MASTER

The informational abilities and opportunities of Manufacturing Execution System data

Stassen, I.T.J.

Award date:
2016

Eindhoven, April 2016

The informational abilities and opportunities of Manufacturing Execution System data

By I.T.J. Stassen

Bsc. Industrial Engineering

Student identity number: 0718738

in partial fulfilment of the requirements for the degree of

Master of Science
in Operations Management and Logistics

Supervisors:
Dr. ir. R.M. Dijkman, Eindhoven University of Technology
Dr. R.J. De Almeida, Eindhoven University of Technology
Ir. I. Rothkrantz, Deloitte Consulting Nederland
L. Boerman Msc, Deloitte Consulting Nederland

TUE. School of Industrial Engineering

Series Master Theses Operations Management and Logistics

Subject headings: Design Science, Framework, Information System, Knowledge Discovery, Manufacturing Execution System, MES, Manufacturing Operations Management

Abstract

Manufacturing Execution Systems (MES) operate in the Manufacturing Operations Management (MOM) environment and focus on the actual execution of production. MES manages, measures, analyses and optimizes the production operating. The market for MES is a billion dollar market, though MES systems still seem relatively unknown to many people in production. Also, in practice MES gets vaguely described. This research aims to provide a structured overview of the possibilities and opportunities of extracting information from MES system data that provide insights for improving the MOM. This overview is, to the best of the writer's knowledge, not present in current literature yet.

In order to establish this overview, two frameworks named as Informational Matrices were established. First, the Current Informational Matrix, focused on current MES functionality, and second the Future Informational Matrix, focused on advanced data analysis methods. Both matrices exist of Informational items that are pieces of information that one could extract from MES. Because MOM considers four areas (Production Operations, Quality Operations, Maintenance Operations and Inventory Operations), there are separate informational matrices for each area. All Informational items in the matrix have a set of properties that provide more information about the item. Usefulness percentage score, established by a questionnaire among MES experts, generates a ranking among the informational items in both matrices. In the Current Informational Matrix, the time frame of the informational item showed that in current MES the focus is past and present oriented which makes the MOM reactive. Additionally, the degree of standardization for data capturing and (performance analysis) per MOM area showed that in the field of production operation the highest degree of standardization is present which makes the configurability of the informational item most generalizable among different MES implementations. In the Future Informational Matrix, main and sub groups were defined in order to structure the findings from the literature search for knowledge discovery with manufacturing operations data. The methods used for each informational item showed that a wide range of methods can be used in order to extract the informational element. The year of publication of the articles, on which the informational elements are based, indicate the amount of time still needed before it becomes available in MES systems as theory takes time to be translated to practice.

The case study demonstrated that the established Informational matrices are both applicable and usable. It is possible to extract the informational items from the data and that there are many opportunities in the presentation of this informational item to enable fast and reliable decision making. The Current Informational Matrix enables a company to assess their own MES related choices and to better discover their own informational needs. Also the ranking provides an opportunity to benchmark their MES (choices) against what is considered useful by the MES experts. The Future Informational Matrix enables a company to be prepared for the (possible) future abilities of data analysis, in other words for future MES. This is very useful for companies to take into account when making decision about data capturing, data structures and MES today.

In a real-life environment different factors will affect how MES is integrated in a company and within its IT structure. The amount of legacy systems and the managerial choices of which information is important to monitor and control, will affect the actual information captured within a MES. Also, every production company has specific processes, specific needs and specific company questions. This makes every MES somewhat different. However, having an overview of the abilities and opportunities of a typical MES can be useful for every company regardless the specific process. It provides company insights in MES abilities and it triggers them to re-evaluate their own MES and MES related choices in order to further optimize their Manufacturing Operations Management.

Preface

This thesis is the result of a graduation project that has been conducted at Deloitte Consulting Nederland in completion of the Master Operations Management & Logistics at the faculty Industrial Engineering at the Eindhoven University of Technology.

I would like to thank my supervisor from Remco Dijkman and Rui de Almeida e Santos Nogueira for their guidance and support during the process of this thesis. Remco, for your guidance and advices during the whole process. I really enjoyed working with you and I particular appreciate your flexibility and expectation management in the preparation of our appointments. Rui, I would like to thank you for the fresh outside view and you extensive knowledge of data mining related topics.

I would also like to thank Deloitte for giving me the opportunity to conduct this research in their professional and energetic environment. In particular I would like to thank Ivo and Linda. Ivo, thank you for you critical view challenging me to keep the business value in mind. Linda, thank you for guiding me through the process and for keeping me on track when I was seeking for a focus in the project. Also I would like to thank the team of Operational Excellence (OE). I have felt very welcome and I really enjoyed working at the office, especially on Fridays. The table football game with my fellow intern and other OE workers was an excellence way to spend a short break and to renew energy.

Also, I would like to thank my friends and family for the support during my student time, and this last period. My friends for making my student time a great journey to remember. We did not only have a lot of fun, but there was also support, serious conversations and learning how to organize memorable events. My parents for giving me the opportunity to study and support me along the way. And last but not least Arianne, for listening to all the stories about this thesis with great interest and your patients and support along the way.

I hope you all enjoy reading my thesis,

Iris

Contents

List of Figures

List of Tables

Abbreviations

B2MML	Business To Manufacturing Markup Language
CRM	Customer Relation Management
EH&S	Environment Health & Safety
ERP	Enterprise Resource Planning
HMI	Human Machine Interface
IT	Information Technology
ISA	International Society of Automation
KPI	Key Performance Indicator
MES	Manufacturing Execution System
MESA	Manufacturing Enterprise Solution Association
MOM	Manufacturing Operations Management
MS SQL	Microsoft Structured Query Language
OPC-UA	Open Platform Communications Unified Architecture
PLM	Product Lifecycle Management
RDBMS	Relational Database Management System
SCP	Supply Chain Planning
XML	Extensible Markup Language

1. Introduction

Manufacturing Execution Systems (MES) are information system that are widely implemented in the production industry and focus on the operations management or the actual execution of a production process. A MES operates between planning software at enterprise level and control software to floor level. MESA (Manufacturing Enterprise Solutions Association) defined MES in 1992 as follows:

> "*a dynamic information system that drives effective execution of manufacturing operations: Using current and accurate data, MES guides, triggers and reports on plant activities as events occur. The MES set of functions manages production operations from point of order release into manufacturing to point of product delivery into finished goods. MES provides mission critical information about production activities to others across the organization and supply chain via bi-directional channels"* (MESA International, 1992).

Generally speaking, MES is concerned with supporting the execution processes in a production plant or plant area while focusing on the use of machines, people and equipment. According to MESA's survey (MESA International, 1997), MES have provided manufacturing enterprises with some of the most impressive benefits of any manufacturing software, such as an average 45% reduction in manufacturing cycle time, a significant improvement of the flexibility to respond to customer demands, the realization of certain degrees of agile manufacturing and customer satisfaction. The great impact of MES is due translating the data to information at the right time to help making the right decisions. Therefore the information that can be extracted from a MES, in order to improve the manufacturing operations, is the focus of this research.

1.1. Problem introduction

MES have a lot of unused potential though they have been widely implemented since the early 1990s. Three reasons explain this gap:

1. The MES and MES's functionality is still relatively unknown to the production field
2. The is no comprehensive and/or consistent overview of the capabilities of a MES
3. The MES data analysis is still very basic with limited use of advanced analytics

The MES Industry has developed over the years in terms of formalization and in terms of market growth, yet the term MES is still relatively unknown in the manufacturing domain. On the MESA International website, a small survey with among 252 respondents shows that 41,9% of the voters are most unfamiliar with the term MES/MOM compared to terms like ERP, PLM, SCP, SH&S, CRM and other (MESA International, 2015). Though, the number of participants is not significantly high, it does indicate that the term MES/MOM is relatively unknown. Especially since this survey was conducted at a website where MES/MOM is a frequent and significant topic. This unfamiliarity with the acronym MES is enhanced by many vendors who do not brand their product with the name 'MES'. Only 15 out of the 71 participants of a MES research conducted by Iskamp and Snoeij (2015) have a brand name with the acronym "MES" in it. This could be due to the fact that the early MES did not have a good reputation in the manufacturing business. Many early MES were closed narrow built systems that lacked the configurability and flexibility it actually needed in order to adapt to changing business needs (NearSoft Europe, 2013). This created long lingering implantation processes and high service costs. Therefore, MES earned a reputation early on as "an expensive and risky endeavor that often did not deliver on initial return on investment goals" (Littlefield, 2012). Though, MES have evolved, since the advancement of computing technologies since the mid-nineties, into more powerful and more integrated software applications (Saenz de Ugartea, Artibab, & Pellerina, 2009), MES vendors today might want to prevent clients thinking back of these "expensive and risky" MES when considering their products. More information about the MES functions, architecture and background is provided in Chapter 2.

For those who are familiar with the MES, many still do not have a comprehensive and/or consist overview of the MES functionality and what added value it can bring to a company. There are organizations who have tried to establish a standardized framework for the MES functions (see Chapter 2), but all implemented MES are different from each other. There is some frequently implemented functionality but most organizations have to fit in into their own IT organization and want to do this in their own way. Kletti (2007) states that "the relevant data model used for the MES will be guided by the sector of industry and the production processes". This is supported by vendors (Van Veen, 2015) who states that every company has different views on what they believe is important information and in what format they believe it should be provided in. "In addition to a set of standardized key figures, the MES must also be able to calculate project-specific KPIs at the user's request" (Meyer, Fuchs, & Thiel, 2009). In other words there is a lot of customization present and no constant or comprehensive overview in terms of information that can be extracted from MES even though there are many similarities for every business.

The data present in MES, which can be translated to information through analysis, consists of data from the planning level, the control level, and data stored by the MES itself. "It is evident that the amount of information collected from control systems increases greatly with the degree of increased automation on the shop floor" (Saenz de Ugartea, Artibab, & Pellerina, 2009). Al this data capturing creates a potentially big and rich database. However, existing analytics in MES, "are coined by major short-comings considerably limiting continuous process improvement. In particular, they do not make use of data mining to identify hidden patterns in manufacturing-related data" (Gröger, Niedermann, & Mitschang, 2012). Current performance analysis in MES mostly aims at current efficiency levels and historic trends and does not use many advanced techniques from the big data field. This means that there is great potential in exploring these options.

1.2. Relevance

Since the 1990s MES have been implemented in manufacturing environments. In 2015, the market for MES is still rapidly growing and developing and is estimated to reach $12.6 billion by the end of 2020 at a CAGR of 10.85% between 2015 and 2020 (Markets And Markets, 2015a). Due to this size and revenue generation, it is a relevant market to research.

Also, a global trend is expected to disrupt many industries, including manufacturing. This trend is called the Internet of Things (IoT). The basic idea of this concept is "the pervasive presence around us of a variety of things or objects which, through unique addressing schemes, are able to interact with each other and cooperate with their neighbors to reach common goals" (Giusto, Iera, G., & Atzori, 2010). In the manufacturing industry, the IoT will generate a new revolution in ways of working when all machines, parts and products become interconnected. This is often referred to as Industry 4.0 or Smart Manufacturing. This is will not be happing in a far future, but it is already happening right now. The IoT in manufacturing market is estimated to grow from USD 4.11 Billion in 2015 to USD 13.49 Billion by 2020, at a compound annual growth rate (CAGR) of 26.9% (Market And Markets, 2015b). This IoT or Industry 4.0 will have great impact on how the information systems within factories are used, especially MES. MES has a unique position the IT architecture as it is a central point for data collection with both production management and data analysis functionality. This could make it very suitable to play a crucial role in the interconnectivity of devices as part of IoT. Some researchers state that planning systems will be more integrated so the need for a separate MES will disappear (INFOR, 2014). However, others strongly believe that MES will improve with (self-) learning from the past and improving or forecasting the current conditions (Critical Manufacturing, 2014a) (Cisco, 2014). MES and analytical power should be taken seriously when considering the strategy mix (Critical Manufacturing, 2014b). In other words, MES could play a crucial role in Industry 4.0 if it is able to incorporate smarter analytics and informational abilities.

Last, when exploiting the MES correctly it can generate significant savings for the MES user's company. An example from Schneider Electric's MES shows that for a bottle manufacturer, measuring downtime information and the status of all equipment automatically in MES enabled them to find the root cause for downtime during the changeover process. This led to a reduction of 50% in changeover time. This created savings in downtime, and additional savings in raw materials and packaging. A saving of $78,500 annually was established (Schneider Electric Software). This was established by only making use of current MES functionality and basic data analysis. As the analysis improves with advanced analytics the financial impact in terms of savings could increase as well.

Concluding, the market for MES is big and still growing. Also, the market conditions are changing rapidly and a need for improved analytics and informational abilities is present. Additionally, MES can generate significant savings when exploding correctly. Given these reasons and it is relevant to research this topic now.

1.3. Research goal

As stated in the problem introduction, there is a lack of overview of the current informational abilities of MES and there is a lot of unused potential of the MES database. Therefore the research goal of this research is formulated as follows:

Research Goal
Provide an overview of the possibilities and opportunities of extracting information from MES data that provide insights for manufacturing operations management.

The defined research goal has multiple aspects that will be discussed successively. First, the aim is to provide an overview. Currently, the possibilities of MES are too unknown or vaguely described and referred to as 'implementation specific'. Therefore, there is a need for a clear overview. Second, the overview needs to show both the possibilities and the opportunities. It is important to first research what is already possible in current MES tooling, how it uses MES data and what can be the added value. Also, there is a great unused potential in MES data. This unused potential of the MES data is researched in the opportunities part of the research where advanced data analytics are explored. Third, the aim is to extract information, this generally means that the data needs to be cleaned, structured, analyzed and/or interpreted in order to translate it from data to information. Fourth, it is explored how the information can contribute to the manufacturing operations management and with that generating added value for a company.

1.4. Research questions

In order to establish the research goal, several research questions are defined. First, the current situation is assessed by researching the current MES informational abilities and its attributes. Second, the opportunities for MES data that arise from big data and data analysis tools used in literature in the manufacturing industry will be researched. In both steps, a MES expert's opinion is important for a ranking of the information. Finally, a case study will provide a proof-of-concept for several current possibilities and found opportunities.

In order to structure these steps, the following research questions are defined:

Research Question 1.	**What information can be derived from MES for manufacturing operations management purposes?**
Research Question 2.	**What other information relevant for manufacturing operations management can be derived from the MES database by making use of knowledge discovery?**
Research Question 3.	**How can we derive this information in a real world situation?**
Research Question 4.	**What relevant insight are provided and what challenges can be encountered in a real world situation?**

In the next section the theoretical framework will be explained and the research questions will be mapped on the research steps. Also the tasks corresponding to each step or question will be explained.

1.5. Research approach

The methodology used is based on the Design Science Research Process (DSRP) of Peffers et al. (2006) which is created for design science in information systems research. In their paper Peffers et al. (2006) created a model for design science in the field of information systems that is consistent with prior research and practice, provides a nominal process sequence for the execution of the research and provides a mental model for how the research's output should look like. The research process with corresponding research questions for this research exists out of three phases and is represented in Figure 1.

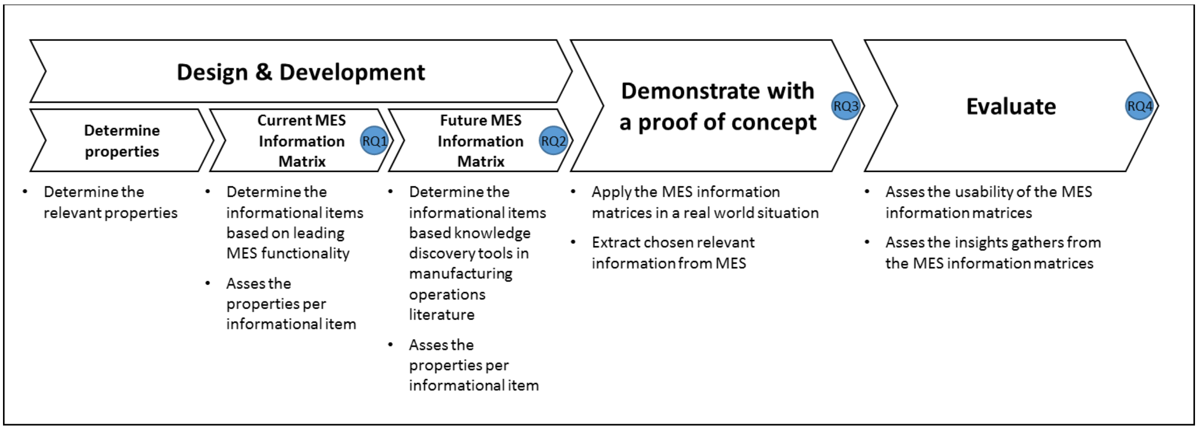


Figure 1 Sequential research process and corresponding moments of the research questions

Per research phase, the approach is discussed separately in the next section.

1.5.1. Design & Development

The aim of this phase is to create two frameworks that provide a structured, complete and comprehensive overview of all information that can be extracted from MES by making use of either current leading MES functionality or knowledge discovery tools. The information elements identified are referred to as the informational elements of the framework. For each informational element, a set of properties is given. The framework itself is referred to as the Current/Future Informational Matrix for the current MES abilities or future possibilities. The establishment of the Informational Matrices consists out of three subsequent steps which are discussed separately.

1.5.1.1. Determine properties

The properties defined for both MES information Matrices are presented below. Some properties are only relevant for one of both matrices.

- Operational area of MES implementation: As the operational processes in different operational areas of a production facility are different, the information to extract differs as well.
- Field expert's ranking: Field experts are asked questions about the informational elements in order to rank them based on (predicted) use which indicates the value of that informational element for practitioners. Field experts are people actively working with MES or a MES related organization or those who are interested in the field of MES.
- Degree of standardization: In the current MES environments the degree of customization is relatively high which could harm the generalizability of the matrix. Therefore, the degree of standardization is incorporated as a generalizability indication.
- Time horizon: Information from the MES can have implications about the past, present or future.
- Source: The origin of the source is provided as there might be differences between theoretical and practical sources of MES.
- Method used: Information is extracted by making use of knowledge discovery tools is non-typical therefore the methods used are mentioned.
- The year of publication: The publication year for Informational items found in literature could indicate a research focus in time and an indication for the time to market of an information item.

1.5.1.2. Current MES information Matrix

First, the informational items from current leading MES functionality are determined based on MES sources. These sources need to represent both the leading theoretical and leading practical perspectives for completeness reasons. The theoretical sources are organizations defining MES by research and by creating conceptual standards; these are the ISA-95 and MESA. The practical sources are leading MES vendors as they implemented MES informational abilities into their based on actual customer demand who use it in practice.

Second, the properties for each information items are assessed. For each relevant property the method is provided

- Operations area of MES implementation: Some theoretic sources already provide this information. For other theoretic sources, as well as the practical sources the area is determined based on the explanation provided about the informational item.
- Field expert's ranking. An online survey about the informational item is held among MES experts in MES related LinkedIn discussion groups. The following information is gathered:

    - General information for the classification of the respondents (relation to MES, country, age)
    - General questions about their knowledge of MES standards, functionality and operational areas.
    - Questions about which MES information items are most frequently used. It is assumed that the most frequent used information is the most valuable.
- Degree of standardization: A recent study of Iskamp and Snoeij (2015) has researched the degree of standardization of data collection and performance analysis in MES functionality per operational area. This is research is used to assess the degree of standardization.
- Time horizon: For each informational item the time horizon of is determined based on the description of the information item.
- Source: The source indicating the information item incorporated.

1.5.1.3. Future MES information Matrix

First the information items from knowledge discovery tools in MES related, manufacturing operations areas are determined by a literature review. This literature review takes both knowledge discovery (which makes use of data mining), and data mining alone into account. As a literature review provides a wide range of informational items, the items are grouped in main- and sub-groups with similar information uncovered.

Second, the properties for each information items are assessed. For each relevant property the method is provided

- Operations area of MES: Based on background of the process analyzed and data used in the literature articles.
- Field expert's ranking: A second online survey is held among the same target group as the first online survey. The set-up of the survey is also similar only different informational items are asked.
- Method used: The methods in literature are presented in the corresponding informational item group.
- Year of publication: The year of publication of the literature is presented in the corresponding informational item group.

1.5.2. Demonstrate with a case study

In order to demonstrate the applicability and the usability of the MES information matrices, they will be assessed in a real world situation. A partnership is established with a company that uses an operating MES.

The applicability is assessed by demonstrating informational items based on the partner company's MES. For this case study a root cause analysis and other small informational items. Both current (MES) analysis methods and knowledge discovery methods are used for this. The usability is assessed by discussing the MES information matrices with the company and discussing how the information items relate to a real world environment. The possible value of the informational items is discussed as well as the challenges. Also the presentation of the information is discussed.

1.5.3. Evaluate

As a conclusion of this research the MES Information Matrices are evaluated with the partner company on the applicability, usability and the possible value created.

1.6. Scope of the research

The focus of this research is given in Table 1 where per item, the aspects in and out of scope are addressed. Note that some of the aspects that are considered out of scope are shortly addressed in the background section of this research in section 2 for clarification reasons.

Table 1 Overview of the scope of this research

What	In Scope	Out of Scope
Industry	Manufacturing	Other industries where MES could be implemented
Type of Manufacturing Industry	Discrete Manufacturing	Other types of manufacturing industries (for example the process industry)
Information system	MES	Other information system. As well other level systems like enterprise planning systems of control layer systems, as systems comparable to MES
MES functionality	Data acquisition and performance analysis	Other MES functionally (for example functionality aimed at facilitating the actual production)
Information	Relevant to manufacturing operations management	Relevant to other areas or non-relevant information

1.7. Contribution of the research

This research provides an overview of the possibilities and opportunities of MES which is currently not present in literature and practice. With this, the research relates the MES systems data to information that can provide insights that help effectively improving the manufacturing operations management.

Current research about MES in literature is focused on the functionality of MES and how the IT structure with other systems can be integrated. There is no literature that provides a consistent and comprehensive overview of the informational abilities of MES and no research that relates business insights to MES abilities. This indicates an opportunity especially when comparing this to other widely implemented information systems, like for example enterprise resource planning systems for which there are ample researches. Additionally, in practice a comprehensive overview is also not present. There is a leading standard for MES functionality but this is a conceptual standard focused at functionality. Also, MES vendors tend to emphasize that every company with MES is different and that they can custom-build almost any client demand if necessary. This research is positioned in that gap and does provide a comprehensive overview of MES informational abilities which can be linked to business insights.

1.8. Structure of the document

The document starts with a background section on MES. Next, the Current MES Information Matrix is established in Chapter 3. Then, in Chapter 4, the establishment of the Future MES Informational Matrix is presented. The demonstration with a case study is provided in Chapter 5. Finally the research ends in Chapter 6 with the conclusion, limitations and further research.

2. Background section

In this section more information about MES is provided as background of this research. Information for the background section is gathered from MES related organizations, literature and websites. The MES vendor's websites have been researched as well. Also several calls, meetings and email conversations have taken place with people who know much about MES or surrounding system. An overview of this can be found in Appendix 1.

In this chapter a short introduction in MES is provided first and how two associations have tried to standardize MES with their reference models. In practice MES vendors have adapted these standards to some extent though they differentiate from this standard and/or each other as well. Reasons for this are elaborated on. Also, the interaction of MES with other manufacturing systems is discussed as well as the data present in MES. Finally, a small insight in the global MES market is provided.

2.1. The MES

As stated in the introduction, MES are information systems that focus on the actual operations management or execution of a production plant/process while operating between the enterprise planning software and the floor control software. Before MES was formally called MES, systems that considered shop floor management where already present. Because there was a need for a more formal description of MES and its functions two organizations played a major role in formalizing it: MESA International and the International Society of Automation (ISA). MESA International focusses on formalizing the core functions of MES. ISA also describes this, but focusses on the cooperation and communication of MES functions with each other and with other systems layers their formalized standard the ANSI/ISA95. The two organizations work closely together in their research. The two organizations and their models are described in the following sections.

2.1.1. MESA (Manufacturing Enterprise Solutions Association)

MESA was originally established as the Manufacturing Execution System Association but when the ANSI/ISA95 gained popularity with the term MOM and the need for a broader definition of MES rose, MESA changed its name the Manufacturing Enterprise Solutions Association. Currently MESA usually refers to the combined term "MES/MOM" when talking about MES(-like) software.

MESA conducts research in order to improve business results and production operations. They created the MESA model, which reflects to the research areas of MESA and its strategic objectives. They define five strategic objectives of their research areas in their most recent model of 2008:

1. Asset performance management (APM)
2. Lean manufacturing
3. Quality and regulatory compliance
4. Product lifecycle management (PLM)
5. Real-time enterprise

The sixth strategic initiative is defined "Additional initiatives" which consist of all of the subjects that do not fit in the five main areas. The graphical MESA model shows which business operations are important for research, as well as which manufacturing or production operations.

In a previous model in 1997 MESA defined the MESA-11 model, which is shown in Figure 2. In this model the 11 core functions of a MES can be found. In their published model, the relationships to external enterprise systems and functional areas are also described.

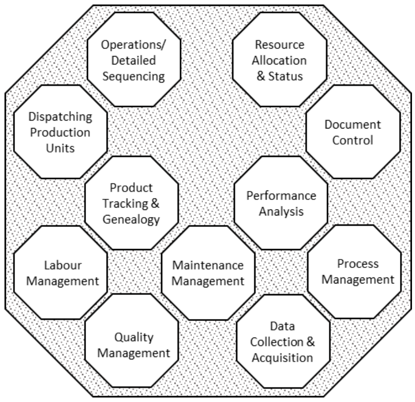


Figure 2 MESA-11 model, established in 1997 with the 11 core functions of MES (MESA International)

The eleven core tasks of MES are defined by MESA in their whitepaper "MES Functionalities and MRP to MES Data Flow Possibilities" (1997). Though this whitepaper is not publically available, the book Manufacturing Execution Systems – MES by Kletti (2007) explained the eleven core functions of MES based upon the original whitepaper as:

1. Operation/ detailed Sequencing
   Sequence and time optimization of the orders finely tuned to the performance of the machines including their finite capacity and to other resources
2. Resource Allocation & Status
   Management and monitoring of resources, such as machines, tools, and so on. Also, registration and display of the current status of resources
3. Dispatching Production Units
   Management of the input materials and intermediate products used in production, this in some cases being for the purpose of documenting material consumers.
4. Document Control
   Management and distribution of product, process, design or order information as well as work instructions which help secure quality.
5. Product Tracking & Genealogy
   Documentation of all events connected with the creation of a product. Recording details of the input materials and ambient conditions.
6. Performance Analysis
   Comparison and evaluation of measured and recorded actual values for installations or areas against operational targets, customer targets, etc.
7. Labor Management
   Control and definition of operations and dispatching to work centers and personnel.
8. Maintenance Management
   Planning and implementation of suitable measures aimed at enabling machines and installations to meet their performance targets.
9. Process Management
   Control and management of the workflow in a production facility in accordance with the planned and current loads and specifications.
10. Quality Management
    Recording, tracking and analysis of the product and process, and verification against ideal values.

11. Data Collection & Acquisition
    Visualization, recording, collection and organization of process data, of material and raw materials, of personnel handling, of machine functions and their control.

All of these function groups, or a reasonable combination of them, can form a total MES solution (Kletti, 2007). In 2004 MESA published a new model for MES, the c-MES. The name c-MES stands for Collaborative MES. The defined eleven functions are redefined or merged into eight main functions. Additionally the collaboration with other enterprise systems is again defined and now also shown in the figure in Figure 3.

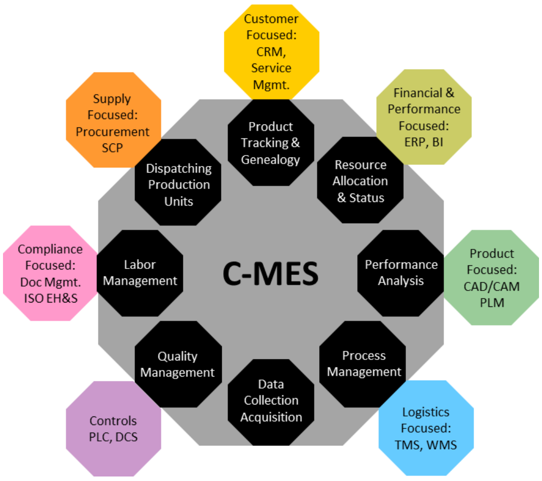


Figure 3 c-MEs model, established by MESA in 2004 with the 11 core functions of MES and how they co-operate with other systems. Source: MESA International

In c-MES, MES is again defined as a connection between automation and corporate management. However, now MES is now also defined as a data and information hub. In other words, MES is defined as an integration platform within the manufacturing company. Though this might be a very broad interpretation, the role of MES for the overall company objectives does become clear.

When looking at the practical side, early MES were on-site applications that solely represented the current as-is process. This had some drawbacks as they were somewhat isolated. Also, they were typically rigid and required a high initial investment both in terms of coding and for on-site hardware (Manufacturing, 2013). This made it a risky investment for some clients.

As MES kept evolving it became more flexible. MES did not only create on-line web-based applications, but it also became more modular so that the client could choose which functionality they needed the most. As stated above, any combination of MEs functions could be a MES.

MES evolved even more and gained functionality outside of the 'Execution' domain. With the launch of the c-MES model, a need for a broader definition raised. This is when the term MOM (Manufacturing Operations Management) was defined by the International Society of Automation (ISA). ISA created an industry standard to define the functional hierarchy of a manufacturing environment in terms of functions, activities and systems, the ISA-95. Later in this chapter, a more detailed description of ISA-95 is provided.

MESA however, also kept evolving and launched a new model in 2008. In this new model, shown in Figure 4, the focus is no longer solely on MES. The new model ranges from enterprise level's strategic initiatives, to business operations, to manufacturing/production operations (plant operations), to the actual manufacturing/production. The model shows how the interrelationship between the levels and shows how events trigger other events that lead to information. The Manufacturing/Production Operations layer describes possible functions of a MES. When looking at the 10 described functions, they are similar to previous defined function of MES in earlier MESA models.

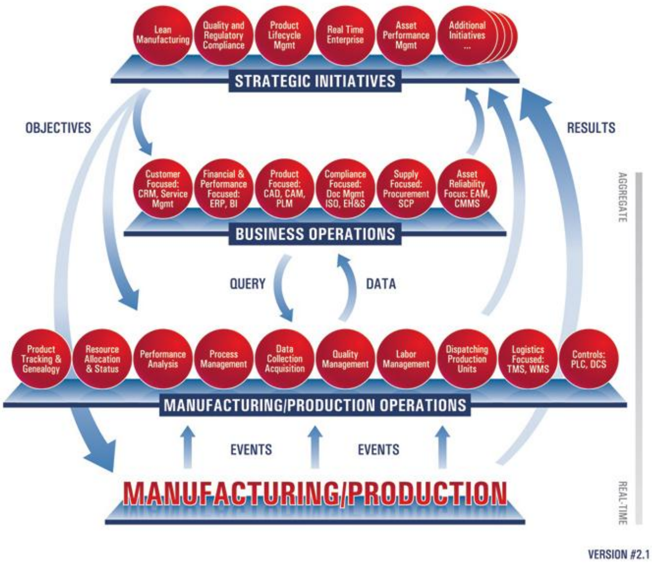


Figure 4 MESA model of 2008. Source: MESA International

2.1.2. ISA (International Society of Automation)

The International Society of Automation is “a nonprofit professional association that sets the standard for those who apply engineering and technology to improve the management, safety, and cybersecurity of modern automation and control systems used across industry and critical infrastructure” (ISA). ISA sets standards, conducts research and provides training and education for industrial automation.

A more formalized and structured hierarchy of how all systems should be integrated and which function belongs to which system level and how they should interact, is formalized by the International Society of Automation in the ANSI/ISA-95.00.02.2013(IES 62264-3 Modified), referred to as the ISA-95.

The ISA-95 is an international standard for the integration of enterprise and control systems. The standard is developed for the global manufacturing industry and can be applied in all types of processes and industries. ISA-95 sets a conceptual foundation for the terminology and communication between the systems in the different functional levels. With this, ISA-95 is the most successful in the Industry with its standards for the vertical integration, though it also sets guidelines for horizontal integration. ISA also works closely together with MESA to keep up to date with new research and developments. The ISA-95 terminology and models have an academic and conceptual character. Many vendors use the basis of ISA-95 but implement variations around this basis as well. Furthermore, clients also often ask for custom made alterations for their specific processes.

For the vertical integration, ISA-95 refers to a functional hierarchy model with five levels. Each level provides different functions and work with different timeframes. The hierarchy model is presented in Figure 5.

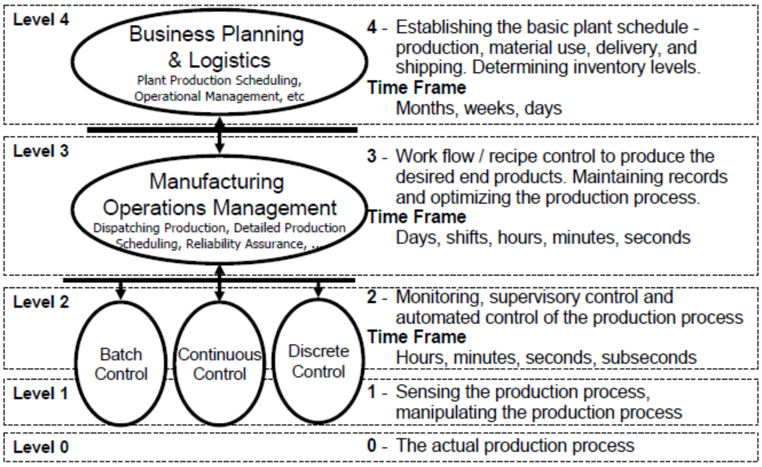


Figure 5 ISA-95 Functional hierarchy model with examples of corresponding system types. Source: ANSI/ISA95.00.01-2010

In this standard the different levels display different business functions. These levels could be linked to different systems types as well, though ISA-95 does not explicitly mention system types as it is a conceptual guideline.

- Level 0 defines the actual physical processes, like the physical production process on a machine.
- Level 1 defines the activities related to the sensing and manipulating the physical production process of level 0. The timeframe of this level is typically seconds or faster. In term of systems, this can be sensors to sense the process or PLCs to manipulate the process based on programmed rules, usually in an "If-Then-Else" format.
- Level 2 defines the activities related to the monitoring and supervisory and automated controlling of the physical process. The timeframe of this level can hours, minutes, seconds and/or sub-seconds. In terms of systems, this level is can be for example a SCADA system.
- Level 3 defines the activities related to the work flow or recipe control of the production steps that reach the desired end-product. This includes coordination, maintaining records and optimization. This level is called Manufacturing Operations Management (MOM). The timeframe of this step is typically days, shifts, hours, minutes and seconds. In terms of systems this level can be for example a MES.
- Level 4 defines the business related activities that are needed to manage the (manufacturing) organization. This includes scheduling production, materials, employee etc. as well as determining inventory levels. The information from level 3 is critical for level 4 to function. The timeframe of this level is typically months, weeks and days. In terms of systems this can be for example an ERP system.

Concluding, MES operates between the business systems and the production control systems. In formalized terms, by ISA, MES operates in the Manufacturing Operations Management level.

In the ISA-95, level 3 is described as the Management Operations Management (MOM) layer. In this layer, typically a MES could be implemented. Before focusing on the MES, a more detailed description of MOM will be provided.

MOM is defined in the ISA-95 and describes the activities and business processes in level 3 of the ISA-95 architecture. In the MOM model, four main operation management areas are defined:

- Production operations management, associated with production control and partly production scheduling.
- Quality operations management, associated with quality assurance.
- Maintenance operations management, associated with maintenance management.
- Inventory management, associated with partly material and energy control and partly product inventory control.

Manufacturing Execution Systems (MES) is one possible system in the MOM level of the ANSI/ISA-95. MOM is has a broad definition in which MES is one of the most common systems. However some parts of MOM can also be executed by other systems, as MES can be different combinations of different functions.

Thought both the theoretical model of MESA and the ISA-95 framework provide theoretical guideless, the implementation in practice can vary. The next section will provide information about this.

2.2. MES implementation in practice

The practical implementation of a MES can vary from the theoretical guidelines as described in section 2.1. This has three main reasons. First, all implementations are different because every manufacturing organization and its processes, systems and people are different. Second, because MES needs to be integrated with the other (manufacturing) systems present with the specific manufacturing organization. Third, due to the origin of the MES vendor because the MES offered itself can vary between the MES vendors as well. These three points will be discussed in more detail in the next sections.

2.2.1. Differences due to variation of a client's MES choices

When implementing a MES there are some important variables to consider. First, the functions and combination of MES can be seen as building blocks, from which the user can freely choose. Second, the areas in which MES is implemented and the configuration to other systems is also an important user's choice. Third, the systems already present in the organization, on which MES has to function together with is important to consider due to amount of configuration needed for them to interact. Fourth, the degree of customization the client wishes for its MES product will play an important role. This degree of customization tends to be relatively high in the MES market. Figure 6 and Figure 7 are from the annual MES Product Survey by Iskamp and Snoeij (2015) and present a graphical overview of these building blocks the customization for a template version of MES and an actual implemented version.

Figure 6 Template with building blocks for MES core. (Ipskamp & Snoeij, 2015)

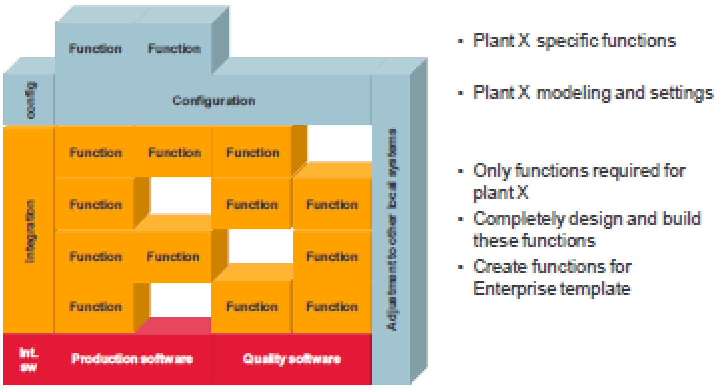


Figure 7 MES architecture for an example plant (X). (Ipskamp & Snoeij, 2015)

2.2.2. Differences due to integration with other manufacturing systems

In the total manufacturing environment, many systems operate together in order to produce the desired end products within a certain time frame. At enterprise level, production gets planned and in the factory the actual production takes place. When looking into these systems many systems abbreviations will arise in different layers of the organization. At enterprise level ERP (Enterprise Resource Planning), CRM (Customer Relationship Management) or PLM (Product Lifecycle Management) systems are frequently used. In the plant however MES, HMI (Human Machine Interface) or SCADA (Supervisory Control And Data Acquisition) systems are very common. This might be very confusing at first but it is clear that these systems all co-exist, have their own added value and need to be integrated in order to work together. Figure 8 gives a graphical overview of how a few common systems interact on the axis of Business versus Production and Suppliers versus Customer side. It is clear that the amount of other systems, the type of other systems and the implantation of the other systems affect how MES needs to be configured in order to cooperate with them.

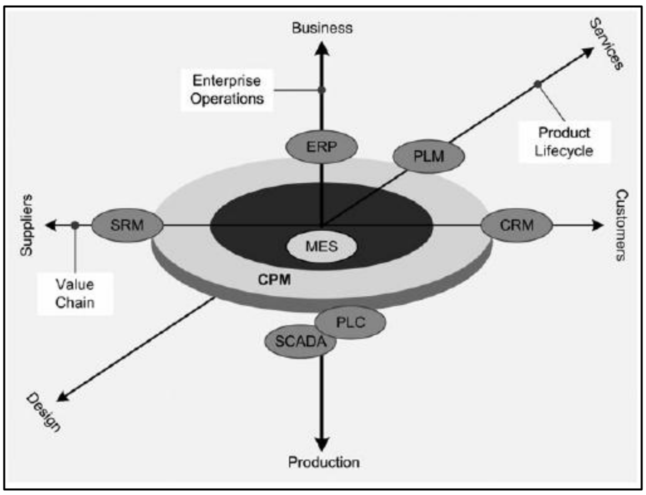


Figure 8 Scope of CPM in the context of adjacent systems. Source: (Meyer, Fuchs, & Thiel, 2009)

2.2.3. Differences due to the origin of the MES vendor

MES where raised from two different system types. On the one hand there where enterprise software providers who wanted to get more detailed information about what is happening in production. On the other hand, there were production control software providers who found that adding functionality opened many possibilities. Currently there are many MES providers but this distinction in origin can still be found in several MES. The major difference is usually the point of view the MES takes. Some take the point of view of machines, where products run through, while others take the point of view of products which run through machines. Though the outcomes are similar, the data structures and way of working can be very different.

Even though there can be differences in the implementation of the MES, all MES work with data. Therefore, more information about the data stored in the MES is provided in the next section.

2.3. Data stored in MES

As MES operates between level 2 and level 4 software, both aspects are present within MES in terms of data. From a top level point of view this data and information can be: production orders, serial numbers, bill of materials, routing, work instructions, inventory locations and many more. From a floor level point of view this can be machine status, product defects, operator ID and many more.

To fully understand what happens with this data, the data structure and basics of all three levels will be explained shortly:

- In an ERP system, the production data consists of object usually linked to a type of order or command which is sent to the MES. For example a production or for a certain amount of items of a specific product scheduled at a specific time.
- In a SCADA system the PLC's of the machines constantly sent messages of a certain variable. This can be as often as multiple times per second to monitor what is happening in a real-time manner. The data in SCADA is 'flat' as it can be compared to a list of data entries that describe a current behavior.
- MES combines both data parts and translates them to MES data. The data can be seen as multi-dimensional as the objects have multiple attributes and data object which can be part of a greater hierarchy. A machine for example will compose of several parts, data objects and variables as well as the products, quality control and other objects do. Additionally the database of MES is updated in an event based manner. Not like in SCADA it is constantly updating, but it only records event logs when changes trigger the system.

Essentially for MES there are two types of data collection. First, data can be collected through the connection with the control or the enterprise layer. Second data can be collected by the MES itself. This can be done automatically by installing an automatic sensor or scanner. However this can also be done manually by letting the operator scan materials or by making the operation enter values into the MES manually.

The data in MES has a XML (Extensible Markup Language) structure if it relates to the ISA-95 standard. XML is a markup language that defines a set of rules for encoding documents while being readable for both computers/machines and humans. The data models used in ISA-95 are represented in the by ISA-95 presented B2MML (Business To Manufacturing Markup Language) which is an XML implementation. In practice 61% of the MES support B2MML which makes it the second largest language after OPC-UA with 65% according to the MES annual product survey.

MES captures the data by itself within the MES database or uses a historian. The MES database where this data is stored in are mostly MS SQL servers or Oracle RDBMS. According to the Annual

MES Product Survey 2015, 84% of the MES supports MS SQL and 76% Oracle RDBMS. Both use a SQL-type language to communicate. MS SQL uses Transactional SQL where Oracle RDBMS uses Procedural Language SQL. These languages have many similarities. The main difference between the two languages is how they handle variables, stored procedures, and built-in functions (Stansfield, 2014). Ideally there would be one big MES database for an entire plant or subject. However, various databases are often found within "one" MES (Meyer, Fuchs, & Thiel, 2009). This makes can make it hard to find the desired data or a standard database structure. Historians are interfaced with MES products when large volumes of data need to be stored that exceed the MES database capabilities. Additionally when data needs to be collected from outside the MES product, historians are used as well.

As all data entries contain several data objects or attributes and always a time stamp, they can be referred to as event logs. Event logs are structured pieces of information that most information systems store during operation. An event log typically contains information about events referring to a user, a timestamp and a case (Van der Aalst, 2005). In ISA-95 these data entries, or event logs, are also described on a conceptual basis. ISA-95 advices that all data is structured with several tags like ID, start time, end time, value, category, description etc.

Aside from data entries, also information about the production execution is communicated between MES and the ERP level. This is information like a bill of material, work instruction and other. This is out of scope for this research.

In Figure 9 a graphical representation of the data entries in MES is provided.

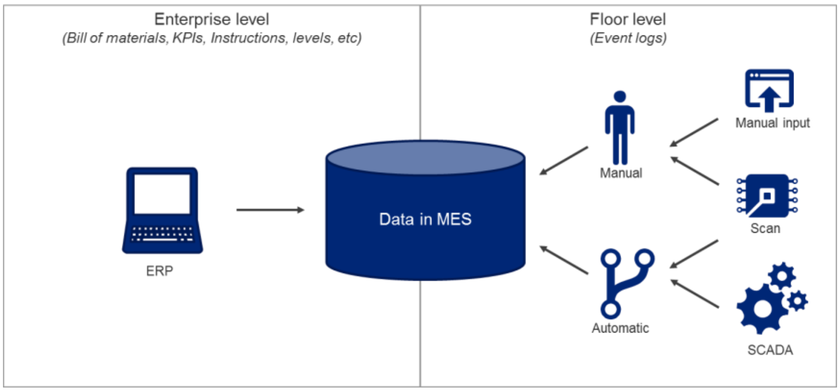


Figure 9 Representation of data in MES

Like in all systems, data quality is highly relevant for MES. The impact of data quality on the information chain has been widely recognized since the onset of large-scale data processing (Sadiq, 2013). Low quality data is more time consuming to analyze, as the data has to be cleaned first. Additionally, low quality data will provide low quality insights. Finally low data quality will prevent algorithms from working properly. In literature the importance of data quality is described as well. Data quality is important to organizations because that it impacts customer satisfaction, operational costs, effectiveness of decision making, and strategy creation and execution (Redman, 1998).

On order to position this background information about MES in the industry and to know which key players are active in the market, more information about the global MES market is provided in the next section.

2.4. MES global market

As stated in the introduction, the global MES market is a billion dollar market that is still rapidly growing.

Historically, MES is well implemented the process industries like pharmaceutical and food where MES realized the traceability needs imposed by the authorities (Saenz de Ugartea, Artibab, & Pellerina, 2009). However MES is currently implemented in many industries, even outside of manufacturing. In 2014, the process industry was still the largest industry for the MES market with a market share of around 56 (Markets And Markets, 2015a).

The total number of MES vendors and MES products in the global MES market is unknown yet some estimates indicate that globally more than 300 MES products might exist (Ipskamp & Snoeij, 2015). The major players in the manufacturing execution systems market are represented in Table 2.

Table 2 Biggest companies in the global MES market (Markets And Markets, 2015a)

Company	Country of origin
ABB Ltd.	Switzerland
Andea Solutions	Poland
Dassault Systemes SA	France
Emerson Electric Co.	USA
General Electric Co.	USA
Honeywell International Inc.	USA
Rockwell Automation, Inc.	USA
SAP AG	Germany
Schneider Electric SE	France
Siemens AG	Germany
Werum IT Solutions GmbH	Germany

3. Establishment of the Current MES Informational Matrix

In this section the current analysis tools of a MES is assessed in order to answer the following research question:

Research Question 1: What information can be derived from MES and is considered useful?

The research question is answered by establishing the Current MES Informational Matrix which provides a structured, complete and comprehensive overview of the information that can be derived from MES and is considered useful.

The information is extracted from MES data by analyzing it. In MES, this is usually referred to as performance analysis functionality. First, a short introduction is provided on what performance analysis in MES means from a theoretical point a view in MES. Second, the current performance analysis tools are investigated by which information they uncover, these are the informational item. This is investigated by assessing two leading theoretical sources and leading practical sources. Third, the values of the information items properties, as defined in the research approach, are determined per informational item. This is assessed by researching the informational items and by a survey held among MES experts. Last, the values are mapped and the Current MES Informational Matrix is established and presented.

3.1. Performance analysis in MES from a theoretic point of view

In the two leading MES books the following definitions of performance analysis are presented:

- From the manufactured sizes to down time, disruptions, piece counters, etc., managerial key figures are produced promptly, in real time, if feasible, in order to allow for simple assessment of production efficiency, detection of problems, etc. Display in various diagram formats is made available to the user (Meyer, Fuchs, & Thiel, 2009).
- Comparison and evaluation of measured and recorded actual values for installations or areas against operational targets, customer targets, etc. (Kletti, 2007)"

What comes back in both definitions is that reporting by a MES is necessary to assess the performance. In these reports, certain defined metrics will be assessed on. A MES will usually provide a couple of 'standard' metrics which will be combined with customer specific metric during implementation.

What is notable of the definition of Kletti is the comparison of actual data with pre-defined operational targets. Key Performance Indicators (KPIs) are becoming ever more important for assessing the competitiveness of production companies relative to others from around the world. MESA has conducted a research about 'Metric That Matter' in manufacturing in which they defined 18 important manufacturing KPIs. Naturally, more KPIs are used within manufacturing, but these are important to consider.

The reports that MES generate can be standard reports, usually generated automatically, or ad-hoc reports, generated manually. Automatic reports are usually able to generate real-time results and insights. Some ad-hoc reports can also be real time but usually, ad hoc reporting is conducted offline. This means that data is exported and will be analyzed by making use of analysis tooling. Additionally, the MES database can also be connected to a separate Business Intelligence (BI) Engine to generate more in depth reports. This BI Engine is usually not a standard feature in MES but a separate module or server. Many MES vendors offer this module or service separate on their websites.

3.2. Informational items from MES provided by leading MES functionality

The informational items derived by MES provided by leading current MES functionality represent the information that should be currently available for MES users, given they implemented their system close to the standards. This is researched by combining multiple leading sources:

- ANSI/ISA-95.00.02.2013(IES 62264-3 Modified): In this framework the possible areas of Manufacturing Operations Management, in which MES operates, are described. These areas are Production management, Quality management, Maintenance management and Inventory management. Additionally, performance analysis in each area is described. As ISA-95 is an industry standard, this will be the basis of the framework.
- MESA International: The research of MESA about Metric that matter provided a funded background on important KPIs in manufacturing. As MESA is leading in MES research, these metrics can be important for MES as well. As in MES some metrics are pre-installed while others are configured during implementation, it is reasonable to add the manufacturing operations specific metrics of the metric that matter report in the framework. The metrics that are out of scope for the plant floor, are not included. The full list of the Metrics that matter is included in Appendix 2.
- MES Vendors: The ten biggest MES vendors according to the research of Markets And Markets (2015) are assessed to check what they sell as the main performance analysis tools. This is checked, by exploring the vendors' websites. Not all analyses possibilities will be on the website, therefore the results of the ten biggest vendors is combined. Also, it is not the case that MES vendors cannot provide the information they do not mention directly on their website. The list of vendors in included in Table 3.

Table 3 List of top ten MES vendors

Company	MES product
ABB Ltd	ABB MES
Andea	-
Dassault Systemes	-
Emerson Electric	Syncade
General Electric	Proficiancy
Honeywell International	Intuition
Rockwell Automation	FactoryTalk & MES
SAP AG	SAP ME
Schneider Elextric SE	Wonderware MES
Siemens AG	MES Simatic
Werum	PAS-X

3.3. Determining the informational items' properties

In this section, the values for each informational item's properties are assessed. The actual values per informational item can be found in section 3.4. , but the additional information is assessed in this section.

3.3.1. Operational area of MES implementation

According to the leading standard in MES, the ISA-95, MES is implemented in the Manufacturing Operations Management area in (a combination of the) four areas:

1. Production Operations Management
2. Quality Operations Management
3. Maintenance Operations Management
4. Inventory Operations Management

For each informational item, the operational area is determined either by the source, or derived from the explanation of the informational item by the sources.

3.3.2. Field expert's ranking

In order to assess the field experts ranking of the informational item, a survey was held. The goal of the survey was to discover to what extent MES experts believe that the informational items are used in practice and would be useful to improve the manufacturing operations management. The survey was held as an online questionnaire in MES related LinkedIn discussion groups. It is assumed that people active in these groups are MES experts. The questionnaire was open from December 24th 2015 to March 13th 2016. During this period of time, 54 respondents filled in the questionnaire.

First, classification questions are asked in order to understand the background of the respondents. The relation to MES, age and country were asked. Also it is asked in which Manufacturing Operations Management area MES is used in the most (multiple answers possible). The answers are summarized in Figure 10. It can be found that most respondents are (working for) MES vendors and between 30 and 40 years old. There is much variation in the countries with India, The Netherlands and USA as largest groups. As expected, the respondents believe that MES is most used in production management.

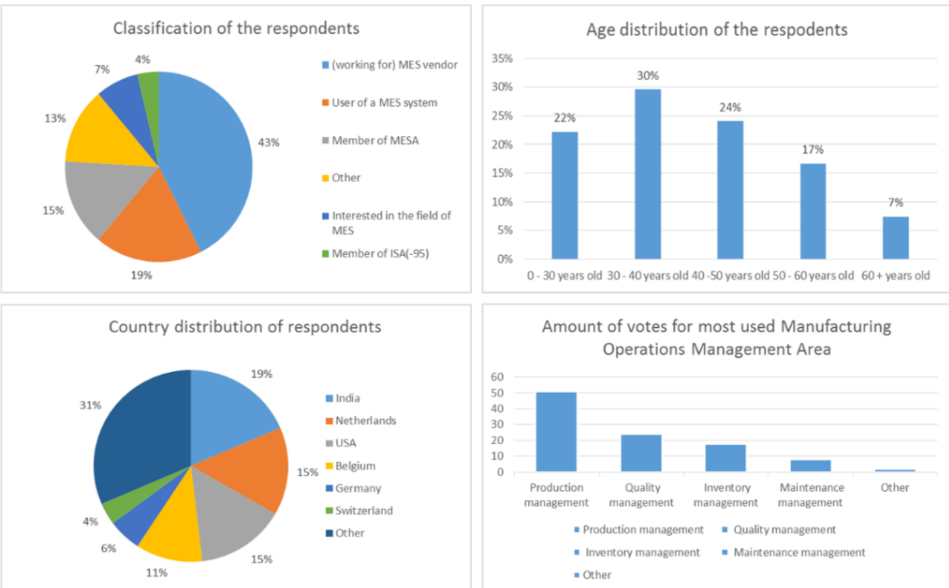


Figure 10 Summary of respondent's background and vote for MOM area

Next, the respondents are asked how familiar they are with some MES related topics in order to validate whether the respondents are MES experts. On a scale of 1 (not familiar) to 5 (very familiar), they are asked to scale the following 6 topics:

1. The ANSI/ISA95

2. The MESA Model
3. MES functionality for Production management
4. MES functionality for Maintenance management
5. MES functionality for Quality management
6. MES functionality for Inventory management

The average score of these six topics was a 4,18 which indicates that the respondents are good familiar with the topic and can be labeled as MES experts. When segmenting this score among the groups some interesting finding can be uncovered. In Figure 11 the segmented results are presented. It can be found that the users of MES have the second least knowledge of the MES related topics. This reflects on the statements of the problem introduction in 1.1 where it was stated that the MES are still unknown in the field. Second, it can be found that some countries, like Switzerland and Germany, have very high scores which implicate relatively low scores in other countries in order to get to the average score. This indicates a difference between countries and knowledge of MES. Last, it is interesting that the older respondents have more knowledge of MES topics then the younger respondents. This could be due to the higher average age in production companies but is interesting to be aware of for the continuity of MES knowledge in the industry.

Group	Average score
(working for) MES vendor	4,27
User of a MES system	3,50
Member of MESA	4,63
Interested in the field of MES	3,42
Member of ISA(-95)	5,00
Other	4,55

Group	Average score
India	3,75
Netherlands	4,46
USA	3,98
Belgium	3,94
Switzerland	4,92
Germany	4,78

Group	Average score
0 - 30 years old	3,76
30 - 40 years old	3,99
40 -50 years old	4,46
50 - 60 years old	4,52
60 + years old	4,50

Figure 11 Average score of MES topics, segmented among the groups

Last, the respondents per MOM area which of the informational items they believed are most useful and would be used most in practice. The amount of times an informational item gets ticketed in this question divided by the total amount of respondents calculated a percentage for each information item. This percentage indicates how useful the information item is according to the MES experts. Comparing these percentages generates a field experts ranking. These score are added to the information item as a property in the Current MES Informational Matrix.

More information about the survey questions can be found in Appendix 6 and results can be found in Appendix 7.

3.3.3. Degree of standardization

The degree of standardization indicated the generalizability of the information among different MES. Performance analysis is the main function of MES that uncovers information from the data. Therefore, this function combined with the data collection is important to investigate when it comes to standardization.

CGI conducts an annual MES survey in which they do research on current MES trends and developments. In this research they also examine whether vendors offer certain MES function with standard 'out-of-the-box' (standard) or 'configurable' (limited configuration needed) functionality or that extensive programming effort is needed. Below the results of each function of MES, according to the ANSI/ISA95, can be found in Figure 12. For the MES Analysis function on average around 64% have out-of-the-box or configurable functionality. For Data Collection this amount is significantly higher with on average 81%. Additionally, the amount of out-of-the-box or configurable functionality is highest for production management within MES.

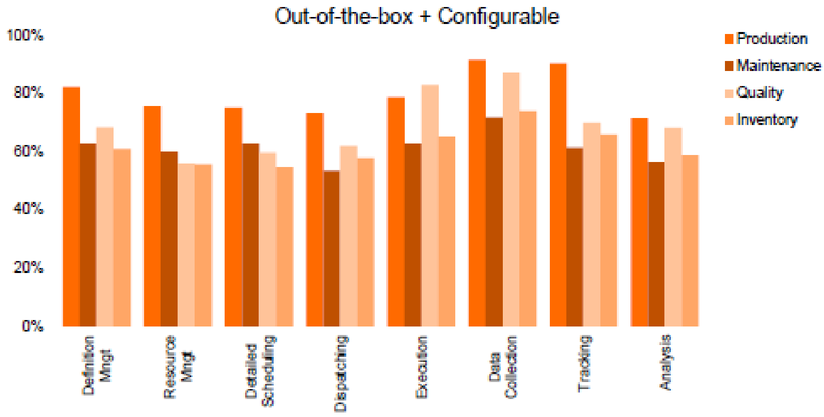


Figure 12 Percentage of out-of-the-box + configurable solutions. Source: (Ipskamp & Snoeij, 2015)

From this, it can be derived that MES functionality is partly standard and partly custom made to client's implantation desire. This is also confirmed by a MES Expert (Ogura, 2015), as he stated that for most his most familiar vendor, from the metrics and KPIs measured by MES, about 40% is standard implemented. This number can rise and be different for specific vendors and client markets, though about 30% will also be custom made metrics.

The percentages derived from the CGI research are used in the Current MES Informational Matrix as a property values per operational area. It is important to mention that the out-of-the-box solutions between different vendors might also differ. What one vendor considers 'standard' might not be standard functionality for another.

3.3.4. Time horizon

The time horizon the informational item assesses can be either the past, the present (also referred to as current status) or the future. The value of this property is derived from the explanations of the informational items provided by the sources.

3.3.5. Source

The source from which the informational item is derived is also provided. This can be the ANSI/ISA95 standard, MESA International's Metrics that matter survey or MES vendors websites.

3.4. The Current MES Informational Matrix

The informational items are listed and the values for each informational item's properties are added. A split is made between the operational areas because these are different processes in the manufacturing environment and because the MES user makes choices in which areas they implement the MES. Also the top three informational items per area is highlighted which will be elaborated on later. The Current MES Informational Matrix will be provided for Production Operations Management first in Table 4, second for Quality Operations Management in. Also the standardization for both data collection and performance analysis is high. This could be due to the fact that MES is mostly used in production operations.

Table 5, Next for Inventory Operations Management in Table 6 and last for Maintence Operations Management in

Table 7. This secuence is chosen because this is the ranking according to experts in which area MES is used most.

Table 4 Current MES Informational Matrix for Production Operations Management

Production Operations Management Degree of Standardization Data Collection 90% - (Performance) Analysis 70%							
		Time Horizon			Source		
Ranking	**Informational Item**	**Past**	**Present**	**Future**	**ISA95**	**MESA**	**Vendor**
78%	**Resource traceability***	**X**			**X**		**X**
76%	**Operational Equipment Efficiency (OEE)**	**X**	**X**			**X**	**X**
67%	**Work In Process (WIP) data**		**X**				**X**
65%	*Real-Time plant and production status*		*X*				*X*
57%	*Equipment/Resource performance*	*X*	*X*				*X*
56%	*Production variability*	*X*	*X*		*X*		*X*
46%	*Schedule or production attainment (time target vs actual)*	*X*	*X*		*X*		
46%	*Equipment/ Resource utilization*	*X*	*X*			*X*	
43%	*Throughput*	*X*	*X*		*X*	*X*	*X*
41%	*Production unit cycle times*	*X*	*X*		*X*		*X*
41%	*Root cause analysis*	*X*					*X*
41%	*Material compatibility & availability*	*X*	*X*		*X*	*X*	
37%	*Weight and dispense support*	*X*	*X*				*X*
30%	*Notification management*	*X*	*X*				*X*
26%	*Personnel tracking*	*X*	*X*				*X*
22%	*Tracking non-productive activities*	*X*					*X*
17%	*Other*						

*Material, equipment, personnel both forward and backward

The list of informational items is largest for production operations management. It can be found that all informational items either reflect on the past or on the current situation. The top three informational items are not surprisingly ranked high. Traceability is one of the most frequent heard uses of MES, as it enables the user to structured recall the information about a product when needed. This can significantly increase recall speed and reduce cost when a defect in of the product in the field occurs. The OEE is also frequently used in manufacturing as is broad measure that takes into account three variables when assessing the efficiency of the equipment. This enables the user to effectively see in which machines could be causing bottle necks. Last, work in process data is a key item for effectively managing the production operations. Also the standardization for both data collection and performance analysis is high. This could be due to the fact that MES is mostly used in production operations.

Table 5 Current MES Informational Matrix for Quality Operations Management

Quality Operations Management							
Degree of Standardization Data Collection 80% - (Performance) Analysis 60%							
		Time Horizon			Source		
Ranking	Informational Item	Past	Present	Future	ISA95	MESA	Vendor
67%	**Quality variability and deviations**	X			X		X
57%	**Yield (analysis)**	x	X			X	
52%	**Batch quality trend analysis**	X					X
46%	*Resource traceability analysis*	X			X		X
41%	*Quality indicator analysis*	X			X		X
35%	*Quality department/operations cycle times*	X	X		X		
31%	*Quality equipment utilization*	X	X		X		
24%	*Quality resource utilization*	X	X		X		
7%	*Other*						

It can be found that again the time horizon for all items is past of present oriented. Additionally, it can be found that for quality operations management the top three informational items revolve around monitoring the quality of the products. Information about the stability of the product quality, the yield percentage and the quality of product batches for batch analyses. It is important for a production company to monitoring the quality of products. First, because errored products can cause problems if they are sold to customers, so this has to be prevented. Second, if the error is detected fault products cost money when repairing or rejecting. Therefore, this needs to be minimized by making use of the quality information. The degree of standardization for data capturing is relatively high, this could be due to the fact that errors can be classified and counted or to the fact the MES is second most used in quality operations. The standardization of performance analysis is middle high.

Table 6 Current MES Informational Matrix for Inventory Operations Management

Inventory Operations Management							
Degree of Standardization Data Collection 70% - (Performance) Analysis 53%							
		Time Horizon			Source		
Ranking	Informational Item	Past	Present	Future	ISA95	MESA	Vendor
64%	**Inventory movement analysis**	X	X		X		
54%	**Received material quality and time**	X			X	X	
48%	**Inventory efficiency**	X	X		X	X	
38%	*Inventory waste analysis*	X			X		
28%	*Inventory Resource usage*	X	X		X		
7%	*Other*						

It can be found that again the time horizon for all items is past of present oriented. The top three is more diverse in Inventory operations management. The movements, quality of received materials and efficiency are ranked most important. These informational elements revolve around knowing where the products in the production area are and around quality. Both are very important aspects of manufacturing operations management. The degree of standardization is lower than for production and quality operations but still the data collection is relatively high. Standardization for performance analysis is the lowest among the four MOM areas. For Inventory operations management many companies chose to incorporate inventory related functionality in their ERP system.

Table 7 Current MES Informational Matrix for Maintenance Operations Management

Maintenance Operations Management Degree of Standardization Data Collection 70% - (Performance) Analysis 55%							
		Time Horizon			**Source**		
Ranking	**Informational Item**	**Past**	**Present**	**Future**	**ISA95**	**MESA**	**Vendor**
61%	**Downtime in proportion to operating time**	**X**	**X**			**X**	**X**
59%	**Status equipment and maintenance schedule**	**X**	**X**		**X**		**X**
43%	**Status materials**	**X**	**X**		**X**		**X**
43%	**Percentage planned vs emergency maintenance**	**X**	**X**			**X**	
30%	*Status assets and maintenance schedule*	*X*	*X*		*X*		
13%	*Status maintenance personnel*	*X*	*X*		*X*		*X*
9%	*Other*						

It can be found that again the time horizon for all items is past of present oriented. The top three is a top four of informational items due to a tie in the third rank. The downtime in proportion to the operating time is an important measure when optimizing the manufacturing operations. This is important for companies as downtime costs money and disturbs the solid production process. The status of the schedule and material is key for day-to-day operations and the percentage planned versus emergency maintenance is again important for optimizing the manufacturing operations. The standardization of data collection is relatively high, and for performance analysis only slightly higher than in Inventory Operations. For maintenance some companies also chose a separate maintenance system, this could explain the lowest ranking for area in which MES is used in the most.

4. Establishment of the future MES Informational Matrix

In this section new data analysis tools in manufacturing and especially on MES data are researched. This in order to answer the following sub research question:

Research Question 2: What other information could be derived from the MES database by making use of knowledge discovery?

This research question is answered by establishing the future MES Informational Matrix which provides a structured, complete and comprehensive overview of the information that can be derived from the MES database by making use of knowledge discovery.

The information extracted from the MES database is assessed from the literature point of view for data analysis tools in a manufacturing environment. First, short introduction in MES and knowledge discovery is provided. Second, a literature study is conducted on knowledge discovery tools with a focus for execution or MES related data in a manufacturing environment in order to find all informational items. Third, the values of the information items properties, as defined in the research approach, are determined per informational item. This is assessed by researching the informational items and by a survey held among MES experts. Last, the values are mapped and the Future MES Informational Matrix is established and presented

4.1. Introduction in MES and Knowledge discovery

MES has a very big and potentially rich database with capturing production specific data and information. Also MES itself has already internal analytic tools to provide manufacturing process information to the user. However, these analytics in MES have limitations. In particular, they do not make use of data mining to identify hidden patterns in manufacturing-related data (Gröger, Niedermann, & Mitschand, 2012). In other words, there are many other analytic possibilities with MES data to discovery information that is now unknown or not accurate.

This creates potential value for new data analysis tools like Knowledge Discovery. Knowledge Discovery in Databases (KDD) is “the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data” (Fayyad, Piatetsky-Shapiro, Smyth, & Uthurusamy, 1996). One of the steps of KDD is Data Mining (DM) and therefore KDD makes use of a particular data mining algorithm. Therefore literature on KDD and literature on data mining can be combined.

4.2. Informational items from MES data by making use of knowledge discovery tools

A literature study is conducted in order to find a complete list of informational items from MES by making use of knowledge discovery tools. For the literature study a split has been made between articles before 2009 and articles from 2009 and beyond. This split is made because in 2009, Choudhary et al. published an extensive literature review about of knowledge discovery in manufacturing, based on the type of knowledge. In their literature review, the scope is on knowledge discovery in manufacturing. Though manufacturing is broader than only execution related data and information, they do cover the execution related part of manufacturing extensively. In other words, in their literature review, they cover the scope of this thesis research as well. Therefore, the literature review of Choudhary et al. (2009) is assumed to provide a complete overview of relevant articles before 2009 for this thesis research.

For the literature from 2009 and beyond, a separate literature review is conducted. For this literature review the library search engine of Eindhoven University of Technology “Focus” is used. This search engine provided access to a diverse set of 102 databases which is sufficiently reprehensive for this search. The full list of databases used by “Focus” in added in Appendix 3. For the search in the search engine the following restrictions have been applied:

- Publication date: from 01/01/2009 until 01/01/2016
- Content type: Journal or Journal Article
    - Books are not included because books are usually too broad or will refer to specific articles.
- Language: English
- Keywords: “Manufacturing” always in the abstract combined with:
    - “Data mining” in the abstract
    - “Knowledge discovery” in the abstract
    - “Manufacturing intelligence” in the abstract

This search resulted in articles for the combination with “Data Mining”, 21 articles for the combination with “Knowledge discovery” (excluded Choudhary et al. (2009), and 42 articles for the combination with “Manufacturing Intelligence”. For the long list the top 25 most relevant articles for each search are included. It is assumed that because the top 25 most relevant articles write about similar information to uncover from the data, the list is extensive and complete when addressing the most relevant and common information to be extracted from the data.

When combining the articles from the three search words, 71 articles are found. After deletion of duplicate articles, a long list of 63 articles remains. This long list of articles is converted into a short list by reading the articles abstract and scanning the document. Articles are assigned to the shortlist if the following criteria are met:

- The article is about a Manufacturing environment
- The data used in the article is Execution data related
- The article aims to uncover information
- Uncovered information applies to performance analysis in MES and no other functions of MES.

In Appendix 4, the long list of articles can be found, with a short description of each article and the decision whether or not the article is assigned to the short list. In total, 33 articles are assigned to the short list.

Now all relevant literature within the scope is gathered, all articles are read and for every article the information that is uncovered by the research with the corresponding method used is listed. For the literature review of Choudhary et al. this generates a list of multiple informational items that are uncovered by a range of methods. Some articles of 2009 and beyond also research multiple informational items to uncover. The list of information items is then clustered in groups that research similar information or have similarities in the goal of the information, like supporting decision making. These groups are called: “Main information (purpose) group”. Within these groups, more specific sub-groups are defined where informational items that are very similar are grouped. These sub-groups are called “Information sub-group”. The full overview all articles, their informational items and their corresponding methods, main information (purpose) group and information sub-group can be found in Appendix 5.

4.3. Determining the informational items’ properties

In this section the values for each informational item’s properties are assessed. The actual values per informational item can be found in section 3.4, but the additional information is assessed in this section.

4.3.1. *Operational area of MES implementation*

For each informational item, the operational area is derived from the analyzed process or data in the literature articles.

4.3.2. *Usefulness according to experts*

In order to assess the field experts ranking of the informational items, a second survey was held. The goal of the survey was to discover to what extent MES experts believe that the informational items are expected to be used most practice and would be most useful to improve the manufacturing operations management. The survey was held as an online questionnaire in MES related LinkedIn discussion groups. It is assumed that people active in these groups are MES experts. The questionnaire was open from February 23rd 2016 to March 13th 2016. During this period of time, 21 respondents filled in the questionnaire.

First, classification questions are asked in order to understand the background of the respondents. The relation to MES, age and country were asked. Also the familiarity with big data and data mining related topics was asked on a scale from 1 (not familiar) to 5 (very familiar). The answers are summarized in Figure 13. It can be found that users of MES are the biggest group followed by (working for) MES vendors, who were the biggest group in the first survey. Most respondents were between 30 and 50 years old and most were from Belgium, The Netherlands or USA. The familiarity with big data or data mining related topic is good, as most give this a score of 4. This is important because some familiarity with these topics is convenient.

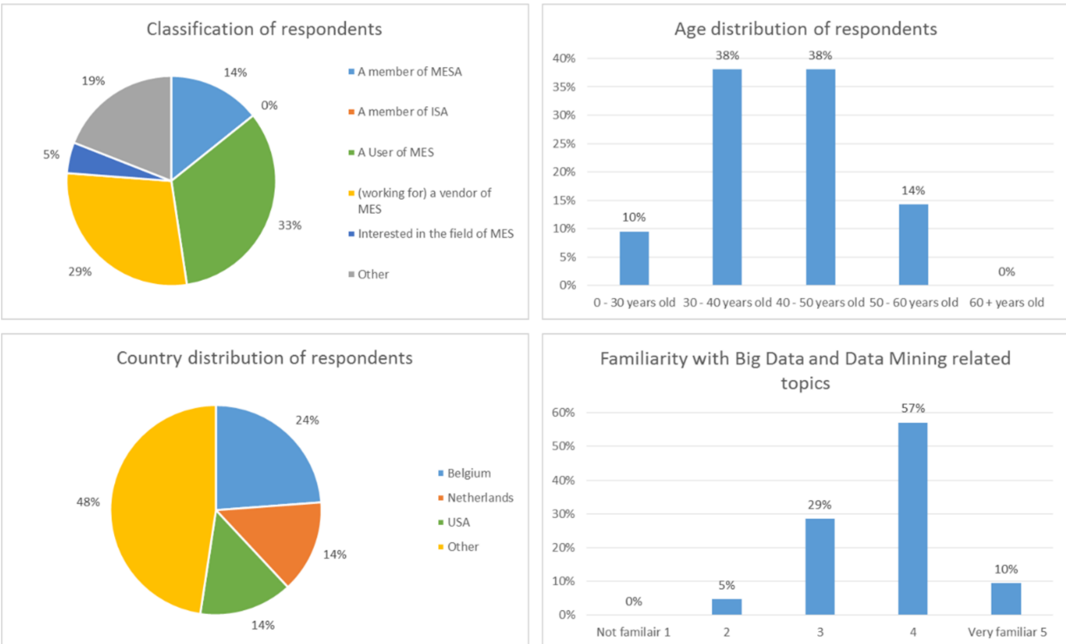


Figure 13 Summary of respondent's background and familiarity with big data and data mining related topics

Next, the respondents are asked to what extend they believe that the main informational items would be used in practice and are considered useful on a scale from 1 to 5. This score is translated to a percentage. Per main group, the respondents could choose which of the sub groups would be the most useful/ relevant. The amount for 'votes' is also translated to a percentage. By scoring the items, the respondents are forced to think about each item individually. Because, the items are not widely available in practice and therefore fast recognized, it is necessary that the respondents take some time to think about each item. The resulted scores are added to the information item as a property in the Current MES Informational Matrix.

More information about the survey questions can be found in Appendix 8 and about the survey answers in Appendix 9.

4.3.3. Method

As knowledge discovery makes use of data mining techniques, the used methods are presented as well. It can be found that a wide range of methods is used.

4.3.4. The year of publication

As the literature review uses another literature review artide of 2009 and separate arti cles of 2009 and later, the year of publication is added. It can be found that there is a wide variety in years over the informational items though some sub-informational item groups were a focus in research in a specific period in time.

4.4. The Future MES Informational Matrix

The informational items are listed and the values for each informational item's properties are added. Again the operational areas are split. The Future MES Informational Matrix is provided for Production Operations Management first in, second for Quality Operations Management and last for Maintence Operations Management in. On Inventory Operations no informational items are found. This can be due to fact that inventory related research with data mining could indude demand patterns analysis. Demand data is usually present in an ERP system and not a MES.

The secuence of the matrices is again chosen because this is the ranking according to experts in which MOM area MES is used in most.

Table 8 Future MES informational matrix for Production Operations Management

Production Operations Management					
%score	**Main informational (purpose) group**	**Sub-Informational Item group**	**% within sub**	**Method**	**Year of publication**
83%	Knowledge of optimal manufacturing settings	Identification of critical process parameters	57%	Regression, Classification, Clustering	2010
		Knowledge of operational process(es)	29%	Association rule mining	2013
				Classification by Neural Networks and decision tree	2012
				Integrated relational databases approach	Before 2009
				data mining with learning classifier	Before 2009
				Two-stage data mining approach	Before 2009
				Decision tree induction, neural network and composite classifier	Before 2009
		Optimization of parameter settings	14%	Range of DM algorithms	2011
				Regression, Classification, Clustering	2010
				Classification (decision tree)	2011
				Set of data mining tools	Before 2009
				Genetic Algorithms and Neural Networks	Before 2009
				Fuzzy c-means clustering	Before 2009
		Grading of (raw)materials	0%	Combination of rule based knowledge representation, fuzzy logic and genetic algorithms	Before 2009
		Improved dispatching rules	0%	Genetic algorithms	Before 2009
				Decision tree based classifications rules	Before 2009
		Improved methods for a specific process	0%	Range of Artificial Intelligence tools	2013
83%	Root Cause analysis	Root cause analysis of nonconformities in the production process	38%	Herfindahl–Hirschman Index (HHI)	2015
		Detection of change points in control charts	19%	Tree based supervised learner	Before 2009
		Root cause analysis for process failure	19%	Bayesian network, Design of Experiment and Statical Process Control	Before 2009

		General Root cause analysis	14%	Regression, Classification, Clustering	2010
		Root cause analysis for unnatural patterns in the data	0%	Fractal dimension based classifier	Before 2009
82%	Condition based monitoring	Monitoring process conditions	67%	Hybrid fuzzy inductive learning	Before 2009
		Monitoring of parameters settings and their effects	29%	Integrated Neural Networks and rough set techniques (other article extended with fuzzy set theory)	Before 2009
				Fuzzy set theory with fuzzy variable rough set	Before 2009
78%	Patterns causing process variations	Identification of critical process parameters	52%	kernel-based approach combined with a maximum margin- based support vector regression algorithm	2010
		Detection of abnormal process behavior	48%	SPC combined with artificial neural networks, support vector regression and multivariate adaptive regression splines	2012
				Hybrid neural network and decision tree	Before 2009
		Identification of process fault classes	0%	Decision tree classification	Before 2009
				Metric Temporal Logic	Before 2009
73%	Decision support	Improved scheduling decisions by insights in options and effects	52%	Genetic algorithms	2014
				Cooperative estimation of contribution algorithm	2014
				Genetic algorithms	2014
				Evolutionary algorithms combined with hybrid planning	2014
		Insights in the effect of to-be-made decisions	48%	Knowledge discovery for databases	2013
				Genetic algorithms	Before 2009
				Workflow mining by Artificial Neural Networks and fuzzy rule sets	Before 2009
72%	Cycle/lead time prediction	Forecasting production cycle time	81%	Gauss-Newton regression method and back-propagation neural network	2012
				Stepwise linear regression and symbolic knowledge acquisition technology	2013
				Classification (decision tree and NN)	2013
				Set of data mining tools (multiple articles extended)	Before 2009
		Forecasting lead time	19%	Regression tree based data mining approach	Before 2009
				Decision tree combined with if-then-else rules	Before 2009
67%	Process performance prediction	Forecasting production process performance	86%	data envelopment and back-propagation neural network	2014
				Bayesian method	Before 2009
				Model selection and cross-validation	Before 2009
		Forecasting of manufacturing process behavior	10%	Metric Temporal Logic	2014
				Decision tree	2010
		Prediction of system output	5%	Data mining and type II fuzzy system	Before 2009

Again most informational items are found for production operations. The most promising according to field experts are the broad group 'Knowledge of optimal manufacturing settings', the 'Root cause analysis' and 'condition based monitoring'. These are all informational items that contribute to a solid production process and effectively improving the Manufacturing Operations Management. A wide range of data mining methods is used for this. It is interesting that condition based monitoring literature was all before 2009. It could be that this informational item and its possibilities are becoming more widely known among the industry.

Table 9 Future MES informational matrix for Quality Operations Management

Quality Operations Management					
%score	**Main informational (purpose) group**	**Sub-Informational Item group**	**% within sub**	**Method**	**Year of publication**
83%	Defect/low quality classification	Detection of a product with quality faults	76%	Clustering by self-organizing maps for classification	Before 2009
				Fuzzy k- & c-means clustering	Before 2009
				Association rule mining	Before 2009
		Classification of product quality	19%	Range of DM algorithms	2011
				Classification (decision tree)	2011
				decision tree, artificial neural network and support vector machines	2013
				Integrated neural network and rough set techniques	Before 2009
				Set of data mining tools	Before 2009
				Hierarchical clustering, k-means partitioning	Before 2009
				Hybrid learning based system with Neural Networks and decision tree	Before 2009
		Product state diagnosis	5%	Cluster analysis and supervised machine learning	2014
83%	Root Cause analysis	Root cause analysis of product quality	10%	Hybrid OLAP-association rule	2013
74%	Low yield factors identification	Identification of characteristics for low yield (product quality failure)	43%	Decision correlation rules and contingency vectors.	2012
				Design of experiment data mining	2014
				Spatial statistics with neural networks	2013
				Chi-square automatic interaction detection (CHAID) algorithm and chi-square test.	2013
				Genetic algorithms	2009
				Hybrid OLAP-association rule	2013
				Self-organizing maps, Neural Networks and rule induction	Before 2009
				Genetic programming	Before 2009
				Rough set theory	Before 2009
		Identification of characteristics product quality	29%	Rough sets theory, attribute relevance analysis, anomaly detection analysis, decision trees and rule induction	2012
				Range of DM algorithms	2011
				Rough set theory	Before 2009
		Suggested improvements for next generations based on quality failure	29%	Bayesian Networks	2011
74%	Yield/Low quality prediction	Prediction of product quality	76%	Range of DM algorithms	2011
				Hybrid OLAP-association rule	2013
				Clustering and Artificial Neural Networks	2014
				Feature set decomposition methodology based algorithm	Before 2009
		Yield prediction	24%	Genetic programming	Before 2009
				Decision trees and Neural Networks	Before 2009

The most promising of Quality Operations informational items are focused on the classification of product quality and the identification of what is causing low product quality or defects. This is interesting because it is logical that one needs to know the cause before it can solve a problem. This also ranked high in the Current Informational Matrix. These elements are again focusing on the past and the present. The pro-active prediction of product quality is ranked the lowest. This indicates that the industry is still in the phase of improving knowledge of the past and the present, before they can become proactive and focus on the future.

Table 10 Future MES Informational Matrix for Maintenance Operations Management

Maintenance Operation Management					
%score	**Main informational (purpose) group**	**Sub-Informational Item group**	**% within sub**	**Method**	**Year of publication**
83%	Machine (component) failure prediction	Preventive maintenance schedule recommendations	43%	Decision tree based data mining	Before 2009
		Forecasting machine/equipment failure	29%	Metric Temporal Logic	2014
				Decision tree	2010
				Regression, Classification, Clustering	2010
				Decision tree	Before 2009
				Recurrent Neural Networks model	Before 2009
				Agent based model and data mining tools for prediction	Before 2009
		Forecasting component failure	24%	Set of data mining tools (Decision trees, rough sets, regression and Neural Networks)	Before 2009
		Machine performance prediction	5%	Neural Networks based estimation model	Before 2009
		Forecasting tool wear	0%	Rough set theory based classifier	Before 2009
		Probability for machine failure	0%	Classification by decision tree	Before 2009
82%	Condition based monitoring	Monitoring tool wear	5%	Rough set theory classifier	Before 2009
				Neural Networks and Support Vector Machines	Before 2009
72%	Machine fault diagnostics	Diagnostics of machine part wear and correlations between parts	33%	association rule mining	2015
		Diagnostics of machine failure	24%	Hybrid case based reasoning	Before 2009
				Data mining approach for concept description	Before 2009
				Hybrid rough set theory and a genetic algorithm	Before 2009
		Identification of characteristics of machine failure	19%	Decision theoretic approach to mine the data combined with greedy value for information	Before 2009
		Identification of machine failure	14%	Association rules	Before 2009
		Classification of machine fault types	10%	Rough set theory approach	Before 2009

For maintenance operations the predicting of failure is ranked the highest, this is in contrast with the results of the quality operations where it was still re-active. This can be due to the high cost of downtime and to the fact that downtime is noticed by everyone, while having some defect product is not. The industry knows the importance of reducing downtimes which was supported by the Current MES Informational Matrix where it showed that downtime related informational items scored high in maintenance operations. It can also be found many publication dates of the maintenance operations related articles are before 2009. This indicates that much research has already been conducted in the past which could indicate that they will sooner be possible and integrated in software functionalities. This because there is always a delay between theory and practice.

A general observation is that for all four areas of MOM many knowledge discovery and data mining methods are used. Many methods are also used within the sub-groups, which have similar goals. This indicates that there are many possibilities and opportunities in extracting an informational element. This could be due to the variety of production processes and production data or to the variety of possible methods and techniques.

5. Demonstrate with a Case Study

In this chapter the applicability of both the Current and Future MES informational matrix is demonstrated with a case study in a real-life environment. For the case study it is researched whether the informational elements can indeed be extracted from MES and how useful the information can be for a company in order to get insights for their manufacturing operations management. Also the usefulness for insights provided by the Informational matrices themselves is researched.

The case study starts with the motivation of the case study as a research method. Next, the goal and the scope of the case study are explained. Then, background information is provided about the case study. From this point the case study proceeds by making a split between the applicability case study and usability case study goal. In both parts the case study design, data collection and analysis are discussed. The results of both case studies are then evaluated as well as the choice for a case study as a research method for this research.

5.1. Motivation for a case study as a research method

A case study is a suitable method as it focuses on research questions related to "how, why?" It requires no control of behavioral events and it is focused on the contemporary events (Yin, 2009). For the case study a partnership with a company is necessary. One partnership was established with an engine factory, and therefore it will be a single-case study.

5.2. Goal of the case study

The goal of the case study is to demonstrate both the applicability and the usability of the Informational Matrices. First the applicability is demonstrated by assessing informational items by making use of a MES data analysis. The informational items demonstrated are chosen by the partner company. Second, the usability is demonstrated by assessing how useful the matrices are in general (to have an overview of MES informational capabilities and possibilities), as well as the demonstration of the informational item.

The outcomes of this case study are the start of establishing a portfolio of practical examples and best practices for the applicability and usability of the Informational matrices.

5.3. Scope of the case study

The research focusses on the MES of the partner company. In this MES system the focus is on a restricted part of the manufacturing process and corresponding machines. The data in MES for this production part of the year 2015 is used.

5.4. Background of MES situation in the case study's company

The partner company for the case study is a global engine manufacturer. In order to monitor and control their processes they have an operating MES integrated in their production operations area and a small part of their maintenance operations. The use of the MES is currently aimed at three main activities:

1. Operational process control: The MES actively controls the process and sends bill of materials and work instructions of the specific engine to the corresponding machines and operators. Partly automatic and partly activated by a signal or manual operation.
2. Tracking and tracing of products and materials: The MES tracks the production steps and corresponding engines. Also the system stores all data of operations and parts of each

specific engine so backward tracing of all steps can be conducted when a problem occurs with the engine in a later process stage or in the field.

3. Root cause analysis when a problem occurs: The manufacturing company has employees working with the MES data to analyze problems in their operational processes or products. These analyses are conducted by making use of the historic MES data.

5.5. The case study execution

This case study consists out of two parts which are discussed separately. The first part of the case study focusses on the applicability of both the Current and the Future Informational Matrix by demonstrating informational items in a real-life setting. The second part of the case study focused on the usability and the added value of the Informational matrices, in other words the usefulness.

For confidentiality reasons, not all information might be presented and information might be anonymized.

5.5.1. The applicability case study

First the design of the case study, for this part, is presented. Next, the data collection is presented. Then the actual analysis is presented in two parts. First the demonstration of the informational items by making use of the Current Informational Matrix and second the demonstration of the informational items by making use of the Future Informational Matrix.

5.5.1.3.1. The applicability case study design

As the partner company has to make time and resources available for this research, the informational items for the demonstration are chosen by the company. The informational item chosen is 'Root Cause analysis' for a problem they have with some caps of their engines in production. The Root Cause analysis consists out of two parts as there are two informational matrices. The first part will make use of current analysis techniques like data analysis with Excel and SPSS. The second part will make use of knowledge discovery techniques like data mining by making use of KNIME. KNIME is chosen as the data mining program because it is a powerful but easy to use program. Also because it can be linked to an MS SQL server on which MES operates as well. Therefore it could be used in a real-life MES environment. For the first part, the root cause analysis consists out of four steps:

1. Understand the problem and the production process relevant to the problem
2. Hypothesis generation for the possible root causes
3. Data analysis per hypothesis
4. Evaluate the results and conclude about the root cause and next steps.

Additional informational elements that can be (relatively easy be) extracted from the MES data are demonstrated as well. For this an example dashboard is created in Qlik Sense. This enables the partner company to get a feeling of how the information could be presented and used on a daily basis instead of providing just the plain information output. Qlik Sense was chosen because it can be linked to an MS SQL server, just like KNIME.

For the second part, the root cause analysis consists out of four of the six steps in the Cross Industry Standard Process for Data Mining (CRISP-DM) steps (Chapman P. , et al., 2000).

1. Data preparation
2. Modeling
3. Evaluation

The step business and data understanding (first steps) will already be covered before this phase and the deployment step is out of scope as this is for demonstration purpose only.

5.5.1.3.2. The applicability case study data collection

For the problem with the engine caps, the data from three machines is collected. The torqueing machine, the tightening machine and the fine drilling machine. Also the data from the rejected engines is collected. This data is collected is it was available in MES and because it considers operations related to the caps. The fine drilling is chosen as it is the first operation after the operations considering the caps. Table 11 provides an overview of the data.

Table 11 Overview of data collected for the root cause analysis

Operation name	Description	Measurement	Data period
Cracking	Cracking of the cap from the casting block	Force used for cracking	2015
Tightening	Tightening of the cap back on the casting block with bolts	Tightening moment and rotation angle	2015
Fine drilling	Accurate fine drilling of the cylinder	Diameter 131 and 105	2015
Tracking of errors document (not in MES)	Error and cause per unique engine code	Error and causes	2015

5.5.1.3.3. The applicability case study analysis

The analysis consists out of two parts. The first part will make use of current analysis techniques like data analysis with Excel and SPSS for a root cause analysis and additional informational elements. The second part will make use of data mining techniques for the root cause analysis by making use of KNIME. The two parts are discussed separately.

5.5.1.3.1. Root cause analysis with current data analysis tools

The root cause analysis with current data analysis tools consists out of four steps which are discussed subsequently. As the fifth step, additional informational elements and dashboards are presented.

5.5.1.3.1.1. Understand the problem and the production process relevant to the problem

In order to understand the problem, conversations about the problem have taken place with several employees within the company. Also the production process is analyzed. A simplified representation of the production process involved is provided in Figure 14. Also knowledge of the dependencies in the data is necessary as there are two suppliers but also as each engine has seven caps tightened by two bolts each. This is represented in Figure 15.

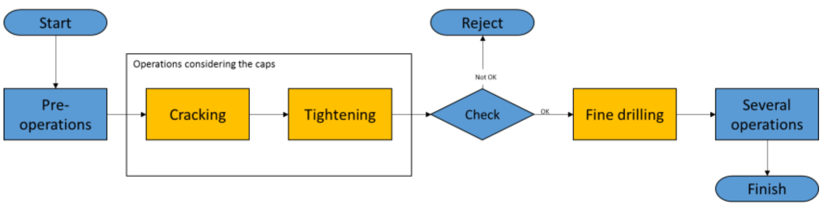


Figure 14 Simplified production process representation relevant to the case study production line part

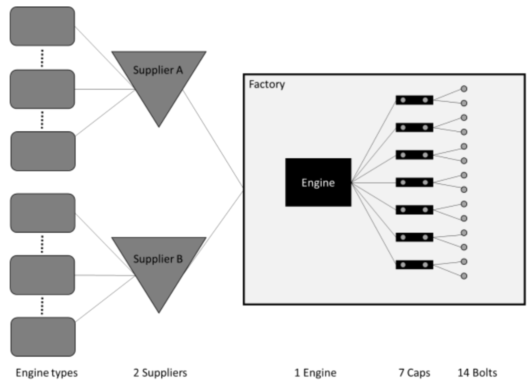


Figure 15 Representation of data dependencies

The engine factory produces seven engine part numbers according to the MES data. From these seven engines parts numbers, the errors arise in the three biggest groups: engine type X, Y and Z. These three part number account for 98% of the production according to the MES data.

The engine factory produces 120 engines per shift, two shifts a day. Most of these engines are present in the MES data, though some gaps are present in the data due to downtime of the MES.

In order discover whether all engines with cap errors can be compared, a statistical study for the three operations on the engine is conducted by making use of SPSS. First, a general analysis for the data has taken place and a standardization of all measurement values is created. All values with corresponding (standardized) Z-value smaller than -3 of greater than 3 are determined to be outliers and are removed from the data. Second, in order to compare the significance of difference of means between two groups the assumption of normality has to be checked. This is checked by creating Q-Q plots. For the operation cracken, the values are not completely normally distributed but it could approach it and for tightening and fine drilling it approaches the normal distribution nicely.

The following groups have been compared for each operation by making use of a t-test:

- Engine types X, Y, Z in groups of two at a time
- Supplier A and Supplier B

It was found that the means between the part numbers and the means between the suppliers are significantly different for most operations on the cap number. Therefore these groups cannot be compared and need to be addressed separately. This was also checked with a chi-square test which also indicated different populations. Also the correlation is tested for the operations between the cap numbers for cracken and fine drilling and between the bolts for tightening. It was found that they are correlated which makes sense as they are from one and the same engine.

The results of all statistical test used in this chapter can be found in Appendix 10.

5.5.1.3.1.2. Hypothesis generation for the possible root causes

After conversations with the process manager and the project manager at the company and brainstorming about the problem, the hypotheses were established.

- Hypothesis 1: Errors in caps arise more often in a specific engine type and/or supplier
- Hypothesis 2: Errors in caps are always present in a specific engine cap number
- Hypothesis 3: Errors in caps occur during specific period in time
- Hypothesis 4: When there is an error in a cap, the processing time is longer
- Hypothesis 5: A higher forces in cracken causes the error in the cap

- Hypothesis 6: When there is an error in the cap, the bolt will need more rotations and has more force applied on it.
- Hypothesis 7: When there is an error in the cap, the cylinder measured in fine drilling is smaller.

The first four hypotheses can conveniently be researched broadly. Hypotheses five to seven are engine type and supplier specific as these are different populations in the data. These hypotheses require much manual work and therefore will only be analyzed for one specific population.

5.5.1.3.1.3. Data analysis per hypothesis

The analyses are completed and an overview of the hypotheses and the findings can be found in Table 12. For clarity reasons only hypothesis one, two, three and six are included in this section. Hypothesis one and two as they provided background for the problem and hypothesis three and six as these had interesting results. The extended results of all hypotheses can be found in Appendix 11.

Table 12 Overview of the hypothesis and the corresponding findings

Hypothesis	Result	Findings
Hypothesis 1: Errors in caps arise more often in a specific engine type and/or supplier	Rejected	
Hypothesis 2: Errors in caps are always present in a specific engine cap number	Not rejected	Interesting period during the summer
Hypothesis 3: Errors in caps occur during specific period in time	Rejected	
Hypothesis 4: When there is an error in a cap, the processing time is longer	Rejected	
Hypothesis 5: A higher forces in cracken causes the error in the cap	Rejected	
Hypothesis 6: When there is an error in the cap, the bolt will need more rotations and has more force applied on it.	Not rejected approved for some bolt numbers	Further research in differences left and right
Hypothesis 7: When there is an error in the cap, the cylinder measured in fine drilling is smaller.	Rejected	

Hypothesis 1: Errors in caps arise more often in a specific engine type and/or supplier

In total 87 engines with an error in a cap where detected in 2015. From these 36 raised at type X, 37 at type Y and fourteen at type Z. The engines with errors are traced in the MES data in order to find the corresponding supplier. 61 engines where found and mapped per type and supplier. This is represented in Figure 16.

Supplier / Engine Type	Supplier A		Supplier B	
Type X	F: 23 T: 3189	0,72%	F: 0 T: 238	0,00%
Type Y	F: 0 T: 0	0,00%	F: 29 T: 39010	0,07%
Type Z	F: 9 T: 901	1,00%	F: 0 T: 238	0,00%

Figure 16 Error engines per engine types and supplier. F = number of false/errors engines and T = the total amount of engines.

It can be found that the errors occur at both suppliers and in all three e ngines types. The hypothesis is therefore rejected.

Hypothesis 2: Errors in caps are always present in a specific engine cap number
The errors occur at the caps of the engines. Usually at only one, but sometimes at two or more caps. Per cap numbers the amount of engines with an error on that cap number are counted per engines type. This is done for each engine type. Not for all engines with errors the error location was saved, therefore that total differs from the total found in hypothesis 1. The errors occur in almost all caps, only never cap 7. The hypothesis is rejected.

Table 13 amount of engines with an error detected on a specific cap

Cap number	Type X	Type Y	Type Z
1	1	6	0
2	2	5	1
3	10	3	2
4	6	3	1
5	1	3	0
6	5	3	0
7	0	0	0

Hypothesis 3: Errors in caps occur during specific period in time
The engines errors per engine type are mapped during the year. The engine errors are counted per week. In Figure 17 Number of errors per week per engine type the result is presented. It can be found that the errors occur during the whole year. It is interesting that during the summer only engines of Type Y occur. The Type Z engines only occur at the end of the year. However, this is due to the fact that the part number for this engine type only exists at the end of the year. The hypothesis is not confirmed as the errors occur during the whole year. However the summer is and interesting period, which does not directly lead to rejecting the hypothesis.

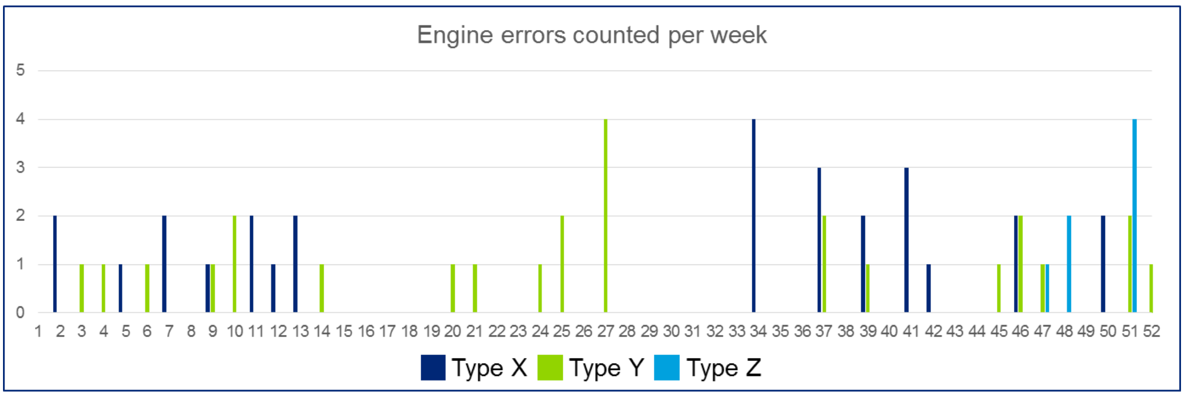


Figure 17 Number of errors per week per engine type

Hypothesis 6: When there is an error in the cap, the bolt will need more rotations and has more force applied on it.

The data is of Type X and supplier A is used for this analysis. First the average values for each measurement item is plotted. This is presented in Figure 18.

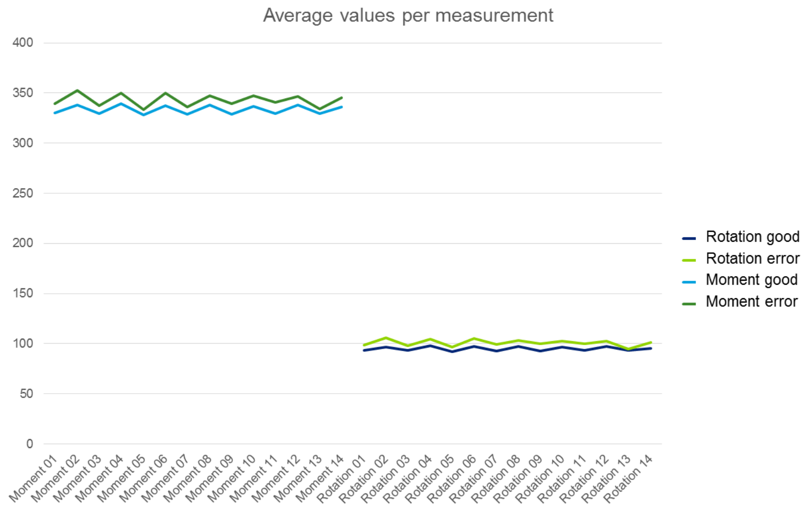


Figure 18 General analysis tightening measurements

Is can be observed that there is a saw-pattern in the graphs. All even values are higher than the odd values. This could indicate a difference in left and right as cap one has bolt 1 and 2, cap two has bolt 3 and 4 and so forth. When splitting the data into even and odd numbers, the pattern disappears. These graphs can be found in Appendix 11.

To further research the differences between the good engines and the engines with errored caps, boxplots are created. These are presented in Figure 19.

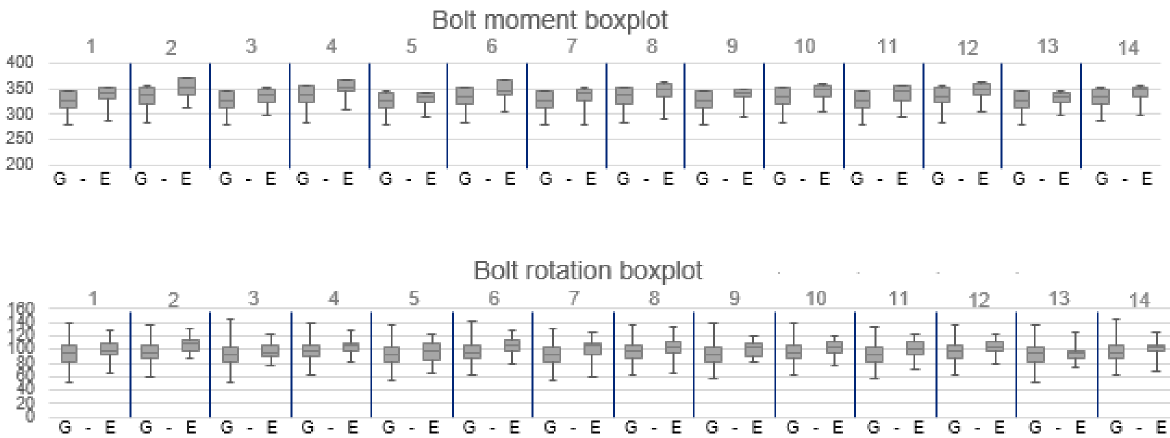


Figure 19 Boxplots for tightening

The differences are again researched whether they are significant by comparing the means by making use of the t-test in SPSS. The results are presented in Table 14. No variances where significantly different.

Table 14 Significant differences between the means of good engines and error engines

Rotation	Moment	Engine		Rotation	Moment		Cap	Errors
X	X	Bolt 1	Bolt 2	Sign	Sign		Cap 1	1
X	X	Bolt 3	Bolt 4	Sign	Sign		Cap 2	2
X	X	Bolt 5	Bolt 6	Sign	Sign		Cap 3	10
X	X	Bolt 7	Bolt 8	X	X		Cap 4	6
Sign	Sign	Bolt 9	Bolt 10	X	Sign		Cap 5	1
X	X	Bolt 11	Bolt 12	X	X		Cap 6	5
X	X	Bolt 13	Bolt 14	X	X		Cap 7	0

It can be found that for some bolt numbers there are significant differences between the means. For example bolt number 6 is significantly different for engines with errors compared to good engines. Also in the corresponding cap number, there are many errors. However, in cap 5 it can be found that four out of three values are significant different for engines with errors compared to good caps, though there are no actual errors found on cap 5. Further research to the differences between left and right might generate interesting insights. Especially setting up an experiment on which side most errors are found (left of right) would be interesting.

The hypothesis is not rejected but not confirmed as well. Additional research is needed.

5.5.1.3.1.4. Evaluate the results and conclude about the root cause and next steps.

The results of the seven hypotheses can be found in Table 12 in section 5.5.1.3.1.3. The actual root cause of the problem is not found, though some interesting insights are raised by the MES data analysis. Further research on the summer period and the differences between the left and right side of the cap seem the most promising. Also repeating the hypotheses five to seven for the engine types Y and Z is recommended.

5.5.1.3.2. Additional informational elements and dashboards

In order to demonstrate that more informational elements can be extracted from the MES data the other informational elements are analyzed as well.

- Operational Equipment Efficiency
- Work in process data
- Real-time plant and production status
- Schedule vs production attainment
- Throughput
- Yield

With the data available only the yield and the thought put could be correctly calculated but the other could be approximated for exemplary purposes. An example dashboard with this information is created.

In the Future Informational Framework it showed that identifying critical process parameters is ranked very high as well as finding nonconformities in the process and change points in the control charts. Monitoring these process conditions is ranked third. In order to provide insights in the future, two more dashboards where created. In these dashboards the operations cracken is monitored. The values in MES per cap can be monitored over time and compared to periods of time where errors occurred. Also the average values for a specific period in time can be compared to the average in other periods of time split per engine type. These are examples demonstrate the intractability of dashboards and the fast forwards clicking trough the screens for monitoring and first sight analyses. The example dashboards are included in Appendix 12.

5.5.1.3.4. Root Cause analysis with KDD (future) data analysis tools

For the second part of the root cause analysis, data mining tools from the Future Informational Framework are demonstrated. The Root cause analysis can be mapped as a root cause analysis, defect low quality classification or low yield factors identification. This leads to a list of various data mining methods to use. For demonstration purposes, four data mining algorithms have been chosen. A short description and a motivation for the choice can be found in Table 15. All four algorithms will be used as a classification model. Classification aims to identify to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known, in a supervised learning technique (Wilbik, 2014). The classification will be a binary class problem with class 0 as good engines and class 1 as engines with errors.

Table 15 Overview of chosen data mining algorithms with a short explanation and motivation

Data mining method	Explanation	Motivation
Decision Trees	The creation of a model that will predict the value of a target variable based on several previous analyzed input variables. The leaves, represent class labels and the branches represent a conjunct of features that lead to those class labels (Chauhan, 2013)	An easy to use but powerful algorithm. Also the branches of the model might indicate important values of parameters that have a significant impact on the classification and ultimately the root cause of the error
Random Forest	A combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Breiman, 2001).	This is more robust to noise than the decision tree but is still a very powerful method.
(Probability) Neural Networks (PNN)	Probabilistic neural network (PNN) can compute nonlinear decision boundaries which approaches the Bayes optimal (Specht, 1990)	The PNN has a fast learning speed and provides probabilities based on a Bayesian classifier model. The probabilities might provide useful insights when the actual classification is not adequate as is provides a probability to which class the data string belongs to.
Support Vector Machines	A machine learning algorithms for classification that maps non-linear input vectors in a high-dimensional space to construct a linear decision (Cortes & Vapnik, 1995)	This can handle non-linear data which makes it generalizable for multiple data sources.

5.5.1.3.4.1. Data preparation

In order to create the right data samples, the data is researched. As established in section 5.5.1.3.1 the three engine types are three different populations in the data and therefore considered separately. The hypotheses five to seven focused on engine type X and supplier A and therefore this data mining demonstration will do this as well.

In engine type X of supplier A, only 23 of the 3189 engines had an error. In other words the groups with good engines (class 0) is significantly bigger then the group with errors (class 1). The data is unbalanced. Though this is also a main attribute of the problem, as it represents its uniqueness, it makes it difficult for the algorithms to distinguish between the two classes.

No articles of the literature research handled unbalanced data so no methods could be copied from literature. In order to handle the unbalanced data, the data is sampled. Then the data mining methods of the articles will be used.

The data is sampled with an oversampling technique. Oversampling is chosen because under sampling would leave to little data left for an appropriate analysis. A risk in oversampling is that the model would eventually over fit. Therefore a cross-validation technique is used where two groups of data are created. This is represented in Figure 20. Group 1 consists of 17 fault engines (originated from 18 engines, but one had too many missing data point and is therefore removed) copies 50 times and a random sample of 2000 of the 3189 good engines. This data is used to learn a model. Group 2 consist out of 5 category 1 engines, and a random sample of 1000 category 0 engines. Outliers are not removed from the data is they might be useful attributed for the model.

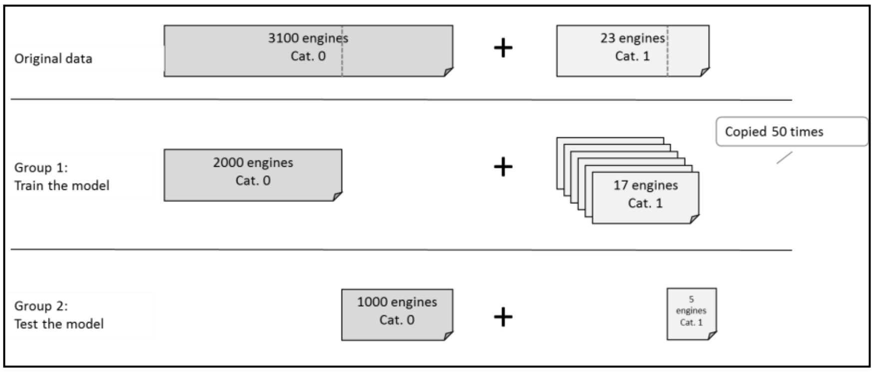


Figure 20 Data preparation and split between group 1 and group 2

5.5.1.3.4.2. Modelling

The data mining models are created by making use of KNIME. KNIME is an open source business intelligence tool that works with 'nodes' that can be dragged into the modeling screen. The nodes can be connected and configured before the model runs. The four models will be discussed separately in the next section. The print screens of the KNIME models are added in Appendix 13**Error! Reference source not found.**. KNIME has a dot as decimal separator; this has to be altered to commas for every KNIME model.

Decision Tree model
The decision tree in KNIME can handle missing values so no alteration on the missing values is necessary. The group 1 data is used for the decision tree learner and the group 2 data is used for the decision tree predictor. At first no pruning was used. Pruning is a method that reduces the size of the tree by eliminating the section with only little power. The model learns the target variable category which holds the class of the engine as a string. The predictor node, predicts the category of the new data. In a second run pruning was used.

Random Forest model
The random forest algorithms in KNIME cannot handle missing values. Therefore a missing values node is added to the model. The missing values only occur at the class 0 data. Because this group is very big, the rows with missing values are deleted. The group 1 data is used for the random forest

learner and the group 2 data is used for the random forest predictor. The model learns the target variable category which holds the class of the engine as a string. The predictor node, predicts the category of the new data.

Probability Neural Network model
The PNN node in KNIME can handle missing values. However, the option to delete the rows is not available. Therefore, the missing value rows are deleted by a missing values node. The group 1 data is used for the PNN learner and the group 2 data is used for PNN predictor. The model learns the target variable category which holds the class of the engine as a string. The predictor node, predicts the category of the new data.

Support Vector Machine (SVM) model
The PNN algorithm in KNIME cannot handle missing values. Therefore a missing values node is added to the model. The missing values only occur at the class 0 data. Because this group is very big, the rows with missing values are deleted. The group 1 data is used for the SVM learner and the group 2 data is used for SVM predictor. The model learns the target variable category which holds the class of the engine as a string. The predictor node, predicts the category of the new data.

1.5.3.3.3.3. Evaluation
The models predict the class of the group 2 data based on the model they 'learned' by making use of the group 1 data and the algorithm. The output provides the table presented in Table 16.

Table 16 Output representation of KNIME scorer which is directly connected to the predictor

Category \ Predicted category		
	1	0
1	True Positive (TP)	False Negative (FN)
0	False Positive (FP)	True Negative (TN)

The models are evaluated on the following criteria:

- Accuracy: How often is the classifier correct?
  - This value should be as high as possible
  - Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$
- Misclassification rate: How often is the classifier wrong?
  - This value should be as low as possible
  - Misclassification rate = $\frac{FN+FP}{TP+TN+FP+FN}$
- False positive ratio: How often is it classified good when there was an error?
  - This value should be as low as possible, though an extra check on a good engine is not bad dependent on the extra work or related costs.
  - FP ratio = $\frac{FP}{TP+FP}$
- False negative ratio: How other is it classified as error when it was good?
  - This value should be as low as possible as it means that an engine with an error will go further in the process. This can have related costs like defect engines or even reputation cost, when the engines get to the field.
  - FN Ratio $\frac{FN}{TN+FN}$
- Cohen's Kappa: How well does the classifier perform compared to chance?
  - This should be as low as possible as it compares the predicted output to the null error rate. The null error rate is the amount of time the model would be wrong when it would just predict all items to the biggest category.

The results of all models are summarized in Table 17.

Table 17 Summary of data mining model results

Decision Tree

Category \ Predicted category				
	1	0	Accuracy	98,40%
1	0	5	Misclassification rate	1,60%
0	11	982	FP Ratio	100,00%
			FN Ratio	0,51%
			Cohen's Kappa	-0,007

Decision Tree (with Pruning)

Category \ Predicted category				
	1	0	Accuracy	97,70%
1	0	5	Misclassification rate	2,30%
0	18	975	FP Ratio	100,00%
			FN Ratio	0,51%
			Cohen's Kappa	-0,008

Random Forest

Category \ Predicted category				
	1	0	Accuracy	99,50%
1	0	5	Misclassification rate	0,50%
0	0	990	FP Ratio	
			FN Ratio	0,50%
			Cohen's Kappa	0

Probability Neural Network

Category \ Predicted category				
	1	0	Accuracy	99,50%
1	0	5	Misclassification rate	0,50%
0	0	993	FP Ratio	
			FN Ratio	0,50%
			Cohen's Kappa	0

Support Vector Machine

Category \ Predicted category				
	1	0	Accuracy	54,87%
1	4	1	Misclassification rate	45,13%
0	448	542	FP Ratio	99,12%
			FN Ratio	0,18%
			Cohen's Kappa	0,119

Table 18 Overview of the probability of belonging to a specific class according to the PNN model.

Actual Class	Probability(Class=0)	Probability(Class=1)	Predicted class
0	0,775894	0,224106	0
0	0,803595	0,196405	0
0	0,830991	0,169009	0
0	0,83215	0,16785	0
0	0,843315	0,156685	0
.	.	.	.
.	.	.	.
1	0,925472	0,074528	0
1	0,967191	0,032809	0
1	0,978897	0,021103	0
1	0,990669	0,009331	0
1	0,998967	0,001033	0
.	.	.	.
.	.	.	.

In Table 18 the five items with the highest probability of belonging to class 1 are presented and the actual class 1 engine. The engines in between are replaced by dots for clarity reasons.

When looking at these results it can be found that no model is close to being an adequate predictor for the engine errors. However there are some differences between the models.

The Decision Tree model, both for with and without pruning, does not classify any of the error engines correctly and has a false positive rate of 100% which means that all engines it indicates as error, is actually a good engine. The Cohen's Kappa is negative for these models which is rare. This means that the model performs worse than just classifying all engines as class 0. However, this is due to the uniqueness of the class 1 as there are only 5 class 1 engines compared to 1000 class 0 engines. The decision tree does try to classify some engines as class 1, though wrongly.

The Random Forest and the PNN both have the same behavior. Both models classify all engines as good (class 0). This generates a high accuracy but this is misleading as the model does not help with the classification problem. For this reason the Cohen's Kappa is 0 for both models. The probabilities of the PNN are also not good for the actual class 0 engines as these probabilities are very low and also not higher compared to others.

The SVM model seems like the best model as it correctly classified 4 out of the 5 class 1 engines. However, the FP rate is 99,12% so many engines would get an extra check based on this result, while they are actually good engines.

It can be concluded that for this demonstration the data mining methods did not find the root cause or generate a model that can correctly classify the (unique) engines with the errors.

5.5.2. The usability case study

First the design of the case study, for this part, is presented. Next, the data collection is presented. Then the analysis of the usability of the matrices is presented based on the data, case interviews, gathered.

5.5.2.1. The usability case study design

The usability of the Informational Matrices is assessed in two ways. First, the results of the informational item demonstration (the root cause analysis) are presented. The feedback and first impression of this is gathered. Second, small semi-structured interviews are conducted with attendees of the presentation and a MES expert at the company. Semi-structured interviews are chosen because it gathers answers on a set of questions but also creates the possibility for an open discussion to get broader information. The interview questions are added in Appendix 14.

For confidentiality reasons, not all information might be presented and information might be anonymized.

5.5.2.2. The usability case study data collection

For the usability first the feedback and first impression of the Informational matrices and the informational item demonstration is gathered during a presentation given at March 21st 2016 at 13:00h at the company location. Second, semi-structured interviews are taken from relevant people within the company. Table 19 provides an overview of the interviews. The full interview transcripts can be found in Appendix 15.

Table 19 Overview of interviewees for data capturing

Function	Interview Date/Time	Present at the presentation	Relation to MES
Managing director engine factory	30-03-16 11:45	Yes	Final responsible for the whole factory, including MES
Senior PE Project Manager	29-03-16 11:00	Yes	Uses MES data frequently in problem analyses and improvement projects. Part of the 'new MES' team
PE Project Manager	29-03-16 10:00	No	Part of the 'new MES' team
Area Manager Engine factory	30-03-16 11:30	Yes	Uses the reports others generate from MES frequently
Head of PE machining process	31-03-16 12:00	Yes	Involved with establishing the current MES
Supervisor Engine Factory Machining process line 2	31-03-16 18:00	Yes	Uses a small part of the operating MES

5.5.2.3. Informational Matrices usability: the analysis

Six interviews are conducted among different functions of the organization. In these interviews questions were asked about the usability and added value of the frameworks. From this it can be derived which insights can be gathered for the company by making use of the framework. The two frameworks and the proposed example dashboards are discussed separately.

The Current Informational Matrix did not have surprising Informational items though some were not expected in the MES environment. It also provided an overview for the company that was not present before. The company's managing director recognized the lack of overview and said this framework helps to start the discussion within his company about MES and what they would want in their own MES. It was also stated that there is always a difference between what is possible in MES, so the Current Informational Matrix, and what is actually implemented in a real-life situation. Some informational items are not possible to configure due to the IT infrastructure and legacy systems, while other informational items are not configured by choice. This information can be tracked in another system but also not measured automatically at all as they the company does not see the added value of measuring this information automatically. What is configured in MES is always demand driven, so what does the company want and how does is related to their specific process. Concluding, the actual implantation and configuration of MES is company specific but having an overview which was not present before is useful to get an insight in the possibilities which can help in the discussion of what the company would want within its MES.

The Future Informational Matrix provided an insight in the future. Some interviewees found this useful to have an insight in what will be possible in a near future. Especially one interviewee who is in the team of developing the new MES for the partner company was interesting in improve identification of critical process parameters and monitoring those to become more pro-active based on the (MES) data. Others also indicated that these future abilities did not represent the current challenges of the company. It is interesting to know what is coming in the future but not relevant now. They did indicate that it might be more useful for more high tech companies. Concluding, it is interesting to have an overview of what is coming in the near future but the relevance depends on the company and their challenges.

There were two types of proposed example dashboards, a general overview dashboard and an operating parameters dashboard. The general dashboard was very interesting for the partner company. It provided an example of what several interviewees already have been looking and asking for. These types of dashboards will be implanted in the new MES as well. The operational parameters dashboards are not relevant for all interviewees. Some believe it is useful to be able to click through screens and compare parameters in time and with each other while other prefers analytical programs for this. Also some so not deal with this in their daily work so they do not have a strong opinion about this. Concluding, the example dashboards are useful but especially the general dashboard.

5.6. Case study results evaluation

With the evaluation of the results of the case study, the fourth research question is answered.

Research Question 1. What relevant insight are provided and what challenges can be encountered in a real world situation?

The insights provided and challenges encountered are different for both parts of the case study. First, the demonstration of informational items is discussed and next the usability of the Informational Matrices.

For the demonstration of an informational item, a root cause analysis is conducted for an error with engine caps. Data from the machining operations regarding the engine caps provided insights in the possible causes of the error. Having improved insights in this of this error, could help the company to reduce the amount of errors which ultimately saves money. There were challenges in this part. First, not all information that was needed was present in the MES. In a real-world environment many system co-exist and not all data is documented in the same format or documented at all. Second, when the data is available the data quality is not always good due to different formats or downtime of the MES. This can make the analysis challenging. Third, it is important to have knowledge about the production process and the product in order to establish the hypotheses but also to explain some of the behavior found in the data.

The additional extracted informational elements, and example informational elements that were represented in the example dashboards provided an example of what the company would want in practice. Having accessible, reliable and accurate information about the production process can help the company with creating a more solid process which will lead to increased process performance, reduced errors and much more. In other words it can help a company improve its Manufacturing Operations Management. This will ultimately lead to increased revenues or reduced cost. The challenges for this part are again that not all information is present in MES (or any system) and the data quality. Also for dashboards it is important to have deep understanding in what information is needed and what is the best format to present it.

The Informational Matrices provided insights for the company as well. The Current Informational Matrix was very useful to get an overview of what is possible in MES and how experts rank these. For the partner company this enables them to start the discussion within their own company of what they want for their MES. There are also some challenges when applying this Matrix. First, not all informational items are possible to extract from a MES as mentioned before. Second, extracting certain information from MES, in other word configuring MES in such a way that is measures, analyzes and presents the information is always a managerial choice based on the demand of that company. It is very difficult to generalize the informational element to direct benefits of a company.

Third, not all informational items can be extracted from MES because of the IT structure or because of the IT in the machines and PLCs.

The Future Informational Matrix provided an insight in the Future of MES abilities. This enables the company to think about these upcoming opportunities when making data related choices in their current systems. The challenges with this matrix is that the actual extracting the informational elements is more difficult than in the current informational matrix. The literature on which the elements are based are all full researches themselves. Also the attitude towards newer data analysis techniques where more conservative in the partner company as they are focused on fundamental challenges first and want to see results of this first before they believe the 'advanced' methods.

5.7. Case study as a research method evaluation

The case study has both a quantitative and a qualitative aspect. The quantitative aspect is evaluated during a presentation at the partner company. At the presentation a project manager, process managers and the managing director of the engine factory were present. All attendees had knowledge of either a process part or the MES data involved. Therefore, the attendees all had a critical view in reviewing the results. During the presentation, and the case study evaluation the qualitative case study analysis was evaluated as solid and representable.

For the qualitative part of the research four test criteria from literature are used. These tests are often used in empirical social research but as this case study also aim to demonstrate and understand a phoneme (the MES and its abilities), these criteria can be used as well. The criteria, based on the article of Yin (2009):

- Construct validity: Whether the key operational measures are used for the purpose to the case study
- Internal validity: Whether (causal) relationships have been searched for in the case study (only for exploratory and causal studies)
- External validity: Whether the case study can be generalized to other cases within a specified domain
- Reliability: Whether the operation the case study is repeatable

For the construct validity multiple employees within the partner company have been interviewed. All of these employees had knowledge of MES but they had different functions and responsibilities considering MES. For the internal validity the status of the company considering their MES was researched in combination with the interviews. Having knowledge of the current MES, its functionality is key to understand the answers in the interviews. Also the developments in their MES environment, that they are developing a new MES, are key information to relate to the interview answers. The external validity or the generalizability is valid for the basics of MES are basic challenges. The conclusions about the overview that the informational matrices provided, which was not present before, are generalizable. Another conclusion about MES usage and the future possibilities are more company specific. This because first, each company has its own operation process and IT architecture and second because there can be a difference in attitude towards data and information automation. A high tech company with a very precise process would have a different attitude towards MES, and a different benefit, than the partnering company of the case study.

6. Conclusion

In this section the conclusion of the research is presented. Also the limitations of this research and the suggestions for further research.

6.1. Conclusion of the research

The developed Informational matrices of this research provide an overview of the possibilities and opportunities of extracting information from MES data that provide insights for manufacturing operations management. The Informational Matrices provide a comprehensive and consistent overview which is not present yet. Also it provides insights in MES for the production field and it includes opportunities derived from advanced data analytics methods. Moreover, the Informational Matrices are both applicable and useable in a real-life scenario as demonstrated by a case study.

The overview is established by creating two Informational Matrices which reflect on both the current MES functionality and advanced data analytics like knowledge discovery and data mining. Both matrices exist of Informational items that one could extract from MES given current MES functionality and standards, or by making use of knowledge discovery tools. Because Manufacturing Operations Management considers four areas (Production operations, Quality operations, Maintenance Operations and Inventory Operations), there is a separate informational matrix for every area.

The Informational matrices consist of Informational that have a set of properties which provide more information about the specific Item. Both have an indicated usefulness percentage score which was established by a questionnaire among MES experts. This generates a ranking of the informational items. For the Current Informational Matrix the time frame of which the informational item provides information is added which always is past or present oriented. Also the source from which the informational item was found is added. The degree standardization of data collection and performance analysis in MES applications of the specific MES area is included to indicate the generalizability of the configuration of the informational items. It was found that data acquisition is more standardized than performance analysis and most production operations have the highest degree of standardization. For the Future Informational Matrix, main and sub groups are defined and the knowledge discovery method used in literature is added as a property per sub group. It was found that a wide range of methods can be used to extract information from the data. Also, the year of publication is added to identify focus areas in time and to estimate the time it will take before the informational item could be widely available in practice. It was found that some sub groups have had research conducted on for a longer time than others.

The case study demonstrated that the Informational matrices are both applicable and usable. The case study demonstrated that it is possible to extract the informational items from the data and that there are many opportunities in the presentation of these informational items to enable fast and reliable decision making. The Current Informational Matrix enables a company to assess their own MES related choices and their own informational needs. Also the ranking provides an opportunity to benchmark their MES and choices to what is considered useful by the MES experts. The Future Informational Matrix enables a company to be prepared for the possible future abilities of data analysis, in other words for future MES. This is very useful to consider when making decision about data capturing, data structures and MES today.

6.2. Limitations of the research

The first limitation of the research is in the survey among MES experts. The number of respondents for both surveys could be higher to increase the reliability of the answers provided by them.

Second, the literature study conducted for the Future Informational Matrix has limitations. This literature study was focused at knowledge discovery and data mining techniques. However, there are also other 'advanced' data analysis techniques that might provide informational items, for example Monte Carlo simulation. This was not included in this research. Moreover, for the literature of 2008 and before, a literature review was used. This has as a consequence that this research is dependent on how complete this literature review conducted by an external person is.

The case study used in this research only considered one company. This makes the influence of this company and this company's vision relatively big. Also the MES system this company used has limited informational functionality which could be different in other companies. Furthermore, the articles used for the literature study where mostly researched conducted in a high tech company, like a semiconductor company. This makes the Future Informational Matrix less applicable to the case study company. Last, the company's desire for the root cause analysis affected the demonstration of informational items. The case study was conducted on a very specific problem. For this MES (data) can be used, however, the majority of the MES use is about general process conditions and improving problems like bottlenecks and other process obstacles. MES is very suitable to detect these problems and the cause of these problems.

The last limitation is that this research was not aimed at fully exploiting the data mining techniques. Therefore the data mining analyses are limited.

6.3. Future Research

First, the surveys could be extended to more sources to get more respondents. For this cooperation with the Annual MES Survey of Iskamp and Snoeij could be searched as they conduct a very wide MES research every year. This cooperation would be interesting in all aspects of the research.

Next, the literature research could be extended to other advanced data analysis fields. Also an additional search for articles of 2008 and before could provide interesting results.

It would also be interesting to test the MES Informational Matrices in more real world environments with more case studies. This needs to involve more case studies in the discrete manufacturing industry which are both high tech and less high tech. Also the scope can be wider in future research and case studies could be conducted in other industries to check the generalizability of the informational matrices.

Also the case study of this research could be extended by an extensive data mining research as the data mining opportunities have not been exploited fully. A data mining expert could further improve the data mining part of the case study which could provide interesting results.

Last it would be interesting to research how measuring certain information in a MES environment relates to management initiatives like Lean and Six Sigma which are getting widely implemented in the manufacturing industries. Researching whether gathering more information in MES enhances or counterwork these initiatives could be interesting.

Bibliography

Breiman, L. (2001, Oct). Random Forest. *Machine Learning, 45*, 5-32.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide.* The CRISP-DM consortium.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide.* The CRISP-DM consortium.

Chauhan, R. S. (2013, May). Predicting the Value of a Target Attribute Using Data Mining. *International Journal of Soft Computing and Engineering (IJSCE), Vol. 3*(Issue 2).

Choudhary, A., Harding, J., & Tiwari, M. (2009). Data mining in manufacturing: A review based on the kind o knowledge. *Journal of Intelligent Manufacturing, 20*(5), 501-521. doi:10.1007/s10845-008-0145-x

Cisco. (2014, Mar-Arl). How IoT is affecting industrial automation. (InTech, Interviewer) Retrieved from https://www.isa.org/standards-and-publications/isa-publications/intech-magazine/2014/mar-apr/cover-story-internet-of-things/

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*(20), 273-297.

Critical Manufacturing. (2014a, 04 10). The Future of MES. Retrieved from http://www.criticalmanufacturing.com/en/newsroom/blog/posts/blog/the-future-of-mes#.VoOnyY1IiM-

Critical Manufacturing. (2014b, May 09). *Critical Manufacturing*. Retrieved from IoT will change the way manufacturing is performed today: http://www.criticalmanufacturing.com/en/newsroom/blog/posts/blog/manufacturing-and-the-internet-of-things#.Vqss8I32Ydl

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining.*

Gartner. (2015, May 18). Gartner Supply Chain Management Market Share Update.

Giusto, D., Iera, A., G., M., & Atzori, L. (. (2010). *The Internet of Things.* Springer.

Goonatilake, S., & Khebbal, S. (1995). *Intelligent Hybrid Systems.* Wiley.

Gröger, C., Niedermann, F., & Mitschand, B. (2012). Data Mining-driven manufacturing process optimization . *Proceedings of the World Congress on Engineering*.

Gröger, C., Niedermann, F., & Mitschang, B. (2012, July 4-6). Data Mining-driven Manufacturing Process Optimization. *World Congress on Engineering*.

INFOR. (2014, Mar-Apr). How IoT is affecting industrial automation. (Intech, Interviewer) Retrieved from https://www.isa.org/standards-and-publications/isa-publications/intech-magazine/2014/mar-apr/cover-story-internet-of-things/

Ipskamp, G., & Snoeij, J. (2015). *MES Product Survey: The invaluable guide to accelerate the evaluation of your potential MES solution.* CGI.

ISA. (n.d.). About ISA. Retrieved January 2016, from https://www.isa.org/about-isa/

Keen, P., & Morton, M. (1979). *Decision support systems: an organizationa perspective.* Addison-Wesley Reading.

Kletti, J. (2007). *Manufacturing Execution Systems – MES.* Springer.

Littlefield, M. (2012, February 6). *LNS Research*. Retrieved December 2015, from Manufacturing Operations Management: MOM vs. MES and the Winner Is...: http://blog.lnsresearch.com/bid/116384/Manufacturing-Operations-Management-MOM-vs-MES-and-the-Winner-Is

Manufacturing, C. (2013, november 19). *A brief history of Manufacturing Execution Systems*. Retrieved from Critical Manufacturing: http://www.criticalmanufacturing.com/en/newsroom/blog/posts/blog/mes-history-59#.Vjh-YY2FMic

Market And Markets. (2015b). *Internet of Things (IoT) in Manufacturing Market by Software (Analytics, Network Management, Remote Monitoring, Security), Application (Predictive Maintenance, Logistics & Supply Chain Management), Platform, Service, & by Vertical-Global Forecast to 2020.* Marketandmarkets.com. Retrieved from http://www.marketsandmarkets.com/Market-Reports/iot-manufacturing-market-129197408.html

Markets And Markets. (2015a, September). Manufacturing Execution System Market by Deployment Type, Process Industry (Chemicals, Food & Beverages, Life Science, Oil & Gas, Power), Discrete Industry (Aerospace & Defense, Automotive, Medical Devices) and by Region - Global Forecast to 2020. Retrieved from http://www.marketsandmarkets.com/PressReleases/mes.asp

MESA International. (1997). *The Benefits of MES - A Report from the field.* Manufacturing Enterprise Systems Association. Arizona: Chandler.

MESA International. (2015). *MESA International Homepage Quick Poll*. Retrieved November 2015, from MESA International: www.mesa.org

MESA International. (n.d.). MESA Model. Retrieved from http://www.mesa.org/en/modelstrategicinitiatives/MESAModel.asp

Meyer, H., Fuchs, F., & Thiel, K. (2009). *Manufacturing Execution Systems: Optimal Design, Planning and Deployment.* McGraw-Hill.

NearSoft Europe. (2013). *MOM vs MES*. Retrieved December 2015, from NearSoft Europe: http://nearsoft.eu/manufacturing_it/mom-vs-mes/

Ogura, A. (2015, November 16). Interview for general knowledge about MES with MES experts. (I. Stassen, Interviewer)

Pawlak, Z. (1982). Rough Sets. *International Journal of Computer and Information Sciences, Vol. 11*(No. 5).

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2006). The Design Science Research Process: A Model For Producing And Presenting Information Systems Research. In *1st International Conference on Design Science in Information Systems and Technology (DESRIST)* (pp. 83-106).

Piatetsky-Shapiro, G. (1991). Discovery, analysis, and presentation of strong rules. In G. Piatetsky-Shapiro, & W. J. Frawley, *Discovery, analysis, and presentation of strong rules.* AAAI Press.

Redman, T. C. (1998, February). The Impact of Poor Data Quality on the Typical Enterpise. *Communications of the ACM, 41*(2).

Sadiq, S. (2013). *Handboek of Data Quality.* Springer. doi:DOI: 10.1007/978-3-642-36257-6

Saenz de Ugartea, B., Artibab, A., & Pellerina, R. (2009, September 6). Manufacturing execution system - a literature review. *Production Planning & Control*, 525-539. doi:DOI: 10.1080/09537280902938613

Schneider Electric Software. (n.d.). Pepsi Bottling Ventures of Idaho. Retrieved January 2015

Specht, D. F. (1990). Probabilistic Neural Networks. *Neural Networks, Vol 3*, 109 - 118.

Stansfield, J. (2014, March 13). Microsoft SQL Server vs. Oracle: The Same, But Different? Retrieved from http://www.seguetech.com/blog/2014/03/13/Microsoft-SQL-Server-versus-oracle

Van der Aalst, W. (2005). Business Alignment: Using Process Mining as a Tool for Delta Analysis and Conformance Testing. *Requirements Engineering Journal, 3*(10), 198-211.

Van Veen, N. (2015, November 25). Interview for MES general knowledge from MES vendors. (I. Stassen, Interviewer)

Wilbik, A. (2014). 1BM56: Business Intelligence lecture slides.

Yin, R. K. (2009). *Case Study Research: Deisgn and Methods* (Vol. 5). SAGE.

Appendix 1. Non written courses for background section

Date	Contact person	Company	Contact method	Link to MES	Gathered information
08-09-2015	Dina Hadžiosmandovic	Deloitte	Meeting	Expert in SCADA systems also some knowledge of industrial control systems in general	How the link from PLC to SCADA to MES works.
19-09-2015	Tracje Dimkov	Deloitte	Meeting	Expert in SCADA systems and experience with both SCADA and MES in practice	Basics of MES and SCADA, history introduction of MES and developments in the industrial control systems field
22-09-2015	Pieter van Klooster	ISA-95	Email	Creators of the ISA-95 standard	Information about the ISA-95; the use and the content of it.
23-09-2015	Michiel Mennen	Deloitte	Call	Experience with implementing and working with a MES	Information about how a specific MES works and the process of implementation. Also information about why some MES related choices are made by this organization.
29-09-2015	Joost Verbeek	Location "the Edge"	Meeting	Works daily with the building control system of 'the Edge'	A building control system is a SCADA system with a HMI. Information about how this works and what information is gathered provides information about data from the control layer in MES
08-10-2015	André van Barneveld Binkhuysen	Deloitte	Meeting	Expert on the Industry 4.0	Information about developments in Industry 4.0
13-10-2015	Jan Snoeij	CGI	Call	Co-Author of the annual MES product survey	Information about MES, the MES market and how information is gathered, analyzed and returned to the user of MES. Also information about current developments in MES
10-11-2015	Edwin Binnenheim	Eindhoven University of Technology	Meeting	Responsible for building management of the TU/e campus, works with building control system	A building control system is a SCADA system with a HMI. Information about how this works and what information is gathered provides information about data from the control layer in MES
16-11-2015	Alberto Ogura	Deloitte	Call	Expert in Oracle solutions. Mostly in Oracle ERP but also experience with Oracle MES.	Information about how Oracle MES works and MES works together with the enterprise planning layer (of the Oracle ERP product). Also general information about KPI's measured and present in Oracle MES solutions.

17-11-2015	Leon de Groot	SAP	Call	SAP is known as a vendor of ERP systems and also has a MES product named SAP ME	Information about how their MES product works and how a MES works together with the enterprise planning layer (of the SAP ERP product).
19-11-2015	Andre Bokma	Deloitte	Meeting	Expert in SAP ERP systems	Information about how SAP ERP works and the layers close to MES. Also how they interact
25-11-2015	Nico van Veen	MESBuilder	Meeting + email	Former MES advisor and founder of MESBuilder	Information about how MES work and which problems companies encounter when implementing. Emphasis on that MES needs much customization because every company has their own processes , IT and MES need. Also explained the tailor made focus of MESBuilder
05-01-2016	Erik Tenbült	PROMAS ST	Call	PROMAS ST is a MES combined with a portable control system	PROMAS ST works in the animal-food-industry. Information about how the MES works and which information is gathered from the MES data. Also information about how this developed over the years as the animal-food-industry used to be more traditional but is slowly getting more data oriented.

Appendix 2. MESA Metric that matter, 28 manufacturing KPI's

Improving Customer Experience & Responsiveness

1. On-Time Delivery to Commit – This metric is the percentage of time that manufacturing delivers a completed product on the schedule that was committed to customers.

2. Manufacturing Cycle Time – Measures the speed or time it takes for manufacturing to produce a given product from the time the order is released to production, to finished goods.

3. Time to Make Changeovers – Measures the speed or time it takes to switch a manufacturing line or plant from making one product over to making a different product.

Improving Quality

4. Yield – Indicates a percentage of products that are manufactured correctly and to specifications the first time through the manufacturing process without scrap or rework.

5. Customer Rejects/Return Material Authorizations/Returns – A measure of how many times customers reject products or request returns of products based on receipt of a bad or out of specification product. (OUT OF SCOPE)

6. Supplier's Quality Incoming – A measure of the percentage of good quality materials coming into the manufacturing process from a given supplier.

Improving Efficiency

7. Throughput – Measures how much product is being produced on a machine, line, unit, or plant over a specified period of time.

8. Capacity Utilization – Indicates how much of the total manufacturing output capacity is being utilized at a given point in time.

9. Overall Equipment Effectiveness (OEE) – This multi-dimensional metric is a multiplier of Availability x Performance x Quality, and it can be used to indicate the overall effectiveness of a piece of production equipment, or an entire production line.

10. Schedule or Production Attainment – A measure of what percentage of time a target level of production is attained within a specified schedule of time.

Reducing Inventory

11. WIP Inventory/Turns – A commonly used ratio calculation to measure the efficient use of inventory materials. It is calculated by dividing the cost of goods sold by the average inventory used to produce those goods.

Ensuring Compliance

12. Reportable Health and Safety Incidents – A measure of the number of health and safety incidents that were either actual incidents or near misses that were recorded as occurring over a period of time. (OUT OF SCOPE)

13. Reportable Environmental Incidents – A measure of the number of health and safety incidents that were recorded as occurring over a period of time. (OUT OF SCOPE)

14. Number of Non-Compliance Events / Year – A measure of the number of times a plant or facility operated outside the guidelines of normal regulatory compliance rules over a one-year period.

These non-compliances need to be fully documented as to the specific non-compliance time, reasons, and resolutions. (OUT OF SCOPE)

Reducing Maintenance

15. Percentage Planned vs. Emergency Maintenance Work Orders – This ratio metric is an indicator of how often scheduled maintenance takes place, versus more disruptive/un-planned maintenance.

16. Downtime in Proportion to Operating Time – This ratio of downtime to operating time is a direct indicator of asset availability for production.

Increasing Flexibility & Innovation

17. Rate of New Product Introduction – Indicates how rapidly new products can be introduced to the marketplace and typically includes a combination of design, development and manufacturing ramp up times. (OUT OF SCOPE)

18. Engineering Change Order Cycle Time – A measure of how rapidly design changes or modifications to existing products can be implemented all the way through documentation processes and volume production. (OUT OF SCOPE)

Reducing Costs & Increasing Profitability (ALL OUT OF SCOPE)

19. Total Manufacturing Cost per Unit Excluding Materials – This is a measure of all potentially controllable manufacturing costs that go into the production of a given manufactured unit, item or volume.

20. Manufacturing Cost as a Percentage of Revenue – A ratio of total manufacturing costs to the overall revenues produced by a manufacturing plant or business unit.

21. Net Operating Profit – Measures the financial profitability for all investors/shareholders/debt holders, either before or after taxes, for a manufacturing plant or business unit.

22. Productivity in Revenue per Employee – This is a measure of how much revenue is generated by a plant, business unit or company, divided by the number of employees.

23. Average Unit Contribution Margin – This metric is calculated as a ratio of the profit margin that is generated by a manufacturing plant or business unit, divided into a given unit or volume of production.

24. Return on Assets/Return on Net Assets - A measure of financial performance calculated by dividing the net income from a manufacturing plant or business unit by the value of fixed assets and working capital deployed.

25. Energy Cost per Unit – A measure of the cost of energy (electricity, steam, oil, gas, etc.) required to produce a specific unit or volume of production.

26. Cash-to-Cash Cycle Time – This metric is the duration between the purchase of a manufacturing plant or business unit's inventory, and the collection of payments/accounts receivable for the sale of products that utilize that inventory – typically measured in days.

27. EBITDA – This metric acronym stands for Earnings Before Interest, Taxes, Depreciation, and Amortization. It is a calculation of a business unit or company's earnings, prior to having any interest payments, tax, depreciation, and amortization extracted for any final accounting of income and expenses. EBITDA is typically used as top-level indication of the current operational profitability of a business.

28. Customer Fill Rate/On-Time delivery/Perfect Order Percentage - This metric is the percentage of times that customers receive the entirety of their ordered manufactured goods, to the correct specifications, and delivered at the expected time.

Appendix 3. List of databases in the Focus search engine

ABI / Inform	Depatisnet	Narcis
ACM Digital Library	Derwent Innovations Index	National Center for Biotechnology Information
ACM Guide to Computing Literature	Digitale Bibliografie Nederlandse Geschiedenis	Natural product updates
AfricaBib	DRIVER - Digital Repository Infrastructure Vision for European Research	OAISTER
Agricola	EconLit	OpenDOAR
Airbase	Elsevier Science Direct	OPmaat Sdu Wettenbank
American Chemical Society	ERIC	Philosopher's Index
Analytical abstracts	Espacenet	PhilPapers
Arbozone.nl	Essential Science Indicators	ProQuest Annual Reports
Archidat bouwkosten	EthicShare	PsycArticles
Autotechnisch handboek	FOCUS on scientific literature	PsycINFO
Avery Index to Architectural Periodicals	Gartner research library	PubMed
Base - Bielefeld Academic Search Engine	Google Scholar	Reaxys
Basiskaarten Eindhoven, Amsterdam, Rotterdam	GreenFile	Reference Manager
Bedrijfsinformatie NL	Groenekennis	Rehva HVAC Dictionary
Beheer en Onderhoud	Historische collectie CBS	Repository TU/e
Bibliografie van de Nederlandse Taal- en Literatuurwetenschap	Hydrotheek	SAE Digital Library
Biografisch Woordenboek van Nederland	Iconda	SciFinder Scholar
Bouwkosten	Ideas	Scopus
Bouwregels in de praktijk	IEEE-IET Electronic Library	Scripties Online
BRIS Warenhuis	Inorganic Crystal Structure Database (ICSD)	SPIE Digital Library
Catalogus TU/e	Inspec	Springer Journals
Catalysts and Catalysed Reactions	Internet encyclopedia of philosophy	SSRN eLibrary Database
Chemical hazards in industry	Journal Citation Reports (JCR)	Statline
ChemIDplus	JSTOR	TU/e in Beeld
Chemiekaarten Online	Keesings Historisch Archief	Ulrichs XML Data
ChemSpider	Krantenbank	USPTO - Patent Full-Text and Full-Page Image Databases
Chemwatch(GoldFFX)	Laboratory hazards bulletin	Van Dale woordenboeken
CiteseerX	LISTA: Library, Information Science & Technology Abstracts	Web of Science
Collection of Computer Science Bibliographies	MathSciNet	Wijsbegeerte in Nederland
Company.info	Medline	Wiley Online Library
CuminCAD	Mendeley	World Factbook
The DBLP Computer Science Bibliography	Methods in organic synthesis	WorldCat
Delpher boeken tijdschriften kranten	Module Kengetallen	zbMath

Appendix 4. Long list of literature for literature review of 2009 and beyond

The literature search generated a long list of 63 articles. From this list, 33 articles are determined to be within the scope of this research. The articles of the long list and the explanation why they are determined to be within or out of the scope can be found in this appendix.

Nr.	Source	In Scope?	Explanation
1	Abramovici, M., & Lindner, A. (2011). Providing product use knowledge for the design of improved product generations. *CIRP Annals-Manufacturing Technology*, *60*(1), 211–214.	Yes	Focus on improving next generation products by making use of the information in the current generation like faults and quality indications.
2	Archimede, B., Letouzey, A., Memon, M., & Xu, J. (2014). Towards a distributed multi-agent framework for shared resources scheduling. *Journal of Intelligent Manufacturing, 25*(5), 1077–1087.	No	Article proposes a framework for shared resources and information exchange with partners. Not information to uncover from MES
3	Aussem, A., de Morais, S., & Corbex, M. (2012, Jan). Analysis of nasopharyngeal carcinoma risk factors with Bayesian networks. *Artificial intelligence in medicine, 54*(1), 53–62.	No	Field of medicine where patients work in manufacturing
4	Azhar Ramli, A., Watada, J., & Pedrycz, W. (2014). A combination of genetic algorithm-based fuzzy C-means with a convex hull-based regression for real-time fuzzy switching regression analysis: application to industrial intelligent data analysis. *IEEJ Transactions on Electrical and Electronic Engineering*, *9*(1), 71–82.	No	Focus on dealing with heterogeneous data and improve regression with a combination of algorithms for fuzzy switching regression analysis. This can be on any type of information, so not specific enough for the research.
5	Borangiu, T., Raileanu, S., Trentesaux, D., Berger, T., & Iacob, I. (2014). Distributed manufacturing control with extended CNP interaction of intelligent products. *Journal of Intelligent Manufacturing, 25*(5), 1065–1075.	No	Focus on adding local intelligence to either the machines (PLCs) or the products itself. Not adding to MES
6	Brito, P., Soares, C., Almeida, S., Monte, A., & Byvoet, M. (2015). Customer segmentation in a large database of an online customized fashion business. *Robotics and Computer-Integrated Manufacturing, 36*, 93–100.	Yes	Improved understanding of customer segments based on customized product orders and their specifications
7	Burlacu, A., Copot, C., & Lazar, C. (2014). Predictive control architecture for real-time image moments based servoing of robot manipulators. *Journal of Intelligent Manufacturing, 25*(5), 1125–1134.	No	Focus on improving the actual industrial robot.
8	Carpanzano, E., Ferrucci, L., Mandrioli, D., Mazzolini, M., Morzenti, A., & Rossi, M. (2014). Automated formal verification for flexible manufacturing systems. *Journal of Intelligent Manufacturing, 25*(5), 1181–1195.	Yes	Formal verification of the control design of manufacturing system behavior (for example detect errors upfront) in a flexible manufacturing system
9	Casali, A., & Ernst, C. (2012). Discovering correlated parameters in semiconductor manufacturing processes: A data mining approach. *Semiconductor Manufacturing, IEEE Transactions on, 25*(1), 118–127.	Yes	Focus on improving the yield by discovering hidden relationships between numerous complex process control parameters.
10	Charaniya, S., Le, H., Rangwala, H., Mills, K., Johnson, K., Karypis, G., & Hu, W.-S. (2010, Jun). Mining manufacturing data for discovery of high productivity process characteristics. *Journal of biotechnology, 147*(3-4), 186–97.	Yes	Process data-driven knowledge discovery aimed at finding performance parameters. Also able to identify and rank process parameters to their relevance in predicting the process outcomes
11	Chien, C.-F., Chang, K.-H., & Wang, W.-C. (2014). An empirical study of design-of-experiment data mining for yield-loss diagnosis for semiconductor manufacturing. *Journal of Intelligent Manufacturing, 25*(5), 961–972.	Yes	Focus on developing a design of experiment data mining that matches with the potential design with huge amount of automatically collected data. Aims at effective and meaningful knowledge from the data to improve the yield
12	Chien, C.-F., Chen, Y.-J., & Peng, J.-T. (2010). Manufacturing intelligence for semiconductor demand forecast based on technology diffusion and product life cycle. *International Journal of Production Economics, 128*, 496–509.	No	Focus on demand forecasting incorporating seasonal factors, market growth, price, repeat purchase and technology substitution. This is more ERP level and not at MES level
13	Chien, C.-F., Gen, M., Shi, Y., & Hsu, C.-Y. (2014). Manufacturing intelligence and innovation for digital manufacturing and operational excellence. *Journal of Intelligent Manufacturing, 25*(5), 845.	No	Is the introduction of the managing that explains what topic will be addresses later
14	Chien, C.-F., Hsu, C.-Y., & Hsiao, C.-W. (2012). Manufacturing intelligence to forecast and reduce semiconductor cycle time. *Journal of Intelligent Manufacturing, 23*(6), 2281–2294.	Yes	Forecast the cycle time of the production line, with input factors WIP, capacity, average layers, utilization a throughput. Also an adaptive model to respond to changes of the production line status
15	Chien, C.-F., Hsu, S.-C., & Chen, Y.-J. (2013). A system for online detection and classification of wafer bin map defect patterns for manufacturing intelligence. *International Journal of Production Research, 51*(8), 2324–2338.	Yes	MI solution that spatial statistics and neural network for the detection and classification of Wafer Bin Maps patterns. This will enable online monitoring and visualization of failure percentages with corresponding patterns that are causing the failures.
16	Chien, C.-F., Zheng, J.-N., & Lin, Y.-J. (2014). Determining the operator-machine assignment for machine interference problem and an empirical study in semiconductor test facility. *Journal of Intelligent Manufacturing, 25*(5), 899–911.	Yes	A methodology to optimize the assignment of test machines and operator in different product mixes. This to improve the utilization and optimize the systems performance
17	Chou, J.-S., Cheng, M.-Y., Wu, Y.-W., & Tai, Y. (2011). Predicting high-tech equipment fabrication cost with a novel evolutionary SVM inference model. *Expert Systems with Applications, 38*(7), 8571–8579.	No	Focus on predicting the fabrication cost. Though some variables could be present in the MES environment, this contribution is too small for MES related research
18	Chou, J.-S., Tai, Y., & Chang, L.-J. (2010). Predicting the development cost of TFT-LCD manufacturing equipment with artificial intelligence models. *International Journal of Production Economics, 128*, 339–350.	No	Focus on predicting product cost in conceptual stages. Not MES or actual production related yet.
19	Çiflikli, C., & Kahya-Özyirmidokuz, E. (2010). Implementing a data mining solution for enhancing carpet manufacturing productivity. *Knowledge-Based Systems, 23*(8), 783–788.	Yes	Improvement of the manufacturing process by data mining. It detects and predicts behavior like breakdowns
20	Çiflikli, C., & Kahya-Özyirmidokuz, E. (2012). Enhancing product quality of a process. *Industrial Management \& Data Systems, 112*(8), 1181–1200.	Yes	Improve performance of manufacturing quality control by discovering hidden patterns.
21	Davidson, I., & Tayi, G. (2009, sep). Data preparation using data quality matrices for classification mining. *European Journal of Operational Research, 197*(2), 764–772.	No	Aims at data preprocessing for data mining when considering the quality of the database (imprecise database)
22	Di Orio, G., Cândido, G., & Barata, J. (2015). The Adapter module: A building block for Self-Learning Production Systems. *Robotics and Computer-Integrated Manufacturing, 36*, 25–35.	No	Though it focusses on the evolution of manufacturing production system, to reduce faults and improve cycle times. It focusses on the monitoring and control systems which are level 2 software instead of MES
23	Donauer, M., Peças, P., & Azevedo, A. (2015). Identifying nonconformity root causes using applied knowledge discovery. *Robotics and Computer-Integrated Manufacturing, 36*, 84–92.	Yes	Aims to make identifying root causes for nonconformities more simple and agile
24	Hao, X.-C., Wu, J.-Z., Chien, C.-F., & Gen, M. (2014). The cooperative estimation of distribution algorithm: a novel approach for semiconductor final test scheduling problems. *Journal of Intelligent Manufacturing, 25*(5), 867–879.	Yes	Improves semiconductor final test scheduling. This already makes use of data mining techniques though this paper incorporates interdependent relations of group decision making in a complex and large problem with local constrains

25	Hsu, C.-Y. (2014). Integrated data envelopment analysis and neural network model for forecasting performance of wafer fabrication operations. *Journal of Intelligent Manufacturing, 25*(5), 945–960.	Yes	Predict performance based on the results of the present performance by integrating data envelopment and back-propagation neural networks.
26	Huang, C.-Y., & Lin, Y.-H. (2013). Applying CHAID algorithm to investigate critical attributes of void formation in QFN assembly. *Soldering \& Surface Mount Technology, 25*(2), 117–127.	Yes	Diagnosing void formation with causes errors and determine what is causing this by making use of data mining
27	Jia, S., Tang, R., & Lv, J. (2014). Therblig-based energy demand modeling methodology of machining process to support intelligent manufacturing. *Journal of Intelligent Manufacturing, 25*(5), 913–931.	No	The focus is to determine energy demand on machine level. This is not at MES level
28	Kamsu-Foguem, B., Rigal, F., & Mauget, F. (2013). Mining association rules for the quality improvement of the production process. *Expert systems with applications, 40*(4), 1034–1045.	Yes	Improvement of the operations processes by extracting knowledge about for example causes of operation dysfunctions or lost production time by making use of association rule mining
29	Kim, S., Jitpitaklert, W., Park, S.-K., & Hwang, S.-J. (2012). Data mining model-based control charts for multivariate and autocorrelated processes. *Expert Systems with Applications, 39*(2), 2073–2081.	Yes	Active monitoring and detection of abnormal behavior by combining statistical process control (SPC) tools with data mining techniques. This in order to analyze large scale multivariate and auto correlated processes
30	Kubler, S., Derigent, W., Thomas, A., & Rondeau, É. (2014). Embedding data on "communicating materials" from context-sensitive information analysis. *Journal of Intelligent Manufacturing, 25*(5), 1053–1064.	No	Focus on applying "Internet of Things" intelligence to the products itself and propose an information dissemination process.
31	Kwak, D.-S., & Kim, K.-J. (2012, feb). A data mining approach considering missing values for the optimization of semiconductor-manufacturing processes. *Expert Systems with Applications, 39*(3), 2590–2596.	No	Focus on data preprocessing and propose a method to handle missing data for process improvement analysis.
32	Kwong, C., Chan, K., & Tsim, Y. (2009). A genetic algorithm based knowledge discovery system for the design of fluid dispensing processes for electronic packaging. *Expert Systems with Applications, 36*(2), 3829–3838.	Yes	Using a genetic algorithm to gain knowledge of the fluid dispensing process in the form of rules. This enables optimizing the settings for a high-yield environment.
33	Köksal, G., Batmaz, İ., & Testik, M. (2011, sep). A review of data mining applications for quality improvement in manufacturing industry. *Expert Systems with Applications, 38*(10), 13448–13467.	Yes	Review of data mining for quality improvement
34	Lamond, B., Sodhi, M., Noël, M., & Assani, O. (2014). Dynamic speed control of a machine tool with stochastic tool life: analysis and simulation. *Journal of Intelligent Manufacturing, 25*(5), 1153–1166.	No	Optimizing machine equipment by dynamic programming. This is not data mining related
35	Lee, C., Choy, K., Ho, G., Chin, K., Law, K., & Tse, Y. (2013). A hybrid OLAP-association rule mining based quality management system for extracting defect patterns in the garment industry. *Expert Systems with Applications, 40*(7), 2435–2446.	Yes	Real-time hybrid OLAP association rule mining to detect patterns for quality failure with a root cause analysis, quality prediction and formulation of pro-active measures
36	Legat, C., Schütz, D., & Vogel-Heuser, B. (2014). Automatic generation of field control strategies for supporting (re-) engineering of manufacturing systems. *Journal of Intelligent Manufacturing, 25*(5), 1101–1111.	No	Focus on the adaptability of the process in open-loop control software in manufacturing. Not about information acquiring
37	Li, C.-D., Xie, T., & Tang, Y.-L. (2014). GMVN oriented S-BOX knowledge expression and reasoning framework. *Journal of Intelligent Manufacturing, 25*(5), 993–1011.	No	Focus is on global manufacturing market on a level higher than the MES level. Demand and supply chain as a whole focus.
38	Liang, C.-J., Chen, M., Gen, M., & Jo, J. (2014). A multi-objective genetic algorithm for yard crane scheduling problem with multiple work lines. *Journal of Intelligent Manufacturing, 25*(5), 1013–1024.	No	Scheduling and planning related
39	Liu, Y., & Harding, J. (2009). Editorial for the special issue of knowledge discovery and management in engineering design and manufacturing. *Journal of Intelligent Manufacturing, 20*(5), 499–500.	No	Editorial of the magazine so not a specific research
40	Moharana, U., & Sarmah, S. (2015). Determination of optimal kit for spare parts using association rule mining. *International Journal of System Assurance Engineering and Management, 6*(3), 238–247.	Yes	Optimize associated spare parts mix for corrective or preventive maintenance
41	Mozafary, V., & Payvandy, P. (2014). Application of data mining technique in predicting worsted spun yarn quality. *The Journal of The Textile Institute, 105*(1), 100–108.	Yes	Prediction of spun yarn quality by data mining
42	Negahban, A., & Smith, J. (2014, December). Simulation for manufacturing system design and operation: Literature review and analysis. *Journal of Manufacturing Systems, 33*(2), 241–261.	No	About discrete event simulation. The data mining that is present, is used on planning and scheduling which is more ERP related
43	Perzyk, M., Kochanski, A., Kozlowski, J., Soroczynski, A., & Biernacki, R. (2014). Comparison of data mining tools for significance analysis of process parameters in applications to process fault diagnosis. *Information Sciences, 259*, 380–392.	No	Focus on determining the relative significance of input variables. Though this can be useful for fault diagnosis, the paper focusses on the technical aspect of the input variables and not on the actual fault diagnosis.
44	Polczynski, M., & Kochanski, A. (2010). Knowledge Discovery and Analysis in Manufacturing. *Quality Engineering, 22*(3), 169–181.	Yes	Article about knowledge discovery and analysis in manufacturing. The current statement, future and deployment is explained with examples.
45	Roy, R., Shehab, E., Tiwari, A., Mey Goh, Y., & McMahon, C. (2009). Improving reuse of in-service information capture and feedback. *Journal of Manufacturing Technology Management, 20*(5), 626–639.	No	Article is about knowledge management systems and how the management of information should be within a manufacturing organization, not about the specific knowledge to be discovered.
46	Russell, B., Shapiro, D., & Vining, A. (2010). The evolution of the Canadian mining industry: The role of regulatory punctuation. *Resources Policy, 35*, 90–97.	No	Article about the Canadian mining Industry. Not manufacturing execution, or data mining related.
47	Sajadfar, N., & Ma, Y. (2015). A hybrid cost estimation framework based on feature-oriented data mining approach. *Advanced Engineering Informatics, 29*(3), 633–647.	No	Focus is on data associated with ERP systems and not MES.
48	Sanders, D., & Gegov, A. (2013). AI tools for use in assembly automation and some examples of recent applications. *Assembly Automation, 33*(2), 184–194.	Yes	Artificial intelligence (AI) tools for assembly automation. Some data mining techniques are used, as well as execution data
49	Stockton, D., Khalil, R., & Mukhongo, M. (2013). Cost model development using virtual manufacturing and data mining: part I—methodology development. *The International Journal of Advanced Manufacturing Technology, 66*(5-8), 741–749.	Yes	The automation of het identification of the virtual manufacturing process time is the focus of this article. This is later used as a basis for the process cost model. They use manufacturing process parameters and apply data mining techniques.
50	Tirkel, I. (2013). Forecasting flow time in semiconductor manufacturing using knowledge discovery in databases. *International Journal of Production Research, 51*(18), 5536–5548.	Yes	Flow time (noted cycle time) forecasting by making use of data from MES and data mining techniques.
51	Uchino, E., Koga, T., Misawa, H., & Suetake, N. (2014). Tissue characterization of coronary plaque by kNN classifier with fractal-based features of IVUS RF-signal. *Journal of Intelligent Manufacturing, 25*(5), 973–982.	No	Not manufacturing related but about a syndrome present in human patients.
52	Vin, E., & Delchambre, A. (2014). Generalized cell formation: iterative versus simultaneous resolution with grouping genetic algorithm. *Journal of intelligent manufacturing, 25*(5), 1113–1124.	Yes	Algorithm for allocation of machines to operations of machine grouping into cells. This to analyze the choice of iterative of simultaneous resolution.
53	Visintin, F., Porcelli, I., & Ghini, A. (2014). Applying discrete event simulation to the design of a service delivery system in the aerospace industry: a case study. *Journal of Intelligent Manufacturing, 25*(5), 1135–1152.	No	Article is focused on simulation of a service delivery system for a long term service contract by making use of Monte Carlo simulation.
54	Wu, C.-H., Wang, D.-Z., Ip, A., Wang, D.-W., Chan, C.-Y., & Wang, H.-F. (2009). A particle swarm optimization approach for components placement inspection on printed circuit boards. *Journal of Intelligent Manufacturing, 20*(5), 535–549.	No	A rectification with an added acknowledgement. Not the full article

55	Wuest, T., Irgens, C., & Thoben, K.-D. (2014). An approach to monitoring quality in manufacturing using supervised machine learning on product state data. *Journal of Intelligent Manufacturing, 25*(5), 1167–1180.	Yes	Describing a product's state and corresponding characteristics during the entire process step in order to increase quality
56	Xiao, J., & Huang, Y. (2011). Design of Sustainable Multifunctional Nanocoatings: A Goal-driven Multiscale Systems Approach. *Chinese Journal of Chemical Engineering, 19*(4), 666–673.	No	Article focusses on the design on sustainable multifunctional Nano coatings from microscopic molecular modeling to classical continuum modeling. Data mining is not included to gain knowledge.
57	Yeh, C.-W., Li, D.-C., & Zhang, Y.-R. (2012). Estimation of a data-collection maturity model to detect manufacturing change. *Expert Systems with Applications, 39*(8), 7093–7101.	Yes	Data collection maturity model that threat the data in three phases, instead of as one big population. The article focuses on determining the two critical points where the phases split by making use of neural networks. This improves understanding of the manufacturing process
58	Yeh, D.-Y., Cheng, C.-H., & Hsiao, S.-C. (2011). Classification knowledge discovery in mold tooling test using decision tree algorithm. *Journal of Intelligent Manufacturing, 22*(4), 585–595.	Yes	Article is about creating a classification method by making use of knowledge discovery in the mold tooling test. By making use of a decision tree knowledge was gained and parameters where adjusted for improving the classification in the test.
59	Yu, H.-C., Lin, K.-Y., & Chien, C.-F. (2014). Hierarchical indices to detect equipment condition changes with high dimensional data for semiconductor manufacturing. *Journal of Intelligent Manufacturing, 25*(5), 933–943.	No	Article is focused on establishing a framework for real-time equipment monitoring by the decrease of the hierarchical indices. The focus on on the hierarchy and the framework and not the actual data analysis. However a point is made that their framework could improve (predictive)maintenance policies and root cause detection
60	Yu, Q., & Wang, K. (2013). 3D vision based quality inspection with computational intelligence. *Assembly Automation, 33*(3), 240–246.	Yes	Focus on accurate classification of products with a 3D vision method to prevent low quality product to go to the customer.
61	Zapcevic, , & Butala, . (2013). Adaptive process control based on a self-learning mechanism in autonomous manufacturing systems. *The International Journal of Advanced Manufacturing Technology, 66*(9-12), 1725--1743.	Yes	Research aimed at knowledge discovery in large real time operating data bases of manufacturing organizations like MES and SCADA databases.
62	Zhang, M., Miesegaes, G., Lee, M., Coleman, D., Yang, B., Trexler-Schmidt, M., . . . Chen, Q. (2014, Jan). Quality by design approach for viral clearance by protein a chromatography. *Biotechnology and bioengineering, 111*(1), 95--103.	No	About modifying a virus for medical purposes
63	Zhang, W., Gen, M., & Jo, J. (2014). Hybrid sampling strategy-based multiobjective evolutionary algorithm for process planning and scheduling problem. *Journal of Intelligent Manufacturing, 25*(5), 881–897.	Yes	Improved process planning and scheduling by making use of hybrid sampling strategy -based multi objective evolutionary algorithms

Appendix 5. Overview of all informational items

In article number refers to the article number given in Appendix 4.

Article No.	Short reference	Information uncovered	Used method	Main informational (purpose) group	Informational sub-group
X	Choudhary et al. (2009)	Online monitoring of causal relationship between process parameters and output quality	Integrated Neural Networks and rough set techniques (other article extended with fuzzy set theory)	Condition based monitoring	Monitoring of parameters settings and their effects
X	Choudhary et al. (2009)	Online monitoring of causal relationship between process parameters and output quality (extended)	Fuzzy set theory with fuzzy variable rough set	Condition based monitoring	Monitoring of parameters settings and their effects
X	Choudhary et al. (2009)	Monitoring process conditions by classification	Hybrid fuzzy inductive learning	Condition based monitoring	Monitoring process conditions
X	Choudhary et al. (2009)	Offline tool wear monitoring	Rough set theory classifier	Condition based monitoring	Monitoring tool wear
X	Choudhary et al. (2009)	Tool wear condition monitoring	Neural Networks and Support Vector Machines	Condition based monitoring	Monitoring tool wear
X	Choudhary et al. (2009)	Lead time prediction	Regression tree based data mining approach	Cycle/lead time prediction	Forecasting lead time
X	Choudhary et al. (2009)	Lead time prediction	Decision tree combined with if-then-else rules	Cycle/lead time prediction	Forecasting lead time
X	Choudhary et al. (2009)	Cycle time prediction	Set of data mining tools (multiple articles extended)	Cycle/lead time prediction	Forecasting production cycle time
X	Choudhary et al. (2009)	Analysis of the effects of decision making	Genetic algorithms	Decision support	Insights in the effect of to-be-made decisions
X	Choudhary et al. (2009)	Decision support for workflow related decisions	Workflow mining by Artificial Neural Networks and fuzzy rule sets	Decision support	Insights in the effect of to-be-made decisions
X	Choudhary et al. (2009)	Online/Real-Time classification of quality faults	Integrated neural network and rough set techniques	Defect/low quality classification	Classification of product quality
X	Choudhary et al. (2009)	Automatic defect classification to find patterns and derive rules for yield improvement	Set of data mining tools	Defect/low quality classification	Classification of product quality
X	Choudhary et al. (2009)	Automatic classification of defect patterns	Hierarchical clustering, k-means partitioning	Defect/low quality classification	Classification of product quality
X	Choudhary et al. (2009)	Real-Time classification of product quality	Hybrid learning based system with Neural Networks and decision tree	Defect/low quality classification	Classification of product quality
X	Choudhary et al. (2009)	Defect product detection	Clustering by self-organizing maps for classification	Defect/low quality classification	Detection of a product with quality faults
X	Choudhary et al. (2009)	Flaws in product detection	Fuzzy k- & c-means clustering	Defect/low quality classification	Detection of a product with quality faults
X	Choudhary et al. (2009)	Fault detection in assembly operations	Association rule mining	Defect/low quality classification	Detection of a product with quality faults
X	Choudhary et al. (2009)	Determine the defective machine in a set of machines	Association rules	Identification of machine Failure	Identification of machine failure
X	Choudhary et al. (2009)	Accurate grading of materials	Combination of rule based knowledge representation, fuzzy logic and genetic algorithms	Knowledge of optimal manufacturing settings	Grading of (raw)materials
X	Choudhary et al. (2009)	Improved dispatching rules	Genetic algorithms	Knowledge of optimal manufacturing settings	Improved dispatching rules
X	Choudhary et al. (2009)	Improved dispatching rules	Decision tree based classifications rules	Knowledge of optimal manufacturing settings	Improved dispatching rules
X	Choudhary et al. (2009)	Combine different databases for improved knowledge extraction	Integrated relational databases approach	Knowledge of optimal manufacturing settings	Knowledge of operational process(es)
X	Choudhary et al. (2009)	Establishing rules to support manufacturing system	data mining with learning classifier	Knowledge of optimal manufacturing settings	Knowledge of operational process(es)
X	Choudhary et al. (2009)	Improved understanding of the processing method by automatic detection and recovery of a process flaw and provide information about the recovered flaw	Two-stage data mining approach	Knowledge of optimal manufacturing settings	Knowledge of operational process(es)
X	Choudhary et al. (2009)	Improved understanding of the cleaning process	Decision tree induction, neural network and composite classifier	Knowledge of optimal manufacturing settings	Knowledge of operational process(es)
X	Choudhary et al. (2009)	Prediction of product parameters for improved quality	Set of data mining tools	Knowledge of optimal manufacturing settings	Optimization of parameter settings

X	Choudhary et al. (2009)	Optimization of factory conditions based on extracted knowledge	Genetic Algorithms and Neural Networks	Knowledge of optimal manufacturing settings	Optimization of parameter settings
X	Choudhary et al. (2009)	Identify operational spaces to optimize manufacturing process and minimize lost	Fuzzy c-means clustering	Knowledge of optimal manufacturing settings	Optimization of parameter settings
X	Choudhary et al. (2009)	Identification of poor yield factors	Self-organizing maps, Neural Networks and rule induction	Low yield factors identification	Identification of characteristics for low yield (product quality failure)
X	Choudhary et al. (2009)	Automated discovery of factors that cause low yield	Genetic programming	Low yield factors identification	Identification of characteristics for low yield (product quality failure)
X	Choudhary et al. (2009)	Identify causes of defect products	Rough set theory	Low yield factors identification	Identification of characteristics for low yield (product quality failure)
X	Choudhary et al. (2009)	Significant features that cause quality control issues	Rough set theory	Low yield factors identification	Identification of characteristics product quality
X	Choudhary et al. (2009)	Predict component failure	Set of data mining tools (Decision trees, rough sets, regression and Neural Networks)	Machine (component) failure prediction	Forecasting component failure
X	Choudhary et al. (2009)	Predicting (machine) faults	Decision tree	Machine (component) failure prediction	Forecasting machine/equipment failure
X	Choudhary et al. (2009)	Prediction of equipment failure	Recurrent Neural Networks model	Machine (component) failure prediction	Forecasting machine/equipment failure
X	Choudhary et al. (2009)	Prediction of equipment failure	Agent based model and data mining tools for prediction	Machine (component) failure prediction	Forecasting machine/equipment failure
X	Choudhary et al. (2009)	Prediction of tool wear	Rough set theory based classifier	Machine (component) failure prediction	Forecasting tool wear
X	Choudhary et al. (2009)	Machine performance prediction	Neural Networks based estimation model	Machine (component) failure prediction	Machine performance prediction
X	Choudhary et al. (2009)	Preventive maintenance schedule recommendations	Decision tree based data mining	Machine (component) failure prediction	Preventive maintenance schedule recommendations
X	Choudhary et al. (2009)	Probability of machine failure or product failure	Classification by decision tree	Machine (component) failure prediction	Probability for machine failure
X	Choudhary et al. (2009)	Distinguish fault types for machines	Rough set theory approach	Machine fault diagnostics	Classification of machine fault types
X	Choudhary et al. (2009)	(Machine) fault diagnosis	Hybrid case based reasoning	Machine fault diagnostics	Diagnostics of machine failure
X	Choudhary et al. (2009)	Fault analysis to know where to focus attention when repairing	Data mining approach for concept description	Machine fault diagnostics	Diagnostics of machine failure
X	Choudhary et al. (2009)	Fault diagnosis reporting	Hybrid rough set theory and a genetic algorithm	Machine fault diagnostics	Diagnostics of machine failure
X	Choudhary et al. (2009)	Patterns for (machine) failure	Decision theoretic approach to mine the data combined with greedy value for information	Machine fault diagnostics	Identification of characteristics of machine failure
X	Choudhary et al. (2009)	Discovery of typical unnatural control chart patterns causing process variation	Hybrid neural network and decision tree	Patterns causing process variations	Detection of abnormal process behavior
X	Choudhary et al. (2009)	Identify classes of process faults	Decision tree classification	Patterns causing process variations	Identification of process fault classes
X	Choudhary et al. (2009)	Process fault classification	Metric Temporal Logic	Patterns causing process variations	Identification of process fault classes
X	Choudhary et al. (2009)	Prediction of the probability of performance (and optimal settings of control factors)	Bayesian method	Process performance prediction	Forecasting production process performance
X	Choudhary et al. (2009)	Prediction of the performance of the manufacturing process	Model selection and cross-validation	Process performance prediction	Forecasting production process performance
X	Choudhary et al. (2009)	Prediction of system output	Data mining and type-II fuzzy system	Process performance prediction	Prediction of system output
X	Choudhary et al. (2009)	Detection of change points in control charts	Tree based supervised learner	Root Cause analysis	Detection of change points in control charts
X	Choudhary et al. (2009)	Root causes for failure in a process stage	Bayesian network, Design of Experiment and Statical Process Control	Root Cause analysis	Root cause analysis for process failure
X	Choudhary et al. (2009)	Discovery of unnatural patterns in process data	Fractal dimension based classifier	Root Cause analysis	Root cause analysis for unnatural patterns in the data
X	Choudhary et al. (2009)	Prediction of quality of a product or batch based on the manufacturing parameters	Feature set decomposition methodology based algorithm	Yield/Low quality prediction	Prediction of product quality

X	Choudhary et al. (2009)	Yield prediction	Genetic programming	Yield/Low quality prediction	Yield prediction
X	Choudhary et al. (2009)	Yield predication	Decision trees and Neural Networks	Yield/Low quality prediction	Yield prediction
1	Abramovici, M., & Lindner, A. (2011)	Product improvements (reduce faults in next generation products)	Bayesian Networks	Low yield factors identification	Suggested improvements for next generations based on quality failure
8	Carpanzano, E., Ferrucci, L., Mandrioli, D., Mazzolini, M., Morzenti, A., & Rossi, M. (2014)	Forecasting machine errors	Metric Temporal Logic	Machine (component) failure prediction	Forecasting machine/equipment failure
8	Carpanzano, E., Ferrucci, L., Mandrioli, D., Mazzolini, M., Morzenti, A., & Rossi, M. (2014)	Forecasting of manufacturing system behavior	Metric Temporal Logic	Process performance prediction	Forecasting of manufacturing process behavior
9	Casali, A., & Ernst, C. (2012)	Identification of critical factors for low yield	Decision correlation rules and contingency vectors.	Low yield factors identification	Identification of characteristics for low yield (product quality failure)
10	Charaniya, S., Le, H., Rangwala, H., Mills, K., Johnson, K., Karypis, G., & Hu, W.-S. (2010, Jun)	Identification of critical process parameters	kernel-based approach combined with a maximum margin- based support vector regression algorithm	Patterns causing process variations	Identification of critical process parameters
11	Chien, C.-F., Chang, K.-H., & Wang, W.-C. (2014)	Improve the yield	Design of experiment data mining	Low yield factors identification	Identification of characteristics for low yield (product quality failure)
14	Chien, C.-F., Hsu, C.-Y., & Hsiao, C.-W. (2012)	Forecasting of production cycle time	Gauss-Newton regression method and back-propagation neural network	Cycle/lead time prediction	Forecasting production cycle time
15	Chien, C.-F., Hsu, S.-C., & Chen, Y.-J. (2013)	Patterns for product failure	Spatial statistics with neural networks	Low yield factors identification	Identification of characteristics for low yield (product quality failure)
16	Chien, C.-F., Zheng, J.-N., & Lin, Y.-J. (2014)	Optimizing utilization by operator to machine assignment/schedule optimization	Genetic algorithms	Decision support	Improved scheduling decisions by insights options and effects
19	Çiflikli, C., & Kahya-Özyirmidokuz, E. (2010)	Predicting machine Breakdowns	Decision tree	Machine (component) failure prediction	Forecasting machine/equipment failure
19	Çiflikli, C., & Kahya-Özyirmidokuz, E. (2010)	Predict manufacturing process behavior	Decision tree	Process performance prediction	Forecasting of manufacturing process behavior
20	Çiflikli, C., & Kahya-Özyirmidokuz, E. (2012)	Patterns for low quality	Rough sets theory, attribute relevance analysis, anomaly detection analysis, decision trees and rule induction	Low yield factors identification	Identification of characteristics product quality
23	Donauer, M., Peças, P., & Azevedo, A. (2015)	Root causes for nonconformities in the production process	Herfindahl–Hirschman Index (HHI)	Root Cause analysis	Root cause analysis of nonconformities in the production process
24	Hao, X.-C., Wu, J.-Z., Chien, C.-F., & Gen, M. (2014)	Improved final test execution scheduling with multi resources and effect of interdependent relationships in group decision making activities	Cooperative estimation of contribution algorithm	Decision support	Improved scheduling decisions by insights options and effects
25	Hsu, C.-Y. (2014)	Predict process performance	data envelopment and back-propagation neural network	Process performance prediction	Forecasting production process performance
26	Huang, C.-Y., & Lin, Y.-H. (2013)	Root cause for product failure	Chi-square automatic interaction detection (CHAID) algorithm and chi-square test.	Low yield factors identification	Identification of characteristics for low yield (product quality failure)
28	Kamsu-Foguem, B., Rigal, F., & Mauget, F. (2013)	Knowledge of operations and information management for the improvement of operations processes	Association rule mining	Knowledge of optimal manufacturing settings	Knowledge of operational process(es)
29	Kim, S., Jitpitaklert, W., Park, S.-K., & Hwang, S.-J. (2012)	Detection of abnormal behavior	SPC combined with artificial neural networks, support vector regression and multivariate adaptive regression splines	Patterns causing process variations	Detection of abnormal process behavior
32	Kwong, C., Chan, K., & Tsim, Y. (2009)	Yield improvement (process knowledge of how the product reacts on system	Genetic algorithms	Low yield factors identification	Identification of characteristics for low yield (product quality failure)

		settings)			
33	Köksal, G., Batmaz, İ., & Testik, M. (2011, sep)	Classification of quality	Range of DM algorithms	Defect/low quality classification	Classification of product quality
33	Köksal, G., Batmaz, İ., & Testik, M. (2011, sep)	Parameter optimization	Range of DM algorithms	Knowledge of optimal manufacturing settings	Optimization of parameter settings
33	Köksal, G., Batmaz, İ., & Testik, M. (2011, sep)	Characteristics for product quality identification	Range of DM algorithms	Low yield factors identification	Identification of characteristics product quality
33	Köksal, G., Batmaz, İ., & Testik, M. (2011, sep)	Prediction of product quality	Range of DM algorithms	Yield/Low quality prediction	Prediction of product quality
35	Lee, C., Choy, K., Ho, G., Chin, K., Law, K., & Tse, Y. (2013)	Identification of patterns for quality failure	Hybrid OLAP-association rule	Low yield factors identification	Identification of characteristics for low yield (product quality failure)
35	Lee, C., Choy, K., Ho, G., Chin, K., Law, K., & Tse, Y. (2013)	Root Cause analysis (product quality)		Root Cause analysis	Root cause analysis of product quality
35	Lee, C., Choy, K., Ho, G., Chin, K., Law, K., & Tse, Y. (2013)	Quality predictions	Hybrid OLAP-association rule	Yield/Low quality prediction	Prediction of product quality
40	Moharana, U., & Sarmah, S. (2015)	Machine part wear and correlations between parts	association rule mining	Machine fault diagnostics	Diagnostics of machine part wear and correlations between parts
41	Mozafary, V., & Payvandy, P. (2014)	Quality prediction	Clustering and Artificial Neural Networks	Yield/Low quality prediction	Prediction of product quality
44	Polczynski, M., & Kochanski, A. (2010)	Identification of critical process parameters	Regression, Classification, Clustering	Knowledge of optimal manufacturing settings	Identification of critical process parameters
44	Polczynski, M., & Kochanski, A. (2010)	Prediction of effects of manufacturing process changes	Regression, Classification, Clustering	Knowledge of optimal manufacturing settings	Optimization of parameter settings
44	Polczynski, M., & Kochanski, A. (2010)	Prediction of equipment breakdowns	Regression, Classification, Clustering	Machine (component) failure prediction	Forecasting machine/equipment failure
44	Polczynski, M., & Kochanski, A. (2010)	Root Cause detection	Regression, Classification, Clustering	Root Cause analysis	General Root cause analysis
48	Sanders, D., & Gegov, A. (2013)	Improved methods for automation in the assembly	Range of Artificial Intelligence tools	Knowledge of optimal manufacturing settings	Improved methods for a specific process
49	Stockton, D., Khalil, R., & Mukhongo, M. (2013)	Identification of the process cycle time	Stepwise linear regression and symbolic knowledge acquisition technology	Cycle/lead time prediction	Forecasting production cycle time
50	Tirkel, I. (2013)	Flow time (cycle time) predictions	Classification (decision tree and NN)	Cycle/lead time prediction	Forecasting production cycle time
52	Vin, E., & Delchambre, A. (2014)	Improved insight in scheduling options for cellular manufacturing	Genetic algorithms	Decision support	Improved scheduling decisions by insights options and effects
55	Wuest, T., Irgens, C., & Thoben, K.-D. (2014)	Increase quality by describing the product state in each step	Cluster analysis and supervised machine learning	Defect/low quality classification	Product state diagnosis
57	Yeh, C.-W., Li, D.-C., & Zhang, Y.-R. (2012)	Improved understanding of the manufacturing process and interdependencies	Classification by Neural Networks and decision tree	Knowledge of optimal manufacturing settings	Knowledge of operational process(es)
58	Yeh, D.-Y., Cheng, C.-H., & Hsiao, S.-C. (2011)	Classification of product quality	Classification (decision tree)	Defect/low quality classification	Classification of product quality
58	Yeh, D.-Y., Cheng, C.-H., & Hsiao, S.-C. (2011)	Improve suggested parameter settings	Classification (decision tree)	Knowledge of optimal manufacturing settings	Optimization of parameter settings
60	Yu, Q., & Wang, K. (2013)	Classification of product quality	decision tree, artificial neural network and support vector machines	Defect/low quality classification	Classification of product quality
61	Zapcevic, , & Butala, . (2013)	Support decision making for adaptive process control	Knowledge discovery for databases	Decision support	Insights in the effect of to-be-made decisions
63	Zhang, W., Gen, M., & Jo, J. (2014)	Improved process planning and scheduling	Evolutionary algorithms combined with hybrid planning	Decision support	Improved scheduling decisions by insights options and effects

Appendix 6. The field expert survey 1 questions

Target audience

The target audience of the surveys are MES field experts. In order to reach this group, the surveys are created as an online form and posted in MES related LinkedIn discussion groups. In these groups people are active who are working or are interested in the field of MES. The targeted groups are:

- LinkedIn group: MESA International (2.046 members)
- LinkedIn group: MES – Manufacturing Execution Systems (16.586 members)

Survey questions

The survey starts with classification questions for the respondents. This is added to research if there are differences between respondent groups. The classification questions are:

- Please classify yourself.
    - Member of ISA(-95)
    - Member of MESA
    - User of a MES
    - (working for) MES vendor
    - Interested in the field of MES
    - Other:
- From what country are you?
    - Open answer
- What is your age?
    - 0 - 30 years old
    - 30 - 40 years old
    - 40 -50 years old
    - 50 - 60 years old
    - 60 + years old

Next, the respondents are asked first to what extent they are familiar with MES related terms. They could answer on a scale from 1 to 5. The MES related terms asked are:

- The ANSI/ISA95
- The MESA Model
- MES functionality for Production management
- MES functionality for Maintenance management
- MES functionality for Quality management
- MES functionality for Inventory management

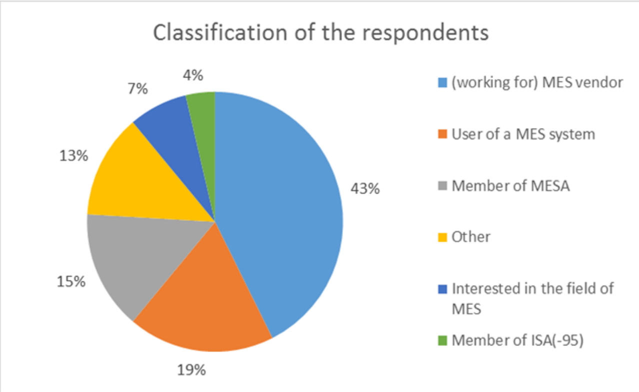
Then, the respondents are asked what about what they believe is used the most in practice. First generally with area (production, maintenance, quality or inventory) and next for each of these areas the individual information blocks.

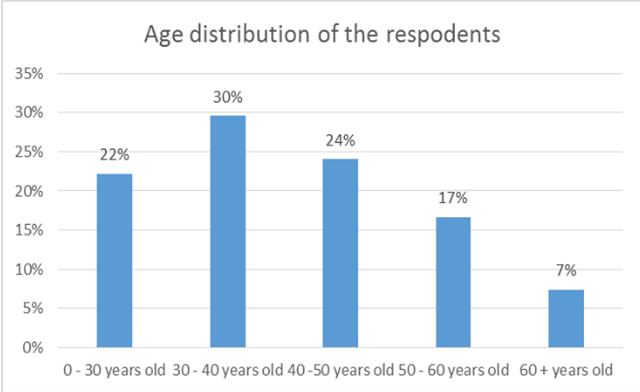
Last, the respondents are asked what they believe Big Data will bring operations management. The options respondents can chose from are both based on section 4.2 as Big Data expectation for the MES Annual survey by Iskamp and Snoeij (2015).

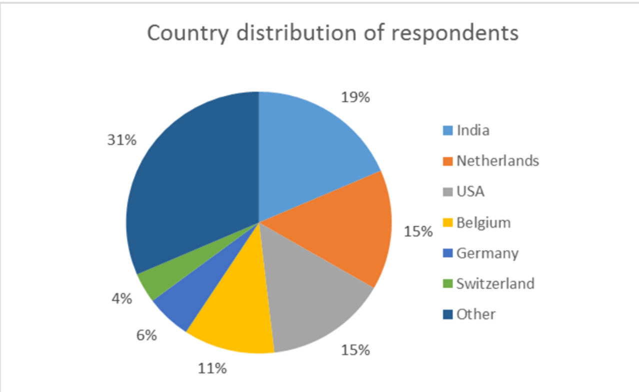
Appendix 7. The field expert survey 1 answers

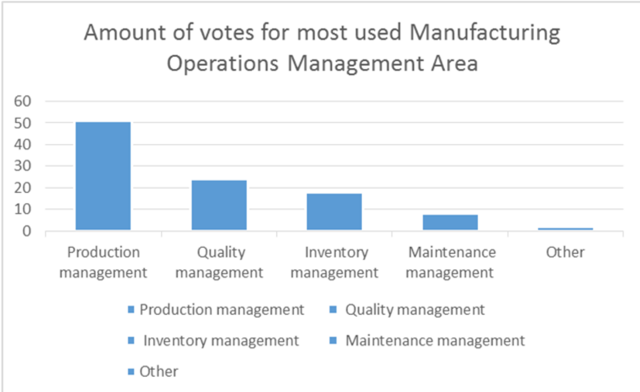
General Information

- Answers where gathered during the time period: 24-12-2015 to 13-3-2016
- The number of respondents is 54





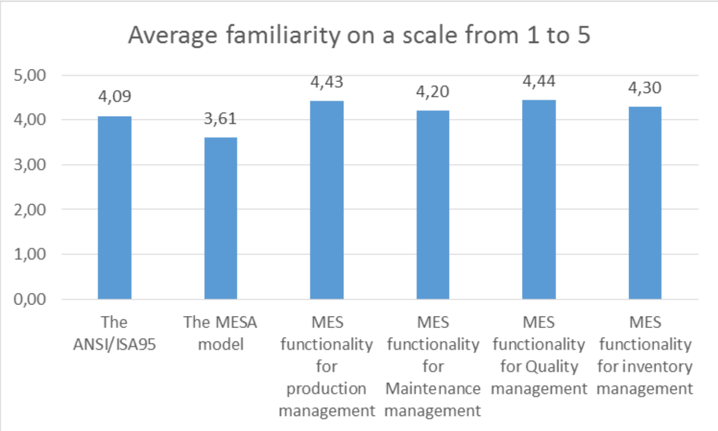




In the classification of respondents, the respondents who ticket the box 'Other' added the following categories:

- MES Team leader
- System Integrator
- MES Project Manager
- Consultant

Familiarity of MES terms and topics



These scores can be segmented among the groups.

	Count	The ANSI/ISA95	The MESA model	MES functionality for production management	MES functionality for Maintenance management	MES functionality for Quality management	MES functionality for inventory management
Average	54	4,09	3,61	4,43	4,20	4,44	4,30
0 - 30 years old	12	3,00	2,75	4,17	4,17	4,33	4,17
30 - 40 years old	16	4,13	3,06	4,19	3,94	4,38	4,25
40 -50 years old	13	4,46	4,31	4,69	4,31	4,62	4,38
50 - 60 years old	9	4,56	4,33	4,56	4,56	4,56	4,56
60 + years old	4	5,00	4,50	5,00	4,25	4,25	4,00
(working for) MES vendor	23	4,35	3,74	4,57	4,17	4,52	4,26
User of a MES	10	3,10	2,30	3,60	3,70	4,30	4,00
Member of MESA	8	4,88	4,63	4,88	4,38	4,38	4,63
Interested in the field of MES	4	3,00	3,00	3,75	3,75	3,50	3,50
Member of ISA(-95)	2	5,00	5,00	5,00	5,00	5,00	5,00
Other	7	4,14	3,86	4,86	4,86	4,86	4,71
India	10	3,60	3,40	3,80	3,80	3,90	4,00
Netherlands	7	4,63	4,00	4,88	4,13	4,63	4,50
USA	7	3,88	3,63	4,25	4,13	4,00	4,00
Belgium	6	4,17	3,67	3,67	3,67	4,33	4,17
Germany	3	4,67	4,67	5,00	4,67	5,00	4,67
Switzerland	2	5	4,5	5	5	5	5
Belarus	1	4	2	5	5	5	5
Singapore	1	4	5	4	4	4	4
New Zealand	1	5	3	5	3	5	3
France	1	5	5	5	5	5	5
Qatar	1	5	1	5	4	5	5
UK	1	5	5	5	5	4	4
Tunisia	1	5	5	5	5	5	5
Italy	1	3	2	5	5	5	4

Czech Republic	1	4	1	5	5	5	5
Denmark	1	5	5	5	5	5	5
Israel	1	5	4	5	5	5	5
Greece	1	1	1	5	5	5	5
South Africa	1	4	4	4	4	4	4
Zenith	1	4	4	4	4	5	4
Indonesia	1	2	1	5	5	5	3
France	1	3	3	4	4	4	4
Unknow n	2	4	4	5	4	5	5

Informational items per Manufacturing Operations Management Area

Production Operations Management

Informational Item	#	%
Resource traceability (material, equipment, personnel & forward and backward)	42	78%
Operational Equipment Efficiency (OEE)	41	76%
Work in Process (WIP) data	36	67%
Real-Time status of plant and production	35	65%
Resource/Equipment performance	31	57%
Production variability	30	56%
Resource/Equipment utilization	25	46%
Scheduled time target performance	25	46%
Throughput	23	43%
Material compatibility & availability	22	41%
Production unit cycle times	22	41%
Root cause analysis	22	41%
Weight and dispense support	20	37%
Notification management	16	30%
Personnel tracking	14	26%
Tracking non-productive activities	12	22%
Other	9	17%
None of the above	2	4%

The items mentioned as 'Other' where: Data Integrity, yield waste and quality/ supplier contract /support of logistic feedback, Production planning, Product safety, Data Acquisition (measurements etc.), this is all dependent on the workflow, waste and rework, and IPC alarms.

Maintenance operations management

Informational Item	#	%
Downtime in proportion to operating time	33	61%
Status of equipment and maintenance schedule	32	59%
Percentage planned versus emergency maintenance	23	43%
Status of materials	23	43%

Status of assets and maintenance schedule	16	30%
Status of maintenance personnel	7	13%
Other	5	9%
None of the above	3	6%

The items mentioned as 'Other' where: Status of orders, Electronic Batch Records, all dependent on the workflow, status of batches/lots

Quality Operations Management

Informational Item	#	%
Quality variability and deviations	36	67%
Yield analysis	31	57%
Batch quality trend analysis	28	52%
Resource traceability analysis	25	46%
Quality indicator analysis	22	41%
Quality department/ operations cycle time	19	35%
Quality equipment utilization	17	31%
Quality resource utilization	13	24%
Other	5	9%
None of the above	3	6%

The items mentioned as 'Other' where: Seasonal quality analysis, SPC, SPC Six Sigma, Quality traceability, al dependent on the workflow

Inventory Operations Management

Informational Item	#	%
Inventory movement tracking	34	63%
Received materials quality and time	28	52%
Inventory efficiency	24	44%
Inventory waste analysis	20	37%
Inventory resource usage	14	26%
Other	4	7%
None of the above	3	6%

The items mentioned as "Other' where: Inventory turnover, finished good stock, shipping and handling including logistics, carrying cost of inventory, al dependent of the workflow.

Appendix 8. The field expert survey 2 questions

Target audience

The target audience of the surveys are MES field experts. In order to reach this group, the surveys are created as an online form and posted in MES related LinkedIn discussion groups. In these groups people are active who are working or are interested in the field of MES. The targeted groups are:

- LinkedIn group: MESA International (2.046 members)
- LinkedIn group: MES – Manufacturing Execution Systems (16.586 members)

Survey questions

The survey starts with classification questions for the respondents. This is added to research if there are differences between respondent groups. The classification questions are:

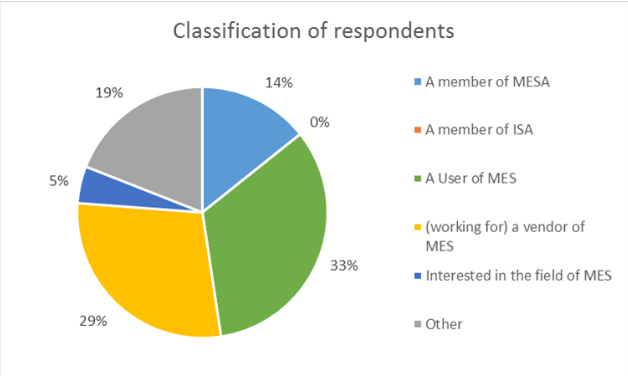
- Please classify yourself.
    - Member of ISA(-95)
    - Member of MESA
    - User of a MES
    - (working for) MES vendor
    - Interested in the field of MES
    - Other:
- From what country are you?
    - Open answer
- What is your age?
    - 0 - 30 years old
    - 30 - 40 years old
    - 40 -50 years old
    - 50 - 60 years old
    - 60 + years old

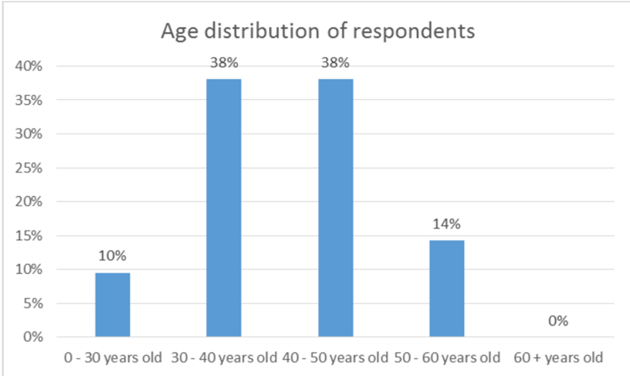
Next, the respondents are asked to how familiar on a scale from 1 to 5 the respondents with big data and data mining related topics.

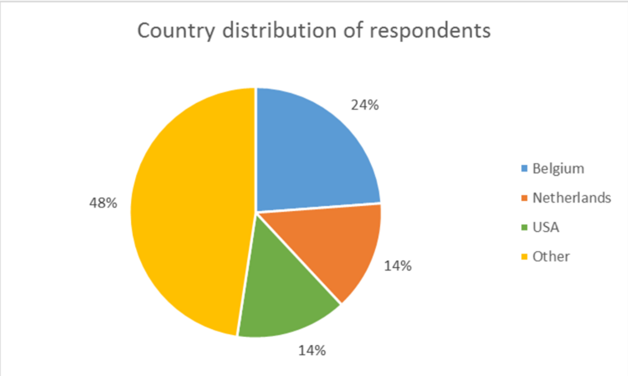
Then, the informational items are asked. The respondents are asked to score the main informational group on a scale from 1 to 5 for how useful and relevant they predict it would be. After each main informational item group, a question about the sub group items of that specific main group is asked. The respondents are asked which sub group they believe would be most useful and relevant.

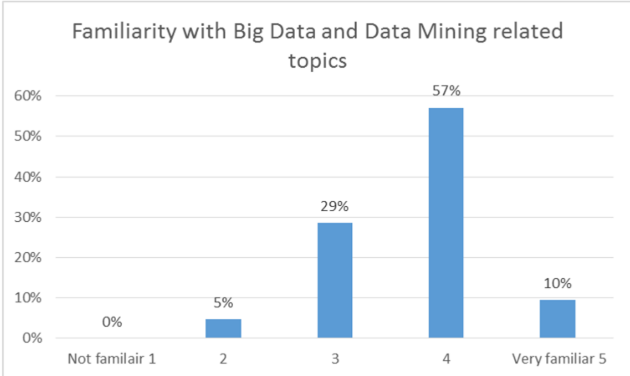
Appendix 9. The field expert survey 2 answers

- Answers where gathered during the time period: 23-2-2016 to 13-3-2016
- The number of respondents is 21









The familiarity with big data and data mining related topics is segmented among the group. The results are presented below.

Overall score	21	3,71

Classification	Count	Avg
A member of MESA	3	3,33
A member of ISA	0	x
A User of MES	7	3,86
(working for) a vendor of MES	6	4,00
Interested in the field of MES	1	4,00
Other	4	3,25

Country	Count	Avg
Belgium	5	3,60
Netherlands	3	3,67
USA	3	3,67
Other	10	3,80

Age group	Count	Avg
0 - 30 years old	2	4,00

30 - 40 years old	8	3,88
40 - 50 years old	8	3,75
50 - 60 years old	3	3,00
60 + years old	0	x

The score per main group is a score given for how useful/relevant the main item is according to the respondent on a scale from 1 to 5. This score is translated to a percentage. Per main group, the respondents could choose which of the sub groups would be the most useful/ relevant. The amount for 'votes' is also translated to a percentage. The results are presented below.

Main group average score	Main group percentage	Main informational (purpose) group	Sub-Informational Item group	Sub group votes	Sub group Percentage
4,10	82%	Condition based monitoring	Monitoring of parameters settings and their effects	6	29%
			Monitoring process conditions	14	67%
			Monitoring tool wear	1	5%
3,62	72%	Cycle/lead time prediction	Forecasting lead time	4	19%
			Forecasting production cycle time	17	81%
3,67	73%	Decision support	Improved scheduling decisions by insights options and effects	11	52%
			Insights in the effect of to-be-made decisions	10	48%
4,14	83%	Defect/low quality classification	Classification of product quality	4	19%
			Detection of a product with quality faults	16	76%
			Product state diagnosis	1	5%
4,24	85%	Knowledge of optimal manufacturing settings	Grading of (raw)materials	0	0%
			Identification of critical process parameters	12	57%
			Improved dispatching rules	0	0%
			Improved methods for a specific process	0	0%
			Knowledge of operational process(es)	6	29%
			Optimization of parameter settings	3	14%
3,71	74%	Low yield factors identification	Identification of characteristics for low yield (product quality failure)	9	43%
			Identification of characteristics product quality	6	29%
			Suggested improvements for next generations based on quality failure	6	29%
4,14	83%	Machine (component) failure prediction	Forecasting component failure	5	24%
			Forecasting machine/equipment failure	1	5%
			Forecasting tool wear	0	0%
			Machine performance prediction	1	5%
			Preventive maintenance schedule recommendations	9	43%
			Probability for machine failure	0	0%
3,62	72%	Machine fault diagnostics	Classification of machine fault types	2	10%

			Diagnostics of machine failure	5	24%
			Diagnostics of machine part wear and correlations between parts	7	33%
			Identification of characteristics of machine failure	4	19%
			Identification of machine failure	3	14%
3,90	78%	Patterns causing process variations	Detection of abnormal process behavior	10	48%
			Identification of critical process parameters	11	52%
			Identification of process fault classes	0	0%
3,33	67%	Process performance prediction	Forecasting of manufacturing process behavior	2	10%
			Forecasting production process performance	18	86%
			Prediction of system output	1	5%
4,14	83%	Root Cause analysis	Detection of change points in control charts	4	19%
			General Root cause analysis	3	14%
			Root cause analysis for process failure	4	19%
			Root cause analysis for unnatural patterns in the data	0	0%
			Root cause analysis of nonconformities in the production process	8	38%
			Root cause analysis of product quality	2	10%
3,71	74%	Yield/Low quality prediction	Prediction of product quality	16	76%
			Yield prediction	5	24%

Appendix 10. Statistical tests for Root Cause Analysis

The results of the statistical test are grouped per focus area. First the population determination. Next, the operations cracken, tightening and fine drilling are examined for Type X, supplier A.

Population determination

Seven engines types are found in the data. The errors only occur in the three biggest engine types. The data of these three engine types is combined and with a Chi-square test it is checked whether the amount of errors is significant different among the three types.

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Type * Category	35425	100,0%	0	0,0%	35425	100,0%

Type * Category Crosstabulation

			Category		Total
			0	1	
Partnr	Type X	Count	3198	23	3221
		% within Type	99,3%	0,7%	100,0%
		% within Category	9,0%	37,7%	9,1%
		% of Total	9,0%	0,1%	9,1%
	Type Y	Count	31645	29	31674
		% within Type	99,9%	0,1%	100,0%
		% within Category	89,5%	47,5%	89,4%
		% of Total	89,3%	0,1%	89,4%
	Type Z	Count	521	9	530
		% within Type	98,3%	1,7%	100,0%
		% within Category	1,5%	14,8%	1,5%
		% of Total	1,5%	0,0%	1,5%
Total		Count	35364	61	35425
		% within Type	99,8%	0,2%	100,0%
		% within Category	100,0%	100,0%	100,0%
		% of Total	99,8%	0,2%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	138,790[a]	2	,000
Likelihood Ratio	70,227	2	,000
N of Valid Cases	35425		

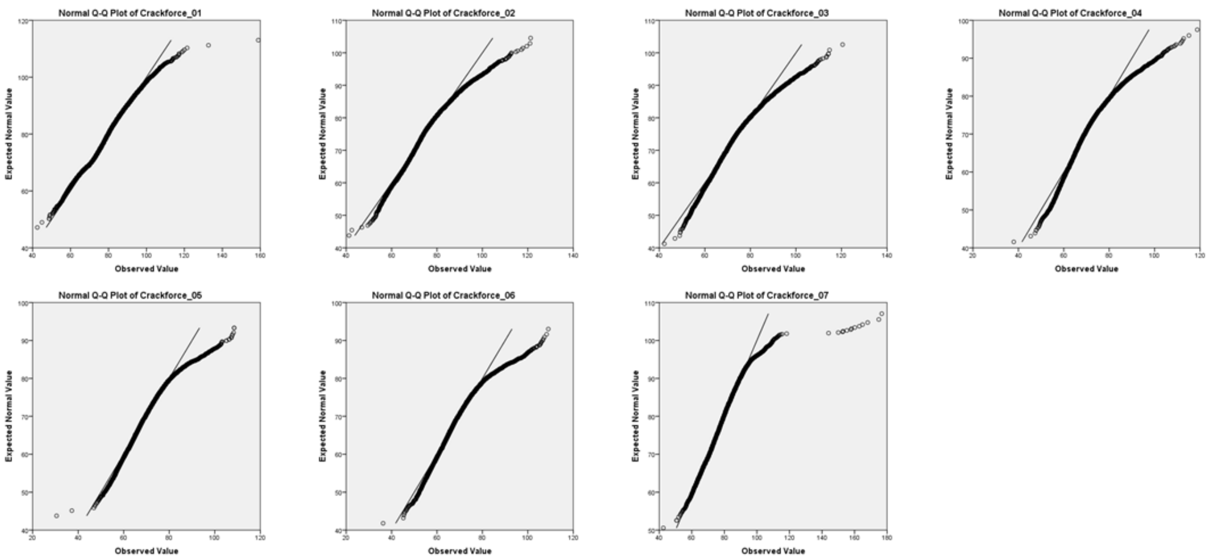
a. 1 cells (16,7%) have expected count less than 5. The minimum expected count is ,91.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,063	,000
	Cramer's V	,063	,000
N of Valid Cases		35425	

It can be found that the amount of errors per group are significantly different per group.

Next the behaviour is checked. For the operation cracken it is analysed whether there is a significant difference between the mean values of the three part numbers. In order to conduct a mean value comparison with a t-test the assumption of normality is checked with Q-Q plots.



It can be found that the distributions approach normality. The normal distribution is assumed and the t-test for compare means can be executed.

The means of the three partnumbers are compared on pairs of two for the operation cracken. Next, the two groups of suppliers are compared. For before the comparison the outliers are removed. The values for cracken are standardized and all rows with a Z-value above 3 or below -3 are removed. Last a correlation test is performed to see whether the values are correlated among the caps.

Mean comparison Type X versus Type Y

Group Statistics

	Partnr	N	Mean	Std. Deviation	Std. Error Mean
Crackforce_01	Type X	3167	68,318440813703870	8,416652250188170	,149559964076704
	Type Y	31367	81,344080613350630	6,820618084080358	,038511216748362
Crackforce_02	Type X	3167	66,634240530786270	7,566940594951237	,134460986376877
	Type Y	31367	75,010270965630180	6,847971272332553	,038665660898225
Crackforce_03	Type X	3167	65,187204339753660	6,979464448684928	,124021810714185
	Type Y	31367	72,626897439408820	7,099568288286503	,040086251685629
Crackforce_04	Type X	3167	65,560083454689060	6,749005228912675	,119926658436803
	Type Y	31367	69,853632099973990	6,606956939279617	,037304823052557
Crackforce_05	Type X	3167	63,968813867698330	6,317486515689713	,112258773233799
	Type Y	31367	69,535755822675270	5,796877349357737	,032730875312583
Crackforce_06	Type X	3167	61,745684907483440	6,459781545901664	,114787289201215
	Type Y	31367	68,060711977296880	5,841486041525686	,032982749115186
Crackforce_07	Type X	3167	71,980241838964550	7,523354499263883	,133686481893209
	Type Y	31367	79,537088183758510	6,281830908070877	,035469065808263

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Crackforce_01	Equal variances assumed	88,757	,000	-100,057	34532	,000	-13,0256398	,1301824419	-13,2808016	-12,7704780
	Equal variances not assumed			-84,342	3598,163	,000	-13,0256398	,1544386502	-13,3284358	-12,7228438
Crackforce_02	Equal variances assumed	11,107	,001	-64,947	34532	,000	-8,37603043	,1289676414	-8,62881115	-8,12324972
	Equal variances not assumed			-59,867	3708,688	,000	-8,37603043	,1399099360	-8,65033839	-8,10172248
Crackforce_03	Equal variances assumed	22,113	,000	-56,290	34532	,000	-7,43969310	,1321678829	-7,69874639	-7,18063981
	Equal variances not assumed			-57,079	3857,813	,000	-7,43969310	,1303392386	-7,69523349	-7,18415271
Crackforce_04	Equal variances assumed	3,185	,074	-34,785	34532	,000	-4,29354865	,1234320511	-4,53547942	-4,05161787
	Equal variances not assumed			-34,186	3804,734	,000	-4,29354865	,1255947978	-4,53978826	-4,04730903
Crackforce_05	Equal variances assumed	,013	,909	-51,069	34532	,000	-5,56694195	,1090088530	-5,78060280	-5,35328111
	Equal variances not assumed			-47,608	3724,452	,000	-5,56694195	,1169330679	-5,79620106	-5,33768285
Crackforce_06	Equal variances assumed	1,785	,182	-57,398	34532	,000	-6,31502707	,1100218829	-6,53067349	-6,09938065
	Equal variances not assumed			-52,876	3707,820	,000	-6,31502707	,1194319199	-6,54918577	-6,08086837
Crackforce_07	Equal variances assumed	59,877	,000	-63,272	34532	,000	-7,55684634	,1194341399	-7,79094109	-7,32275160
	Equal variances not assumed			-54,636	3625,598	,000	-7,55684634	,1383117134	-7,82802285	-7,28566984

The means of all seven caps are significantly different between Type X and Type Y. All with a significance value of 0,000.

Mean comparison Type X versus Type Z

Group Statistics

	Partnr	N	Mean	Std. Deviation	Std. Error Mean
Crackforce_01	Type X	3167	68,318440813703870	8,416652250188170	,149559964076704
	Type Z	521	75,961005011516290	9,562671217646567	,418948232517310
Crackforce_02	Type X	3167	66,634240530786270	7,566940594951237	,134460986376877
	Type Z	521	75,267008341650720	6,075398804031182	,266168054182360
Crackforce_03	Type X	3167	65,187204339753660	6,979464448684928	,124021810714185
	Type Z	521	73,125037518234180	5,890110684587889	,258050434285110
Crackforce_04	Type X	3167	65,560083454689060	6,749005228912675	,119926658436803
	Type Z	521	74,631352385796420	6,843601873288786	,299823641701247
Crackforce_05	Type X	3167	63,968813867698330	6,317486515689713	,112258773233799
	Type Z	521	68,370855719769700	6,075851883794183	,266187903967172
Crackforce_06	Type X	3167	61,745684907483440	6,459781545901664	,114787289201215
	Type Z	521	65,483810809980770	5,775418886009571	,253025695359657
Crackforce_07	Type X	3167	71,980241838964550	7,523354499263883	,133686481893209
	Type Z	521	75,502011276391490	8,824364230115116	,386602416119201

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Crackforce_01	Equal variances assumed	28,683	,000	-18,824	3686	,000	-7,64256420	,4059983019	-8,43856763	-6,84656077
	Equal variances not assumed			-17,180	659,226	,000	-7,64256420	,4448435729	-8,51604527	-6,76908312
Crackforce_02	Equal variances assumed	4,283	,039	-24,760	3686	,000	-8,63276781	,3486615669	-9,31635639	-7,94917923
	Equal variances not assumed			-28,949	810,603	,000	-8,63276781	,2982032695	-9,21810947	-8,04742615
Crackforce_03	Equal variances assumed	1,499	,221	-24,560	3686	,000	-7,93783318	,3232021850	-8,57150590	-7,30416046
	Equal variances not assumed			-27,725	781,125	,000	-7,93783318	,2863065423	-8,49985453	-7,37581183
Crackforce_04	Equal variances assumed	4,981	,026	-28,373	3686	,000	-9,07126893	,3197093823	-9,69809363	-8,44444423
	Equal variances not assumed			-28,091	696,774	,000	-9,07126893	,3229189055	-9,70527966	-8,43725820
Crackforce_05	Equal variances assumed	3,268	,071	-14,817	3686	,000	-4,40204185	,2970886402	-4,98451615	-3,81956755
	Equal variances not assumed			-15,238	717,688	,000	-4,40204185	,2888910390	-4,96921438	-3,83486932
Crackforce_06	Equal variances assumed	1,136	,287	-12,417	3686	,000	-3,73812590	,3010472745	-4,32836153	-3,14789027
	Equal variances not assumed			-13,454	750,840	,000	-3,73812590	,2778455043	-4,28357233	-3,19267948
Crackforce_07	Equal variances assumed	46,539	,000	-9,649	3686	,000	-3,52176944	,3649895741	-4,23737084	-2,80616804
	Equal variances not assumed			-8,609	650,268	,000	-3,52176944	,4090641803	-4,32501556	-2,71852332

The means of all seven caps are significantly different between Type X and Type Z. All with a significance value of 0,000.

Mean comparison Type Y versus Type Z

Group Statistics

	Partnr	N	Mean	Std. Deviation	Std. Error Mean
Crackforce_01	Type Y	31367	81,344080613350630	6,820618084080358	,038511216748362
	Type Z	521	75,961005011516290	9,562671217646567	,418948232517310
Crackforce_02	Type Y	31367	75,010270965630180	6,847971272332553	,038665660898225
	Type Z	521	75,267008341650720	6,075398804031182	,266168054182360
Crackforce_03	Type Y	31367	72,626897439408820	7,099568288286503	,040086251685629
	Type Z	521	73,125037518234180	5,890110684587889	,258050434285110
Crackforce_04	Type Y	31367	69,853632099973990	6,606956939279617	,037304823052557
	Type Z	521	74,631352385796420	6,843601873288786	,299823641701247
Crackforce_05	Type Y	31367	69,535755822675270	5,796877349357737	,032730875312583
	Type Z	521	68,370855719769700	6,075851883794183	,266187903967172
Crackforce_06	Type Y	31367	68,060711977296880	5,841486041525686	,032982749115186
	Type Z	521	65,483810809980770	5,775418886009571	,253025695359657
Crackforce_07	Type Y	31367	79,537088183758510	6,281830908070877	,035469065808263
	Type Z	521	75,502011276391490	8,824364230115116	,386602416119201

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Crackforce_01	Equal variances assumed	139,649	,000	17,728	31886	,000	5,383075602	,3036512439	4,787907696	5,978243507
	Equal variances not assumed			12,795	528,824	,000	5,383075602	,4207145533	4,556598679	6,209552525
Crackforce_02	Equal variances assumed	1,141	,286	-,850	31886	,395	-,256737376	,3019707962	-,848611541	,3351367885
	Equal variances not assumed			-,955	542,174	,340	-,256737376	,2689618307	-,785072300	,2715975484
Crackforce_03	Equal variances assumed	11,198	,001	-1,592	31886	,111	-,498140079	,3128121638	-1,11126373	,1149835756
	Equal variances not assumed			-1,908	545,394	,057	-,498140079	,2611454273	-1,01111408	,0148339256
Crackforce_04	Equal variances assumed	3,258	,071	-16,361	31886	,000	-4,77772029	,2920235199	-5,35009741	-4,20534316
	Equal variances not assumed			-15,813	536,223	,000	-4,77772029	,3021355092	-5,37123463	-4,18420594
Crackforce_05	Equal variances assumed	4,984	,026	4,546	31886	,000	1,164900103	,2562720046	,6625972958	1,667202910
	Equal variances not assumed			4,344	535,841	,000	1,164900103	,2681926741	,6380621398	1,691738066
Crackforce_06	Equal variances assumed	3,653	,056	9,988	31886	,000	2,576901167	,2579894612	2,071232080	3,082570254
	Equal variances not assumed			10,099	537,819	,000	2,576901167	,2551663462	2,075656309	3,078146025
Crackforce_07	Equal variances assumed	159,785	,000	14,427	31886	,000	4,035076907	,2796817881	3,486890040	4,583263774
	Equal variances not assumed			10,394	528,790	,000	4,035076907	,3882260717	3,272422194	4,797731621

The means of cap 1, 4, 5, 6, and 7 are significantly different with a significance value of 0,000. Cap 2 and 3 are not significantly different with a significance value of subsequently 0,340 and 0,111.

Supplier A versus Supplier B

The means of both suppliers are compared. All seven part types are included in the data.

Group Statistics

	Supplier	N	Mean	Std. Deviation	Std. Error Mean
109001 Crack-kracht 01	Supplier A	3464	68,80388858	10,26449185	,1744008160
	Supplier B	32453	81,56493189	8,057038325	,0447247585
109002 Crack-kracht 02	Supplier A	3465	68,01944532	10,04307219	,1706141178
	Supplier B	32418	75,07784745	7,978284046	,0443114932
109003 Crack-kracht 03	Supplier A	3463	66,55408802	9,726132047	,1652775687
	Supplier B	32419	72,71578544	8,210746438	,0456018890
109004 Crack-kracht 04	Supplier A	3461	66,71130238	9,210975786	,1565686623
	Supplier B	32409	70,09086458	7,641640781	,0424476648
109005 Crack-kracht 05	Supplier A	3464	64,66442226	8,891061294	,1510652809
	Supplier B	32415	69,64785985	6,843963162	,0380132192
109006 Crack-kracht 06	Supplier A	3462	62,36111068	8,810203463	,1497346826
	Supplier B	32419	68,20253131	6,973657726	,0387311882
109007 Crack-kracht 07	Supplier A	3462	72,14066340	9,698552269	,1648327024
	Supplier B	32407	79,77366683	7,419924772	,0412173520

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
109001 Crack-kracht 01	Equal variances assumed	52,821	,000	-86,062	35915	,000	-12,7610433	,1482780854	-13,0516727	-12,4704139
	Equal variances not assumed			-70,877	3931,656	,000	-12,7610433	,1800442964	-13,1140323	-12,4080543
109002 Crack-kracht 02	Equal variances assumed	96,567	,000	-48,159	35881	,000	-7,05840213	,1465651658	-7,34567417	-6,77113008
	Equal variances not assumed			-40,042	3945,159	,000	-7,05840213	,1762744611	-7,40399975	-6,71280451
109003 Crack-kracht 03	Equal variances assumed	9,081	,003	-41,183	35880	,000	-6,16169742	,1496174812	-6,45495210	-5,86844275
	Equal variances not assumed			-35,938	4006,686	,000	-6,16169742	,1714532211	-6,49784111	-5,82555374
109004 Crack-kracht 04	Equal variances assumed	22,821	,000	-24,208	35868	,000	-3,37956221	,1396060962	-3,65319428	-3,10593014
	Equal variances not assumed			-20,833	3985,027	,000	-3,37956221	,1622206838	-3,69760550	-3,06151891
109005 Crack-kracht 05	Equal variances assumed	18,079	,000	-39,446	35877	,000	-4,98343759	,1263343803	-5,23105670	-4,73581848
	Equal variances not assumed			-31,991	3913,761	,000	-4,98343759	,1557745933	-5,28884463	-4,67803055
109006 Crack-kracht 06	Equal variances assumed	18,907	,000	-45,556	35879	,000	-5,84142063	,1282240226	-6,09274350	-5,59009777
	Equal variances not assumed			-37,769	3937,748	,000	-5,84142063	,1546627949	-6,14464734	-5,53819392
109007 Crack-kracht 07	Equal variances assumed	41,265	,000	-55,662	35867	,000	-7,63300343	,1371312409	-7,90178471	-7,36422215
	Equal variances not assumed			-44,924	3905,718	,000	-7,63300343	,1699078865	-7,96612000	-7,29988686

The means of all seven caps are significantly different between supplier A and B. All with a significance value of 0,000.

Correlation test among the caps

Correlations

		Crackforce_01	Crackforce_02	Crackforce_03	Crackforce_04	Crackforce_05	Crackforce_06	Crackforce_07
Crackforce_01	Pearson Correlation	1	,814**	,745**	,678**	,630**	,610**	,693**
	Sig. (2-tailed)		,000	,000	,000	,000	,000	,000
	N	35675	35675	35675	35675	35675	35675	35675
Crackforce_02	Pearson Correlation	,814**	1	,907**	,841**	,751**	,739**	,677**
	Sig. (2-tailed)	,000		,000	,000	,000	,000	,000
	N	35675	35675	35675	35675	35675	35675	35675
Crackforce_03	Pearson Correlation	,745**	,907**	1	,891**	,767**	,737**	,645**
	Sig. (2-tailed)	,000	,000		,000	,000	,000	,000
	N	35675	35675	35675	35675	35675	35675	35675
Crackforce_04	Pearson Correlation	,678**	,841**	,891**	1	,768**	,727**	,626**
	Sig. (2-tailed)	,000	,000	,000		,000	,000	,000
	N	35675	35675	35675	35675	35675	35675	35675
Crackforce_05	Pearson Correlation	,630**	,751**	,767**	,768**	1	,890**	,765**
	Sig. (2-tailed)	,000	,000	,000	,000		,000	,000
	N	35675	35675	35675	35675	35675	35675	35675
Crackforce_06	Pearson Correlation	,610**	,739**	,737**	,727**	,890**	1	,791**
	Sig. (2-tailed)	,000	,000	,000	,000	,000		,000
	N	35675	35675	35675	35675	35675	35675	35675
Crackforce_07	Pearson Correlation	,693**	,677**	,645**	,626**	,765**	,791**	1
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	
	N	35675	35675	35675	35675	35675	35675	35675

**. Correlation is significant at the 0.01 level (2-tailed).

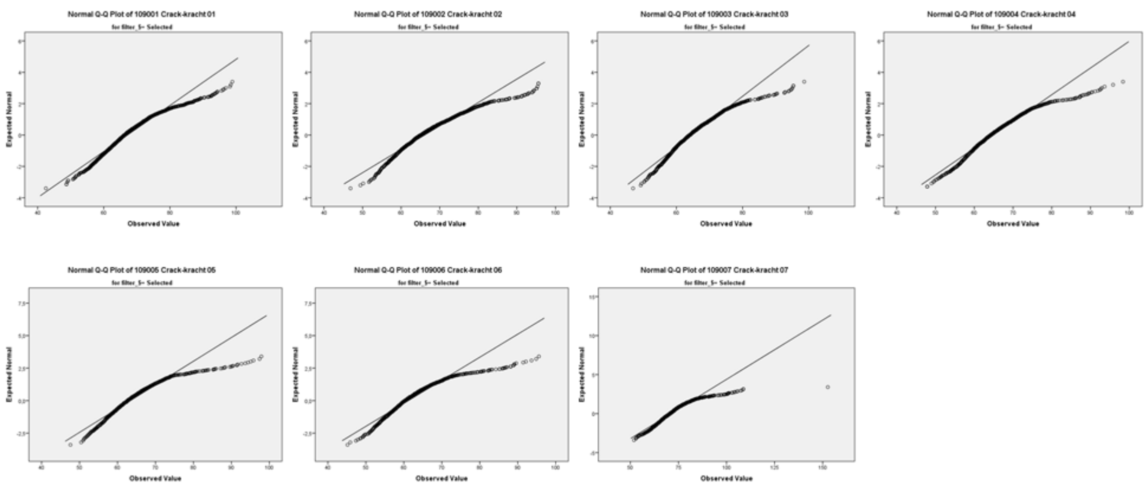
The crack force per cap are significantly correlated among the caps. This makes sense as they all are from the same casting block.

In the next section, a focus is applied on engine Type X and Supplier A. Also outliers with a Z-value greater than 3 or lower than -3 are removed.

Cracken

It is analyzed whether engines with an error are significantly different from engines without an error. Engines without an error are group A and engines with an error are group B.

First a test for normality with Q-Q plots



The Q-Q plots do deviate slightly from the normal distribution line, but assumption of normality is possible. The t-test for means comparison is executed.

Group Statistics

	Category2	N	Mean	Std. Deviation	Std. Error Mean
109001 Crack-kracht 01	A	2945	67,14663031	6,821700299	,1257042571
	B	23	68,02500857	6,027606092	1,256842753
109002 Crack-kracht 02	A	2945	66,09744738	6,727506198	,1239685316
	B	23	66,50709074	5,329907663	1,111362573
109003 Crack-kracht 03	A	2943	64,83186240	6,721332782	,1238968507
	B	23	65,58890974	5,329879995	1,111356804
109004 Crack-kracht 04	A	2940	64,91857000	5,875484344	,1083602525
	B	23	65,37130604	4,961915611	1,034630926
109005 Crack-kracht 05	A	2942	63,44984150	5,973756611	,1101352115
	B	23	63,82685135	5,020732805	1,046895159
109006 Crack-kracht 06	A	2939	61,18127516	5,802660379	,1070353811
	B	23	62,25055196	5,041488951	1,051223115
109007 Crack-kracht 07	A	2940	71,08943191	6,590828870	,1215531926
	B	23	72,61591374	5,848881402	1,219576078

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
109001 Crack-kracht 01	Equal variances assumed	,374	,541	-,616	2966	,538	-,878378253	1,426804821	-3,67600596	1,919249456
	Equal variances not assumed			-,695	22,442	,494	-,878378253	1,263113322	-3,49492475	1,738168240
109002 Crack-kracht 02	Equal variances assumed	,833	,362	-,291	2966	,771	-,409643364	1,406303175	-3,16707218	2,347785455
	Equal variances not assumed			-,366	22,551	,718	-,409643364	1,118255322	-2,72548285	1,906196122
109003 Crack-kracht 03	Equal variances assumed	,301	,583	-,539	2964	,590	-,757047336	1,405021082	-3,51196303	1,997868359
	Equal variances not assumed			-,677	22,550	,505	-,757047336	1,118241645	-3,07286218	1,558767507
109004 Crack-kracht 04	Equal variances assumed	,246	,620	-,368	2961	,713	-,452736039	1,228594885	-2,86172248	1,956250399
	Equal variances not assumed			-,435	22,485	,668	-,452736039	1,040289911	-2,60746891	1,701996836
109005 Crack-kracht 05	Equal variances assumed	,018	,894	-,302	2963	,763	-,377009844	1,249110051	-2,82622104	2,072201348
	Equal variances not assumed			-,358	22,490	,724	-,377009844	1,052672427	-2,55736636	1,803346670
109006 Crack-kracht 06	Equal variances assumed	,023	,879	-,881	2960	,378	-1,06927680	1,213556386	-3,44877660	1,310222999
	Equal variances not assumed			-1,012	22,459	,322	-1,06927680	1,056658227	-3,25806083	1,119507234
109007 Crack-kracht 07	Equal variances assumed	,112	,737	-1,107	2961	,268	-1,52648183	1,378558518	-4,22951178	1,176548122
	Equal variances not assumed			-1,245	22,439	,226	-1,52648183	1,225618615	-4,06537763	1,012413969

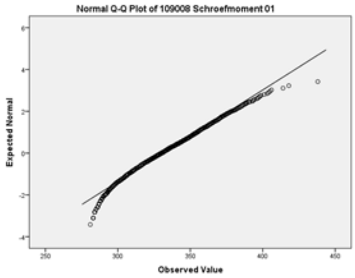
The crack force on a cap is not significantly different between engines with and engines without an error for all caps.

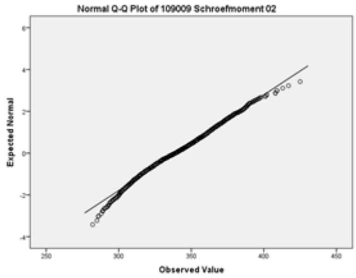
Tightening

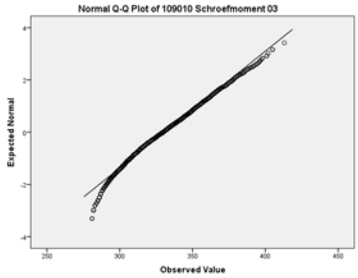
It is analyzed whether engines with an error are significantly different from engines without an error. Engines without an error are group A and engines with an error are group B.

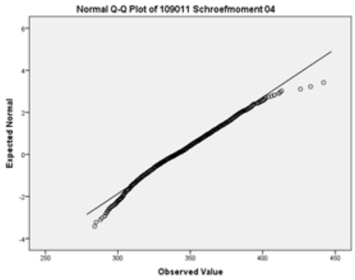
First a test for normality with Q-Q plots. First for the tightening moment (Schroefmoment) and second for the degree of rotation (verschroevingshoek)

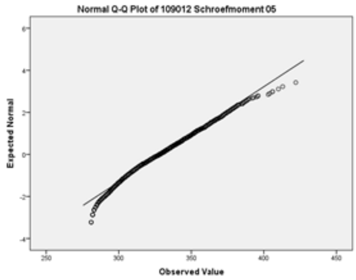
Tightening moment:

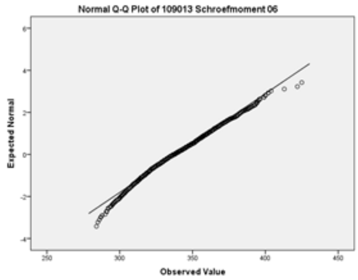


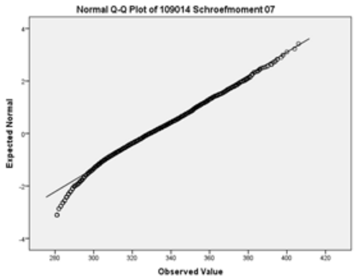


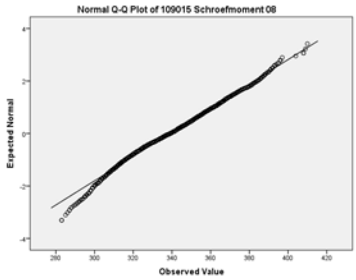


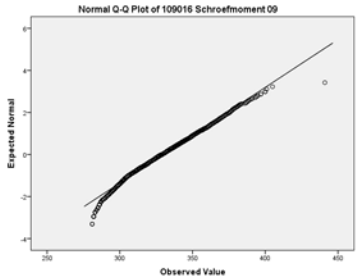


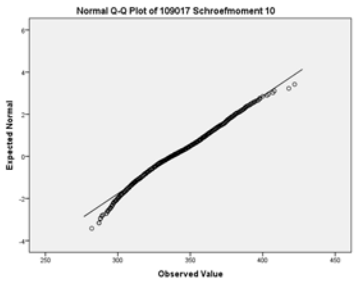


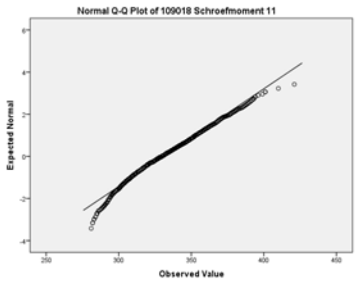


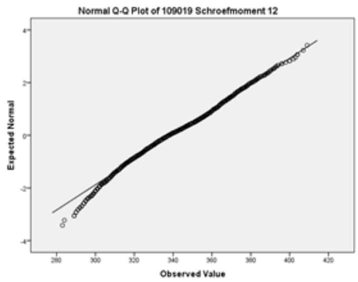


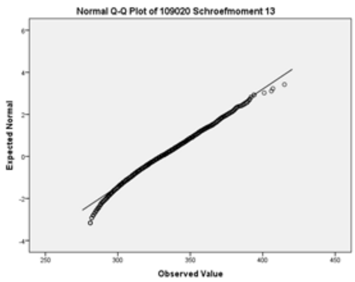


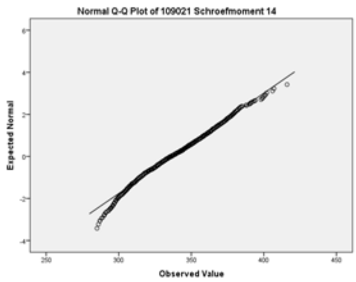




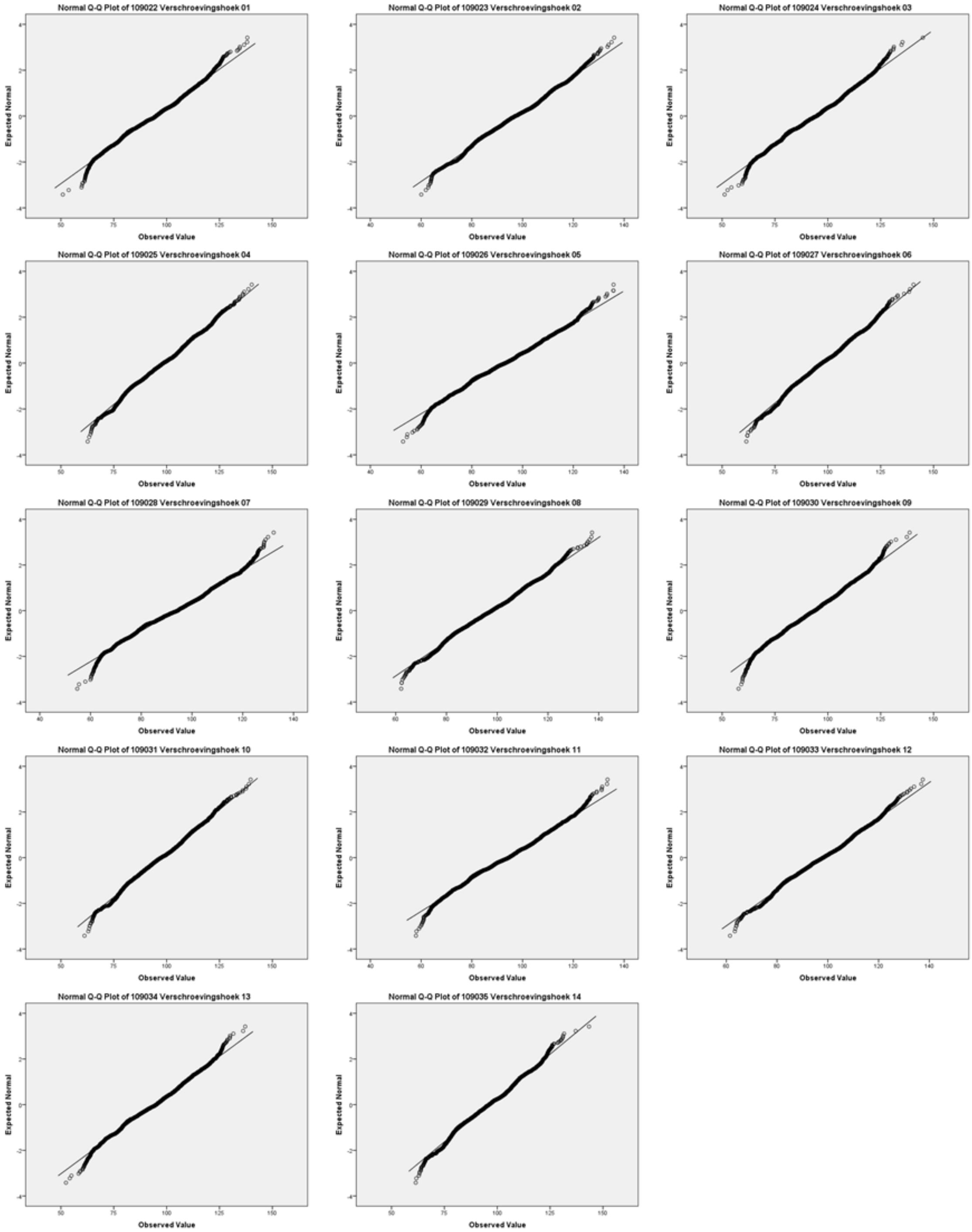








Degrees of rotation:



From these Q-Q plots the it can be found that the plots approach the line of normality. Therefore a normal distribution can be assumed and the t-test for compare means can be executed.

Group Statistics

	Category 2	N	Mean	Std. Deviation	Std. Error Mean
109008 Schroefmoment 01	A	3161	331,149	22,7844	,4053
	B	22	339,136	19,8477	4,2316
109009 Schroefmoment 02	A	3161	339,063	21,8452	,3885
	B	22	352,500	20,3651	4,3419
109010 Schroefmoment 03	A	3161	330,621	22,3584	,3977
	B	22	337,409	20,4582	4,3617
109011 Schroefmoment 04	A	3161	340,756	21,7703	,3872
	B	22	349,773	19,8012	4,2216
109012 Schroefmoment 05	A	3160	328,902	22,0843	,3929
	B	22	333,727	20,0741	4,2798
109013 Schroefmoment 06	A	3160	338,488	21,2913	,3788
	B	22	349,955	21,9295	4,6754
109014 Schroefmoment 07	A	3162	330,114	22,5375	,4008
	B	22	335,955	26,0978	5,5641
109015 Schroefmoment 08	A	3162	339,018	21,6591	,3852
	B	22	347,455	22,5129	4,7998
109016 Schroefmoment 09	A	3161	329,790	22,0111	,3915
	B	22	339,091	18,6443	3,9750
109017 Schroefmoment 10	A	3161	338,242	21,5863	,3839
	B	22	347,318	19,0423	4,0598
109018 Schroefmoment 11	A	3159	330,750	21,5245	,3830
	B	22	340,636	23,7838	5,0707
109019 Schroefmoment 12	A	3159	339,091	20,8447	,3709
	B	22	346,955	18,9296	4,0358
109020 Schroefmoment 13	A	3159	330,765	21,6335	,3849
	B	23	334,043	19,8666	4,1425
109021 Schroefmoment 14	A	3159	336,896	20,9538	,3728
	B	23	345,435	20,8039	4,3379

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
109008 Schroefmoment 01	Equal variances assumed	1,910	,167	-1,640	3181	,101	-7,9872	4,8706	-17,5371	1,5627
	Equal variances not assumed			-1,879	21,387	,074	-7,9872	4,2509	-16,8177	,8433
109009 Schroefmoment 02	Equal variances assumed	,164	,685	-2,876	3181	,004	-13,4370	4,6716	-22,5967	-4,2774
	Equal variances not assumed			-3,082	21,338	,006	-13,4370	4,3592	-22,4938	-4,3803
109010 Schroefmoment 03	Equal variances assumed	,333	,564	-1,420	3181	,156	-6,7876	4,7808	-16,1614	2,5862
	Equal variances not assumed			-1,550	21,351	,136	-6,7876	4,3798	-15,8868	2,3116
109011 Schroefmoment 04	Equal variances assumed	,688	,407	-1,937	3181	,053	-9,0166	4,6549	-18,1436	,1103
	Equal variances not assumed			-2,127	21,355	,045	-9,0166	4,2394	-17,8240	-,2093
109012 Schroefmoment 05	Equal variances assumed	1,974	,160	-1,022	3180	,307	-4,8257	4,7220	-14,0842	4,4328
	Equal variances not assumed			-1,123	21,355	,274	-4,8257	4,2978	-13,7544	4,1030
109013 Schroefmoment 06	Equal variances assumed	,031	,861	-2,517	3180	,012	-11,4667	4,5560	-20,3997	-2,5337
	Equal variances not assumed			-2,445	21,277	,023	-11,4667	4,6907	-21,2139	-1,7196
109014 Schroefmoment 07	Equal variances assumed	,207	,649	-1,210	3182	,226	-5,8407	4,8271	-15,3053	3,6239
	Equal variances not assumed			-1,047	21,218	,307	-5,8407	5,5785	-17,4345	5,7531
109015 Schroefmoment 08	Equal variances assumed	,060	,806	-1,820	3182	,069	-8,4365	4,6350	-17,5244	,6514
	Equal variances not assumed			-1,752	21,271	,094	-8,4365	4,8152	-18,4425	1,5695
109016 Schroefmoment 09	Equal variances assumed	2,576	,109	-1,977	3181	,048	-9,3010	4,7047	-18,5255	-,0764
	Equal variances not assumed			-2,329	21,409	,030	-9,3010	3,9942	-17,5977	-1,0042
109017 Schroefmoment 10	Equal variances assumed	1,688	,194	-1,967	3181	,049	-9,0763	4,6148	-18,1247	-,0280
	Equal variances not assumed			-2,226	21,377	,037	-9,0763	4,0779	-17,5478	-,6049
109018 Schroefmoment 11	Equal variances assumed	,406	,524	-2,145	3179	,032	-9,8866	4,6084	-18,9223	-,8509
	Equal variances not assumed			-1,944	21,240	,065	-9,8866	5,0852	-20,4545	,6813
109019 Schroefmoment 12	Equal variances assumed	1,372	,242	-1,764	3179	,078	-7,8631	4,4570	-16,6019	,8758
	Equal variances not assumed			-1,940	21,356	,066	-7,8631	4,0528	-16,2828	,5567
109020 Schroefmoment 13	Equal variances assumed	1,618	,203	-,725	3180	,469	-3,2788	4,5248	-12,1507	5,5930
	Equal variances not assumed			-,788	22,382	,439	-3,2788	4,1603	-11,8983	5,3406
109021 Schroefmoment 14	Equal variances assumed	,504	,478	-1,947	3180	,052	-8,5391	4,3848	-17,1365	,0583
	Equal variances not assumed			-1,961	22,326	,062	-8,5391	4,3539	-17,5609	,4827

Group Statistics

	Category 2	N	Mean	Std. Deviation	Std. Error Mean
109022 Verschroevingshoek 01	A	3161	94,230838342296980	15,022382909655242	,267193925030146
	B	22	98,436818181818180	14,532443633655282	3,098327393092836
109023 Verschroevingshoek 02	A	3161	97,311913951281470	13,078963793617099	,232627519372952
	B	22	106,064545454545450	12,237876693856085	2,609124077801747
109024 Verschroevingshoek 03	A	3161	93,900161341347830	14,936529723984757	,265666906993554
	B	22	97,949545454545440	12,486083864036000	2,662042024288704
109025 Verschroevingshoek 04	A	3161	98,482644732679660	13,072260755005184	,232508296530137
	B	22	104,263181818181820	11,622039156526887	2,477827073684487
109026 Verschroevingshoek 05	A	3160	92,973920886075970	14,924544871035152	,265495738605305
	B	22	96,708636363636340	15,684401500627960	3,343925635542892
109027 Verschroevingshoek 06	A	3160	97,589591772151800	12,982125709860613	,230941652411495
	B	22	104,923181818181820	13,169088765656825	2,807659158588657
109028 Verschroevingshoek 07	A	3162	93,341446869070140	14,946038814993043	,265793999988486
	B	22	99,477727272727310	17,084205719463096	3,642364897756149
109029 Verschroevingshoek 08	A	3162	97,686231815306950	13,179563921661511	,234379761500973
	B	22	103,233181818181820	15,372350375287644	3,277396112080762
109030 Verschroevingshoek 09	A	3161	93,350218285352820	14,640121872216444	,260394882055831
	B	22	99,861818181818170	10,456319409066245	2,229294788455996
109031 Verschroevingshoek 10	A	3161	97,501175260993620	13,072586960681900	,232514098550738
	B	22	102,478181818181820	11,350132634577921	2,419856408423205
109032 Verschroevingshoek 11	A	3159	93,772139917695400	14,354749208972107	,255399943240860
	B	22	99,694090909090900	14,723208326309633	3,138998562222102
109033 Verschroevingshoek 12	A	3159	98,016139601139710	12,745886464011160	,226775029787929
	B	22	102,834545454545460	11,362578399236150	2,422509854363933
109034 Verschroevingshoek 13	A	3159	94,136198163975990	14,599574537468417	,259755879669397
	B	23	94,440434782608680	12,787541692195740	2,666386763178609
109035 Verschroevingshoek 14	A	3159	96,194441278885960	13,022478349571596	,231696157342624
	B	23	101,445652173913060	12,632987419697889	2,634159969608702

Independent Samples Test

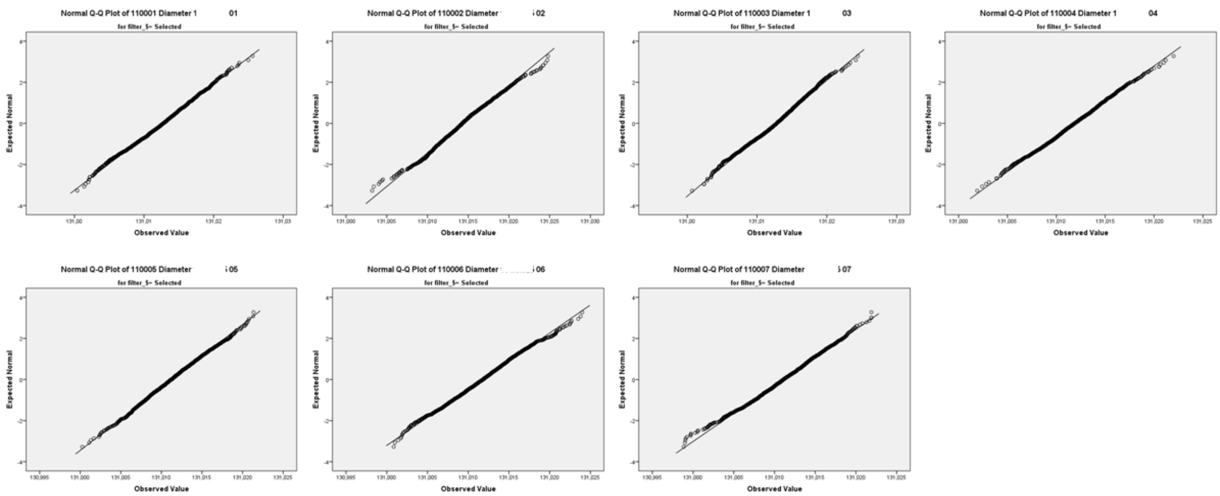
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
109022 Verschroevingshoek 01	Equal variances assumed	1,090	,297	-1,309	3181	,191	-4,20597984	3,213228106	-10,5061884	2,094228725
	Equal variances not assumed			-1,352	21,314	,190	-4,20597984	3,109827202	-10,6674321	2,255472372
109023 Verschroevingshoek 02	Equal variances assumed	,057	,811	-3,129	3181	,002	-8,75263150	2,796981194	-14,2367006	-3,26856243
	Equal variances not assumed			-3,341	21,335	,003	-8,75263150	2,619473996	-14,1949194	-3,31034362
109024 Verschroevingshoek 03	Equal variances assumed	1,445	,229	-1,268	3181	,205	-4,04938411	3,192362691	-10,3086816	2,209913423
	Equal variances not assumed			-1,514	21,420	,145	-4,04938411	2,675265715	-9,60626302	1,507494794
109025 Verschroevingshoek 04	Equal variances assumed	1,448	,229	-2,068	3181	,039	-5,78053709	2,794761795	-11,2602546	-,300819613
	Equal variances not assumed			-2,323	21,371	,030	-5,78053709	2,488711939	-10,9506253	-,610448875
109026 Verschroevingshoek 05	Equal variances assumed	,156	,693	-1,169	3180	,242	-3,73471548	3,194081444	-9,99738374	2,527952784
	Equal variances not assumed			-1,113	21,266	,278	-3,73471548	3,354448784	-10,7053725	3,235941529
109027 Verschroevingshoek 06	Equal variances assumed	,045	,833	-2,640	3180	,008	-7,33359005	2,777682566	-12,7798208	-1,88735934
	Equal variances not assumed			-2,603	21,285	,016	-7,33359005	2,817141104	-13,1873810	-1,47979912
109028 Verschroevingshoek 07	Equal variances assumed	,152	,697	-1,917	3182	,055	-6,13628040	3,200805472	-12,4121310	,1395702266
	Equal variances not assumed			-1,680	21,224	,108	-6,13628040	3,652049904	-13,7262509	1,453690124
109029 Verschroevingshoek 08	Equal variances assumed	,183	,669	-1,965	3182	,050	-5,54695000	2,823002238	-11,0820381	-,011861872
	Equal variances not assumed			-1,688	21,215	,106	-5,54695000	3,285766143	-12,3758540	1,281953967
109030 Verschroevingshoek 09	Equal variances assumed	2,838	,092	-2,082	3181	,037	-6,51159990	3,127058568	-12,6428550	-,380344810
	Equal variances not assumed			-2,901	21,577	,008	-6,51159990	2,244451102	-11,1716041	-1,85159571
109031 Verschroevingshoek 10	Equal variances assumed	1,269	,260	-1,781	3181	,075	-4,97700656	2,794493490	-10,4561980	,5021848463
	Equal variances not assumed			-2,047	21,390	,053	-4,97700656	2,431001408	-10,0269504	,0729373181
109032 Verschroevingshoek 11	Equal variances assumed	,000	,989	-1,928	3179	,054	-5,92195099	3,071608506	-11,9444860	,1005840462
	Equal variances not assumed			-1,880	21,279	,074	-5,92195099	3,149371541	-12,4662036	,6223016106
109033 Verschroevingshoek 12	Equal variances assumed	1,634	,201	-1,768	3179	,077	-4,81840585	2,725028592	-10,1613980	,5245863091
	Equal variances not assumed			-1,980	21,370	,061	-4,81840585	2,433101089	-9,87299238	,2361806745
109034 Verschroevingshoek 13	Equal variances assumed	3,017	,083	-,100	3180	,921	-,304236619	3,052822156	-6,28993635	5,681463108
	Equal variances not assumed			-,114	22,420	,911	-,304236619	2,679009423	-5,85414018	5,245666944
109035 Verschroevingshoek 14	Equal variances assumed	1,431	,232	-1,927	3180	,054	-5,25121090	2,724686017	-10,5935307	,0911089373
	Equal variances not assumed			-1,986	22,342	,059	-5,25121090	2,644330133	-10,7303568	,2279349942

The means are significantly different between good engines and error engines for:
Tightening moment (schroefmoment) of cap: 2, 4, 6, 9 and 10
Degrees of rotation (verschroevingshoek) of cap: 2, 4, 6 and 9
For the other caps the difference in not significant between the means of the good engines and the error engines.

Fine Drilling

It is analyzed whether engines with an error are significantly different from engines without an error. Engines without an error are group A and engines with an error are group B.

First a test for normality with Q-Q plots.



From these Q-Q plots the it can be found that the plots approach the line of normality. Therefore a normal distribution can be assumed and the t-test for compare means can be executed.

Group Statistics

	Category2	N	Mean	Std. Deviation	Std. Error Mean
110001 Diameter 01	A	1880	131,0126270	,0038743445	,0000893551
	B	15	131,0119167	,0042401847	,0010948110
110002 Diameter 02	A	1880	131,0143912	,0030599573	,0000705727
	B	15	131,0143333	,0030075302	,0007765410
110003 Diameter 03	A	1880	131,0125665	,0035473618	,0000818138
	B	15	131,0124933	,0037935032	,0009794783
110004 Diameter 04	A	1880	131,0118682	,0029088493	,0000670876
	B	15	131,0111733	,0036008861	,0009297448
110005 Diameter 05	A	1880	131,0112804	,0032562445	,0000750997
	B	15	131,0085967	,0035247323	,0009100820
110006 Diameter 131/0.025 06	A	1880	131,0117326	,0036562820	,0000843259
	B	15	131,0114200	,0038519383	,0009945662
110007 Diameter 131/0.025 07	A	1880	131,0110775	,0036518555	,0000842238
	B	15	131,0080067	,0042368227	,0010939429

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
110001 Diameter 01	Equal variances assumed	,003	,956	,707	1893	,480	,0007103399	,0010050685	-,001260819	,0026814982
	Equal variances not assumed			,647	14,187	,528	,0007103399	,0010984514	-,001642692	,0030633717
110002 Diameter 02	Equal variances assumed	,010	,918	,073	1893	,942	,0000578844	,0007931236	-,001497604	,0016133725
	Equal variances not assumed			,074	14,232	,942	,0000578844	,0007797412	-,001611938	,0017277067
110003 Diameter 03	Equal variances assumed	,136	,712	,080	1893	,937	,0000731733	,0009200597	-,001731264	,0018776109
	Equal variances not assumed			,074	14,196	,942	,0000731733	,0009828892	-,002032187	,0021785333
110004 Diameter 04	Equal variances assumed	,824	,364	,920	1893	,358	,0006948793	,0007555351	-,000786890	,0021766482
	Equal variances not assumed			,745	14,146	,468	,0006948793	,0009321621	-,001302473	,0026922318
110005 Diameter 05	Equal variances assumed	,117	,732	3,177	1893	,002	,0026837028	,0008446419	,0010271759	,0043402297
	Equal variances not assumed			2,939	14,191	,011	,0026837028	,0009131753	,0007276106	,0046397949
110006 Diameter 06	Equal variances assumed	,001	,981	,330	1893	,742	,0003125593	,0009481917	-,001547051	,0021721698
	Equal variances not assumed			,313	14,202	,759	,0003125593	,0009981346	-,001825373	,0024504920
110007 Diameter 07	Equal variances assumed	,130	,719	3,240	1893	,001	,0030708487	,0009478697	,0012118696	,0049298277
	Equal variances not assumed			2,799	14,166	,014	,0030708487	,0010971804	,0007202223	,0054214750

The means for the operation fine drilling are significantly different between good engines and error engines for cap number 5 and 7. For the other cap numbers the difference is not significant.

Appendix 11. Data analysis for all hypotheses

In this section the data analysis per hypothesis is discussed.

Hypothesis 1: Errors in caps arise more often in a specific engine type and/or supplier

In total 87 engines with an error in a cap where detected in 2015. From these 36 raised at type X, 37 at type Y and 14 at type Z. The engines with errors are traced in the MES data in order to find the corresponding supplier. 61 engines where found and mapped per type and supplier. This is represented in Figure 21 Error engines per engine types and supplier. Here F = number of false/errors engines and T = the total amount of engines.

Supplier / Engine Type	Supplier A		Supplier B	
Type X	F: 23 T: 3189	0,72%	F: 0 T: 238	0,00%
Type Y	F: 0 T: 0	0,00%	F: 29 T: 39010	0,07%
Type Z	F: 9 T: 901	1,00%	F: 0 T: 238	0,00%

Figure 21 Error engines per engine types and supplier.

It can be found that the errors occur at both suppliers and in all three engines types. The hypothesis is therefore rejected.

Hypothesis 2: Errors in caps are always present in a specific engine cap number

The errors occur at the caps of the engines. Usually at only one, but sometimes at two or more caps. Per cap numbers the amount of engines with an error on that cap number are counted per engines type. This is done for each engine type. Not for all engines with errors the error location was saved, therefore that total differ from the total found in hypothesis 1. The errors occur in almost all caps, only never cap 7. The results can be found in Table 20 amount of engines with an error detected on a specific cap. The hypothesis is rejected.

Table 20 amount of engines with an error detected on a specific cap

Cap number	Type X	Type Y	Type Z
1	1	6	0
2	2	5	1
3	10	3	2
4	6	3	1
5	1	3	0
6	5	3	0
7	0	0	0

Hypothesis 3: Errors in caps occur during specific period in time

The engines errors per engine type are mapped during the year. The engine errors are counted per week. In Figure 22 the result is presented. It can be found that the errors occur during the whole year. It is interesting that during the summer only engines of Type Y occur. The Type Z engines only occur at the end of the year. However, this is due to the fact that the part number for this engine type only exist at the end of the year. The hypothesis is not confirmed as the errors occur during the whole year. However the summer is and interesting period, which does not directly lead to rejecting the hypothesis.

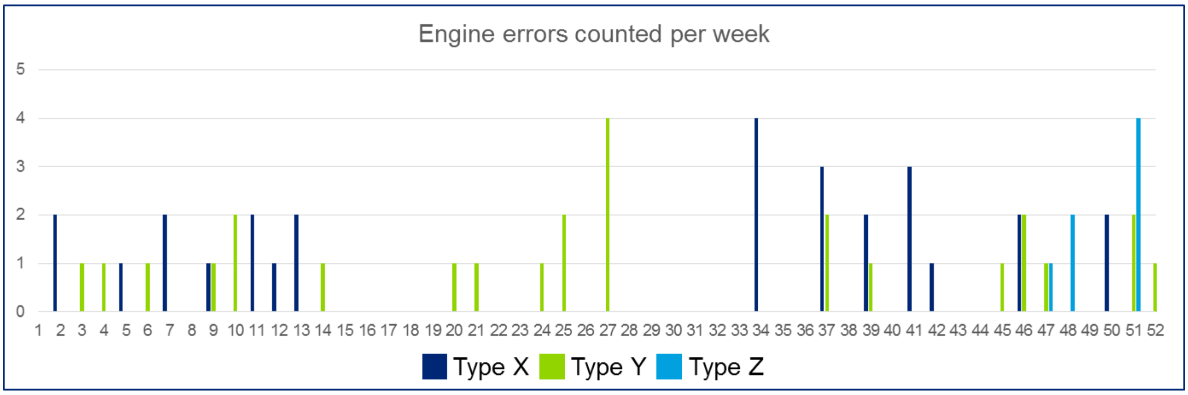


Figure 22 Engines errors during the year, counted per week

Hypothesis 4: When there is an error in a cap, the processing time is longer

For the processing time the data and time stamp of each cap operation is needed. The time between the first and the last operations on an engine is the processing time. When exporting the MES data, the data and timestamp get random formats. This leads to a lot of manual work when performing this analysis. Therefore, the analysis is conducted on a data sample of two months. If there are large differences to observe, an analysis of a full year could be considered.

The data is split in two group. All engines that where produced 'good' and all engines with errors. The outliers in processing time where removed. For fine drilling all data and time stamps where the same for an engines and therefore the processing time could not be calculated. The results are presented in Figure 23. It can be found that the differences in processing time are very small. Also the processing time of the error engines are within the variation of the good engines. A year analysis is not needed. The hypothesis is rejected.

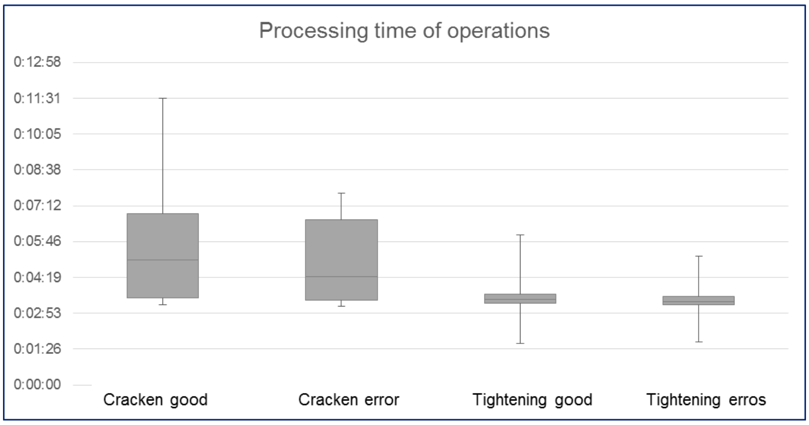


Figure 23 Processing time analysis

For the next part the analysis will focus on a specific engine type and a specific supplier. The Type X engine of Supplier A is chosen. This is chosen because the relative amount of error engines is higher than the Type Y engines and the absolute amount of error engines is 2,5 times bigger than type Z.

Hypothesis 5: Engines with an error in a cap have had a higher force when in the operation cracken

The data is of Type X and supplier A is used for this analysis. The average values for cracken per cap cap number is plotted in a graph presented in Figure 24.

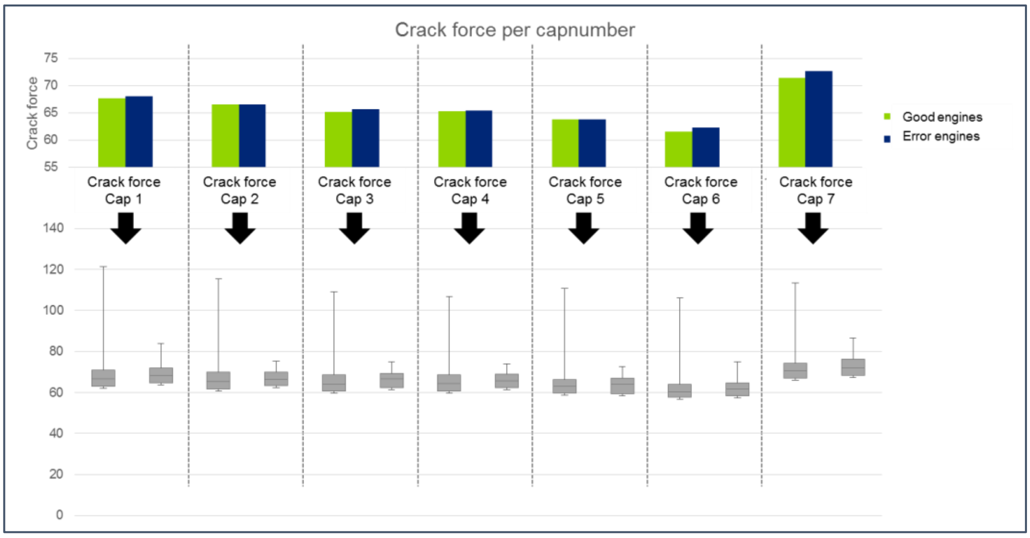


Figure 24 Crack force analysis

A difference can be observed. In order to check the variance and spread of the crack force, boxplots where created. Again on the left the good engines and on the right the engines with an error. Again a small difference can be observed. The results are also presented in Figure 24.In order to check whether this difference is significant, the means are compared by making use of a t-test in SPSS for which the results can be found in Appendix 10.The differences are not significant. And the hypothesis is therefore rejected.

Hypothesis 6: When there is an error in the cap, the bolt will need more rotations and has more force applied on it.

The data is of Type X and supplier A is used for this analysis. First the average values for each measurement item is plotted. This is presented in Figure 25.

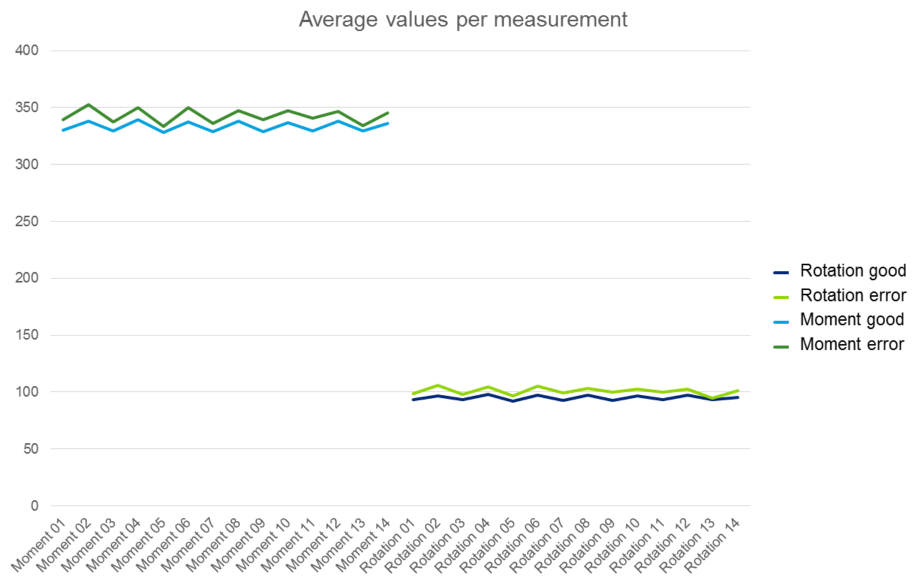


Figure 25 General analysis tightening measurements

Is can be observed that there is a saw-pattern in the graphs. All even values are higher than the odd values. This could indicate a difference in left and right as cap one has bolt 1 and 2, cap two has bolt

3 and 4 and so forth. When splitting the data into even and odd numbers, the pattern disappears as can be found in Figure 26 and Figure 27.

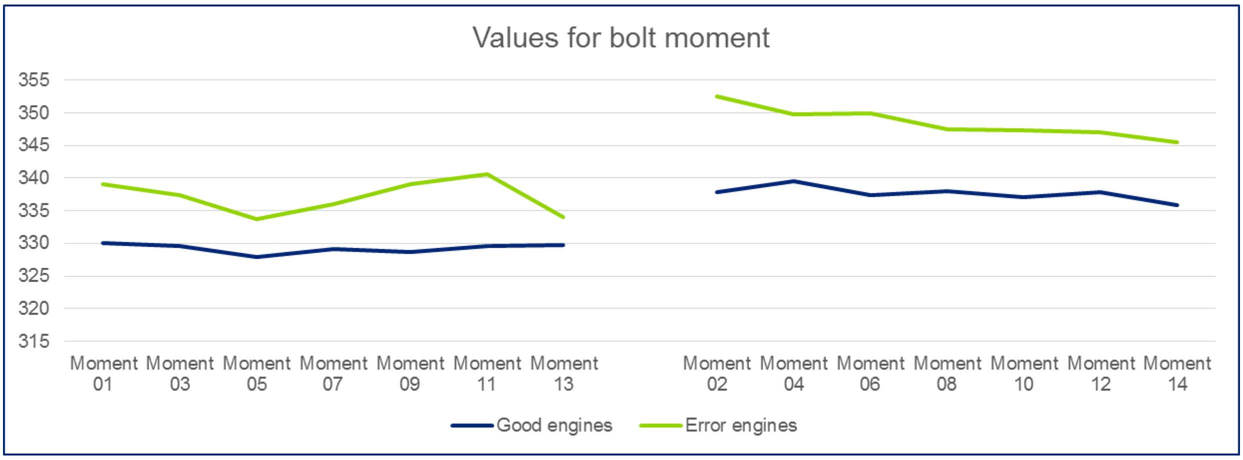


Figure 26 Values for tightening moment

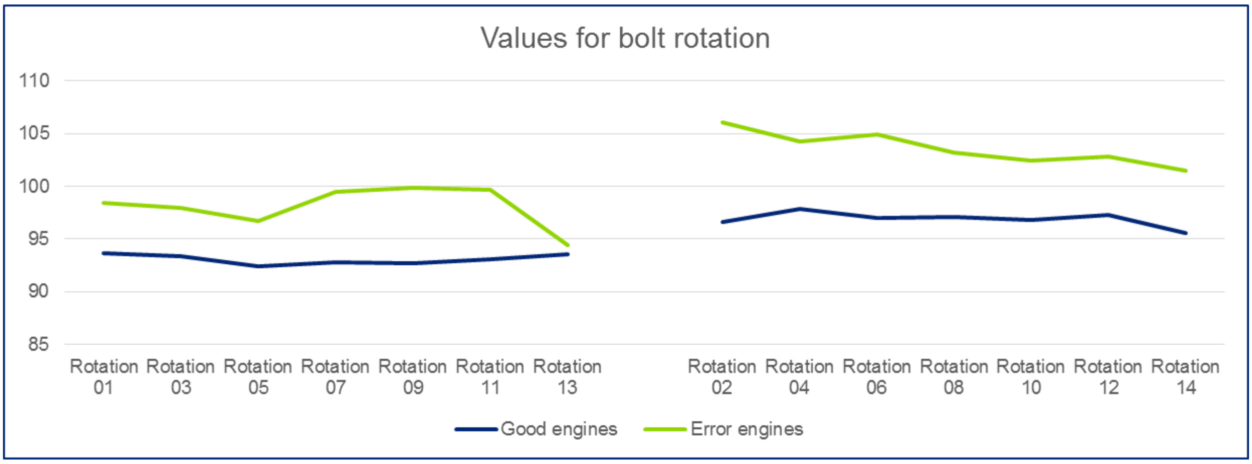


Figure 27 Values for tightening rotation

To further research the differences between the good engines and the engines with errored caps, boxplots are created. These are presented in Figure 28.

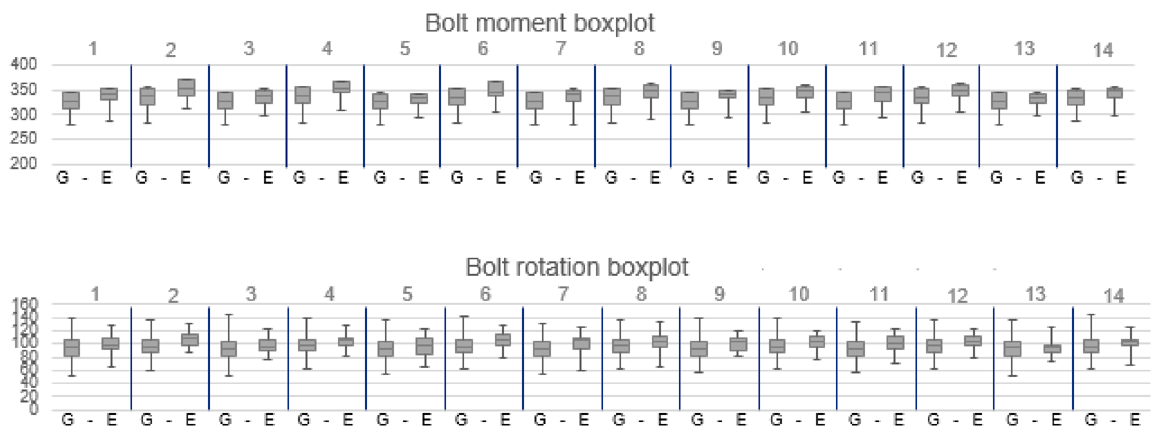


Figure 28 Boxplots for tightening

The differences are again researched whether they are significant by comparing the means by making use of the t-test in SPSS which can be found in Appendix 10. The results are presented in Table 21. No variances where significant different.

Table 21 Significant differences between the means of good engines and error engines

Rotation	Moment	Engine		Rotation	Moment		Cap	Errors
X	X	Bolt 1	Bolt 2	Sign	Sign		Cap 1	1
X	X	Bolt 3	Bolt 4	Sign	X		Cap 2	2
X	X	Bolt 5	Bolt 6	Sign	Sign		Cap 3	10
X	X	Bolt 7	Bolt 8	Sign	X		Cap 4	6
Sign	Sign	Bolt 9	Bolt 10	X	Sign		Cap 5	1
X	Sign	Bolt 11	Bolt 12	X	X		Cap 6	5
X	X	Bolt 13	Bolt 14	X	X		Cap 7	0

It can be found that for some bolt numbers there are significant differences between the means. For example bolt number 6 is significantly different for engines with errors compared to good engines. Also in the corresponding cap number, there are many errors. However, in cap 5 it can be found that four out of three values are significant different for engines with errors compared to good caps, though there are no actual errors found on cap 5. Further research to the differences between left and right might generate interesting insights. Especially setting up an experiment on which side most errors are found (left of right) would be interesting.

The hypothesis is not rejected but not confirmed as well. Additional research is needed.

Hypothesis 7: When there is an error in the cap, the cylinder measured in fine drilling is smaller.

For this operation also the average values are research and box plots are created. These are presented in Figure 29.Figure 29 Boxplots for fine drilling

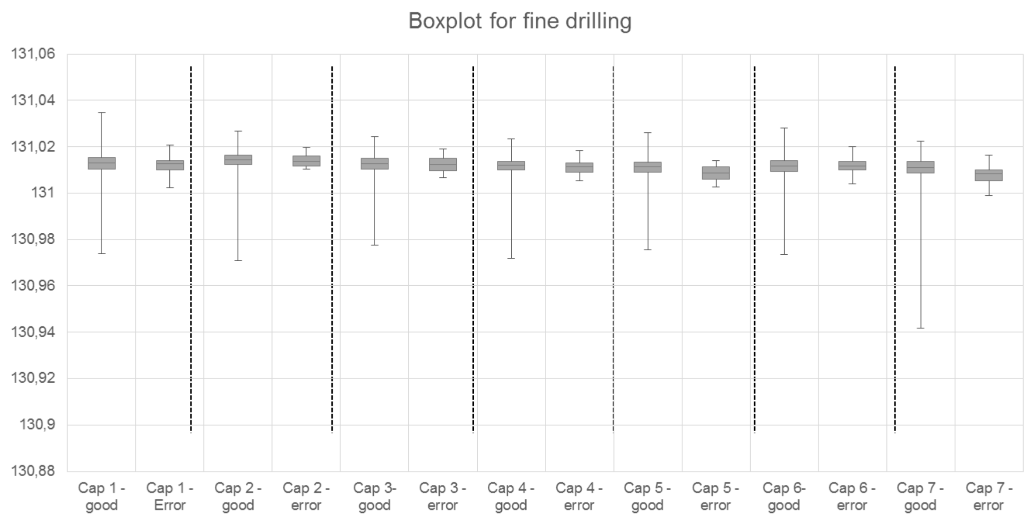


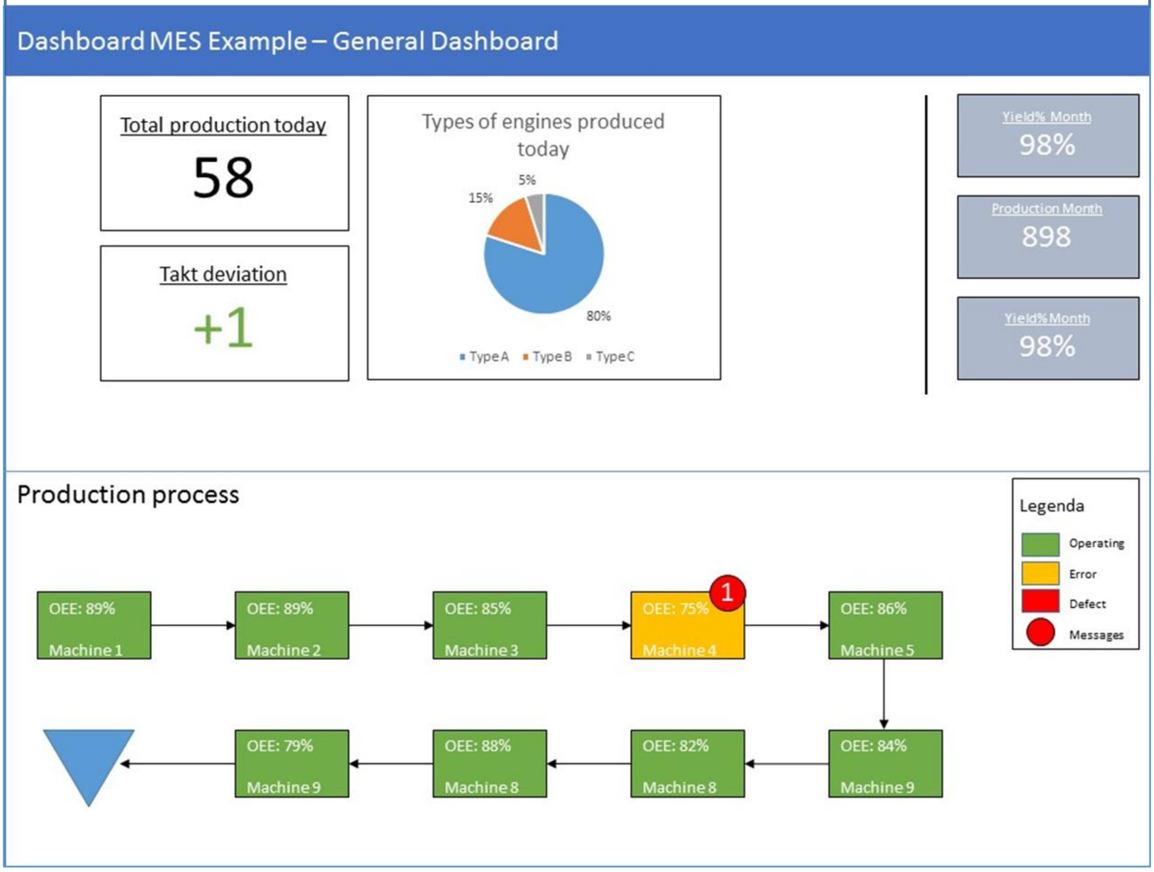
Figure 29 Boxplots for fine drilling

The differences are again research in SPSS and the means are compared between the good engines and the engines with errors with a t-means test which can be found in Appendix 10. Surprisingly the means where different for caps 5 and 7. These are the caps with almost no, or no errors. An explanation for this would require further research. However, the hypothesis is rejected.

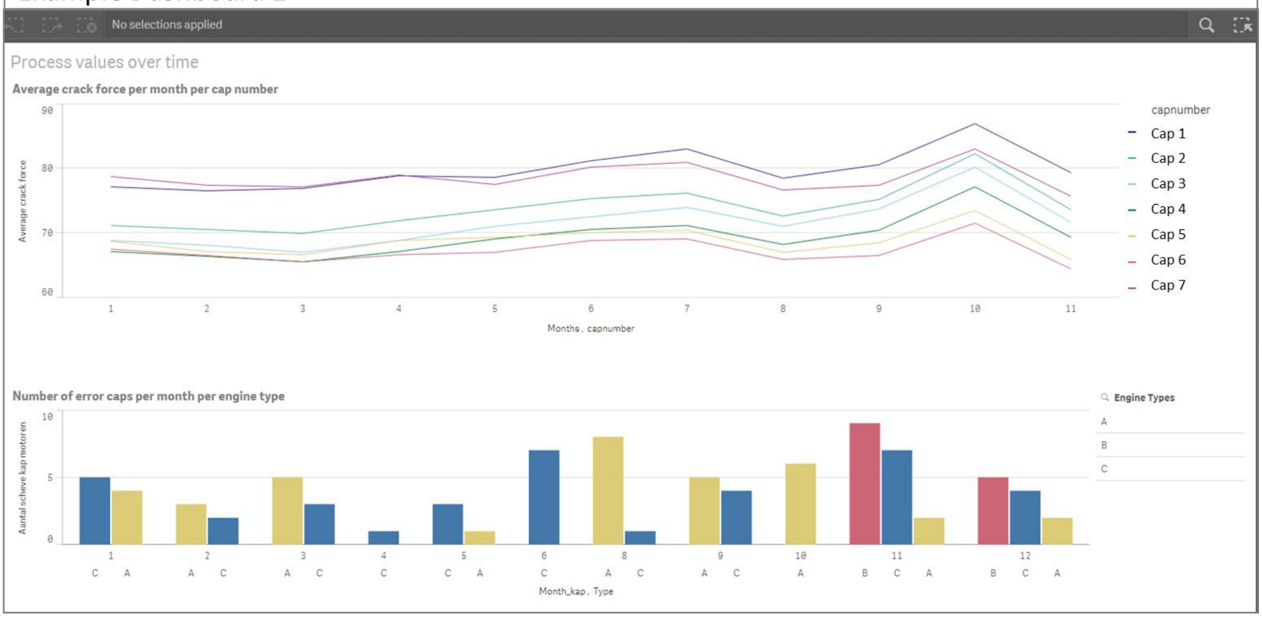
Appendix 12. Example Dashboards

The values in the dashboards are example values for confidentiality reasons.

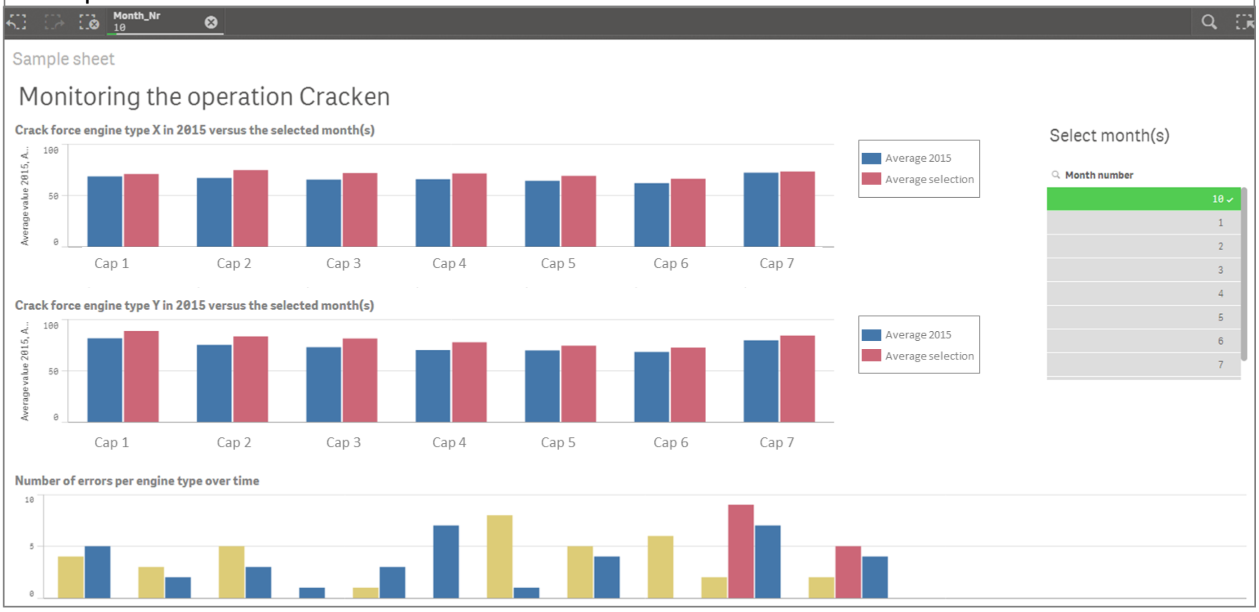
Example dashboard 1



Example Dashboard 2

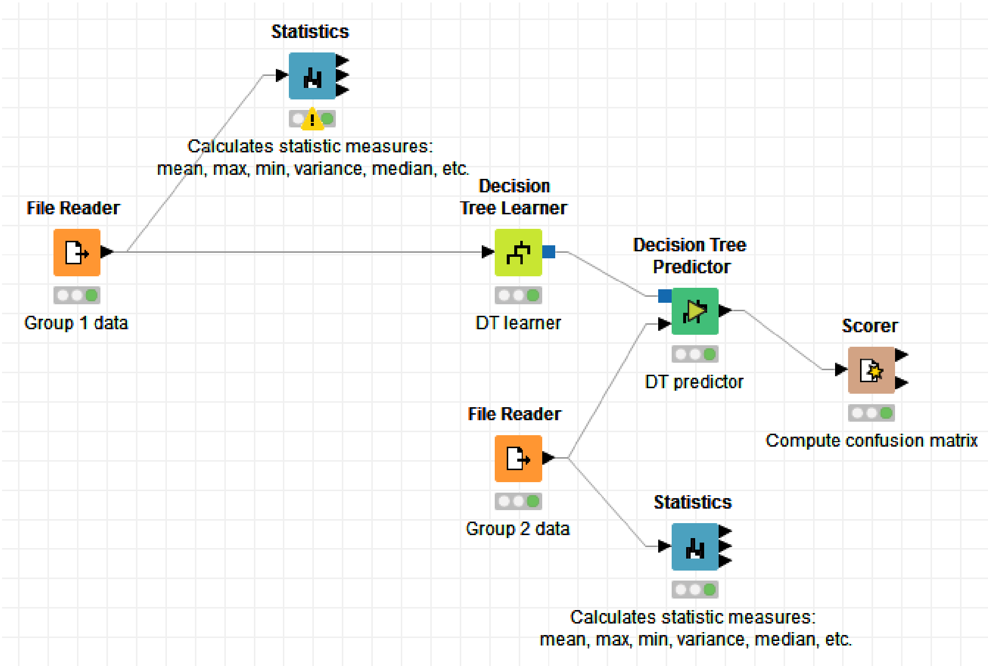


Example Dashboard 3

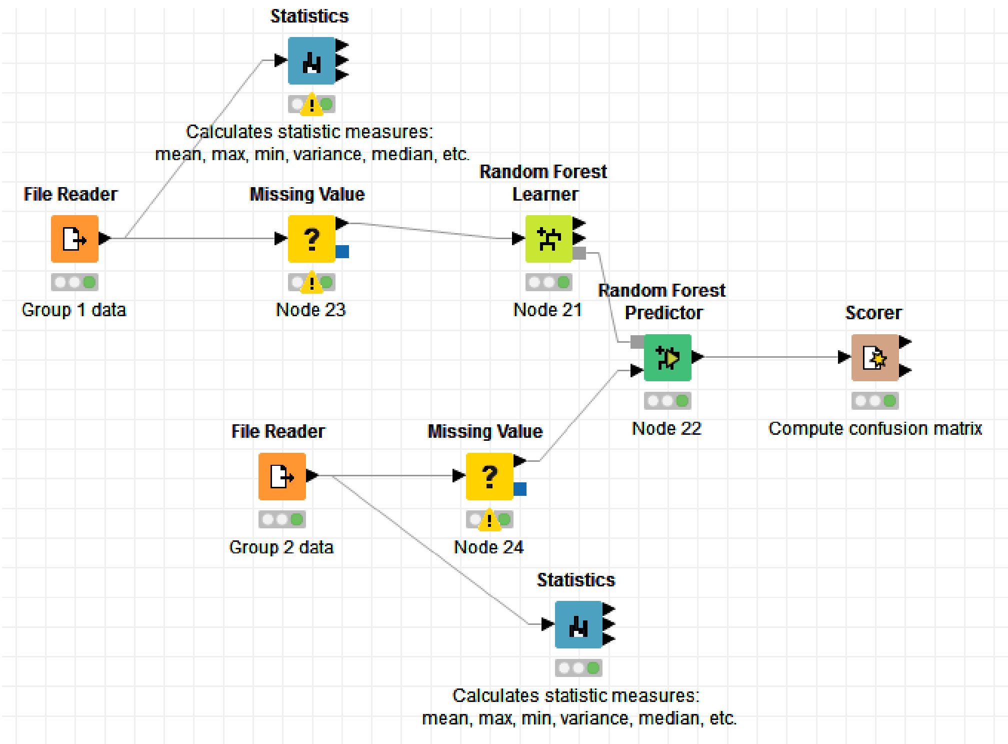


Appendix 13. KNIME models

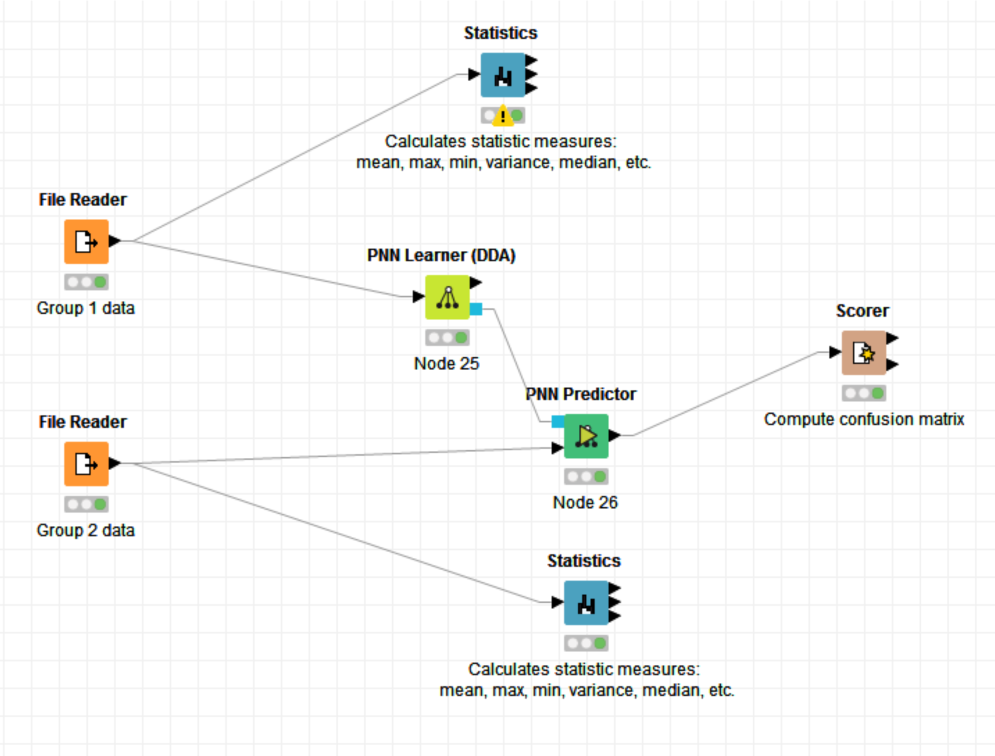
Decision tree model



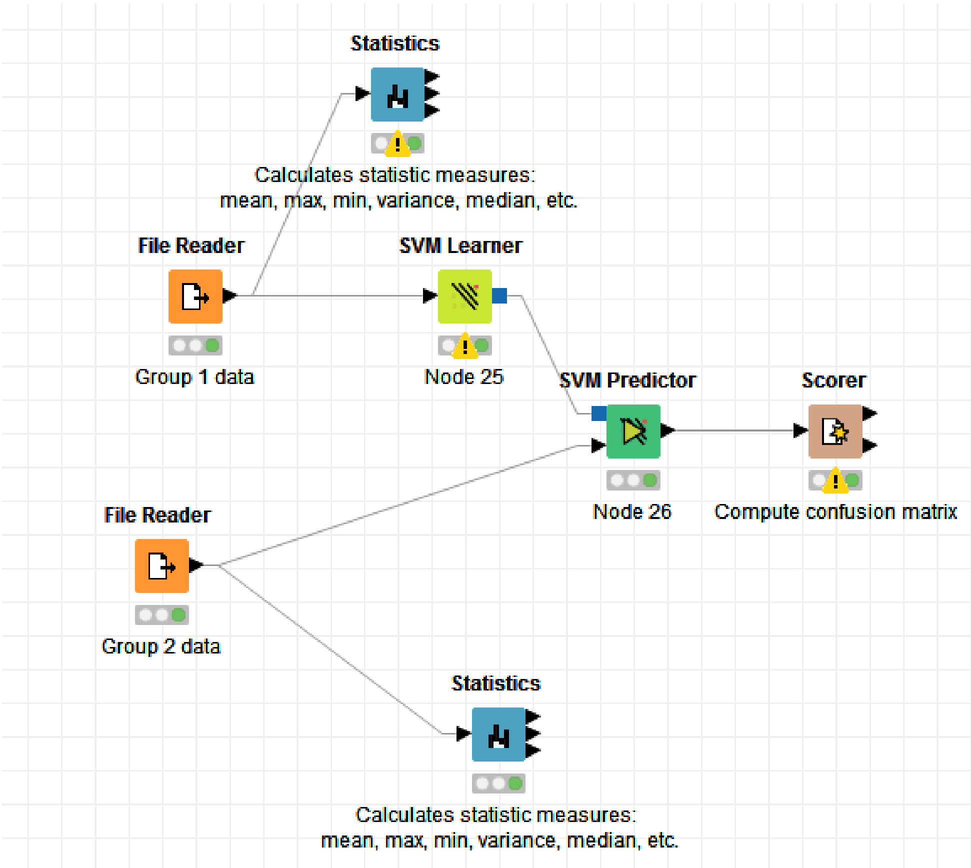
Random Forest model



Probability Neural Network



Support Vector Machines



Appendix 14. Interview Information Matrices usability questions

Questions about the Current MES Informational Matrix

1. **Do you feel that the Current MES Informational Matrix reflects on reality and how you experience MES?**
2. **Does it help with creating an overview of the informational abilities of a MES?**
3. **Do you feel that the Current MES Informational Matrix is useful to your company and in what way?**
4. **Do you feel that the Current MES Informational Matrix is useful to any company and in what way?**

Questions about the proposed dashboards

5. **Do you feel that the proposed dashboards (of dashboards in general) help to get an overview of the (most important) information and does it help with the production management?**

Questions about the Future MES Informational Matrix

6. **Do you feel that the Future MES Informational Matrix helps with getting insights in the advanced possibilities with MES data and help you prepare for the future?**
7. **Which informational elements from the Future MES Informational Matrix would be most interesting for your company?**

Questions about the Root Cause analysis

8. **How do you reflect on the Root Cause analysis conducted and presented at your company, both the approach and the results?**

Appendix 15. Interview transcripts

Interview 1

Function: Project Manager PE Engines
Date and Time: 29-03-2016 10:00h

<u>Questions about the Current MES Informational Matrix</u>

1. **Do you feel that the Current MES Informational Matrix reflects on reality and how you experience MES?**

Yes it conform what I would expect. There are no surprises in the informational items from which I would not expect to be possible to extract from a MES.

2. **Does it help with creating an overview of the informational abilities of a MES?**

As I am in the middle of a project which concerns developing a new MES for our company, I already have a lot of knowledge of what is possible with MES. Therefore it does not bring any new information for me. The ranking with what are important elements according the experts/industry is useful for us to be able to benchmark our MES. This helps us in assessing whether we are complete and if we are missing important elements.

For people within our company who are working less with MES it is also useful to have an overview of the possibilities.

3. **Do you feel that the Current MES Informational Matrix is useful to you company and in what way?**

As mentioned before, mostly as a benchmark a check whether we thought about everything in the new MES project. It could give insights for adding elements or at least of thinking about each elements and deciding why we should or should not add it.

4. **Do you feel that the Current MES Informational Matrix is useful to any company and in what way?**

It could be useful for every (production) company with MES to have an overview of the informational capabilities and the ranking of which are considered most useful/ important.

<u>Questions about the proposed dashboards</u>

5. **Do you feel that the proposed dashboards (of dashboards in general) help to get an overview of the (most important) information and does it help with the production management?**

Yes these could be very useful indeed. We want to make more use of dashboards in our new MES as well. However we chose not to add the dash boarding functionality to the MES itself but to link a business intelligence tool to the MES. So the MES will provide the data, but the BI tool will present the dashboards and do the analyses.

We believe it is very useful to proactively use the dashboards because when you only use them when there is a problem, you will always be too late. It is important to determine which parameters are most important and to closely monitor these. This will be in multiple screen that will vary between general information and process part specific information. Then we can monitor the process and see when it is getting near critical boundaries so we can proactively take action.

<u>Questions about the Future MES Informational Matrix</u>

6. **Do you feel that the Future MES Informational Matrix helps with getting insights in the advanced possibilities with MES data and help you prepare for the future?**

Is looks useful at first sight. We missed some important aspects of MES in our previous MES so this would help to be prepared for the future and to what will be possible with MES data. It is important for us to learn from what we missed in our former MES to prevent mistakes in our new MES. It is useful to know what is possible according to literature though we are already far in the process of developing the new MES. Therefore it is a bit too late for us to be highly relevant right now.

7. **Which informational elements from the Future MES Informational Matrix would be most interesting for your company?**

- Identification of critical process parameters
- Knowledge of operational process(es)
- Root cause analysis
- Detection of change points in control charts
- Monitoring process conditions (especially in assembly operations)
- Detection of abnormal process behavior
- Forecast production cycle time
- Predict process performance

Interview 2

Function: Senior PE Project Manager
Date and Time: 29-03-2016 11:00h

<u>Questions about the Current MES Informational Matrix</u>

1. **Do you feel that the Current MES Informational Matrix reflects on reality and how you experience MES?**

Yes they it is conform what one could expect. I can relate more to some element names then others. Most informational elements I would consider useful. [We go through the list and extra explanation for some items is provided]. I believe for the root cause analysis it is always conducted by humans by extracting information from a MES, for us by making use of Minitab. Material compatibility would be something I would measure on the floor and not in MES. Also weight and dispense responds would only be useful when having a recycling process for the waste or dispense. Tracking non-productive feels like over automating a waste step.

2. **Does it help with creating an overview of the informational abilities of a MES?**

It is an overview but for me I would not use it. What information you extract from MES is always demand driven and not driven by possibilities and a MES expert ranking. That something is ranked high according to experts does not mean that we need it. It is always demand driven. For example we have a problem, in order to solve that we need certain information and then we research whether our vendor or a different vendor can provide it as a software tool/package or that they can custom build it (dependent of the costs off course).

3. **Do you feel that the Current MES Informational Matrix is useful to your company and in what way?**

As mentioned before, it is an overview but we have incorporated the functionality in MES that we found useful and did not incorporate what we did not found useful. For adding functionality it is always demand driven.

4. **Do you feel that the Current MES Informational Matrix is useful to any company and in what way?**

It is difficult to guess whether it would be useful to other companies. The desired to add (informational) functionality to a MES always rises from the specific process.

Questions about the proposed dashboards

5. **Do you feel that the proposed dashboards (of dashboards in general) help to get an overview of the (most important) information and does it help with the production management?**

I believe that dashboard functionality is one of the main reasons to have a MES next to logging data and using it for traceability reasons. Dash boards are especially useful for continuous processes (like ours) where a constant flow of product is produced in a line. With a dashboard we can monitor this process continuously and proactive. We need to know the critic elements and monitor those. In our new MES we want to improve the dashboards (currently there are limit dashboard available) with WIP, Machine status, process factors and these critical parameters.

Questions about the Future MES Informational Matrix

6. **Do you feel that the Future MES Informational Matrix helps with getting insights in the advanced possibilities with MES data and help you prepare for the future?**

I believe that here the same applies as before. There has to be a need for information (demand driven), and then you research how you can obtain this information by research vendors and software possibilities. It would be more useful for us to have a list of vendors and what they have to offer in this field. The ranking of experts is not that useful for as the demand for information is different for each company. We have to focus on our process and our demand for information and research how we can get this.

7. **Which informational elements from the Future MES Informational Matrix would be most interesting for your company?**

We already research critical process parameters. When we know these it is interesting to monitor those with a dashboard or other functionality. The root cause analysis will always be conducted by humans as the system only provides the information that humans request and the humans will always conclude. Human interpretation is always key.

Questions about the Root Cause analysis

8. **How do you reflect on the Root Cause analysis conducted and presented at your company, both the approach and the results?**

We also use a similar approach with hypothesis driven research. We brainstorm about possible causes (cause and effect diagram) and determine which are most likely. Then we gather the data from MES and analyze the data hypothesis driven. If we find the root cause we use this knowledge to change the process to prevent the problem from happening in the future.

Interview 3

Function: Area Manager Engine factory
Date and Time: 30-03-2016 11:30h

Questions about the Current MES Informational Matrix

1. **Do you feel that the Current MES Informational Matrix reflects on reality and how you experience MES?**

I do not use all of our current MES functionality so I am no experts but there are no major surprises in the informational elements. I usually do not use MES data, I only look at the conclusions of a MES data analysis.

2. **Does it help with creating an overview of the informational abilities of a MES?**

I do not know whether we have an overview already and which item we have in MES and which we do not have in MES. So yes this overview seems useful but it is hard for me to tell.

3. Do you feel that the Current MES Informational Matrix is useful to your company and in what way?

As I mentioned earlier, it is hard for me to determine that.

4. Do you feel that the Current MES Informational Matrix is useful to any company and in what way?

Depending on what they already have in MES and how much they are missing it could be useful. It is hard for me to determine.

<u>Questions about the proposed dashboards</u>

5. Do you feel that the proposed dashboards (of dashboards in general) help to get an overview of the (most important) information and does it help with the production management?

A while ago I asked out IT Department whether something like this was possible. We feel it is important for people on the floor to know what is happening with the process. We have a few people responsible for multiple machine so it can be hard for them to keep the overview. A dashboard which can be viewed from every angle of the factory would help them to get a fast overview and to prioritize what is important. Especially since they have to walk quite a distance between the multiple machines. Our IT Department proposed a dashboard that looks similar to the example. Linking this dashboard to MES could be an added value. We will have a pilot soon with the dashboard on a screen and in a cloud environment so that it can be accesses from anywhere.

It might also be useful to add the control cards to a MES environment. Then MES could predict and we can be triggered by MES to take action instead of doing test and measurement on every machine every once in a while.

<u>Questions about the Future MES Informational Matrix</u>

6. Do you feel that the Future MES Informational Matrix helps with getting insights in the advanced possibilities with MES data and help you prepare for the future?

It is nice to know what will be possible in a near future in practice. Sometimes it seems that if everything is working perfect, then the options are endless. However, the real world has though me that not everything goes as smoothly as proposed. Also many opportunities will never be implemented. There is always this feeling of that I want to see it before I believe it.

7. Which informational elements from the Future MES Informational Matrix would be most interesting for your company?

I think condition based monitoring would be most useful.

<u>Questions about the Root Cause analysis</u>

8. How do you reflect on the Root Cause analysis conducted and presented at your company, both the approach and the results?

I think it was a useful analysis. Upfront I thought that the data would provide quite some useful insights, but unfortunately it did not. This probability indicates that the cause to the problem is broad or that is has multiple causes. It was a relative fast analysis that eliminated some possible causes. In the factory we usually think of what the problem could be and do research on testing that. Now you started with addressing multiple hypothesis and researched them subsequently. This is also the approach of the black belt employee who does more of these analyses.

Interview 4

Function: Managing director engine factory
Date and Time: 30-03-2016 11:45h

<u>Questions about the Current MES Informational Matrix</u>

1. **Do you feel that the Current MES Informational Matrix reflects on reality and how you experience MES?**

It is conform what I would expect. There are no big surprises of informational items in the list. There are some eye openers I did not think about for integrating it in MES. For example the personnel tracking, I would think of another system for this.

2. **Does it help with creating an overview of the informational abilities of a MES?**

Yes it is useful to have an overview. However there is always a differences between what theoretically possible and what is possible in practice. Also not all information in MES is useful to measure with MES. The technology push in these subjects is very high but the demand can be different. I think that all items that are rated 50% or higher by experts are useful to integrate in MES, the other are less useful.

3. **Do you feel that the Current MES Informational Matrix is useful to your company and in what way?**

We are already far developed in our journey of developing the new MES and the new MEs landscape. We will re-scope our MES and limit it to core functionality and measure information in our main frame and other systems.

4. **Do you feel that the Current MES Informational Matrix is useful to any company and in what way?**

If you do now have a MES (or are the beginning of redeveloping one), it is very useful to identify the information need based on this list. There is no overview of the abilities of MES. When contacting a vendor for the possibilities, you immediately get asked back what you are looking for. This leads to a cycle because it is hard to identify your needs when you do not know what is possible and the vendor cannot tell you what is possible because he does not know your needs. Having this overview with a ranking can help a company with establishing a discussion to identify the needs.

Also I believe that MEs is a strategic choice which is not ROI driven. MES has benefits in efficiency, traceability, monitoring etc. but is hard to create a summation of benefits to calculate a (positive) ROI. Because you cannot simply calculate the ROI, the discussion about MES has to be more fundamental in why you want this system and what for. Then you can test potential requirements better.

<u>Questions about the proposed dashboards</u>

5. **Do you feel that the proposed dashboards (of dashboards in general) help to get an overview of the (most important) information and does it help with the production management?**

The first dashboard is an example of what I would like to have implemented tomorrow. We already collect this information now but it is reported by team leader manually. This consumes a lot of time. This is what I am looking for. A fast overview of the process status, generated in an automated way. I would like to add some zoom in functions per process step or buffer position so that I can see which products are where.

The second and thirds example frameworks is what we want to establish outside our new MES with another system and an integrated business intelligence engine. We are still discussing how this would look like. This is an example for it but it depends of the requirement we have for that environment.

Questions about the Future MES Informational Matrix

6. **Do you feel that the Future MES Informational Matrix helps with getting insights in the advanced possibilities with MES data and help you prepare for the future?**

Both Yes and No. No because it is hard to oversee it at this point. Yes because I see there are interesting informational items in the main group root cause analysis. With advanced data analysis techniques that take into account multiple parameters, you can find relationships that you could not uncover before.

However I do feel there are also many informational items in the list that I feel should not be an output of MES but should be knowledge present among the workers. People should have exhaustive process based on experience and learning the process, not because MES told them. However new techniques could generate additions and triggers for research to expand that knowledge. I can imagine that this is specific to our company and process as you can see very well what is happening. I more high tech industries where this is harder, and the knowledge is more complicated, I can imagine that more knowledge could be produced by a system.

7. **Which informational elements from the Future MES Informational Matrix would be most interesting for your company?**

Our current challenges are still fundamental and therefore we are not working towards big data or data mining related solution. First have to make sure that what we do now, we do good. Also that the data we store now is usable for our current challenges. When we have that under control we will think of adding big data and data mining related functionality. However we do discuss these items to know what will be the future in order to be prepared.

Questions about the Root Cause analysis

8. **How do you reflect on the Root Cause analysis conducted and presented at your company, both the approach and the results?**

I think you followed the same steps that our project managers and black belts follow as well though you had less process knowledge so it was based on first observations. I think it would have been interesting to incorporate data from the assembly operations. It is a pity you did not find the solution for us.

Interview 5

Function: Head of PE machine process Engine Factory
Date and Time: 31-03-2016 12:30h
Questions about the Current MES Informational Matrix

1. **Do you feel that the Current MES Informational Matrix reflects on reality and how you experience MES systems?**

Yes, though I do see some informational elements that we do not have in our MES. I missed some MES capabilities like production support, code number alterations and ready notifications (these are out of scope as they do not provide information but are part of the operational functions of MES). Overall it is very representative to what I would expect.

2. **Does it help with creating an overview of the informational abilities of a MES system?**

I know that the actual MES implementation can be very different and very company and process specific. As was involved in developing the current MES I already have a view of the current possibilities in MES.

3. **Do you feel that the Current MES Informational Matrix is useful to your company and in what way?**

Yes with this overview I can see what is possible which will help me in the internal discussions and in discovering our specific needs.

4. **Do you feel that the Current MES Informational Matrix is useful to any company and in what way?**

The same as for us. It help giving you an insight in the possibilities, with an experts ranking and then a company can decide for themselves what they need. Then they can choice what they want to configure in their MES.

<u>Questions about the proposed dashboards</u>

5. **Do you feel that the proposed dashboards (of dashboards in general) help to get an overview of the (most important) information and does it help with the production management?**

Yes I think it is very useful to have your informational available in an interactive manner. However I do miss the control limits in the graph so that I can see how far the values are from the critical limits.

<u>Questions about the Future MES Informational Matrix</u>

6. **Do you feel that the Future MES Informational Matrix helps with getting insights in the advanced possibilities with MES data and help you prepare for the future?**

It is similar as the first matrix. It generates an overview which can be helpful for discussion but what you implement in MES is a choice.

7. **Which informational elements from the Future MES Informational Matrix would be most interesting for your company?**

For Production Operations:

- Identification of critical process parameters
- Knowledge of operational process(es)
- Detection of change point in control charts
- Monitoring process conditions
- Identification of critical process parameters
- Detection of abnormal process behavior
- Forecasting production cycle time
- Forecasting production process performance

For Quality Operations

- Defect/low quality classification (all three)

For Maintenance Operations

- Forecasting machine/equipment failure
- Forecasting component failure
- forecasting machine performance
- Forecasting tool wear
- Diagnostics of machine failure

<u>Questions about the Root Cause analysis</u>

8. **How do you reflect on the Root Cause analysis conducted and presented at your company, both the approach and the results?**

There were no big surprises during your presentation. I also expected that I would be very difficult to find the actual root cause so it is no surprise that you could not find it. Your approach is also how we do it at the office. The project manager and black belt also use this hypothesis testing approach. In the factory we do more experiments when we think something could be causing it.

Interview 6

Function: Supervisor Engine Factory Line 2
Date and Time: 31-03-2016 18:00h

Questions about the Current MES Informational Matrix

1. **Do you feel that the Current MES Informational Matrix reflects on reality and how you experience MES systems?**

I mostly use the traceability functionality for engines that we sent to other parties. Before we sent them, we check whether all operations where ok. When I look at your matrix I see no big surprises. I think this is functionality I would expect from a MES system. In the lowest ranked scores I see some less recognizable items.

2. **Does it help with creating an overview of the informational abilities of a MES system?**

Yes absolutely! I kind of see ERP functionality in this like work in process and OEE. Also knowing where the engines blocks are in the production line. Now I see that this is also possible in MES. However what is possible and what we have can be different as we have many systems co-operating.

3. **Do you feel that the Current MES Informational Matrix is useful to your company and in what way?**

Yes I think it is useful. Especially for production operations.

4. **Do you feel that the Current MES Informational Matrix is useful to any company and in what way?**

Yes I think this is relevant to every company that uses MES, especially in producti on.

Questions about the proposed dashboards

5. **Do you feel that the proposed dashboards (of dashboards in general) help to get an overview of the (most important) information and does it help with the production management?**

These dashboards are great so I would definitely say yes. Now we have self-build excel files as 'dashboards' to monitor our process. If we can do this in MES it will be automatically filled plus it will be interactive and fast, especially example 1 I would like to see the amount of errors per process part to quickly have an overview and to act upon it. Example 2 and 3 would be useful to enable us to detect when a process is heading for its critical boundaries. This could be combined with control charts. Overall dashboards would enables us to effective executive the Total Productive Maintenance (TPM) as we have the informational we need everywhere and always real-time.

Questions about the Future MES Informational Matrix

6. **Do you feel that the Future MES Informational Matrix helps with getting insights in the advanced possibilities with MES data and help you prepare for the future?**

It cautiously say yes because it is difficult for me to recognize all terms and have a good understanding of what everything means. When I look at the terms and the scores I do agree with the ranking. I would have provided a similar ranking as the high scoring element help creating a more solid process.

7. **Which informational elements from the Future MES Informational Matrix would be most interesting for your company?**

I think it data can help us with our challenges and questions it is always good. Especially the root cause analysis is very important in our continues improvement focus.
Element that I think are relevant most relevant are:

- Identification of critical process parameters
- Improved knowledge of the production process
- Optimization of parameters settings
- Conditional based maintenance (!)
- Patterns for variation
- Decision support (!)

<u>Questions about the Root Cause analysis</u>

8. **How do you reflect on the Root Cause analysis conducted and presented at your company, both the approach and the results?**

It was conform what I expected. I was surprised by the amount of data that you had for your analysis, it did not know we had all the data and information in MES (or that we were able to get extract it from MES data). I experience that my team and myself have some assumptions about the process and how thinks are. Your fresh view was refreshing. I saw some results that I assumed as well, that some caps have more errors then others, though it happens in all caps. In practice I sometimes experience that we have an idea of what is the problem/cause and that we immediately start with an experiment to test it. Your approach was more structured. However we both did not find the solution yet.