerably important aspects to gain and secure market shares (Wang et al., 2015). CPS combines deeply embedded computation and communication to interact with physical processes as to add new capabilities beyond the original physical systems (Wang, Törngren, & Onori, 2015). With the multitude of stakeholders, processes, and equipment involved in production, Wang et al. (2015) believes that CPS shows promise of integrating communication across all levels of production. The characteristics of communicated or connectedness, along with intelligence and responsiveness of CPS, is also highlighted by Monostori et al. (2016):

- Intelligence, i.e. being able to acquire information from their environment and act autonomously.
- Connectedness, i.e. the ability to connect to the other parts of the system for collaboration, including human beings.
- Responsiveness, the ability to react to unexpected internal and external changes.
Moreover, the expectations towards CPS are manifold, and interesting features in regard to manufacturing includes: robustness at every level, self-maintenance, safety, remote diagnosis, real-time control, predictability, and efficiency (Monostori et al., 2016). In addition, Wang et al. (2015) mentions the sustainability and energy efficiency of production systems as recent drivers for CPS. As the levels of autonomy increases in production, the future of manufacturing systems will depend on realistic models of the physical world, i.e. a digital twin (Rosen, Wichert, Lo, & Bettenhausen, 2015). With the increase availability of data through sensors, virtual simulations in near real-time of physical manufacturing systems are made possible (Coronado et al., 2018). Veridical simulations and analysis are therefore enabled to control the manufacturing processes with, ultimately leading to productivity increases (ibid.).

**Implementation of CPS**

According to Monostori (2014) the implementation of CPS evokes fundamental questions regarding the relations of autonomy, cooperation, optimization, and responsiveness. Moreover, the author elaborates, the integration of analytical and simulation-based approaches is projected to become more prevalent in the future. As such, challenges emerge including the operating of sensor network and big data, as well as information retrieval, representation and interpretation that has to be handled with emphasis on security aspects. The consideration for safety and security is also raised by BMBF (2013). As a critical success factor, it is essential that data and information contained within facilities and products is protected against misuse and unauthorised access. The issue of security is such a notable domain of interest that Wang et al. (2015) consider it to make or break future advancement within CPS.

Because of a societal pressure and expectations of transitioning towards CPS quickly, most solutions presented seem simple and attainable. However, according to Ribeiro (2017) that is a fundamental problem with the development of CPS as the discussion on a conceptual level is deceivingly simple. Another barrier presented by Wang et al. (2015) is the industry itself. According to the authors, its characterised by a conservative culture – a result of operating under incredibly tight margins. As such, allowing for major uncertainties at a strategic level is difficult.

However, some visionaries like Lee et al. (2015) believes that CPS forms the necessary foundation for Industry 4.0. According to these authors, CPS consists of two main functional components, namely: advanced connectivity to ensure real-time data acquisition; and intelligent data management & analytics. In order to form a CPS, Lee et al. (2015) propose the following 5-level architecture visualised in Figure 3.1.
Figure 3.1 – The 5C architecture for CPS-implementation. Adapted from Lee et al. (2015)

I. **Smart Connection**: The acquisition of correct and reliable data from machines as well as its components marks the first level of CPS-implementation. Methods for collection includes sensors as well as seamless integration of data systems.

II. **Data-to-Information Conversion**: Second level revolves around making sense of data and algorithms for prognostics and health management applications, i.e. condition based maintenance. This involve analytics of multivariate data correlation.

III. **Cyber**: At this stage, cyber and physical systems start to merge into a digital twin. Accurate analytics provide machines-optimisation as well as precise predictability upon historical information.

IV. **Cognition**: Moving up in the hierarchy hereafter means that machines start to act upon all information on its own (e.g. integration simulation).

V. **Configuration**: Finally, as the feedback-loop from the cyber-space closes, configuration makes machines resilient and autonomous. As such, machines apply corrective and preventive decisions.

3.2 **Data Quality**

"On two occasions I have been asked, ‘Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?’… I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.” – Charles Babbage (1791-1871)

The concept of big data seems to have many promising effects on data analysis, but it still struggles with the same core fragility as any data analysis does, namely: the usefulness of the analysis heavily relies on the underlying data quality (Wand & Wang, 1996). Data quality is often broken down into two categories of dimensions. The first category being the intrinsic nature of the data, meaning an objective view of the data. The other category is the context in which the data is gathered, also called the external view (Wang & Wand, 1996). The contextual view of data can typically be subjective opinions gained from questionnaires or self-reported surveys. Hazen, Boone, Ezell, and
Jones-Farmer (2014) classify this kind of data more under the research term information quality, while the more objective dimensions fall under the term data quality. It is worth mentioning that data and information are used interchangeably in the literature, but in this study we will use Hazen et al. (2014) view of data meaning the more raw and objective form. This is primarily due to this study’s focus on data management on processes in production which does not involve subjective data to much extent.

The intrinsic dimensions of data are being categorized differently in the literature (Lee, Strong, Kahn & Wang, 2002). However, measurable dimensions such as accuracy, timeliness, consistency and completeness is consistently mentioned as important factors for the raw data quality (Wand & Wang, 1996; Lee et al., 2002; Hazen et al., 2014). The dimensions are further explained by Hazen et al. (2014) as:

- **Accuracy** answers to the question whether the data reflect the real-world object, is the data correct?
- **Timeliness** refers to the time between the recording of data and analysis, is the data up-to-date?
- **Consistency** emphasize the importance of consistent formatting of data between and within systems.
- **Completeness** refers to the inclusion of necessary data, is any important data missing?

Ensuring the quality of these characteristics creates a solid base of which analysis and decisions can be made of. Even the more advanced methods, such as machine learning, requires a solid data quality foundation to be able to produce a relevant output (Wuest, Wiemer, Irgens & Thoben, 2016).

3.2.1 Making Sense of Data

A dataset of high quality is of little value to companies and organisations in its raw form. It is the process of extracting useful insights from the dataset that is creating high value (Wang & Wand, 1996). This process is referred to as knowledge discovery in scientific literature. In relation to this topic a common encounter is data mining. These two concepts can sometimes be used interchangeably and it is therefore important to understand them correctly. Fayyad, Piatetsky-Shapiro and Smyth (1996) defines data mining as one step of the knowledge discovery process. The knowledge discovery process is the whole process from data creation to useful knowledge, while data mining is the step where data is transformed into patterns. Simplified, data mining is the analysis phase where analytical tools are applied.

There are several models describing the knowledge discovery process which vary in detail, an overview of frequently used models in both academia and industry are presented in a survey by Kurgan and Musilek (2006). The most referenced model in academia is made by Fayyad et al. 1996) and contains an in-depth description of nine steps. These steps are categorized by the authors as selection, pre-processing, transformation, data mining and interpretation. The model is presented with the result of each activity in Figure 3.2.
Figure 3.2 – The knowledge discovery process. Adapted from Fayyad et al. (1996)

Fayyad et al. (1996) highlights that the value creating activity from their model is considered to be the data mining-phase. However, the selection, pre-processing and transformation are crucial ground work for the data mining to create that value. Much like infrastructure are important in ordinary mines, these supporting activities play a key role. The importance of the generated knowledge cannot be stressed enough, as Harding, Shahbaz, Stinivas and Kusiak (2005) states:

“Knowledge is the most valuable asset of a manufacturing enterprise, as it enables a business to differentiate itself from competitors and to compete efficiently and effectively to the best of its ability.”

Achieving knowledge through this process is therefore desirable for companies. However, to even start the process, we must start in the other end of the spectrum, the raw data pool. What we put in here is determining the potential for any future analysis.

3.2.2 What to Measure?

To be able to proceed with the knowledge discovery process, to ultimately extract useful knowledge, data first have to be collected. This is where the difficult question of which data to be collected arise. A common practice in the industry is to measure on key characteristics on both the product and the process, which practice-oriented literature also suggests (Thornton, Donnelly, & Ertan, 2000). Key characteristics has even an own standard in the aerospace industry, which is the standard AS9100D. The standard is issued by the International Aerospace Quality Group (IAQC) and they define key characteristics as:

“An attribute or feature whose variation has a significant effect on product life, form, function, performance, service life, or producibility; that requires specific actions for the purpose of controlling variation.”

Moreover, a subset of the industry standard (AS9103) exists which sole focus is the handling of variation management of key characteristics. This includes a methodology for KC-identification (such as SPC, Design of Experiments, and failure mode effect analysis [FMEA]). This method is strengthened by Thornton et al. (2000) which advocates the use of key characteristics in measurements of both the product and process life-cycles.

Another methodology first introduced in the automotive industry is the Production Part Approval Process (PPAP). PPAP is an industry-oriented methodology to ensure high quality of customer specified characteristics and is part of high-tier supplier programmes in the aerospace industry (GAS internal documents, 2017). It consists of eleven steps and these activities are there to ensure that the design works in practice and that the company can produce these products reliably. Among
these steps there are both the design failure mode effect analysis (D-FMEA) and process failure mode effect analysis (P-FMEA). These activities are there to early identify the key characteristics in order to control them during the production and product life-span.

**A Multivariate Mind-set**

As can be seen from the description of what to measure from the viewpoint of the aerospace industry standards the requirement is to measure key characteristics in the product and the process. This mind-set is good in a univariate setting where the most significant variables are of interest. However, in a multivariate setting this mind-set changes. Eriksson (1999, p. 24) states that in order to unleash the full potential of multivariate tools, all information can be useful. The more information you feed these tools the more powerful they become. As such, in order to ascend to data-driven manufacturing measuring everything possible will be of importance.

### 3.3 Statistical Process Control

*A powerful collection of problem-solving tools useful in achieving process stability and improving capability through the reduction of variability.* – Montgomery (2013, p. 188)

A prominent technique for quality control is statistical process control (SPC) founded upon the Shewhart control chart from 1920. Montgomery (2013, p. 190) explains the basics of the control chart as a set-up with a centre line, upper control limit (UCL) and lower control limit (LCL). The sample points of the process is later mapped on the control chart and as long as the points fall in between the control limits no actions are required. However, if a point falls outside of the control limits, further investigation is required to understand the anomaly. Figure 3.3 illustrates the concept of monitoring a process.

![Figure 3.3](image)

*Figure 3.3 – A standard univariate control chart for monitoring a process. White dot indicates an alarm to be investigated*

The upper- and lower control limits are usually assigned 3 standard deviations apart from the centre line (Montgomery, 2013, p. 28). This creates a probability of recognizing 27 errors of 10 000 attempts on a stable process. However, Montgomery (2013, p. 29) points out that no process is truly stable and if the mean shift ±1.5 sigma from target the error estimation is 668 errors in 10 000 attempts.
The implementation of control charts on processes are usually conducted in two phases. The objective of Phase I is according to Montgomery (2013, p. 206) to retrospectively analyse data and construct trial control limits. This, in order to determine whether the process have been in control under the period and if the control limits can be used to analyse future production. In the Phase I scenario there are often large shifts to be detected and therefore the Shewhart control chart is most effective (Montgomery, 2013, p. 207). Before going to Phase II, assignable causes for large process shifts are meant to be solved.

In Phase II, the current data output is compared to the control chart resulting from Phase I in order to monitor the process. The focus in Phase II is detecting small process shifts and the Shewhart control chart is not optimal in those cases (Montgomery, 2013, p. 207). Alternative control charts such as the Cumulative Sum (CUSUM) and the Exponentially Weighted Moving Average (EWMA) are good candidates in these scenarios due to their ability to incorporate previous data points.

3.3.1 Applications in Low Volume Production
When using statistical tools, the sheer quantity is an important parameter when trusting the output of the analysis. In situations where the batches are small or the production series short this becomes a problem. Montgomery (2013, p. 451) suggests the use of the deviation from normal (DNOM) control chart to combat this. In this type of control chart only the deviation is monitored and it is therefore possible to measure the output of different parts from the same machine in the same control chart. The different parts can together make a bigger statistical population than the individual subgroup parts, making for better monitoring of the machine’s performance. CUSUM and EWMA are also good applications in low volume scenarios. Their ability to detect small shifts even in small batches with individual measurements are very suitable for this environment (Montgomery, 2013, p. 453).

3.3.2 Multivariate Statistical Process Control
Traditional univariate control charts as those mentioned above defines product quality by monitoring of separately key characteristics. MacGregor and Kourti (1995) problematizes this approach as quality variables (key characteristics) seldom are independent of one another. Moreover, the authors states that it is a rare occasion that these variables adequately define product quality alone. Hence, they state that product quality could only be properly defined through the simultaneous measurement of several key characteristics, i.e. a multivariate property. Figure 3.4 illustrates a scenario whereas $Y_1$ and $Y_2$ shows no sign of being out-of-control when plotted separately in univariate Shewhart-diagrams. However, the true situation of the multivariate correlation is revealed by the elliptical joint region of $Y_1$ and $Y_2$. 
However, the case illustrated in Figure 3.4 with a control chart for two dependent variables are impractical for two reasons (Montgomery, 2013, p. 516). First, the sequence of time of the plotted points is lost as oppose to the scenario of a traditional univariate chart. Moreover, construction of an ellipse for more than two quality characteristics becomes geometrically difficult. Possible solutions include more complex control charts such as the chi-square control chart, the Hotelling $T^2$, Multivariate-EWMA and Multivariate-CUSUM.

According to Montgomery (2013, p. 512) the usage of multivariate methods has increased greatly in recent years as a response to the popularity of automatic inspection procedures. Today, it is not uncommon for industries to maintain manufacturing databases with process and quality data on hundreds of variables and millions of individual records. Thus, monitoring and analysis of these data sets with univariate SPC becomes tedious and ineffective.

The most common procedure for multivariate process control is the Hotelling $T^2$ control chart (Montgomery, 2013, p. 514). It is a direct analogue of the univariate Shewhart-chart for monitoring the process’ mean vector. Moreover, this control chart is directionally invariant, i.e. its ability to detect a shift in the mean vector only depends on the magnitude, and not the direction, of the shift (Montgomery, 2013, p. 517). In addition, both the chi-square and Hotelling $T^2$ methods relies on information only from the current sample. As a consequence, they are insensitive to moderate shifts in the mean vector. However, alike the univariate case, it is possible to extend both the CUSUM and EWMA control charts to a multivariate state in order to detect smaller shifts (Montgomery, 2013, p. 524).

To describe the mathematical fundamentals behind each of these in detail would be an immense task and is therefore beyond the theoretical scope of this thesis. Instead the following works are recommended on the subject for proper explanation: Hotelling (1947); Crosier (1988); Lowry, Woodall, and Champ (1992); MacGregor and Kourti (1995); Montgomery (2013). Furthermore, an extensive overview of multivariate quality control is presented by Bersimis, Psarakis, and Panaretos (2007).
3.4 Multivariate Analyse Techniques

To understand the world around us, as well as ourselves, we need to measure many things, many variables, many properties of the systems and processes we investigate. [...] Multivariate data, well measured on intelligently selected variables contain much more information than univariate data, and hence an adequate multivariate characterization of samples, systems, and processes of interest, is a necessary first step in their investigation. This is so both for intellectual and practical purposes, i.e., both in basic research and applied technology. – Eriksson (1999, p. 1)

Latent Structure Methods

Large and complex data sets increases the number of process variables, as such the average run-length performance to detect a specified shift increases as well (Montgomery, 2013, p. 533). Thus making it difficult to detect shifts with the control charts describe in the previous section. Moreover, if the number of measured quality variables is large, it is likely that the correlation among them is strong (MacGregor & Kourti, 1995). This causes the covariance matrix to be nearly singular as a result. Principal Component Analysis (PCA) is therefore a useful procedure to reduce the dimensionality of the quality variable space by linear combinations of eigenvalues explaining the variability of the data set (Montgomery, 2013, pp. 533-534).

The brilliance of PCA is that in practice, one rarely needs to calculate all eigenvalues since most of the variability of the data set captured in the first few principal components are sufficient to explain variations in the product (MacGregor & Kourti, 1995). Jeffers (1967) summarizes the practical objectives of PCA, whereas the basics includes the following:

1. The examination of the correlations between the variables of a selected set.
2. The reduction of the basic dimensions of the variability in the measured set to the smallest number of meaningful dimensions.
3. The elimination of variables which contribute relatively little extra information.

To exemplify this Jeffers (1967) describes a practical case study wherein PCA is used to consolidate the analysis. The test regarded a product testing in form of pit-props, i.e. the supporting timber used in mines. As such, it was of great importance to determine which population of species that had a sufficient compressive strength to pass the load. In the study, timber was collected from over the country and a set of 13 variables were tested. Through the usage of the methodology described in the list above, they successfully reduced the dimensionality to include only five variables, accounting for a total of 64 per cent of the variability in compressive strength. Furthermore, Jeffers (1967) argue that there are many similar examples in the field of product testing where the advantages of PCA gives clear guidance as to the selection of variables necessary for further studies.

The methods for discovering sub-dimensional process shifts are sometimes denoted latent structure methods, and another approach besides PCA is Partial Least Square (PLS). By classifying variables into x’s (inputs) and y’s (output) and applying regression adjustment, a set of weighted averages can predict the y’s or linear combinations of the y’s (Montgomery, 2013, p. 538). In a way, this procedure maximizes the covariance much alike how principal components directions maximize variance. In practise, PLS shows promising application in the field of chemometrics, due to the immense number of variables (Montgomery, 2013, p. 538). Wold, Sjöström and Eriksson (2001) also points out the inherent ability of PLS to effectively handle many variables with even incomplete or noisy data, which can be the case in complicated problems. The authors conclude
that PLS is a simple but powerful tool to analyse complex problems and that multivariate methods are effective when understanding the behaviour of complicated systems.

**Machine Learning**

One promising solution to analysis in a complex manufacturing environment is machine learning (ML). This method has already been shown successful in several applications, e.g. process optimization (Harding et al., 2005). The term learning is defined by Simon (1983) as:

>“Learning denotes changes in the system that is adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more effectively the next time”

As such, ML aim to learn the machine better approaches each attempt, ultimately making the machine an expert. The learning process consists of a learner and teacher. The learning process is generally divided into three different categorizes and Monostori (2003) describes them as follows:

- **Supervised learning:** the correct response is provided by a teacher,
- **Reinforcement learning:** less feedback is given, since not the proper action, but only an evaluation of the chosen action is given by the teacher,
- **Unsupervised learning:** no evaluation of the action is provided, since there is no teacher.

The two latter types have not been researched to the same extent as supervised learning, where so-called neural networks are trained in a specific task. LeCun, Bengio & Hinton (2015) describes neural networks as three different types of layers: input layer, hidden layers and an output layer. According to Wuest et al. (2016) the strength of ML is to find highly complex non-linear patterns in diverse types of data, which are used for prediction, detection, classification, regression, or forecasting. Thus, ML offers great applicability to problems within manufacturing (ibid.).

### 3.5 Frame of Reference Conceptualised

A part of this thesis aim was the clarification of the data-driven perspective of Industry 4.0 within manufacturing. In order to efficiently use a data-driven approach towards decision-making demands information available to be of impeccable condition (Wand & Wang, 1996). Hence, the thesis’ theoretical framework solidifies its foundation with a basis of data quality. A qualification for data is proposed by Hazen et al. (2014) and the origin of adequate data is presented in the knowledge discovery process (Fayyad et al., 1996).

Statistical process control has long been a cornerstone in quality management and its usefulness in the aiding of achieving process stability and capability is well documented (Montgomery, 2013, p. 188). Traditionally univariate, i.e. the determination of product quality separately by its characteristics, the trending’s of data collection allows for enhanced analysis in a multivariate setting (Montgomery, 2013, p. 513). As quality variables tends to be dependent on other characteristics (MacGregor & Kourti, 1995), this approach allows for an increased process knowledge as latent structures are exposed.

However, in order to fully elevate into these promising domains of multivariate process knowledge, one must accept the future trends of digitalisation (Lee et al., 2013). With abundance of (big) data gathered through embedded sensors, i.e. the internet of things, the challenge of interpreting it increases (Barnahagi et al., 2012). The assurance of an adequate degree of data quality is therefore still very vital. The concept of Industry 4.0 and data-driven manufacturing is in essence an expression of the merge between the cyber- and physical world. To achieve the hopes of real-time
control, predictability, and self-optimizations (Monostori et al., 2016) it is necessary for the future of manufacturing systems to achieve veridical simulations, i.e. a digital twin (Rosen et al., 2015). According to Lee et al. (2015) CPS is what is effectively forming Industry 4.0, and its two main dependencies are advanced connectivity and intelligent data management & analytics. As such, this thesis’ presents a framework (see Figure 3.5) solidified upon the traditional data management tools, i.e. data quality, SPC, and data analysis. It is necessary to master these techniques in order to make use of the enhancements available through a CPS-implementation. The implementation-plan presented in this thesis is inspired by the works of Lee et al. (2015). However, it has replaced the levels with stages representing the tools and services used. In addition, instead of a pyramid, the thesis proposes an upside-down cone representing the increased amplitude in process knowledge gained with these advancements. Ultimately, as technology advances it is sure to bring additions of even more intelligence. As such, the framework tops of with a blank stage to illustrate the ambiguity brought by disruptive concepts, i.e. CPS or Industry 4.0.

![Diagram](image)

**Figure 3.5** – The thesis' theoretical frame of reference conceptualised
4 **EMPIRICAL FINDINGS**

4.1 **A Paradigm Shift in Quality Control**

Historically GAS has always had a keen eye for documentation and pressed on the importance of traceability. Development of the old in-house quality system, KPS, started almost 30 years ago. In combination with the fact that personnel have shown a great longevity at the site ensures that the culture of keeping detailed records is well established. The modus operandi during the KPS-era involved a lot of paper work as outcome registration were stored in binders following the products. Figure 4.1 illustrates the previously perspective on quality at GAS.

![Previous Quality Perspective](image)

*Figure 4.1 – An illustration over the historical approach to quality control at GAS*

A clear drawback of KPS was the notification-delay when processes operated out-of-control. Due to the nature of the system, quality reports were not conducted until the very end, as the detail’s characteristics were compiled and uploaded to KPS. As such, the time between discovery of an error and its origination could potentially be severe for operations positioned early in the process. The traditional way of working meant that deviations were seen as normal and unavoidable. Hence, periods of intermittent production occurred, where problems aggregates and delays were not uncommon. Therefore, the old perspective of quality control revolved a lot around the handling of deviation and its prevention through lessons learned. An obvious sub-optimisation and GAS found themselves in need of a more proactive approach in regard to quality control.

4.1.1 **The Road 2 Zero Initiative**

Change seldom comes without reason and for GAS, the paradigm shift in quality control comes as a response to an increased expectation on its capabilities from their customers. With fierce competition and lower time-to-market, accepting deviations and late deliveries is a thing of the past.

The internal Operational Management System (OMS) describes the elements of the PPAP (AS9145) and the standards for various quality tools (e.g. MSA, PMFEA, 8D). The current focus is to establish a setting of data-driven manufacturing with closed loop SPC securing the processes:
"To ensure that our manufacturing processes are stable and capable, we continuously follow the process outcomes and react before the process becomes unstable and, at worst, generates non-conformance. The use of SPC is the key to drive and visualize robustness. From 2018 and forwards, the SPC data will increasingly turn into ‘Real Time SPC’ to enable direct adjust and secure robust processes.” (GAS internal documentation, 2018)

Internally this promotion is called the Road 2 Zero (R2Z) with the ultimate goal of producing zero defects. Much of its core elements is based upon robustness and have already been integrated into the current perspective of quality control. This perspective is presented in Figure 4.2.

![Current Quality Perspective](image)

**Figure 4.2 – The current approach to quality control at GAS**

While developing their new proactive approach GAS also acquisitioned a new database for quality control. At the time of this thesis, the final transitions from KPS to QSYS were underway. A program developed by an external company for monitoring a product’s lifecycle. Being more-or-less up to date with the technological advancement, QSYS allows for greater flexibility and scalability than KPS. As such, GAS believe that the system is able to tackle upcoming challenges related to the increased complexity and volumes of future data collection.

### 4.2 Quality Data Management

The transition to QSYS meant a shift in the general management of quality data at GAS. Now all data collection (except process data) is stored in QSYS by default. An important remark is that QSYS acts as a database for quality data rather than a holistic solution for quality management. This is illustrated in Figure 4.3. Albeit QSYS itself offers additional licenses for tools such as SPC-modules and analysis tools, GAS choose to complement it with its own analytical tool.
Figure 4.3 – An illustration over the general quality data management at GAS

Figure 4.3 also illustrates the exclusion of process or machine data in the storage and analysis phase. Some of the newer machines already record most of the machines critical operating variables. This data is not stored in the QSYS-database, but can in some cases be stored locally within the machines internal programmes. This data is not stored in with part identification, therefore making it difficult to correlate the process data with the product data. In many machines data is not stored at all, but can be observed during the operating state. The reasoning behind this is expanded upon by one interviewee:

“QSYS is created to be able to handle machine data as well. But it requires a sensor that reports this kind of data. It is not technically difficult to incorporate the machine data into the system, the reason is that nobody is requesting this type of data. Therefore we do not store it.”

Furthermore, the addition of the analysis tool to QSYS is developed internally and could been seen as a fork containing the “best-of” from KPS. This tool reads and compose visual aids from QSYS such as graphs of statistical process control and distributions. However, the configurations in terms of tolerance and specification limits are managed through a products inspection plan within QSYS. An example of the analytical tool is presented in Figure 4.4.

Figure 4.4 – A compilation of data visualised in the analysis tool
In addition to the generated diagrams, a table summary of the chosen characteristic’s tolerances, nominal, and process capability ($C_p$ & $C_{pk}$) values are also calculated and displayed in the same report. Thus, the overall output is quite extensive and allows for a detailed review of the characteristic. A remark is that the import system is standardized and does not differentiate depending whether a characteristic is numeric or categorical.

As mentioned earlier, the collected data is used to keep track of day-to-day operations and a warning system is in place to track trends. By default, this system alerts affected process technician if a key characteristic uses more than 80 percent of its total tolerance span. The frequency of system alerts varies depending on the criticality of the specified characteristic. Upon alarm, a team of engineers and operators are led by the process technician to examine the cause and search for a suitable solution.

4.3 Planning Ahead: Industry 4.0

All generated data does not reside in one system. Instead there are several systems that stores data that are relevant to understand all active variables in the manufacturing phase. The currently gathered data is currently stored in four categories. The first system is the Enterprise Resource Planning (ERP) where GAS use software from SAP. This system tracks business commitment, payrolls, deliveries et cetera. Beneath the ERP system is the Project Life Management (PLM) system, which is focused on optimising production lifespan for the product. Here GAS uses the Siemens system Teamcenter. In addition to this there are several Manufacturing Enterprise Systems (MES). It is here where the previously mentioned system QSYS fits in. Lastly, there are several automation systems that can vary depending on the machine. In this category all the PLC programming is done for the machines and sensor data stored.

All systems are today basically standalone, meaning that they do not communicate with each other. This is starting to shift. In a recent project the ERP system and the MES system QSYS have started communicating some data with each other. For example, if a value outside the control limits is obtained in QSYS, an automatic report is created in the EPS-system. Before the operator had to manually create this report and fill in all details manually. The integration between the systems both saves time and makes the life easier for the operators. The foundation for a smarter production is talkative systems. This is a statement reinforced by one interviewee, who describes GAS transition to Industry 4.0 as:

"Industry 4.0 is somewhere where all software programs communicate with each other in a smart and sensible way."

To illustrate the above quote, see Figure 4.5.

Figure 4.5 – Industry 4.0 visualised on a system level
Before the work with integrating the system can begin, GAS must decide which system to work with. Due to the GKN’s 14 global aerospace factories the systems at each site and within each site can vary drastically. Each have optimized their systems for the task at hand, not really thinking holistically for all sites. An ongoing project is working on standardizing the digital platforms, which means that GAS might have to adapt to different systems than what they are currently using. This global initiative towards a standardised system platform between all sites have only just begun and it will take many years to implement. Therefore, the integration between systems are projected further in to the future.

4.4 Internal Case Studies of Quality Data Management
To resolve the thesis’ aim of internal process examination in terms of a data-driven quality management applicability, a sample of three has been studied and described.

4.4.1 Process A
The purpose of Process A is to make a specific detail’s critical area heat-resistant and to protect it from wear and tear. This is achieved by coating the sensitive area by a surface of nickel through electroless plating. A procedure which requires the underlying material to be free from oxides. However, due to restraints stemming from the detail’s design and material, the operation is seen as especially intricate. For a schematic overview of Process A, see Figure 4.6.

![Figure 4.6 – An illustrative overview of Process A](image)

As can be seen in Figure 4.6 the product goes through three quality controls. If a defect nickel layer is detected in either of these controls the product will start over from the blasting operation. The root cause to several defects are from previous analysis traced back to the blasting operation. The blasting operation is difficult because of two reasons: it is difficult to measure the outcome of the process (the oxide thickness level) and the blasting robot has limitations in reachability.

Data Collection and Storage
Due to the increase in defects an improvement project has been launched in order to analyse the affecting factors and find the root cause. In collaboration with the project team current data output has been gathered and through the interviews an identification of data storage has been done, see Figure 4.7.
In total six different data storage system exist for this process. None of these systems communicate with each other, meaning that the information in one system is only obtainable there or maybe manually copied to another system. The blasting operation uses both the quality system QSYS for reporting some details over media usage and nozzle equipment. In addition, there are also an analogue log book where the operator keeps track over robot settings and tools. The chemical baths are processed through an entirely different database called TrueChem. This system handles all chemical data and provides the laboratory personnel with tools to analyse the process. In combination to this there is an analogue log book, where the operator notes specific values from the baths. The furnace uses its own monitoring system where it keeps track of time and temperature. Finally, the process is monitored by the company’s business system from SAP. This system keeps track of supply chain details, such as material flow and inventory level. The only data that is transmitted from the quality controls is a pass or fail.

**Data Analysis**

Continuous analysis of data are done on the chemical baths. The TrueChem system provides laboratory employees with individual control charts to determine whether or not measured levels are approved for production. The furnace monitoring system also utilizes control charts to make sure the temperature stays within limits, these are controlled afterwards to make sure the process was stable.

The analysis of the chemical baths from the laboratory is communicated to the production floor via a system giving the current bath a green or red light on a dashboard. This gives the operator information if the bath is usable or not. The simple yet effective visualisation tool is embraced by both operators and chemical analysts.

**Multivariate Analysis of Process A’s Data**

As part of this thesis, application of multivariate tools was tested on data from Process A. The data used for analysis was gathered from all data storage system except the furnace. This data collection
was done in combination with an ongoing improvement project for Process A. The data was combined into a single excel file by the improvement project team. The performance measurement of the data was whether or not the product had passed the quality test. The other variables were available data inputs such as, nozzle wear, type of blasting media and blasting angle. Most of the dataset contained categorical datatypes and four of them was numerical. In total, information containing 13 variables from 130 products was collected. An excerpt from the data file can be seen in Appendix E. The analysis is presented in greater detail in Section 5.3.

4.4.2 Process B

In comparison to Process A, Process B differs vastly in its characteristics. Process B’s goal is to assemble a part that belongs to a jet engine. The assembly is done through welding and several minor details is processed on the part in different operations. The customer requires tight tolerances on the products geometry, which means that the product’s geometry is measured extensively. In this specific assembly line the product goes through 24 operations. These operations include activities, such as welding, cleaning and quality control measurements through machine probing.

Even though Process B has been reasonably stable the recent years, some quality issues have occurred. These have only required some rework on the product and has not been a problem due to stable production volumes. However, the production rate has recently been elevated to almost twofold the original production volume with plans to make it threefold in the coming year. To be able to handle these volumes the amount of reworks must decrease. In addition, the contract contains a hefty fine if the parts are not delivered on time making this process gain high priority as of late.

Data Collection and Storage

During the development of the process key characteristics for the product have been identified. These key characteristics consist of requirements from the customer and internal requirements that make sure the product is geometrically accurate. These are controlled by machine probing measuring machines. The collected measurements are all stored in the QSYS system. QSYS current database structure stores the measured data point with associated data to secure the traceability over the process. A typical raw data file over one measurement point can therefore contain over 20 columns, with information about different tolerance levels, inspection plans et cetera.

If further analysis’ are to be made, the data has to be extracted from QSYS, a process perceived as tedious by two interviewees. Due to QSYS being built like a database it does not present explanations to information found there. For any comprehensibility of the information in the database, both the blueprint and each requirements corresponding inspection plan have to be read in conjunction. These documents are found in other systems, also expressed difficult to navigate by the same interviewees.

The company’s ERP system is used to control the supply chain for the process. It also functions as the deviation system for the process. The operator reports deviations with different degrees of importance and also classifies the problem type. A problem with this classification system is that over the majority of the deviations are classified as “other”. According to one interviewee the root cause is that the operator does not have time to thoroughly fulfil these forms and it might be that the predefined problem areas are ill-suited.
What is important to notice is that no process parameters (e.g. robot settings and parameters) are obtained in the process, only product parameters are gathered. The machine park is of modern standard and therefore stores data in their own system. This data is not associated with the products serial number. Correlation with the process’ parameters and the resulting product is therefore not possible. The process data is in comparison to the measurement data in QSYS only accessible to personnel with the machine knowledge and permission.

Data Analysis
The measurements stored in QSYS is automatically processed by GAS own AT. It outputs individual control charts for each measured point and presented the normal distribution over the data. Due to the process many measurements nobody is monitoring all of them, instead they use the warning system described in Section 4.2, which simply notify the right people in case of deviations. Any further analysis is not done regularly, only in case of root cause projects.

Multivariate Analysis of Process B’s Data
Due to the recent ramp-up in production rate for Process B, a project group is looking at potential improvements for the process. They have gathered a dataset of chosen measurements of the product and summarized it in a spreadsheet. In total there is 271 measurement points that are ranging over 333 products, forming a set of 90,243 data points. Due to the relatively big volume of data, an attempt of applying multivariate tools to extract useful knowledge was conducted during this thesis. The origin of the dataset is described here, while further analysis is presented in Section 5.3.

The measurement points were collected from five of the 24 operations in the process. These five operations consist of machine probing. A schematic overview over the process and where the measurements take place are presented in Figure 4.8.

![Figure 4.8](image)

Figure 4.8 – An overview of Process B’s operations. Measurement points (MP) shows how many data points that is collected in each Measurement operation (M1-M5)

From Figure 4.8, the number of inputs are displayed in relation to the operation. Most of the data is gathered in the final quality control, measurement operation 5 (M5), where 170 individual measurements are done on every part. The measurement operations M1 and M2 are in consecutive
order because the part switches fixture and machine between the measurements. The same logic applies to M3 and M4. The dataset these measurements represent were cleaned and prepared for analysis. The cleaning process involved removing incomplete measurement points. An excerpt of the spreadsheet is presented in Appendix F.

4.4.3 Process C

Process C involves two main operations which are blasting and thermal spraying. What differs greatly in comparison to the other two processes is its data management. In this method they have developed their own process monitoring program. This program is called PKR and was custom built for the process in the 1990's. It is used to monitor process key characteristics to verify the process state before and under the process. With this type of data monitoring the process performance is controlled beforehand, which helps to assure the products outcome. A schematic picture of the process and product data from Process C is illustrated in Figure 4.9.

![Figure 4.9 – A schematic overview of Process C](image)

The operator collects process data through either reading of the state of the machine or measuring the process state. This data is manually inputted to the PKR system, which then analyse the data and visualize the result through an individual control chart. Depending on the process outcome the process gets a no or go decision. If the process gets a go, the process can begin and the process parameters are supervised by the operator during the process. The supervised parameters are predetermined in the design phase of the process and marked as key characteristics in the programme. This way, it is easy to follow the critical parameters of the operation. The same procedure is executed before both the blasting and thermal coating operation.

The outcome of the operation is measured through test pieces that go along the real products in the operations. These test pieces are taken apart and lab tested after each operation and the result is reported through the quality system QSYS. They are kept track of in a similar way as Process A and B, which means that QSYS primarily works like a database with some added analyse tools in the form of individual control charts. These lab controls are performed in sample test with the sample frequency decided by the prior variance in the quality test. Due to Process C's collection of both product and process data correlation analysis are commonly done, for example in root-cause
analysis projects. They have not delved deeper into more advanced multivariate tools, mainly because they see that more data is needed. Indifferent to Process A and B, Process C will not undergo any multivariate attempts in this thesis as no presentable dataset was accessible at the time being.

Today the thermal coating operation relies on the operator’s manual input. There is already an ongoing project into learning more about the process with the future goal being an automatic feedback loop, which will reduce the operator dependency. As the interviewee from Process C stated:

“We need to learn more about sensors and automation to be able to make a connection directly into the process and into equipment that automatically control it. We are actively working with this, but we cannot do everything at the same time.”

The increased surveillance with sensors go hand in hand with another ongoing project that aim to foresee the process outcome through simulation. The results from both these projects will not be finished before this thesis end, however, the projects highlight current focus areas.

4.5 Case Example I: Electrolux Home Care & SDA

To broaden the thesis’ perspective on the subject of Industry 4.0 and its application, an interview was held with Electrolux, a company which operates in a completely different sector than GAS. Electrolux provides a wide range of commercial products and services. As a global company they have dispersed its factories and areas of interest throughout the years. Today, most of its manufacturing is conducted in Asia, while software development and strategic operations maintain in the West.

The interviewee held the position of Quality Director within Advanced Analytics, Intelligence, and Reporting (AIR). A hive-off soon to celebrate two years; which differentiates from their ordinary quality departments as its sole purpose is to strengthen the market shares through extensive analytics. AIR is founded upon two pillars: the Internet of Things and Data Science. The first year of operations concluded little of practical value as recruiting of competence as well as definition and creation of infrastructure took quite some time. Now however, the team consist of 15 people and Electrolux is starting to reap the rewards of AIR’s work. Indifferent from GAS, the current focus for Electrolux’s “Industry 4.0”-efforts were aimed directly towards understanding their end users’ commercial behaviour.

Initially the project focused on semantic analysis, both from a learning perspective as well as a predicative approach. By crawling retailer webpages and understanding customer reviews through text mining, AIR could analyse immense data sets and correlate product features. As such, both negative and positive feedback were incorporated into coming updates and models. Moreover, Electrolux were now able to base their trend prediction on a larger population and categorise customer profiles. Of course, these methods are not uniquely tied to the digital era. Instead one should view it as a set of new tools for market research enabled by a shift in their customer’s behaviour of using digital tools as well.

With the success of the semantic analysis, AIR is now transitioning to incorporate these techniques to improve the post-purchase phase for their customers. Many of their products includes electrical components able to transmit anonymous run data back to Electrolux. With these capabilities AIR can identify sub-optimal behaviour and/or offer personalised customer-offerings per product and
apply data-driven software updates post-purchase. Moreover, as the amount of data grows, Electrolux is able to conduct more precise life-cycle analysis and trace key characteristics with rigor.

To start off the project an external company were consulted who could show a proof-of-concept and establish “game rules”, i.e. the standards for data collection and storage. Furthermore, a lesson-learned for Electrolux was the importance of data quality. This concluded an extensive measurement system analysis in order to ensure the trustworthiness of the input data. The incremental expansion and success of pilot-projects leads Electrolux to now put faith in the usage of a data-driven approach at grand scale. With continuous fine-tuning the company hopes to implement multivariate analysis onto their manufacturing sector as well. According to the interviewee potential areas were these techniques shows promise includes Design of Experiments, decision models for machine learning, and validation of processes.

4.6 Case Example II: GKN Aerospace Norway AS

During the thesis an opportunity to compare GAS to one of its sister sites (GAN) within the concern presented itself. As such, the manufacturing situation and conditions were similar. The projects discussed were aimed at solving problems on one of their highest prioritized product.

**Multivariate Analysis Project**

Due to the complex process of the product a project was launched for multivariate analysis of the process in order to gain increasing knowledge over the process. The project plan was spanning over two years, with the initial 6 months dedicated to data collection and pre-processing of the data. The remaining 18 months was planned for analysis of the data and implementation of identified changes to optimize the process. The actual project ended in data collection and pre-processing data stretching over the entire project time. Some multivariate analysis, such as PCA and lead-time prediction, was also possible during the end of the projects lifecycle. The results produced low levels of degree of explanations which made it difficult to make process changing decisions on it.

Although no game-changing information was obtained, the project resulted in many important findings and landed in other improvements. The biggest realisation from the project was the underwhelming amount of data and its questionable quality. The collection methods were often done manually on paper and later stored in physical achieves. Therefore, part of the project instead became to digitise the current collection and storage solutions. The main problem for multivariate analysis was the low amount of data collected and the only viable solution to collect the data needed is to first build a data infrastructure that can handle it.

**Green Monitor Project**

The sister site is part in a collaboration project between companies and research institutes called Green Monitor. The projects initial goal was to establish methodologies on measuring machine energy consumption in order to lower the power consumption, hence the name. The project software has developed more sophisticated functions which now makes it possible to monitor several machine health parameters. In this system, machine data is stored with their corresponding parts serial number. This makes it possible to correlate it with product data and realistically make useful multivariate analysis. However, the project is in an early stage and is not incorporated to this extent yet.
**Geometrical Part-fit Optimisation**

An interesting project currently undergoing development at GAN revolves around part fit optimisation. One of their main products is constructed through an assembly of 13 parts made in six different configurations. Due to customer requirements and the geometrical constraints of the main part that they are being assembled onto, some parts have specified positions. Others however move freely and can be positioned according to fit depending on the distribution of their outcome. In this case, the subparts of one production unit can be assembled in 24 unique combinations as Figure 4.10 illustrates.

During one production week GAN fabricates enough subparts to reach an astonishing possibility of $2.4 \times 10^{47}$ assembly configurations. Clearly, it is not humanly possible to test all configurations through trial-and-error. Instead, the idea behind the fit optimisation project is to take advantage of modern scanners and computation power to test each configuration beforehand. Each subpart is 3D-scanned and configurations run through an algorithm based upon a desired score of least disturbance. Albeit still in a premature stage, the results show great promise and the algorithm quickly computes sets of part configurations with substantially better outcomes than traditional assembly.
5 ANALYSIS

To recap research area one and two, the data-driven perspective of Industry 4.0 seems plausible and a key focus area for improvements (Lee et al., 2013; GTAII, 2014; Wang et al., 2015). The increased usage of connected sensors will inevitably create big data volumes that create opportunities for new process knowledge to be acquired (Barnaghi et al., 2012). Through the usage of already well-established analysis methods such as ML, PCA and PLS, the theory suggests that data can be successfully analysed (MacGregor & Kourti, 1995; Monostori et al., 2016). The increased collection and application of multivariate tools is seen as a well-suited solution to complex manufacturing situations (Wuest et al., 2016). However, the introduction of these concepts in reality is not clearly mapped, due to the high level of customization needed in each situation. As such, further examination of GAS applicability will be analysed in this chapter.

5.1 Assessment of Quality Data Management at GAS

There is no need to stress the necessity of adequate quality management to GAS. The employees take great pride in their achievements and product quality alone. A statement which is reinforced both by internal documentations and our empirical findings. The recent shift in quality perspective is a response both to the challenge of the Road 2 Zero-initiative as well as an acclimatisation to digitalisation of manufacturing.

Today, the main hinder of data-driven manufacturing is the abundance of quality data systems regarding some processes, like Process A. To gain sensibility and promote the usage of data-driven manufacturing, the system needs to be streamlined as to avoid perplexity. Storing bits of data in a handful of systems makes it near impossible to piece together a true picture of an earlier event. As such, root-cause analysis and other improvements is potentially severely reduced in effectiveness when based on half-truths.

5.1.1 Collection & Storage

The current GAS mind-set of collecting data is interpreted as it is primarily done to control customer requirements, aviation standards and critical areas identified by the internal design team. These data points are all found on the product itself, which makes the collection method heavily weighted towards measurements of product characteristics. The data is categorized in key characteristics and non-key characteristics. Altogether the collected product data is stored in the common system QSYS, which everybody have access to. The availability of the data is therefore high. Worth noticing is the absences of focus on process data, which is illustrated in Figure 4.3. Their data philosophy regarding process data is to only monitor it in an adequate manner. The potential value multivariate process analysis can create is not mentioned in their strategy.

In combination with the increasing use of the method PPAP, process FMEA’s will become more common. The process risk analysis will help identify critical process parameters and incorporate a more process focused thinking. However, the method does not specify that the process parameters should be controlled or stored in any specific way. The method lacks a multivariate analysis perspective, which clearly contradicts the perspective presented in the theoretical framework. Since this method heavily affects which parameters that will be monitored in future production lines, making changes to the standard is one way to incorporate process monitoring.
**Internal Processes**

When the focus shifts towards the specific processes this mind-set on collecting and storing data is further strengthened by the previous analysis. Especially Process B is an example of this. Product data is measured and stored automatically in measurements machines. This is very effective, because it does not require any manual input from operators, which removes many problems the other processes are plagued with. However, all data that is being created by the machines are only monitored and not stored in an integrated system or with part number identification.

Process C is an outlier to the previous mind-set of process data. Instead they have in the design phase of their production line identified the key characteristics of the process as well. And they do not only monitor these characteristics but also store the data. Their standalone process system (PKR) is a great way of accomplishing this. Worth noticing is that not all available process data is gathered, only process key characteristics. Due to the process age the data collection is not automated, which is both cost ineffective and increase errors in the quality of the data.

Process A is plagued by a complex data storage problem. Storing information in six different systems make any attempt of analysis very difficult. The storage is very ineffective due to the same data being stored in several locations. This is likely due to the old age of the process. The collection method is primary manual with possess several downsides. However, there is good things to learn from this process as well. The chemical process team have sub-optimised their part of the process. Which means that extensive process data is collected and stored from the chemical baths. This makes it possible for the chemical team to assure the output of the product and therefore work in a reactive way. The reason for the tight control is due to the increased regulations regarding chemical operations. The overall mind-set of GAS is therefore once again prominent, if something must be measured it is.

**Case Examples**

Comparing GAS mind-set with its sister site in Norway many identical factors is found. During GAN’s multivariate project, they run into the same realisation that far too little data is collected for any extensive analysis. They also identified reasons as to why data is collected, with the conclusion that it is mostly because of requirements, not because of the great value deep analytics can generate.

When comparing the mind-set towards the case example of Electrolux, things change. They seem to have acknowledge the importance of data analytics, due to their department of 15 people working with “advanced analytics, intelligence and reporting”. As the title suggests one might think that they are working with advanced analytics, such as machine learning or multivariate analysis in their department, but the efforts of the team have so far been on creating the required data infrastructure data analytics require. Their biggest problem has not been the analytics itself, instead the problem is the underlying foundation of data, consisting data of low quality and volume. Even though their industry of consumer electrics differs greatly from the aerospace industry, their investment into analytics and data management is inspiring.

5.1.2 Quality Control & Analysis

In comparison to collection and storage of product data, which has been a strong point of GAS culture, analysis is underdeveloped and omitted. With that said, stating that quality control is non-existing would be a blatant lie, instead it is conducted rigorously and therefore GAS never gives the benefit of doubt on uncertain details. Certainty a policy worth of earnest consideration and an exemplary practice. This comes as no surprise for a first-tier supplier of products in a demanding
sector such as the aerospace industry. However, due to the strict demands and policies, one might believe that GAS lacks audacity to realise its true potential from a data-driven perspective.

As previously discussed, the coming (or is it already expected?) paradigm shift towards Zero Defects manufacturing puts immense pressure on aircraft suppliers. Either way you twist and turn it, the transformation of raw material into finalised products will always result in variation. It could be due to a poor delivery upwards in the supply stream or an unexpected shift in the process outcome. Therefore, it is inevitable that a defect ultimately slips through. However, the task of Zero Defects presents an interesting challenge for the aerospace industry. When faced with such goal, there is little doubt that an unexpected and disruptive solution will emerge as time presses on. The first action taken in order to meet the increased target goal for GAS is a remodelling of their way of working. From a traditional approach of reactive acting upon deviation, to a more proactive stance is now flowing top-down through the organisation.

A critique of the proactive approach is that yet again that the majority of its strategy revolve around the measurement of key characteristics. You can pin-point KC all you want in a univariate control chart, but without proper accounting of the process data, the analysis runs the risk of falling flat and result in sub-optimisations. In addition, one might wonder how much of a proactive approach the current quality perspective holds. Surely, it is an improvement from previous perspective, but without actively using inline measurements and automatically compensation during processing, you are still measuring upon completion. Even if the reaction time now is substantially quicker than before.

With extensive assessment of correlations between both product and process parameters, a multivariate analysis could enhance its output through increased knowledge of latent structures (MacGregor & Kourti, 1995). Successful clustering of process outputs would help root-causes analysis and proper PCA and PLS lays the foundation of more advanced analytics and Machine Learning. A development which the future surely holds in its grasp as we advance further into the era of digitalisation.

**Internal Processes**

Due to the nature of the different process examined at GAS, the amount of analysis conducted and its necessity differentiates. Process A is characterised as a chemical process coupled with semi-automatic operations. For the time being, the accompanied robot stores little to no data and it might not be economically viable to invest into modern technology here due to the limited production volume. The complex situation of documentation around Process A, quality control and analysis faces an unnecessary uncertainty in regard to data quality. Moreover, the chemical aspect of the process introduces an aspect of intricacy which needs to be considered when conducting process improvements.

However, a couple of possible improvements surrounding quality control and analysis are plausible. The process’ current situation, wherein the blaster is used by more products than the current one specified in Process A, alternative control charts than the default Shewhart-diagram is of interest. For example, Montgomery (2013, p. 453) proposes DNOM-charts for machines that handle several products, each characterised by a low volume. Furthermore, the intricacy of the later chemical baths is measured and analysed by default Shewhart’s as well. Even if this section of Process A currently is stable and robust, it would be of interest to try application of multivariate analysis tools such as PCA or PLS to uncover latent correlations. Montgomery (2013, p. 538) argues that PLS shows promising application in chemometrics due to the immense number of variables in use.
Process B is of great importance to GAS and the product produced here has a central position in the Road 2 Zero-strategy. With a great economic outlook for a foreseeable future, Process B is expected longevity. Moreover, with volumes far higher than your average process, Process B is capital intensive and the machines near state-of-the-art. As such, it exists a fine balance between the need to produce and the will to improve. Every bit of efficacy results in a magnitude in output, but as there is little room for error, GAS is sure to tread carefully. As previously discussed, and illustrated in Figure 4.8, Process B already does extensive product measurements. However, in order to strengthen the analysis, ensuring the addition of process data (i.e. machine data) would potentially uncover latent structure of correlation between different process steps. If so, this knowledge would serve greatly as to understand variation and aid in the implementation of machine optimisation and compensation.

Process C is currently showing stable outputs and the quality control and analysis has a fine record. However, the situation at hand is univariate, and to improve beyond current status Process C would have to extend their efforts into the multivariate domain. As mentioned earlier, the collection of data is today very depending on operators. Therefore, a first step would be to incorporate more sensors and correlate process settings.

*Case Examples*

Due to the similar conditions found in Norway at GAN, parallels can be drawn between their work and GAS. In general, the current situation of quality control and analysis at GAN could be described as of comparable to GAS. However, the practical procedures still differ. For example, GAN does not use QSYS, instead one could loosely compare their internal system to KPS. In that aspect, they have simultaneously realised the value of quality data management. In line with the global strategy that is now being developed, the chances of software standardisations between sites are high. As such, through knowledge exchange the implementation of whichever software should be relatively smooth between sites.

Albeit promising, the initial pilot projects conducted at GAN seriously points out the need for adequate data quality. Seeing how the attempt at multivariate analysis greatly exceeded the planned timeline, it goes to show that it is easy to underestimate the difficulties of collection. Even though the results from the analysis were questionable in terms of statistically accuracy, one should not lose hope. With its newfound knowledge of requirements and improvements to the existing data infrastructure, the next attempt is sure to fare better. Moreover, with the success currently underway with the Green Health Monitoring as well as the Part-Fit Optimisation projects, data-driven manufacturing gains momentum for each passing day at GAN.

As the interview held with Electrolux was primarily about Industry 4.0 from their perspective, proper data regarding quality control and production analysis is lacking. Therefore it would be ill-advised to guess their procedures. From a more general perspective of analytics, the quick growth of the AIR-department goes to show that Electrolux finds the potential of big data analytics as important.

5.2 Industry 4.0 and Internal Process’ Applicability

With the future possibilities of digitalisation advancements, the question eventually boils down into whether it is imaginable to incorporate these tools to GAS’ internal processes. From a theoretical perspective one would easily believe it so. Analysis of the conducted interviews ultimately breakdown to a somewhat modest, but hopeful attitude towards the concept of Industry 4.0. The outcome is presented in Figure 5.1. For further details about their knowledge, see Appendix D.
Generally speaking, the employees at GAS believe that the future certainly holds a lot of digital improvements that will effect production positively. For instance, streamlining the usage of production and planning systems will lead to higher outputs as efficiency increases. The interviewees agree that ultimately, the effects will have positive impact on production in one way or the other in terms of product quality.

A portion of the employees approached sees the future as merely enhanced by the recent developments. At the end of the day, the aircraft sector is stuck with an endless amount of demanding certifications not easily sidestepped without rigorous justification. As such, the majority of production will remain the same and the revolution of Industry 4.0 will fail to come. The improvements following the digitalisation will be no different from the usual advancements one could expect over time.

Furthermore, a majority of the employees puts a lot of faith into the Industry 4.0-concept, albeit not totally convinced of a total revolutionary upheaval. However, it is clear to those people that the most prominent tools following the digitalisation will effectively change the prerequisites of manufacturing domain. Therefore to be able to keep up with the changing customer demands and general competition, it is necessary to ensure a certain forward-position and amenability in regards to these advancement. Else, the risk of becoming obsolete and seen as too conservative could potentially jeopardise future revenue streams.

The visionary-minority on the far right believe that the factories of the future will be autonomous and self-driven, transforming the manufacturing domain completely. With that said, the perception of these employees does not necessarily mean that they believe that production will be transformed come Monday morning. Instead, these individuals are able to imagine futuristic scenarios and extrapolate upon the rapid development pace seen in other technological areas. Moreover, these employees possess a convincing faith in the ability of Industry 4.0 to solve the problems of today.

Albeit visionaries can put their trust into the concept, it seems to be necessary to persuade both employees and stakeholders not yet convinced through incremental demonstration of its
capabilities. Initially, much of these demonstrations include the picking of low-hanging fruit, i.e. increased traceability (Sadeghi et al., 2015) as well as increased situation awareness through internet of things. Thus, enabling both the people and machines at GAS to make more informed and intelligent decisions (Barnaghi et al., 2012; Shin et al., 2014; Biswas & Sen, 2017).

When, and if, these demonstrations succeed one would surely believe that it results in an increased demand for these levels of sensor and data-driven production. As the empirical findings showed, even though the capability to capture and store process data already exists today. It is not being acted and capitalized upon because the need for this kind of data is not explicitly specified. As a result, the quality data management systems in place today becomes underutilized and sub-optimal seen to its potential. If one views the adaption towards Industry 4.0 as incremental rather than discrete, it becomes obvious that it is necessary to start in Act I (see Section 3.5) and utilize industrial big data via internet of things as key enabler (Lee et al., 2013).

Moreover, the employees see the ability of advanced analytics as a possibility in near-reality. Through proper caution to the absolute imperative of data quality in these scenarios (Hazen et al., 2014) and with respect to the successive refining of data through the model of the knowledge discovery process (Fayyad et al., 1996). It is not unlikely that advanced techniques such as predictive analytics and machine learning could be taken advantaged of (Shin et al., 2014). Examples of this curiosity for techniques like this could be seen at Case Example II: GKN Aerospace Norway AS, as presented in Section 4.6.

Further advancements through Industry 4.0 towards digital twins, machine cognition, and self-optimization have yet to show practical implementations according to the findings of this thesis. As such, these ideas have only been discussed with the most inveterate employees on a purely conceptual level. To conclude the general consensus at GAS, it seems to be leaning towards that the revolution of Industry 4.0 is not a one-time investment. A mind-set surely sound to have, as it is necessary to proceed these unknown times with caution. These findings confirm with the thoughts of both Ribeiro (2017), i.e. the pressure and expectations towards CPS, as well as the industry norm of operating under tight margins presented by Wang et al. (2015). However, with the inherent curiosity present at GAS it seems likely that once the deployment of Industry 4.0 gains momentum, the enthusiasm will take hold. If the possibilities outweigh potential drawbacks, bureaucracy matters and standards will adapt over time.

5.3 A Multivariate Attempt on Process A & B

Process A

Due to Process A’s data containing many categorical input parameters, it was difficult making use of multivariate tools. In addition, the output of the process was categorical, which means that it only got a pass or deny in the quality control. Most of the analysis consisted of analysing specific parameters in relation to their respective output, see Appendix E for example. In Appendix E, it is evident that very low tray angle resulted in more negative outcomes. When combing the information with which blasting media used it became evident that this happened with only one specific media. With further investigation the reason was that this blasting media created a greater wear on the nozzle equipment, which created this low angle. Similar realisations could be seen from the data, however no real new insight was created from the analysis.

The lack of depth on the analysis, is mainly due to the quality of the data input. Since there were only process parameter settings that was recorded, a design of experiment would be more suiting
in this situation. However, this has already been tested on this process, but unfortunately without generating much knowledge on the affecting parameters. The design of experiment project used similar input variables as the ones in this analysis and it strengthens the idea of the lacking input variables.

Our analysis is that the process steps, mainly the blasting operation, contains too much internal noise that is not being measured. For example, if the process uses the exact same settings, two widely different outcomes can still occur. Therefore, both analysis and design of experiments are of no use with this dataset. Instead more parameters need to be monitored in the process. These parameters are likely not as easy to measure and might require extra equipment due to the machine being old. Moreover, it would be of great use a gradation when measuring the outcome.

**Process B**

Process B used only numerical values and the dataset was relatively large. Therefore, a PCA was conducted and the output from the PCA is presented in Appendix F. The data in the PCA was further divided into different clusters. However, without further information this data is difficult to comprehend. The analysis lacks information about the outcome of the product. With that information it is possible to cluster the PCA in desirable regions and undesirable. Luckily this is possible with the deviation reports from the ERP-system which can act as a measurement of the outcome. Due to restricted access to the system and lack of time, this analysis was not performed.

If the deviation report was used as quality measurements in the PCA, it is still difficult to draw conclusions. It would still be unclear on what to actually change in the process to prohibit this result, since the only information available is product outcome. To overcome this issue the analysis could have been performed on the process data instead. The PCA would instead show these process parameters, which would be clustered towards the product quality measurements. This way optimal machine settings and complex multivariate correlations can be found. Altogether creating more knowledge over process and taking several steps towards reaching the company’s Zero Defects vision.
6 CONCLUSIONS AND RECOMMENDATIONS

A conclusion of the thesis’ findings at GAS is that a consensus of optimism towards the future and Industry 4.0 can be found. As visualised in Figure 5.1, the majority of the interviewees believe that a selection of the tools contained within Industry 4.0 could positively affect their industry and products. However, this mind-set does not necessarily prove that the transition towards a data-driven manufacturing will go down smoothly. Seeing how the majority view the concept of Industry 4.0 as a futuristic project, rather than something achievable today. Resulting in a somewhat conservative outlook compared to other industries and the literature findings. As such, it puts a lot of pressure on the inevitable implementation attempt and its initial success. Aside from a couple of visionaries, the view of Industry 4.0 as futuristic relieves the procedures of today from immediate development, as one could view it as a solution brought in by the coming generations. Ergo, creating a situation wherein a desire of improvement exists, but current regulations and standards acts as hindrance of experimentation and allowance of mistakes. Yet, all in all it is obvious that a shift in the quality perspective and initiatives such as Road 2 Zero is starting to pave the way towards more innovative process methodology in general.

From the analysis of the three internal processes at GAS it is clear that some processes are not developed alongside technology advancements, resulting in old processes using an outdated collection, storage and analysis method. Big changes (e.g. machine replacement or methodology) require extensive work due to flight safety standards and customer contracts which have to be renegotiated in that case. This is affecting GAS company culture in the area of continues improvement negatively and may be one of the reasons the area lacks focus.

The outdated data management is most visible in Process A. Due to not being continuously developed, the procedure still contains data collection and storage on paper in some instances. The process relies on operators’ expertise to a great extent and its process correlations are not fully understood or mapped. The quality control consists of a visual judgement with no gradation. Except for the chemical baths, little data is stored. The storage of the data resides in a handful different systems which makes its accessibility for analysis difficult. Any root-cause analysis will therefore both be difficult and time consuming. Because of this, the applicability of multivariate analysis tools and eventual Industry 4.0-tools is at the moment low.

The newest process, Process B, has adequate data collection, storage and analysis in place for the product’s quality data. The modern machine park also offers the ability to rather easily store the machine data with part identification, something that is not done today since nobody is requesting the data. Connecting the machines’ process data together with the quality data creates a solid data-driven foundation for multivariate analysis, machine learning algorithms and even a cyber-physical system in the form of a digital twin. Resulting in greatly enhanced knowledge about complex relationships in the process. Thereby, the potential to advance towards more Industry 4.0-related tools is promising for Process B.

Process C’s unique mind-set to process data management can be seen as good role model for the other processes. Continual measurements of the process state before and under the production that are stored with part identification makes it a good procedure. However due to their manual collection of data, there are still things to improve and to expand their analysis to the next industrial revolution as more data has to be collected from the process. In that scenario collecting the data manual is too time consuming and automating the process is inevitable. With an addition of more
sensors capable of measuring what is currently done manually would propel the process towards Industry 4.0 rapidly.

**Lessons Learned from the Multivariate Analysis**

The importance of the data collected became evident when analysing Process A’s data. It lacked the data infrastructure to easily collect the data needed. Only the compilation of data used in this analysis was very time consuming and difficult to gather, which is understandable from the unstructured data storage system presented in Figure 4.7. The analysis concluded that the collected data did not represent the Process A’s real outcome to the required extent. More parameters should be measured in order to better understand the process. For example critical blasting machine parameters, such as blasting media speed. Moreover, the reliability of the measurements can be controlled through a measurement system analysis.

Process B’s analysis taught a different lesson. Even though the dataset was extensive in volume, the analysis needed other types of information in order to draw useful conclusions. One solution is to look at the deviation reports and correlate them with the conducted PCA shown in Appendix F. Even if this alone would not result in process improvement recommendations, it could at least increase the knowledge of the process and important variables. Another solution is to connect the machines process data to the quality measurements. Due to the modern machines, this seems possible. This information has the potential to result in fine-tuned process parameters and will facilitate root-cause identification in the case of deviations. This also paves the way for more advanced tools, such as machine learning algorithms.

### 6.1 Recommendations

In order to propose a sensible roadmap towards progression and to differentiate the pressing concern between recommendations, a division is made between the coming five and 15 years respectively.

#### 6.1.1 The Beginning of an Era: Within 5 Years

**Collect All Data, Especially Process Data**

As can be seen from the theoretical framework the start of the next industrial revolution is the increased data collection through IoT. From this study it is clear that too little data is gathered from the studied processes and that GAS is not in line with this mind-set today. The strategy for data management should be changed to emphasize the importance of all kinds of data and especially process data.

**Create Easily Accessible Data**

The IT systems today are already perceived as difficult to navigate and difficult to extract data from. If the volume of data is increased by the tenfold, the difficulty would increase. In order to extract value from the data it is absolutely core that the data is easily obtainable and understandable, preferably in one integrated system.

**Use Advanced Analytics**

With the prior recommendations in place, it is time to expand upon the usage of analysis tools. Multivariate tools, such as PCA, and machine learning algorithms can be very effective in optimising the processes and gaining increased knowledge over complex correlations in the process.
These three suggestions together create an important aspect of the future quality data management at GAS. The future quality data management mind-set is presented in Figure 6.1. The figure highlights the importance of both product and process data, as well as other types of data that might affect production, e.g. environmental data, such as temperature and pressure. These are all stored in an accessible integrated system and later analysed with advanced analytical tools.

![Figure 6.1 – A proposal for a future state of quality data management at GAS with increased data collection, streamlined storage as well as advanced analytics in place](image)

6.1.2 Pyramid Ascendance: Within 15 Years
The next step beyond the initial steps is to create a digital twin of the production line. With extensive usage of real-time sensors in place, creating a virtual real-time manufacturing facility which can simulate complex future scenarios is possible. This will immensely decrease variation and a scenario of zero defects will not be far off. From here on the horizon is difficult to foresee, the machine cognition phase may rapidly decrease the need for operators, at the same time this might be more difficult in practice than expected and might be far off. However, the foundation for this future still starts with extensive data management, something attainable today.

6.1.3 Other Recommendations

**Use Appropriate Control Charts**
The usage of ordinary control charts in every situation is not optimal. Some stable operation can utilize EWMA or CUSUM-charts to easily detect trends in production, which are not detectable on ordinary control charts. Another example is when the machine operates on several different low volume products, then DNOM-charts are more justifying to the machines performance.

**Use 3 Sigma Control Limits**
Currently the control limits at GAS are estimated from the tolerance limits, which means that the control charts typically act as ways to detect flawed products. They do not function as indications on when the process is statistically out of control. Therefore, continuous improvements are not part of the daily routine and the root-cause identifications are mostly done on flawed parts.
7 DISCUSSION

7.1.1 Evaluation of Credibility
Due to the descriptive nature of the thesis much of what is studied is depended on time and place. That is, if one would try to repeat the same case study, the procedure would not necessarily generate exact response due to potential technological development or terminations between studies. However, this thesis aimed to fulfil its academic commitment and is therefore as transparent as possible to enable eventual longitude studies at GAS or meta-analysis. Yet, as described in Section 2.4, data collection was mainly conducted through confidential interviews, and albeit a summary of interviewees can be seen in Table 2.1, it hinders the repeatability of later studies. Moreover, the quantitative sub-analysis of the thesis and its datasets are restricted by company policy and the accessibility of these files are therefore limited outside GAS.

Seeing how the topic of Industry 4.0 is relatively niched and perhaps not common knowledge between all employees at GAS it was necessary (as always) to ensure an adequate degree of validity. To assure that the thesis actually studied what it sought to do, the sampling of interviews was aided both by its supervisor as well as reference group. An overview of the interviewees’ general knowledge about the topic can be seen in Appendix D. Furthermore, multiple sources were used to reinforce the thesis representation of the subject. As such, the thesis’ external validity is somewhat strengthened by its widened state, however the limited number of case examples are not deemed sufficient to characterise the entirety of the manufacturing sector.

Ultimately, the presentation following the finalization of the thesis at GAS were greatly appreciated and received positive feedback. As one member of the audience proclaimed: “I have been to several presentations regarding Industry 4.0 and this was the most pedagogical rundown of the subject yet”. As such, we strongly believe that the thesis achieved its aim fulfilment.

7.1.2 Implications of the Study
To overcome the increasingly complex manufacturing environment, the future seems to behold several solutions. When reading about Industry 4.0 the transition can be seen as demanding and far into the future. This study aimed, through research area one and two, to clarify the data-driven transition in to Industry 4.0 and to identify analysis tools used in this future scenario. These two research areas have together helped breaking down the awaiting revolution into more manageable pieces, which is rare to come by in the literature. Therefore, this thesis clarified the data-driven perspective of Industry 4.0, both from an academic and organisational standpoint – leading to an overall increased knowledge of the subject.

In the case study at GAS, the thesis found an increasing need for more process knowledge due to their intricate processes. Through research area three, three different internal processes were analysed in correlation to their specific needs and by research area four, specific recommendations were given customized to each process and the organisations overall strategy. As the study found, an overall shift in GAS mind-set over process data is needed. As such, the study provides increased knowledge on how GAS already today can prepare for future technology advancements. The studies implications do not limit itself to GAS, the study thoughtfully described each process in detail in order for other readers to see applicability in their own processes. Due to the distinct disparity in process design between the internal case studies as well as the similar conclusions found in the case examples. As such, we believe that the general recommendations of this thesis stand true and applicable in other complex manufacturing situations.
7.1.3 Future Research Proposals

Seeing how this thesis began as exploratory and later descriptive, it leaves with a potential for a longitude study at GAS to follow-up its journey towards Industry 4.0. Both to conclude whether this thesis bore fruit or if GAS decided to continue on a different route. As shown throughout this thesis, the subject of Industry 4.0 is of a delicate nature and its fragility of implementation in already existing processes raises questions of its applicability from an economic standpoint. Therefore, it would be of great interest to continue this case study from a multidisciplinary perspective.

The project of Industry 4.0 could be seen as a worldwide interest for the manufacturing sectors as well as a political drive force. As such the possibility of a global meta-analysis exists and should surely deliver interesting findings and current best-practices. However, due to the economical incitements of business excellence, it could potentially hinder a researcher’s accessibility as companies decides to keep its practices secret.
REFERENCES


48


APPENDIX A: INTERVIEW GUIDE SYSTEM OWNER

Introduction of the master thesis and information as to why this interview is of importance.

Is it okay if the interview is recorded?

Disclaimer: All information provided will be treated confidentially.

Background

Tell us about your position in the company?
Tell us about your area of responsibility regarding system XX?
For how long have you been responsible for system XX?
Have you had any other experiences from systems other than system XX?

System

What is the purpose of the system?
   Is it meant to be an overall solution to data management?
Which type of data is stored in the system?
   Which type of data is meant to be stored there?
   What is the purpose of the stored data?
Is the system adaptable for changes?
What limitations does the system have?
   How well can the system be integrated with other systems?
   Is the system compatible with older machines?

Data collection

How is different data collected?
   Manually or automatically?
How is data reported manually?

Analysis

How is data structured in the system?
   Are there limitations for the user?
Is data easily available for analysis?
Which analytical tools does the system provide?
   What is the purpose of the tools?
   Which conclusions are can be drawn from these?
Can analysis in real-time be performed in the system?
Can the system perform multivariate analysis? Correlation analysis?

How is the result from the analysis communicated back to production?
For the analysis that you are not able to perform in the system, are alternative software available?

About SPC tools,
Which control charts are used?
How are the control limits calculated?

In comparison with previous systems, which are the advantages/disadvantages with the current system?
Which features will future system need in order to surpass the current system?

How do you feel with future systems? Buy or develop internally?

Is the system part of your work towards Industry 4.0?
Are there a strategy for developing the system to be able to handle more data, which will be of use in Industry 4.0?

If no, do you wish to see a strategy for integration of Industry 4.0 in the system?

**Vision for Industry 4.0**

How important do you think data analysis currently is for the company?
How important do you think data analysis is in five years for the company?

Have you heard about Industry 4.0/Smart manufacturing/Factory of the future and what is your perception of the concept?
Are you aware of any strategy for the adaption of Industry 4.0-tools within the company?

If no, do you think the company needs a strategy for Industry 4.0?

How far away do you see an eventual implementation of Industry 4.0?

Do you see any negative sides with the future scenario Industry 4.0 illustrates?
APPENDIX B: INTERVIEW GUIDE METHOD OWNER

Introduction of the master thesis and information as to why this interview is of importance.

Is it okay if the interview is recorded?

Disclaimer: All information provided will be treated confidentially.

Background

Tells us about your position in the company

Tell us about your responsibility regarding method xx

For how long have method xx been your responsibility?

Do you have any experience of methods others than your current?

Method

What is the purpose of method xx?

What data do you generate and collect from method xx?

What is the reason behind these measurements? Have they proven noteworthy via Design of Experiments or other methods?

How do you measure data today?

Manual or automatically?

How do you report manual collected data?

For what later usage is data collected?

Do you wish to measure more data from the method?

Analysis

What types of analysis is done on the method?

Purpose?

Who conducts these analysis?

Which conclusions are you able to take based on this?

Is data easily accessible for analysis?

How is the results from analysis reported back to production?

Statistical Process Control

Is SPC used?

Which control charts? Why?

How do you choose control limits?
In general, how do you consider the awareness of SPC at GAS?
What is your view on the general competence regarding SPC?
In your opinion, is there a need for more education?
How important is analysis of data today for the method in general?
Is there a strategy in place to develop the method?
If no, do you wish to see a strategy for improvements?
In your opinion, how do you wish to develop the method further?

Vision for Industry 4.0

How important do you think data analysis currently is for the company?
How important do you think data analysis is in five years for the company?
Have you heard about Industry 4.0/Smart manufacturing/Factory of the future and what is your perception of the concept?
Are you aware of any strategy for the adaption of Industry 4.0-tools within the company?
If no, do you think the company needs a strategy for Industry 4.0?
How far away do you see an eventual implementation of Industry 4.0?
Do you see any negative sides with the future scenario Industry 4.0 illustrates?
APPENDIX C: INTERVIEW GUIDE INDUSTRY 4.0-VISION

Introduction of the master thesis and information as to why this interview is of importance.

Is it okay if the interview is recorded?

Disclaimer: All information provided will be treated confidentially.

**Background**

Tell us about your position in the company?

For how long have you worked here?

**Vision for Industry 4.0**

How important do you think data analysis currently is for the company?

How important do you think data analysis is in five years for the company?

Have you heard about Industry 4.0/Smart manufacturing/Factory of the future and what is your perception of the concept?

Are you aware of any strategy for the adaption of Industry 4.0-tools within the company?

If no, do you think the company needs a strategy for Industry 4.0?

How far away do you see an eventual implementation of Industry 4.0?

Do you see any negative sides with the future scenario Industry 4.0 illustrates?

**Other topics of interest**

Process vs. product data.
APPENDIX D: RESPONDENT’S KNOWLEDGE LEVEL

Knowledge of the term Industry 4.0
Have you heard of the concept?

Classification of respondent’s knowledge level of Industry 4.0
## APPENDIX E: DATASET & JMP OUTPUTS FOR PROCESS A

### Dataset from Process A

<table>
<thead>
<tr>
<th>Date</th>
<th>Result</th>
<th>Kat-Ni-Bath</th>
<th>Kat-HCICl Change</th>
<th>Kat-CuChange</th>
<th>Kat-Nozzle</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017-01-04</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-04</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-04</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-04</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-04</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-04</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-04</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-04</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-04</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-04</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-04</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-04</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-05</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-05</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
<tr>
<td>2017-01-05</td>
<td>1</td>
<td>HM1</td>
<td>HCl1</td>
<td>Cu2</td>
<td>5</td>
</tr>
</tbody>
</table>

### Tray angle vs. Result

![Tray angle vs. Result](attachment:image.png)
Tray angle and Media vs. Result
APPENDIX F: DATASET & JMP OUTPUT FOR PROCESS B

Dataset from Process B

PCA with colour clustering Process B