DTU Library

Model-based Process Surveillance and Optimisation for Fault Detection and Diagnosis

Rønsch, Georg Ørnskov

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Industrial PhD Thesis
Doctor of Philosophy in Data utilisation within Industry 4.0

DTU Compute
Institut for Matematik og Computer Science

Model-based Process Surveillance and Optimisation for Fault Detection and Diagnosis

Georg Ørnskov Rønsch

Kongens Lyngby, 2022

Supervisor: Associate Professor, Murat Kulahci
Co-supervisors: Hans Martin Stage & Anita Friis Sommer

DTU Compute
Institut for Matematik og Computer Science
Technical University of Denmark
Richard Petersens Plads
Building 324
2800 Kongens Lyngby, Denmark

Summary

With the introduction of Industry 4.0, focus on digitisation and digitalisation have become very apparent. Many companies have this high on their agenda, and many resources are spent on adjusting the overall ideas and concepts to fit within the context of the given company. One of the central aspects of Industry 4.0 is the utilisation of data to achieve increased productivity, new products and business opportunities. This industrial PhD thesis has been conducted in close collaboration with an industrial partner and focuses on utilising manufacturing data to achieve increased productivity. The industrial partner is a leading producer of injection moulded construction toys. Therefore, the work is related to injection moulding; however, the findings and learnings generalise well to other industries.

Initially, the formulation of the project scope was related to increasing productivity through monitoring and control of the injection moulding process. The intent was to utilise newly established machine connectivity at the industrial partner to collect readily available machine process data and explore how best to develop and implement a process monitoring solution. The expected and needed data was not collected because of the complexity encountered when initiating the data collection and changing prioritisation. This setback provoked reflections on the purpose and approaches for data utilisation in manufacturing. Eventually, it changed the project's focus, however, still with the overall focus on improving productivity. The new focus became centred around how to become effective in scoping and exploring data utilisation[1].

Utilising data in a manufacturing setting is a demanding task with many obstacles to pass. In the past and during the PhD project, this has caused many frustrations and challenges, which have been condensed and transformed into a data utilisation framework, supporting a practical exploration and implementation approach. The framework was first presented in paper A (*"An investigation of the utilisation of different data sources in manufacturing with application in Injection Moulding"*) and later extended to include more learnings and aspects of data utilisation in a manufacturing context. The central part of the framework is formulating a core business objective, followed by identifying key manufacturing drivers that have to be addressed to improve the objective. For each key driver, a number of applications are identified as potential contributors for optimising the business objective. At this stage, the

[1] A more describing title for the project would therefore be, *"Practical approach and framework for data utilisation and productivity optimisation within injection moulding"*, however, the original title still cover the content and have therefore not been changed.

work moves into a practical exploration of the identified applications. This is proposed to be conducted sequentially, starting with applications that require the least new data integration and collection. As the maturity in the organisation is growing, more complex data integration and utilisation can be initiated. All applications should be evaluated against the total cost of utilising a given data source and the potential value creation.

The thesis is a collection of six scientific papers where five of these include examples of data utilisation applications in a manufacturing context[2]. Three papers address aspects of predictive maintenance to reduce productivity loss caused by mould failure. Two of these papers utilise existing quality data and maintenance history data for mould worn-out evaluation, where the third explores the use of acoustic emissions for real-time condition-based maintenance of injection moulds. The use of acoustic emissions have proven promising for detecting the lack of lubrication and mechanical defects represented by loss latch lock. The remaining two papers explore utilising underlying equipment time-series data to capture process dynamics not detected in readily available process data. The changing process dynamic explored is caused by variations in raw materials and impacts product quality. Using designed experimentation and machine learning, a solution is presented to detect the disturbance and identify the optimal process settings to reduce the effects on product quality. Besides the work presented in the scientific publications, the thesis includes initial exploration using a multivariate approach to monitor product quality and the injection moulding process. This work is included since it represents some of the challenges of utilising data in a manufacturing context and is important for the overall reflections, recommendations, and conclusions.

> *"Data is the new oil. It's valuable, but if unrefined, it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value."*
>
> — Clive Humby, UK Mathematician and architect of Tesco's Clubcard

The quote by Clive Humby is very relevant and widely used in relation to data utilisation within Industry 4.0, and often what upper management in companies refer to when initiating the journey of data utilisation. One thing that is seldom realised is the complexity related to this and the amount of resources and dedication required to make it happen. This should by no means be a reason for not starting the data utilisation journey, just a heads-up, to align expectations.

[2]The last paper collects experiences gathered by our university group (DTU Compute, Section for Statistics and Data Analysis) working in close collaboration with various companies utilising data in manufacturing contexts.

Resumé (Danish)

Med introduktionen af Industri 4.0 er fokus på digitalisering blevet meget tydeligt. Mange virksomheder har det højt på dagsordenen, og der bruges mange ressourcer på at tilpasse de overordnede ideer og koncepter, så de passer ind i konteksten for den givne virksomhed. Et af de centrale aspekter af Industry 4.0 er udnyttelsen af data til at opnå øget produktivitet, nye produkter og forretningsmuligheder. Dette Ph.d.-projekt er udført som en erhvervs-ph.d. i tæt samarbejde med en industriel partner og fokuserer på at udnytte produktionsdata til at opnå øget produktivitet. Den industrielle partner er en førende producent af sprøjtestøbt byggelegetøj, hvorfor arbejdet er relateret til sprøjtestøbning. Resultaterne og læringen generaliserer dog godt til andre industrier.

Oprindeligt var formuleringen af projektet relateret til forbedring af produktiviteten ved introduktion af proces monitorering. Hensigten var at udnytte en nyetableret datainfrastruktur hos den industrielle partner til indsamling af tilgængelige procesdata og undersøge, hvordan disse bedst udnyttes til udvikling og implementering af en løsning til proces monitorering. De forventede og nødvendige data blev ikke tilgængelige på grund af kompleksiteten forbundet med dataindsamlingen samt ændret prioritering. Dette ledte til refleksioner omkring det overordnede formål med dataudnyttelse og hvordan dette blev grebet an. Dette resulterede i et ændret fokus, dog stadig med det overordnede fokus på at forbedre produktiviteten, men nu mere centreret omkring, hvordan en effektiv definering af formålet og strukturering ville kunne forbedre og lette dataudnyttelse i en produktionsvirksomhed[3]

Udnyttelse af data i et produktionsmiljø er en krævende opgave med mange forhindringer og faldgrupper. Dette har under ph.d.-projektet, og i tidligere situationer, medført mange frustrationer og udfordringer. Dette er blevet transformeret og omsat til en model for effektiv dataudnyttelse, der understøtter en praktisk udforsknings- og implementeringsstrategi. Modellen blev først præsenteret i artikel A (*"En undersøgelse af anvendelsen af forskellige datakilder i fremstilling med anvendelse i sprøjtestøbning"*) og senere udvidet til at omfatte flere erfaringer og aspekter af dataudnyttelse i en produktion kontekst. Den centrale del af modellen er at formulere et konkret og centralt forretningsmål, efterfulgt af identificering af kerneområder, der skal behandles for at forbedre målet. For hvert kerneområde identificeres en række

[3]En mere beskrivende titel kunne derfor være: *"Praktisk tilgang og rammer for dataudnyttelse og produktivitetsoptimering inden for sprøjtestøbning"*, den originale titel dækker dog stadig indholdet og er derfor ikke blevet ændret.

applikationer, som potentielt kan bidrage til at optimere forretningsmålet. Næste skridt er praktisk udforskning af de identificerede applikationer. Det anbefales, at dette udføres sekventielt, startende med applikationer, der kræver mindst ny dataintegration. I takt med at modenheden i organisationen vokser, kan mere kompleks dataintegration igangsættes. Alle applikationerne bør vurderes i forhold til de samlede omkostninger ved at bruge en given datakilde og den potentielle værdiskabelse.

Afhandlingen er en samling af seks videnskabelige artikler, hvor fem af disse indeholder eksempler på applikationer der udnytter data i en produktions kontekst[4]. Tre artikler omhandler aspekter af prædiktive vedligehold for at reducere produktivitetstab forårsaget af defekter på støbeforme. To af disse artikler anvender eksisterende kvalitetsdata og vedligeholdelseshistorik til evaluering af udslidte forme, hvor den tredje udforsker brugen af akustiske emissioner til tilstandsbaseret vedligehold i realtid af støbeforme. Brugen af akustiske emissioner har vist sig at være lovende til at detektere manglen på smøring og mekaniske defekter repræsenteret ved klinketræk.

De resterende to artikler undersøger brugen af underliggende tidsseriedata fra produktionsudstyret til at fange procesdynamikker, der ikke kan fanges med standard tilgængelige procesdata. De undersøgte procesdynamiske variationer er forårsaget af variationer i råmaterialer og påvirker produktkvaliteten. Ved hjælp af designede eksperimenter og dataanalyse præsenteres en løsning til at detektere materialevariationerne og identificere de optimale procesindstillinger for at reducere indvirkningen på produktkvaliteten. Udover arbejdet præsenteret i de videnskabelige artikler, indeholder afhandlingen indledende undersøgelse af muligheden for at anvende en multivariat tilgang til procesmonitorering af sprøjtestøbningsprocessen. Dette arbejde er inkluderet, da det repræsenterer nogle af udfordringerne der kan være forbundet med brugen af data i en produktionssammenhæng. Disse læringer er vigtige i forhold til de overordnede refleksioner, anbefalinger og konklusioner.

> *"Data is the new oil. It's valuable, but if unrefined, it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value."*
>
> — Clive Humby, UK Mathematician and architect of Tesco's Clubcard

Citatet af Clive Humby er meget relevant og meget brugt i forhold til dataudnyttelse inden for Industri 4.0, og ofte refereret til af ledelsen i virksomheder når de igangsætter aktiviteter med dataudnyttelse. En ting, der sjældent indses, er kompleksiteten i forbindelse med dette og mængden af ressourcer og dedikation, der kræves for at få det til at ske. Dette bør på ingen måde være en grund til ikke at starte dataudnyttelsen, blot en heads-up, for at afstemme forventningerne.

[4]Den sidste artikel er en sammenfatning af erfaringer opnået som universitetsgruppe (DTU Compute, Sektion for Statistik og Dataanalyse) gennem tæt samarbejde med forskellige virksomheder

Preface

This thesis is a result of an industrial PhD, conducted by Georg Ørnskov Rønsch at the department of Applied Mathematics and Computer Science, Section for Statistics and Data Analysis at the Technical University of Denmark (DTU). All work has been conducted in close collaboration with an industrial partner. Based on company policies, the company wants to be anonymous and will, therefore, only be mentioned as the ”Industrial partner” throughout this thesis. The project is funded by the Danish Innovation Fund and the industrial partner. The project's main supervisor has been Associate Professor Murat Kulahci from DTU Compute, and co-supervisors have been Hans Martin Stage and Anita Friis Sommer from the industrial partner. The external research stay has been conducted at Aalborg University, Department of Electronic Systems, under the supervision of Professor Zheng-Hua Tan.

The PhD thesis deals with aspects of Industry 4.0 within utilising manufacturing data for optimisation of productivity. The thesis includes six research papers with specific examples of data utilisation within injection moulding and reflections on challenges with data utilisation in a manufacturing context.

> *“The Fourth Industrial Revolution is still in its nascent state. But with the swift pace of change and disruption to business and society, the time to join in is now.”*
> — Gary Coleman (2014), *Global Industry and Senior Client Advisor, Deloitte Consulting*

Kongens Lyngby, 28th February 2022

Georg Ørnskov Rønsch

Acknowledgement

The idea of starting as a PhD student was randomly sparked at a MADE meeting with Murat Kulahci and Peter Paasch Mortensen (quote Peter ”Why have you never done a PhD, you should do it, it is never too late..!”). Several chats with Murat convinced me that doing an industrial PhD was an excellent opportunity to learn new skills within machine learning and explore some of the ideas we had on data utilisation in manufacturing. So a great thanks to Murat and Peter for initiating it all. Going into the PhD journey more than 15 years after finishing my master's within chemical engineering, was quite a decision. Looking back, would I do it all again? No way! Am I glad that I started and completed the journey? Yes indeed! This PhD project has been exciting, educational, fun and challenging at times. Thanks to Innovation fund Denmark for believing in the application and supporting the project.

> *"You can't change the world alone - you will need some help - and to truly get from your starting point to your destination takes friends, colleagues, the goodwill of strangers and a strong coxswain to guide them."*
>
> — William H. McRaven

As beautifully worded by William H. McRaven, this journey would not have been possible without outstanding support. I dedicate this section to the people who have influenced, affected, and supported me throughout the past three years. Without these people, this PhD project and thesis would not have been possible.

Murat, thanks for all the support, feedback and inspiring dialogues. It has been fantastic to have you as a supervisor, and I would not have been able to navigate the project without your support and encouragement. We still need to explore so much, so hopefully, we can continue the collaboration.

None of this work would have been possible without all the condition-less and outstanding support from colleagues at the industrial partner. What we can accomplish together being curious, brave, and focused is genuinely unique. I am proud of being part of this and looking forward to continuing the data utilisation journey in the world's most fantastic company. There are a group of colleagues that I need to give a dedicated thanks for the tremendous support. I want to thank Jesper Kjærsgaard Christensen, Birgitte Høst-Madsen and Lars Mose Henriksen for input and feedback, making the explored applications business relevant. A special thanks to Martin Dybdahl for a fantastic collaboration on paper A and B, to Till Böttjer for collaboration on paper F, and Casper Solheim Bojer for joint exploration of possibilities for utilising readily available process data. Bo Skovbjerg, Tommy Tungelund Kristiansen, Anders Andreassen, Anders Lautrup Trans, Zhiyi Cheng, Hjalte Tind Nygaard and Peder Morten Albin Jensen for support conducting all the needed experimentation. Hans Martin Stage and Anita Friis Sommer for great supervision and the Industry 4.0 team, Otto H.A. Abildgaard and Rikke Lind Iversen for input and reflections.

One of the privileges of being enrolled in an industrial PhD project is to have two groups of colleagues. It has been fantastic being part of the Statistics and Data Analysis group at DTU. I want to thank the group for open and inspiring dialogues and the joint effort on paper D. Special thanks to Flavia Dalia Frumosu for always having an open and positive mindset when I have presented some twisted ideas (and for taking the role as the "rubber duck"). Also, great thanks for the great collaboration on paper E. Thanks to Abdul Rauf Khan for great conversations and for pursuing the ideas of an auto ML solution for manufacturing companies.

My external stay was originally planned for Universitat Politècnica de València. Thanks to professor Alberto José Ferrer Riquelme for inviting me to Valencia for my external; I am sure I would have enjoyed it (I hope we will be able to complete the paper on process monitoring). COVID restrictions changed the plans, and great thanks to professor Zheng-Hua Tan for making it possible to have my external stay at Aalborg University, Department of Electronic Systems. Thanks for the great collaboration on paper C, and a special thanks to Iván López-Espejo for the dedicated effort.

Lastly, I would like to thank my beloved wife, Anette and my kids, Jonas and Mathias, for the needed understanding and support in making life as a PhD student fit in a busy family schedule.

List of contributions

The thesis is based on the following publications:

[A] **G. Ø. Rønsch**, M. Kulahci & M. Dybdahl (2021). An investigation of the utilisation of different data sources in manufacturing with application in injection moulding, International Journal of Production Research, DOI:10.1080/00207543.2021.1893853
Published

[B] **G. Ø. Rønsch**, M. Kulahci & M. Dybdahl (2022). Real-time adjustment of injection molding process settings by utilizing Design of Experiment, time series profiles and PLS-DA, Quality Engineering.
Accepted

[C] **G. Ø. Rønsch**, I. López-Espejo, D. Michelsanti, Y. Xie, P. Popovski and Z.-H. Tan (2022). Acoustic from injection moulding, exploration of use-cases within monitoring and mould maintenance, International Journal of Computer Integrated Manufacturing.
In review

[D] M. Kulahci, F.D. Frumosu, A. R. Khan, **G. Ørnskov Rønsch**, and M. P. Spooner (2019), "Experiences with Big Data: Accounts from a Data Scientist's perspective" Quality Engineering, DOI:10.1080/08982112.2019.1686641
Published

[E] F.D. Frumosu, **G. Ørnskov Rønsch**, and M. Kulahci, "Mould worn-out prediction in the plastic injection moulding industry: A case study," Quality Engineering (2019). DOI:10.1080/0951192X.2020.1829062
Published

[F] Till Böttjer, **Georg Ø. Rønsch**, Cláudio Gomes, Devarajan Ramanujan, Alexandros Iosifidis, and Peter Gorm Larsen (2021). Data-Driven Identification of Remaining Useful Life for Plastic Injection Moulds, CARV21 - Lecture Notes in Mechanical Engineering (LNME). DOI:10.1007/978-3-030-90700-6_49
Published

Publications not included in the thesis:

- M. Kulahci, F.D. Frumosu, A. R. Khan, **G. Ørnskov Rønsch**, and M. P. Spooner (2020), ”Rejoinder-Experiences with Big Data: Accounts from a Data Scientist’s Perspective” Quality Engineering.
  DOI:10.1080/08982112.2020.1808223
  Published

The work carried out in the PhD have been presented at:

(1) **G. Ø. Rønsch**, M. Kulahci & M. Dybdahl, ”An investigation of the utilization of different data sources in manufacturing with application in injection moulding”, ENBIS 2021 Spring Meeting (2021-05-17), **Presentation**

(2) **G. Ø. Rønsch** & M. Kulahci ”Indirect measure of material variation in raw materials”, dsk.2020, The biannual conference for the Danish Chemometrics Society (2020-11-06), **Presentation**

(3) **G. Ø. Rønsch**, M. Kulahci & M. Dybdahl, ”An investigation of the utilization of different data sources in manufacturing with application in injection moulding”, MADE - Manufacturing Academy of Denmark (2020-06-23), **Presentation**

(4) **G. Ø. Rønsch**, M. Kulahci & M. Dybdahl, ”An investigation of the utilization of different data sources in manufacturing with application in injection moulding”, InfinIT - innovationsnetværk for IT (2020-01-08), **Presentation**

(5) **G. Ø. Rønsch**, ”SPC in a multivariate system - a practical approach”, DSK conference on process monitoring - Få processen i kontrol med Statistisk Proceskontrol (SPC), (2019-05-13), **Presentation**

Abbreviations

ABS Acrylonitrile butadiene styrene

AI Artificial Intelligence

AGI Artificial General Intelligence

ANI Artificial Narrow Intelligence

ARIMA AutoRegressive Integrated Moving Average

ASI Artificial Super Intelligence

BIMS Brussels Injection Molding Sessions

CAE Computer-Aided Engineering

CPS Cyber-Physical-Systems

CNC Computer Numerical Control

CNN Convolutional Neural Network

DFT Discrete Fourier Transform

ERP Enterprise Resource Planning

FDA Food and Drug Administration

FFT Fast Fourier Transform

GTAI Germany Trade and Invest

IMR Injection Molding Rheometer

IoT Internet of Things

IIoT Industrial Internet of Things

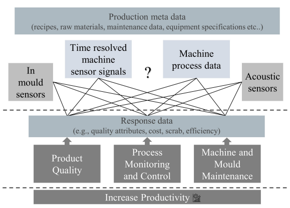
MADE Manufacturing Academy of Denmark

MES Manufacturing Execution Systems

MSPC	Multivariate Statistical Process Control
NCMS	Non-Conformity Management System
PAT	Process Analytical Technology
PCA	Principal component analysis
PLC	Programmable Logic Controller
PLS	Partial least squares
PLS-DA	Partial least squares - Discriminant Analysis
POC	Proof Of Conceptl
QbD	Quality by Design
RBO	Research and Business Objective
RNN	recurrent neural network
RQ	Research question
SPC	Statistical Process Control
SPE	Squared Prediction Error
STFT	Short-time Fourier transform
SVM	Support Vector Machine

Reader's guide

The purpose of this introductory guide is to introduce the reader to the structure of this PhD thesis. The thesis is based on a collection of articles consisting of five peer-reviewed journal articles (three published, one accepted and one in review) and one published peer-reviewed conference paper. Five of the publications cover topics related to data utilisation within injection moulding, and one is a more general paper (paper D) on data utilisation within manufacturing. All six publications are relevant to the overall topic of the thesis. Still, to improve the readability of the thesis, only three of the publication are included as chapters (paper A, B and C) in the thesis and the remaining three (paper D, E and F) are added as appendices. Sections including a paper will start with a two-page summary highlighting the objective, approach and results. The post-print version of the papers are included for the published and accepted papers, and for the paper in review, the pre-print version is included. An illustration of the proposed data utilisation framework will be included throughout the thesis (as the icon to the right). The purpose is to link data sources utilised in the different applications (published articles) to the data utilisation framework.



Chapter 1 is the introduction section where the project motivation and objectives are presented together with an introduction to the topics addressed in the thesis. The developed data utilisation framework is also introduced together with an overview of the results obtained in the project.

Chapter 2 gives an introduction to the fourth industrial revolution (Industry 4.0) and especially data utilisation within Industry 4.0. The data initiatives in Industry 4.0 is compared to previous concepts for data utilisation within manufacturing, where differences and similarities are addressed.

Chapter 3 gives an introduction to the general concepts of injection moulding and a more in-depth explanation of the interaction between material properties, moulding process and quality. This is essential and central for the main part of the work conducted in the PhD project and is not sufficiently covered in the included publications.

Chapter 4 gives an introduction to process monitoring and control. This chapter also presents additional (unpublished) work conducted at the industrial partner related to process monitoring within injection moulding. The chapter includes initial results and reflection on methods used followed by next step initiatives.

Chapter 5 includes paper A, which introduces the data utilisation model (described in chapter 1) on an application within injection moulding. The main objective of paper A has been to demonstrate that different data sources from the same underlying process contain different levels of information and that data utilisation complexity and cost should impact the decision on what solution to implement.

Chapter 6 includes paper B, which is based on the same data as in paper A, but where the objective has been to demonstrate an effective frame of combining observational process data and controlled experiments to capture the causal relationship between uncontrolled process disturbances and final product quality.

Chapter 7 serves as an introduction section for the use of acoustic emissions for condition-based monitoring of injection moulds introduced in paper C.

Chapter 8 includes paper C, which demonstrates the use of acoustic emission for condition-based maintenance within injection moulding. The paper includes an approach for using acoustic emission in cases where a general solution is desired for simple implementation on multiple similar production equipments.

Appendix A includes paper D, presenting reflections on challenges encountered when utilising data in a manufacturing context. The content is based on experiences and lessons learned through collaboration between our university group and various companies.

Appendix B includes paper E, presenting an exploration of existing mould maintenance service data to perform a mould wear-out prediction. This is turned into an interactive dashboard for exploring and monitoring mould state.

Appendix C includes paper F, explores the use of existing product quality data to detect degradation of injection moulds. A solution using MSPC-PCA is presented for supporting and scheduling mould maintenance and for supporting mould worn-out evaluation.

Appendix D is used to present the content of the presentation made at IDA (point 5, in the list of contributions), demonstrating the use of MSPC-PCA on quality data from injection moulding.

> *"Organization is what you do before you do it, so when you do it, it's not all messed up."*
>
> — Winnie The Pooh

Contents

CHAPTER 1

Introduction

This industrial PhD has been conducted in close collaboration with an industrial partner. The industrial partner produces injection moulded elements and is world-leading within the business area. More than 3,000 different element shapes are made, using more than 20 different materials and 30 different colours The injection moulding setup consists of approximately 10,000 injection moulds and more than 3,000 injection moulding machines located at four production sites (situated on three different continents). The company has been producing moulded products for more than fifty years, and injection moulding is a core and business-critical competency.

The thesis focuses on effectively utilising production and process data for optimising productivity. We propose a practical framework for structuring data utilisation in a manufacturing context. The framework has been developed through the PhD project and based on learnings achieved within the project and experiences from previous work conducted with other companies.

The Industrial Partner's focus on data utilisation has developed through the rise of the fourth industrial revolution (Industry 4.0). Initiative introduced in the Industry 4.0 context have guided the exploration of data utilisation at the industrial partner and are further introduced in section 2.

1.1 Motivation and aim

For the past six years, our university group (DTU Compute, Section for Statistics and Data Analysis) and the industrial partner have been part of Manufacturing Academy of Denmark (MADE). Through this collaboration, we have experienced that many companies struggle with utilising data for process optimisation and production improvements. Within this context, the idea for the PhD project was sparked and formulated. Many of the experiences created as part of the collaboration in MADE also serve as the foundation for paper D, where many of the common challenges related to utilising manufacturing data are addressed. We, therefore, found it relevant to propose a framework and approach for practical and effective data utilisation in a manufacturing context. Thus, formulating and testing such a framework has been a central part of the PhD project. The work has been based on an injection moulding setup and challenges apparent at the industrial partner but also reflected upon in a more general context.

In 2014, the industrial partner started the exploration of Industry 4.0, where connectivity, smart production, and data utilisation was defined to be central topics. A series of demonstrators indicated that production benefits could be achieved by utilising data from the production. It was also identified that the injection moulding machine connectivity had to be improved to enable the right level of communication to and from the machines. This connectivity has been established, and approximately 25 standard process parameters (more than 200 are available on the machines) are collected for each moulding cycle. The data are structured and stored in a cloud platform to be combined with other relevant data sources (e.g. production order and quality measures). Based on this improved data collection, the industrial partner gave the overall frame of the project content *"increase productivity within injection moulding by use of manufacturing data"*. The frame has been guiding the formulation of the overall research question:

Research question 1 (RQ1)

How to best utilise data from injection moulding machines to increase the productivity of producing injection moulded elements?

This is a vast question with many options and the next natural questions was, what type of productivity? What data from the machines should be collected? And how should the data be used? As we reflect upon in paper D, we often see a tendency towards a very broad perspective when data have to be utilised in a production context, making it hard to approach the task in the right way and therefore also hard to achieve the intended outcome. Research question 1 has therefore been supported by a research and business objective:

Research and business objective 1 (RBO1)

Propose a framework for the approach of utilising manufacturing data for process optimisation.

Therefore, the first objective of the PhD project has been to set up a framework that breaks down the overall business objectives into tangible business contributions that can be linked to the production shop floor, data collection, and data analytic tasks.

During the PhD, a second research question (RQ2) emerged as the exploration of different data sources progressed, and the industrial partner presented a challenge with quality issues caused by dual sourcing of raw material. Since it is challenging to investigate and solve such a challenge in a production setup (limited variation in process parameters and uncontrollable disturbances), the question is how such investigation can be conducted in an off-line setting still reflecting the disturbance factors impacting the production setup. The second research question is formulated as:

Research question 2 (RQ2)

In an off-line setting, how to develop real-time data analytic models that will help mitigate the impact of the variation in a nuisance factor in a production setup?

At the industrial partner the question is related to mitigation of raw material variations' impact on product quality (both for dual sourcing and introduction of new sustainable materials). This is centred around injection moulding and challenges seen at the industrial partner. However, our clear expectation is that the problem and proposed solution are generalisable to other companies working with injection moulding and other industries in general.

Based on approaches described in literature [1] [2] [3], it became clear that different types of data are available and relevant when looking at injection moulding. Therefore, a central part of the PhD work became to investigate what type of injection moulding data was best suitable for solving the RQ2. This also added a layer of complexity in the data utilisation framework since different data sources have to be evaluated against the specific business application and the sum of business objectives, which have been formulated as additional research and business objective (RBO2).

Research and business objective 2 (RBO2)

How to best test and select data sources to overcome challenges with process and quality variation caused by material variation introduced with dual sourcing and the introduction of new sustainable materials?

Working with the data utilisation framework at the industrial partner (through interviews with subject matter experts) made it clear that mould maintenance had a significant impact on the overall productivity and, therefore, also had to be addressed. Paper E focused on utilising mould maintenance data to develop an early warning system to predict when moulds are approaching a worn-out state. Since this approach only gives worn-out and updated predictions when the mould is maintained, it does not capture the need for maintenance tasks like lubrication, cleaning, or sudden breakages of mould parts (main factors impacting productivity). To overcome this, a maintenance approach has to be based on real-time data from sensors that reflect the condition of the mould. A third research question was therefore formulated:

Research question 3 (RQ3)

How to do real-time condition-based maintenance on injection moulds to reduce the cost of preventive maintenance and reduce the risk of mould malfunction?

In a review of maintenance applications within Industry 4.0 by Çınar et al. [4], 19 of 67 reviewed maintenance applications were using acoustic or vibration data. The acoustic data were used for applications within Industrial pumps, CNC metal milling, wind turbine bearings and rotating machinery. These applications have similar characteristics as injection moulding, making acoustic emissions relevant for monitoring injection moulds. This is supported by results obtained by Kek et al. (Crack detection in injection moulds [5]) and Moreira et al. (maintenance monitoring of injection moulds using a flexible pressure sensor a three-axis accelerometer. [6]) Based on this, we formulated a third research and business objective as:

Research and business objective 3 (RBO3)

Investigate the use of acoustic emissions for condition-based mould maintenance.

The three defined research questions and linked objectives are centred around the use of data for process and production improvements. The questions and objectives will be explored within injection moulding and generalised for other industries.

1.2 Data utilisation framework

The results and contribution achieved through the PhD project have been running in two parallel tracks. One has been focusing on formulating and testing a framework for practical data utilisation within the context of Industry 4.0. The other track has been exploring different data utilisation cases, supporting the defined business objectives. The business objects will be specified in the following and the main emphasis will be on describing the data utilisation framework since it is central for the rest of the work conducted and this description is not included in other sections of the thesis. The results of the investigated data utilisation cases (later referred to as "applications") can be seen in the sections including paper A-F.

Various data utilisation frameworks have been proposed covering many of the aspects of data utilisation in a manufacturing context. Mert et al. [7] address the growing amount of collected data and the challenges of utilising the data due to complicity in the data utilisation tools. Mert et al. also present a conceptual framework for utilising analytics applications in an industrial Big Data setup to overcome these challenges. Based on a survey amongst 92 manufacturing companies, Frank et al. [8] conclude that implementation of basic technologies as Internet of Things (IoT), cloud services, big data and analytics is still low. This is concluded without reflecting on why this is the case. Gröger [9] present the development and implementation of the Bosch Industry 4.0 platform, with a focus on data structure and technical solutions. It is concluded that establishing a data utilisation platform involves various challenges beyond the tools and technologies. It is highlighted that the industrialising analytical solution development on top of the platform is a challenge and the empowering of business domain specialists to do advanced analytics. Based on several surveys, Gajdzik et al. [10] highlight that company decision-makers show an increased willingness to invest in a digital transformation but lack knowledge of their current status in Industry 4.0 and strategic guidance for its implementation. Gajdzik et al. also present an overall framework for improving the implementation of Industry 4.0 initiatives, where analysis of the company capabilities, strategic focus, initial pilot projects, and impact analysis are highlighted as essential steps.

The main focus in the reviewed work is on the technical framework for data collection and analytics. It is interesting that Gröger [9] highlights that the development of analytic applications is the dominant challenge. Based on this, we see that a more explicit and practical approach is needed to guide the implementation of data utilisation, where it is essential to focus on capabilities, strategic focus and pilot project as highlighted by Gajdzik et al. [10]. We created and presented the first version of our data utilisation framework in paper A, where it was centred around evaluating different data sources from injection moulding. The framework has expanded to cover more aspects of the data utilisation journey and has been central throughout the PhD project. The framework will be described in a more general context and later reflected upon for use within other industries.

1.2.1 Data utilisation journey - defining the framework

When starting a data exploration journey in manufacturing, it is essential to anchor it in a solid and clear business objective. Too often, the journey is started with the intention to discover what the collected data can bring (almost like an artist beginning with a blank canvas, just letting the paintbrush determine the destination). Data exploration is essential and critical, but it has to be data exploration with clear intent guided by domain insight. Without the business objective being well defined initially, the destination will be unclear, making it challenging to conduct an effective data exploration and utilisation.

> *"Would you tell me, please, which way I ought to go from here?" Alice asked. "That depends a good deal on where you want to get to," said the Cat. "I don't much care where—" said Alice. "Then it doesn't matter which way you go," said the Cat. "—so long as I get somewhere," Alice added as an explanation. "Oh, you're sure to do that," said the Cat, "if you only walk long enough."*
>
> — Lewis Carroll, *Alice in Wonderland*

Like for Alice, data utilisation with no clear destination will eventually result in some outcome, but it might not support any critical business needs or objectives, and therefore not add the value needed to justify the effort spent and cost of implementing the solution. Therefore, the effort will waste time and resources and have put unnecessary stress on the organisation. Consequently, we see the need for a more structured framework or guiding tool to support an effective data utilisation, where the foundation of the data utilisation framework is the "Business object" (Step 1 in Figure 1.1). In our case, the overall business objective is formulated as "increase productivity".

Other examples could be “reduce energy consumption”, “increase throughput”, “reduce lead time”, ”improve quality” or “become sustainable - zero-impact production”.

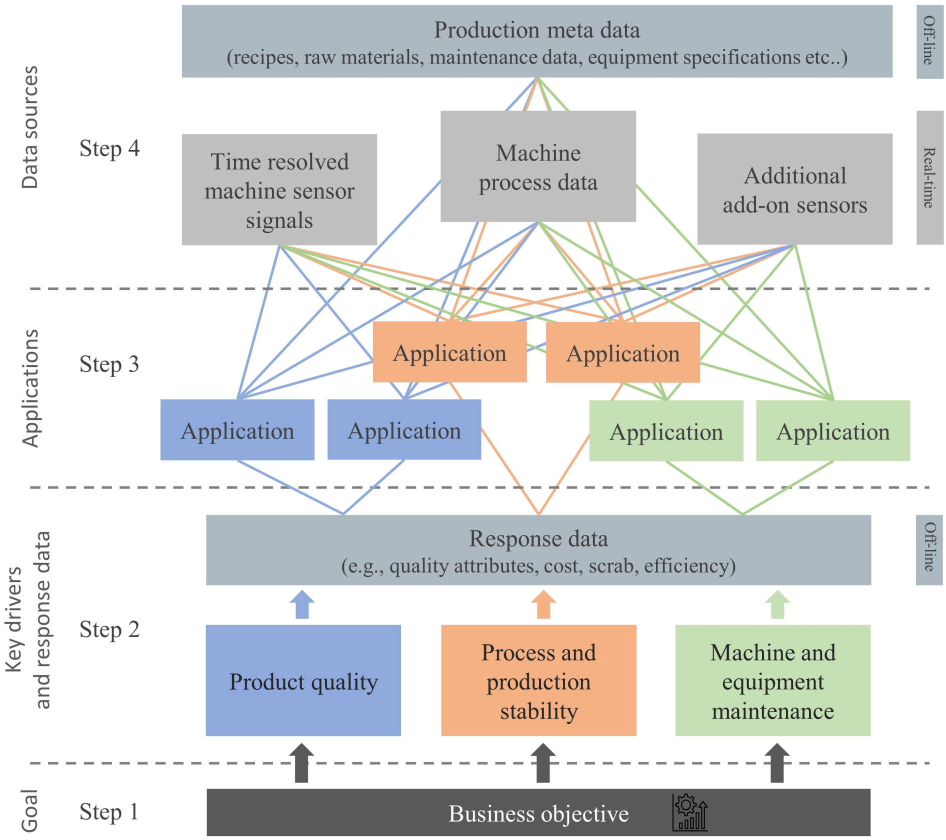


Figure 1.1. Four-step data utilisation framework (business objective, key drivers, applications and data sources). The connecting lines between the applications and different data sources indicate potential data connections. The colouring (blue, orange and green) highlights the connection between the individual key drivers, associated applications and linking data connections. Production metadata and response data are indicated as off-line data and three types of process data (time-resolved machine sensor data, machine process data and additional add-on sensor data) as real-time data.

When the business objective is clear, it will guide identifying key drivers (Step 2 in Figure 1.1) that must be addressed to achieve the desired improvement. The key drives are the main contributors to the loss/gap that must be mitigated. Based on a series of interviews of moulding specialist at the industrial partner, three key drivers of production loss was identified (described in paper A). These drivers will differ from company to company and from business objective to business objective within the same company. When identifying a key driver, it is also essential to determine what losses exist for that specific driver and how improvements can be measured/quantified. The identified measures are often used as response variables when the data modelling is started.

With the key drivers and response data in place, the next stage is to identify relevant applications within each driver (Step 3 in Figure 1.1). For each application, it has to be determined what is driving the behaviour and what data to potentially describe this behaviour. This work requires the involvement of subject matter experts and is essential for identifying all data sources that are relevant to include in step 4 and to ensure commitment from the business experts. It is essential to notice that this is often an iterative process. Exploring the first data identified might reveal that more or other data sources need to be added. Therefore, it is essential to start this data exploration as early as possible so that all required data can be identified and added to the data foundation. In the framework described here, the focus is on data utilisation. However, it is central to be aware of other approaches to address the identified gaps in the key drivers. It is not given by default that the only or best solution is to implement a data utilisation approach; it could be that standard LEAN tools or a change in the production process/setup is more appropriate (*"If the only tool you have is a hammer, you tend to see every problem as a nail." - Abraham Maslow*).

In many cases (ours included), it would be a too extensive task to conduct the data exploration for all the identified applications at once and thereby get the total overview of the needed data. Therefore, it is relevant to start with applications that add the most business value or where most of the data foundation is already in place.

Readily available process data from the process/equipment under investigation is not always enough to achieve the desired business objective. As presented in papers A and B, injection moulding machines have a deeper layer of process data (timer resolved profiles) that contain more information essential for achieving the business objective. Utilising these timer profiles comes with an additional cost since it requires both machine hardware and software update. Therefore, as different data sources are identified as relevant for one or more applications, they have to be evaluated against each other, both in relation to utilisation cost and value creation (cost-benefit analysis). In some cases, the business benefits would not be able to cover the cost of collecting and utilising specific data sources, which primarily can be the case if additional sensors have to be added to the equipment (e.g., as we propose with acoustic sensors in paper D) to obtain the desired business objective.

Starting the practical data utilisation at the industrial partner made it apparent that the constructed data foundation lacked many supporting data sources (metadata and response variables) and variables in the different data sources to link them together. This has demonstrated the benefits of starting small (with one selected application at a time), and scale as useful applications have been built. It has also become evident that data utilisation in a manufacturing setup is not a turn-key solution that can just be applied. It takes a joint effort from subject matter experts, data engineers, data scientists and organisational backup from management.

Our experience using the proposed framework is that it creates organisational clarity of the task at hand and supports the prioritisation of the different initiatives. We see the structure and elements of the framework to be generalisable and, therefore, useful for other companies in their data utilisation journey. Some key points to highlight:

Step 1 - Business objective: Defining a clear business objective is essential and will ensure purposes throughout the project. All activities in the data utilisation have to link back and support the defined business objective. Business objects could be formulated as: "increase productivity", "reduce energy consumption", "increase throughput", "reduce lead time" or "become sustainable zero-impact production".

Step 2 - Key drivers: Identification of the main contributors (key drivers) for achieving the defined business objective are essential to break down the business object into tangible focus areas. Business and subject matter experts have to be involved in this to ensure the proper focus and identification of measurable response data (quantification of business impact) and ensure commitment in the organisation.

Step 3 - Applications: An overview of potential applications (concrete solutions to mitigate the identified gap) has to be made before the exploration/implementation is started. Based on the overview, a few applications are selected to be the first to be explored. Applications with the biggest business impact or smallest effort are selected first. As applications are evaluated to have a positive business impact, they can be implemented, and new applications are explored.

Step 4 - Data sources: Identification of the needed data sources is an iterative process, making the data foundation change over time. The identified application will most likely use some of the same data sources. Therefore, the cost of collecting and utilising the different data sources has to be valued against the sum of estimated benefits. It might be possible to use different data sources and achieve close to the same outcome for an application. In these situations, a cost/benefit evaluation should be used to determine which of the data sources to utilise.

1.2.2 Data utilisation and maturity level

Many illustrations of data maturity and utilisation levels have been presented in the context of Industry 4.0 (see description in section 2.2). In paper D, we presented our interpretation of maturity levels, described as a roadmap. This approach has been developed further and will be presented here and linked to anticipated applications in a manufacturing context.

Our data utilisation framework described in the previous section is centred around identifying and structuring business objectives, key drivers, applications and needed data foundation in the data utilisation journey. This is done while evaluating the business value of utilising different data sources for different applications. Our data utilisation roadmap illustrated in Figure 1.2 is a tool to guide the test and implementation of the identified application based on the data utilisation maturity in the organisation.

The roadmap illustration consists of two halves. The left side (the road) illustrates the data maturity levels and level of data utilisation, whereas the right side illustrates the data foundation and data utilisation complexity. Two paradigm shifts have been indicated in the data utilisation roadmap. The first shift indicates the transformation from aggregated off-line data (e.g., per hour/day/week) to utilising real-time data directly from the production equipment, which require implementation of real-time data ingestion and cleaning functionalities. The needed capabilities for passing the first paradigm shift could be data engineers, Six Sigma capabilities, and fundamental statistic insight. The second paradigm shift indicates the transformation going from simple data usages (visualisation or Statistical Process Control, SPC) to a model-based approach using simple to complex machine learning modelling. Passing the second paradigm shift introduces new complexity that must be handled. This is regarding the demand for an enriched data foundation and an utilisation framework consisting of data model development, monitoring, and maintenance. This requires capabilities within data engineering and data science.

It would, in some situations, be tempting to jump from the lower maturity level, using business intelligence on productivity data, to the implementation of a model-based approach for complete monitoring and optimisation of a process/production. However, this is a giant leap, and it is doubtful that it would work; we, therefore, propose doing this in steps. One of the benefits of taking the journey in steps, following the maturity levels defined in the model, is that as applications are developed and tested, missing data sources and missing data connections will be identified. Attempting to create a complete data foundation from the start will be a complex and overwhelming task, where there is a high risk that wrong data sources and an excessive amount of unneeded data will be included, whereas critical data sources and data connections might be left out. Another benefit is the incremental discovery of needed competencies in the organisation, making it possible to build the competencies (training existing staff or hiring new people) as they become needed.

Based on our experience, we propose a development and implementation strategy based on conducting a series of “Proof Of Concept” (POC) applications (also recommended by Gajdzik et al. [10]). When conducting a POC for a given application, it becomes clear what type of data is relevant for achieving the defined business objective. Some data sources are most meaningful for a specific application, whereas others can be used for multiple applications. As more POC’s are being conducted, an interaction map will appear between the different data sources and applications (see the upper part of Figure 1.1 or Figure 9.1 in the discussion section). At this point, it also becomes relevant to evaluate which of the data sources are essential and which can be substituted by other data sources. At the same time, it must be evaluated what the expected cost of utilising the different data sources are and what the expected value creation will be. This approach supports the creation of the needed data foun-

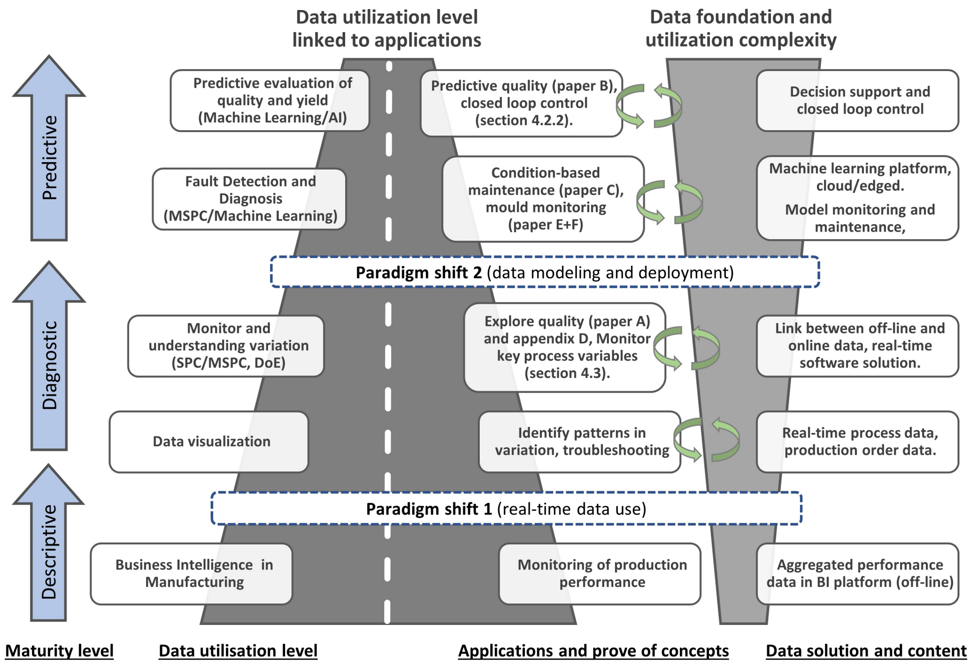


Figure 1.2. Data utilisation roadmap with maturity level to the left, applications in the middle and data foundation and utilisation complicity to the right. The width of the figure to the right (Data foundation and utilisation complexity) illustrates the amount of data needed and the complexity of data utilisation at a given stage of data utilisation.

dation (e.g., what data to include? sampling frequency? how to link different data sources?) and the identification of required capabilities and solutions to succeed with large scale implementation. The iterative process of updating the data foundation based on learnings from the development and test of the applications is illustrated in Figure 1.2 by circular arrows between the applications and the data foundation.

As part of the initial data utilisation work at the industrial partner, a simple data visualisation tool was created where time series of the most critical process variables could be displayed. Despite the simplicity in the solution, just making the dynamics in the crucial process variables visible for subject matter experts made them reflect on the variation and identify potential root causes and solutions to reduce the observed variation. Although the solution was simple in its appearance, creating the required data foundation and software was a demanding task. The way of thinking and the needed data solution changes radically when going from collecting and presenting aggregated data by the hour to collecting and presenting data in real-time, making it a solution and organisational paradigm shift (shift 1 in Figure 1.2). After using the data visualisation solution in a test setup for some weeks, subject matter experts started to reflect on the possibility to add action limits on the time series plots to get an indication of when to react to the observed variation. The fact that this proposed improvement came from the users is essential and a good indication of a growing maturity level in the organisation. Adding action limits (control limits) on the time series is what is done within Statistical Process Control (SPC). There exist multiple publications on SPC in general and for quality and process monitoring in injection moulding, where the most interesting one uses Multivariate SPC (MSPC) on machine process data (e.g. Kazmer et al. [11]). MSPC is interesting for injection moulding since the process and quality variables are highly correlated. We have demonstrated the use of MSPC on quality measures (paper F) to identify remaining useful life for plastic injection moulds[1].

Our data utilisation roadmap has been a key in communication with company stakeholders. This has been towards IT stakeholders (product owners, data engineers and data scientists) when communicating with business owners, application owners, and subject matter experts. It has been a visual tool to communicate the complexity of starting a data utilisation program within a manufacturing environment. We have received good feedback on creating proof of concept for applications within different maturity levels and focusing on the overall data foundation and evaluation of different data sources against each other. In order to conduct POC's across the different maturity levels without having the support for ready data infrastructure and foundation, the main part of the data handling has been done manually.

We have experienced lower resistance to predictive support applications despite these being black box solutions. However, we see a higher resistance when it comes to predictive solutions that should take over key production areas. This means that the illustrated roadmap has to be considered dynamic in the sense that when the overall maturity level for business-critical parts of the production might be on a data visualisation level, supporting predictive closed-loop solutions could be implemented in the same production area for non-business critical applications. Linked to the industrial partner's work, condition-based maintenance using acoustic is likely to be implemented soon. In contrast, closed-loop adjustment for material variations will take longer since it changes the moulding approach used over many years (making the guiding approach described in paper B a more acceptable method).

1.2.3 Data utilisation summary and reflections

Companies are at different stages in utilising data from manufacturing. Some companies have started the transition by implementing MES systems, digital shop-floor management [12], paperless production [13], making data utilisation for production optimisation a natural next step. Other companies are still executing production orders manually, collecting production performance on paper and do not collect any equipment process data. This means that our proposed data utilisation framework

[1] In section 4.3 I present some unpublished work, where MSPC is tested for process monitoring using readily available process data

would look very different depending on the company trying to use it. The starting point could be very different, where some companies already have a data foundation that can be utilised and expanded, others have to start from scratch. Despite this very different starting point, the structure with defining a clear business objective, identification of key drivers and POC-based exploration of identified applications are still the same. To summarise the proposed data utilisation framework, the main takeaways are collected in the following:

- **Data utilisation framework:** The first step in a data utilisation journey should be the creation of a data utilisation framework as described in section 1.2.1. This is crucial to ensure a solid and transparent business objective from the start and understand how the introduced application will eventually create business value. Our experience is that the data utilisation journey often is initiated by data enthusiastic subject matter experts in an R&D or innovation department who have the courage to challenge the status quo. This makes the journey spark as a bottom-up rather than a top-down approach (Gaynor [14]), where the business objectives are not always in focus (driven by curiosity rather than value creation). The bottom-up approach is often needed to initiate the journey and spark the interest in management by showing what is possible by utilising data. This could lead to the founding and formulation of more formalised work around data utilisation. At this point, it becomes crucial to follow the steps in the data utilisation framework to ensure the involvement of business subject matter experts and management.
- **Data utilisation level and capabilities:** We recommend doing an initial assessment of the data maturity level in the organisation before stating any exploration of the defined applications (also highlighted as important by Gajdzik et al. [10]). This ensures that the needed competencies are available and that the required data can be provided. It is crucial to have capabilities within data engineering, data science, and dedicated subject matter experts who can identify business opportunities and how to realise these. Data engineering and science capabilities can be external resources in the initial phase, but it should be considered how these resources should be integrated as exploration and implementation progress.
- **Applications and proof of concept:** It is our recommendation to start with simple non-model based exploration (e.g. data visualisation), which can be achieved by manual data extraction (does not have to be running in real-time). The next step could be individual model-based applications that can be explored, implemented and scaled separately. Each application has to be evaluated regarding the expected return of investment. The identified data sources might be utilised across different applications, which must be included in the overall evaluation. It is also central to reflect on the complexity of the identified data and if simpler or lower-cost data could be used instead, thereby reducing the implementation cost. This could impact the quality of

the individual application, but the quality might still be sufficient to achieve an acceptable business outcome with a better cost/benefit trade-off. It is important to notice that it is rare that a single application can support the creation of the data foundation (machine connection, data collection, data storage and tools for data visualisation and modelling). Therefore, it can be necessary to conduct several POC's before a combined benefit is sufficient to cover the cost of creating the data foundation needed to implement the identified applications.

- **Data foundation:** The creation of the data foundation needed to support the implementation of the first identified and profitable application will eventually enable the exploration of other applications. As more applications are identified, evaluated and potentially implemented, it is good practice to review the generated data foundation. It should be investigated if an individual application can be improved utilising data intended for other applications, if redundant data can be eliminated, or data sources combined to achieve new applications. Solutions for data collection, storage and distribution are necessary to consider but are out of scope for the current work, and therefore not addressed further.

  With the implementation of a structured and continuous data collection, a question that eventually will arise is; What data needs to be stored for a longer time? And for how long? At this point, it is necessary to reflect on the purpose of storing data; is it for documentation? If so, for how long? is it to enable exploration of new applications? Alternatively, is it for model maintenance and optimisation? Again it comes down to a cost/benefit consideration, need for accessibility, and, e.g. regulatory requirements (especially relevant for food and pharmaceutical companies). Hsu et al. [15] present an interesting reflection on hot-cold data storage[2] and how this can be utilised to reduce storage cost. Before sending the collected data for storage, it could also be considered if all data have to be stored or if the data could be reduced before storage. One option could be to aggregate the data (e.g., per hour, per batch or production order) and thereby reduce the number of saved observations, or only save a subset of the collected variables. Using these approaches, there is a risk that valuable information is lost. Tonglin and Baijian [16], present principal component analysis (PCA) as a powerful tool for dimension reduction (often used for data prepossessing before applying regression and classification algorithms) before data storage. PCA is often a better option for data reduction than manually sub-selecting variables since the retrieved principal components are a linear combination of all the collected variables. No trivial answer can be given for how best to optimise the amount of information stored while reducing the storage cost, while this has to addressed case by case.

[2]Hot data storage is used for data that has to be fast to access and that are frequently used. This flexibility comes with an additional storage cost compared to cold data storage, where data retrieval is slow and inflexible

- **Data and modelling solution:** Depending on the available capabilities within the company, it should be considered if the strategy should be to go for commercial software solutions or an open-source framework and custom-made software solutions. Commercial software solutions for data visualisation could be Power BI [17] or Tableau [18] and for SPC/MSPC solutions like SAS JMP [19] or NWA Quality Analyst [20]. Entering the upper level of data utilisation applications (model-based concepts), it becomes harder to identify suitable commercial solutions since the application is more dedicated to the specific production setup. Non-code auto-ML solutions exist (like DataRobot [21]) and could be an option to use. However, these solutions still require data engineering and data science capabilities to prepare the needed data, formulate the data analytic task and interpret the results. Going for an open-source solution makes the development and implementation more extensive, but at the same time, makes the shift towards model-based concepts less challenging (since the same platform and tools can be used). It is also possible to automate part of the model development with open-source by using auto ML approaches. The research group behind AutoML.org [22] (research groups at the University of Freiburg and the Leibniz University of Hannover) has collected and developed a series of tools for machine learning automation.

The summary and reflection presented above are very general and have to be adapted to fit the individual use-case within different companies. One of the fundamental steps is creating the data foundation for the data utilisation journey. It is always a question of when to start creating a manufacturing data foundation. In the above description, the creation is done as profitable applications are identified and implemented. In some cases, initial data collection and data foundation is needed before exploring and evaluating a given application is possible. This has been the case for testing model-based process monitoring at the industrial partner. With a very consistent manufacturing setup, data had to be collected over a long period (and from multiple machines) to create the needed data foundation for creating and validating the proposed solution. The estimated value creation of the initial proof of concepts at the industrial partner (data visualisation, a process confirmation solution and supported changeover, all conducted before the PhD, therefore described further here) was insufficient to cover the cost of the needed upgrade of connectivity and creation of a data collection platform. Despite this, the top management decided to approve the budget for getting the needed infrastructure and platform in place. This was a clear signal that data utilisation was a priority turning the approach from bottom-up to top-down.

1.3 Contributions

The application tested and developed in the PhD is reflected in the six submitted publications (paper A-F). The applications will be introduced in the following, focusing on motivation, achieved results and contributions (academic and industrial). A more in-depth summary of the six papers can be found as the first section in the chapters and appendices that contain a paper.

Paper A - An investigation of the utilisation of different data sources in manufacturing with application in Injection Moulding

The work in paper A is linked to RQ1. It presents a framework for evaluating the use of different data sources for achieving real-time prediction of element dimensions from injection moulding. The framework has been demonstrated to effectively evaluate the cost/benefit of utilising different data sources for the same objective. Also, the paper presents an evaluation of three levels of data from injection moulding (readily available process data, time-resolved machine pressure profiles and cavity pressure profiles) and concludes that time-resolved pressure profiles are the best options when modelling element dimensions.

Achieved results

- Understanding the information contained in different data sources from injection moulding (readily available process data, time-resolved machine pressure profiles and cavity pressure profiles).
- Proposed solution for utilising machine pressure profiles to detect raw material variations (originating from dual sourcing of raw material).

Contribution

- Framework for comparing information content in different data sources and benchmarking the cost and value creation.
- Highlighting the possibility to use underlaying machine/equipment data for process optimisation by extracting the essential process dynamics.
- Proven that within injection moulding, variation in raw material can be captured by the change in injection and dosing pressure profiles.

Paper B - Real-time adjustment of injection moulding process settings by utilising Design of Experiment, time-series profiles and PLS-DA

Paper B is linked to RQ2, focusing on the effective use of designed experimentation to observe the causal relation between disturbances factors, process settings and product quality. Variation caused by dual sourcing of raw material is the investigated disturbance factor.

Achieved results

- Demonstrate that changes in injection and dosing pressure profiles can be used to detect a change in the origin of raw material.
- It has been demonstrated that knowing the origin of the material (classification based on the pressure profiles), holding pressure and mould temperature can be used to reduce variation on element quality caused by material variations.

Contribution

- Approach for systematical exploration (off-line) of disturbances factors impact on product quality and linking these findings back to a concrete solution that could be implemented in an industrial context (real-time).
- Demonstrating that under-laying equipment signals contain relevant information process information that can be utilised for improving product quality.

Paper C - Utilisation of acoustic signals from injection moulding for predictive maintenance

Paper C is linked to RQ3, exploring the use of acoustic emissions from injection moulding for condition-based maintenance of injection moulds. The overall focus is to propose a solution that can be developed on a subgroup of moulds and then scaled to new unseen moulds. This is desired since it is impossible to collect faulty data on new moulds and reduce the modelling workload by having global models instead of models for all mould/machine combinations.

Achieved results

- Using generative Gaussian modelling, it has been shown that acoustic emission from injection moulds can be used for detecting the need for lubrication and loose/defect latch lock.
- Implementing a simple model adaptation make the proposed solution generalisable to new unseen moulds.

Contribution

- Exploration of acoustic emissions from five different injection moulds.
- A practical, effective and scalable modelling concept for achieving condition-based maintenance on injection moulds using acoustic emissions.

Paper D - Experiences with big data: Accounts from a data scientist's perspective

Paper D is linked to RQ1 through reflections on challenges encountered when utilising data in a manufacturing context. The work is a collection of experiences and lessons learned from collaboration with numerous companies exploring and developing Big Data applications for production analytics.

Achieved results

- Discussion of pitfalls and challenges in approaching data utilisation in a manufacturing context.

Contribution

- Collection of crucial insights and learnings gained through practical experience with data utilisation within different industries.
- Reflection on how to effectively engage with the utilisation of manufacturing data for process and production optimisation.

Paper E - Mould wear-out prediction in the plastic injection moulding industry: a case study

Paper E is related to RQ1 (by mitigating productivity loss caused by mould malfunction) and indirectly related to RQ3 by identifying the need for additional sensor data to obtain real-time condition-based maintenance. In paper E, we utilise existing mould maintenance service data to perform a mould wear-out prediction to enable re-order of new moulds in due time. The work was motivated by a need to detect when an injection mould needed to be exchanged with a new mould. Failing to detect this might cause production stop (not having a functional mould) or production of defective products.

Achieved results

- Increased understanding of the information contained in the used maintenance data and how best to utilise this information.
- Interactive dashboard for exploration and monitoring of mould state.

Contribution

- Demonstration of survival curves used for prediction mould wear-out.
- Demonstrate effective use of mixed categorical and numeric data using an unsupervised random forest approach.

Paper F - Data-Driven Identification of Remaining Useful Life for Plastic Injection Moulds

As stated in the outline of future work in paper E, the product metrology data is expected to be essential for assessing mould wear and tear. Paper F addresses this by exploring metrology data to detect long-term degradation of injection moulds.

Achieved results

- Analysis of mould degradation pattern reflected in element metrology measures.
- Scaling approach to enable monitoring of samples consisting of elements from multiple moulds, cavities and different metrology measures.
- Unsupervised MSPC approach for monitoring the change in product quality for products with multiple quality attributes.

Contribution

- Comparison of unsupervised and supervised approach for detecting out-of-control behaviour for plastic injection moulds.
- The proposed solution can be utilised for supporting and scheduling mould maintenance and support mould worn-out evaluation.

CHAPTER 2

Industry 4.0 and Data utilisation

The fourth industrial revolution (Industry 4.0) originates from Germany. The basic concept was first presented at the Hannover Fair in 2011 and later published in the ”Recommendations for implementing the strategic initiative INDUSTRIE 4.0” [23] report in 2013. The initiative was formulated by a group of German industry, academic and governmental representatives with the scope to increase the German competitiveness in a more competitive industry and secure the future of the German manufacturing industry. Since 2011, the concepts of Industry 4.0 have been a central topic at many universities and companies, where more and more content has been proposed and investigated.

2.1 Industrial revolutions

Historically, there have been defined four industrial revolutions, where the fourth was defined and initiated with the German initiative, see illustration in Figure 2.1.

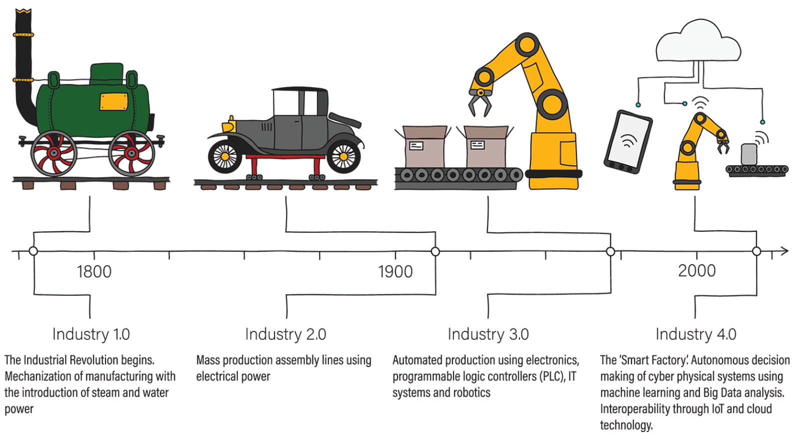


Figure 2.1. Illustration of the four industrial revolutions. Source: www.pngwing.com [24]

The Industrial Revolution started in the 18th century. The first phase is defined as the first industrial revolution and is characterised by the introduction of manufacturing machinery powered by steam and water. This radically changed the way of production and made it favourable to gather production at dedicated manufacturing sites. In the same period, William Cooke and Samuel Morse invented the telegraph [25], which revolutionised long-distance communication, and later developed into the communication we know and benefit from today. The second revolution started when Henry Ford [26], in 1913, introduced the first assembly line, and electricity became the primary power source for all manufacturing machinery. The introduction of electricity and assembly lines was the foundation for mass production. The third revolution is defined as the introduction of automation and digitisation using electronics and computers. One of the crucial inventions impacting the third revolution has been the programmable logic controller (PLC) invented by Richard Morley in 1968 [27]. The PLC is still central for manufacturing communication and an integrated part of most manufacturing machinery. The invention of the industrial robot by George Devol [28] in 1954 has also had a significant impact on the third revolution and how manufacturing and assembly are done today.

2.1.1 Industry 4.0 as a framework

Industry 4.0 can be seen as an umbrella enclosing digital solutions for the manufacturing industry. Industry 4.0 encapsulates more than "just" digital manufacturing solutions since it covers the complete supply chain. Some of the keywords defining Industry 4.0 are included in Figure 2.2.

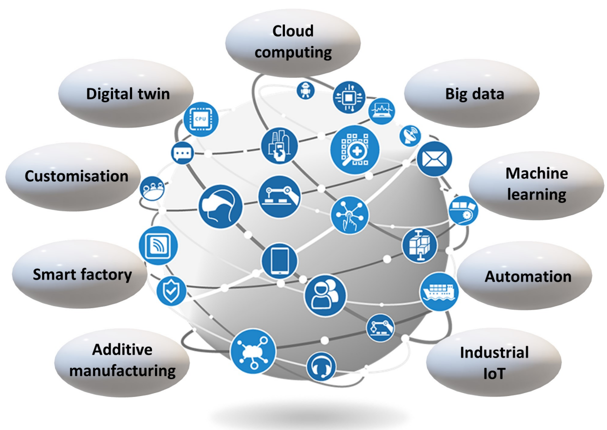


Figure 2.2. Illustration of keywords that defines the fourth industrial revolution.
Source: elements from www.pngwing.com [29] are used in the illustration.

Industry 4.0 addresses the technical aspects of manufacturing and how to utilise digitisation to create new products, business concepts, and models. The level of connectivity is increasing within the complete supply chain. This is interactions from the end-users back to product development. Data are collected from consumer products in many sophisticated ways describing the customer's interaction with the products, which will enable the development of new products and functionalities.

Looking more isolated on manufacturing of products, I see the Industrial Internet of Things (IIoT) as one of the keystones for the changes in the fourth industrial revolution. IIoT is the foundation for the inter-connectivity between production equipment, industrial sensors, data collection and cloud computing. I think the formulation from Germany Trade and Invest (GTAI) presented by Deloitte [30] is an excellent way to interpret this inter-connectivity:

> *"A paradigm shift... made possible by technological advances which constitute a reversal of conventional production process logic. This means that industrial production machinery no longer simply "processes" the product, but that the product communicates with the machinery; to tell it exactly what to do."*
>
> — Deloitte*, Industry 4.0 and manufacturing ecosystems.*

The interconnection between product and manufacturing process is a central part of digital twins or Cyber-Physical-Systems (CPS). Many definitions of Digital twin exist, where Stark and Damerau (2019, [31]) define it as: *"...a digital representation of a unique active product or unique product-service system that comprises its selected characteristics, properties, conditions, and behaviours by means of models, information, and data within a single or even across multiple life cycle phases."* The Cyber-Physical-Systems have been described by Baheti (2011 [32]), as: *"transformative technologies for managing interconnected systems between its physical assets and computational capabilities"*. As described by Lee et al. (2015, [33]) the CPS consists of two parts *"(1) the advanced connectivity that ensures real-time data acquisition from the physical world and information feedback from the cyberspace; and (2) intelligent data management, analytics and computational capability that constructs the cyberspace."*. To make this abstract definition into something more tangible and implementable, Lee et al. (2015, [33]) have defined a 5C model framework (Connection level, Conversion level, Cyber level, Cognition level and Configuration level). The 5C model visualised in Figure 2.3, is an adapted version of the visualisation given by Lee et al. (Fig. 1. in [33]). I think the 5C model provides a good representation of data utilisation within Industry 4.0. The model highlights the major difference between the data utilisation level achieved so far (Data-to-information level to Cognition level) and the desired state (Configuration level). At the configuration level, there is feedback from cyberspace to physical space that acts as supervisory control where equipment and process self-configure and self-adapt based on current production layout, quality performance, equipment maintenance state etc. This can be interpreted as a layer of closed-loop regulation that is derived from sensor-data collected and analytic created in the cognition level, and the abstraction level could be from an individual equipment to interaction between multiple machines.

Summarising and simplifying the above, one could say that both digital twin and Cyber-Physical-Systems aim to replicate the physical manufacturing setup in a digital

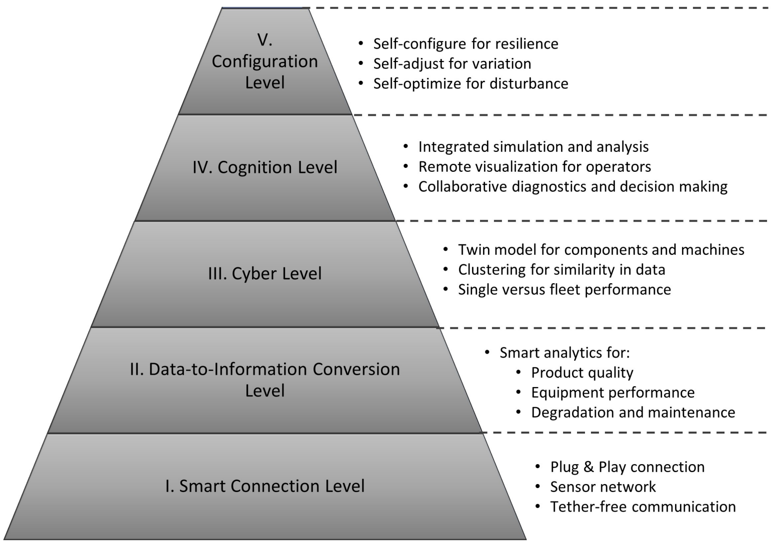


Figure 2.3. 5C model, illustrating five levels of data utilisation. Source: Re-making of visualisation given by Lee et al. (Fig. 1. in [33])

version. Ideally, with real-time interaction and data exchange between the digital version and physical manufacturing setup. The digital version can then be used for simulations, real-time monitoring and optimisation of the manufacturing setup (production layout, equipment settings, product quality, predictive maintenance etc.). Stark and Damerau (2019, [31]) stated that digital twins are not yet fully established, neither in engineering and applied information technology nor in production and maintenance business. It is also noted that digital twins are not a prerequisite for the implementation of Industry 4.0, but as the manufacturing complexity increases, digital twins will become more relevant. My personal view is that much could be achieved with a digital representation of a complete manufacturing/production setup. The construction of such a digital twin will be very complicated and challenging (of course depending on the production complicity). It would require a complete model description of all processes and process dynamics. Looking at an installation of a robotic arm or a simple assembly line, this could be achieved. Still, looking at more complex systems like biological processes (e.g., fermentation), chemical processes (e.g., synthesise) or physical processes like injection moulding it becomes more complicated. One aspect is the complexity in describing the systems, another challenge that is still very apparent is the lack of connectivity and data availability. These have been challenges at the Industrial partner throughout the PhD project and seems to be general challenges. Based on a comprehensive survey Garza (2018, [34]) state that *"Significant challenges are also encountered in the generation and collection of data* *from the floor. Automated data acquisition is critical for the implementation of cyper-physic systems. Yet, most small to mid-size companies still rely significantly on manual methods for data gathering"* and Frank et al. (2019, [8]) state that *"Our results also show that the implementation of the base technologies is challenging companies, since big data and analytics are still low implemented in the sample studied"*. On the more positive side, Yang and Gu (2021, [35]) have made an extensive analysis of the current state of Industry 4.0 initiatives and conclude that many theoretical concepts have expanded to real-world applications. They made an analysis across 14 countries where they focus on governmental policies and enabling initiatives. For Denmark, "MADE" is mentioned as the foundation for driving the Danish Industry 4.0 agenda. This has been successfully achieved with a bottom-up initiative, collaboration between Danish manufacturing companies, five universities, three research and technology organisations.

2.2 Data utilisation in manufacturing

The use of data in a manufacturing context for improving quality and productivity is not a new concept that arrived with the definition of Industry 4.0. The first approach originates back to the early 1920's where Walter A. Shewhart [36] introduced the concepts of Statistical Process Control (SPC). Later, W. Edwards Deming standardised SPC and in 1986 SPC became a central topic within the Six Sigma community (introduced by Bill Smith). Since Shewhart and Deming introduced SPC the use of data in manufacturing have scaled to many other application than SPC.

Over the last 20 years, the utilisation of data from manufacturing has been labelled with different names (mainly driven by the food and pharmaceutical industries). In 1992, Juran [37] introduced the frame of Quality by Design (QbD) that was adopted by Food and Drug Administration (FDA) and included in their guidelines. The idea of QbD was to reach quality goals by increasing the understanding of the interaction between the production process and the product quality. In 2004, the frame was Process Analytical Technology (PAT), defined by the United States Food and Drug Administration (FDA) [38]. The overall focus in PAT was to move away from only doing product release control and instead monitor critical process parameters during production to ensure product quality. The central part of both QbD and PAT was to increase production and process understanding to ensure product quality. As data utilisation is a central part of Industry 4.0, Industry 4.0 could be seen as a prolonging of initiatives in QbD and PAT. I do see a key change with the Industry 4.0 data utilisation, which is that the data analytics used to be conducted by subject matter experts (with a technical or engineering background) and therefore based on process and production insights, where the trend today is that data scientists with a background within computer science and machine learning take over these tasks. I, therefore, see a risk that the focus is going to move away from achieving process understanding to achieve a good prediction accuracy. There is considerable risk if the main focus is on correlation and not on causality. To avoid this, I see great importance in ensuring data science skills among subject matter experts and ensuring close collaboration between data engineers, data scientists, and subject matter experts.

As described above, there are many levels of data utilisation. In my view, it is essential not to neglect the classic and straightforward approaches in favour of machine learning approaches. As I described in the Introduction, simple data visualisation can, in many cases, be powerful and support operators to improve the production setup. At the same time, it is also crucial to be aware of new opportunities arising with machine learning, digital twins and cyber-physical systems. Based on my current and previous experience working with industry I will, in section 4, propose a data utilisation abstraction for injection moulding that combines new and classic approaches for effective data utilisation.

2.3 AI in manufacturing

A new trend within manufacturing analytics is "manufacturing AI" (Artificial Intelligence). Ten years ago, not many machine learning applications were labelled AI. To some extent, it has become a buzzword that is more and more used for everything within machine learning. Looking at funded projects with data utilisation within manufacturing (under the umbrella of Industry 4.0) and titles of special issues within journals that focus on data utilisation, it is clear that AI is becoming THE word to use when working with deep learning or machine learning[1]. It has become more and more popular to use AI for all kinds of computer and data analytic applications. Is that because the prerequisites have changed, or is it hype? To reflect on this, we first have to get the definition straight.

In 1950 Alan Turing [39] formulated the "imitation game" (the Turing test) designed to test a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human. The test is designed so that a spectator is asking a written question to a human and a machine. The machine passes the test if the spectator can't tell which response came from the machine based on the two answers. The first definition of AI was formulated in 1956 [40] by J. McCarthy, M. L. Minsky, H. Simon, A. Newell and C. E. Shannon. Their definition of AI referred to the ability of machines to understand, think, and learn similarly to human beings, indicating the possibility of using computers to simulate human intelligence. A central aspect when defining AI is the definition of intelligence. One definition of human intelligence is given by Colom (2010) *"Intelligence enables humans to remember descriptions of things and use those descriptions in future behaviours. It is a cognitive process. It gives humans the cognitive abilities to learn, form concepts, understand, and reason, including the capacities to recognise patterns, innovate, plan, solve problems, and employ language to communicate. Intelligence enables humans*

[1]Part of this text is taken from an assignment I handed in for the course - "AI for the people" at AAU in 2021.

to experience and think" and Kaplan and Haenlein (2019, [41]) are suggesting this definition of AI *"a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation"*. Should artificial intelligence be seen as the capability of mimicking human intelligence (not according to the two definitions above), or should artificial intelligence be seen as the capability to do beyond the capabilities of humans, or just the ability to detect patterns in data?

A general representation of AI is presented in Figure 2.4, where machine learning and deep learning are seen as sub-spaces of AI. This means that all machine learning is to be considered as AI, but not all AI as machine learning.

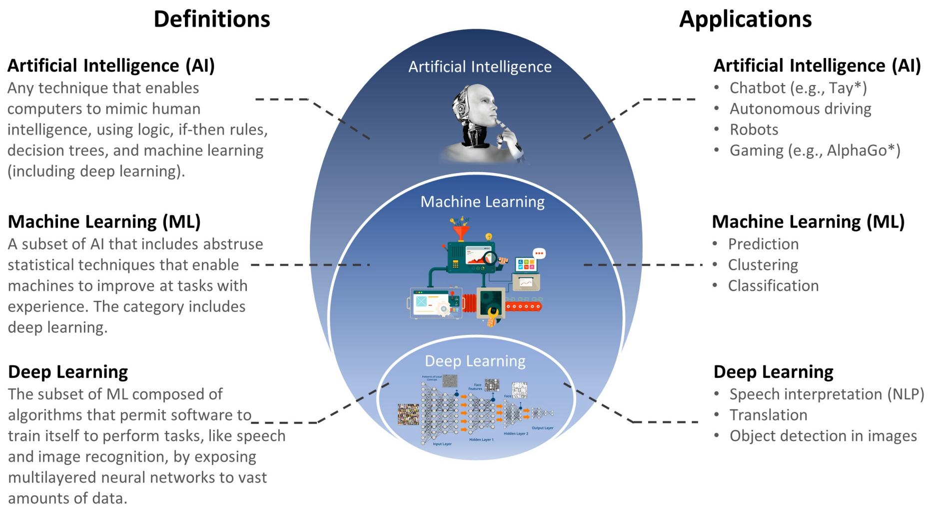


Figure 2.4. Illustration of the interaction between AI, ML and deep learning.
Source: The inspiration to the illustration is from [42] and the definitions to the left are from Claire D. Costa [43]. References to applications marked with*: Tay [44], AlphaGo [45]

Three levels are often used to classify AI applications; Narrow, General or Super. Kaplan and Haenlein (2019, [41]) has the following reflection on the three levels in relation to applications *"Today, first generation AI applications - those that apply AI only to specific tasks and are generally referred to under the label artificial narrow intelligence (ANI)—are near ubiquitous. They enabled Facebook to recognize faces in images and tag users, they allowed Siri to understand your voice and act accordingly, and they enabled Tesla to develop self-driving cars. In the future, we may see the second generation of AI, artificial general intelligence (AGI), able to reason, plan, and solve problems autonomously for tasks they were never even designed for. And we might possibly see the third generation, artificial super intelligence (ASI), which are truly self-aware and conscious systems that, in a certain way, will make humans redundant"*. Based on this, I see the definition by McCarthy et al. (1956 [40]) as artificial general intelligence, making narrow intelligence more machine learning than AI. I think that AI is used too widely and that the use of the term AI should be limited to real AI applications, where the aim is to mimic human intelligence (like in, e.g., advanced chatbots like Tay [44] and autonomous vehicles). With the more broad use of AI, the term is diluted and the impression of the application is confused.

Based on the above, I see the main part of the presented manufacturing application to be "just" machine learning applications (ML). Machine learning applications can be divided into three main classes (illustrated in Figure 2.5) based on the statistical approach used; Unsupervised learning, supervised learning and reinforcement learning, where the latter is the least common in manufacturing applications. Supervised methods are used in cases where both input (X data) and response (Y data) variables are available (also referred to as labelled data). When using supervised learning for a discreet response variable, classification can be applied, and regression can be used for a continuous response variable. If response variables are not present, unsupervised methods can be applied to the X data. In reinforcement learning, the system is given some input data, a set of parameters that can be changed and an objective. The system then tries different settings and, based on a reward system, optimise to find the best way to achieve the defined objective. Reinforcement learning works best in applications with a rapid response from the award system (e.g., gaming or robot applications). Examples of supervised learning can be seen in papers A, B, C, E and F and unsupervised learning in papers A, C, E and F.

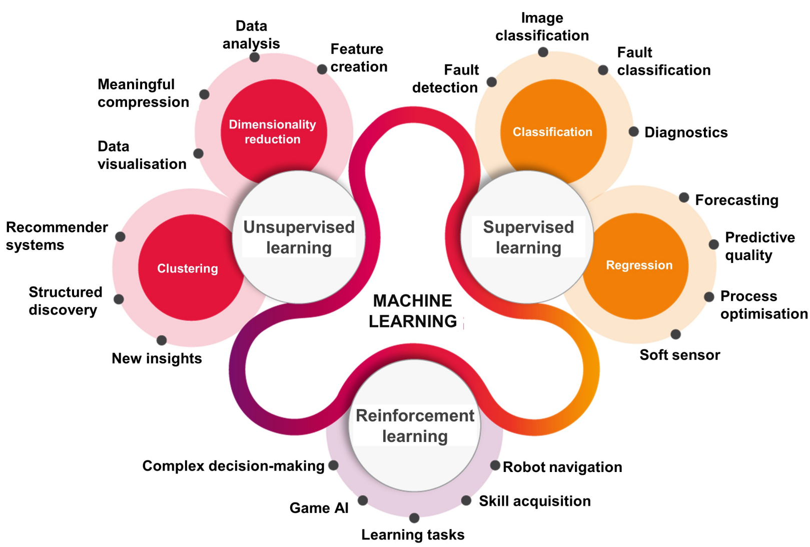


Figure 2.5. Illustration of the three classes of ML, where use-cases are given for each class. Source: Re-creation of figure from pngwing [46]

CHAPTER 3

Injection moulding

This section will introduce injection moulding covering the interaction between materials, quality and the moulding process. There will be a particular focus on data utilisation, monitoring and optimisation of the moulding process. The description in this section will be more comprehensive than the injection moulding descriptions in the included papers (A, B, C, E and F). This section aims to cover all the relevant information without repeating the description in the individual papers.

Injection moulding is a widely used manufacturing technique, and the majority of the plastic parts used everyday have been produced by injection moulding. This is because injection moulding offers large flexibility in size and shape of the produced elements, the option of using many different plastic materials (resulting in different product functionalities) and colours, the ability to make a large number of identical items with high precision and that the technique is scalable and cost-effective. The injection moulding plastics market size was estimated to be approximately USD 284.7 billion in 2021, with an expected annual growth rate of 4.6% from 2021 to 2028 [47], making it a significant and essential industry (as a comparison, the Danish gross national product was USD 355 billion in 2020 [48]).

For manufacturers producing plastic parts by injection moulding, the production setup can roughly be divided into three main areas; plastic materials, the injection moulding machines, and the injection mould that shapes the produced part. All three areas are critical to making plastic parts of high quality and are therefore addressed in the following.

3.1 Injection moulding machine and mould

An injection moulding setup consists of an injection moulding machine and an injection mould (see illustration in Figure 3.1). A moulding machine consists of an injection unit (left side in Figure 3.1) and a clamping unit (right side of Figure 3.1), where the injection mould is placed in the clamping unit. The main functions of the injection unit are to melt and mix the resin and inject the molten plastic into the mould, where the role of the clamping unit is to open/close the mould and ensure a sufficient amount of clamping force (closing force) when the molten plastic is injected into the mould so that the molten plastic stays within the mould cavities (will be described further in section 3.2).

Moulding machines are produced in different sizes and types, where the two main types are hydraulic and electric machines. In a hydraulic machine, the injection and clamping unit movements are powered by hydraulic pressure generated by a hydraulic pump. The movements in an electric machine are controlled with servo drives. The moulding machines used in the PhD project have all been hydraulic machines. Therefore, the difference between the two types will not be described further.

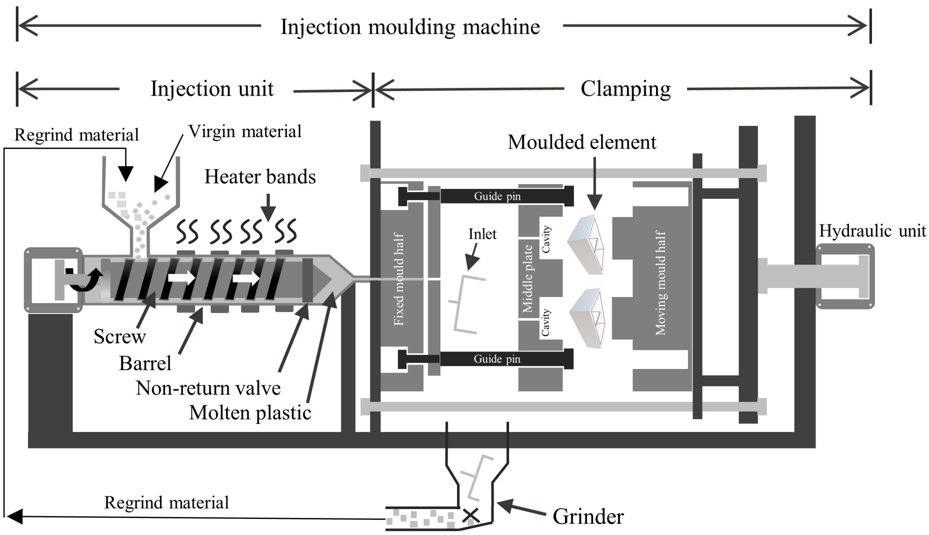


Figure 3.1. Illustration of an injection moulding machine with injection and clamping unit

Injection moulds differ in size and design. Moulds used for moulding large elements typically only have one cavity and therefore only produces one element per moulding cycle (the moulding process/cycle is described in the next section). Moulds producing small elements can produce multiple elements in each moulding cycle. At the industrial partner, moulds are constructed as a modular system where one type of mould box can be used for different insert packages (insert plate), each producing a specific element dimension (see illustration in Figure 3.2). This concept makes it possible to have the main components as a standard component and only customise a small part of the mould (the cavity and core plate). Besides reducing the complexity in mould manufacturing, this also reduced the complexity of condition-based mould maintenance (investigated in paper C, see section 8) since the acoustic emission mainly is impacted by the different mould box types and, to a lesser extent, the layout of the insert plate.

Moulds (the mould box) commonly consist of two or three plates (the mould in Figure 3.2 is a two-plate mould and the mould in Figure 3.1 is a three-plate mould). Using a three-plate mould makes it possible to separate the inlet from the moulded

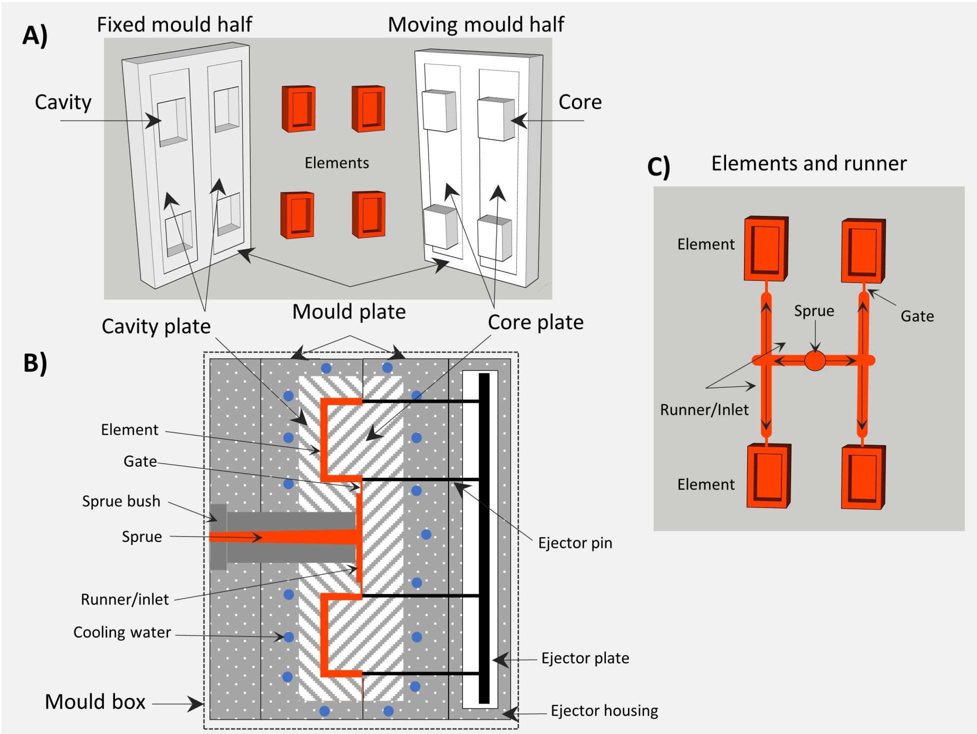


Figure 3.2. Illustration of a mould box with inserts and cores (two-plate mould with four cavities). A) shows the cavity and core plate that is the shape-given part of the mould. B) closed two-plate mould, where the cavity and core plate are seen in the middle of the mould and are the only two parts not part of the mould box. The cooling channels are placed in the mould plates surrounding the cavity and core plate. C) illustrates the runner/inlet system with elements, gates and sprue.

elements. When using a two-plate mould, the inlets will be attached to the elements after ejection. To overcome the challenge with the attached inlets, it is possible to use a hot-runner system (the alternative is a cold-runner mould), where the runner/inlet is kept warm and therefore doesn't solidify during cooling of elements. An additional benefit of using a hot-runner mould is that no inlets are generated (minimum of waste material). When using a cold-runner three-plate mould, the inlets are often collected and transferred to a grinder standing next to the moulding machine. The ground inlets are then mixed with virgin material in a mixer unit and reused in the moulding machine. This can often be done without any problems. Still, since the heating of the material in the barrel causes material degradation, this can be a challenge for some specific materials (Bernasconi et al. [49] have demonstrated that melt properties change as a function of the amount of regrind and that this also impacts the tensile strength on the produced parts). In work conducted as part of this thesis, we have only used cold-runner three-plate moulds where the inlets are ground and re-used in a mix with virgin resin.

3.2 Moulding cycle

A moulding cycle can be seen as a batch process that consists of a series of events. To simplify the description of a moulding cycle, the events are first seen as being serial events that don't overlap. An illustration of the main events/steps can be seen in Figure 3.3 and the four main steps are described below.

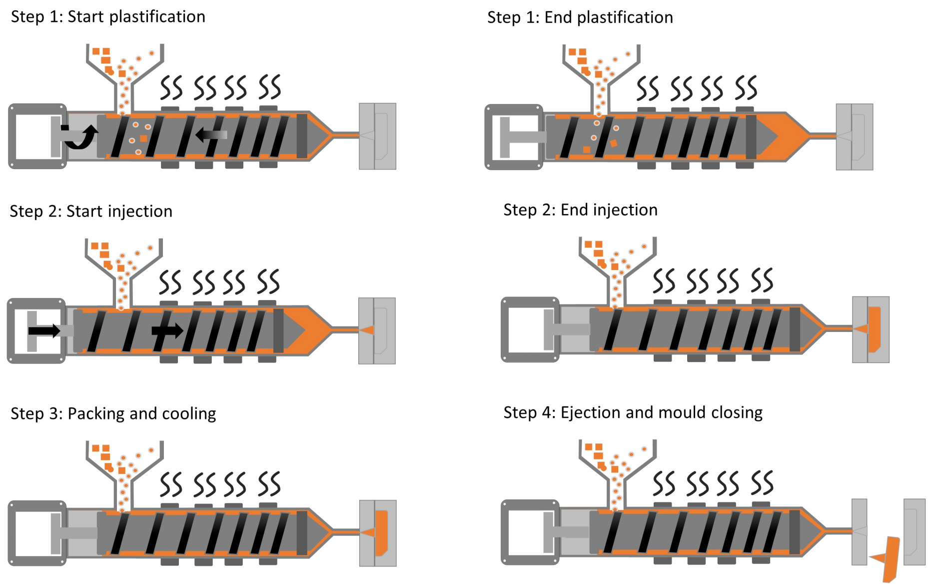


Figure 3.3. Illustration of screw movements during the moulding cycle

Step 1: **Plastification** - The first phase is preparation of the molten plastic. The plastic granulate enters the injection machine barrel through the hopper. The granulate is moved from the barrel entrance towards the nozzle by rotation of the screw (the rotation also mixes the granulate to ensure even melt quality). As the plastic granulate is moved through the barrel, it is melted by a mix of heat from the heating bands and friction heat. As the molten plastic moves forward in the barrel, the screw is forced backwards, leaving space in front of the screw for the molten plastic (the non-return valve at the tip of the screw ensures that molten plastic cannot flow backwards). The position of the screw is monitored and used to determine the amount of molten plastic (melt) in front of the screw. When enough melt is in front of the screw, the process will move into the next phase (awaiting the signal that the mould is closed and ready for injection).

Step 2: **Injection** - The next phase is the injection of the melt into the mould. This is done by moving the screw forward where the non-return valve makes the screw work like a piston. The motion of the screw is controlled so that a constant injection velocity is achieved (resulting in a varying injection pressure). When the screw reaches the switchover point (often when 95% of the melt is injected), the control strategy changes from velocity to pressure-driven control to obtain a constant pressure (start of the packing phase).

Step 3: **Packing and cooling** - As the melt cools down in the mould, the material will start to solidify and shrink. The core of the runners and elements will remain fluid longer than the surface. Keeping constant pressure on the melt in the packing phase will ensure that more melt is pressed into the elements, preventing the elements from being underfilled (see description of elements defects in section 3.3). As the melt cools further down, the viscosity will increase, and at some point, it cannot flow anymore (solidifies) and is considered frozen. The opening (gate) into the elements is often the narrowest place in the runner system, and therefore the place where the melt solidifies first (the gate freeze time). At this point, the process goes from packing to cooling since the elements cannot be impacted anymore. The length of the cooling period is determined by the demoulding temperature (described in more detail in section 3.4). The position of the screw at the end of the packing phase is called the material cushion and is often monitored to ensure that the same amount of melt is injected in each moulding cycle.

Step 4: **Ejection** - The last step of the moulding cycle is the ejection of the moulded elements out of the mould. As the elements cool down, they will shrink and release from the cavities' surface and bind to the cores. The elements and the cores are designed so that the elements will stay on the cores when the mould is opened, which will ensure that the elements are separated from the inlets. When the mould is fully opened, ejector pins in the mould will push the elements of the cores. The elements can now be collected.

In the description above, the moulding process is described as a series of events, but to reduce the total cycle time, some of the moulding events can be run in parallel, illustrated in Figure 3.4. Since the pressure in the barrel doesn't impact the elements after gate freeze, the plastification can be started when the gate freeze is reached. Plastification is, therefore, often run in parallel with the cooling phase. Since the machine hydraulic pressure is needed for mould opening (not the case for all machines), the plastification has to be completed before the end of the cooling time.

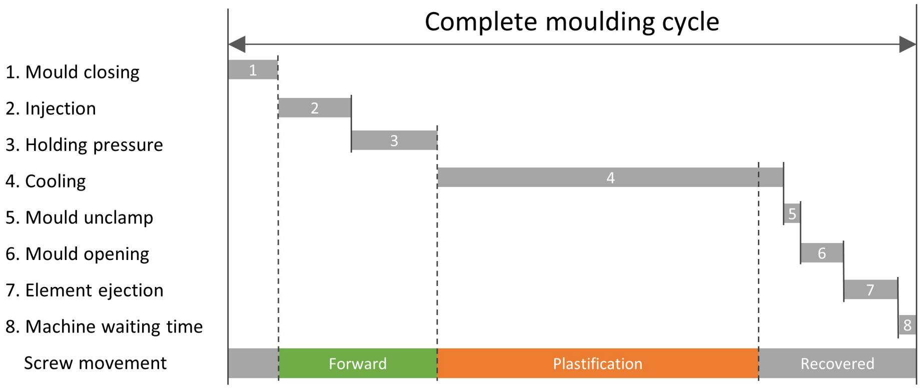


Figure 3.4. Illustration of the phases in an injection moulding cycle

A good illustration of the screw movements and phases between mould opening and closing can be seen in paper C, Figure 2 and 3, where the acoustic recording is labelled with screw movements and mould events.

3.3 Element quality and defects

The quality of moulded elements can be defined by different characteristics, where physical dimensions, element weight, functionality and visual appearance are among the most used.

3.3.1 Metrology

Element dimension and weight are impacted by both the moulding process and the condition of the injection mould. As shown in paper B, the holding pressure affects both element length, width and element weight, and in paper F, we demonstrated that the element dimension was changing over the mould's lifetime due to wear and tear (e.g., the element length increased due to removal of surface material in the cavities). Element dimensions can be categorised as mould bound or process affected. A mould bound measure is controlled by the shape of the cores (preventing the elements to reduce in size caused of shrinkage) and could, e.g., be the inside length of the element in Figure 3.2 whereas the outside length is process affected. The difference is that during cooling in the mould, the shrinkage in the element cannot reduce the inside dimensions since the element cannot become smaller than the core. Still, a considerable shrinkage after gate freeze can reduce the outside dimensions (a gap will appear between the surface of the cavity and the surface of the element). If the shrinkage occurs after the ejection of the elements from the mould (not cooled sufficiently in the mould), all element dimensions can be impacted.

3.3.2 Visual quality

The most common visual defects experienced at the industrial partner (relevant for the work conducted in the PhD) are collected in Figure 3.5. The visual appearance characteristics, main cause and potential counteractions are described in the list below.

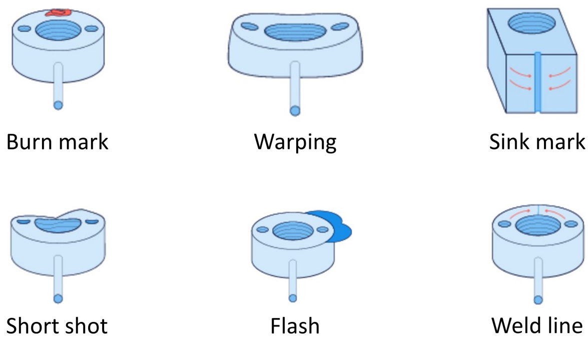


Figure 3.5. Illustration of some common defect types on injected moulded elements (Image Source: https://mechanicalenotes.com/injection-moulding-process/)

- **Burn marks** can appear as black spots (or degraded material) on the moulded elements and can be caused by different effects. Often an insufficient venting of the mould can cause burn marks (trapped air/gas in the mould cavities can combust) [50]. Inadequate venting can be due to the wrong design of air vents in the mould, wax or material trapped in the air vents or running with a too high injection velocity.
- **Warping** is often caused by uneven shrinkage in the element. If the shrinkage is even within the element, it just becomes smaller during cooling, but if the element shrinks at different rates, internal stress can cause the element to warp when ejected from the mould. It is essential to consider shrinkage and warpage when designing the elements since, e.g., the position of the inlet and differences in element thickness impacts the shrinkage and thereby warping. As shown by Chiang and Chang (2007) [51], process parameters such as packing pressure, mould temperature and packing time significantly impact warpage and, therefore, can be used to compensate for warping.
- **Sink marks** are seen as a small sink on the surface of the elements, often in areas with large element thickness. A common cause to sink marks is that the

elements are not sufficiently filled in the packing phase, and then when the core of the elements shrinks/contracts during cooling, it will leave a sink on the surface of the element. Mathivanan and Parthasarathy (2009) [52] observed that a higher melt temperature in combination with lower mould temperature and higher pack pressure resulted in minimisation of sink depth.

- **Short shot** is an incomplete filling of the element and is caused by too little melt injected into the mould and is often seen on the edges of the elements (part of the element is missing). Moayyedian [53] identifies gate type, melt temperature and packing pressure to be impacting the level of short shot. If a short shot occurs during production, it can often be detected by a change in the material cushion (monitor the volume injected melt into the mould).

- **Flash** on an element occurs when the material escapes the cavities and gets captured in the mould closing surface, resulting in a flap of excess material on the side of the element (seen in the closing/parting line between the two mould halves). Flashing can be caused by too low closing force on the mould, if the material is stuck on the closing surface so that the mould does not close entirely or by mould wear.

- **Weld line** is seen as a line on the element where two melt flow fronts meet (often opposite the inlet point). It is a challenging defect to eliminate but can be minimised by doing flow simulations in the elements design phase (investigating the placement of the inlet). Besides the visual impression of the elements, weld lines also reduce the strength of the elements (weld lines are a weak spot). Sreedharan and Jeevanantham (2018, [54]) demonstrated that the backpressure, holding time, and melt temperature are all important in reducing the level of weld lines. These process parameters impact the material flow speed in the mould where the earlier the two flow fronts meet, the warmer they most likely are and, therefore, better melt together.

3.4 Materials

As described in paper A and B, the material properties impact the quality of the produced elements. Some introductory material understanding must be introduced to fully understand the interaction between material properties, moulding process, and element quality.

Some of the central findings in the conducted work are related to classifying or characterising material variation and utilising this information to reduce variation in element quality. In paper A and B, we have shown that machine pressure profiles can be used as an indirect measure of material variation in ABS coming from different suppliers. Therefore, this section links these findings to plastic material science to better understand the pressure profiles' use for indirect characterisation of material properties and differences. This will be limited to only including ABS since we have only used ABS in the PhD work, and since ABS plays an essential role as the industrial partner.

3.4.1 Acrylonitrile Butadiene Styrene (ABS)

Plastics can either be thermoplastics or thermosets, where thermoplastics become soft when heated and hard when cooled (no chemical reaction). In contrast, thermosets plastics react chemically when heated and form crosslinks (curing process), which is irreversible. This means that it is hard to recycle/reuse thermosets plastics, whereas thermoplastics can be reused multiple times (e.g., in injection moulding, the runners can be reused as thermoplastics). Thermoplastics can either have an amorphous or crystalline structure. Crystalline material forms highly ordered crystalline structures when cooled (imagine boiled spaghetti going back to its unboiled state when cooled). In contrast, amorphous material keeps its unstructured order when cooled (boiled spaghetti that just go stiff). Plastics can also be semi-crystalline, where part of the polymers form ordered crystalline structures when cooled, and the reaming stay unstructured [55]. ABS is a thermoplastic and amorphous polymer.

ABS have a glass transition temperature at around 105°C (structure turns “viscous liquid or rubbery”). ABS can be used for injection moulding with a melt temperature in the range of 210-270°C and a mould temperature of 40-70°C. An injection Pressure of 500 - 1000 bar is often used with a moderate to high injection speed. ABS has a shrinkage of 0.4-0.8%, which must be considered when designing the element and injection mould. ABS is hard and tough in nature, has high dimensional stability, good impact resistance, high surface brightness and resistance to dilute acid and alkalis. ABS has a poor weathering resistance; stress cracking can occur in the presence of some greases, poor solvent resistance (particularly aromatic, ketones and esters) [56]. ABS can be used for many different element types, where some examples are; electrical enclosures, computer keyboards, vacuum cleaners, 3D construction toys, automotive door liners handle and instrument panels.

3.4.2 Rheology properties for plastic materials

When it comes to doing injection moulding, some essential material rheology properties come into play. Molten plastic is a non-newtonian fluid, meaning that the melt properties are depended on temperature, pressure and shear rate. For sticky fluids flowing in a pipe (polymer stick perfectly to a metal surface), the velocity will be zero at the walls and max in the centre. For newtonian fluids, the flow-front will have a parabolic shape (Figure 3.6 A, whereas non-newtonian fluids will be flatter. The flow of polymers is always laminar, so moving in layers (called shear flow). The viscosity and the shear rate will determine the flow front shape. For newtonian fluid, the shear rate (velocity gradient) is defined as [57]:

$$\dot{\gamma} = \frac{\partial v_x}{\partial y} = \frac{\partial \gamma}{\partial t} \tag{3.1}$$

where γ is the shear strain defined as: $\gamma = \frac{\partial x}{\partial y}$, ∂v_x is the change in velocity and ∂y is the distance from the wall to the centre. The shear stress is defined as [57]:

$$\tau = F/A = \eta \cdot \dot{\gamma} \tag{3.2}$$

where F [N] is the force pushing the flow-front, A is the area $[m^2]$ and η the viscosity or resistance to flow $[Pa{\cdot}s^{-1}]$. Since there is symmetry around the centre of the pipe, gamma dot will be zero at the centre since there is no velocity gradient (same velocity for both layers around the centre). For a newtonian fluid, the flow-front is a parabola, and by taking the derivative of the function, the shear rate ($\dot{\gamma}$) is found to be a straight line with zero at the centre and max at the wall.

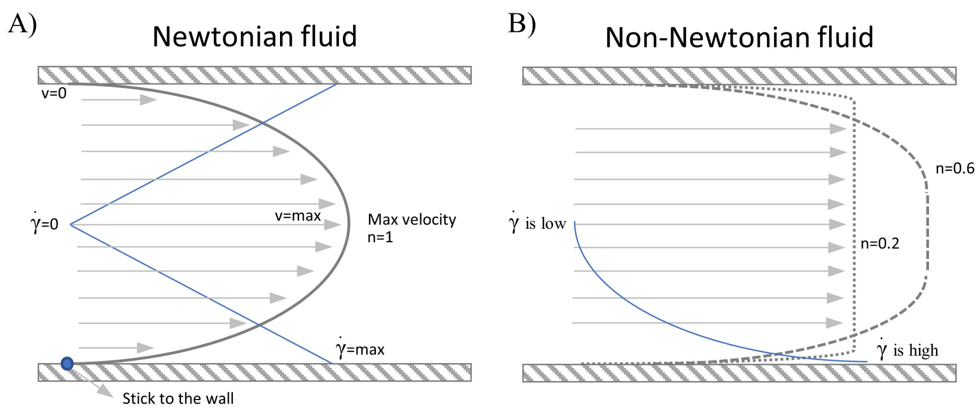


Figure 3.6. Velocity profile of material flow, A) being a newtonian fluid and B) being a non-newtonian fluid. n is calculated from the slope of the viscosity function in Figure 3.7 B) (n = 1 - slope). Source: Inspired by content of BIMS course by Dr. Vito Leo [57]

Viscosity can be measured using a capillary rheometer. A Capillary rheometer consists of a cylinder containing molten material and a capillary at the bottom of the cylinder. The molten material is, by a piston, pushed out through the capillary using different flow rates, and the corresponding pressure drop is measured. The shear rate ($\dot{\gamma}$) can be calculated from the flow rate, knowing the radius of the capillary and the pressure can be approximated to be the shear stress (τ). Plotting the calculated $\dot{\gamma}$ versus shear stress, the viscosity is found as the slope of the line. For a newtonian fluid, the relationship will be linear, whereas the relationship is decaying for a non-newtonian fluid (Figure 3.7 A). The viscosity curve is generated by plotting the viscosity as function of the shear stress (Figure 3.7 B), where low shear stress results in newtonian behaviour. As the shear stress is increasing, the viscosity will drop due to disentanglement of polymer chains (shear thinning). This means that the more the polymers are pushed, the more aligned they become, and thereby exhibits less resistance to flow. The shear thinning reduces the viscosity by a factor of a thousand, making injection moulding possible.

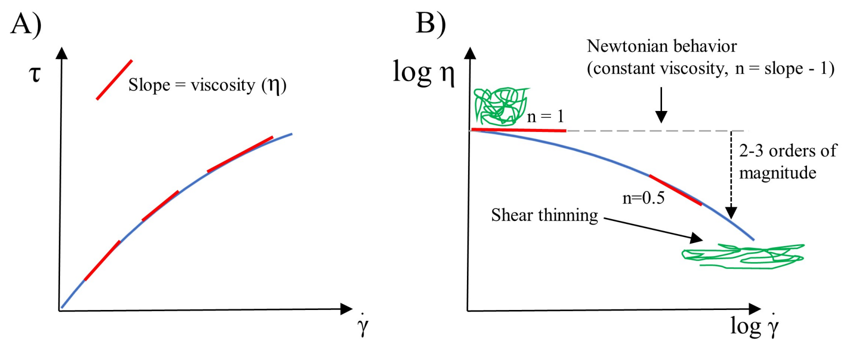


Figure 3.7. Material viscosity as a function of shear rate, B). The green spaghetti-like molecules in the upper part of plot B), illustrates long chains of polymers randomly in tangled. As the shear rate increases, the polymers become disentangled, illustrated in the lower part of plot B). Source: Inspired by content of BIMS course by Dr. Vito Leo [57]

Looking back at equation 3.2, we assumed that the viscosity was constant (true for newtonian fluids), so for a non-newtonian fluid we need to change the equation to take the change in viscosity into account [57]:

$$\tau = \eta(\dot{\gamma}) \cdot \dot{\gamma} \tag{3.3}$$

In Figure 3.6 A the shear rate was linear (first derivative of a parabolic function). With the shear rate not being linear for a non-newtonian fluid, the shape of the flow front has to be something else than parabolic. From Figure 3.7 it can be seen that at a low shear rate, the viscosity is high (steep slope), and with the shear rate being lowest in the centre of the flow, there will be a high resistance to flow. This results in a flat flow front, as illustrated in Figure 3.6 B. As the shear rate increases, as we approach the side of the pipe, the viscosity drops and making the flow easier (increasing the velocity).

The viscosity discussed here is measured at a steady-state (after 5-10 seconds of flow in the capillary rheometer). In real applications in injection moulding, there will be a transient region before a steady state is reached (the polymers need time to disentangle). In the transient region, shear acts opposite (activation energy), so moving with a high flow will increase the transient viscosity compared to the steady-state viscosity (illustrated in Figure 3.8). Applying higher pressure to the polymer to get them to move faster (the velocity is achieved by applying pressure) makes it harder for the polymer to disentangle, resulting in the need for more pressure. As the polymers become disentangled, the viscosity will drop, resulting in a lower viscosity than achieved using a lower pressure/injection speed.

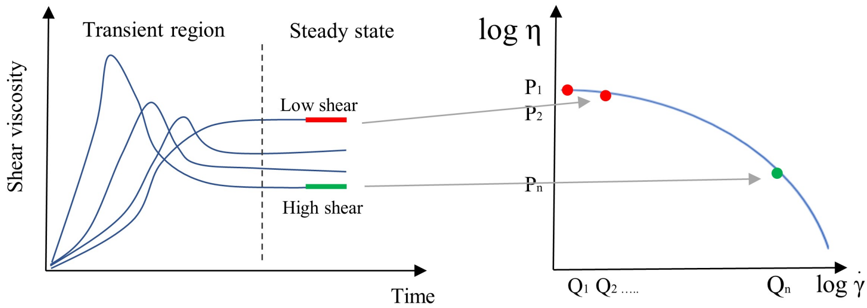


Figure 3.8. Shear viscosity in the transient region of the injection. Source: Recreation of content from BIMS-1 course by Dr. Vito Leo [57].

Temperature and pressure also impact the polymer viscosity. Increased temperature will increase the mobility of the polymer resulting in lower viscosity, whereas increased pressure will decrease the mobility of the polymer resulting in higher viscosity. Equation 3.3 also needs to be updated to include the temperature and pressure dependency:

$$\tau = \eta(\dot{\gamma}, T, P) \cdot \dot{\gamma} \tag{3.4}$$

The impact of temperature on the viscosity differs from material to material, and for some, it can almost be neglected (for semi-crystalline materials). For amorphous materials like PVC and ABS, the temperature impact is significant. At a low shear rate, a change in barrel temperature of 5-10 °C can reduce the viscosity by a factor of two [57]. The pressure impact on viscosity is highly dependent on the shear rate. At high shear rates (during injection), the pressure impact is very low, whereas, during packing (low shear rate), the effect of pressure is very high. This means that increasing packing/holding pressure to get the last molten plastic into the cavities will increase viscosity (by a factor of 2-3), making it even harder to fill the cavities. This is why reaching a high filling of the cavities during the injection is crucial.

Based on the above, it is clear that material properties are essential for controlling injection moulding and the quality that is produced. The capillary rheometry described is one of the best approaches for describing the material properties impacted by injection moulding. Still, at the same time, the test is conducted at a much lower shear rate than used in injection moulding. Therefore, rheometry methods have been developed utilising an injection moulding machine, Injection Molding Rheometer (IMR). In this case, the injection moulding machine is substituting the cylinder in the capillary rheometer, making it possible to simulate production-like conditions regarding injection pressure, speed and shear rates. Therma-flo$^{\mathrm{TM}}$ [58] is an example of a commercial solution where a specially designed mould with cavity sensor, combined with machine profiles, are used to measure and analyse the material proprieties. The solution can be used to compare and quantify the difference between materials and evaluate a material's flow sensitivity to temperature and pressure. With this advanced setup, it should be possible to characterise material viscoelastic properties. A setup like this is only available for doing screening tests of materials since it utilises a dedicated mould and sensors (still a very relevant and essential application). Being limited to a single mould, it can not be used in combination with different production moulds with different runner systems and element designs. Therefore, the question is how much of the "material characterisation" could be archived only using pressure profiles from a standard injection moulding machine. Using such an approach, it would be possible to utilise the "material characterisation" across different moulds and during element manufacturing (for monitoring or closed-loop control).

3.4.3 Injection moulding pressure profiles and material properties

Paper A and B demonstrated that some of the material differences impacting element quality could be identified using the pressure profiles from the injection moulding machine. In Paper A and B, it was found that both the pressure profile from injection and the dosing phase could be related to relevant material differences. A plot of dosing and injection profiles from work done in paper B is presented in Figure 3.9. The plotted profiles represent data using ABS from two different vendors, where minor differences in material properties are expected. Looking at the dosing and injection phases, it is evident that the injection phase has the most considerable resemblance to the process of measuring material viscosity properties seen in Figure 3.8.

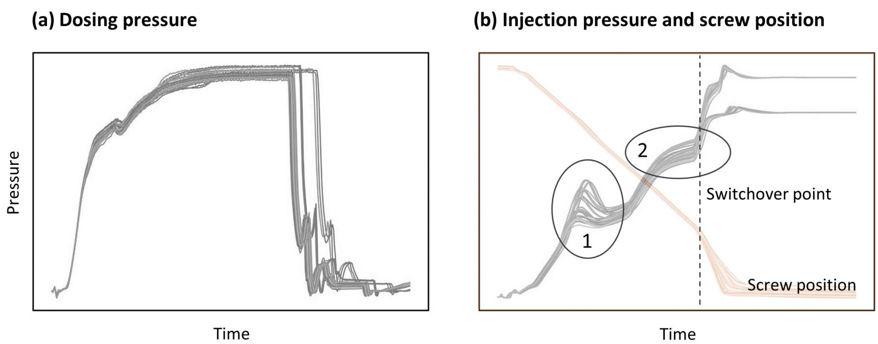


Figure 3.9. Machine profiles from paper B, a) dosing pressure and b) injection pressure overlaid with screw position (orange lines). The peak marked 1, is likely to represent the activation energy, and 2, the steady state region.

The injection phase has been run with the same injection speed, limiting the information generated about the material properties. To achieve viscosity-like results (like in Figure 3.7) it would be necessary to do experiments with different injection speeds resulting in varying shear rates. Each flow rate (or shear rate) gives one resulting steady-state pressure using a capillary rheometer. The most likely position to get a steady-state pressure reading on a moulding machine during a normal moulding cycle would be at switchover (area marked with 2 in Figure 3.9). Bozzelli (2012 [59]) have demonstrated that a plot of the relative viscosity can be generated for any material, mould and injection moulding machine combination running a simple test. In the test, several injections are conducted with varying injection velocities (shear rate is then calculated as 1 over the resulting injection time), and the resulting pressure at switchover is captured. The relative viscosity is then calculated as the pressure at switchover, times the injection time, plotted against the shear rate. The test is used to determine the optimal injection velocity (point on the curve with a low slope) for the given material, mould and machine combination.

At the industrial partner, there is a dedicated process for creating a moulding process for all new moulds (will be described in the next section). One step in this is to find the optimal injection pressure, achieved using step-wise change of the injection velocity. Collecting the injection profiles in this step would enable the calculation of the relative viscosity. These profiles would also be a valuable dataset making it possible to better understand the interactions between mould design, machine conditions and different materials. For the individual mould, this could be used to improve troubleshooting when production issues occur. When production issues are encountered, the test can be repeated on the production machine (new mould, machine and material combination) and compared to the original test results. This will indicate what has changed and how this could be corrected.

The same approach (varying injection velocities) could be used in a material test scenario to compare materials of different grades or from various supplies. Where one outcome would be the relative viscosity and an other would be the machine pressure profiles. All materials are tested in a standard moulding setup (identical moulds) meaning that the profiles can be compare directly.

3.5 Moulding process - creating a mould card

When designing a new element for injection moulding, the first step is selecting a plastic material that can be used to create the given element shape with the desired properties. The next step is to develop a mould that can produce the element. Often one of the standard mould boxes can be used, limiting the design to the inserts and cores. For more complex elements a new mould construction might be needed. The production inserts and cores (or complete mould) is then produced with the given specification. The next step is to create a moulding process for the given element shape, mould and material combination. The moulding process for a new mould is developed in a systematical running-in process where all critical process parameters are determined (e.g., barrel temperature, injection speed, holding pressure, mould temperature, cooling time, switchover point). At the industrial partner, a dedicated mould qualification team is responsible for this. Due to confidentiality, only an overall description of the running-in procedure will be described here. As a reference, Moser (2012, [60]) presents an alternative running-in approach using design of experiment. The outcome of the running-in process is the creation of a ”mould card” that specifies the settings for creating a moulding process that will produce moulded elements within the defined quality (element functions, appearance and dimensions). Before process settings can be explored, a series of mechanical settings have to be made to ensure safe and effective moulding. This could be mould closed position, mould open/close speed, ejector force and speed. When this is in place, the following process settings can be determined (the list is not complete but contains the most relevant process settings):

- **Melt temperature -** The first step in the running-in process is to set the barrel heating bands so that the desired melt temperature is achieved. The melt temperature is measured by ejecting melt into a cup and measuring the temperature with a temperature probe while stringing.
- **Injection speed -** Now running with the correct melt temperature, the next step is to find the injection speed (cm3/s). The optimal point is where there is a good relationship between injection speed and the pressure used to achieve this (often, the injection speed that requires the least injection pressure is selected).
- **Switchover point -** Knowing the injection speed the next step is to determine the switchover point. As described in the material section 3.4, it is desired to achieve approximately 95% filling in the injection phase using velocity control (utilising the low viscosity at high shear rate) before entering the packing phase. The reason for changing to pressure control is that there is a high risk of overfilling using velocity control to completely fill the cavities. The 95% filling is found by turning off the packing phase and then changing the amount of material injected to produce underfilled elements. It is evaluated by visually inspection when the elements are 95% filled and thereby, when the switchover point is reached.
- **Packing pressure -** The next phase to optimise is the packing phase, ensuring that the elements are entirely filled without being overfilled. The theoretical volume of the moulded elements is known from the geometry of the cavities, and the material density is known from the material specification. From this, the theoretical element weight can be calculated. From a starting pressure (resulting in underfilled elements), the packing pressure is increased in steps until the weight of the elements reach the theoretical weight (a high packing time is used to ensure that the time is not the limiting factor). The pressure at this point is selected as the packing/holding pressure.
- **Packing time -** Having the packing pressure in place, the packing time can be optimised to reduce the total cycle time. This is done by keeping the packing

pressure constant and in steps reducing the packing time (often in steps of 0.5 seconds). When the packing time becomes too low, the elements will start to loose weight. The point of time immediately before this happens is selected as the packing time.

- **Cooling time** - The last step in the running-in procedure is to find the cooling time that results in the optimal demoulding temperature. The cooling time has to be long enough to ensure that the temperature in the hottest part of the moulded element is well below the glass transition temperature (Tg). This is crucial to reduce the risk of deformation at and after demoulding. The element temperature is measured with an infrared thermometer, just after ejection, at the thickest part of the element. The cooling time is often the largest contributor to the total cycle time. Therefore, it is essential to reduce it as much as possible without compromising the element quality.

New trends are going towards conducting the initial running-in in a digital twin of the moulding machine and mould. Moldflow [61] and Moldex3D [62] are two commercial software solutions for doing mould flow simulations that include simulation of the injection phase in the moulding machine. One of the biggest challenges is getting a sufficiently good model description of the relation between element quality, material rheological properties and process settings. As these tools become better, they will potentially have a considerable impact on element and mould design and the running-in procedure. One of the challenges today is that the running-in of the mould is the first evaluation of the relation between the mould design, element design, moulding process and element quality. For known element shapes and materials, this rarely causes any problems. Still, when creating significantly new element shapes or using new materials (e.g., new sustainable materials), it occurs that a moulding process can not be designed in such a way that acceptable quality is achieved. In these situations, the injection mould might have to be scraped, or adjustments have to be made in the mould to fix the problems. This is costly and increases the lead time from element design to element production. Having a digital twin setup of the moulding machine and the new mould will make it possible to create and test an initial moulding process before the injection mould is manufactured. Issues can then be fixed in the 3D model of the element/mould, and a new simulation can be conducted. This will increase the likelihood for first time through using new materials or element shapes. As part of the PhD project, an initial test of this concept has been conducted together with the CAE (Computer-Aided Engineering) department at the industrial partner. Convincing results were obtained, and the CAE department carried on the work.

3.6 Injection moulding process data

When looking at process data within injection moulding, they can be split into two groups: data used for creating the moulding process for the specific mould (described in section 3.5) and data used for process monitoring and adjustment during production. The process parameters collected in this section are based on industrial partner experience and a review of scientific work on data utilisation within injection moulding (Thyregod et al. 2001 [63], Kazmer 2008 [11] and Hopmann 2021 [64]). It differs from company to company how the split is between fixed parameters on the mould card and flexible parameters that the moulding operator can adjust during production. The main part of the process parameters at the industrial partner is fixed on the mould card, leaving only a few to be adjusted during production and only within specified narrow limits. If more adjustment is needed, this has to be done by an engineering department or the mould qualification department. The list below contains the essential process variables collected at the industrial partner and a reflection on how these can be used for monitoring and process adjustment.

- **Injection pressure** is the force applied by the screw to inject the melt (molten resin) into the mould. Up to 95% of the machine injection pressure capacity is often used for the injection. The injection pressure is not measured by a pressure sensor in the barrel. Still, it is calculated based on the system pressure[1] and the ratio between the hydraulic cylinder area and the area of the screw tip. The injection pressure can be collected at different places in the moulding cycle, e.g., max injection pressure, injection pressure at switchover. The injection pressure is not a setting but a result of the used injection speed and material properties. Since the injection speed is kept constant, the injection pressure can be used to monitor changes in the material properties.

- **Holding (or packing) pressure** is used to obtain the last filling of the cavities. The holding pressure is constant and lower than the injection pressure. The holding pressure is applied until gate freeze (often 2-4 seconds). The holding pressure is a set value and cannot be used to monitor the moulding process. At the same time, the holding pressure is one of the most essential process parameters for compensating for material variations (as demonstrated in papers A and B) and, therefore, highly relevant to include when analysing moulding data. At the industrial partner, the holding pressure can, during production, be adjusted within medium pressure +/- 70 bar.

- **Clamp pressure (clamping force)** is the force applied to keep the mould closed during injection and packing (needs to be higher than the injection pressure).

[1] System pressure is the hydraulic pressure used by the injection moulding machine. The pressure is measured in the hydraulic system.

- **Back pressure** is the pressure generated when the screw is dosing new material (move backwards in the barrel, step 1 in Figure 3.3). As demonstrated in papers A and B, the back (or dosing) pressure profile (pressure as function of time) can be used to monitor material variations. The back pressure process value available on the moulding machine is one value on this profile (usually the max value).

- **Barrel temperature** is often measured at multiple positions on the barrel and can be used for controlling a desired temperature profile through the barrel. The temperatures are measured between the barrel surface and the barrel's insulation. The temperature, therefore, doesn't give a true reflection of the melt temperature. Instead of only collecting the barrel temperature at the different zones, I think it could be interesting also to collect the energy consumed by the heating bands (this is possible on some moulding machines) since this will give an indirect measure of the amount of sheer heat generated in the barrel. This measure could potentially be used to monitor changes in the material properties caused by material degradation or variation in the virgin material.

- **Material cushion** measures the material in front of the screw after the injection and packing phase. The cushion can be seen as a consequence measure of the moulding process and is essential to monitor. Changes in the cushion can be caused by a change in material properties, wear on the return valve, or an element stuck in a cavity (less material can be injected into the mould).

- **Plastification Time** is the time it takes to dose new melted plastic in front of the non-return valve. As the transport of the melted plastic is achieved by rotating the screw at a constant speed, the plastification can vary. Since the time variations are mainly caused by a change in material properties, monitoring the plastification time is an indirect measure of variation in the material mix.

- **Injection Time.** When running with a constant injection speed, the injection time will reflect variation in material properties. The injection time is therefore also relevant to monitor.

- **Cycle time** is a sum of all time contributions in the injection moulding cycle and, therefore, is not optimal for monitoring. Since the cycle time is a sum, it won't reveal what is causing the variation. At the same time, there is a risk that variation in one time contribution can be evened out by an opposite variation in another time contribution. Therefore, the cycle time is not used for monitoring but can be used for comparison of similar moulds running with a too long cycle time (could then potentially be optimised).

Besides the variables listed above (collected at the industrial partner), the following variables are highlighted to be important by Kazmer 2008 [11] and Hopmann 2021 [64].

- **Mould temperature** is often presented as the cooling water temperature that leaves the mould. This temperature will not indicate the temperature in the mould cavities (where the temperature is relevant to know). To have a good representation of the temperature in the cavities, additional sensors have to be installed in the mould. One thing is to monitor the cooling water temperature, but I find it more interesting to monitor the amount of energy removed from the mould. This is possible if the mould cooling system is equipped with a flow and temperature sensor on the cooling water[2].

- **Switchover point.** The relevance of using the switchover point for monitoring depends on the switchover strategy. When using pressure to trigger the switchover, it is relevant to monitor the position since this will reveal variation in the material properties. In contrast, if the screw position is used to trigger the switchover, monitoring the pressure at the switchover is more relevant since the position will be close to constant [65].

The process variables listed above are readily available on standard moulding machines and, therefore, often used for process monitoring. This will be described in details in section 4. As described in paper A and B additional data can be collected from in-build sensors in the moulding machines or by adding additional sensors to the mould and moulding machine. In paper C, this has been exemplified by adding acoustic sensors for condition-based maintenance. These data sources are described in the respective papers and will not be described further here.

[2] $Q = m \times Cp \times \Delta T$ where, Q is the Energy (J), Cp is the specific heat capacity of coolant (J/g/°C), m is the mass (g) of coolant and ΔT is the temperature difference of coolant in and out of the mould.

CHAPTER 4

Process monitoring and control

The purpose of this section is to give an introduction to process monitoring with a particular focus on monitoring within injection moulding. The overall goal of process monitoring is to ensure a production process that stays within the conditions to produce products that comply with specified product specifications. If the production process ran in an isolated environment with constant inputs and no disturbances, the process output would also be constant. This is an ideal state that never occurs in a production setup. In practice, there will always be disturbances that impact the production process. The most critical disturbance factor affecting an injection moulding process is illustrated in Figure 4.1. In the illustration, inputs have been defined as mould type, element shape, type of material and moulding machine settings (mould card settings). The type/configuration of the moulding machine is seen as a disturbance (the same moulds are used in different moulding machines). This reflects the production setup at the industrial partner, but it is also expected to represent injection moulding setups used in general.

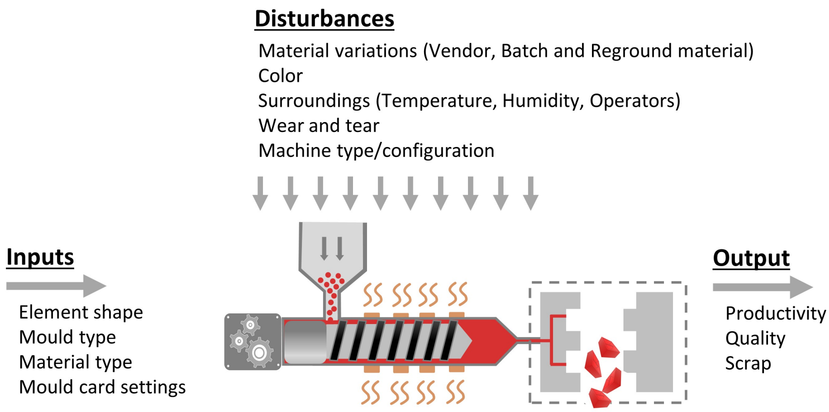


Figure 4.1. Illustration of typical input parameters and disturbances impacting the output of an injection moulding process (Image Source: Moulding machine illustration is from paper A)

The moulding machine contains many internal control loops designed to ensure a constant and stable moulding process. The standard control loops will e.g. try to ensure constant injection speed, constant pressure in the packing phase, precise movement of the screw and constant barrel temperature. This level of control is crucial to produce elements with consistent quality continuously. The standard inbuilt control loops in an injection moulding machine will only to some extent be able to compensate for the disturbance factors listed in Figure 4.1. It is, therefore, often necessary to add an additional layer of control to ensure consistent quality and productivity output. This has been described by Kazmer (2003, [66]), where he distinguishes between machine control, state variable control and setpoint control (illustrated in Figure 4.2). The machine control is the closed-loop control described above. It is highlighted that many of the machine default settings for the controllers (set initially by the machine manufacturer) are seldom changed, which can lead to an insufficient moulding process. The inbuilt machine control is continuously improved to ensure a constant moulding output. Kazmer (2003, [66]) states that it is often beneficial to add additional monitoring or control, where two different approaches are mentioned. The first is based on monitoring process state variables. The state variables reflect the melt condition (e.g. melt pressure, melt volume, flow rate, inlet pressure or element stress). The state variables are the resulting effects of a given injection moulding process. An example could be changing the injection speed where the resulting state variable would be change in melt viscosity. Many of the state variables listed by Kazmer can not be measured directly, making state variable feedback and control a challenge. A different challenge is that some relationships between set points and apparent state variables can be misleading. One example is changing the input setting of barrel temperature, where the resulting state variables would be the melt temperature. Here, the pitfall is that more than 50% of the melt heating originates from shear heating (impacted by polymer properties and screw design). One typical state variable control example is utilising cavity sensors to make real-time adjustments to compensate for pressure and viscosity changes.

Kazmer's last level of control is setpoint control, where process performance is used as feedback to adjust machine set points. Kazmer describes this as evaluating quality from one moulding cycle that can be used to adjust the settings for the next moulding cycle. However, Kazmer doesn't reflect on what such quality measures should be and how they can be measured/determined close to real-time. One approach could be introducing a soft-sensor that, based on available process variables, could predict the element quality and then use this prediction to control the moulding process. This approach is closely linked to the work conducted in paper B. The utilisation of state variable and setpoint control is generally limited by the need for a model relation between critical element quality attributes, process variables and uncontrolled process disturbances.

The process disturbances introduced in Figure 4.1 are to some extent uncontrollable, but restrictions/limitations can be introduced to limit their negative effect. Adding constraints to a production setup will always have drawbacks in the form of reduced flexibility. At the industrial partner, a specific mould type should be able to produce approvable element quality in all available moulding machines of a given size

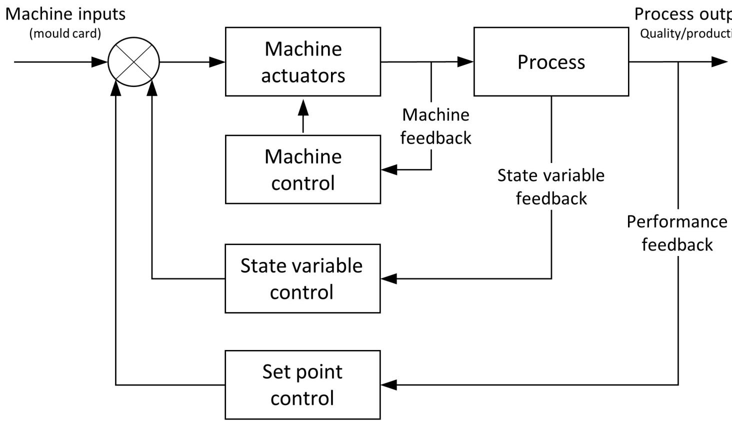


Figure 4.2. Levels of control in an injection moulding setup (Image Source: Modified version of figure 8. by Kazmer (2003, [66])

(clamp force). This will cover machines from different vendors, machines of different ages, maintenance states and configurations (all introducing uncontrolled variations). Reducing this uncontrolled variation by limiting which mould can be used on which machines will make production planning difficult and potentially introduce production bottlenecks. In paper A and B, the focus was on mitigating impact by dual sourcing raw material from multiple material vendors to improve quality. This could potentially also be achieved by limiting the use of specific vendor materials to specific moulds with an optimised mould card for that particular material. Again, this will increase the risk of introducing production bottlenecks. An alternative approach has been used at the industrial partner, where carefully designing the elements, moulds, and the moulding process has made the element quality somewhat robust against introduced disturbances. As production demands increase and new sustainable materials are introduced, it is becoming more and more challenging to maintain high throughput and quality using this approach. Therefore, it is becoming more and more relevant to introduce a monitoring or control approach where the process can be adjusted to compensate for uncontrollable disturbances.

4.1 Process monitoring approach

It is common practise to monitor a production process and various approaches can be applied. As described in the section 2.2, Statistical Process Control has been used in various industries for many years. The food and pharmaceutical industry is transitioning towards more process monitoring to improve the process and product understanding. Monitoring can generally be applied to the produced products and/or production processes. Often, it is easiest to introduce monitoring of product quality measures since it doesn't require data collection from the production equipment. Monitoring the product quality is very important since it directly measures what is essential for the customer/consumers. There is some limitation to only doing quality monitoring. Only a sub-sample of the produced products is measured, resulting in periods with no monitoring. Secondly, a sampling and measuring chain often takes time, causing a delay between an out-of-control situation and an opportunity to react/interact. Lastly, suppose only the quality is monitored. In that case, an out-of-control situation will require an extensive investigation to estimate the root causes, which primarily is the case if no process data is collected. When monitoring is applied to process variables, it is possible to achieve real-time monitoring without sub-sampling and delay. It is thereby possible to react immediately when an out-of-control situation occurs. Exclusively focusing on monitoring process variables, there is a risk that process variation indicates an out-of-control situation, which might have no impact on product quality. Without a solid understanding of the relationship between product quality and process variables, unnecessary time can be spent investigating process variations that don't impact product quality. It can always be discussed if not all process-related out-of-control situations are relevant since they could be linked to poor quality not captured by the implemented quality measures.

4.2 Selecting a monitoring/control strategy

Collecting the different approaches for monitoring and control, it is clear that multiple options are available. The following exemplification of a potential monitoring and control approach is centred around the manufacturing setup at the industrial partner. Based on prior work done at other companies, I assess that the approach will be generalisable to other productions with single unit manufacturing. The approach must be somewhat adjusted for application within fermentation and other batch processes.

The overall split between different approaches is linked to the frequency of the inspections, where one is cycle to cycle evaluation (often referred to as process monitoring), and the other is limited to evaluation of sub-samples (often used for quality inspections).

4.2.1 Sample by sample evaluation

The industrial partner has a two-level quality inspection setup. The most frequent inspection is performed two to three times a day (depending on the mould type) and consists of visual inspection of an element sample consisting of elements from one moulding cycle. Visual defects are the most dominant defect types why it makes sense to conduct this inspection daily. Data from the visual inspection is only collected when a critical deviation in element appearance is detected. This approach

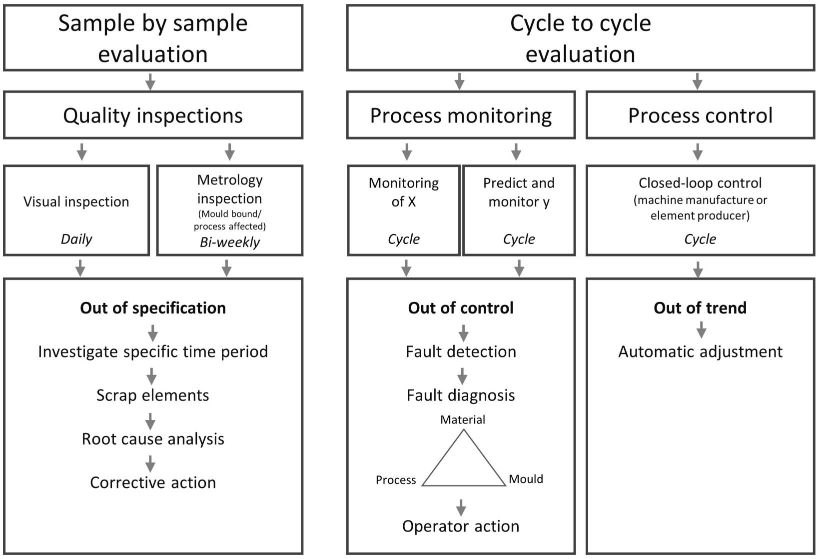


Figure 4.3. Illustrating different approaches for monitoring and control. The illustration relates to the injection moulding setup used at the industrial partner.

removes the possibility of introducing SPC for monitoring visual quality trends. It could be interesting to collect and monitor the severity of defects and the number of minor defects detected to enable quality monitoring (e.g. monitoring time between rare events). One moulding cycle is sampled for metrology inspection every third week, where critical geometric properties are measured (measurement data are stored digitally). The measures are evaluated against element specifications. If there are no deviations from the specifications, the production will continue. In cases where quality issues (metrology or visual quality) are detected (out of specification, in Figure 4.3), it has to be determined when the quality issues have started. This is done by collecting produced elements from the high-bay warehouse. During production, the produced elements are collected in plastic boxes (p-boxes) containing elements from a few moulding cycles up to several hundred cycles. The first p-box to be collected from the warehouse is the p-box containing elements from the last inspection. If defect elements are found in this p-box, the p-box from the previous inspection is collected, and so forth. If no quality defects were found in the first p-box, the p-box produced between the two inspected p-boxes is retrieved and inspected. This procedure is continued until the origin of the defect has been detected. The affected elements are scraped (or accepted despite), and a root cause analysis is conducted to identify the problem causing the production of defective elements. This is followed by corrective actions to eliminate the problem. There is no systematical implementation of statistical monitoring (SPC) of the element dimensions. Therefore, it is impossible to be proactive and react to quality shifts towards an out of specification situation. Paper F presented a multivariate approach using element metrology data to predict mould degradation. The same approach (with minor modifications) could be used to monitor element quality continuously. An example of this was presented at the one-day conference "Få processen i kontrol med Statistisk Proceskontrol (SPC)" at IDA 2019 [67]. The work from that presentation has been collected in appendix D.

4.2.2 Cycle to cycle evaluation

A different and often more attractive approach is to do cycle to cycle evaluation. Cycle to cycle evaluation in real-time makes it possible to be proactive and react to trends towards out-of-control situations. Moving from sample by sample evaluation to cycle to cycle evaluation makes it possible to potentially eliminate the need for conducting extensive re-sampling of p-boxes to identify the origin of the out-of-control situation. It should be noted that moving to the cycle to cycle level will require that all critical process variables are collected digitally. This can be a considerable effort, which I will elaborate more on in the next section. At cycle to cycle level, it can be considered to monitor the process, highlighting out-of-control situations, or go further and implement control feedback to the process. It would often be preferred to start with monitoring and guide operators towards what to adjust in the process. When this has proven to work, it can be considered to automatise some of the adjustments. As discussed previously, it can be a disadvantage only to monitor the process variable ($\mathbf{X}$) without knowing the relation between the process variables and the produced element quality ($\mathbf{y}$). Within injection moulding, it is also challenging to collect the state variables (reflecting the condition of the melt) that best reflect the relationship between the process and the quality produced. The results achieved in paper B, where the machine pressure profile is used as an indirect state variable, could be utilised to build a soft sensor for predicting element metrology. SPC or MSPC could then be introduced for monitoring the predicted quality. This concept has only been demonstrated to work for metrology responses, so additional work has to be conducted to explore the possibility of utilising the same method for visual element quality. Since visual element quality is closely related to the melt properties, it is expected that the machine profiles would improve the ability to predict visual quality. A reliable and well-tested soft sensor predicting element quality could be considered to introduce closed-loop control. This requires the identification of process variables that can be used to adjust the moulding process (impact the state variables) to compensate for the predicted changes in element quality. In paper B, potential control variables were identified as mould temperature and holding pressure. The approach has been tested further at the industrial partner, achieving promising results. The work has not been published and will therefore not be presented further.

4.2.3 Cost of inspection

In the introduction, it was emphasised that it is essential to have a clear objective and to have a cost/benefit approach when utilising data. The objective of introducing quality or process monitoring is to become proactive and react to shifts in quality or process variables, reducing or eliminating the production of elements out of specifications. This could be achieved by doing sample by sample or cycle to cycle evaluation as addressed above. The latter would be the most effective, considering the potential capability to reduce the amount of scrap produced, lost production time and resources used for root cause investigations. The question is then, what would be the most cost-effective approach? An abstract illustration of aspects to take into account for this evaluation can be seen in Figure 4.4. The cost of scrap and lost productivity related to process issues are expected to be reduced, going from no monitoring to some level of monitoring. A first reduction is envisioned as a sample by sample monitoring of quality is introduced, followed by a further reduction introducing cycle to cycle process monitoring. An additional cost reduction is expected to be achieved by introducing a closed-loop control. Another potential benefit is reducing human interaction as more monitoring and control is introduced. In this context cost of "human interaction" is defined as resources used for sample collection, sample evaluation/measuring, root cause investigations and implementation of corrective actions. The reduced cost of scraped products and lost productivity combined with the cost of human interaction drives the financial benefits.

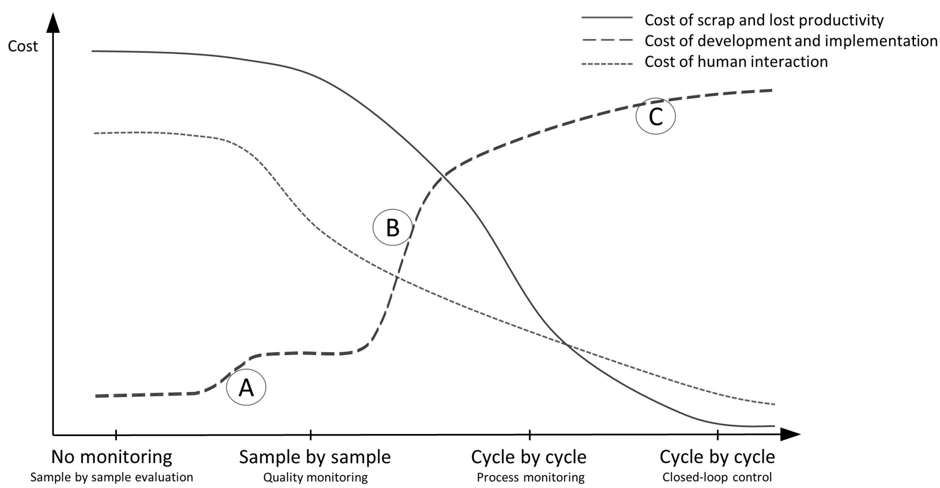


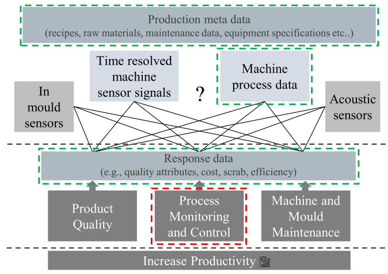
Figure 4.4. Abstract illustration of cost and benefits implementing process monitoring. The x-axis represent the level of monitoring and control. The y-axis represent an abstract cost measure.

Besides the benefits, the cost also has to be identified. The curve representing the cost of development and implementation, in Figure 4.4 has three cost points highlighted. Point A, represents the implementation cost of an IT solution for collecting data for the visual quality inspections and an IT solution for monitoring these data and the metrology measures already collected. The cost related to this is expected to be less than the potential benefits. When going from sample by sample to cycle to cycle monitoring, a significant investment has to be made (point B, in Figure 4.4). The investment is related to enabling data collection from the injection moulding machines. This includes installing needed communication hardware for the data collection and a data storage solution. Besides this, there might be a need for edge devices and monitors for data visualisations (e.g. control charts). With the data infrastructure in place, the next step would be designing and developing a working monitoring solution (classic SPC, MSPC or monitoring based on machine learning). This could potentially be a significant investment, and the associated benefits will most likely not cover the cost. This has been the case at the industrial partner, where additional data utilisation applications had to be identified before the data collection could be established. The investment going from process monitoring to closed-loop control (point C, in Figure 4.4) is expected to be small compared to moving to cycle to cycle monitoring. An essential cost to consider is the implementation of a data science pipeline that can be used to develop and deploy machine learning models. There will also be a considerable cost related to human resources within data engineering and data science.

Many considerations have to be made when moving into data utilisation for process and quality monitoring or control. It is always essential to make a small scale installation for proof concept. This will make it possible to estimate the potential benefits and costs of the complete implementation (installation, development, and maintenance of the systems). At the industrial partner, it has been necessary to identify and prove the benefits of multiple data utilisation applications before the decision was taken to implement the needed infrastructure to enable data collection and utilisation for the identified applications.

4.3 Industrial use case - Process monitoring within injection moulding

As part of the PhD project, there has been ongoing work related to the exploration of injection moulding process monitoring. The work is still ongoing, and it is expected that a publication will be submitted in the first half of 2022[1]. I have decided to include part of this work in the thesis since it is highly relevant for the overall reflections and conclusions made on data utilisation within injection moulding and data utilisation in general.



4.3.1 Abstract

The current work addresses process monitoring in an injection moulding setup consisting of multiple injection moulding machines. The process variables collected are highly correlated, which is handled using a MSPC-PCA based approach for monitoring. To overcome autocorrelation between consecutive moulding cycles, the process data is aggregated in sub-batches corresponding to boxes (p-boxes) used for collecting the produced elements. When nonconforming products are detected, p-boxes are investigated to determine the starting time for the quality problem. The objective of the investigation has been two-folded. One has been to investigate if the starting time of quality issues can be detected in the collected process data. Secondly, to investigate if MSPC-PCA based monitoring can be used with the collected process variables. Since faulty moulding cycles overlap with in-control cycles, it is found that using the available data, it is not possible to use MSPC-PCA monitoring. Additional process variables need to be included, and aggregation per p-box has to be reevaluated to improve this. A classification (PLS-DA) approach has been tested to detect quality issues. For models built for individual moulds, it has been possible to detect faulty p-boxes with a classification accuracy of only 50%. More faulty data needs to be included in the training model to improve the classification accuracy. To archive this, data needs to be combined from multiple moulds. This will introduce additional complexity caused by different moulding processes used for different elements and mould designs.

[1] The work is conducted together with;
Casper Solheim Bojer, PhD student, Department of Materials and Production, Aalborg University,
Till Böttjer, PhD student, Department of Electrical and Computer Engineering, Aarhus University,
Murat Kulahci, Department of Applied Mathematics and Computer Science, DTU Compute,
Alberto José Ferrer Riquelme, Universitat Politècnica de València, Departamento de Estadística e Investigación Operativa Aplicadas y Calidad (DEIOAC), Spain

4.3.2 Industrial context

The industrial partner uses standard injection moulding machines from Engel, Arburg and Battenfeld that all are European producers of injection moulding machines. The moulding setup consists of more than 3,000 moulding machines, as described in the introduction. It was initially planned that this work should include data from more than 1,000 moulding machines to investigate the potential benefits of utilising data across different moulds and moulding machines. After an initial data exploration, it became clear that the data collection protocol had not been correctly configured, resulting in essential process variables missing from the majority of the moulding machines. It has, therefore, only been possible to collect a sufficient (still not complete) dataset from 60 Engel Victory 330/60HL moulding machines, where the process data have been collected over six months. Besides process data, information on nonconformity cases from the Non-Conformity Management System (NCMS) has also been included in the data for the investigation. NCMS is used to record and document all nonconformity cases following a standard procedure. As mentioned in section 4.2.1 all elements are collected in p-boxes when produced. The quality sampling is therefore done on p-box level. Elements in a p-box can not be linked to a specific production time but only a production interval. Quality issues are labelled on p-box level, with a defect type and an action (e.g. exemption or scraped). Neither the severity of the quality problem or the number of detective elements per p-box is recorded, but can indirectly be deduced from the action label. When a quality problem is detected, extensive sampling and quality inspection procedure is initiated to determine the starting point of the quality problem (described in section 4.2.1). Since this requires considerable human resources, it is interesting to investigate if this starting point can be identified using existing process data from the moulding machines. The objective for this has been formulated as:

> **Research and industry objective 4 (RIO4)**
>
> Is it possible to detect the starting point of a nonconformity case utilising the available data And can this detection be included in a process monitoring setup?

Currently, machine process data is not used systematically for process monitoring. The control strategy used is based on adding fixed action limits on critical process variables. Limits are set for material cushion, dosing time and barrel temperature, and when limits are violated, auto scrap is activated to reject potential faulty products. To avoid machines running with long periods of auto scraping products, an andon[2] signal is raised when the machine has been scraping 20 consecutive moulding cycles. To capture machines randomly scrapping a high number of moulding cycles, the machine productivity is evaluated hourly (actual production over theo-

[2] Andon is a LEAN system which notifies operators, maintenance staff, and other workers of a quality or process problem

retical production). This setup has been effective in ensuring high product quality and equipment productivity. As production complexity increases (new sustainable materials, material dual sourcing, etc.), it is uncertain if the system will be sufficient. Hence it is of interest to investigate if a monitoring setup can be implemented to mitigate the negative impact of increased complexity.

4.3.3 Introduction

The overall goal of process monitoring is to detect when the process is moving out of conditions defined as ”normal operation” or ”in control” state. One approach for process monitoring is Statistical Process Control (SPC). Implementation of SPC is a two-step process where the first step is to describe the ”in control” state (phase I), and the second step is the evaluation of new observations against the ”in-control” state (phase II). A classic approach for describing phase I is to collect data for a given period where the process has been in control (only containing common causes variation) and calculate the mean and standard deviation (when assuming the variable to be normally distributed). The in-control region is often defined as three standard deviations around the mean (Montgomery, 2012 [68]). In phase II (monitoring phase), it is evaluated if new observations are belonging to the same distribution (in-control), which often is visualised in a control chart. SPC is mainly useful for single variables or uncorrelated sets of multiple variables. When data from a process consist of multiple highly correlated variables an alternative approach must be used (still using the same approach with phases I and II). Multivariate SPC (MSPC) has been introduced to handle situations with multiple process variables simultaneously. However the classic approach of MSPC can not handle many and correlated process variables (MacGregor and Kourti, 1995 [69]). Principle Component Analysis (PCA) is therefore introduced in combination with MSPC, to transform the raw correlated variables into a reduced subspace of uncorrelated latent variables [69]. This is often referred to as MSPC-PCA. The latent variables are selected so that the first explains the most of the variation in the original data. The second latent variable is selected so that it is orthogonal to the first and in the direction second largest variation (and so on for the following latent variables). The in-control state is then defined as a multivariate space consisting of a selected number of latent variables, where the number of included latent variables could be selected so that a total of 80% of the original variation is explained. In phase II, new multivariate observations are projected onto the multivariate space defined by the selected latent variables. From this, the latent projection, score values and residuals are calculated. The process is monitored using two control charts. The first control chart is used to monitor the within model variation, using Hotelling's T^2 statistics. The second control chart is used to monitor the distance to the model, by calculating the squared prediction error (SPE or the Q-statistics). This approach can be applied individually to both process variables $\mathbf{X}$ and quality variables $\mathbf{Y}$, when multiple quality variables describe the product quality (see example in appendix D). When both process variables and quality measures ($\mathbf{y}$ or $\mathbf{Y}$) exist for the same sample, PLS-MSPC can be used, where the decomposition of $\mathbf{X}$ is done with respect to the variation in $\mathbf{y}$. This then relates the phase II monitoring of $\mathbf{X}$ to the quality variables. The introduction of PCA and PLS makes it possible to handle and monitor correlated variables under the assumption that the process is stationary and no autocorrelation (time-dependent correlation) is present in the data. De Ketelaere et al. (2015) [70] present an overview of different approaches to handle autocorrelation (Dynamic-PCA) and non-stationary (Recursive-PCA and moving-window PCA) data individually and concludes that no methods effectively handle the combination of both autocorrelation and non-stationary data.

The work presented in appendix D and the initial process monitoring work present here both use MSPC-PCA, hence this will be explained in the following section.

4.3.4 Methodology, MSPC-PCA

Principle component analysis can be used to handle correlation between variables in the data to be monitored by projecting the original variable into a latent and uncorrelated space. This can be expressed as:

$$\mathbf{X} = \mathbf{Z}\mathbf{P}^T, \tag{4.1}$$

where, $\mathbf{X}$ is the original $n \times m$ data matrix, $\mathbf{Z}$ is the $n \times m$ score matrix, $\mathbf{P}$ is the $m \times m$ loading matrix. The relevant part of the variation in $\mathbf{X}$ will be contained in the first principal components (PCs), and the noise contributions in $\mathbf{X}$ will be contained in the remaining PCs. The number of, k, relevant PCs to keep in the model can be found by the amount of explained variance or cross-validation.

The PCA model containing k principal components (PCs) for an $n \times m$ input matrix $\mathbf{X}$, can be written as:

$$\mathbf{X} = \mathbf{Z}_k\mathbf{P}_k^T + \mathbf{E}, \tag{4.2}$$

where, k is the number of selected PCs, $\mathbf{Z}$ is the $n \times k$ score matrix, $\mathbf{P}$ is the $k \times m$ loading matrix and $\mathbf{E}$ is the $n \times m$ residual matrix. In a monitoring context (phase II), new observations ($1 \times m$) are then projected unto the defined model space, and the Hotelling T^2 distance to the model, for the new observation i, can be calculated as:

$$T_i^2 = (z_{k,i} - \bar{z}_k)^T \mathbf{S}^{-1} (z_{k,i} - \bar{z}_k) \tag{4.3}$$

were z is the sample score vector, $\bar{z}$ is the score mean vector from phase I and $\mathbf{S}$ the sample covariance matrix also from phase I. The evaluation of the calculated T^2 value is done against a control limit (for phase II) calculated as:

$$UCL_{T^2} = \frac{k(m+1)(m-1)}{m^2 - mk} F_{\alpha,k,m-k}, \tag{4.4}$$

were $F_{\alpha,k,m-k}$ is the upper $100\alpha\%$ critical point of the F distribution with k and $m-k$ degrees of freedom [69]. Monitoring the within model variation using T^2 will capture if extreme variation occurs in the variation captured by the PCs contained in the model. If variation occurs in the latent variables dominating the residual, this will not be detected. Hence monitoring the residual also has to be implemented. The residuals are monitored using squared prediction error (SPE) or the Q-statistics [71]:

$$Q_i = \tilde{x}_i \tilde{x}_i^T = x_i(\mathbf{I} - \mathbf{P}\mathbf{P}^T)x_i^T \tag{4.5}$$

were $\mathbf{I}$ is the identity matrix, $\mathbf{P}$ the loading matrix, x_i is the i^{th} observation vector and $\tilde{x}_i$ is its projection into the residual subspace. The control limits for Q can be approximated as:

$$UCL_Q = \theta_1 \left[\frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0 - 1)}{\theta_1^2} \right]^{1/h_0} \tag{4.6}$$

were z_α is the standardised normal distribution value with a significance level of α, and with:

$$\theta_i = \sum_{i=k+1}^{m} \lambda_i^j \tag{4.7}$$

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_{12}^2} \tag{4.8}$$

and λ_j is the j^{th} eigenvalue of the covariance matrix $\mathbf{S}$. A high T^2 value indicates that a sample is extreme but that the variation pattern is well-accounted for by the PCA model. A high Q value indicates that the sample represents a variation pattern that is not well accounted for by the PCA model.

Since PCA is scale-dependent, it is crucial to scale the input data if the variables are on different scales (e.g. if some variables are in millimetres and others in kilometres). A typical scaling is to mean centre and scale all variables to unit variance. In phase II, it is crucial to use the mean and variance computed in phase I for the scaling.

4.3.5 Data structure

The available process data is presented in table 4.1 where the variables are compared with process variables used in selected reference studies. From this comparison, it can be seen that mould temperature is used in all three reference studies. This supports the findings of paper A and B, where it was shown that the variations in mould temperature (more extreme variation than expected during production) is important for element dimensions. The cooling water temperature used in the moulds at the industrial partner is kept constant and controlled central. Nevertheless, it can not be guaranteed that the resulting mould temperature is constant, and without collecting data, the impact can not be accounted for.

	Industrial partner	**Thyregod 2001 [63]**	**Hopmann 2021 [64]**	**Kazmer 2003 [66]**
	Cushion	Melt cushion		
		Screw stroke		Shot size
Position/Stroke		Change over position	Switchover point	
			Injection velocity	Injection velocity
			Screw RPM	Screw RPM
	Clamp force			
	Back pressure	Back pressure		
Pressure	Holding pressure	Holding pressure	Packing pressure	Packing pressure
	Max injection pressure	Injection pressure		Max injection pressure
		Pressure at point		
		of change over		
	Barrel temperature	Barrel temperature	Barrel temperature	Barrel temperature
	Feed fhroat temp.			
		Mold temperature	Mould temperature	Mould temperature
Temperature		Heater temp 1		
		Heater temp 2		
		Oil temp		
		Crosstemperature		
			Packing time	Packing time
	Plastification time	Charging time		
Time		Dwell time		
	Injection time	Injection time		
	Cooling time	Cooling time	Cooling time	Cooling time
	Cycle time	Cycle time		

Table 4.1. Comparison of collected process variables at the industrial partner and reference studies. The variables are grouped by variable type, and comparable variables are aligned row-vice (despite different naming).

Both Hopmann [64] and Kazmer [66] include the injection velocity and screw RPM. Both the industrial partner and Thyregod [63] have included the injection time, which to some extent will capture the same variations as the injection velocity (since it is intended to keep the injection velocity constant). The screw RPM is an important parameter when defining the mould card, but is kept constant during production, so not expected to exhibit significant variation. Going forward, variables that could be beneficial to collect at the industrial partner would be switchover point, pressure at switchover, mould temperature and packing time.

The dynamic changes in a moulding process occur relatively slow resulting in autocorrelation between cycles. A simple approach to mitigate autocorrelation is by sub-sampling the collected data or aggregating the data in time. In the specific case where quality evaluation is linked to p-boxes, it makes sense to aggregate the collected data per p-box. Aggregation can be calculated in different ways, where the most common is to calculate the average per group. This will only reveal the central tendency, which might not be sufficient. A measure of variation within each group can be included as standard deviation, variance, range or quantile range (e.g. 1% lower and 99% upper quantile). A measure of extreme values can be included as the number of observations outside a quantile range. In the initial work, the aggregation of process data per p-box includes the mean and 1% lower and 99% upper quantile.

Besides autocorrelation, it is also essential to consider the correlation between the process variables. A correlation plot of selected process variables can be seen in Figure 4.5 (mean values per p-box). From the correlation plot, is it evident that correlation exists among the process variables. The six temperature measures on the barrel create a light coloured square in the centre of the plot indicating some positive correlation between the measures (also expected). There are correlations among other process variables as well, where one is a negative correlation between cushion and plastification time. When the material cushion increases (less material is injected into the mould), less material has to be dosed in the next dosing phase resulting in a reduced dosing time. MSPC-PCA is a reasonable option for process monitoring with the seen correlation between the process variables.

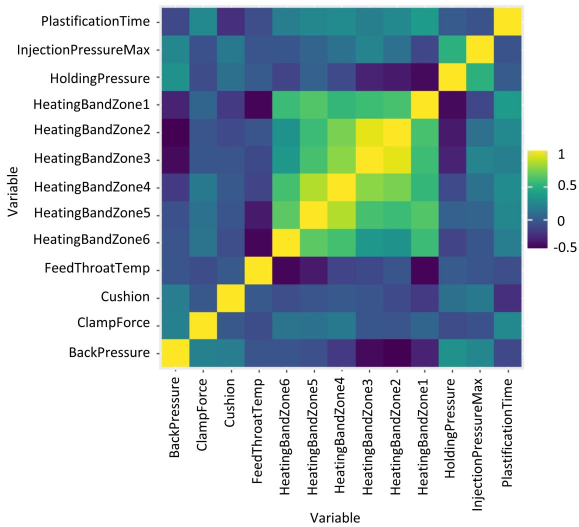


Figure 4.5. Correlation plot with mean values per p-box

The total data collected contains approximately 60 moulds running for six months. In the dataset, there are 32 moulds with quality issues registered in NCMS. In all cases, only one type of quality issue has been registered and only within a single production order. Data from 60 moulding machines, collected over six months, would in many cases be sufficient to do an initial proof of concept. With the production setup used at the industrial partner (many production orders and production problems occuring rarely) this has been a challenge. First, let us list the possible approaches to utilise the available data. One approach for utilising the process data for monitoring would be to introduce MSPC-PCA neglecting the quality data. In this case, a phase I study has to be defined and data reflecting "normal" operation included from a period without any quality problems. An out-of-control situation in phase II should then preferably include any upcoming quality problems. This will be tested in the next section together with an alternative approach where the process data is used in classification to predict if a p-box would contain a quality problem. With the implementation of a classification-based approach monitoring could be included by monitoring the predicted probability of a p-box containing detective products.

4.3.6 Data exploration and MSPC-PCA

The investigation presented in the following is conducted on data from one mould running with five different production orders over five months (in total, including data from 2750 p-boxes). In the third production order, quality issues (flash, detected on elements) are reported for 46 consecutive p-boxes, corresponding to an overall error rate of 1.7%.

An initial exploration using PCA can be seen in Figure 4.6. The left version of the score plot (PC1 vs PC2) is coloured according to production order, and the right version is coloured according to quality evaluation on the p-boxes. There seems to be a shift in the data correlation between the different production orders, especially from the first (red points), over the second (light green points) to the last three production orders (phase II). It seems that the faulty p-boxes are grouped close together with the remaining p-boxes from the same production order (11755522).

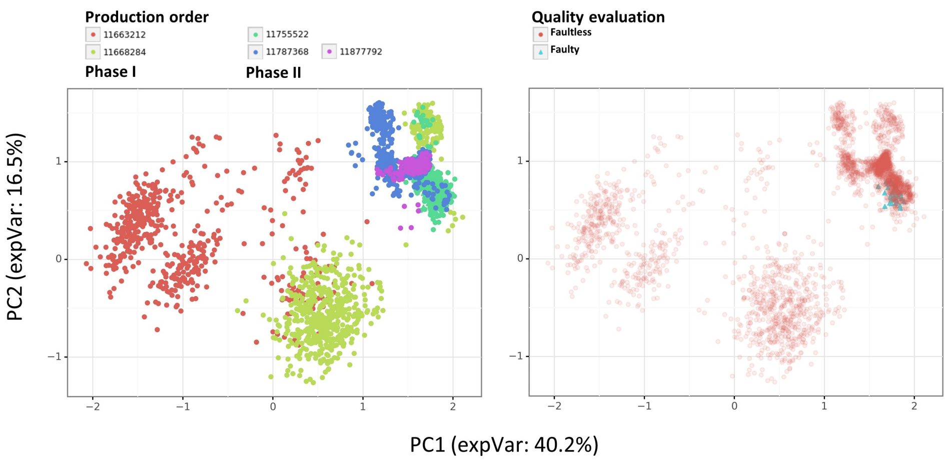


Figure 4.6. Score plot representing the first two latent variables, together explaining 56% of the variation in the process data. Left plot coloured according to production order and right plot according to quality evaluation.

The score values for the first four principal components are plotted in Figure 4.7 as function of time. It becomes clear that faulty p-boxes are well separated within the given production order when looking at PC3 and PC4. At the same time, it is also clear that the variation change related to the faulty class is small compared to the variation seen in the entire period. To investigate what variables impact the faulty class, it would be necessary to make a separate PCA model for production order 11755522 (not presented here).

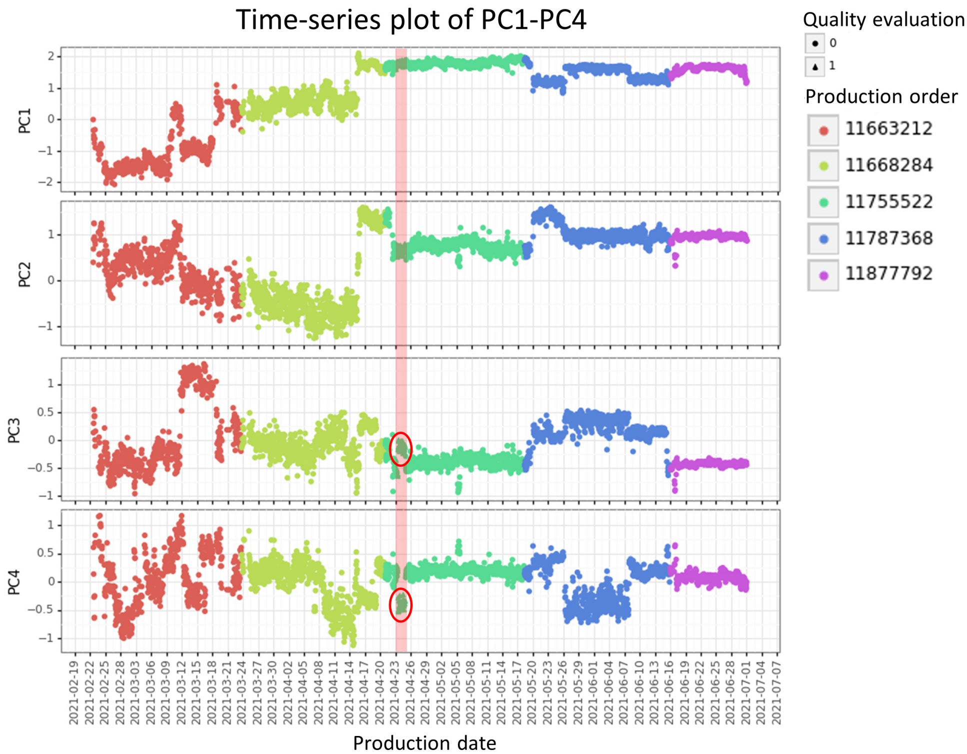


Figure 4.7. Time-series plot of the first four principal components explaining 40%, 16%, 8% and 6% of the variation in the process variables. The period with the production of faulty elements is highlighted with a red area and a red ellipsis.

Since the scope is to investigate MSPC-PCA, the data have to be split into a phase I period followed by a phase II. Since the quality issue occurred at the start of the third production order, the first two production orders will be used as phase I and the remaining three production orders as phase II. A PCA model is therefore calculated on phase I data (including six PC, explaining 81% of the total variation). The phase II data is then projected onto this model to calculate the scores for phase II (and related SPE and T^2 statistics).

A collected view of the variation across phase I and II can be seen from the T^2 and SPE plot in Figure 4.8. From both the T^2 and SPE plot, it can be seen that phase I seems to be relatively stable without any extreme outliers above the control limit and can therefore be assumed to be in control. T^2 has a significant shift from phase I to phase II, indicating that the within model correlation changes. This could be seen as a shift to a new operating regime. Investigating the production order history, did not explain this significant shift. The relative stable SPE indicates that the residual remains fairly constant going into phase II. The p-boxes in the red area, showing the period with faulty elements, are behaving as the neighbouring data point, indicating the MSPC-PCA approach implemented on data from this specific mould does not capture the faulty quality. Similar results have been achieved using the same approach on the remaining 31 moulds with quality issues.

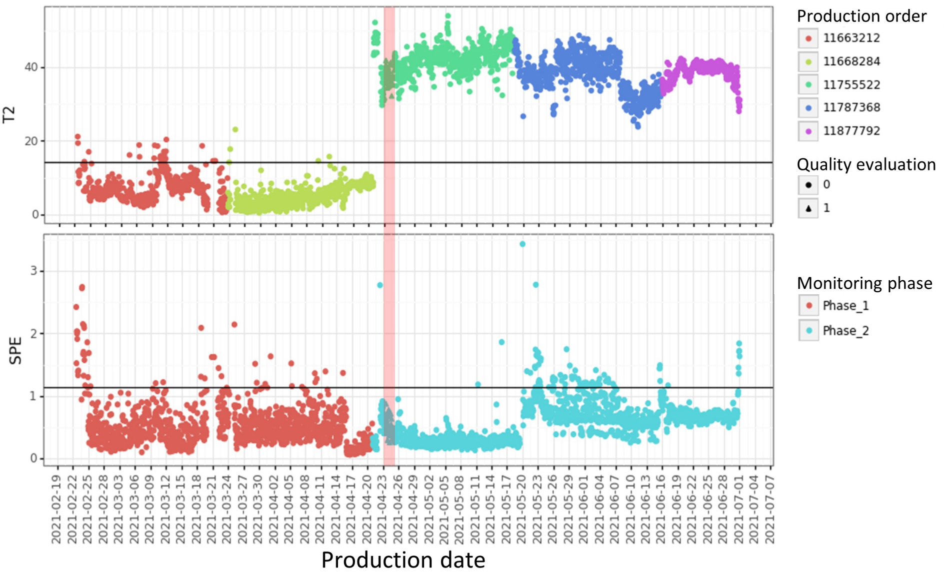


Figure 4.8. Time-series plot of SPE and T^2 with phase II control limits. The T^2 is coloured according to production order, and the SPE is coloured according to the monitoring phase. The period with production of faulty elements is highlighted with a red area.

4.3.7 Classification of p-boxes with faulty elements

Making a classification model to predict if a p-box contains faulty elements has implications using the available data. In general, when developing a machine learning model, it is good practice to test the model on an independent test set, which is not possible in the current case. An alternative approach for an initial model evaluation (validation) is to do cross-validation; however, it is also preferable to have independent data. In the current case, it would be fair to consider all the p-boxes with faulty elements as one sample containing dependent sub-samples. One approach to overcome this would be to combine data from different moulds, which will allow for an independent test set. Combining process data from different moulds require that the data are scaled since the moulds are running with different moulding processes (individual mould cards). This has initially been achieved by scaling the process data according to the mould card values (when available) or alternatively by subtracting the average values for the individual moulds. This approach has been tested for moulds with similar fault type (six moulds causing flashes on elements), where a classifier[3] was trained using a leave-one-mould out validation, where five moulds were used for training the classifier and the sixth mould as an independent test set. A classification error close to 100% was achieved (all p-boxes classified as faultless), indicating that a combined model requires a different approach, more training cases, or additional process variables. The prediction achieved using production order 11755522 as a test set can be seen in the right side of Figure 4.9.

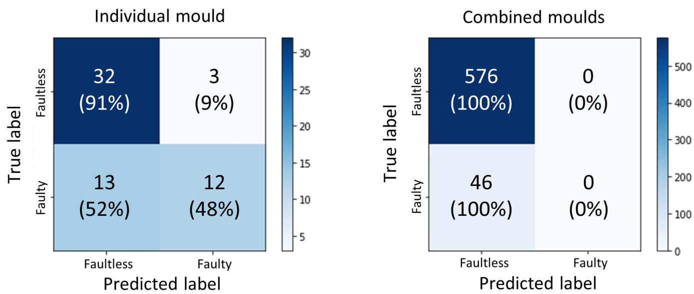


Figure 4.9. The left confusion matrix represents accuracy for classifying p-boxes with flash in production order 11755522, using data from the same production order for training the classifier (PLA-DA). The results in the right confusion matrix are achieved using training data from five other moulds (with production orders containing elements with flash detects) and using all p-boxes in production order 11755522 as a test set.

When creating a classification model for each mould individually, all faultless p-boxes are classified correctly, whereas only around 50% of the faulty p-boxes are classified correctly (see an example for production order 11755522 in the left side of Figure 4.9). This is somewhat surprising since the faulty p-boxes are from the same population (same moulding period). One factor that could be impacting the results is that the fraction of defective elements in the faulty p-boxes is unknown. Suppose the level of defective elements is low (only a few moulding cycles causing defective elements). In that case, the aggregation of data across the p-box might be diluting the issue causing the faulty elements, resulting in misclassification. It is not expected that including a few additional production orders (same mould) with the same faulty types would improve the results due to the more or less random shifts seen within and between productions orders (Figure 4.7 and Figure 4.8). It is therefore expected that

[3]tested using PLS-DA, SVM and XGBoost achieving similar results

additional process variables, or an alternative data aggregation is needed to improve the classification accuracy.

4.3.8 Discussion and reflections

The work presented here is initial, and more exploration has to be made before conclusions can be made. The findings achieved so far will be summarised and discussed in the following.

Unsupervised monitoring (MSPC-PCA)

In the majority of the investigated cases, faulty moulding cycles overlap with in-control cycles (both in time-series of scores, control charts of T^2 and SPE). This indicates that the correlation structure in the used process data doesn't change in an actual out-of-control situation. At the same time, it is seen that out-of-control is flagged when there are no apparent quality issues. This is an unwanted situation and makes it difficult to use the concept for process monitoring. The general assumption in SPC is that phase I is performed on a period where the process is in control, and this is confirmed as part of the analysis in phase I. To have a phase I estimate of the future true mean and variance (used as a reference for evaluating future process variations), phase I should contain all in-control behaviour expected in the future. If this is not the case, a new in-control state (or regime) might wrongly be declared out-of-control resulting in an increased false-negative rate. Comparing the results from the 60 moulds investigated, we see different operating regimes for the individual moulds, where the faulty p-boxes seem to be randomly distributed among them. It could be considered to investigate the use of Recursive-PCA and moving-window PCA to overcome the challenges with non-stationary (shifts between operating regimes) as proposed by De Ketelaere et al. (2015) [70].

Supervised monitoring - fault classification

To enable a classification model to classify expected error types correctly, the training data must contain examples of these error types. This can be hard to obtain for the individual moulds as errors occur rarely. This means that data have to be combined across moulds for a supervised approach to work. It then becomes a requirement that the correlation structure in the process data related to the different error types has to be comparable across moulds. With the initial results, it is clear that combining data across mould has to be optimised (see reflections on this in the next section).

Individual or combined models

At the industrial partner, it would not be possible to include the needed fault types early enough to obtain the potential benefits of having real-time classification on individual moulds. If supervised monitoring, as described above, is to be implemented, it would be necessary to combine data from multiple moulds. The data used in the initial work has been combined based on the type of fault seen for a given mould. The classification results are expected to be improved by grouping the moulds in meaningful groups (e.g. by mould, machine and material type), thereby reducing the data complexity. This can not be tested with the current data set with only a few moulds representing the different defect types.

Another limiting aspect of using moulds individually is that this requires a dedicated model for all active moulds. At the industrial partner, approximately 10,000 active injection moulds are producing elements. The resources necessary to build and maintain data for that many moulds will outweigh any potential benefits. This will be regardless of implementing a supervised or unsupervised approach.

In paper C, we presented an approach (using acoustic recordings as input data) where an initial model was trained across moulds and adjusted at the start of the production order. A similar approach (model transfer) could be explored using process data from the moulding machine.

Data from individual cycles or aggregated data

Using the p-box as aggregation level in the initial work was chosen since the quality evaluation is done on p-box level. It is more the exception than the rule that elements from all cycles in a p-box are faulty. This being the case, it can be augured that aggregating the process data per p-box is a wrong approach or that new aggregation features have to be included to reflect the within p-box variation better. It might also be possible to improve the outcome of using this aggregation approach by enriching the information level of the quality evaluation. Instead of having a label stating that a p-box is faulty because too many elements have too severe defects, it could be recorded how many defective elements are present in the p-box and to what level they are defective (evaluated on a numeric scale). This could be utilised to investigate the variation in the process variables collected within a p-box and thereby maybe filter out the most likely cycles causing the faulty elements or to generate alternative aggregation features. If faulty cycles within the p-box could be identified, these cycles could then be used in building a classification model. Alternatively, the problem could be changed into a regression problem, trying to predict the quality level in the p-box and then monitor the predicted quality level.

Moving entirely away from aggregating the data before using it for supervised or unsupervised monitoring creates different challenges. In supervised learning cases, the challenges become linking the quality evaluation to the specific moulding cycle. A starting point could be to register (set a flag in the data table containing the collected process data) when samples are collected for quality inspections and at the same time, quantify and digitally register the quality measures. It would then be possible to collect a data set consisting of quality evaluations and process data on individual moulding cycles. In an application using consecutive moulding cycles in a training context, it also becomes challenging to handle the high level of autocorrelation between cycles and approaches as Dynamic-PCA might have to be considered [70].

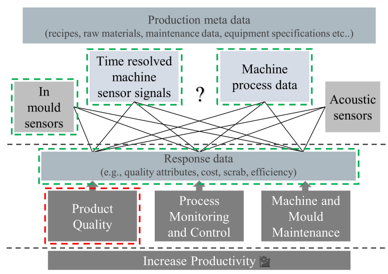
General reflections

With the initial results obtained using the readily available process data collected, it is questionable if these data contain the needed information related to moulding dynamics affecting the produced element quality and, therefore, useful for developing a monitoring setup. The work in paper A and B demonstrated that underlying pressure profiles from the moulding machines contain more relevant information for describing variation in element dimensions than readily available process data. The explanation is that the pressure profiles better capture the dynamics of the moulding process that reflect the material variation impacting the element dimension. Since some of the same material variations also affect the visual quality attributes, it would be expected that the prediction of visual defects could be improved using the machine pressure profiles. Therefore, it is proposed that the machine profiles are collected for a number of production moulding machines and tested in a process monitoring context. The moulding machines at the industrial partner do not support the collection of machine profiles and therefore need updating to enable this. This comes with a relatively high cost, which has to be considered when evaluating the monitoring solution utilising this data type.

When comparing the findings from the initial investigation with solutions and results achieved in scientific publications, one apparent difference is that published work is often based on a designed study (as we have done in paper A, B and C). When a designed study is used, it is possible to eliminate all the factors that are making real-life applications challenging. Besides eliminating disturbance factors in a designed study, it is also possible to introduce additional systematical variation that might never be present in real-life. These two factors make it easier to achieve clear results comparing different methods and modelling concepts. At the same time, it makes it difficult to use the achieved results in real-life applications directly. This said, designed studies are crucial for learning causal relationships, which can often be exploited in a production setup.

CHAPTER 5

Paper A, An investigation of the utilisation of different data sources in manufacturing with application in injection moulding



Starting a data utilisation journey, it is essential to ensure a well-defined business objective to guide the focus and selection of data utilisation applications. We propose a systematical framework for linking the business objective, key drivers and applications to data collection and utilisation. The first version of the framework was presented in paper A and have since expanded to the framework described in the Introduction. With a defined business objective and identification of potential applications, the question often becomes, what data to collect and utilise to achieve the desired business value? Different data sources can be available from a manufacturing setup, and it might not be obvious what data to collect and use for the specific application and at the same time be used to support other applications. This has in RQ1 been formulated for a specific case within injection moulding.

Research question 1 (RQ1)

How to best utilise data from injection moulding machines to increase the productivity of producing injection moulded elements?

5.1 Summary

The work in paper A is linked to RQ1 and presents a framework for evaluating the use of different data sources for achieving a specific business objective. The main scope of this work is to investigate the level of information in the most commonly used data sources within injection moulding that can be linked to variation in element quality. The three data types included in the work is readily available process data, underlying continuous process readings used internally in the injection moulding machine and mould pressure sensors which are the most commonly used data source in scientific literature. The proposed framework has been demonstrated to effectively evaluate the cost/benefit of utilising various data sources for the same objective. Also, the paper presents an evaluation of three levels of data from injection moulding and concludes that time-resolved pressure profiles are the best options when modelling element dimensions (soft-sensor approach).

Achieved results

- Understanding the information contained in different data sources from injection moulding (readily available process data, time-resolved machine pressure profiles and cavity pressure profiles). It was confirmed that readily available process data do not contain state variables that reflect material property variations that impact element quality. It was also demonstrated that machine pressure profiles contain the same level of material related information as the cavity pressure profiles. Since the utilisation cost is significantly lower for machines profiles (compared to cavity profiles), it makes the machine profiles the optimal data source to use in large scale injection moulding setups.
- Based on the available data, it was proposed to utilise machine pressure profiles in a two-step approach. First, the machine profiles were used to classify the material used. The classification result was then combined with readily available process data to predict element dimensions.

Contribution

- Framework for comparing information content in different data sources and benchmarking the cost and value creation.
- Highlighting the possibility to use underlying machine/equipment data for process optimisation by extracting the essential process dynamics.
- Proven that variation in raw material can be captured by the change in injection and dosing pressure profiles within injection moulding.

5.2 Paper A

Paper A

An investigation of the utilisation of different data sources in manufacturing with application in Injection Moulding

G. Ø. Rønsch[a*], M. Kulahci[a] and M. Dybdahl[b]

[a]DTU Compute, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby; [b]Department of Business Development and Technology, Aarhus University, Denmark.

ARTICLE HISTORY
Compiled December 7, 2021

ABSTRACT
This work focuses on the effective utilisation of varying data sources in injection moulding for process improvement through a close collaboration with an industrial partner. The aim is to improve productivity in an injection moulding process consisting of more than 100 injection moulding machines. It has been identified that predicting quality through Machine Process Data is the key to increase productivity by reducing scrap. The scope of this work is to investigate whether a sufficient prediction accuracy (less than 10% of the specification spread) can be achieved by using readily available Machine Process Data or additional sensor signals obtained at a higher cost are needed. The latter comprises Machine Profile and Cavity Profile Data. One of the conclusions is that the available Machine Process Data does not capture the variation in the raw material that impacts element quality and therefore fails to meet the required prediction accuracy. Utilising Machine Profiles or Cavity Profiles have shown similar results in reducing the prediction error. Since the cost of implementing cavity sensors in the entire production is higher than utilising the Machine Profiles, further exploration around improving the utilisation of Machine Profile Data in a setting where process variation and labelled data are limited is proposed.

KEYWORDS
Industry 4.0, Data Evaluation, Machine Learning, Deep Learning, Predictive Quality, Injection Moulding

* CONTACT G. Ø. Rønsch. E-mail: geor@dtu.dk

1. Introduction

Process equipment is typically furnished with sensors to collect Machine Process Data that are deemed to be relevant for process safety, performance and maintenance. There is an increasing push towards more effective utilisation of this data for obtaining better insight towards understanding, improving and optimising the process. In this work, we primarily focus on the injection moulding process, but we expect parallels to be easily drawn to other types of production equipment. The characteristics of Machine Process Data are very much dependent on the type of equipment and process as for example in the case of continuous or batch processes. Injection moulding (see description in section 2.1) is a batch process. However, as opposed to other batch processes - like fermentation that can take hours due to its short duration (which is often seconds), it is hard to make adjustments to the process during an injection moulding cycle. The different cycles are therefore often seen as discrete events, where information from one cycle can lead to adjustments in the next cycle. The available data in injection moulding is therefore aggregated data from the individual cycles in the form of selected features such as maximum values (e.g., of pressure or temperature), values at a specific time in the cycle (measured on time or distance), time measures, distances, integrals or other.

To obtain this aggregated data, the Machine Profiles for some essential variables such as pressure and screw position is collected and stored. While aggregated data is readily available, the Machine Process Data in its raw form is only made available to the machine operator as a plot on the machine control screen (plotted as the cycle progresses) and in many cases discarded after the cycle is complete. Only in certain cases, the injection moulding equipment is modified/designed to extract (e.g. via OPC UA) the continuous Machine Profiles in its raw form and store it to an external system. The structure and use of the collected data are described in more details in section 2).

It has been a growing trend since the rise of Industry 4.0 to mount pressure and temperature sensors in the cavities of injections moulds to capture the dynamics of the moulding process. Ageyeva et al. (2019) presents an extensive overview of avialible mould sensors and Gao et al. (2014) has demonstrated the correlation between cavity sensor signals and element quality. It has been speculated that additional information can be captured, and that the information can be used for process monitoring (Mao et al. 2018). The use of cavity sensors is a viable solution for e.g., mould concept development (evaluation of balance in the mould) and material test, but can become costly if multiple injection moulds and injection machines have to be equipped with these sensors.

In the quest of utilising data within Industry 4.0, it is paramount that a proper mapping between the purpose and the potential data sources is established. Too often a task/project is started with the question 'How can we utilise this data?' and not 'What type of data can support the given challenge?' (Kulahci et al. 2020). An example of such a data mapping is illustrated in Figure 1. The underlying aim in this mapping is the challenge to increase productivity. Once this mapping is established, the next step is identification of the means with which the productivity can be improved. In the investigated case product quality, process monitoring, and maintenance have been identified as three areas that can support efforts towards improving productivity:

- The loss of productivity related to product quality occurs when a quality error is detected during quality inspection before the production order is released. When an error is detected, the produced products have to be checked and potentially scraped. In that sense, real time prediction of quality could give an early indication that the product quality is deteriorating and thereby reduce scrap related loss of productivity.
- The current monitoring setup at the industrial partner is based on productivity loss. Hard limits on for example material cushion and barrel temperature are used to trigger auto rejection and the productivity per hour is then used to identify machines that need attention. By introducing Multivariate Statistical Process Control (MSPC) techniques, it will be possible to achieve a more effective process monitoring and more importantly a tool for supporting fault detection and diagnosis (Sanchez-Fernandez et al. 2018). This will result in increased productivity by early detection and reduced time for troubleshooting.
- At the industrial partner all machine and mould maintenance is based on LEAN TPM (Total Productive Maintenance) concept. Performing time-scheduled maintenance can lead to over-maintained equipment (loss of productivity) and delay in needed maintenance that can cause decline in quality and/or preventable stoppage. By switching to predictive maintenance (optimally in combination with TPM), real time schemes can be established to perform maintenance in due time.

The next step would have been to evaluate which of the identified productivity gaps had the highest potential and then focus on that. However, for illustration purposes, in this work the product quality is selected as the target for improvement as it is the easiest to demonstrate within a short time frame.

After selecting the focus area, the task is to identify what type of data is expected to be relevant to reduce the identified productivity gap and what is actually available. In this example the standard machine data (Process Data) is identified together with the time resolved machine sensor signals (Machine Profiles) and the in-mould cavity sensors (Cavity Profiles) as additional sensors. The Machine Process Data is obtained at the lowest cost as it is available in almost all moulding machines and can easily be collected. The structure of the data is also simple. This makes the effort required for data handling and analysis relatively low. The Machine Profiles come at a higher cost as they in most cases are not extractable from the injection moulding machine. Moreover, it is more cumbersome to extract the contained information from Machine Profiles as the data structure is complicated to handle and analyse. Adding additional sensors, in this case cavity sensors, is the solution with the highest cost, since it demands investment in new equipment, customised integration in the mould and additional data collection. The cost of extracting the information in the data collected from the cavity sensors is on the same level as for the Machine Profiles since the data from both sources have the same structure.

When the mapping is completed, a cost-benefit analysis has to be performed to identify what the best trade-off is between the gain in productivity and the cost of data collection, handling and analysis. Unfortunately, this part has in many cases been neglected with the focus being firmly on getting the best predictive power irrespective of the extra cost for sensors, data collection, analysis and not least the continuous maintenance of the whole data utilisation chain. Zhang and Alexander (2008) concludes that cavity signals can be used as process fingerprints and that these can be used for process monitoring, This might be a valid conclusion but it is not considered if the same results could be achived using less expensive data source, such as machine profiles. Farahani et al. (2019) do a relevant evaluation of different data scources againt the cost of computation (by reducing data complexity), but ignores the cost of data collection and therefore recommend the use of cavity sensors to enable significant prediction opportunities. Zhang et al. (2016) presents quality prediction results obtained using in-built machine sensor, but do not include cavity sensors in the study and therefore lack the comparesion against results obtained using cavity sensors.

In the current work, we will investigate whether additional prediction power (for online prediction of quality) can be achieved by utilising Machine or Cavity Profiles compared to what can be achieved by using the available aggregated Machine Process Data. We also evaluate if the cost of the additional prediction power can be justified by the expected gain in productivity.

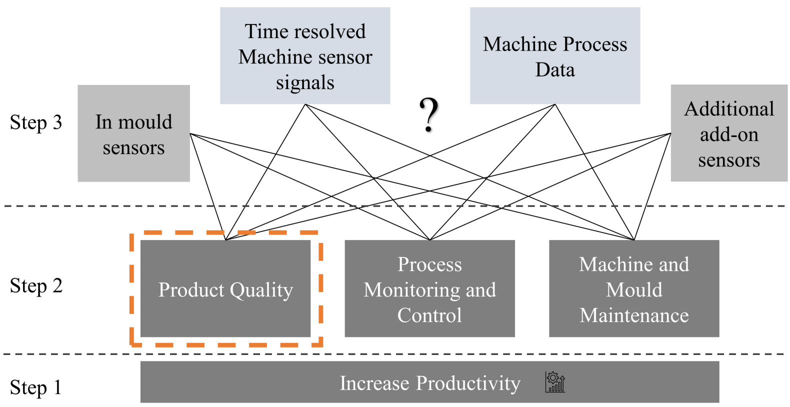


Figure 1. Overview of data utilisation within injection moulding, where the focus is to improve productivity. From the connections it can been seen that it could be different data sources that are needed for different aspects related to improving productivity.

2. Methodology

This section will cover an introduction to injection moulding, the data collected as well as, Machine Learnings (ML) and Deep Learning (DL) techniques used in the paper.

2.1. *Injection moulding*

The injection moulding process is commonly used to transform plastic pellets into moulded parts (Rauwendaal 2000; Whelan and Goff 1996). This produces a number of plastic elements for each 'shot' (injection moulding cycle). The basic layout of an injection moulding machine is presented in Figure 2. The moulding machine and the injection mould are two separate units where many different moulds can be used in the same moulding machine.

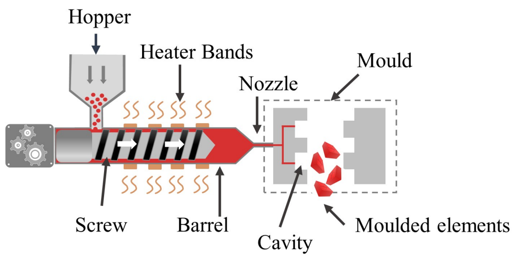


Figure 2. Sketch of a Injection moulding machine with the key components named.

The moulding process consists of roughly four phases:

(1) First, plastic pellets are fed into the hopper and the barrel. The plastic pellets in the barrel start to gradually melt due to the rotation of the screw that generates shear heat and due to external heat, that is provided by the heater bands wrapped around the barrel. As the screw rotates, the molten plastic is pushed in front of the screw (dosing phase) and the screw is forced backwards.

(2) The next phase is the injection phase, where the screw is moved forward (as a piston) and the molten plastic is pushed through the nozzle into the mould. A return valve on the screw ensures that the molten plastic flows into the mould and not back in the barrel.

(3) The plastic elements are cooled while the mould remains closed (cooling water is flowed though cooling channels placed inside the mould).

(4) When the plastic elements have solidified, the mould is opened and the plastic elements are ejected from the mould, as illustrated in Figure 2.

The injection moulding process is a complex process during which various pressure, temperature and speed measurements are collected and used for controlling the process through engineering control. The injection moulding process can be seen as a fast batch process (8-30 seconds), as each 'shot' yields time series data that is expected to follow a predetermined profile, see examples of injection pressure profiles in Figure 3. It has been well established that quality assurance in injection moulding requires the analysis of multiple process variables simultaneously as in the multivariate statistical process control (MPSC)(Ambrozic and Hutson 2006; Hazen and Kazmer 2007; Kazmer et al. 2008).

2.2. *Data acquisition and preparation*

As described in section 1, three different data sources (predictors) have been in focus for the exploration of injection moulded element quality (metrology). The three data sources can be divided in two main groups, one derived from sensors integrated in the injection moulding machine (Machine Sensor Signals) and one derived from additional sensors added to the injection mould (Cavity Sensor Signals). The origin, content/structure and pre-treatment of the different data sources will be described in the following sections.

2.2.1. Machine Sensor Signals

The Machine Sensor Signals consist of two separate datasets, one containing the Process Data that by default is available on any moulding machine and high-resolution continuous signals (Machine Profiles) that are used within the moulding machines, which are not by default available.

2.2.1.1. Machine Process Data. For injection moulding there are five main types of variables collected as Machine Process Data; Temperature, Pressure, Speed, Positions and Movements (delta positions). The available Machine Process Data has a n $\times$ m matrix structure with n being the number of moulding cycles collected and m being the number of selected variables.

The process variables are designed/selected by the moulding machine manufactures to give a good representation of the underlying continuous moulding process dynamics described by the Machine Profiles shown in the next section. Pressure variables for example are sampled at specific points from the underlying continuous reading of the hydraulic pressure. This is the case for the injection pressure, (the peak pressure in the melt during the injection), the switch over pressure, (the pressure at switch over point (where the injection phase switches to the packing or holding phase)) and the holding pressure, (sampled during the phase with constant holding pressure). Examples of other key process variables are material cushion, barrel temperature, injection time. In total, 28 process variables are collected and used in this work.

The Machine Process Data is used as input data to a simple Linear Model, described in section 2.4.2, and since the variables are numeric and on different scales, they will be mean centered and scaled to unit variance before used in the analysis.

2.2.1.2. Machine Profiles. The Machine Profiles are the underlying continuous data used within the machine to generate many of the Process Data variables. The Machine Profiles can be extracted from the moulding machine at the end of each moulding cycle. Since the Machine Process Data only represent fractions of the profiles it is expected that the complete profile will contain more information that reflects the underlying dynamics in the moulding process.

On the Arburg moulding machine used in this exploration, there is a limitation on how many data points that can be sampled per moulding cycle. E.g., 501 points per exportable dataset, called 'plot window'. To increase the resolution of the collected profiles the moulding cycle is divided into three plot windows, where each plot window is set to capture a specific part of the moulding cycle; injection phase, dosing phase and de-moulding. Within each of these plot windows, there is a set of continuous variables that can be selected to reflect the target dynamic. For the three plot windows the continuous variables shown in Table 1 were selected.

Table 1. The nine moulding profiles collected, with variable name, ID, measure unit, number of time point collected and resolution in sec between data points.

Moulding phase	Variable	ID	Unit	Time points	Resolution
Injection	Injection pressure	InjectPress	bar	501	0.006
Injection	Injection Speed	InjectSpeed	mm/s	501	0.006
Injection	Screw Position	InjectScrewPosition	mm	501	0.006
Dosing	System Pressure	DosingSystemPress	bar	368	0.006
Dosing	Dosing Speed	DosingSpeed	mm/s	368	0.006
Dosing	Screw Position	DosingScrewPosition	mm	368	0.006
De-moulding	Ejector Force	EjectorForce	kN	501	0.002
De-moulding	Ejector Speed	EjectorSpeed	mm/s	501	0.002
De-moulding	Ejector Movement	EjectorMove	mm	501	0.002

An example of each of the collected profiles can be seen in Figure 3. The variation seen in the profiles represents the total variation of the profiles in this work.

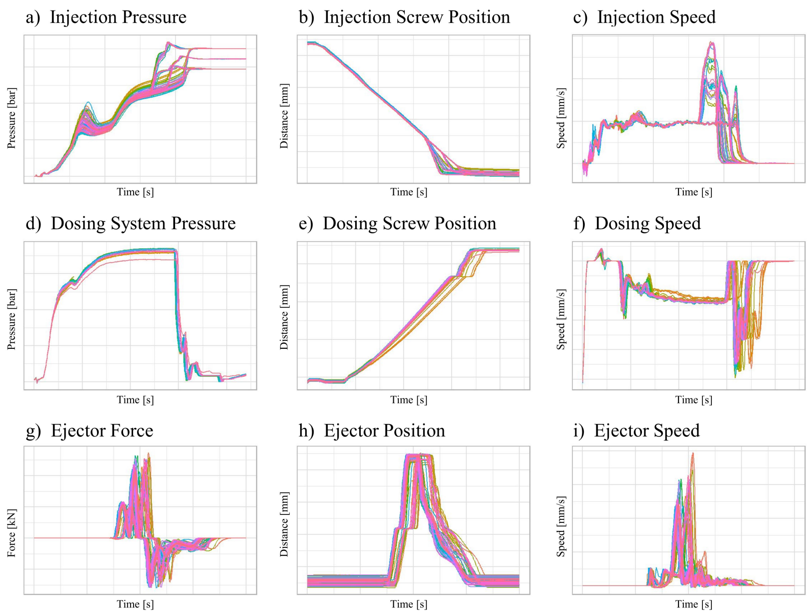


Figure 3. Collected injection moulding Machine Profiles.

The profiles displayed in Figure 3 can be structured in different ways when utilised for prediction of element quality. They can be utilised individually, but it is expected that combining the information from more profiles will improve the prediction. In this work, the profiles are used as individual profiles (total of 9 profiles) and where 3 profiles are combined by stitching them together in the time dimension into 3 new 1D profiles (illustrated in Figure 4), 3×[88,1370] (samples , time points). The combined profiles are referred to as a 'channel' and is created in the following way:

- Channel 1 - Pressure/Force (**InjectPress** + **DosingSystemPress** + **EjectorForce**)
- Channel 2 - Speed (**InjectSpeed** + **DosingSpeed** + **EjectorSpeed**)
- Channel 3 - Movements (**InjectScrewPosition** + **DosingScrewPosition** + **EjectorMove**)

These 3 channels will also be combined into a 3D data structure [88,1370,3] (samples, time points, channels). By combining the three channels in a 3D array, it is possible to extract correlation between pressure/force, speed and mould movement.

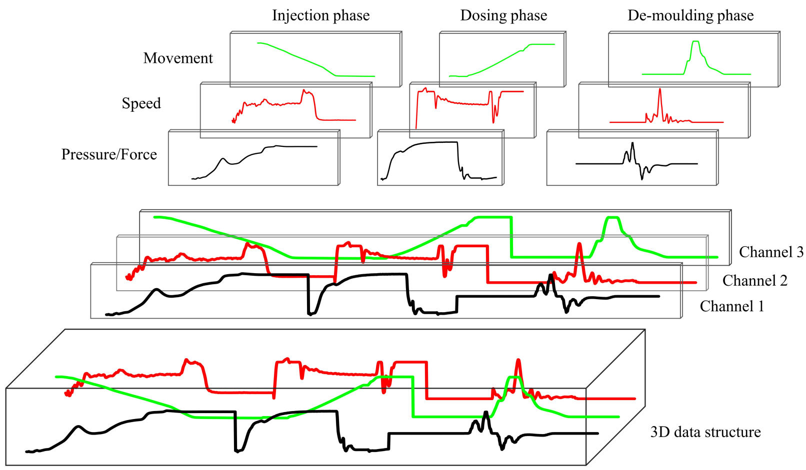


Figure 4. Illustrating the process going from single profiles into combined Machine Profiles in a 3 dimensional structure.

As described in the section 3, 8 consecutive shots were collected to represent each quality measure. In the utilisation of the profiles an average profile is therefore used. All the profiles are on different scales and since they are to be utilised with Deep Learning, they are normalised to the range [0,1] with a minmax scaler. The scaling is done as

$$x_i^{'} = \frac{x_i - \min(x)}{\max(x) - \min(x)}, \tag{1}$$

where x is the matrix containing all related profiles, x_i is the original profile and $x_i^{'}$ is the scaled profile.

Since profiles are to be used for different Deep Learning methods, and not all of those methods allow for 3D array as the input, this array will in specific methods be unfolded to a 2D array [88,4110], which causes the time correlation between the profiles to be lost.

The Machine Profiles will be used both in autoencoders, described in section 2.5.2, to extract the features from the profiles, and directly as the input to Deep Learning regression methods, described in section 2.5.1.

2.2.2. Cavity Sensor Signals

The cavity dataset (Cavity Profiles) consists of temperature profiles, measured by four (one in each cavity) PRIAMUS[1] cavity temperature sensors, and four pressure profiles, measured by PRIAMUS cavity pressure sensors, with a resolution of 3000 measures pr. cycle, resulting in 2×[4,3000] for each cycle (Figure 5). Since in each cycle, the four profiles from temperature sensors and four profiles from pressure sensors are without any significant variations, the four profiles for each are averaged into one mean profile, to remove potential noise in the individual profiles, resulting in 2×[1,3000] profiles. This procedure is done for all 88 experiments: 2×[88,3000].

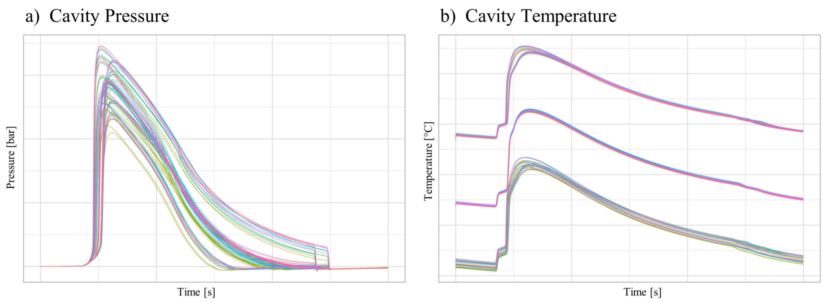


Figure 5. Cavity profiles showing the total variation in the collected data.

The profiles from each cycle are combined in the third dimension [88,3000,2], but like mentioned in section 2.2.1.2, not all of the learning methods allow 3D inputs, and the profiles are therefore in specific cases flattened into [88,6000].

Similar to the Machine Profiles, the resulting Cavity Profiles are to be used for Deep Learning, and are therefore scaled to range in [0,1] as in Equation 1, and are combined in a similar manner as the Machine Profiles, illustrated in Figure 4.

2.2.2.1. Standard Cavity Features (SF). The classical approach of utilising the information from the Cavity Profiles is to use the two Cavity Standard Features that by default are extracted in the PRIAMUS system (Kent 2016). The two features are the peak and the integral of the profiles. This is available for both pressure and temperature and will in the text be referred to as 'The Standard Cavity Features'. In

[1]PRIAMUS System Technologies is a manufacturer of cavity sensors.

this paper, these features are used directly, like the Machine Process Data, but we also investigate the advantages of using the full profiles particularly for the applications using Deep Learning methods.

2.3. *Explorative analysis*

Principal Component Analysis (PCA) is a well-known technique for dimension reduction and for visualising structures in large and complex datasets (Kégl et al. 2008; Jolliffe 2011). PCA is a projection method where the original data is projected onto new latent variables (Principal Components) that are orthonormal, uncorrelated and represented in decreasing order the directions of the highest variance in the data. Each sample/observation is in the latent space represented by a set of score values that can be plotted in a score plot (scatter plot). The structure/patterns in the score plot can be interpreted by looking at the corresponding loading plot that contain information of the correlation between the variables. Variables that are grouped together are positive correlated whereas variables that are placed in the opposite directions are negative correlated. The higher the loading value, the more impact the variables have (placed further away from origin).

2.4. *Regression methods*

This section describes the ways in which Machine Learning techniques are used to analyse the data (an overview can be seen in Figure 7).

2.4.1. Evaluation criteria (Prediction error in % of specification spread)

For the online prediction of quality to be useful, it is desired to obtain a prediction error less than 10% of the specification spread on the response measures, which is a rule of thumb often used in Measurement System Analysis for example (Pyzdek 2003). From this it is decided to use Root Mean Square Error (RMSE) in percent of the specification spread (RMSE_Pct) as evaluation criteria when comparing results for the different models. RMSE is calculated in the following way (Esbensen et al. 2002):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_{i,ref}\right)^2}{n}}, \tag{2}$$

where $\hat{y}_i$ is the predicted value for observation i, $y_{i,ref}$ is the reference value and n is the number of observations. RMSE returns the prediction in the original measuring units.

RMSE_Pct is then calculated as a ratio between RMSE and the specification spread,

$$RMSE\_Pct = \frac{RMSE}{USL - LSL} \cdot 100, \tag{3}$$

where USL is the upper specification limit and LSL is the lower specification limit.

2.4.2. Baseline models

To build a baseline for benchmarking the performance of the feature learning methods, a linear regression model (LM) is used for the quality measures described in section 4.1, using the Machine Process Data, described in section 2.2.1.1 as the input. Since there are multiple predictors, a Multiple Linear Regression (MLR) model will be used for modelling the relationship between the response variable and the independent predictors.

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_px_p + \varepsilon, \tag{4}$$

where y is the response, b_0 is the intercept, b_p is the slope for predictor x_p and ε is the error term. The justification of using MLR is that only few predictors are used, and that the collinearity among the predictors is low.

2.4.3. Utilising features from Deep Learning in regression

The gradient boosting machine (GBM) is an extension of the boosted decision trees, making it usable for regression problems. A decision tree splits the input into distinct regions aiming to minimise the loss of the prediction. Essentially the GBM algorithm is an ensemble method, that iteratively combines several of these relatively simple decision trees into a learner with improved accuracy. To use this for a regression problem, the model improves in regions where it does not perform well by iteratively fitting a new tree to the current residual, until some stopping criteria is met (Touzani et al. 2017). In this paper, the GBM model is used on autoencoded data, described in section 2.5.2, with the response variables.

2.5. *Feature learning method for profile data*

When looking into utilising the machine and cavity time resolved profiles, in the form of short time series, there are a variety of options. Zhang and Alexander (2008) have used fast wavelet transformation for feature extraction of Cavity Profiles, Mao et al. (2018) have successfully used Convolution Auto Encoder (CAE) for feature extraction from cavity and Machine Profiles for use in process monitoring, Fawaz et al. (2019) provide a comprehensive comparison of Deep Learning methods for time series classification and mention CNN methods as particularly effective in capturing discriminative shapes regardless where they occur in the time series, Liu et al. (2018) have demonstrated good results from using multivariate CNN for time series classification. Based on all this prior work, it is decided to focus on utilising Deep Learning for analysing the time resolved Cavity and Machine Profiles.

2.5.1. Deep Learning

Earlier work on raw signal data processing has shown good results using LSTM (Long Short Term Memory) networks (Nagorny et al. 2017), and CNN (Convolutional Neural Networks) (Ordóñez and Roggen 2016). LSTM belongs to the Recurrent Neural

Network (RNN) family, which is specialised in processing sequential data (Goodfellow et al. 2016). More specifically LSTM networks have gates that control the flow of information, which enables them to store time dependency information (Abbasimehra et al. 2020), and are therefore expected to be able to model the dynamics in the time profiles. CNN is another type of neural network, specialised in processing grid-like datasets (Goodfellow et al. 2016). An example is digital images, which essentially are 2D grids of pixel values, and CNN has shown good results in this field (Dabeer et al. 2019). In the same way as an image can be thought of as a 2D-grid of pixels, a time-profile can be thought of a 1D-grid of samples over a fixed time interval (Goodfellow et al. 2016). CNN can therefore be a promising method for extracting informative patterns from the profile data.

These Deep Learning methods will be used both as classifiers of selected settings in the designed experiment (DoE), and regression on quality measures of the moulded plastic elements. These DoE settings and quality measures are described in section 3.

2.5.2. Autoencoders

In manufacturing, it is often the case that the data representing the response variables (e.g. product quality data) is harder to obtain than the predictors. This is also the case in this work, yielding a small training dataset for predictive modelling. As proposed by Frumosu and Kulahci (2019) a way to overcome this is by using a semi-supervised approach. A similar approach in Deep Learning will be through the use of autoencoders (AEs).

The AE is a technique in neural networks, consisting of two phases: encoding, where the input is mapped to a hidden layer, and decoding, where the hidden layer is used to regenerate the input (Sagheer and Kotb 2019). AEs are proven to capture useful properties of the input data and store them in the hidden layer (Goodfellow et al. 2016), in order to be able to regenerate the input. In this paper, the AEs will be trained on the unlabelled data (see description at the end of section 3), as shown in the top of Figure 6, and afterwards used to encode the labelled data into a reduced feature space, and run regression on this encoded feature space. Since the AE is trained on unlabelled data, the main benefit is that a lot more data often is available for training the many weights and constants in the neural network, which can be a challenge with a small data set. Another benefit is that the hidden layer in the AE contains a reduced number of data points (features) compared to the input data. This can be utilized to reduce model complexity when different profiles have to be combined into one joint regression model. The main risk when using a technique where the predictors are reduced before being used in a regression model (as in the Principal Component Regression), is that the response is not considered when the features are generated and that important variation in the predictors that correlate to the response therefore is dropped from the feature space.

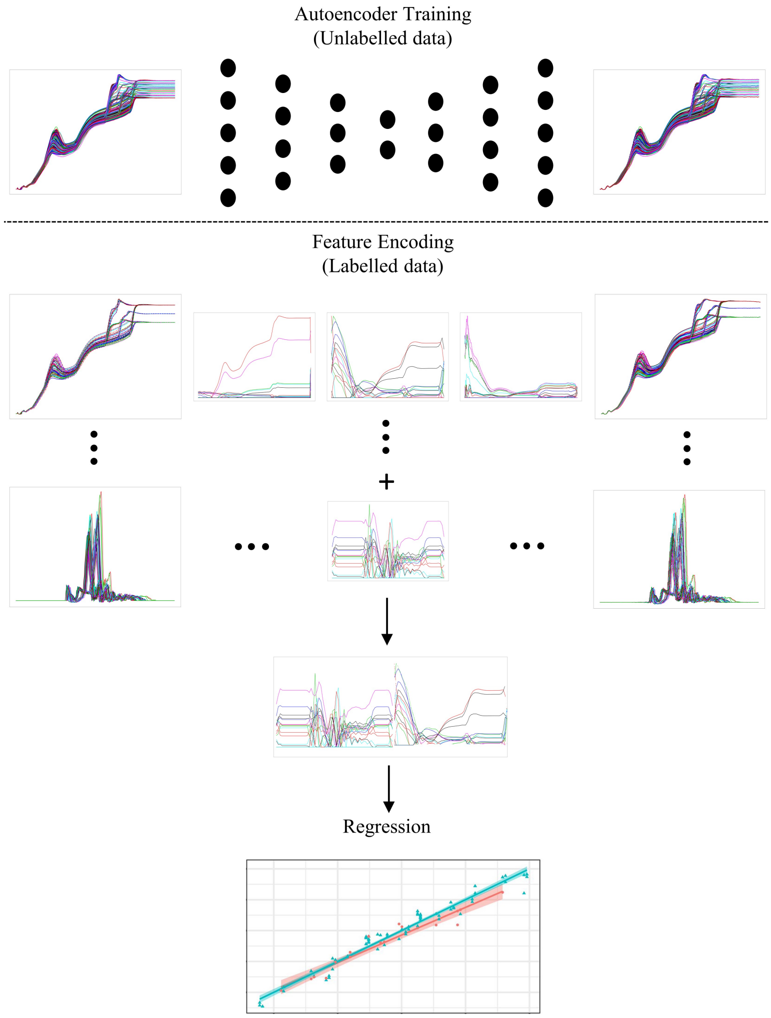


Figure 6. Illustration of unsupervised approach using autoencorders (AE).

The AEs will be built with a Feedforward, Convolutional, and LSTM structure. Combining the profiles before running them through the AE results in a high regeneration error. Even though this does not necessarily mean that the encoded space does not contain the needed information to use as input to the regression methods, it indicates that some information in the profiles is missing in the encoded space. Therefore, as the input of the AEs, each individual profile is encoded into the reduced feature space, and these encoded features are afterwards combined, before being used as input to the regression model. This combination is illustrated in Figure 6 for two profiles but are done for all nine profiles. The approach of combining the feature space for the profiles results in more features as input to the regression compared to combining the profiles before the encoding. However, since the goal is not to lower the number of features, and each individual profile is regenerated with a lower error this way, the feature space is expected to contain more information of the original profiles and thereby reducing the risk of leaving out variation correlated to the responses.

Figure 7 gives an overview of how the four datasets are utilised in the paper, resulting in six different outputs, that are benchmarking the section 4.

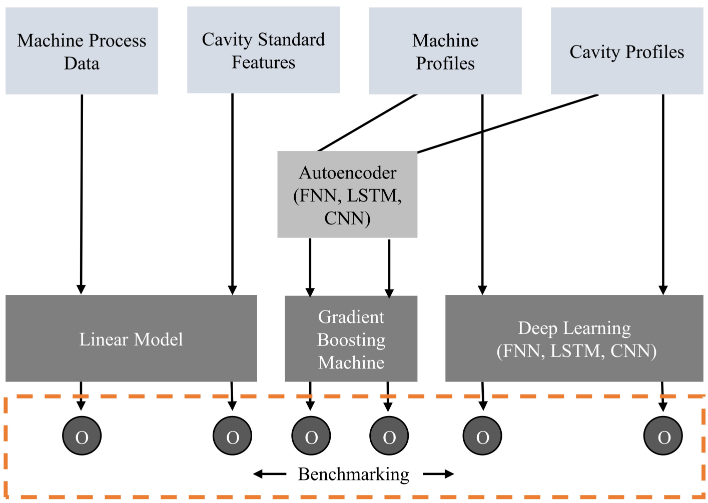


Figure 7. Overview of the usage of datasets and data modelling techniques.

3. Experimental setup

The information content has been a concern with Machine Process Data particularly when the primary goal of production is to keep the processes as stable as possible. The consequent lack of variation in the Machine Process Data requires large amounts of data to be collected, stored and analysed. This certainly has its own challenges. An alternative solution is to run a controlled study where the process is deliberately changed to create the needed variation. Plastic injection moulding is a very stable production process that often runs with constant settings. This makes it challenging to conduct controlled experiments to generate enough variation in the data yet still stay within the feasible operating conditions.

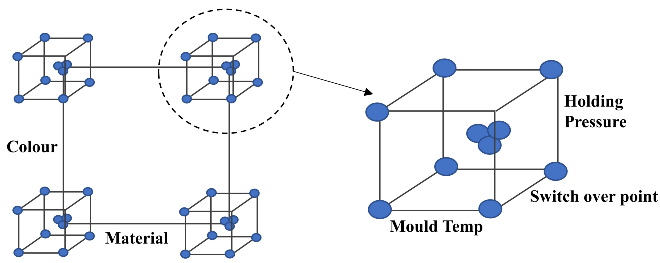
To create a dataset that contains enough variation in both the element quality and the generated input profiles, an experimental study is performed. Five factors were identified as relevant to include in the study; material, colour, mould temperature, holding pressure and switch over point. The variation in material is generated by using the same quality grade ABS plastics from two different vendors simulating the dual sourcing and colour variation was introduced by using red (organic colour) and white (inorganic colour). The mould was optimised using 99% filling and the switch over points for 70% and 85% fillings were calculated to simulate different running-in of the mould. Mould temperature was varied between 20 °C and 50 °C, which is a large variation compared to normal operation range. Holding pressure was varied with +/- 70 bar from the normally used holding pressure, which represents the allowed variation under normal production. The test settings and run order can be seen in Figure 8a and the structure of the experimental design can be seen in Figure 8b.

Since changing material and colour is time consuming, it was decided to use a split plot design (Kowalski et al. 2007). Material and colour were treated as 'hard to change' factors and mould temperature, holding pressure and switch over point were treated as 'easy to change' factors (see Figure 8). The experiment was conducted over a period of 8 days, with one combination of the hard to change factors and all combinations of the easy to change factors each day (11 test runs each day and 88 test runs in total).

WP	Run	Material	Color
1	1 - 11	ABS 1	Red
1	12 - 22	ABS 1	White
1	23 - 33	ABS 2	Red
1	34 - 44	ABS 2	White
2	45 - 55	ABS 2	White
2	56 - 66	ABS 2	Red
2	67 - 77	ABS 1	Red
2	78 - 88	ABS 1	White

Mould Temp	Holding Pressure	Switch over
20	Low	70%
20	Low	99%
20	High	70%
20	High	99%
50	Low	70%
50	Low	99%
50	High	70%
50	High	99%
35	Med	85%
35	Med	85%
35	Med	85%

(a) Settings in the experimental setup. Split plot with two hard to change factors and three easy to change factors. The displayed settings for the easy to change factors are the standard settings, in practice they were randomised for each hard to change run.



(b) Illustration of the experimental split-plot design used. A two-level factorial design with the two hard to change factors is the base of the experiment. In each of the four corners of this two-level factorial design the three easy to change factors is varied in a three factor two level factorial design with three centre points. This combined is called a whole plot and has here been repeated twice (WP1 and WP2).

Figure 8. Setting (a) and structure (b) of the experimental design used to generate the variation in predictors and responses.

The test was run on an Arburg allrounder 470 S, 600-290 with a 4-cavity test mould. The mould was equipped with PRIAMUS pressure and temperature sensors in each cavity. The element produced was a small rectangular box with outside dimension of 33.45mm×17.40mm×11.50mm, and with a wall thickness of 1.75mm. The element has a supporting rib in the middle of the element with a wall thickness of 1.00mm as shown in Figure 9.

To ensure a representative quality measure, 8 consecutive shots were collected. From the collected samples, 2 elements were randomly sampled from each of the four cavities.

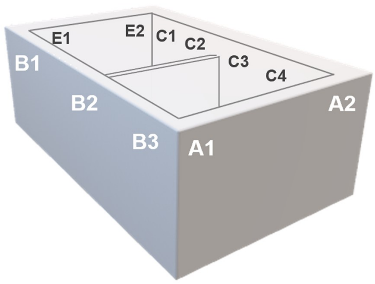


Figure 9. A 3D illustration of the moulded element with the 11 measuring point. All meteorology results were measured with a Coordinate Measuring Machines from Zeiss.

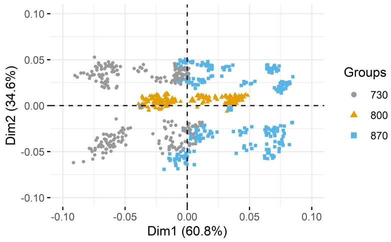
Initial tests show that 140 shots between the easy to change factors are sufficient to obtain a stable moulding process. Elements for quality measurements were therefore sampled from shots 141 to 148 in each test run, leaving data from the remaining 140 unlabelled shots available for training the AE (containing unique information since they are sampled in the transition between one stable moulding process to another). The Design of Experiment (DoE) is only used to introduce variation in the data for the evaluation of the three data sources under investigation. This means that the results will not be evaluated as a classical DoE, but that the data will be treated as observational data coming from a production environment.

4. Results

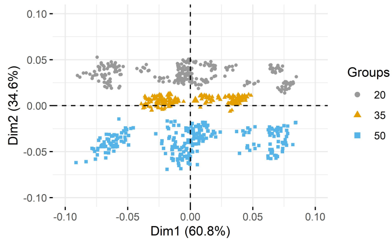
The overall focus is to investigate whether the readily available data from an injection moulding machine is sufficient in effectively describing the variation in the quality measures or significantly better results can be achieved by using cavity sensor or time resolved profile data from the injection moulding machine.

4.1. *Response exploration*

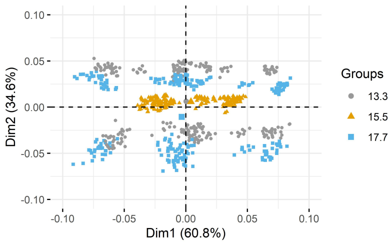
In cases where multiple responses are potentially correlated, a common approach is to use Principal Component Analysis (PCA) to obtain the uncorrelated principal components as the linear combinations of the responses and use these to investigate groupings/similarities among the samples by plotting the scores in a scatter plot. When plotting the scores of the first two principal components as in Figure 10, a clear grouping emerges and by colouring the scatter plot according to the DoE settings the grouping can be interpreted.



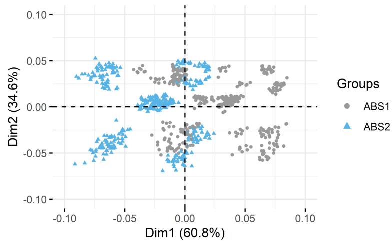
(a) Colour according to **Holding Pressure**



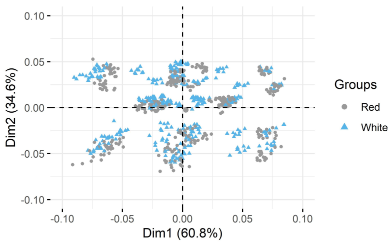
(b) Colour according to **Mould Temperature**



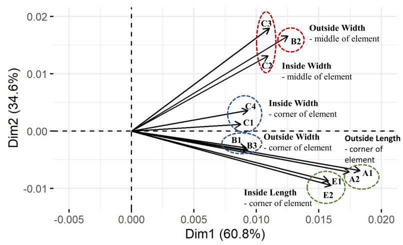
(c) Colour according to **Switch Over Point**



(d) Colour according to **Material**



(e) Colour according to **Colour**



(f) Contribution Plot

Figure 10. Score plot coloured according to settings in DoE. The input data for the PCA is a 704 by 11 matrix consisting of one row for each element measured and 11 columns with the 11 measuring points.Each point in the score plot represents one produced element and the 11 measures on the elements. 94.5% of the total variation in the responses can be explained by only two latent variables.

The first principal component summarises around 60% of the variation in the responses and is highly impacted by the change in raw material and pressure (Figure 10d and

10a). The second principal component summarises around 35% of the variation and is highly impacted by mould temperature (10b) and to some extent the switch over point where small subgrouping are seen in the direction of the second principal component, (10c). The change in colour (red to white) does not seem to have much of an effect on the metrology measurements.

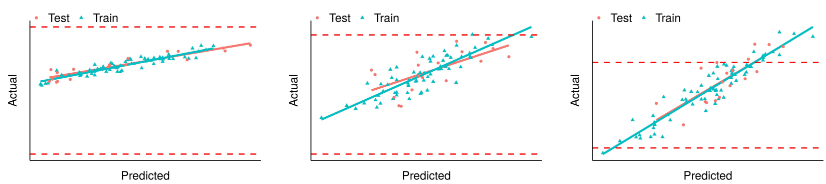
From the loading plot in Figure 10f, it can be seen that the changes in the DoE settings have different impact on the different location of the element. The inside and outside dimensions for the same measures are placed close together indicating high correlation. It is also seen that the width at the middle of the element is impacted mostly by the change in the second principal component, which is affected by changes in the mould temperature. This also makes sense since warpage is highly impacted by the mould temperature.

Since the 11 responses are highly correlated (as illustrated above), we decided to reduce the number of responses to three measures that reflect different phenomena (the three directions of the loadings). This is done by averaging the metrology measures in the following way (the link between the measuring position on the moulded element and the coding for the measuring points can be seen in Figure 9):

- **Outside length (OutL)**: The average of A1 and A2
- **Inside Width end (InWend)**: The average of C1 and C4
- **Inside Width middle (InWmid)**: The average of C2 and C3

4.2. *Prediction of element quality*

To establish a base model for comparison, a linear regression model is fitted to responses defined in section 4.1 using Machine Process Data. The scatter plots of the predicted responses shown against the actual values are given in Figure 11. Product specification limits are added to the plots to put the variation in the data in a better perspective.



(a) OutL
Test R^2: 0.87
Test RMSE Pct.: 3.6%

(b) InWend
Test R^2: 0.43
Test RMSE Pct.: 13%

(c) InWmid
Test R^2: 0.63
Test RMSE Pct.: 16.1%

Figure 11. Actual vs Predicted (both training data and test set) for the baseline models for OutL, InWmid and InWend. A Linear Model (LM function in R) was used for the regression. The basic statistics can be seen below the three plots. The specification spread for OutL is a factor of 4 larger than for InWmid and InWend.

The variation generated for inside width at the middle of the element (InWmid) span the hole specification band and exceeds the upper specification limit. The main reason for this is the large variation in mould temperature that causes warpages in the elements. The mould temperature is in normal production kept fixed around the centre point setting with only small variation. It should therefore be noted that the variation seen here is generated for experimental purposes and is not representative for normal production conditions.

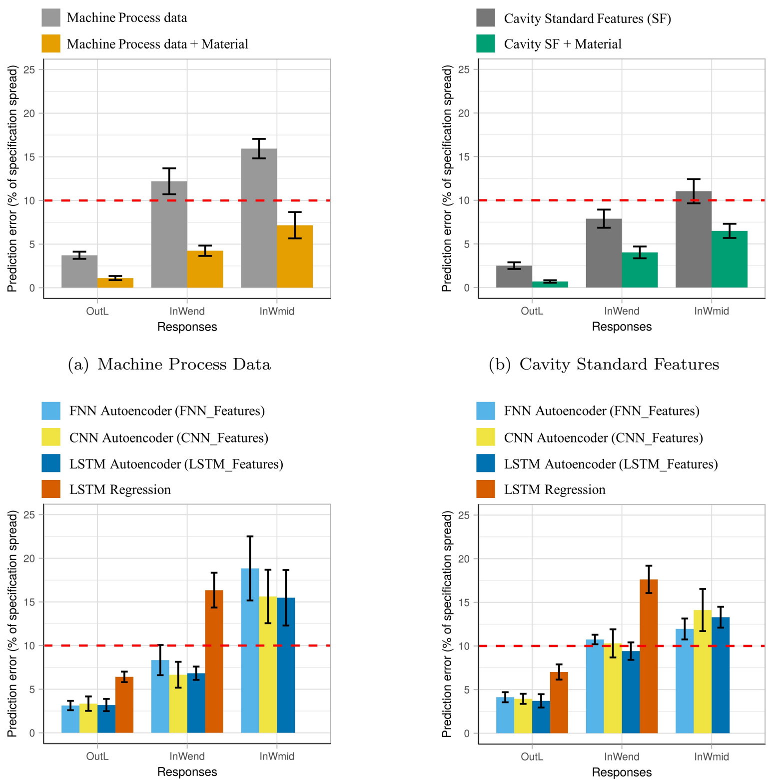
Figure 12a shows a comparison of the prediction error for the three baseline models against the upper limit for the prediction error, which is 10% of the specification spread (This error measure is explained in section 2.4.1). The prediction error for the baseline model based on the Machine Process Data is less than 10% of the specification spread for OutL, but exceeds the set threshold for InWmid and InWend.

Since from section 4.1 it was concluded that material had a significant impact on the responses (this aligns with the findings of Farahani et al. (2019)), the material information is added to the Machine Process Data, which improved the prediction error as seen in Figure 12a. This shows the importance of the material information and that none of the process variables reflect the change in the raw material. It should be noted that the material information is only available for the experimental setup. The actual production setup however is running with dual sourcing of ABS using one silo for the material[2]. This means that there is no possible way of tracking the brand of ABS used in the moulding machine. The material information is therefore unknown and cannot be included when using actual Machine Process Data. The cost of changing the raw material storage and feeding systeme is to high for it to be a realistic solution.

Therefore, the material information needs to be measured or extracted from a different data source. As demonstrated by Dumitrescu et al. (2005) and McLauchlin et al. (2014), Near Infrared spectroscopy (NIR) could be used for inline classification of plastic variations. Since it is costly to implement NIR probes in all moulding machines it will be investigated if the cavity or machine profile data can be utilized to indirectly measure the change in raw material and thereby be used to improve the predictive models.

Figures 12b through 12d show the prediction results of the models using Standard Cavity Features of the Cavity Profiles (described in section 2.2.2.1), Machine Profile data and Cavity Profiles, respectively. The FNN_Features, CNN_Features and LSTM_Features correspond to the models from the feature extraction of Feedforward AE, Convolutional AE and LSTM AE respectively, where the extracted features are used in a regression with the Gradient Boosted Machine, described in section 2.4.3. LSTM Regression is the Deep Learning method on both the Machine Profiles and Cavity Profiles.

[2]New supply of material is added in the top of the silo and material used in the production is emptied out though the bottom. This means the materials from the two suppliers are layered in the silo.



(a) Machine Process Data (b) Cavity Standard Features

(c) Machine Profiles (d) Cavity Profiles

Figure 12. Comparison of prediction accuracy with different input data. In order to get an estimate of modelling error a 5-fold cross validation with new random splitting of data into train and test set (20% of the initial data) were used. The variation from doing the cross validation is represented with the error bars. For illustrative purposes, the models which exceeds 25% of specification spread on the responses, are removed from the plots (FNN and CNN regressing models and LSTM regression for InWmid), to focus on the best performing methods. The upper limit for the prediction error (10% of the specification spread) is indicated with the red dotted line.

From Figure 12, it is noted that none of the Deep Learning methods perform better than the baseline model on Machine Process Data with the material variation included, but the encoded features with GBM regression offer promising results for both the OutL and InWend responses. The Deep Learning LSTM regression using the profiles has a higher prediction error than that of the baseline models, and the encoded features from the profiles. The fact that the DL methods perform worse than the linear baseline model indicates that either the relationship between the Machine Process Data and the responses can be approximated sufficiently well with a Linear Model, or the more complex nonlinear models suffer from the lack of training data, which is in agreement with Hastie et al. (2017). This is also supported by the fact that the AE features perform better than the LSTM regression, since these are built using more training data.

The results obtained using AE on the Cavity Profile data in Figure 12d have a lower prediction error than the results obtained using the Machine Profiles (12c). The reason for this could be that the Cavity Profiles are less complex and therefore the models do not require the same amount of training data as in the case of more complex Machine Profiles. Another explanation could be that the cavity sensors better capture the material variation. This can be tested by performing an FNN classification of the material type, with the Machine and Cavity Profiles as inputs.

As shown in Figure 13, only a few of the collected profiles seem to contain information related to the variation in material. An interesting observation is that profiles related to the dosing reflect the material variation the most as in the case of dosing speed[3]. This correlation can be used to reduce the model complexity by building a surrogate variable from the profiles that is correlated with the material variation (dosing speed). The surrogate variable can thereafter be infused into the Machine Process Data to match the results from Machine Process Data + material in Figure 12a. It is also noteworthy that the Cavity Profiles do not seem to capture the material variation.

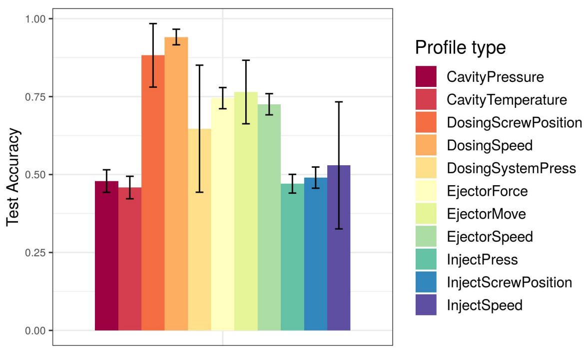


Figure 13. Classification accuracy obtained from FNN classification model with Machine Profiles as input and the settings in the DoE as responses.

Since the process variables are usually readily available, it should be possible to obtain the same prediction results when using Machine Process Data + material by obtaining the 'material information' from first classifying the dosing speed in the dosing phase and combining this result with the Machine Process Data. The result obtained from this can be seen in Figure 14.

[3]This can be explained by small changes in viscosity of the melt depending on material properties.

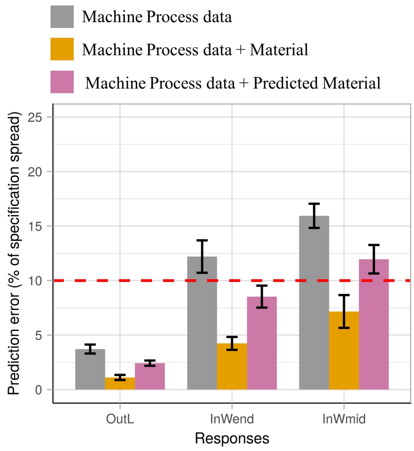


Figure 14. Comparison of prediction accuracy with material information from experimental settings and obtained from classification using the dosing speed profile (using a LM for all three models, and errors bars estimated as in Figure 12).

The results show that the models using Machine Process Data + the predicted material type outperform the similar models with Machine Process Data alone, but still with a significant higher prediction error than the Machine Process Data + the actual material type. This is expected due to inaccuracies in the classification as shown in Figure 13, which is made based on only 88 data points. It is therefore expected that this can be improved by adding more training samples. If it then is possible to get below the desired threshold of 10% prediction error of the specification spread this could be considered a valid approach moving forward.

5. Discussion

Our main goal throughout the article was to compare different data sources in injection moulding in terms of establishing a predictive model for the quality characteristics of the moulded elements. The data is obtained from an experimental campaign during which systematic variation is introduced in both the input variables and responses. Since the actual production is a very stable process and data sources available for the experimental setup are either not available or costly to obtain in practice, our ultimate aim was to determine the information content of these various sources of data towards establishing a predictive model that is also feasible to implement in actual production.

The results from the experiments have shown that the material type as a predictor greatly improves the predictive power of the base model built on the readily available Machine Process Data from the moulding machines, and therefore is highly relevant to include this type of information in the analysis. Also, the simple Linear Model with the material information as one of the predictors outperforms more complex Machine learning models. We speculate that the reason behind this is: (1) the relationship between the process variables and quality measures can be well approximated with a Linear Model and/or (2) the more complex nonlinear models used in the study suffer from the lack of training data. The latter is being supported by the fact that the model based on the encoded features extracted from autoencoding all the unlabelled data performs nearly as well as the Linear Models, and the fact that the Machine Profiles did not seem to correlate well with the material variation. This information could on the contrary be extracted when running classification directly on the individual profiles. It is therefore expected to be able to produce better results, if the profile information could be simplified. A suggestion to this is to examine which profiles do not contribute with additional information and exclude those from the combined profile data.

Furthermore, in general it seems to be crucial to collect more data points in order to obtain the desired accuracy of the quality measures, regardless of which approach is used. It can though be concluded that the investment in cavity sensors for the full production does not generate the desired value in improving the prediction compared to utilising solely the information in Machine Profiles. This study suggests more investigation being needed on the use of other data sources than the readily available Machine Process Data. In many industrial applications as in the case of this study, different sources of data are available at different costs. In order to justify the effectiveness of an expensive data source, a sufficiently good model needs to be obtained. This creates a dilemma on what comes first: the expensive and comprehensive data collection for building a good model, or the proof of the effectiveness of such data? This dilemma will in some cases result in a limited data collection in the initial phase of the projects, which can unfortunately lead to unsatisfactory or ambiguous results, and therefore to a weak foundation to prove the value.

Another concern in a data analytics study in stable processes is the lack of variation in data to train complicated models. The data for this article comes from dedicated experiments, and not from the production line itself. However, experimental data does not always represent well the real production conditions. For instance, in this study the temperature in the experiments varies in a greater range than it does in real production. This leads to yet another question: Should the data generated from the experiments emulate the real-life situation the best it can with the possibility of the low variation, or should it prioritise variation to build the best possible dataset for analysis?

6. Conclusion

This paper presents a scheme for selecting and evaluating different sources of manufacturing data using injection moulding for demonstration. It has been demonstrated that a structured and thorough evaluation of different data sources is important to select cost effective solutions for manufacturing. Marginal better prediction accuracy was obtained using cavity sensors compared to machine profile data, but not enough to justify an investment in cavity sensors for a complete injection moulding manufacturing setup. Variation from dual sourcing ABS from two suppliers was identified as the main uncontrolled factor impacting elements dimensions. Attempts were therefore made to classify the material used by utilising the machine profiles and include the classification results (92% classification accuracy was obtained) in the regression model for element quality. This improved the prediction accuracy for element dimension, but additional work has to be conducted to improve the results.

In the end, we offer a trivial conclusion; data is the most important component in a data analytics study, but ultimately the study has to yield a model that is implementable. In that sense, as tempting as it may be, demanding more data of higher quality may simply be futile. Manufacturing provides many opportunities for data analytics yet also demands to deal with reality on reality's terms. In that sense, a thorough cost/benefit analysis of acquiring more data towards achieving modelling goals is needed to be in compliance with the production requirements and limitations.

References

Hossein Abbasimehra, Mostafa Shabanib, and Mohsen Yousefic. An optimized model using lstm network for demand forecasting. *Computer & Industrial Engineering*, 2020.

Tatyana Ageyeva, Szabolcs Horváth, and József Gábor Kovács. In-mold sensors for injection molding: On the way to industry 4.0. *Sensors — Open Access Journal*, 2019.

Chris Ambrozic and Lee Hutson. Multivariate analysis (mva) for quality detection in injection molding systems in the medical device community. *Annual Technical Conference–ANTEC, Conference Proceedings*, 2006.

Sumaiya Dabeer, Maha Mohammed Khan, and Saiful Islam. Cancer diagnosis in histopathological image: Cnn based approach. *Informatics in Medicine Unlocked*, 2019.

Oana R. Dumitrescu, Dave C. Baker, Gary M. Foster, and Ken E. Evans. Near infrared spectroscopy for in-line monitoring during injection moulding. *Polymer Testing*, 24(3):367 – 375, 2005. ISSN 0142-9418. . URL `http://www.sciencedirect.com/science/article/pii/S0142941804001576`.

Kim H. Esbensen, Dominique Guyot, Frank Westad, and Lars P. Houmoller. *Multivariate Data Analysis: In Practice : an Introduction to Multivariate Data Analysis and Experimental Design*. CAMO, 2002. ISBN 82-993330-2-4.

Saeed Farahani, Nathaniel Brown, Jonathan Loftis, Curtis Krick, Florian Pichl, Robert Vaculik, and Srikanth Pilla. Evaluation of in-mold sensors and machine data towards enhancing product quality and process monitoring via industry 4.0. *The International Journal of Advanced Manufacturing Technology*, 105(1):1371–1389, Nov 2019. ISSN 1433-3015. . URL `https://doi.org/10.1007/s00170-019-04323-8`.

Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 2019.

Flavia D. Frumosu and Murat Kulahci. Outlier detection using an iterative strategy for semi-supervised learning. *Quality and Reliability Engineering International*, 2019.

Robert X. Gao, Xinyao Tang, Guthrie Gordon, and David O. Kazmer. Online product quality monitoring through in-process measurement. *CIRP Annals*, 63(1):493 – 496, 2014. ISSN 0007-8506. . URL `http://www.sciencedirect.com/science/article/pii/S0007850614000444`.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. An MIT Press book, 2016. `http://www.deeplearningbook.org`.

Trevor Hastie, Robert Tibshirani, and Herome Friedman. *The Elements of Statistical Learning - Data Mining, Inference,and Prediction*. Springer Series in Statistics, 2017.

D. Hazen and David O. Kazmer. Multivariate process analysis for the prediction of injection molded part quality. *Annual Technical Conference-ANTEC, Conference Proceedings*, 2007.

Ian Jolliffe. *Principal Component Analysis*, pages 1094–1096. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-04898-2. . URL `https://doi.org/10.1007/978-3-642-04898-2_455`.

David Kazmer, S. Westerdale, and D. Hazen. A comparison of statistical process control (spc) and on-line multivariate analyses (mva) for injection molding. *International Polymer Processing Journal of the Polymer Processing Society vol. 23 (05): 447-458*, 2008.

Robin Kent. *Quality Management in Plastics Processing - Strategies, targets, techniques and tools*, chapter 10.12. Elsevier, 2016.

Scott M. Kowalski, Peter A. Parker, and G. Geoffrey Vining. Tutorial: Industrial split-plot experiments. *Quality Engineering*, 2007.

Murat Kulahci, Flavia Dalia Frumosu, Abdul Rauf Khan, Georg Ørnskov Rønsch, and Max Peter Spooner. Experiences with big data: Accounts from a data scientist's perspective. *Quality Engineering*, January 2020. ISSN 0898-2112. .

Balázs Kégl, Donald Wunsch, and Andrei Zinovyev. *Principal Manifolds for Data Visualisation and Dimension Reduction, LNCSE 58.* Springer, 01 2008. ISBN 978-3-540-73750-6.

Chien-Liang Liu, Wen-Hoar Hsaio, and Yao-Chung Tu. Time series classification withmultivariate convolutional neural network. *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*, 2018.

Ting Mao, Yun Zhang, Yufei Ruan, Huang Gao, Huamin Zhou, and Dequn Li. Feature learning and process monitoring of injection molding using convolution-deconvolution auto encoders. *Elsevier: Computers and Chemical Engineering*, 2018.

Andrew R. McLauchlin, Oana Ghita, and Ali Gahkani. Quantification of pla contamination in pet during injection moulding by in-line nir spectroscopy. *Polymer Testing*, 38:46 – 52, 2014. ISSN 0142-9418. . URL `http://www.sciencedirect.com/science/article/pii/S0142941814001330`.

Pierre Nagorny, Maurice Pillet, EricPairel, Ronan Le Goff, Jérome Loureaux, Marlène Wali, and Patrice Kiener. Quality prediction in injection molding. *IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 2017.

Francisco Javier Ordóñez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *MDPI Sensors*, 2016.

T. Pyzdek. *The Six Sigma Handbook: A Complete Guide for Green Belts, Black Belts, and Managers at All Levels.* Mcgraw-hill Education - Europe, 01 2003.

Chris Rauwendaal. *SPC: Statistical Process Control in Injection Molding and Extrusion.* Hanser Publishers, 2000.

Alaa Sagheer and Mostafa Kotb. Unsupervised pre-training ofa deep lstm-based stackedautoencoder for multivariate timeseries forecasting problems. *Sci Rep 9, 19038*, 2019.

A. Sanchez-Fernandez, F.J. Baldan, G.I. Sainz-Palmero, J.M. Benítez, and M.J. Fuente. Fault detection based on time series modeling and multivariate statistical process control. *Chemometrics and Intelligent Laboratory Systems*, 2018.

Samir Touzani, Jessica Granderson, and Samuel Fernandes. Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, 2017.

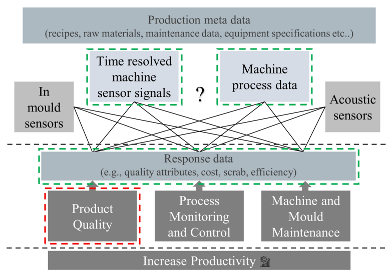
Tony Whelan and John Goff. *Quality Moulding.* Applied Market Information, 1996.

J. Zhang and S. M. Alexander. Fault diagnosis in injection moulding via cavity pressure signals. *International Journal of Production Research*, 2008.

Yun Zhang, Ting Mao, Zhigao Huang, Huang Gao, and Dequn Li. A statistical quality monitoring method for plastic injection molding using machine built-in sensors. *The International Journal of Advanced Manufacturing Technology*, 85(9):2483–2494, Aug 2016. ISSN 1433-3015. . URL `https://doi.org/10.1007/s00170-015-8013-2`.

CHAPTER 6

Paper B, Real-time adjustment of injection molding process settings by utilizing time series profiles and PLS-DA



It can often be challenging to investigate the impact of disturbance factors in a production setup. This can be because the disturbances are unknown, cannot be controlled, or because the data reflecting the disturbance are not collected. If data reflecting the disturbance factors can be collected and analysed, it is often challenging to move from data correlation to causality. Causality can be achieved by systematically changing a factor and observing the response (foundation for Design of experiments). Conducting experiments in a production setup can be challenging since controlling all factors impacting the outcome can be difficult. Hence, designed experiments are often conducted in a pilot plant or test setup. This can also be a challenge since it can be hard to link the results back to the actual production setup. This has been formulated into research question 2.

Research question 2 (RQ2)

In an off-line setting, how to develop real-time data analytic models that will help mitigate the impact of the variation in a nuisance factor in a production setup.

6.1 Summary

Paper B is linked to RQ2, focusing on the effective use of designed experimentation to observe the causal relation between disturbance factors, process settings and product quality. In the specific case, the variation in raw materials is the investigated disturbance factor. Two levels of material variation are known at the industrial partner, where the most apparent is introduced with dual sourcing of the same material type from multiple vendors. The second level of variation is represented by shipment to shipment from the same vendor. Only the vendor to vendor variation is investigated in paper B. As described in RQ2, the objective is to develop a solution that can be used in real-time to mitigate the impact of vendor-related material variations on the produced element quality. Besides including the relevant material disturbance in the experimentation, it is also applicable to include and vary process variables that can be used to compensate for the expected material variations. Last but not least, it is crucial to collect process state variables that reflect the material condition impacting the element quality. We utilise the learnings from paper A and include the machine pressure profiles as the state variable for the current work. These aspects of experimental work have been combined using a split-plot design where colour and material have been handled as hard-to-change factors and mould temperature and holding pressure as easy-to-change factors. From this, it was found that all main effects, except colour and one 2-factor interaction (pressure*temperature), had a significant impact on the element dimensions (represented by the outside length and inside width of the element). Material classification, utilising machine pressure profiles, was tested using Partial Least Squares (PLS-DA), Adaptive Boosting (AdaBoost) and Feed-Forward Neural Net (FNN). The performance of the three methods was comparable (classification accuracy between 94.5% and 97.5%), with PLS-DA performing best (pressure profile from the dosing phase).

Achieved results

- Conducting structured experimentation, it has been demonstrated that changes in injection and dosing pressure profiles can be used to detect a change in the origin of raw material. This has been achieved using PLS-DA with the machine profiles as input and the source of material as the response.
- It has been demonstrated that knowing the origin of the material (classification based on the pressure profiles), holding pressure and mould temperature can be used to reduce variation on element quality caused by material variations.

Contribution

- Approach for systematical exploration (off-line) of disturbance factors' impact on product quality and linking these findings back to a concrete solution that could be implemented in an industrial context (real-time).
- Demonstrating that underlying equipment signals contain relevant process information that can be utilised for improving product quality. This type of data could potentially be utilised from many different types of equipment, where energy consumption is collected as a function of time.

6.2 Paper B

Paper B

Real-time adjustment of injection molding process settings by utilizing Design of Experiment, time series profiles and PLS-DA

G. Ø. Rønsch[a*], M. Dybdahl[b] and M. Kulahci[a]

[a]DTU Compute, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby; [b]Department of Business Development and Technology, Aarhus University, Denmark.

ARTICLE HISTORY
Compiled February 23, 2022

ABSTRACT
Various production/process equipment's have built-in sensors allowing for continuous collection of process data. However to ease the data processing burden, it is often the case that only certain features such as aggregated measures or peak values are stored. Yet also in some cases such sensor signals can be extracted fully reflecting the dynamics of the process and utilized for process optimization. The aim of this paper is to demonstrate that such data can be utilized in an effective way to optimize an injection molding process using signals from built-in pressure and position sensors. The observational process data is combined with data from controlled experiments to observe the causal relationships between disturbance factors, process settings and the final quality of the products. We demonstrate that signals from built-in injection molding machine sensors can be used for detecting and mitigating quality issues caused by variation in raw material due to the dual sourcing from two suppliers, which cannot actually be identified during production. For this, we show that the origin of raw material can be classified using the time series profiles of dosing pressure and PLS-DA (Partial Least Squares Discriminant Analysis). Through experimental work, we conclude that this classification can be used for increasing the operating window for holding pressure and mold temperature, which ensure production of products within specifications.

KEYWORDS
Design of Experiments, Machine learning, Injection molding, In-built sensors, PLS-DA

* CONTACT G. Ø. Rønsch. E-mail: geor@dtu.dk

1. Introduction

In manufacturing processes, the primary goal of the production process is to effectively transform the raw material into the final product. To ensure the stability of the product quality over time, it is desired to use a raw material with uniform properties and keep external disturbances at a minimum. Depending on the manufacturing process at hand, robustness towards unexpected disturbances can be achieved by adjusting process settings when change in raw material and/or external disturbances are detected. The current work is anchored within plastic injection molding with similar concerns and conducted in close collaboration with an industrial partner. The investigation and experimental work are therefore centered around the manufacturing setup and control strategy used at a specific process, but we believe the overall approach is generalizable to other injection molding setups and to a large extent to other industries. Particularly in process robustness studies where there is a need to mitigate the impact of the nuisance factors that are uncontrollable in production yet controllable in an experimental setup, the proposed approach can be easily adapted.

The injection molding setup under investigation is characterized by having multiple molds used in a given molding machine. The setup therefore consists of many small production orders resulting in many changeovers where molds are changed in the molding machine depending on the order. This requires an effective and reliable running-in procedure that ensures stable production to reach the desired quality. After running-in a mold, the molding process is kept constant with only a few settings that can be adjusted by the molding operators. Despite of all this, there can still be unacceptable variation in the product (also known as element in injection molding) quality. This indicates that there are external disturbances that are strong enough to impact product quality despite the control strategy used.

Often companies use multiple sources for supply of raw material to ensure stable delivery and low cost. The industrial partner in this project uses two different sources for supply of raw material. In this paper, we refer to these sources as “source of raw material”, “material source” or “ABS1 and ABS2”. They come from two suppliers and are delivered in no specific order and stored in the same silo tank. The new shipments are unloaded into the storage silo from the top and material used in the production is extracted from the bottom of the silo. The different shipments are therefore layered randomly in the silo causing the origin of the material used at the machine to be unknown.

By dual sourcing of the raw material, there is a risk of additional disturbances caused by variation in the raw material (beside within supplier variation). The quality of the raw material is often described with a set of univariate specifications on the raw material’s quality attributes. Even if all univariate attributes are within specifications, the correlation among the attributes (as addressed by Duchesne and MacgregorDuchesne and Macgregor (2004)) could be different from supplier to supplier or change over time and thereby cause changes in product quality. At the industrial partner production site, it has been experienced that material variation introduced by dual sourcing affects the product quality, and considerable attention has been paid to reduce this impact. The experience that material variation is impacting quality is supported by Farahani et al. (2019) where variation in element weight is shown to be impacted by the variation in the raw material quality. One solution to overcome this is through modelling the variation and defining which process settings can be adjusted to compensate for it and thereby making it possible to keep product quality stable. This will however not be directly implementable since, as mentioned earlier, during the production the exact source of raw material remains unknow during production. In this work, we develop a real-time solution that can mitigate or reduce the potential impact of the changes in the raw material source by identifying the source of the raw material used at the process and thereby creating an opportunity for adjusting the machine settings accordingly.

Different options for determining the source of the raw material during production have been identified. The most obvious approach is to change the raw material feeding system so that it is possible to track the right source of the raw material at any time. However, this has proven to be an infeasible solution due to the cost of establishing such a system and it would not capture within supplier variation. The second option is about establishing a real-time material monitoring setup (in the current case used for classification of the raw material source). For this to be a feasible solution, it should come at a low-cost, which renders installing new measuring equipment rather infeasible.

For the final product quality, McLauchlin, Ghita, and Gahkani (2014) have demonstrated the Near Infrared spectroscopy (NIR) can be used for inline classification of plastic variations. Gao et al. (2014) show that it is possible to predict element dimensions using mold cavity pressure sensors. The implementation of mold cavity sensors is too costly to be an option, but as demonstrated by Chen et al. (2008), the in-built sensors in the molding machine contain similar information as the cavity sensors. Lopes and Ribeiro (2000) have also demonstrated that element dimension could be predicted using standard machine process variables and a neural network. However, none of these articles includes the aspect of material variation classification and how to mitigate this by changing settings on the molding machine. In the current work material variation is represented by dual sourcing of raw material, but this could be extended to seasonal variation, geographical variation or other variations in raw materials. In cases of more diffuse variation contributions (where samples can't be labeled) a regression approach could be consider instead of classification (not investigated further in the current work).

1.1. *Injection molding*

Injection molding is a widely used manufacturing technique that transforms plastic granulate into a plastic element with a given shape and properties. Molded elements vary in shape and size, from very small elements used for example in a hearing aid or in a pharmaceutical device to very large elements used for example in the automobile industry. Most molded elements are manufactured in a setup consisting of an injection molding machine equipped with a mold that defines the shape of the molded element as depicted in Figure 1. Besides the barrel, the molding machine consists of either a hydraulic system or electric servo drives that are used to move the screw and the moving part of the mold, in-built sensors that generate input signals to the control system and a control panel used by the operator to configure the molding machine and to monitor the state of the molding process.

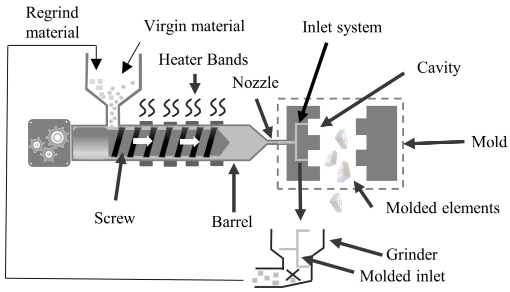


Figure 1. An injection molding setup where the left part is the barrel of the molding machine and the mold is represented on the right. The grinder used to grind the inlet from the mold is placed underneath the mold.

1.2. *Injection mold*

An injection mold consists of two halves that together form the cavity in which the molded element is created. There can be multiple cavities in one mold. The fixed half of the mold, which is closest to the molding machine, contains an inlet system that is used to transport the melted plastic from the barrel into the cavities. A mold is designed with either a cold or hot inlet system. In a hot runner mold, the inlet channels are kept at a temperature where the plastic in the inlet channels remains fluent between the injections. In a cold runner mold, the inlet channels are cooled in the same way as the molded elements. In this case, the inlets are removed when the elements are ejected from the mold. The inlets are often grinded, mixed with virgin material and reused as raw material in the molding process.

1.3. *Molding cycle*

The injection molding process is a fast batch process where the cycle time can vary from a few seconds to minutes. The number of cavities in the mold defines how many elements are produced in each cycle, which vary from 1 to more than 100 elements. One molding cycle consists of several steps. First, raw material pellets from the hopper enter the barrel at the back end of the screw. The screw rotates and thereby transports the plastic material towards the nozzle. As the material moves though the barrel, it is melted due to sheer stress and external heat from the heater bands. As more melt build up in front of the screw, the screw is pushed backwards. When the screw reaches a given position, it is ready for injection. This determines the end of the dosing phase. The next step is the injection phase, where the screw acts as a piston. A return valve at the tip of the screw ensures that melt is not pushed back along the screw. The screw hence presses the melt though the nozzle and mold inlet system into the mold cavities. At this point the two mold-halves are kept together by a high pressure applied to the moving half. When 95-98% of the melt is injected into the mold as measured by the position of the screw, the so-called switch-over point, the control strategy is changed from being based on screw position to a constant pressure, also known as the holding pressure. The next step is the cooling of the elements in the mold cavities. This is done by circulating water around in cooling channels inside the mold. As the molded elements cool down, the plastic starts to shrink. The constant holding pressure ensures that more melt can be pushed into the mold as the plastic shrinks while the core of the inlets and element remain liquid at this point. When the plastic has solidified, the mold is opened, and the elements are ejected out of the mold. The mold is then closed and ready for the next injection.

2. Methodology

The methodology contains an overall introduction to split-plot design, an introduction to the machine data used and an overall description of the machine learning methods used.

2.1. *Design of Experiments using split-plot design*

Design of Experiments (DoE) is a classical approach used for investigating casual relations between experimental factors and responses (Montgomery 2012). The causal relations are obtained by systematically changing the experimental factors and evaluating the impact on the responses. One of the fundamentals in DoE is that more than one factor can be varied in each test run. This is done to ensure that interactions between factors are captured as opposed to the one-factor-at-a-time experimentation. Usually the experiments are executed in a completely randomized manner to eliminate any possible bias. In some situations, however, it is a challenge to conduct the experiments in a fully randomized order. This is often the case in manufacturing where the costs of changing the settings of the experimental factors can vary greatly. One example is within plastic injection molding where change of raw material or color can take hours since the mold and molding machine must be emptied, cleaned and operated with the new material for the running-in period. In many molding applications, using a cold runner mold, the molded inlet is reground and mixed with the virgin material. When this is the case, an important part of the running-in is to ensure the steady state in the mix of reground plastic and virgin plastic since this might impact the molded element. If an injection molding experiment includes different materials to be tested, it is therefore desired to conduct multiple test runs using the same material and thereby reduce the needed time and resources for the experimental work. One approach to overcome the requirement of complete randomization is to use a split-plot design. Kowalski, Parker, and Vining (2007) present an extensive introduction to the use of split-plot design in an industrial context. Split-plot design is a special case of a factorial design, where for a given setting of the hard-to-change (HTC) factors (known as a whole plot) experiments in the easy-to-change (ETC) factors (known as subplots) are executed. This way hard-to-change factors are changed for fewer experiments than in the case of the completely randomized design. The analysis of split-plot designs, however, requires additional care as there are two error terms: between whole plots and within whole plot as depicted in Figure 2. In the current work, it is decided to include up to two-factor interactions while taking into account the split-plot structure of the design.

Whole plot to Whole plot Variation

WP	Material	Color	
1	ABS 1	Red	
2	ABS 1	White	
3	ABS 2	Red	
4	ABS 2	White	
5	ABS 2	White	
6	ABS 2	Red	
7	ABS 1	Red	
8	ABS 1	White	

Mold Temperature	Holding Pressure
20	Low
20	Low
20	High
20	High
50	Low
50	Low
50	High
50	High

Subplot to Subplot Variation

Figure 2. DoE structure showing the split between whole and subplot factors. The description of the factors can be seen in Section 3.

2.2. *Data*

This section contains an overall introduction to the structure of the data used in this article and a description of the two types of data generated in the experimental work and how it is used in the classification of the source of raw material.

2.2.1. Data structure

The overall structure of the data used in the current work can be seen in Figure 3. The analysis of the experimental study involves two responses as described in Section 3.2. Additionally, for each experiment, machine profiles and process data are both collected from the molding machine. The former is the full machine profiles describing the full dynamics of the molding process as described in Section 2.2.2, and the latter refers to readily available machine settings and aggregated process data, defined and generated by the molding machine producer, which is explained in further detail in Section 2.2.3. These will be used together with labels from the experimental factors in the classification of the raw material source described in Section 4.1.

In Figure 3, The experimental design run ID is the central and connecting ID in all data sources. The dimensions of the different data matrixes correspond to:

- n: Number of experimental runs as described in Section 3.
- m: Number of response variables in experimental investigation as described in Section 3.2.
- p: Number of factors in experimental investigation as described in Section 3.
- r: Number of injection molding cycles with data collection for each experimental setting[1].
- q: Number of collected injection molding cycles for the classification of raw material source $(n \times r)$

[1] r equals 100, since only the last 100 shots is collected for the classification of raw material source. The first cycles after changing the DoE settings is omitted since the molding process is very unstable at this point.

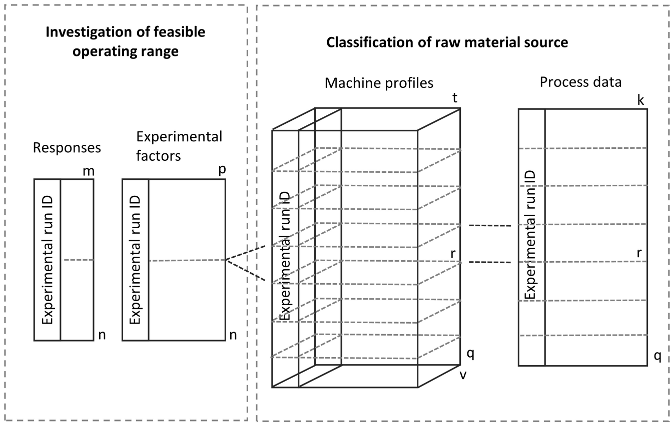


Figure 3. Data structure used in both experimental analysis and classification of the source of raw material.

- k: Number of collected process variables from the injection molding machine as described in Section 2.2.3.
- v: Time series profiles collected from essential part of the injection molding cycle as described in Section 2.2.2.
- t: Duration of collected time series profiles as described in Section 2.2.2.

2.2.2. Machine profiles

The machine profiles are continuous time series data extracted from internal sensors on the molding machines. The pressure and screw position sensors are the internal sensors used in current work (as proposed by Mao et al. (2018) and Nagorny et al. (2017)). The data from the screw position sensor are further processed to generate speed profiles. This results in three types of profiles (pressure, position and speed) which are collected for both the injection phase and the dosing phase as described in section 2.2.2, resulting in the collection of six profiles used for the classification of the material type.

The datapoints in each profile are collected for every 0.006 second. The profiles from the injection phase consist of 501 measurements, whereas the profiles from the dosing phase consist of 368 measurements. The decoupled collection of the profiles from the injection and holding phase is done to increase the resolution. It should be noted that a maximum of 501 points can be collected per profile. Examples of machine profiles can be seen in Figure 4.

The profiles are used with the classification methods described in Section 2.3. Each variable is used for classification individually in order to lower the cost of data collection in real life applications when the analysis method is to be implemented in actual production. The profile data is in the form of a v by q by t array, where v is the number of collected profiles, q is the number of captured cycles, and t represents the

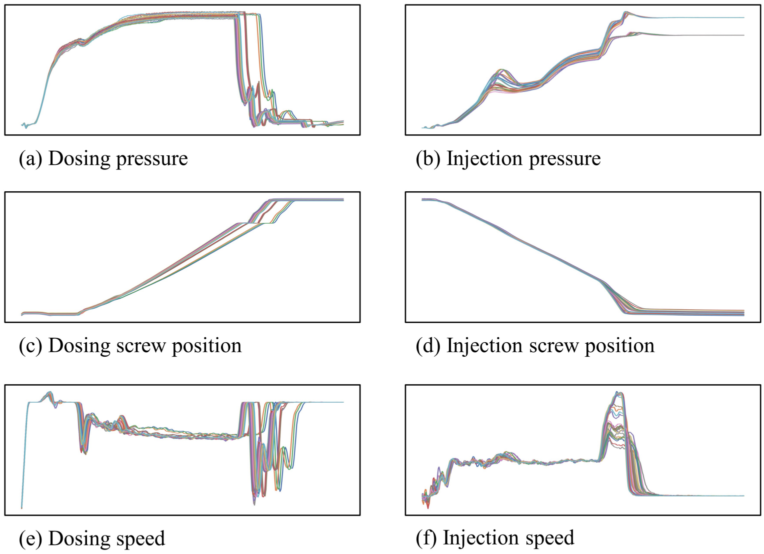


Figure 4. Visualizations of Machine Profiles. The pressure profiles in (a) and (d) is measured in bar. The speed profiles in (b) and (e) is measured in mm/s and the position profiles in (c) and (f) is measured in mm from a reference position.

duration. To bring the numeric values in the profiles to a common scale, a min-max normalization is applied, scaling the profiles to the range of [0,1]. The scaling is done as

$$\mathbf{Z}_{v,i} = \frac{\mathbf{X}_{v,q} - \min(\mathbf{Z}_v)}{\max(\mathbf{X}_v) - \min(\mathbf{X}_v)} \text{ for } i \leq q, \tag{1}$$

where $\mathbf{Z}_{v,i}$ is the scaled profile and $\mathbf{X}_{v,i}$ is the original profile.

2.2.3. Process data

The second dataset used for the source of raw material classification is the machine process data. These process variables are predetermined by the supplier of the molding machines. Most of these process variables are derived from the machine profiles. These could e.g., be maximum values, values at a specific time in the molding cycle or difference in screw position in the molding cycle. These process variables have been selected by the molding machine producer to give a good representation of the underlying process dynamic captured in the profiles. Besides this, the process data consists of time constants (e.g., injection time, dosing time and cooling time), barrel temperatures, energy consumption, etc. In this work a subset 28 of these process variables are selected for the analysis. These process variables can be categorized as temperature, pressure, time or position related data.

The data is structured in a q by k array, with q being the number of cycles and k

being the selected process variables. This data is to be used in machine learning and therefore scaled to zero mean and unit variance, as

$$\mathbf{Z}_j = \frac{\mathbf{X}_j - \hat{\mathbf{X}}_j}{\sigma(\mathbf{X}_j)} \text{ for } j \leq k, \tag{2}$$

where $\mathbf{Z}_j$ is the scaled feature column vector, $\mathbf{X}_j$ is the original feature vector, $\hat{\mathbf{X}}$ represents the mean of $\mathbf{X}$ and σ is the standard deviation.

As described earlier, from the implementation of the analytics methods perspective, the main difference between machine profiles and process data is that while the former is available in the current study, in general only process data is available. Obtaining machine profiles for all injection molding machines will in most cases require considerable investment that needs to be clearly justified with the information gain through this data.

2.3. *Classification methods*

This section describes the three methods used for the classification of the source of the raw material. The machine profile and machine process data will be used as predictors in the classification, and the material variable from the experimental study as described in Section 3 will be used as the response.

2.3.1. PLS-DA

PLS is a chemometrics method that transitions the input data ($\mathbf{X}$) into a latent space to reduce collinearity and noise. PLS is an extension to Principal Component Regression (PCR), where PCR first translates the $\mathbf{X}$ matrix to a reduced uncorrelated latent space that can be used for linear regression for the response variable $\mathbf{y}$. The drawback of PCR is that the response is not considered in the reduction of $\mathbf{X}$ into the latent space and that useful information that correlates $\mathbf{X}$ to $\mathbf{y}$ can be lost. PLS overcomes this by letting $\mathbf{y}$ affect the decomposition of $\mathbf{X}$ so that the latent variables chosen also possess the highest correlation with $\mathbf{y}$.

PLS is a common model used to fit cross-correlated multivariate data with many examples of implementations that can be found in the literature. One example is Zeaiter, Knight, and Holland (2011) who use PLS in injection molding on cavity sensor signal to predict the element dimensions. In this study, we use the machine profiles in a classification problem to predict the source of raw material.

PLS was originally intended for regression problems. The variation of the method used as a classifier in this paper is PLS-DA (Partial Least Squares Discriminant Analysis, Brereton and Lloyd (2014)), where the predicted real numbers are translated into one of two classes. For a two-class classification, the two classes are assigned to 1 or 0 and a standard PLS model is generated. When classifying new samples, the prediction will be a value on a continuous scale. The evaluation of the result in relation to the two classes is done by setting a threshold (e.g., 0.5) and all prediction above 0.5 is assigned to class 1 and predictions below 0.5 are assigned to class 0. The threshold can be set in different ways, but in the current work 0.5 is used.

2.3.2. AdaBoost

AdaBoost (Adaptive Boosting) algorithm is not a classifier in itself, but more precisely a method where weak classifiers are combined into a single strong classifier, by reweighing the contributions of each individual classifier. By combining these weak learners, it is possible to introduce non-linearity into the classifiers, which may result in a more expressive and complex classifier than the simple linear parts from which the strong classifier is build see Duchi (2016). In this study, the chosen weak classifiers are small decision trees with a maximum depth of 3.

The AdaBoost algorithm has previously shown great results in binary classification problems. Charest, Finn, and Dubay (2018) use AdaBoost with decision trees on profile data from the injection molding process in order to classify the molded element into one of three classes: Good part, Poor Quality or Short shot. Nagorny et al. (2017) use profiles from injection machines on AdaBoost as a regression method to predict the element dimensions. In that study, the AdaBoost algorithm performed better than the other tested boosting methods but was surpassed by the deep learning methods.

2.3.3. Deep learning

Deep Learning methods are well known for their ability to capture non-linear relationships between the inputs and the response. Feed Forward Neural Nets (FNN) is used as the deep learning classifier to compare against AdaBoost and PLS. FNN has previously been applied in injection molding with great success. Costa and Ribeiro (1999) use machine process data and profiles from the molding machine as input to FNN as a process monitoring system. Ali and Yu-To Chen (1999) use a dataset similar to the machine process data used in this study, as input to FNN for mapping the element dimensions of a molded element. Even though these examples have a different problem from the classification in this study, it shows that the FNN is able to extract information from the similar input data sets.

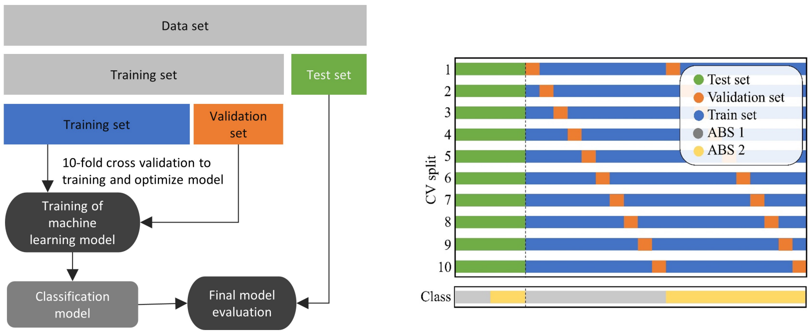
2.4. *Training and test split*

In any predictive model building study, the trained model is ultimately tested against the data (a.k.a. test data) that model has not seen before Xu and Goodacre (2018). Consequently, it is a common practice to split the available data into a training set used for training the model and a test set used for establishing how well the model performs on unseen data. This is illustrated in Figure 5 (a). It is therefore important to ensure that the test set is independent of the training set.

In the current situation, where the data is generated in a designed experiment, extra attention is needed. As illustrated in Section 2.2.1, data is collected from multiple injection cycles (r) for each experimental run (n). Since the r cycles are coming from the same experimental run, they are correlated and must be considered as one group when splitting the data into a training and test set. If the data (q) is seen as one pool of data and randomly split, there is a high risk that data from the same experimental run will end in both the training and testing set. To ensure the validity of the test set, it is therefore decided to do a test split based on unique experimental runs resulting in n groups of data.

To ensure even distribution of the raw material source in both the training and test set, as illustrated in Figure 5 (b), the splitting into train and test is done with a stratified split. The full dataset is first split in two subsets; one with all ABS1 and one with all ABS2. 10% of the experimental runs is then randomly drawn from each subset and collected to one test set. The remaining experimental runs are collected in the training set. Similarly, the validation set is for validating the models during training drawn with a stratified 10-fold cross validation split (90/10% split in training and validation set), again to ensure even distribution of ABS1 and ABS2 in both datasets.

To be able to generate a confidence interval on the test performance, the data split illustrated in Figure 5 (a) is repeated 5 times with a stratified 5-fold split, on the whole data set, ensuring 5 unique test sets, covering all experimental runs, and the model performance is reported as the mean model performance and confidence interval based on the variation between the 5 repetitions.



(a) Overall data flow in model training, validation and evaluation

(b) Illustration of stratified splitting in test and validation

Figure 5. Illustration of splitting the data with visualization of the proportion of material type in the sets.

2.5. *Evaluation of the methods*

The classification methods are validated with 10-fold cross-validation. The quality of the classifications is measured with the accuracy measure as following:

$$Accuracy(\hat{\mathbf{y}}, \mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{c}_i, \tag{3}$$

$$\mathbf{c}_i = \begin{cases} 1, & \text{if } \hat{\mathbf{y}} = \mathbf{y} \\ 0, & \text{if } \hat{\mathbf{y}} \neq \mathbf{y} \end{cases} \tag{4}$$

where $\hat{\mathbf{y}}$ is the predicted class, $\mathbf{y}$ is the actual class and n is the number of samples in the dataset.

Since the accuracy only explains the proportion of correctly classified, the test set is further evaluated with a confusion matrix to give a more detailed analysis of the classification. The confusion matrix consists of four fields; true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The TP and TN fields contain the instances of the correctly classified labels. Likewise, the FP and the FN contain the instances of misclassified labels, where FP contains the instances that are misclassified as the positive class, and FN contains the instances that are misclassified as the negative class.

3. Experimental setup

As mentioned in the Introduction, the aim of the experimental work is to investigate the feasible operating range for two selected response factors (described in section 3.2) with varying source of raw material. Besides the source of raw material, three other factors (color, holding pressure and mold temperature) have been identified as relevant to include in the experimental study:

- ABS is used as the plastic material in this experiment and the variation is introduced by using ABS from two suppliers (ABS1 and ABS2). The ABS from both sources are considered the same quality grade where the quality of both is within the specified limits (evaluated from randomly sampled quality inspections).
- The molded part is produced in different colors. The color is included to test for the potential interaction between material and color. To cover the variations in colors used in production, an organic color (red) and an inorganic color (white) are used in the experimental study.
- Holding pressure is the most important control factor used to adjust the molding process to ensure consistent quality. The mold used is optimized to run with a specific holding pressure with an allowed variation of +/- 70 bar. This range is therefore used as levels for this factor.
- The mold temperature is normally kept constant with only very small variation (+/- 1 °C). Despite this, the mold temperature is included in the experiment because it is known to have an important interaction with material and product quality and therefore can potentially be used as a control variable. The temperature is varied between 20 °C and 50 °C, where 35 °C is its normal operating setting.

Of the four factors included in the experiment, raw material and color are hard to change since changing the levels of these factors demand a cleaning of the grinder, material mixing unit and molding machine. To reduce the experimental cost, it is therefore decided to conduct the experiment using a split-plot design with raw material and color as hard-to-change factors and holding pressure, mold temperature as easy-to-change factors. Both raw material and color are discrete factors with two levels, whereas holding pressure and mold temperature are continuous and are tested at two levels. The design used is a replicated split plot design in the form of $2^2 \times 2^2$ (illustrated in Figure 6) resulting in 32 test runs.

3.1. *Experimental procedure*

For the experiments, an Arburg allrounder 470 S, 600-290 with a 4-cavity mold is used. The experiment was conducted over a period of eight days, running one of the eight

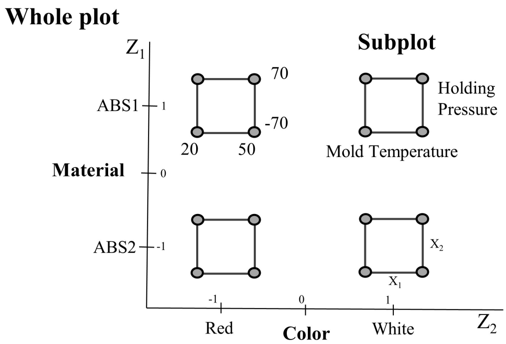


Figure 6. The whole plot is shown at the outer axes and the subplot at each whole plot combination. The subplot is run in all four experimental combinations in the whole plot.

whole plots each day. The molding machine was first cleaned and then connected to a mixing unit containing the desired ABS type and color. The molding process was started and run for one hour using the standard settings, which are the center settings to ensure stable production with a steady state of reground inlet material in the virgin material used in the molding process. The experiments are started only after ensuring that the molding process is stable. Results from an initial test showed that the molding process was stable after 120 molding cycles (shots) for a new experiment. It was therefore decided to let the molding machine run for 140 shots after each experiment before the products were collected (illustrated in Figure 7). Products from 8 consecutive injections are collected. Once the products were retrieved, the next experiment is started.

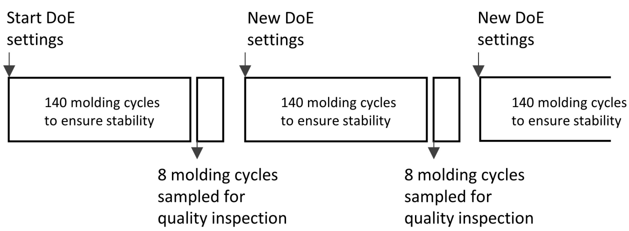


Figure 7. Illustration of the experimental procedure used.

From each of the collected samples consisting of 8 consecutive injections, three elements are randomly collected from each of the four cavities. It should be noted that the cavity numbers are visible on the element. Geometric properties of the 12 elements collected following the described scheme are obtained.

3.2. *Molded part and quality measures*

The injection molded element used in the test is depicted in Figure 8. Four different geometric properties are obtained using a Coordinate Measuring Machines from Zeiss. These metrics are expected to define the "quality" of the element. The average inside width in the middle of the element (InWmid) is calculated as the average of B1 and B2, and the average length of the element (OutL) is calculated as the average of A1 and A2 as depicted in Figure 8. These two average measures will be used as responses in the analysis of the experiments. The size of the element is 33.45mm $\times$ 17.40mm $\times$ 11.50mm, and it has a supporting rib in the center of the element to reduce warpage.

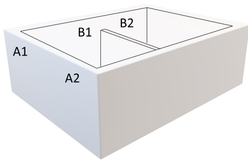


Figure 8. 3D illustration of the small rectangular box with identification of the four selected measuring points that represent the element quality.

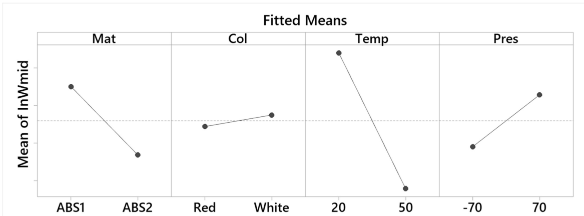
4. Results

The focus in the first part of the analysis is to fit and evaluate a model of the experimental settings against the two quality responses. The residual analysis shows no violation of the model assumptions. ANNOVA table for the analysis of the experimental data is given in Table 1. It is concluded that all main effects except for color are identified as significant at 5% level. One of the 2-factor interactions (pressure*temperature) is also found significant at the same significance level.

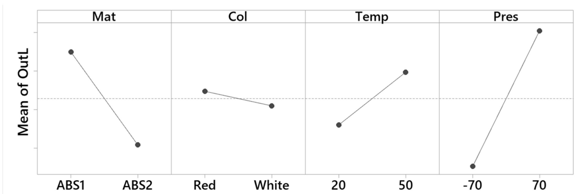
		InWmid				**OutL**			
Source	DF	Adj SS	Adj MS	F-Value	P-Value	Adj SS	Adj MS	F-Value	P-Value
Mat	1	0,00264	0,00264	95,71	**0,001**	0,00462	0,00462	70,94	**0,001**
Col	1	0,00008	0,00008	2,73	0,174	0,00011	0,00011	1,73	0,258
Mat*Col	1	0,00001	0,00001	0,48	0,526	0,00001	0,00001	0,1	0,768
WP Error	4	0,00011	0,00003	3,81	0,022	0,00026	0,00007	13,35	0
Temp	1	0,01041	0,01041	1435,97	**0**	0,00148	0,00148	303,64	**0**
Pres	1	0,00151	0,00151	208,88	**0**	0,00985	0,00985	2018,48	**0**
Mat*Temp	1	0,00003	0,00003	4,1	0,059	0,00000	0,00000	0	0,962
Mat*Pres	1	0,00000	0,00000	0,25	0,623	0,00001	0,00001	1,64	0,217
Col*Temp	1	0,00002	0,00002	2,21	0,155	0,00001	0,00001	1,57	0,228
Col*Pres	1	0,00000	0,00000	0,07	0,795	0,00000	0,00000	0,04	0,844
Temp*Pres	1	0,00007	0,00007	10,02	**0,006**	0,00001	0,00001	1,03	0,325
SP Error	17	0,00012	0,00001			0,00008	0,00001		
Total	31								

Table 1. Analysis of variance for InWmid to the left and OutL to the right. Mat is the material, Col is the color, Temp is the mold temperature and Press is the holding pressure. P-values for all significant terms are depicted in bold.

Furthermore, from the main effects plots in Figure 9, it can be seen that material and pressure have similar effects on both responses whereas the mold temperature has different effects in direction for the two responses. The different behavior for OutL and InWmid is attributed to different shrinkage/warpage of the element caused by the supporting rib in the middle of the element. The variation caused by using ABS from two different suppliers has the same effect on both quality measures. ABS2 results in reduced dimensions compared to when ABS1 is used. This can be explained by difference in shrinkage properties for the two materials.



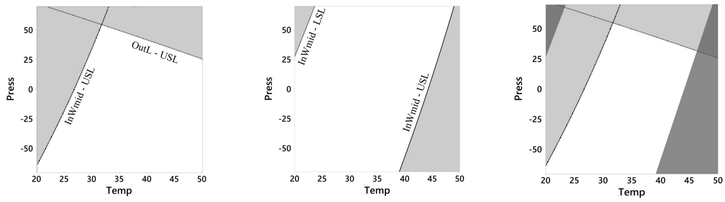
(a) Main Effects Plot for InWmid



(b) Main Effects Plot for OutL

Figure 9. Plot of main effect from the analysis of the split plot design.

As in any experimentation in production, we need to consider statistical significance against practical significance. For that we introduce the specifications for the quality measures as contour lines in a contour plot, see Montgomery (2012). The white areas in the contour plots in Figure 10 represent the feasible operating range for pressure and mold temperature. The gray areas represent operating condition causing quality to be outside the specifications. The difference in feasible regions for ABS1 and ABS2 can be seen in Figure 10(a) and (b), where the feasible operating range is largest for ABS2. Running with the standard mold temperature of 35 °C the whole operating range for pressure is available when using ABS2, where only the low to medium pressure range is available for ABS1 at that temperature. This is an important insight since it is currently unknown which material the molding machine is using. Therefore, since the temperature is often kept at 35 °C and the pressure is a control parameter for the operators to overcome process related issues, they are also risking to produce out of specification products when setting the pressure at a high level. When the source of the raw material is unknown, the feasible operating region should reflect the restrictions associated with that uncertainty as depicted in the double overlaid contour plot in Figure 10(c). This limits the operator's ability to compensate for other process disturbances using pressure as control variable. If the full operating area should be available at operator's disposal, the source of the material has to be known or predicted at any time during the production.



(a) ABS1 (b) ABS2 (c) ABS1 and ABS2 contour lines overlaid in same contour plot

Figure 10. Overlaid contour plot of InWmid and OutL. The contour plot is showing pressure vs. temperature where color is held constant at red since mostly organic colors are used and switch-over point at 99% filling.

4.1. *Classification of the source of the raw material*

From the analysis of the data in the previous section, it becomes clear that the identification of the source of the raw material is essential for performing real-time adjustments to machine settings to ensure uniform product quality. The focus in the following is therefore to investigate if data from the molding machine can be utilized to identify which source of raw material is used at any given time during the production The description of the model and the data used for this purpose can be found in Section 2.2.

The training of the classification models uses 80% of the data generated from the experiments, where the remaining 20% is used for testing as described in Section 2.4. The model performance for the 3 different models, PLS-DA, AdaBoost and FNN, and the 7 different data sources is compared in Figure 11.

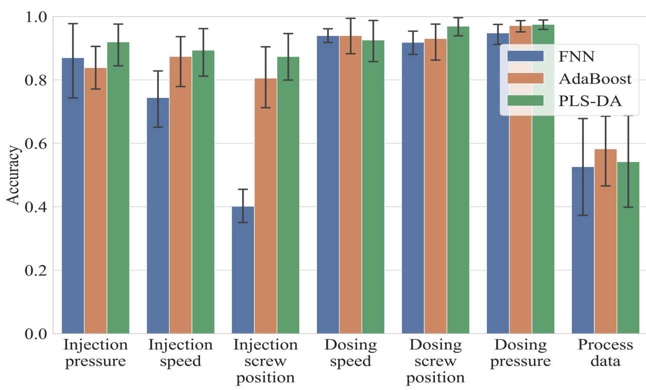


Figure 11. Evaluation of mean accuracy for classification methods on material type. A classification accuracy of 0.5 corresponds to a random guess.

The best results are obtained with PLS-DA on the dosing pressure profile with 97.5% average accuracy on the 5 test splits. The dosing pressure profile in general contains more information on the material source than the other variables, since all three models perform well using this variable, with FNN being the lowest with 94.7% average accuracy. The intuition behind the dosing pressure profile performing well is that the dosing pressure is a result of the screw rotation, conveying material in front of the screw, and that this could be affected by material viscosity. Variation in viscosity is a known variation in different sources of ABS.

Considering the two data sources, machine profiles and process data, it is most desirable to utilize the process data for the classification since this dataset is readily available on the molding machines with lower complexity and size. In Figure 11, the process data performs poorly, with an average accuracy around 50%, which corresponds to a random guess, and with high deviation between the test splits. This is somewhat surprising since the process data should contain the key features of all 6 machine profiles plus additional machine sensor data, e.g., barrel temperature and machine settings. This indicates that the variation in the material properties is represented in the dynamic shape of the profiles and not in the key features.

Figure 12 shows the confusion matrices on the test set for the three selected models with all test splits considered. PLS-DA on the dosing pressure profile has 2.38% misclassified observations with 1.21% false positives and 1.17% false negatives. AdaBoost on dosing pressure profile misclassified 1.43% as false positives and 1.24% as false negatives. Finally, FNN on dosing pressure profile has 0.15% false positives and 4.93% false negatives.

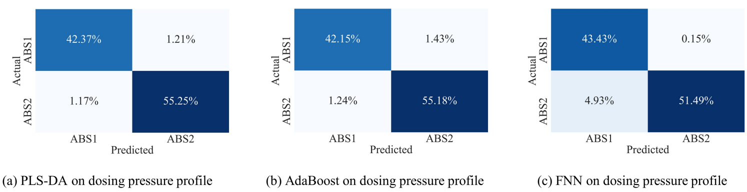


Figure 12. Confusion matrices on test data for best performing method and data combination.

Any of these three models using the profile data of the specified variable can be used to identify the source of raw material in production and hence help the operator use the right operating window for adjusting the pressure to avoid production problems.

5. Discussion

The overall scope in this work has been to investigate how to mitigate quality issues caused by variation in raw materials (in our case, when dual sourcing ABS material from two suppliers). The proposed approach was to utilize injection molding machine sensor signals to identify the origin of the raw material used in the molding machine and ultimately use this information to adjust the molding process to compensate for the differences in material properties.

PLS (Partial Least Squares), AdaBoost (Adaptive Boosting) and FNN (Feed-Forward Neural Net) have been used for classifying the source of the raw material. It is shown that PLS-DA performed the best in predicting the source of raw material. This classification offers very valuable information in production since it makes it possible for the operators to adjust the process settings, i.e. holding pressure and even mold temperature, depending on the source of raw material used, and thereby ensure satisfactory production. A combination of temperature and pressure that satisfies the specifications of both types of material exists, but by knowing the source of the raw material, the operators are able to utilize the full individual process window to adjust the process settings to ensure produced elements within specifications and safeguard against other production concerns that are beyond the scope of this study.

It should be noted the source of raw material was well determined in the experimental setup. This is, however, not the case in the real production, where dual-sourced raw material is stored in the same silo. This also means that there will be a transition period where a combination of the two types of raw material is used in the molding machine. We did not include this classification in our study. It could, however, be done in an extended study, where experiments with different combinations of the two types of raw material are used and learning models are trained to predict the percentage of either material. This would introduce increased complexity to the problem but would be a more precise representation of the production. This is, however, beyond the scope of this work where we were also aiming at laying down the groundwork for upcoming investigations.

Demonstrating that injection molding machine profiles can be used to differentiate between the same grade of ABS material from two suppliers has sparked the idea that it might be possible to utilize this further. One potential case is to use the machine profiles to detect changes in raw material properties over time (from delivery to delivery). Another could be to investigate if the profiles can be used to detect change in the material water content which is a critical material quality attribute. Once again, these are beyond the scope of the current study.

6. Conclusion

The main objective in this paper was to develop real time data analytic models that will help mitigate the impact of the variation in a nuisance factor in production. It can therefore be seen as a process robustness study where we specifically investigate the possibility of using injection molding machine signals to classify the source of raw material used in the molding machine. This has been proven possible using the dosing pressure profile and PLS-DA (Partial Least Squared Discriminant Analysis). Through experimental work, it has also been shown that the holding pressure and mold temperature could be used as control variables to center the molding process within the specifications of two selected quality measures. The proper classification of the source of raw material through the proposed classifier allows the operator to gain the vital information on the actual operating window for the adjusted variables such as pressure and mold temperature. Furthermore, the experimental work and predictive modeling performed in this study have initiated a very timely conversation and interest in data driven methods and their potential use not only in extending the current work but also in other aspects of the production process. The same approach and considerations could be utilized for other industries/applications, where an indirect measure could be used to identify the variation in a nuisance factor(s) during production and based on this, change critical process variables to compensate for the detected variation.

References

Ali, I. G., and Yu-To Chen. 1999. "Design quality and robustness with neural networks." *IEEE Transactions on Neural Networks*10 (6): 1518–1527.

Brereton, Richard G., and Gavin R. Lloyd. 2014. "Partial least squares discriminant analysis: taking the magic away." *Journal of Chemometrics*28 (4): 213–225. https://onlinelibrary.wiley.com/doi/abs/10.1002/cem.2609.

Charest, M., R. Finn, and R. Dubay. 2018. "Integration of artificial intelligence in an injection molding process for on-line process parameter adjustment." In *2018 Annual IEEE International Systems Conference (SysCon)*, April, 1–6.

Chen, Wen-Chin, Pei-Hao Tai, Min-Wen Wang, Wei-Jaw Deng, and Chen-Tai Chen. 2008. "A neural network-based approach for dynamic quality prediction in a plastic injection molding process." *Expert Systems with Applications*35 (3): 843 – 849. http://www.sciencedirect.com/science/article/pii/S0957417407002977.

Costa, N., and B. Ribeiro. 1999. "Artificial neural networks for data modelling of a plastic injection moulding process." In *ICONIP'99. ANZIIS'99 ANNES'99 ACNN'99. 6th International Conference on Neural Information Processing. Proceedings (Cat. No.99EX378)*, Vol. 3, 1081–1087 vol.3.

Duchesne, C., and J.F. Macgregor. 2004. "Establishing Multivariate Specification Regions for Incoming Materials." *Journal of Quality Technology*36 (1): 78–94. Cited By 34.

Duchi, John. 2016. "Boosting." CS229 Supplemental Lecture notes.

Farahani, Saeed, Nathaniel Brown, Jonathan Loftis, Curtis Krick, Florian Pichl, Robert Vaculik, and Srikanth Pilla. 2019. "Evaluation of in-mold sensors and machine data towards enhancing product quality and process monitoring via Industry 4.0." *The International Journal of Advanced Manufacturing Technology*105 (1): 1371–1389. https://doi.org/10.1007/s00170-019-04323-8.

Gao, Robert X., Xinyao Tang, Guthrie Gordon, and David O. Kazmer. 2014. "Online product quality monitoring through in-process measurement." *CIRP Annals*63 (1): 493 – 496. http://www.sciencedirect.com/science/article/pii/S0007850614000444.

Kowalski, S., P. Parker, and G. Geoffrey Vining. 2007. "Tutorial: Industrial Split-plot Experiments." *Quality Engineering*19: 1 – 15.

Lopes, Noel, and Bernardete Ribeiro. 2000. "Part Quality Prediction in an Injection Moulding Process Using Neural Networks." .

Mao, Ting, Yun Zhang, Yufei Ruan, Huang Gao, Huamin Zhou, and Dequn Li. 2018. "Feature learning and process monitoring of injection molding using convolution-deconvolution auto encoders." *Computers & Chemical Engineering*118: 77 – 90. http://www.sciencedirect.com/science/article/pii/S0098135418307026.

McLauchlin, Andrew R., Oana Ghita, and Ali Gahkani. 2014. "Quantification of PLA contamination in PET during injection moulding by in-line NIR spectroscopy." *Polymer Testing*38: 46 – 52. http://www.sciencedirect.com/science/article/pii/S0142941814001330.

Montgomery, Douglas C. 2012. *Design and Analysis of Experiments.* 8th ed. Wileyr. International Student Version.

Nagorny, P., M. Pillet, E. Pairel, R. Le Goff, J. Loureaux, M. Wali, and P. Kiener. 2017. "Quality prediction in injection molding." In *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 141–146.

Xu, Yun, and Royston Goodacre. 2018. "On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning." *Journal of Analysis and Testing*2.

Zeaiter, Magida, Wendy Knight, and Simon Holland. 2011. "Multivariate regression modeling for monitoring quality of injection moulding components using cavity sensor technology: Application to the manufacturing of pharmaceutical device components." *Journal of Process Control*21 (1): 137 – 150. http://www.sciencedirect.com/science/article/pii/S0959152410002131.

CHAPTER 7

Acoustic emission for monitoring

After a year of working with the utilisation of injection moulding process data, I presented some of the findings to a moulding specialist at the industrial partner. As part of the presentation, it was discussed how to best support the operators based on results from the real-time process monitoring. There was consensus that monitoring the process and notifying the operators when the process was out of control could add value. It was then discussed what type of diagnostics needed to be included and that it might be limited what could be achieved using the collected process data. The moulding specialist then said, ”often, we troubleshoot the moulding machine by listening to it”. It was then discussed what type of diagnostics an operator could provide by listening to the moulding machine and that it could be beneficial if these diagnostics could be automated and used in real-time to monitor the moulds and moulding machines. It was at this point decided to investigate acoustic monitoring of moulding machines as part of the PhD project.

This chapter will introduce some basic concepts within the analysis of acoustic signals to support the work done in paper C. There will be a short introduction to; acoustic signals, recording of acoustic emissions, signal treatment, some examples of feature generation and how this can be used in machine learning applications. This will all be centred around acoustic emissions from injection moulding. Paper C presents work on utilising acoustic emission from the injection mould for condition-based maintenance. Besides this work, it has also been investigated to what extent acoustic emissions from the injection moulding machines can be used for process monitoring. Because of confidentiality, this work has not been published. The work has now reached a stage where publishing part of the concept is possible. It will be introduced at the end of this chapter since it is relevant for the overall conclusion on utilising and evaluating different data sources for the same application.

7.1 Acoustic emissions

In its basic form, acoustics are vibrations propagated in, e.g. air where it will reach the human ear as differences in air pressure. These differences in air pressure make the eardrum vibrate. The vibrations are turned into signals transported to the brain and interpreted as sound. Humans can register air pressure changes vibrating at up to 20 kHz (20,000 times per second). Therefore, this is often defined as acoustics (what humans can perceive). Vibration above 20 kHz (up to several gigahertz) is defined as ultrasound. Some animals can perceive ultrasound above 100 kHz (e.g. a bat and a beluga whale). Since humans are limited to sensing vibration below 20 kHz, a lot of research has been dedicated to this frequency range. Research areas within acoustics include recording of acoustics, sound absorption in different materials (for noise reduction), speech recognition etc.

A vibrating source (e.g. moulding machine) generates waves transmitted in the machine structure and radiated to the surroundings. Vibrations propagated in a solid material is often referred to as structure-borne sound or vibro-acoustics. This type of vibration is utilised in work made on condition-based maintenance. The acoustics are collected using a contact microphone attached directly to the metal surface of the injection mould and moulding machine. The same vibrations will be transferred from the metal surface to the surrounding air, making it detectable by humans. The microphones used for the initial work are contact microphones designed to record acoustics from musical instruments. Therefore, they are optimised to collect acoustic vibration with a frequency below 22 kHz. Compared to room microphones (recording vibration in the air), an advantage of using contact microphones is that surrounding noise would not impact the recordings.

When collecting the acoustic emissions, it is essential to consider the sampling frequency or rate (signal readings per second). The acoustic emissions are continuous vibrations. It must be discretised and saved as discrete samples to record and store this continuous vibration signal digitally. Therefore, the sampled and saved data will be a discretised representation of the continuous signal. As a consequence of this, it is essential to consider how the signal is sampled. Acoustic vibrations can be seen as a combination of different sine waves that generate the resulting vibration in the microphone, collecting the acoustic vibrations. For illustrative purposes, let's assume that a given acoustic vibration only consists of vibrations at a single frequency (the black sine wave in Figure 7.1).

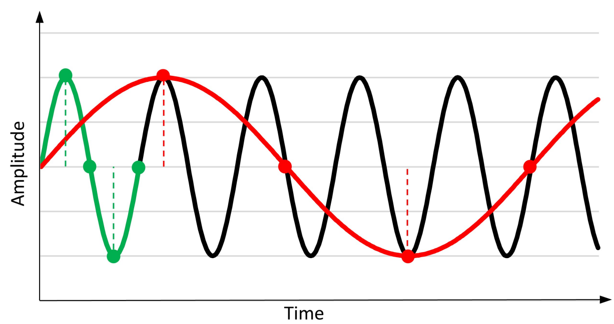


Figure 7.1. Sine wave for illustration of Nyquist Frequency and Aliasing. The black sine wave is the emitted acoustic wave. The green wave represents this using a sampling rate four times the frequency of the emitted acoustic wave. The red wave represents this using a sampling rate of 1.25 times the frequency of the emitted acoustic wave.

If this signal is sampled at four times faster than the signal frequency (vibrations per second), it is seen that good approximations of the original wave can be reproduced (the green sine wave in Figure 7.1). The red sine wave in Figure 7.1 is achieved when sampling with a rate just above the wave frequency. The resulting approximation of the underlying wave will then be different from the original wave. This phenomenon is called aliasing and was initially described by Harry Nyquist in 1928 (formulating the Nyquist Theorem). The Nyquist Theorem state that the sampling rate has to be greater than twice the highest frequency that has to be reproduced. The frequency at half the sample rate is often called the Nyquist frequency. A sampling rate at 44.1 kHz or higher is often used when recording human-perceivable acoustics to avoid aliasing.

7.2 Time and frequency domain

When considering an acoustic recording, the first representation is often a plot of signal amplitude as a function of time (Figure 7.2, B). The wave in Figure 7.2 B is constructed of the three individual sine and cosine waves in Figure 7.2 A. The three individual waves are constructed with a specific amplitude and frequency (220 Hz, 440 Hz and 880 Hz). This amplitude representation (time-domain representation) of the signal is not that informative since it only represents the loudness of the recording as a function of time. Analysing an acoustic recording concerning a specific application (e.g. condition-based maintenance), it is interesting to identify what underlying vibration patterns are present in the signal and how these change over time (a faulty state might contain vibrations at a different frequency than a faultless state). To achieve this, it is necessary to transform the signal into the frequency-domain (frequency-domain representation), which will reveal the different frequencies present in the signal. This transformation can be done using Fourier Transform. Discrete Fourier Transform (DFT) is a mathematical concept that can decompose a discrete signal into its constituent frequencies. The underlying assumption is that any function can be represented as a combination of cosine and sine functions. The discrete Fourier transform is given by: [72]:

$$\hat{f}_k = \sum_{j=0}^{n-1} f_j e^{-i2\pi jk/n} \tag{7.1}$$

and the inverse of the function is (iDFT) given by:

$$f_k = \frac{1}{n} \sum_{j=0}^{n-1} \hat{f}_j e^{i2\pi jk/n} \tag{7.2}$$

Applying the DFT to an input vector f, will result in an output vector of Fourier coefficients $\hat{f}_k$ for each frequency:

$$\{f_0, f_1, ..., f_n\} \xrightarrow{\textbf{DFT}} \{\hat{f}_0, \hat{f}_1, ..., \hat{f}_n\} \tag{7.3}$$

This can be computed using matrix multiplication, where $\omega_n = e^{i2\pi/n}$ is the fundamental frequency:

$$\begin{bmatrix} \hat{f}_0 \\ \hat{f}_1 \\ \hat{f}_3 \\ \vdots \\ \hat{f}_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega_n^1 & \omega_n^2 & \dots & \omega_n^{n-1} \\ 1 & \omega_n^2 & \omega_n^4 & \dots & \omega_n^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega_n^{n-1} & \omega_n^{2(n-1)} & \dots & \omega_n^{(n-1)^2} \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_3 \\ \vdots \\ f_n \end{bmatrix} \tag{7.4}$$

Fast Fourier Transform (FFT) is an optimised implementation of the DFT and is, therefore, referred to when Fourier transform is applied.

For illustrative purposes, the mixed wave (Figure 7.2 B) was repeated for 1 second and then sampled with a sampling rate of 20 kHz (creating an artificial acoustic recording). DFT was then applied to this discrete signal, and the result is plotted in Figure 7.2 C).

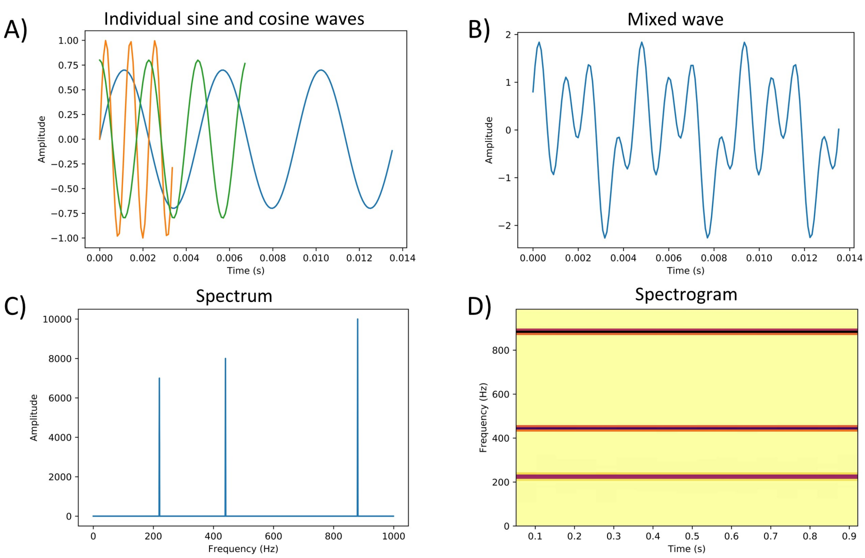


Figure 7.2. A) Three individual cosine and sine signals, blue - amplitude of 0.7, frequency at 220 Hz, green - amplitude of 0.8, frequency at 440 Hz and orange - amplitude of 1, frequency at 880 Hz. B) The three waves in A) are added together in a mixed wave. C) DFT applied to the mixed wave in B). D) spectrogram of a 1-second representation of the mixed wave in B)

From the spectrum, it can be seen that there are three dominant frequencies at 220 Hz, 440 Hz and 880 Hz in the signal (match the design of the mixed wave). The amplitude in the spectrum reflects the amplitude of the individual frequency contributions in the original signal. If the input signal is stationary, the spectrum in Figure 7.2 C) gives a good representation of the information in the frequency domain. A signal coming from rotating equipment (e.g. a bearing or motor running at the same speed) can be seen as stationary. The signals or acoustic emissions from an injection moulding machine is non-stationary since the frequencies change during the moulding cycle (caused by the different events occurring in the moulding cycle).

When the signal of interest is non-stationary, it is relevant to analyse the frequencies as a function of time. This can be achieved by applying a Gabor transformation[1] also known as Short-time Fourier transform (STFT) to the signal. The STFT can be seen as a moving window FFT, where FFT is applied on the windowed signal. The spectrums for the individual windows are then put together in a 2D time/frequency matrix. This can then be plotted as a spectrogram capturing the time/frequency relationship. The resolution of the spectrogram is determined by the windows' size and overlap between the windows. The concept of moving window FFT is illustrated in Figure 7.3 A, and the resulting spectrogram can be seen in Figure 7.3 B.

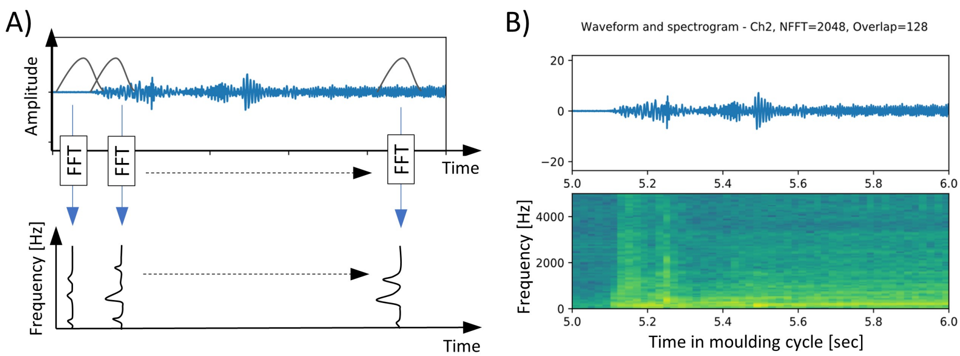


Figure 7.3. A) illustrates STFT as a moving window FFT, where the upper part is a spectrum, and the lower part is a plot of frequency as a function of time with one column for each window in the STFT. B) is an example of a spectrogram of the first second of the dosing phase in an injection moulding cycle (created using librosa [73] in python).

Figure 7.2 D is an example of a spectrogram for a stationary signal. More examples of spectrograms from injection moulding can be seen in paper C, Figure 4. Using a machine learning technique, utilising acoustic emissions for, e.g., event detection, requires some feature extraction/generation from the raw acoustic recording. Feature extraction approaches are often designed for specific applications. One approach commonly used within speech recognition is to generate spectrograms from the raw

[1]For information on Gabor transformation, see [72]

recordings and then use these as images in a deep learning model, as demonstrated by Prasomphan (2015, [74]) for detecting human emotion from speech recordings. Within speech recognition the scale on the spectrogram is often changed to Mel scale (Mel-Spectrogram or Log Mel-Spectrogram), since it changes the Hz scale into a scale that represents the human perceptions of sound [75]. The same approaches can to some extend also be used within machine maintenance applications. In paper C, we explored the use of spectral centroid as features for detecting faulty moulds.

7.3 Utilising acoustic emission in injection moulding

Different applications have been presented using acoustic emissions for monitoring the condition of injection mould and moulding machines. References related to this can be seen in the introduction of paper C. The following sections will present some ideas for potential acoustic applications beyond maintenance applications.

7.3.1 Process monitoring using acoustic emissions

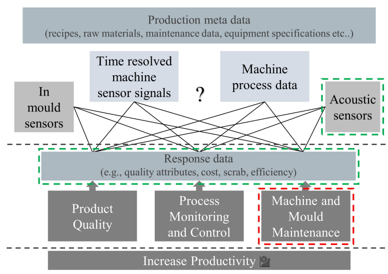
Based on the introduced data utilisation framework, where different data sources are explored across various applications, it has been questioned whether acoustic equipment and recordings could be utilised for more than maintenance applications. The initial exploration has not been published; hence only the overall ideas and concepts will be mentioned here. The experimental work presented in paper A and B has been repeated, using ABS from four different vendors. The machine pressure profiles were collected together with recordings of acoustic emissions from the injection unit of the moulding machine (trying to capture acoustic differences caused by changes in material properties). The predictive power of the machine pressure profiles for material classification have been confirmed in this study. Comparable classification results have been achieved using acoustic recordings. This is seen as an initial indication that the acoustic emissions from the moulding machine change as the material properties are changed.

7.3.2 Future work

Based on the results achieved in paper C and results from additional not published work, the industrial partner has decided to extend the exploration of acoustic emissions for monitoring within injection moulding. Acoustic equipment for 68 moulding machines has been purchased and will be installed in the spring of 2022. The equipment is designed for a manufacturing environment, and the sensors/microphones have a sampling rate of up to 1,000 kHz and, therefore, include the lower ultrasound range. The initial focus will be on condition-based mould maintenance and later extended to also explore condition-based moulding machine maintenance. Besides the aspect of condition-based maintenance, additional work will also include the use of acoustics for process monitoring (monitoring variation in material properties). This will be evaluated against the results achieved utilising machine profile data from the in-built pressure sensors (linked to the work done in paper A and B).

CHAPTER 8

Paper C, Utilization of acoustic signals from injection moulding for predictive maintenance



To keep up with the growing competition within the manufacturing industry, it is crucial to have a production setup that runs smoothly without unexpected breakdowns causing loss of productivity. This can only be achieved with reliable and well-maintained equipment. This is often accomplished using preventive maintenance, to ensure equipment is regularly maintained. The maintenance interval is based on experience, and for simplicity, the same interval is often used across different equipment of the same type. The maintenance interval is a cost/benefit consideration between spending too many resources on preventive maintenance (cost of conducting the maintenance and the lost productivity while performing the maintenance) and the cost of unplanned breakdown (cost of repair and lost productivity). An alternative approach to mitigating this evaluation is using predictive or condition-based maintenance where the condition of the equipment is monitored, and maintenance is performed when needed. Condition-based maintenance can be applied for both the injection mould and injection moulding machine within injection moulding. The most apparent maintenance challenges at the industrial partner are related to mould maintenance, which is reflected in research question 3.

Research question 3 (RQ3)

How to do real-time condition-based maintenance on injection moulds to reduce the cost of preventive maintenance and reduce the risk of mould malfunction?

Predictive or condition-based maintenance is a common topic within Industry 4.0 and data utilisation in general. The readily available equipment data have often been tailored with process monitoring and control in mind. As an effect, these data seldom reflect the maintenance state of the equipment. Therefore, exploration and implementation of condition-based maintenance approaches often require the installation of additional sensors or measuring devices. Çınar et al. (2020, [4]) present a broad overview of different techniques used within condition-based maintenance applications. The use of vibrations or acoustic emissions for condition-based maintenance is well represented in the overview. This have guided the formulation of RIO3 *"Investigate the use of acoustic emissions for condition-based mould maintenance"* and the work presented in paper C.

8.1 Summary

The overall focus is to propose a solution that can be developed on a subgroup of moulds and then scaled to new unseen moulds. This is desired since it is impossible to collect faulty data on new moulds and to reduce the modelling workload by having global models instead of models for all mould/machine combinations. Creating a global model, in general, requires data from multiple moulds containing both faultless and faulty operation. In the current work, the data have been limited to originate from five moulds with only a few faulty moulding cycles. This limitation initiated a generative modelling approach, where only the faultless moulding cycles are modelled since more data are available for this class and the signature of the faultless state is less complex. Useful features are generated using a feature extraction pipeline, consisting of Fast Fourier Transform of the acoustic signal, followed by time-averaging and PCA decomposition. To enhance the generalisation ability, a model adaption approach is proposed, where the mean and standard deviation of the first 14 moulding cycles after maintenance (assuming faultless state) is used to adjust the distribution of faultless state. Using this approach, acceptable classification accuracy is achieved.

Achieved results

- Using generative Gaussian modelling, it has been shown that acoustic emission from injection moulds can be used for detecting the need for lubrication and loose/defect latch lock.
- Implementing a simple model adaptation makes the proposed solution generalisable to new unseen moulds.

Contribution

- Exploration of acoustic emissions from five different injection moulds.
- A practical, effective and scalable modelling concept for achieving condition-based maintenance on injection moulds using acoustic emissions.

8.2 Paper C

Paper C

Utilization of acoustic signals from injection moulding for predictive maintenance

G. Ø. Rønsch[a], I. López-Espejo[b], D. Michelsanti[b], Y. Xie[b], P. Popovski[b] and Z.-H. Tan[b]

[a]Technical University of Denmark, Department of Statistics and Data Analysis, Lyngby, DK
[b]Aalborg University, Department of Electronic Systems, Aalborg, DK

ARTICLE HISTORY
Compiled December 7, 2021

ABSTRACT
Predictive and condition-based maintenance is given more and more attention to further optimize the utilization of manufacturing and production equipment. Utilizing acoustic signals for equipment monitoring and predictive maintenance has been proven effective within many applications. Many manufacturing and production setups consist of multiple alike machines (e.g., within injection moulding) where it would be beneficial to use the same monitoring setup and configuration on all machines. Based on an industrial application within injection moulding (using five different injection moulds), we propose a methodology utilizing acoustic signals from injection moulds combined with generative Gaussian or autoencoder modeling. To improve the generalization ability of generative modeling to moulds not seen at training time, we propose simple, yet effective, model adaptation, which only requires a few faultless moulding cycles at runtime/test time. The best results are obtained using the Gaussian model, where area under the curve values close to one are achieved when employing a model adapted to the specific mould at test time to detect abnormal situations like mechanical-defective moulds (loose latch lock) and the need for lubrication. The proposed framework is light in terms of computation and makes the setup implementation practically feasible in a real industrial context with multiple alike machines.

KEYWORDS
Acoustic signal processing; Predictive maintenance; Injection moulding; Industry 4.0; Machine learning; Autoencoder

CONTACT G. Ø. Rønsch. Email: geor@dtu.dk

1. Introduction

Industry 4.0 and data utilization are a new focus area in many companies, and work is carried out to explore how best to utilize data in the manufacturing environment to support improvement of quality and productivity. Many different objectives are explored, such as predictive quality by Schmitt et al. (2020), fault detection and classification by Park, Fan, and Hsu (2020), process monitoring by Weichert et al. (2019) and predictive maintenance by Çınar et al. (2020). The proposed applications range from utilizing "small" to "big" data and from existing equipment data to the use of new sensors to acquire additional data.

1.1. *Maintenance*

Maintenance is an important topic within manufacturing companies and optimization of existing maintenance procedures is often addressed within Industry 4.0. An effective maintenance setup is the foundation for keeping a high production throughput and to ensure a consistent and high product quality level. As stated by Salonen and Deleryd (2011), maintenance can, in general, be split into two categories: corrective and preventive. *Corrective maintenance* refers to the set of reactive procedures aimed to address failures that occur randomly. These failures are often caused by lack of preventive maintenance or poor equipment reliability. *Preventive maintenance* is performed proactively to uphold high equipment performance and thereby reduce the need for corrective maintenance. There is always a cost/benefit trade-off between the resources spent for preventive maintenance and the resulting cost of corrective maintenance. One way to address this trade-off is to adopt data-driven *predictive maintenance*, where the focus is on continuous monitoring of the equipment performance and detecting when maintenance is needed. As described by Spendla et al. (2017), the overall benefit from using predictive maintenance is that interventions can be conducted on an as-needed basis, thereby optimizing the availability of the manufacturing equipment.

The critical aspect of predictive maintenance is to identify equipment-related signals and data that can be utilized to detect reliably the need for maintenance. Within injection moulding there have been several proposals for predictive maintenance. Frumosu, Rønsch, and Kulahci (2020) demonstrated that some level of predictive maintenance can be obtained by using maintenance records, but this approach is not able to capture sudden mould failures. To be able to detect mould failures in real-time, it is necessary to do shot-by-shot evaluation of mould performance. Many of the signals that are available by default on a moulding machine are designed for monitoring the moulding process and machine performance. As demonstrated by Chen, Guo, and Wang (2020), machine data are widely used for online prediction of element quality. Park et al. (2016) have devised methods to map possible maintenance components in injection moulding machines and process parameters that have affected malfunctioning of these components. Based on this, they proposed to use machine parameters for predictive maintenance, but without showing that this approach could detect malfunction of the injection mould, e.g., defective latch locks, blocking of air vents or surface seizures on guide pins caused by lack of lubrication. The data used for predictive quality and process monitoring can only be used for an indirect measure of predictive mould maintenance and it is not even a guarantee that these signals will capture mould failures. Moreira et al. (2020) have justified that an additional pressure sensor and a three-axis accelerometer can be used for predictive maintenance monitoring. However, since data were only collected on productive cycle, it was not demonstrated that malfunctions could be detected using the proposed setup.

1.2. *Acoustic signals for predictive maintenance*

Acoustic signals and vibrations are used in many different industries for monitoring of equipment and prediction of machine failure. Biswal, George, and Sabareesh (2016) demonstrated that vibration data from accelerometers can be used for fault detection in wind turbines. Both Durbhaka and Selvaraj (2016) and Ahmad et al. (2019) showed that vibration data and machine learning can be used for fault detection of bearings. Using machine learning and features extracted from the frequency domain, Ha et al. (2018) demonstrated that acoustic signals from rotating machine gearbox can be used for detection of misalignment. Pan et al. (2017) presented a good introduction to the use of acoustic signals within IoT and Industry 4.0. The focus in that work was on the needed framework and different approaches utilizing acoustics for monitoring.

In general, the cases that describe the use of vibrations or acoustic signals for predictive maintenance develop and test the solution on only one equipment or production unit. It is therefore unclear if the solutions can be used across multiple similar production units, or if the solutions have to be tailored to the individual production units. When a production environment consists of multiple comparable production units, it is desirable that the concept and models can be utilized across different production units (reducing the number of models to develop and maintain). One industry where this is highly relevant is within injection moulding (see Section 2.1), where a production setup often consists of multiple injection moulding machines and injection moulds.

1.3. *Industrial application with injection moulding*

The current work has been conducted in collaboration with an industrial partner within injection moulding (see the illustration of an injection moulding machine in Figure 1). The production setup consists of 40 injection moulding machines and 200 injection moulds (approximately, 150 different element shapes). The production is characterized by small production orders running from days up to a couple of weeks. This means that there are many changeovers in the production where moulds are replaced in the moulding machines. Two brands of hydraulic injection moulding machines are used: Engel and Arburg. The moulds used are standard three-plate moulds and all moulds can be used in all the moulding machines.

The overall scope of the current work is therefore to explore how best to utilize acoustics for predictive maintenance in a setup consisting of multiple comparable injection moulding machines with frequent change in the equipment configuration (change of injection mould). The aim is to come up with a setup that is simple and only needs a minimum of configuration (if any) when moved from machine to machine or mould to mould.

The contribution of the current work is a demonstration of a combined use of acoustic signals, digital signal processing and machine learning for condition-based maintenance (or predictive maintenance) within injection moulding. The new approach needs to generalize well to reduce the complexity of introducing the solution into a manufacturing setup consisting of more than 200 comparable production units. The scope will be limited to only focusing on lack of lubrication and loose/defective latch locks. The proposed solution also has to generalize enough to be used within applications other than injection moulding. With all of these considerations in mind, in this paper, we explore two different generative models for binary classification of machinery into faultless or faulty. These two different generative methods model faultless machinery and, in turn, are based on Gaussian and autoencoder modeling. While adoption of this methodology is not totally new in the context of machine fault detection (Purohit et al. 2019; Ahmad et al. 2020; Yu, Liu, and Ye 2021), in this work, as a novelty, we propose to perform model adaptation for improved performance/generalization across different production units and under data scarcity conditions. In particular, as carefully explained in Subsection 4.2, we take advantage of the fact that machinery starts to work in optimum conditions to carry out *online* Gaussian and autoencoder model adaptation.

2. Background and system setup

This section contains a short introduction to injection moulding and acoustic signals coming from the injection moulding machine and injection mould during a complete injection moulding cycle.

2.1. *Injection moulding*

Plastic injection moulding is a widely used technique for production of plastic components and products. An injection moulding setup consists of an injection moulding machine and an injection mould (see illustration in Figure 1). The injection mould shapes the produced part, whereas the moulding machine is used for melting the plastic granulate and injecting the melted plastic into the mould.

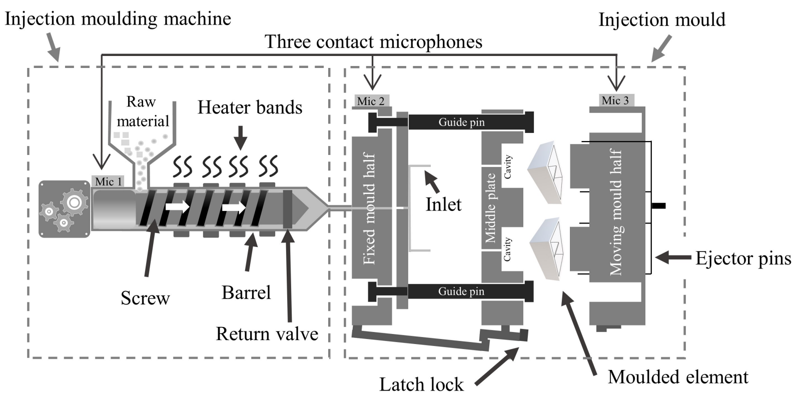


Figure 1. Illustration of an injection moulding machine with a three-plate mould in open position. The position of the three contact microphones can be seen on the injection unit (Mic 1), fixed (Mic 2) and moving mould half (Mic 3).

A moulding cycle can be seen as a series of three distinct phases. The first phase is the injection phase, where melted plastic is injected into the closed mould, followed by a cooling phase, where the plastic in the mould solidifies. The last phase is the demoulding phase, where the mould opens and the moulded elements are ejected out of the mould followed by the closing of the mould. During the cooling phase, new melted plastic is transported through the barrel by rotating the screw (dosing phase). The rotation of the screw is also mixing the granulate coming onto the barrel to ensure a homogeneous melt. When sufficient melt is transported in front of the screw, the rotation is stopped. During the injection of the melt, the screw is moved forward and a return valve at the tip of the screw ensures that the melted plastic is pressed into the mould (working as a piston) filling up the cavities. To ensure cooling of the elements, a constant flow of water is pumped around in cooling channels inside the mould.

2.1.1. Maintenance setup

The industrial partner has a three-folded preventive maintenance strategy that ensures high productivity and quality. The most frequent inspection (every twelfth hour) is a visual inspection where the produced elements (checked for burn marks, flashes, oil on elements, etc.) and the overall state of the mould are assessed. Twice a week, a Total Productive Maintenance (TPM) is performed on the moulds, where the closing surface, air vents and guide pins are cleaned, and new grease is added to the guide pins and other moving parts (this is performed while the mould is placed in the moulding machine). A systematical cleaning of the mould is carried out after every 300,000 moulding cycles (every four to six week). The systematical cleaning is performed in the tool shop and here the mould is disassembled and all mould parts are cleaned. After the cleaning, the mould parts are inspected and greased, and the mould is assembled again.

2.2. *Acoustic signals from injection moulding*

The acoustic signal coming from an injection moulding machine consists of contributions from rotation of the screw, metal against metal at mould opening and closing, flow of plastic during injection, flow of cooling water in the mould, and air flow in the air vents. This makes the signal more complex than a signal coming from, e.g., bearings, where the main contribution comes from rotation. An example of an acoustic signal from one moulding cycle can be seen in Figure 2.

The microphone on the injection unit (Mic 1) mainly captures a signal related to injection and dosing of material as well as some constant background noise from the moulding machine (this signal could be interesting in relation to predictive machine maintenance). The two microphones on the mould do not capture acoustic signals from the moulding machine but only from different mould events between mould opening and closing (some of these events are highlighted in Figure 2). Since the focus in the current work is on mould maintenance, only acoustic signals between mould opening and mould closing from Mics 2 and 3 (referred to as channels 2 and 3 in the following sections) will be used.

2.2.1. Acoustic signals from different injection moulds

One of the goals in the current work is to make a solution that works across different moulds and on different moulding machines. It is therefore interesting to compare the acoustic signals from different moulds. In Figure 3, acoustic signals have been collected from the five moulds used in the current work. The plot is limited to the phase between mould opening and closing, and, for each mould, the waveforms from Mic 2 and Mic

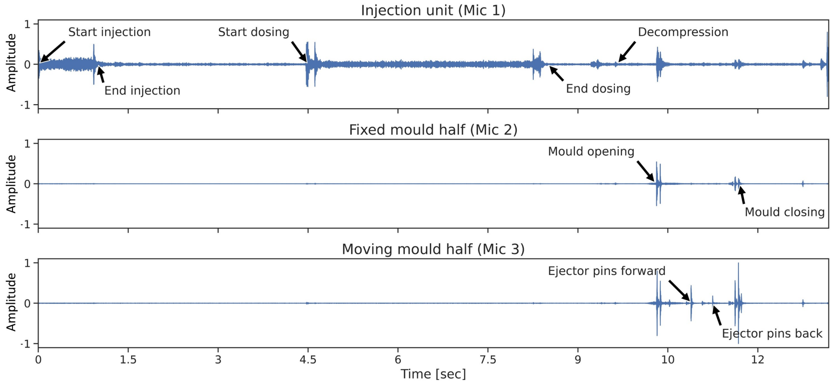


Figure 2. Waveform of an acoustic signal from an injection moulding cycle (cycle time of 13.2 seconds) with microphones at three different positions (injection unit, fixed and moving mould half).

3 are overlaid (fixed mould half and moving mould half, respectively). As it can be seen, the time between mould opening and closing differs from approximately 1.5 to 2.5 seconds.

Comparing the five moulds in Figure 3, it is clear that the events occurring in the mould opening and closing phase are not identical across moulds (shape of waveforms differs) and that different events take place at different time points in the moulding cycle (e.g., the ejector movements are not aligned). This could be a challenge when looking into a solution that works well across moulds (this will be addressed throughout the following sections).

Beside the waveform (i.e., signal amplitude in the time domain), it is also of interest to compare the different moulds in the frequency domain. Spectrograms of Moulds 1 to 4 can be seen in Figure 4, where the vertical axis is in log scale highlighting the lower frequencies. For Mould 1 and Mould 3, there are some clear background signals at around 256 Hz (horizontal lines), and at around 512 Hz for Mould 2. It is expected that these background signals originate from the machine hydraulic system.

3. Experimental setup and data collection

The experimental work has been conducted in two steps. In the first step, three worn-out moulds (Mould 1, Mould 2 and Mould 3) have been tested in a test facility (same moulding machine), where the purpose was to induce different known mould failures, i.e., lack of lubrication on guide pins and loose/missing latch locks. In the second step, recordings have been conducted on two moulds (Mould 4 and Mould 5) in a production environment (two different moulding machines) to verify results from the initial tests.

3.1. *Collection of data for the lubrication case*

When moulding with a three-plate mould, the middle plate is moved over the guide pins in every cycle. To avoid surface seizures, it is important that there is sufficient grease on the guide pins. The following procedure was performed to test the change in

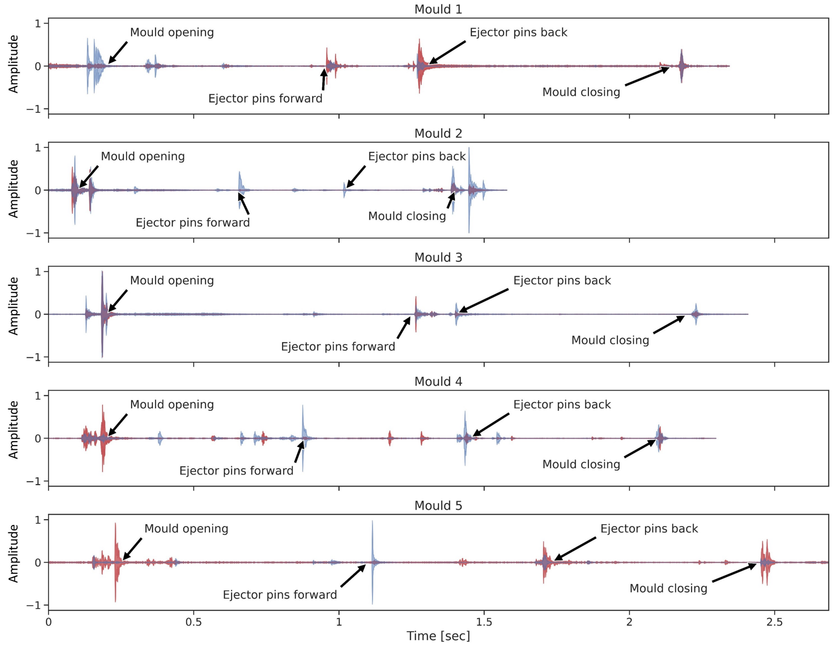


Figure 3. Waveform plot of an acoustic signal from the mould opening and closing phase (five different moulds), where the red waveform is from the fixed mould half (Mic 2) and the blue one is from the moving mould half (Mic 3).

the acoustic signal from a well-lubricated mould to a mould with lack of lubrication:

(1) A baseline of approximately 50 moulding cycles was recorded for a well-lubricated mould (denoted as "*faultless*" moulding cycles).
(2) The grease from the guide pins was removed with a dry cloth and additional 50 cycles were recorded.
(3) The guide pins were cleaned with an organic solvent and around 50 additional cycles were recorded (this step was repeated 2-3 times). The latter 50 cycles are denoted as "*faulty*" moulding cycles.
(4) Grease was added to the guide pins and 50 moulding cycles were recorded.

For the production moulds (continuous running), lubrication will be added twice a week to prevent surface seizures (see maintenance description in Subsection 2.1.1). Recordings have been conducted on two production moulds before (denoted as "*faulty*") and after the maintenance where lubrication was added (denoted as "*faultless*").

3.2. *Collection of data for the latch-lock case*

The latch lock on a mould (see illustration in Figure 5) is exposed to a large force when the mould is opened and closed. There is therefore a risk that the latch-lock head becomes loose and breaks off over time. To simulate this in a controlled way, the

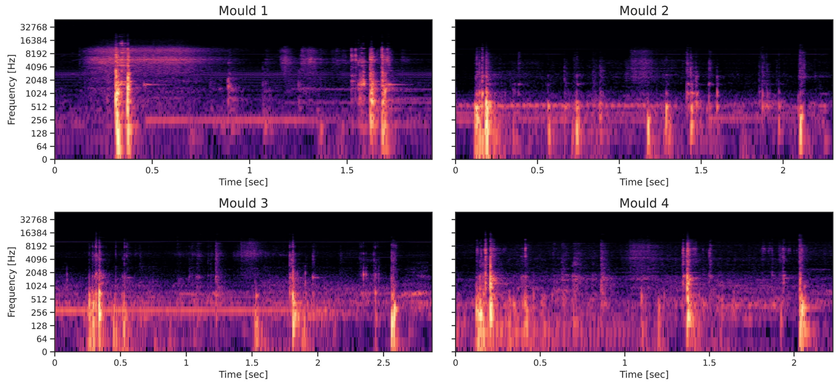


Figure 4. Spectrograms from the mould opening and closing phase (four different moulds).

following procedure was carried out for the three worn-out moulds used:

(1) A baseline of approximately 50 moulding cycles was recorded (denoted as "*faultless*").
(2) The socket screw holding the head of the latch lock was loosened 0.01 mm.
(3) Fifty moulding cycles were recorded and the resulting gap of the latch-lock head was measured.
(4) Step 3 was repeated until it was evaluated that further cycles would break off the latch-lock head (gap of 1.5 to 2.0 mm). The last (approximately) 50 cycles were denoted as "*faulty*".

For the two production moulds, only a reduced test of defect latch locks was possible (to reduce the risk of damaging the moulds). The latch lock was loosened in the same way as for the test moulds but, as soon as there was a risk of making damage to the mould, the machine was stopped and the latch lock was fixed again (the gap was at the end of the test of 0.5 mm only). Approximately, 50 moulding cycles were recorded as a baseline before the latch lock was loosened (denoted as "*faultless*"), and approximately 50 cycles before stopping the moulding machine (denoted as "*faulty*").

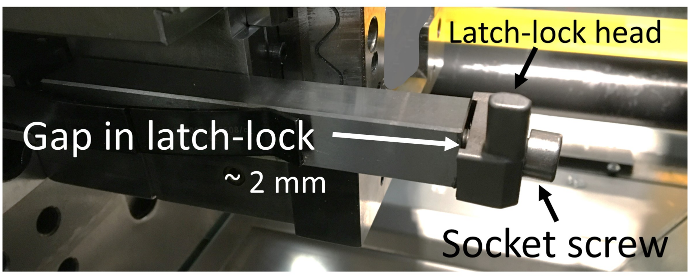


Figure 5. Latch lock at the end of the test. The latch-lock head is loose, where the gap is approximately of 2 mm.

Table 1. Amount of data collected for each of the five moulds. For latch lock and lubrication, the collected dataset is divided into faultless and faulty subsets.

		Lubrication		Latch lock	
	Mould type	*Faultless*	*Faulty*	*Faultless*	*Faulty*
Mould 1	Test	40	82	49	44
Mould 2	Test	53	49	60	25
Mould 3	Test	67	50	50	49
Mould 4	Production	100	98	50	54
Mould 5	Production	100	100	50	43

3.3. *Overview of the collected data*

This subsection serves as an overview of the collected data and the division of the data into training and test datasets. Table 1 contains information of mould number, mould type and the number of recordings collected. The columns under "Latch lock" and "Lubrication" comprise the number of faultless and faulty moulding cycles used for training and testing. For the Gaussian and autoencoder modeling methods, leave-one-mould-out validation is used, where the models are estimated on four of the moulds and tested on the fifth mould. Furthermore, for Gaussian and autoencoder model adaptation, the first τ moulding cycles from the "Latch lock/Faultless" and "Lubrication/Faultless" columns are employed, the motivation of which is carefully explained throughout Subsection 4.2.

Besides, as presented in the next section, a preliminary analysis using support vector machine (SVM) classification is also carried out in this work. For the SVM case, the data are used in two different ways. In the first case, where a model is independently trained for each mould using mould-specific faultless and faulty data, an 80/20 split is used for training/testing. This results in a training dataset containing approximately 80 cycles and a test dataset of approximately 20 cycles. In the second case, leave-one-mould-out validation is used as for the Gaussian and autoencoder modeling case.

3.4. *Data acquisition*

Acoustic signals were recorded using three Schertler Dyn-Uni-P48 contact microphones, and a Behringer UMC404HD audio interface connected to Raspberry PI. The microphones have a frequency range of 20-20,000 Hz, a dynamic range of 139 dB and a theoretical sensitivity of -62 dB. The sampling frequency has been set to 44.1 kHz and the resolution to 24 bits. The contact microphones were attached to the mould using magnets and 3D-printed adapters. The three contact microphones used in the test were placed on the injection unit, on the fixed half of the mould and on the moving part of the mould (see Figure 1).

4. Methodology

Two different approaches will be investigated for detecting lack of lubrication and loose latch locks, where one utilizes features distributed over time and the other utilizes features averaged over time.

As a preliminary study, spectral centroid features and SVM classification will be explored for features distributed over time. In the second approach, time-averaged features are used for Gaussian and autoencoder generative modeling of faultless moulding cycles for classification. The latter approach has been found to be the best in terms of generalization ability across moulds and will therefore be described in more details in the following subsections.

4.1. *Spectral centroid and support vector machine classification*

Different spectral features have been tested, and the spectral centroid, computed by Librosa toolkit (McFee et al. 2015), as input for an SVM classifier was found useful for identifying both loose latch locks and the lack of lubrication.

Calculated on a time frame basis, the spectral centroid simply consists of the mean frequency derived from treating the magnitude spectrogram as a distribution over frequency bins (see Figure 8). This implies that the dimensionality of the spectral centroid feature vector depends on the time length of the input acoustic signal, and we have up to 30% variation in length of recordings from one mould to another. Since SVM requires classifying feature vectors of a fixed dimensionality, linear interpolation is employed to produce 170-dimensional spectral centroid feature vectors (to match the mould with the longest time duration). Notice that linear interpolation yields *warped* time feature vectors.

SVM is used as a classifier since it often gives good results when used with spectral data, as demonstrated, e.g., by Sampaio et al. (2020). SVM is a kernel-based method, and the best results were obtained by using a linear kernel. The regularization parameter, implemented as a squared ℓ_2 penalty, was set to 0.1.

4.2. *Generative approach to fault detection*

This subsection is intended to describe a generative approach to machinery fault detection (in this specific case, faulty moulds), which is based on either Gaussian or autoencoder modeling of faultless mould acoustic signals using a common feature extraction pipeline. The reason for considering a generative approach modeling faultless moulds to solve this classification problem is two-fold: *1)* statistics of faulty mould acoustic signals are more complex and difficult to learn than those of faultless mould signals, since faults consist of a wide variety of error types, and *2)* we suffer from data scarcity (only a few error types are represented in the limited available data). Based on this, the idea is to detect faults as deviations from models of faultless moulds and find suitable thresholds for this decision.

The rest of this subsection is devoted to describe the feature extraction pipeline, Gaussian and autoencoder modeling and classification, and an *online* Gaussian and autoencoder model adaptation procedure for improved fault detection performance.

4.2.1. Feature extraction

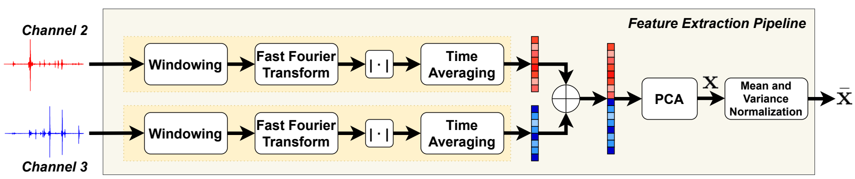


Figure 6. Feature extraction pipeline considered in this work for fault detection.

Figure 6 depicts the feature extraction pipeline considered in this work. Signals from Mics 2 and 3 (denoted as channels 2 and 3) are processed in parallel as follows. First, the input acoustic signals are framed using a 20 ms Hann window with a 10 ms hop. Next, by application of a 2048-point fast Fourier transform (FFT), magnitude spectra are computed. Then, the magnitude spectra are time-averaged to produce a 1025-dimensional summary spectral vector for each of the input channels. The two resulting summary spectral vectors are concatenated prior to dimensionality reduction to $D = 256$ components by means of principal component analysis (PCA). Finally, *only* in case of employing an autoencoder for fault detection (see Subsection 4.2.3), mean and variance normalization is applied in order to stabilize training as well as to improve model generalization (LeCun et al. 2012).

The final $D = 256$-dimensional feature vectors and their normalized versions used for Gaussian- and autoencoder-based fault detection will be denoted by $\mathbf{x}$ and $\bar{\mathbf{x}}$, respectively. In turn, $\mathbf{x}^{\{b\}}$ and its normalized version $\bar{\mathbf{x}}^{\{b\}}$ will refer to feature vectors from acoustic signals corresponding to faultless machinery.

Notice that both the PCA transform and normalization parameters (i.e., mean and variance) are calculated from the corresponding training faultless dataset (see Section 3 on experimental data).

4.2.2. Gaussian-based fault detection

We have found that using a single multivariate Gaussian distribution is adequate to properly model $\mathbf{x}^{\{b\}}$. Therefore, we assume that

$$\mathbf{x}^{\{b\}} \sim p\left(\mathbf{x}^{\{b\}}; \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b\right) = \mathcal{N}_D\left(\mathbf{x}^{\{b\}} \middle| \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b\right) = \prod_{i=1}^{D} \mathcal{N}\left(x_i^{\{b\}} \middle| \mu_{b,i}, \sigma_{b,i}^2\right), \tag{1}$$

where $\mathbf{x}^{\{b\}} = \left(x_1^{\{b\}}, ..., x_i^{\{b\}}, ..., x_D^{\{b\}}\right)^\top$, $\boldsymbol{\mu}_b = (\mu_{b,1}, ..., \mu_{b,i}, ..., \mu_{b,D})^\top$ is the mean vector and $\boldsymbol{\Sigma}_b = \operatorname{diag}\left(\sigma_{b,1}^2, ..., \sigma_{b,i}^2, ..., \sigma_{b,D}^2\right)$ is the diagonal —because $\mathbf{x}^{\{b\}}$ components are optimally decorrelated due to PCA— covariance matrix. It must be noticed that both $\boldsymbol{\mu}_b$ and $\boldsymbol{\Sigma}_b$ are estimated by maximum likelihood from the same training faultless dataset as the one used for PCA transform computation in Subsection 4.2.1.

At test time, we will observe a time sequence of feature vectors $\mathbf{x}$. Each of these observations summarizes the acoustic signal yielded by one moulding cycle of the moulding machine (limited to signal between mould opening and closing). In other words, if $\mathbf{x}(t)$ is the feature vector from the t-th operating/moulding cycle, we will observe a time sequence $\{\mathbf{x}(1), \mathbf{x}(2), ..., \mathbf{x}(t), ...\}$. We are then interested in finding out whether $\mathbf{x}(t)$, $\forall t$, corresponds to an acoustic signal recorded from faultless or faulty machinery. With this goal in mind, we calculate a score $S_G(t)$ simply consisting of the log-likelihood

$$S_G(t) = \log\left(p\left(\mathbf{x}(t); \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b\right)\right). \tag{2}$$

If $S_G(t) > \chi_G$ ($S_G(t) < \chi_G$), where χ_G is a log-likelihood decision threshold to be determined, $\mathbf{x}(t)$ is assumed to come from a faultless (faulty) moulding cycle.

To obtain a value for χ_G, we can take advantage of the fact that, at test time, the moulds will start working in optimum conditions. This means that we can safely

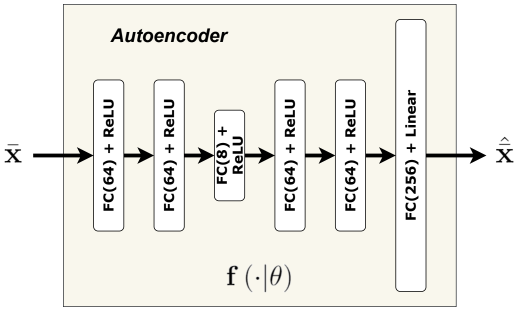


Figure 7. Autoencoder architecture for the modeling of acoustic signals from faultless moulding cycles. "FC(N)" stands for a fully-connected layer with N neurons. "ReLU" and "Linear" refer to rectified linear unit and linear activation functions, respectively.

assume that the first τ moulding cycles are faultless[1]. Hence, χ_G may be approximated as follows Gharghabi et al. (2017):

$$\chi_G = \frac{1}{\tau}\sum_{t=1}^{\tau} S_G(t) - z_{.9995}\sqrt{\frac{1}{\tau-1}\sum_{t=1}^{\tau}\left(S_G(t) - \frac{1}{\tau}\sum_{t'=1}^{\tau} S_G(t')\right)^2}, \tag{3}$$

where $z_{.9995} \approx 3.29$ is the z-score for the 99.95 percentile point of the standard normal distribution.

Finally, we know that the moulds will slowly transit across moulding cycles from being faultless to being faulty. Therefore, we can exploit this information in order to mitigate possible outlier scores by application of a first-order recursive smoothing:

$$\bar{S}_G(t) = \gamma\bar{S}_G(t-1) + (1-\gamma)S_G(t), \tag{4}$$

where $0 < \gamma < 1$ is a remembering factor (in this paper, $\gamma = 0.9$). Preliminary experiments revealed that it is preferable to compare $\bar{S}_G(t)$ with χ_G instead of $S_G(t)$ for improved fault detection.

4.2.3. Autoencoder-based fault detection

Inspired by Purohit et al. (2019), we also decided to experiment with an autoencoder, whose architecture can be visually inspected in Figure 7, for modeling the faultless status of the mould. Particularly, the goal is to learn the set of parameters θ of an autoencoder function $\mathbf{f}(\cdot|\theta) : \mathbb{R}^D \to \mathbb{R}^D$, $\hat{\bar{\mathbf{x}}} = \mathbf{f}(\bar{\mathbf{x}}|\theta)$, in such a manner that the mean squared reconstruction error

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{D}\left\|\bar{\mathbf{x}}^{\{b\}} - \hat{\bar{\mathbf{x}}}^{\{b\}}\right\|_2^2 \tag{5}$$

is minimized during the training phase. The autoencoder of Figure 7 is trained, by using Adam (Kingma and Ba 2015) as an optimizer, for a total of 50 epochs, which is enough to guarantee convergence. The minibatch size is 32 faultless training samples.

[1]Particularly, in this work, $\tau = 14$ was heuristically selected taking into consideration our data scarcity scenario (see Table 1).

Our expectation is that, at test time, the mean squared reconstruction error

$$\varepsilon(t) = \frac{1}{D} \left\| \bar{\mathbf{x}}(t) - \hat{\bar{\mathbf{x}}}(t) \right\|_2^2 \tag{6}$$

is lower when $\bar{\mathbf{x}}(t)$ corresponds to an acoustic signal from a faultless moulding cycle than when it comes from a faulty cycle. In other words, we can construe Eq. (6) as a score $S_A(t) = \varepsilon(t)$ for fault detection as similarly done in the context of Gaussian modeling. In this case, if $S_A(t) < \chi_A$ ($S_A(t) > \chi_A$), where χ_A is a mean squared reconstruction error decision threshold to be estimated, $\bar{\mathbf{x}}(t)$ is assumed to proceed from a faultless (faulty) moulding cycle. Following the same philosophy behind Eq. (3), χ_A may be estimated as

$$\chi_A = \frac{1}{\tau} \sum_{t=1}^{\tau} S_A(t) + z_{.9995} \sqrt{\frac{1}{\tau - 1} \sum_{t=1}^{\tau} \left(S_A(t) - \frac{1}{\tau} \sum_{t'=1}^{\tau} S_A(t') \right)^2}. \tag{7}$$

With the same aim as for the Gaussian case, a smoothed sequence of mean squared reconstruction errors $\left\{\bar{S}_A(1), \bar{S}_A(2), ..., \bar{S}_A(t), ...\right\}$, obtained from the application of a first-order recursion as in Eq. (4), is used for improved fault detection.

4.2.4. Gaussian and autoencoder model adaptation

To enhance the generalization ability of the Gaussian and autoencoder models, training data from a variety of faultless moulds (see description in Section 3) are used in practice. Nevertheless, as aforementioned, our experimental framework suffers from data scarcity, so the Gaussian and autoencoder models may fail at test time when trying to account for the statistics of the acoustic signals from moulds not seen during the training phase. To alleviate this issue, we can again take advantage of the assumption that, at test time, the moulds will work in optimum conditions for, at least, the first τ moulding cycles. Therefore, observations coming from the first τ moulding cycles can be exploited to perform model adaptation as explained below. Despite its simplicity, the usefulness of the proposed adaptation methodology to provide robustness against statistics variations can be assessed in Section 5.

(1) Gaussian model adaptation: Let $\mathcal{X}_G = \left\{\mathbf{x}^{\{b\}}(1), \mathbf{x}^{\{b\}}(2), ..., \mathbf{x}^{\{b\}}(\tau)\right\}$ be the set of feature vectors used for Gaussian adaptation. We want to obtain new Gaussian parameters $\boldsymbol{\mu}'_b = \left(\mu'_{b,1}, ..., \mu'_{b,i}, ..., \mu'_{b,D}\right)^\top$ and $\boldsymbol{\Sigma}'_b = \text{diag}\left(\sigma'^2_{b,1}, ..., \sigma'^2_{b,i}, ..., \sigma'^2_{b,D}\right)$ replacing $\boldsymbol{\mu}_b$ and $\boldsymbol{\Sigma}_b$ in the Gaussian model of Eq. (1) in such a manner that the likelihood of the adapted Gaussian density $p\left(\mathbf{x}^{\{b\}}; \boldsymbol{\mu}'_b, \boldsymbol{\Sigma}'_b\right) = \mathcal{N}_D\left(\mathbf{x}^{\{b\}} \middle| \boldsymbol{\mu}'_b, \boldsymbol{\Sigma}'_b\right)$ is maximized on the adaptation set, where (Digalakis, Rtischev, and Neumeyer 1995; Gales 1998)

$$\begin{aligned} \mu'_{b,i} &= \alpha_{b,i}\mu_{b,i} + \beta_{b,i}; \\ \sigma'^2_{b,i} &= \alpha^2_{b,i}\sigma^2_{b,i}. \end{aligned} \tag{8}$$

In other words, we want to estimate $\alpha_{b,i}$ and $\beta_{b,i}$, $\forall i$ ($i = 1, ..., D$), maximizing the likelihood of $p\left(\mathbf{x}^{\{b\}}; \boldsymbol{\mu}'_b, \boldsymbol{\Sigma}'_b\right)$ on $\mathcal{X}_G$. It is straightforward to show that this adaptation procedure is equivalent to simply estimate the new Gaussian parameters $\boldsymbol{\mu}_b'$ and $\boldsymbol{\Sigma}_b'$ by maximum likelihood from $\mathcal{X}_G$.

(2) Autoencoder model adaptation: Let $\mathcal{X}_A = \left\{\bar{\mathbf{x}}^{\{b\}}(1), \bar{\mathbf{x}}^{\{b\}}(2), ..., \bar{\mathbf{x}}^{\{b\}}(\tau)\right\}$ represent the set of feature vectors used for adapting the autoencoder model. One at a time, each observation $\bar{\mathbf{x}}^{\{b\}}(t)$ $(t = 1, 2, ..., \tau)$ is used once to update the parameters of the autoencoder model $\mathbf{f}\left(\cdot|\theta\right)$ (i.e., one epoch) by backpropagation. Again, Adam is utilized as an optimizer.

4.3. *Evaluation metrics*

Since in this paper we are dealing with binary classification of faultless/faulty machinery, our primary metric is the area under the receiver operating characteristic curve (AUC). The larger the AUC $\in [0,\ 1]$, the better a classifier is. Particularly, an AUC of 1 means perfect classification in the sense that the model used is capable of perfectly distinguishing between the classes.

In real-life applications, we need to use a decision threshold for classifying machinery according to its faultless or faulty status (see Section 4). Therefore, to evaluate the performance of the proposed decision threshold calculation methods of Eqs. (3) and (7), we report sensitivity and specificity (also known as true positive and true negative rates, respectively) measurements. These metrics are defined as follows:

$$\begin{aligned} \text{Sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{Specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \end{aligned} \tag{9}$$

where TP, TN, FP and FN represent, respectively, the total number of true positives, true negatives, false positives and false negatives. The larger Sensitivity $\in [0,\ 1]$ and Specificity $\in [0,\ 1]$, the better the classification performance. Furthermore, it should also be noted that *the faulty class is our positive class*, since our objective is to spot faulty machinery.

Besides, all the results based on autoencoder modeling are reported along with standard deviations across 5 different autoencoders trained with different random model initialization. Let us observe that an equivalent stochasticity is not involved in Gaussian modeling, and this is why results based on Gaussian modeling do not include standard deviations.

Finally, it should be noted that accuracy, defined as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{10}$$

where Accuracy $\in [0,\ 1]$, is specifically reported for our initial analysis employing SVMs for the classification of spectral centroid features. As we will see in Subsection 5.1, accuracy is enough to prove the total lack of generalization of this initial approach and the need for more robust techniques like those based on generative modeling that are proposed in this work.

5. Results

The result section consists of two subsections, where the first one covers a brief description of the results from the initial work using spectral centroid as feature and an SVM for the classification of faultless and faulty moulds. The second subsection covers the results from using the proposed generative modeling for fault detection.

5.1. *Initial data exploration and analysis*

Many feature extraction techniques for acoustic signals have been developed within the domain of speech and music analysis Alías, Socoró, and Sevillano (2016); Yu et al. (2017). A naive application of these techniques has been explored for utilization within predictive mould maintenance. It has been found that spectral centroid captures some of the characteristics of the faultless and faulty machinery behavior for both the latch-lock and lubrication cases.

To get an impression of spectral centroid, some examples (for the lubrication case) are presented in Figure 8. Particularly, in Figure 8a), a spectrogram from Mould 5 is depicted along with the spectral centroid. To illustrate that the spectral centroid captures effects caused by lack of lubrication, the average spectral centroid with 95% confidence interval for faultless and faulty cycles are overlaid in Figure 8b) (Mould 3) and in Figure 8c) (Mould 4). For Mould 3, differences are presented around spectral centroid number 135 and 160 (the 95% confidence intervals for the two classes do not overlap). For Mould 4, differences are seen around spectral centroid number 120 and 150. These two mould examples indicate that it would be possible to use spectral centroid for the classification of faultless and faulty moulding cycles. The two mould examples also illustrate that the discriminative features in the spectral centroid vectors do not completely overlap in warped time across different moulds, which is further illustrated in Figure 8d), where the average spectral centroid with 95% confidence interval from faultless and faulty moulding cycles is collected for all five moulds. From Figure 8d), it is clear that there is no clear separation between faultless and faulty cycles at any individual spectral centroid warped time point. This is caused by the differences in moulding cycle length and differences in events occurring between mould opening and closing.

PCA is an effective tool for data compression and visualization of similarities in large data structures. PCA is therefore used for an initial investigation of the faultless/faulty behavior captured by spectral centroid (only the lubrication case is presented, but similar results are obtained for the latch-lock case). Based on PCA, it is of interest to explore any grouping present in a scoreplot[2]. Two scoreplots from the lubrication case can be seen in Figure 9.

From the scoreplot in Figure 9a), it can be seen that performing PCA on the spectral centroid from one mould leads to clear separation between faulty and faultless moulding cycles (despite the unsupervised approach). When applying PCA on spectral centroid from multiple moulds, it can be seen (Figure 9b)) that the faulty and faultless cycles are overlapping. This clearly indicates that the differences between moulds are more dominant in the spectral centroid than the differences caused by faulty or faultless cycles.

[2] PCA is a projection method where the original data are projected onto new axes (lower-dimensional subspace) in the direction of maximum variation. A scoreplot is a scatter plot of the data on this new projection, where each point is a sample (in this case, one moulding cycle). In a scoreplot, it is possible to visualize the structure of the collected data, where points that group together represent similar variation.

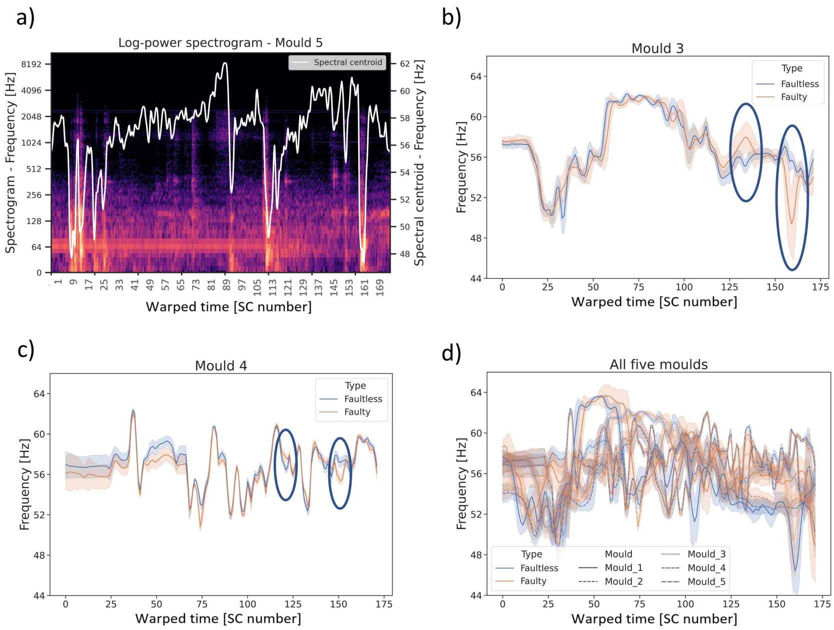


Figure 8. Visualization of spectral centroid (SC). The horizontal axis represents warped time and is given in the SC number, since all moulds have been aligned to have the same number of SCs irrespective of the cycle duration: a) spectrogram from Mould 5 along with SC; b) and c) average spectral centroid with 95% confidence interval for faultless and faulty cycles (differentiated by color) for Moulds 3 and 4, respectively; d) average spectral centroid with 95% confidence interval for faultless and faulty cycles for all five moulds (moulds differentiated by line style).

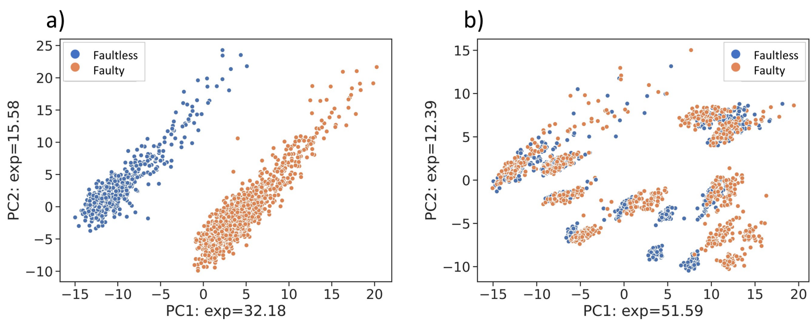


Figure 9. Scoreplot (PC1 vs. PC2) from the application of PCA to spectral centroid from different moulding cycles with faultless and faulty lubrication: a) is a scoreplot of cycles from Mould 4, and b) is a scoreplot of cycles from all five moulds. The amount of explained variance (exp) captured by PC1 and PC2 can be seen from the axis labels.

Based on the findings that some discriminative power is present in the spectral centroid, this feature has been used as input for an SVM classifying faultless and faulty moulding cycles. It has been found that this approach results in a high test accuracy if trained on data from the same mould, where the training data include both faultless and faulty moulding cycles (see results in the two left columns of Table 2). In case of following a leave-one-mould-out strategy (see the two right columns of Table 2), where the mould used as test case has not been part of the training dataset, the concept performs with an accuracy around 50%, which means that the classifier is random guessing (almost all cycles are classified as faulty). This confirms the results from the PCA analysis (no clear separation of faultless and faulty cycles). The main reason being the difference in events (and time position of events) occurring between mould opening and closing for the different moulds, which becomes very present in the used spectral centroid feature vectors, since vector components represent warped time instants.

Table 2. Classification performance in terms of accuracy using spectral centroid and SVM. For "Individual mould", accuracy is based on 80/20 train/test split per mould. For "Leave-one-mould-out", accuracy is from testing on the specific mould not being part of the training data.

	Individual mould		**Leave-one-mould-out**	
	Latch lock	*Lubrication*	*Latch lock*	*Lubrication*
Mould 1	1.00	0.96	0.41	0.47
Mould 2	1.00	1.00	0.45	0.41
Mould 3	0.96	1.00	0.49	0.43
Mould 4	1.00	0.97	0.43	0.45
Mould 5	0.98	0.94	0.49	0.47

Based on the obtained results and the limited experimental data available, we have proposed the feature extraction pipeline of Figure 6 aggregating the time information in order to reduce the impact of differences in the mould events. As previously discussed, this feature extraction pipeline is used along with generative modeling, the experimental results of which are presented in the next subsection.

5.2. *Generative approach*

The different generative methods presented in Subsection 4.2 are evaluated in this subsection for the detection of two different types of mould failure: latch-lock looseness and lack of lubrication. Specifically, we test Gaussian modeling with (Gaussian(Ad)) and without (Gaussian) adaptation, as well as autoencoder modeling with (Autoencoder(Ad)) and without (Autoencoder) adaptation.

5.2.1. Latch-lock looseness detection

Table 3 reports latch-lock looseness detection results, broken down by mould, obtained by the different generative methods presented in this paper. Recall that experimental results are achieved by means of leave-one-mould-out validation.

The first thing that draws our attention is that classification performance is highly mould-dependent when no model adaptation is carried out. This means that, when observing Gaussian and Autoencoder, AUC values are very low for Moulds 1, 2 and 3 (AUC around 0.5, meaning random guess), in contrast with those of Moulds 4 and 5. However, performing Gaussian and autoencoder model adaptation by simply exploiting the first $\tau = 14$ faultless cycles at test time is very effective. Specifically, both Gaussian(Ad) and Autoencoder(Ad) are able to provide perfect class separability,

Table 3. Latch-lock looseness detection results, broken down by mould, obtained by different generative methods. Autoencoder results are provided with standard deviations.

		AUC	*Sensitivity*	*Specificity*
Mould 1	Gaussian	0.56	0.03	1.00
	Autoencoder	0.10 ± 0.08	0.00 ± 0.00	1.00 ± 0.00
	Gaussian(Ad)	**1.00**	1.00	0.20
	Autoencoder(Ad)	**1.00 ± 0.00**	0.96 ± 0.08	0.85 ± 0.34
Mould 2	Gaussian	0.38	0.29	1.00
	Autoencoder	0.60 ± 0.03	0.36 ± 0.01	1.00 ± 0.00
	Gaussian(Ad)	**1.00**	0.99	1.00
	Autoencoder(Ad)	**1.00 ± 0.00**	0.92 ± 0.04	1.00 ± 0.00
Mould 3	Gaussian	0.62	0.00	1.00
	Autoencoder	0.05 ± 0.01	0.00 ± 0.00	1.00 ± 0.00
	Gaussian(Ad)	**1.00**	0.99	1.00
	Autoencoder(Ad)	**1.00 ± 0.00**	0.91 ± 0.09	1.00 ± 0.00
Mould 4	Gaussian	0.97	0.00	1.00
	Autoencoder	**1.00 ± 0.00**	0.00 ± 0.01	1.00 ± 0.00
	Gaussian(Ad)	**1.00**	0.85	1.00
	Autoencoder(Ad)	0.92 ± 0.14	0.31 ± 0.41	1.00 ± 0.00
Mould 5	Gaussian	0.99	0.69	1.00
	Autoencoder	**1.00 ± 0.00**	0.83 ± 0.03	1.00 ± 0.00
	Gaussian(Ad)	**1.00**	1.00	0.80
	Autoencoder(Ad)	**1.00 ± 0.00**	1.00 ± 0.00	0.80 ± 0.00

Table 4. Latch-lock looseness detection: PCA feature absolute mean differences between the training and test datasets, as well as between the adaptation and test datasets. Absolute mean differences are broken down by mould, as well as by the faultless and faulty test subsets. The reported values are obtained by summing absolute mean differences across all the D PCA features.

	Training-Test		**Adaptation-Test**	
	Faultless Test Subset	*Faulty Test Subset*	*Faultless Test Subset*	*Faulty Test Subset*
Mould 1	33.7	32.0	4.3	8.2
Mould 2	97.5	98.1	13.0	24.0
Mould 3	83.5	78.9	4.4	25.6
Mould 4	60.9	59.2	15.5	18.8
Mould 5	54.9	67.4	4.8	20.3

as we can see from the AUC values in Table 4.

We hypothesize that the performance of Gaussian and Autoencoder being highly sensitive to the mould is linked to the following two-fold reason:

(1) The specific (i.e., mould-dependent) acoustic signals yielded by each machinery (see Figures 3 and 4) differently affecting the feature statistics of each mould;
(2) The relatively poor generalization ability of the Gaussian and autoencoder models boosted by our data scarcity problem.

Since assuming Gaussianity of each of the D PCA features on a mould basis is reasonable, and knowing that we also employ Gaussian modeling of our training and adaptation data, we can compare the mean vectors of our training, adaptation and test datasets to endorse our two-fold-reason hypothesis. Table 4 shows PCA feature absolute mean differences between the training and test datasets, as well as between the adaptation and test datasets. Notice that these difference values are obtained by summing absolute mean differences across all the D PCA features. For example, let $\mu_{b,i}^{\{Tr\}}$ and $\mu_{b,i}^{\{Te\}}$ $(i = 1, ..., D)$ be the mean values of the i-th PCA feature as estimated from the training dataset and the faultless test subset, respectively. Then, the first column of Table 4 is calculated as $\sum_{i=1}^{D} \left| \mu_{b,i}^{\{Tr\}} - \mu_{b,i}^{\{Te\}} \right|$. An equivalent procedure is followed for the rest of the columns in this table.

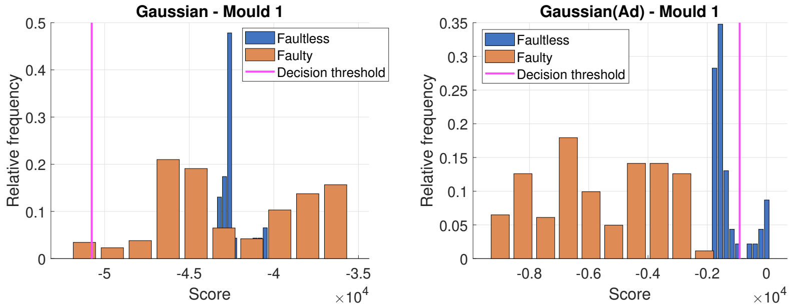


Figure 10. Latch-lock looseness detection: Faultless and faulty score histograms for Mould 1 resulting from employing Gaussian (left) and Gaussian(Ad) (right). The estimated decision thresholds are represented by vertical magenta lines.

From Table 4, we can see that the absolute mean differences between the adaptation dataset and the faultless test subset are significantly smaller than those between the adaptation dataset and the faulty test subset. This is an indication of the higher statistical similarity between the adaptation data and the faultless test data that we want to model in contrast with the faulty test data. This also helps to explain the perfect class separability obtained by both Gaussian(Ad) and Autoencoder(Ad). On the contrary, absolute mean differences between the training dataset and the faultless and faulty test subsets are comparable to a large extent, which can help to explain the worse discrimination ability of both Gaussian and Autoencoder with respect to that of these methods with adaptation. Only for Mould 5, for which both Gaussian and Autoencoder reach an excellent performance, the absolute mean difference between the training dataset and the faulty test subset (67.4) is clearly larger than that between the training dataset and the faultless test subset (54.9).

To assess the quality of the methods proposed in Eqs. (3) and (7) for decision threshold calculation, we can observe the sensitivity and specificity values reported in Table 3. For example, under the more interesting adaptation scenario, we can see that, generally speaking, the estimated decision thresholds provide a decent performance despite the simple philosophy behind Eqs. (3) and (7). That being said, given that both Gaussian(Ad) and Autoencoder(Ad) yield perfect class separability, there is room for improvement in terms of decision threshold calculation.

For illustrative purposes, Figure 10 plots faultless and faulty score histograms for Mould 1 resulting from using Gaussian and Gaussian(Ad). Moreover, we have also overlaid on these histograms the estimated decision thresholds (vertical magenta lines). From this figure, we can visually inspect how adaptation allows us to obtain non-overlapping faultless and faulty score distributions and a better decision threshold, which dramatically improves classification performance.

5.2.2. Lubrication detection

Similarly to the above latch-lock looseness detection case, Table 5 shows lubrication detection results, broken down by mould, achieved by the Gaussian and autoencoder modeling methods with and without adaptation. Again, recall that, due to severe data scarcity, experimental results are obtained by following a leave-one-mould-out strategy.

From the AUC column of Table 5, we can see that, except for Moulds 1 and 5 (AUC

Table 5. Lubrication detection results, broken down by mould, obtained by different generative methods. Autoencoder results are provided with standard deviations.

		AUC	*Sensitivity*	*Specificity*
Mould 1	Gaussian	0.05	0.00	1.00
	Autoencoder	0.00 ± 0.00	0.00 ± 0.00	1.00 ± 0.00
	Gaussian(Ad)	**1.00**	1.00	1.00
	Autoencoder(Ad)	**1.00 ± 0.00**	1.00 ± 0.00	1.00 ± 0.00
Mould 2	Gaussian	0.99	0.43	1.00
	Autoencoder	0.99 ± 0.00	0.02 ± 0.03	1.00 ± 0.00
	Gaussian(Ad)	**1.00**	0.98	1.00
	Autoencoder(Ad)	**1.00 ± 0.00**	0.73 ± 0.38	1.00 ± 0.00
Mould 3	Gaussian	**1.00**	0.98	1.00
	Autoencoder	**1.00 ± 0.00**	0.97 ± 0.02	1.00 ± 0.00
	Gaussian(Ad)	**1.00**	1.00	0.21
	Autoencoder(Ad)	**1.00 ± 0.00**	1.00 ± 0.00	0.92 ± 0.11
Mould 4	Gaussian	0.96	0.00	1.00
	Autoencoder	**1.00 ± 0.00**	0.35 ± 0.34	1.00 ± 0.00
	Gaussian(Ad)	**1.00**	1.00	0.59
	Autoencoder(Ad)	**1.00 ± 0.00**	0.81 ± 0.40	1.00 ± 0.00
Mould 5	Gaussian	0.06	0.00	1.00
	Autoencoder	0.27 ± 0.01	0.00 ± 0.00	1.00 ± 0.00
	Gaussian(Ad)	**1.00**	1.00	0.43
	Autoencoder(Ad)	0.66 ± 0.03	0.99 ± 0.02	0.47 ± 0.07

Table 6. Lubrication detection: PCA feature absolute mean differences between the training and test datasets, as well as between the adaptation and test datasets. Absolute mean differences are broken down by mould, as well as by the faultless and faulty test subsets. The reported values are obtained by summing absolute mean differences across all the D PCA features.

	Training-Test		**Adaptation-Test**	
	Faultless Test Subset	*Faulty Test Subset*	*Faultless Test Subset*	*Faulty Test Subset*
Mould 1	45.8	37.4	2.0	20.7
Mould 2	38.7	39.9	2.2	8.5
Mould 3	24.6	34.7	1.3	21.9
Mould 4	46.5	49.9	2.5	8.3
Mould 5	21.0	20.3	2.8	4.3

values close to 0 mean that the model is consequently making the wrong decision most of the time), both Gaussian and Autoencoder exhibit very good discrimination ability. Once again, adaptation of the Gaussian and autoencoder models by exploiting the first $\tau = 14$ faultless moulding cycles at test time yields perfect class separability except for Mould 5 when Autoencoder(Ad) is used. In the latter case, we hypothesize that there is insufficient data to successfully adapt the autoencoder so it fits the distribution of the faultless test data. In other words, Autoencoder(Ad) could be underperforming as a result of data scarcity.

Indeed, the performance of Gaussian and Autoencoder being highly sensitive to the mould should be, once again, related to the mould-dependent acoustic signal characteristics along with the lack of generalization of the Gaussian and autoencoder models as a result of data scarcity. As we did for latch-lock looseness detection, Table 6 reports PCA feature absolute mean differences between the training and test datasets, as well as between the adaptation and test datasets. As expected, from Table 6, we can see that the absolute mean differences between the adaptation dataset and the faultless test subset are substantially smaller than those between the adaptation dataset and the faulty test subset. Furthermore, it can also be seen from this table that, unlike for Moulds 2, 3 and 4, for Moulds 1 and 5, the absolute mean differences between the training dataset and the faultless test subset are larger than those between the

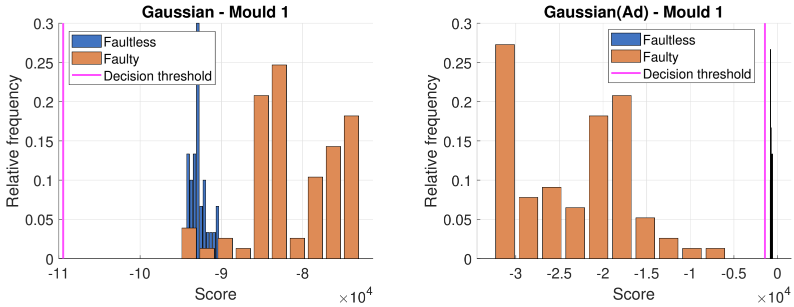


Figure 11. Lubrication detection: Faultless and faulty score histograms for Mould 1 resulting from employing Gaussian (left) and Gaussian(Ad) (right). The estimated decision thresholds are represented by vertical magenta lines.

training dataset and the faulty test subset. This fact helps to explain the deteriorated performance of Gaussian and Autoencoder on the latter moulds.

Then, let us observe the sensitivity and specificity values included in Table 5 in order to assess the quality of the decision threshold calculation methods of Eqs. (3) and (7). The performance provided by the two decision threshold calculation methods heavily depends on both the particular mould and generative method employed. For example, while the estimated thresholds for Mould 1 when considering adaptation yield flawless classification, the threshold chosen for Mould 4 leads to poor performance when no adaptation is considered, taking into account the associated AUC values. These results point out the need for more robust decision threshold calculation methods possibly exploiting other information sources in addition to the scores given by the first τ faultless moulding cycles at test time.

Finally, as we did for the latch-lock looseness detection case, Figure 11 depicts, for the lubrication case, faultless and faulty score histograms from using Gaussian and Gaussian(Ad) also on Mould 1. From the observation of these histograms, a similar conclusion can be drawn: adaptation is of utmost importance in order to improve the separability of the faultless and faulty score distributions as well as the quality of the estimated decision threshold for classification.

6. Discussion and conclusion

In this paper, we have shown that, in order to generate a robust modeling concept (that can be used for new moulds not included in the model), the acoustic signal features have to be aggregated with respect to time to overcome the differences in events occurring within a moulding cycle. We have presented two generative methods using this concept for faulty moulding cycle detection, one based on Gaussian modeling and the other on autoencoder modeling. The motivation behind proposing the use of generative methods modeling faultless moulding cycles has been to better cope with data scarcity, since statistics of faultless acoustic signals should be less complex to model with respect to those of faulty machinery. While the utility of these techniques has been demonstrated for two tasks in this paper (i.e., latch-lock looseness and lubrication detection), in principle, these methods can be applied for detecting any type of mould failure (or to any other type of production equipment or machine).

Our experimental results using Gaussian and autoencoder modeling without adaptation have shown to be highly mould-dependent. This fact reflects that each injection mould generates, to some extent, specific acoustic signals impacting the feature statistics while the generalization ability of the Gaussian and autoencoder models is limited due to data scarcity. A way to improve generalization/performance could be by substantially increasing the amount of data to train a Gaussian mixture model (instead of a single Gaussian density) able to account for the different modes associated to a variety of faultless moulds. Since collecting new training data has not been feasible at this point, a simple, yet smart and effective alternative strategy has been raised: performing Gaussian and autoencoder model adaptation by exploiting the first τ moulding cycles at test time. These moulding cycles can be safely assumed to be faultless given that the machinery will start working in optimum conditions.

We have demonstrated that, when performing adaptation, the model can almost perfectly distinguish between faultless and faulty cycles. Particularly, while we suspect that autoencoder modeling could be underperforming in our data scarcity scenario, adapted Gaussian modeling has reached ideal class separability at test time in all cases. In addition to superior detection performance, adapted Gaussian modeling (which only needs to estimate its mean and variance parameters from a few faultless cycles at test time) is computationally cheaper than autoencoder modeling. This makes the solution appealing for, e.g., direct integration in an IoT device, which often have low computational power, where the alternative would be edge computing in a dedicated industrial PC or cloud computing. It is possible to use the derived model for all new moulds introduced (of same type), meaning that the monitoring can be started after running the first τ moulding cycles on a new mould (we have found that $\tau = 14$ is sufficient). This is an extreme benefit in the industrial context described in this paper (multiple alike setups with changing configuration in machine/mould combination), and it is expected that the same could be achieved within other industries (e.g., tool wear detection in computer numerical control milling, where a large variety of different tools are used).

Regarding the proposed decision threshold calculation methods, our experiments have revealed that there is room for improvement despite these methods, as a trend, provide decent results. Therefore, as future work, we will explore more robust methods for the calculation of the decision threshold possibly relying on other information sources in addition to the scores coming from the first τ faultless moulding cycles at test time.

References

Ahmad, Sabtain, Kevin Styp-Rekowski, Sasho Nedelkoski, and Odej Kao. 2020. "Autoencoder-based Condition Monitoring and Anomaly Detection Method for Rotating Machines." In *Proceedings of Big Data 2020 – IEEE International Conference on Big Data, December 10-13, Atlanta, USA*, .

Ahmad, Wasim, Sheraz Ali Khan, M M Manjurul Islam, and Jong-Myon Kim. 2019. "A reliable technique for remaining useful life estimation of rolling element bearings using dynamic regression models." *Reliability Engineering & System Safety* 184: 67–76.

Alías, Francesc, Joan Claudi Socoró, and Xavier Sevillano. 2016. "A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds." *Applied Sciences* 6 (5): 143.

Biswal, Sailendu, Jithin Donny George, and G.R. Sabareesh. 2016. "Fault Size Estimation Using Vibration Signatures in a Wind Turbine Test-rig." *Procedia Engineering* 144: 305–311.

Chen, Joseph C., Gangjian Guo, and Wei-Nian Wang. 2020. "Artificial neural network-based online defect detection system with in-mold temperature and pressure sensors for high precision injection molding." *The International Journal of Advanced Manufacturing Technology* 110: 2023–2033.

Digalakis, V.V., D. Rtischev, and L.G. Neumeyer. 1995. "Speaker adaptation using constrained estimation of Gaussian mixtures." *IEEE Transactions on Speech and Audio Processing* 3: 357–366.

Durbhaka, Gopi Krishna, and Barani Selvaraj. 2016. "Predictive maintenance for wind turbine diagnostics using vibration signal analysis based on collaborative recommendation approach." In *Proceedings of ICACCI 2016 – International Conference on Advances in Computing, Communications and Informatics, September 21-24, Jaipur, India*, 1839–1842.

Frumosu, Flavia Dalia, Georg Ø. Rønsch, and Murat Kulahci. 2020. "Mould wear-out prediction in the plastic injection moulding industry: a case study." *International Journal of Computer Integrated Manufacturing* 33: 1245–1258.

Gales, M.J.F. 1998. "Maximum likelihood linear transformations for HMM-based speech recognition." *Computer Speech & Language* 12: 75–98.

Gharghabi, Shaghayegh, Yifei Ding, Chin-Chia Michael Yeh, Kaveh Kamgar, Liudmila Ulanova, and Eamonn Keogh. 2017. "Matrix Profile VIII: Domain Agnostic Online Semantic Segmentation at Superhuman Performance Levels." In *Proceedings of ICDM 2017 – IEEE International Conference on Data Mining, November 18-21, New Orleans, USA*, 117–126.

Ha, Jeong-Min, Hyeon-Jung Kim, Yoo-Soo Shin, and Byeong-Keun Choi. 2018. "Degradation Trend Estimation and Prognostics for Low Speed Gear Lifetime." *International Journal of Precision Engineering and Manufacturing* 19: 1099–1105.

Kingma, Diederik P., and Jimmy Ba. 2015. "Adam: A method for stochastic optimization." In *Proceedings of ICLR 2015 – 3rd International Conference on Learning Representations, May 7-9, San Diego, USA*, .

LeCun, Yann A., Léon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. 2012. "Efficient BackProp." In *Neural Networks: Tricks of the Trade*, Vol. 7700, 9–48. Springer.

McFee, Brian, Colin Raffel, Dawen Liang, Daniel Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. 2015. "librosa: Audio and Music Signal Analysis in Python." 01, 18–24.

Moreira, Eurico Esteves, Filipe Serra Alves, Marco Martins, Gabriel Ribeiro, António Pina, Diogo E. Aguiam, Edoardo Sotgiu, Elisabete P. Fernandes, and João Gaspar. 2020. "Industry 4.0: Real-time monitoring of an injection molding tool for smart predictive maintenance." In *Proceedings of ETFA 2020 – 25th IEEE International Conference on Emerging Technologies and Factory Automation, September 8-11, Vienna, Austria*, 1209–1212.

Pan, Zhaotai, Yi Ge, Yu Chen Zhou, Jing Chang Huang, Yu Ling Zheng, Ning Zhang, Xiao Xing Liang, et al. 2017. "Cognitive Acoustic Analytics Service for Internet of Things." In *Proceedings of ICCC 2017 – IEEE International Conference on Cognitive Computing,* *June 25-30, Honolulu, USA*, 96–103.

Park, Chulsoon, Dughee Moon, Namchul Do, and Sung Moon Bae. 2016. "A predictive maintenance approach based on real-time internal parameter monitoring." *The International Journal of Advanced Manufacturing Technology* 85: 623–632.

Park, You-Jin, Shu-Kai S. Fan, and Chia-Yu Hsu. 2020. "A Review on Fault Detection and Process Diagnostics in Industrial Processes." *MDPI Processes* 8: 1–26.

Purohit, Harsh, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi. 2019. "MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection." In *Proceedings of DCASE 2019 – 4^{th} Workshop on Detection and Classification of Acoustic Scenes and Events, October 25-26, New York, USA*, .

Salonen, Antti, and Mats Deleryd. 2011. "Cost of poor maintenance: A concept for maintenance performance improvement." *Journal of Quality in Maintenance Engineering* 17: 63–73.

Sampaio, Pedro Sousa, Ana Castanho, Ana Sofia Almeida, Jorge Oliveira, and Carla Brites. 2020. "Identification of rice flour types with near-infrared spectroscopy associated with PLS-DA and SVM methods." *European Food Research and Technology* 246: 527–537.

Schmitt, Jacqueline, Jochen Bönig, Thorbjörn Borggräfe, Gunter Beitinger, and Jochen Deuse. 2020. "Predictive model-based quality inspection using Machine Learning and Edge Cloud Computing." *Advanced Engineering Informatics* 45.

Spendla, Lukas, Michal Kebisek, Pavol Tanuska, and Lukas Hrcka. 2017. "Concept of predictive maintenance of production systems in accordance with industry 4.0." In *Proceedings of SAMI 2017 – IEEE International Symposium on Applied Machine Intelligence and Informatics, January 26-28, Herl'any, Slovakia*, 405–410.

Weichert, Dorina, Patrick Link, Anke Stoll, Stefan Rüping, Steffen Ihlenfeldt, and Stefan Wrobel. 2019. "A review of machine learning for the optimization of production processes." *The International Journal of Advanced Manufacturing Technology* 104: 1889–1902.

Yu, Hong, Zheng-Hua Tan, Zhanyu Ma, Rainer Martin, and Jun Guo. 2017. "Spoofing detection in automatic speaker verification systems using DNN classifiers and dynamic acoustic features." *IEEE transactions on neural networks and learning systems* 29 (10): 4633–4644.

Yu, Jianbo, Xing Liu, and Lyujiangnan Ye. 2021. "Convolutional Long Short-Term Memory Autoencoder-Based Feature Learning for Fault Detection in Industrial Processes." *IEEE Transactions on Instrumentation and Measurement* 70.

Çınar, Zeki Murat, Abubakar Abdussalam Nuhu, Qasim Zeeshan, Orhan Korhan, Mohammed Asmael, and Babak Safaei. 2020. "Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0." *MDPI Sustainability* 12: 1–42.

CHAPTER 9

Discussion

The contribution of the conducted work is two-fold in the sense that part of the work has been related to formulating a practical approach for effective data utilisation within manufacturing and the other related to investigating concrete applications of data utilisation within injection moulding. The work has been centred around the following three research questions:

RQ1: How to best utilise data from injection moulding machines to increase the productivity of producing injection moulded elements?

RQ2: In an off-line setting, how to develop real-time data analytic models that will help mitigate the impact of the variation in a nuisance factor in a production setup?

RQ3: How to do real-time condition-based maintenance on injection moulds to reduce the cost of preventive maintenance and reduce the risk of mould malfunction?

Results and learnings from the individual applications will in the following be linked back to the data utilisation framework, enabling an overall evaluation of the proposed approach. The discussion is divided in sections covering the main topics in the thesis, where reflections on future research will be included at the end of each section. The last section of the discussion will contain recommendation for the industrial partner.

9.1 Result overview

The project contribution consists of six publications that combined cover the three research questions listed above. Five of the publications link to applications identified in the data utilisation framework and the last cover data utilisation in more general terms. Figure 9.1 connects the explored application and publications to the data utilisation framework by linking the key drivers, applications and used data sources. The first version of our data utilisation framework was created and presented in paper A, centred around evaluating different data sources from injection moulding. Later, the framework was updated to reflect more general aspects related to data utilisation in a complex manufacturing setup. Paper B presents an approach for conducting experiments in an off-line setting and linking the results back to an application implementable in a manufacturing context. Experimental exploration and creation of a potential solution utilising acoustic emissions for real-time condition-based mould maintenance was presented in Paper C. Paper E presents a solution utilising existing

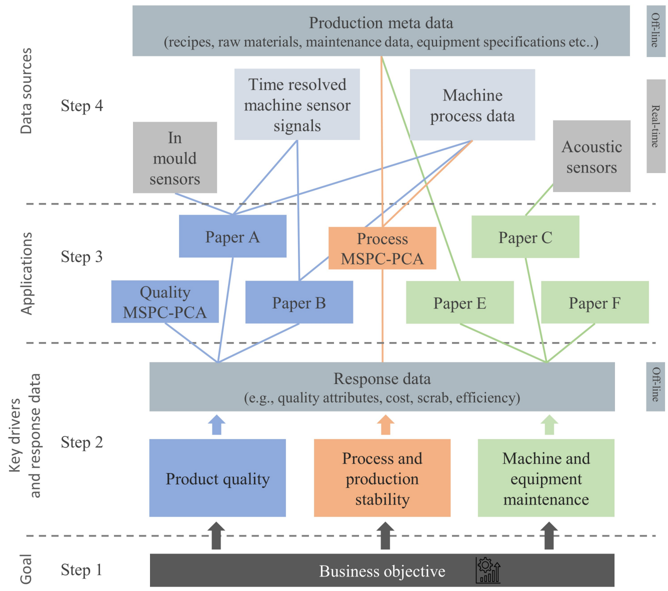


Figure 9.1. Data utilisation framework with contribution in form of application is added.

mould maintenance service data to perform a mould wear-out prediction enabling re-order of new moulds in due time. The same objective was addressed in paper F, were element metrology data was used to detect long-term degradation of injection moulds. Paper D collects experiences gathered though exploration of various data utilisation applications in different companies. Some of the central pain-points highlighted in paper D have been addressed in the presented data utilisation framework (clear objective, forward looking and sequential exploration), where others also have become pain-points in the PhD projects (e.g. data lacking essential information). Besides the publications, examples of using MSPC-PCA have been presented for monitoring element quality and monitoring of the injection moulding process.

9.2 Data utilisation framework

The initial intend of the PhD project was to focus on utilising collected process data for process surveillance and optimisation for fault detection and diagnosis. This was formulated based on an expectation that a comprehensive and sufficient data foundation was in-place at the industrial partner. Looking back through the lens of the learnings collected throughout the project, this was a naive expectation. The complexity[1] involved in establishing a data foundation that would support both fault detection and diagnosis, was neglected and at the end resulting in making the intended exploration impossible. This crucial insight initiated a reflection around how this could have been avoided and resulted in the formulation of the data utilisation framework. The key points addressed with the utilisation framework are:

- Identifying the fundamental business objective will guide the discovery of potential applications to realise the business potential. The identification of different applications supporting the same business objective, makes it possible to implement the most accessible and profitable application first. In many situations, these applications will require the least developed data foundation.
- Starting the journey of investigating and implementing applications requiring the least new data integration enables the organisation to mature as applications are developed. This will increase the readiness level and understanding of what is needed for ingestion and utilisation of new and more complex data sources.
- The structured approach will ensure synergies between utilising different data sources. Data sources might be usable for multiple applications, or new application might be discovered as data sources are combined.
- It is essential to evaluate the total cost of utilising different data against the potential benefits. In some cases it might be very expensive to implement a solution that supports full realisation of the potential business benefit. In such cases, it might be optimal to implement a solution with significantly lower implementation costs, hence only realising part of the potential benefit.

These considerations support research question 1 (RQ1), focused on best approach for utilisation of data within injection moulding. The data utilisation framework and learnings are broadly defined and therefore generalisable outside the area of injection moulding.

After working dedicatedly with data utilisation in manufacturing for three years, I don't see the analytic solutions or ideas on how to apply these in a manufacturing context as the limiting factor for effective implementation. I see the data availability, deployment infrastructure and maturity level in general as being the main limiting factors. The digital systems surrounding a data utilisation application in a modern manufacturing setup are complex and often consist of data sources with different ownership. The different data sources are not structured to support data extraction and combination with other data sources. For all existing and new digital manufacturing

[1] The complexity originates from; connecting and collecting data from around 30 different types of injection moulding machines, combining data from multiple data sources (ERP, MES and quality systems) often not intended for data extraction, creation of a IT infrastructure supporting data utilisation.

products containing data, it has to be considered how the collected data is to be utilised within data analytic applications. With an increasing demand for insights, driven by data utilisation, it is not a question of "if" the collected data is going to be used in an analytic application, but rather "when". It is therefore crucial, at least for new digital applications that the data collection part of the application is designed so that efficient and structured data extraction is possible and that data contains keys that can link the content to other data sources. I therefore see an opportunity for future research linked to protocols for data collection and structuring in digital manufacturing solutions. One aspect is regarding data analytics, an other area where this must be highly relevant is in relation to digital twin solutions, where data interactions and feedback from the physical world is going to be crucial.

9.3 Utilising underlying complex machine data

When starting data utilisation application with a well-defined objective, the question often becomes what data to collect and utilise to achieve the desired business value. In the current case, the business objective was to increase productivity by moving the geometric quality evaluation (conducted every third week) closer to real-time, thereby reducing the risk of producing elements outside specification in-between quality inspections. Material properties and moulding conditions are known to significantly impact the produced element dimensions (mainly by affecting the element shrinkage). It is crucial to monitor the process state variables that reflect the melt conditions to compensate for variation in material properties. As described in section 4, the state variables in a moulding process are hard to capture and therefore not readily available as process variables. Consequently, cavity pressure sensors have become popular to implement since they reflect the material state in the cavities. Closed-loop control, using cavity sensors, have been developed to compensate for material variations. The implementation of cavity sensors is cumbersome and expensive and, therefore, often only used for test applications and small scale productions, leaving a gap for large scale production setups where it would not be feasible to utilise cavity sensors. In paper A and B, it was explored and concluded that build-in machine pressure sensors contain the same level of information as cavity sensors and therefore could be used as a state variable to capture change in material properties. Results from designed experiments indicated that holding pressure could be used to compensate for variation occurring in raw materials. It is therefore expected that a similar closed-loop control setup could be achieved using the build-in machine sensors instead of cavity sensors. This will radically reduce the cost of implementing a real-time monitoring or closed-loop solution, especially, since it is expected that machine build-in sensor signals will be readily available on new injection moulding machines at the industrial partner. I see this development as central for all future development and optimisation within injection moulding at the industrial partner. This is both in relation to real-time prediction of element quality and for general process monitoring. I see the following area of research related to utilising machine pressure profiles:

- Profiles from different machines have to be combined to reduce the number of models to develop and maintain. Because of the different moulding process for different moulds, the profiles will differ significantly in length and shape. This could potentially be handled by introducing recurrent neural networks (RNN) that are unaffected by shifts on the x-axis (time direction).
- In a start-up situation (change of material or colour) machine profiles could potentially be used to monitor when the melt has become stable and homogeneous, hence producing stable product quality. One approach that could be explored is to monitor the variation between a number of consecutive moulding cycles, where the expectation would be that variation should be reduced in time and become stable when the melt conditions are homogeneous.
- The focus in the conducted work has been related to machine profiles and material proprieties. Alternative applications could be imagined within utilising the machine profiles for condition-based maintenance applications. This could be using system pressure to monitor mould condition, or system pressure to monitor the condition of the non return valve.

I see the approach of utilising underlying time-resolved equipment data as applicable for other industries as well. It is especially the case for other mixing processes, where the energy consumption can be collected as function of time[2]. An example was presented by Fertin Pharma in 2016 [76], where the energy consumption in a chewing-gum mixer was utilised to detect material differences caused by change in raw materials. This insight was used for root cause analysis and considered implemented for process monitoring.

9.4 Condition-based mould maintenance

An injection mould is an expensive and intricate construction that requires regular maintenance to continuously produce elements of high quality. The most frequent preventive maintenance is lubrication of moving parts and cleaning of the closing surface (including air vents). This can be performed in a short time while the mould is placed in the moulding machine. On a less frequent interval, the mould is removed from the moulding machine, disassembled, cleaned, individual parts inspected, lubricated, assembled and reinstalled in the moulding machine. This maintenance often requires a production stop of up to 12 hours. At the industrial partner the maintenance schemes are executed on fixed production intervals (twice a week and every six weeks), despite that the optimal interval may differ from mould to mould. To reduce risk of mould malfunction, maintenance intervals are selected to fit moulds requiring

[2]For pumps, it could be collection of energy consumption as function of flow rate.

most frequent maintenance. Regardless of preforming preventive maintenance mould malfunction still occurs. This can be caused by sudden breakage of mould parts (e.g. latch locks as described in paper C) or undetected wear and tear on mould parts. Introducing condition-based maintenance is an option to optimise cost of preventive maintenance and increase likelihood of detecting mould malfunction before a breakdown that could potentially cause severe damage to the mould. The foundation for implementing condition-based maintenance is that signals reflecting the mould condition are available. Readily available process data from the injection moulding machine is not optimised to reflect mould conditions and therefore not likely to contain the needed information. An alternative approach, using acoustic emissions, has therefore been investigated for detection loose/defective latch locks and insufficient lubrication (presented in paper C). It has been found that both error types can be detected on individual moulds using features (spectral centroid) extracted from the acoustic emissions and classification using support vector machine (SVM). To reduce the number of models to develop and maintain, it is desirable to create global models that can be used across multiple moulds. Across moulds large time variations and differences in events occurring in the mould opening and closing sequence are seen. Since SVM can't handle shifts in the signal along the x-axis it can not be used when combining multiple moulds in one model. We therefore present an alternative solution utilising acoustic emissions combined with generative Gaussian modelling. Still this approach lacks the ability to generalise well between moulds. Using the assumption that moulds after maintenance are in faultless condition we introduced a model adaptation (based on the first few faultless moulding cycles at test time) that improves the generalisation ability. The proposed solution requires a minimum of computation resources making it practically feasible in a real industrial context with multiple installations. The data used was generated in a controlled experiment, and therefore needs to be validated in real settings. To improve the model performance, more data needs to be included and the assumption for model adaptation has to be confirmed in practice.

The achieved results have served as proof of concepts at the industrial partner. Based on this, industrial grade acoustic monitoring equipment have been purchased for 68 moulding machines. Exploration and future development of applications within acoustic emissions for condition-based mould and machine maintenance will continue together with an external partner. This will include exploration and comparison of using acoustic emission and machine profiles for process monitoring. With the installation of acoustic recording systems on multiple injection moulding machines new areas of research will appear:

- The amount of data collected introducing high resolution acoustic recording will demand research around distribution between edge and cloud analytics. It will also be relevant to investigate what data to collect and use for model development and maintenance.
- As described for machine profiles it is highly relevant to investigate how global

models can be used to reduce the number of models to develop end maintain. Utilising spectrograms this could potentially be achieved using recurrent (RNN) or convolutional neural network (CNN).

9.5 Experimentation in an off-line setting

In section 2.2, I refereed to the FDA initiative of QbB and PAT, where the objective is to improve product quality and consumer safety though improving understanding of the production setup. I fully support this approach and see ”process understanding” as a prerequisite for all efficient and lasting process and production improvements. Often it can be challenging to achieve the needed level of insight and understanding by collecting process data from the current version of the production process. Production processes are designed to be as stable as possible, hence containing limited amounts of variation. This variation is often driven by underlying dynamics and uncontrollable disturbances not recorded and therefore variation with no context. To enrich the data with structured variation, needed for increasing process understanding, it is necessary to conduct systematical tests. Running this type of tests can often be challenging in a production setup, hence addressed in research question 2.

The investigation in paper A, B and C are all conducted in an off-line setting intended for running designed experiments, whereas the work in paper D, F and investigation of MSPC-PCA on injection moulding was performed on data collected from production systems. The benefit of conducting controlled experimentation is that specific variables can be adjusted in a controlled way, enabling interpretability and causality. The drawback is that it is often difficult to utilise the results directly in the actual manufacturing setup, where many factors are uncontrolled. The benefits of utilising readily available data collected in the running production, is that data is already collected and it is reflecting the actual conditions impacting product quality. The limitations are that the data collected do not reflect all disturbances impacting the process and product quality, reducing the insights to correlation and not causality. The ultimate goal therefore has to be a combined approach where benefits from both manufacturing data and off-line experimentation can be utilised. For this to be possible the off-line experimentation setup must be comparable to the actual manufacturing setup, with the flexibility to move process settings outside the normal range and to control known disturbances factors. Preferable a dedicated part of the existing manufacturing setup can be used for testing of alternative process settings, raw material variations, process disturbances, evaluation of alternative sensors and data utilisation concepts. This makes it possible to combine data from machines used for both standardised production and designed and controlled experiments. With this enabled, new research opportunities will emerge e.g. how to use variation patterns in manufacturing data to guide design of experiments to turn correlation into causality.

9.6 Process and quality monitoring

Process monitoring has proven to be challenging within injection moulding. Many aspects are impacting this; data availability, how to define phase I (what is an in-control state?), and autocorrelation. It has been shown that both quality measures and process variables are highly correlated and that approaches as MSPC-PCA are needed for monitoring. It has been demonstrated that MSPC-PCA can be used to monitor element quality and that this can be linked to mould degradation. MSPC-PCA is doable in a simple and controllable setup, but shown challenging in a setup with multiple operating regimes and unclear links to quality. Some questions and reflections arise based on this:

- It could be questioned if process monitoring based on readily available process data is the approach to pursue? Since readily available process data don't seem to reflect the state of the melt, it might be better to utilise machine profiles for process monitoring. This has not been explored and therefore an area for future research.

- Process data have been aggregated per p-box (production unit containing elements from multiple moulding cycles) to cover autocorrelation and link process data to quality evaluation. Based on the achieved results it is questionable if this is the optimal approach. It might be possible to improve results if the aggregated measures used reflect the within p-box variation better. Alternatively, data from individual cycles should be used in combination with methods designed to handle autocorrelation (e.g. dynamic-PCA or ARIMA[3] models).

- An alternative approach would be to move away from a unsupervised monitoring approach and instead consider machine learning based regression or classification for specific applications. This could be online prediction of element dimensions (mentioned in relation to utilising machine profiles), prediction of non-conformity or applications within condition-based maintenance. This will require experimental work to ensure causality in the models and it must be based on global models to minimise the number of models to develop and maintain.

Despite the fact that the above reflects the production setup at the industrial partner, I see the reflections as relevant for other industries as well.

[3]AutoRegressive Integrated Moving Average, that can be used to model out autocorrelation, where the monitoring will performed on the residuals from fitting the ARIMA model to new unseen data

9.7 Reflections and suggestions for the industrial partner

The work conducted in the PhD is not completely aligned with the original intentions of investigating process monitoring of injection moulding. This said, I think that the resulting content, learnings and reflections are relevant and value adding for the industrial partner. The proposed framework for utilising manufacturing data for process optimisation will enable a much more structured and prioritised implementation of digital solutions (linked to RBO1). The data utilisation framework can at the same time be used to guide the development of the needed data foundation and analytic infrastructure. The work conducted around moulding state variables investigating the relation between information from readily available process variables, machine profiles and signals from cavity sensors brings new insights that can be utilised in various ways (linked to RBO2). The exploration of using acoustic emissions for condition-based mould maintenance has shown promising results and will be continued with installation of sensors on 68 moulding machines (linked to RBO3). I see the following concrete aspects as crucial to consider moving forward:

- Utilisation of machine profile data has proven to be essential moving forward. It is therefore recommended that collection of the profile data is enabled on a series of injection moulding machines in the manufacturing area. This can used to confirm that changes in material properties (caused by dual sourcing) can be detected in a manufacturing context. It will also enable the possibility to explore process monitoring based on machine profiles and compare this against results obtained using readily available process data.

- To improve the business case of installing an acoustic recording solution for condition-based mould maintenance it should be considered how the same equipment can be utilised for other applications. The most straightforward application to explore is to place additional sensors on the moulding machine for condition-based machine maintenance. With initial results indicating that acoustic emissions can be used for monitoring the state of the melt, it should be investigated to what extent this can be utilised and compared with results obtained using machine profiles. This will be essential for the overall cost/benefit evaluation.

- Many of the proposed and future digital solutions would require testing in a real manufacturing environment to estimate the true business potential. It is therefore recommended to dedicate a specific part of the moulding area (e.g. one production module consisting of 68 moulding machines) for developing and testing new sensors and digital solutions. Focusing development and test in one area would make it possible to also explore synergies between different applications. Operators, quality personnel and service technicians in such testing area should be involved the in development and testing of the applications. This would ensure the right focus, commitment and valuable end-user feedback.

- The results achieved utilising existing quality data for monitoring mould degradation, have demonstrated that the collected data can be utilised for more than product release. I see that as an opportunity to create additional value of the resources already use for quality inspection and analysis.

With the complexity represented in the manufacturing setup at the industrial partner, I would propose to select a dominant subgroup of machine types, mould types, element shapes and materials and focus the exploration and development for this subgroup. When a usable solution has been identified, it can be considered if it is feasible to scale the solution to include other subgroups. Aiming for a solution that can cover the full complexity from the start will make it very difficult to develop any solutions. Based on past experiences, I would recommend that the possibility of collecting machine profiles[4] is included as a requirement when ordering new injection moulding machines. Machine profiles should then be collected as default together with a collection of readily available process variables. When the potential business value is fully assessed on new machines it can be evaluated if older machines should be upgraded to also collect machine profiles.

[4]Collection should consists of screw position, injection speed and system pressure for hydraulic machines and screw pressure on new electric machines. These should be collected for the full moulding cycle with a resolution of minimum 500 data points.

CHAPTER 10

Conclusion

With the introduction of Industry 4.0, new focus, and hype, has been placed on data utilisation in manufacturing. Part of the hype is driven by an expectation that data utilisation is a simple solution to obtain substantial productivity gains. That productivity gains can be achieved, is likely. That this should be simple, is far from true. Many tasks related to efficient data utilisation are still not addressed and understood, resulting in failed attempts to introduce analytical applications within manufacturing.

The work presented in the thesis has been focused on improving productivity through utilising data within manufacturing. A practical framework centred around identifying a clear business objective has been presented to explore and implement data utilisation applications. Using this framework, it has been demonstrated that valuable applications can be developed when data suitable for the specific applications are collected in a structured way. The approach of ”just” collecting available data, assuming that machine learning can magically extract meaning and insights that will optimise a manufacturing setup, is naive. Based on an understood business objective, data should be collected with the purpose of data utilisation and thereby be designed to contain the needed information.

The work conducted has demonstrated the benefit of focusing on individual applications and exploring these in a sequential matter. There might be various data sources for a given application that can be utilised to achieve the same business objective. Therefore, the total cost of utilising different data sources has to be evaluated against the potential benefits. The financially most feasible and technically implementable solution can then be selected.

It is crucial not to underestimate the complexity of utilising data in a manufacturing setup. Data needed for effective data utilisation originate from various sources, including; manufacturing equipment, quality systems, ERP and MES systems. This data integration has to be handled to realise the potential benefits. The data infrastructure should be built sequentially as applications are implemented. It is not feasible to start implementing the full expected data foundation; hence, identifying all the needed data sources and interactions between them would not be possible. The focus in the conducted work has been on predictive applications; however, as described in the data utilisation roadmap, business value can also be achieved by implementing simple solutions, like, e.g. data visualisation. Starting the journey by implementing simple digital solutions will support maturing the organisation for implementing predictive solutions. It is crucial to realise that the journey has to be a joint effort, including management, business subject matter experts, data engineers and data scientists. The focus should be centred around a desire to reach sustainable solutions that are based on data-driven insights and process understanding.

I see great potential in utilising manufacturing data for process monitoring, improving of process understanding and process optimisation. With a dedicated effort and focus, I am convinced that this will play a significant and dominating role in future manufacturing.

"As for the future, your task is not to foresee it, but to enable it."
— Antoine de Saint Exupery, *best known for his story, The Little Prince.*

Bibliography

[1] S. Farahani, N. Brown, J. Loftis, C. Krick, F. Pichl, R. Vaculik, and S. Pilla, "Evaluation of in-mold sensors and machine data towards enhancing product quality and process monitoring via industry 4.0.," *International Journal of Advanced Manufacturing Technology*, vol. 105, no. 1-4, pp. 1371 – 1389, 2019.

[2] M. Zeaiter, W. Knight, and S. Holland, "Multivariate regression modeling for monitoring quality of injection moulding components using cavity sensor technology: Application to the manufacturing of pharmaceutical device components," *Journal of Process Control - J PROCESS CONTROL*, vol. 21, pp. 137–150, 01 2011.

[3] T. Mao, Y. Zhang, H. Zhou, D. Li, Z. Huang, and H. Gao, "Data driven injection molding process monitoring using sparse auto encoder technique," in *2015 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 524–528, 2015.

[4] Z. M. Çınar, A. Abdussalam Nuhu, Q. Zeeshan, O. Korhan, M. Asmael, and B. Safaei, "Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0," *Sustainability*, vol. 12, no. 19, 2020.

[5] T. Kek, D. Kusi, A. Han, and J. Grum, "Acoustic emission crack detection in injection molding," in *11th European Conference on Non-Destructive Testing (ECNDT 2014), Prague 2014, Oct 6-11 (ECNDT 2014)*, 2015.

[6] E. Moreira, F. Alves, M. Martins, G. Ribeiro, A. Pina, D. Aguiam, E. Sotgiu, E. Fernandes, and J. Gaspar, "Industry 4.0: Real-time monitoring of an injection molding tool for smart predictive maintenance," in *25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 1209–1212, 09 2020.

[7] M. Gökalp, K. Kayabay, M. A. Akyol, P. Eren, and A. Koçyiğit, "Big data for industry 4.0: A conceptual framework," pp. 431–434, 12 2016.

[8] A. G. Frank, L. S. Dalenogare, and N. F. Ayala, "Industry 4.0 technologies: Implementation patterns in manufacturing companies," *International Journal of Production Economics*, vol. 210, pp. 15–26, 2019.

[9] C. Gröger, R. P. Bosch, and Robert-Bosch-Platz, "Building an industry 4.0 analytics platform. practical challenges, approaches and future research directions," 2018.

[10] B. Gajdzik, S. Grabowska, and S. Saniuk, "A theoretical framework for industry 4.0 and its implementation with selected practical schedules," *Energies*, vol. 14, no. 4, 2021.

[11] D. Kazmer, S. Westerdale, and D. Hazen, "A comparison of statistical process control (spc) and on-line multivariate analyses (mva) for injection molding," *International Polymer Processing - INT POLYM PROC*, vol. 23, pp. 447–458, 11 2008.

[12] P. Clausen, J. B. Mathiasen, and J. S. Nielsen, "Barriers and enablers for digitizing shop floor management boards," in *2018 Global Wireless Summit (GWS)*, pp. 288–293, 2018.

[13] M. Djassemi and J. A. Sena, "The paperless factory: A review of issues and technologies," *International Journal of Computer Science and Network Security*, vol. 6, no. 12, p. 185, 2006.

[14] G. H. Gaynor, "Innovation: top down or bottom up," *IEEE Engineering Management Review*, vol. 41, no. 3, pp. 5–6, 2013.

[15] Y.-F. Hsu, R. Irie, S. Murata, and M. Matsuoka, "A novel automated cloud storage tiering system through hot-cold data classification," in *2018 IEEE 11th International Conference on Cloud Computing (CLOUD)*, pp. 492–499, 2018.

[16] T. Zhang and B. Yang, "Big data dimension reduction using pca," in *2016 IEEE International Conference on Smart Cloud (SmartCloud)*, pp. 152–157, 2016.

[17] "Data visualisation using Microsoft Power BI, Microsoft Power BI." https://powerbi.microsoft.com/en-za/. Accessed: 2021-12-13.

[18] "Data visualisation using Tableau, Tableau." https://www.tableau.com/products/our-platform. Accessed: 2021-12-13.

[19] "Using SAS JMP for statistical process controls to improve and monitor your process, SAS JMP." https://www.jmp.com/en_ph/events/ondemand/technically-speaking/using-statistical-process-controls-to-improve-and-monitor-your-process.html. Accessed: 2021-12-13.

[20] "Using NWA for statistical process control, Northwest analytics." https://www.nwasoft.com/products/nwa-quality-analyst/features-and-benefits. Accessed: 2021-12-13.

[21] "DataRobot AI Cloud Platform, DataRobot." https://www.datarobot.com/. Accessed: 2021-12-13.

[22] "Research area that targets progressive automation of machine learning AutoML, AutoML." https://www.automl.org/. Accessed: 2021-12-13.

[23] H. Kagermann, W. Wahlster, and J. Helbig, "Recommendations for implementing the strategic initiative industrie 4.0," final report of the industrie 4.0 working group, München, 2013.

[24] "Illustration - Industry 4.0 Fourth Industrial Revolution, technology, electronics, mode Of Transport, engineering png, pngwing." https://www.pngwing.com/en/free-png-hoyzq#google_vignette. Accessed: 2021-12-16.

[25] "Morse Code & the Telegraph, HISTORY." history.com/topics/inventions/telegraph.

[26] "Ford's assembly line starts rolling, HISTORY." https://www.history.com/this-day-in-history/fords-assembly-line-starts-rolling.

[27] "Programmable controllers: How it all began, WORKFORCE DEVELOPMENT." https://www.plantengineering.com/articles/programmable-controllers-how-it-all-began/.

[28] "A Brief History of Robots in Manufacturing, Mariane Davids." https://blog.robotiq.com/a-brief-history-of-robots-in-manufacturing.

[29] "Illustration - Keywords that defines the fourth industrial revolution., pngwing." https://www.pngwing.com/en/free-png-yfczb. Accessed: 2021-12-16.

[30] "Industry 4.0 and manufacturing ecosystems, Deloitte." https://www2.deloitte.com/tr/en/pages/manufacturing/articles/industry-4-0-manufacturing-ecosystems-exploring-world-connected-enterprises.html.

[31] R. Stark and T. Damerau, *Digital Twin*, pp. 1–8. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019.

[32] R. Baheti and H. Gill, "Cyber-physical systems," *The impact of control technology*, vol. 12, no. 1, pp. 161–166, 2011.

[33] J. Lee, B. Bagheri, and H.-A. Kao, "A cyber-physical systems architecture for industry 4.0-based manufacturing systems," *Manufacturing Letters*, vol. 3, pp. 18–23, 2015.

[34] H. Ahuett-Garza and T. Kurfess, "A brief discussion on the trends of habilitating technologies for industry 4.0 and smart manufacturing," *Manufacturing Letters*, vol. 15, pp. 60–63, 2018. Industry 4.0 and Smart Manufacturing.

[35] F. Yang and S. Gu, "Industry 4.0, a revolution that requires technology and national strategies," *Complex & Intelligent Systems*, vol. 7, 01 2021.

[36] W. A. Shewhart, "Quality control charts," *The Bell System Technical Journal*, vol. 5, no. 4, pp. 593–603, 1926.

[37] J. M. Juran, *Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services*. Free Press, 1992.

[38] "Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, FDA." https://www.fda.gov/media/71012/download.

[39] A. M. Turing, "Mind," *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[40] Y. Pan, "Heading toward artificial intelligence 2.0," *Engineering*, vol. 2, no. 4, pp. 409–413, 2016.

[41] A. Kaplan and M. Haenlein, "Siri, siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence," *Business Horizons*, vol. 62, no. 1, pp. 15–25, 2019.

[42] "Illustration - artificial intelligence definition. pngwing." `https://www.pngwing.com/en/free-png-tgvpc`. Accessed: 2021-12-23.

[43] "What Is Machine Learning & Deep Learning?, Claire D. Costa, Medium." `https://medium.com/@harish_6956/what-is-machine-learning-deep-learning-7788604004da`. Accessed: 2021-12-23.

[44] G. Neff and P. Nagy, "Automation, algorithms, and politics| talking to bots: Symbiotic agency and the case of tay," *International Journal of Communication*, vol. 10, no. 0, 2016.

[45] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, pp. 484–503, 2016.

[46] "Illustration - Machine learning classes, pngwing." `https://www.pngwing.com/en/free-png-yzvfe`. Accessed: 2021-12-23.

[47] "Injection Molded Plastics Market Size, Share & Trends Analysis Report By Raw Material (Polypropylene, ABS, HDPE, Polystyrene), By Application (Packaging, Automotive & Transportation, Medical), By Region, And Segment Forecasts, 2021 - 2028, Grand View Research, Inc., USA." `https://www.grandviewresearch.com/industry-analysis/injection-molded-plastics-market`. Accessed: 2021-11-16.

[48] "Statistikbanken, tabel NAN1: Forsyningsbalance, bruttonationalprodukt (BNP), beskæftigelse mv. efter tid, prisenhed og transaktion, Danmarks Statistik." `https://statistikbanken.dk/nan1`. Accessed: 2021-11-16.

[49] A. Bernasconi, D. Rossin, and C. Armanni, "Analysis of the effect of mechanical recycling upon tensile strength of a short glass fibre reinforced polyamide 6,6," *Engineering Fracture Mechanics*, vol. 74, no. 4, pp. 627–641, 2007. Fracture and Damage in Engineering Applications.

[50] M. Košík, L. Likavčan, J. Bílik, and M. Martinkovič, "Diesel effect problem solving during injection moulding," *Research Papers Faculty of Materials Science and Technology Slovak University of Technology*, vol. 22, no. 341, pp. 97–102, 2014.

[51] K.-T. Chiang and F.-P. Chang, "Analysis of shrinkage and warpage in an injection-molded part with a thin shell feature using the response surface methodology," *The International Journal of Advanced Manufacturing Technology*, vol. 35, no. 5-6, p. 468–479, 2007.

[52] D. Mathivanan and N. S. Parthasarathy, "Sink-mark minimization in injection molding through response surface regression modeling and genetic algorithm," *The International Journal of Advanced Manufacturing Technology*, vol. 45, no. 9-10, p. 867–874, 2009.

[53] M. Moayyedian, K. Abhary, and R. Marian, "The analysis of short shot possibility in injection molding process," *The International Journal of Advanced Manufacturing Technology*, vol. 91, pp. 1–13, 08 2017.

[54] J. Sreedharan and A. Jeevanantham, "Optimization of injection molding process to minimize weld-line and sink-mark defects using taguchi based grey relational analysis," *Materials Today: Proceedings*, vol. 5, no. 5, p. 12615–12622, 2018.

[55] L. W. McKeen, "1 - introduction to plastics and polymers," in *The Effect of UV Light and Weather on Plastics and Elastomers (Fourth Edition)* (L. W. McKeen, ed.), Plastics Design Library, pp. 1–20, William Andrew Publishing, fourth edition ed., 2019.

[56] "A Detailed Guide on Acrylonitrile Butadiene Styrene, Omnexus." `https://omnexus.specialchem.com/selection-guide/acrylonitrile-butadiene-styrene-abs-plastic?src=prop-cnx`. Accessed: 2021-12-28.

[57] "BIMS-1 (Injection Molding) and BIMS-2 (Mechanical Properties), Dr. Vito LEO." https://bims-seminars.thinkific.com/.

[58] "Therma-flo - Injection Molding Analysis Software, Beaumont." https://www.beaumontinc.com/therma-flo/.

[59] "Understanding polymer flow interpreting the viscosity curve, John Bozzelli, Injection Molding Solutions." https://www.ptonline.com/articles/understanding-polymer-flow-interpreting-the-viscosity-curve.

[60] S. Moser, "Effective run-in and optimization of an injection molding process," 2012.

[61] "Plastic injection and compression mold simulation for design and manufacturing, AUTODESK."

[62] "Moldex3D is the world leading CAE product for the plastics injection molding industry, Moldex3D."

[63] P. Thyregod, H. Spliid, H. Melgaard, and H. Madsen, "Modelling and monitoring in injection molding," 2001.

[64] C. Hopmann and M. Schmitz, "Plastics industry 4.0: Potentials and applications in plastics technology," in *Plastics Industry 4.0* (C. Hopmann and M. Schmitz, eds.), pp. I–XIII, Hanser, 2021.

[65] "Switch Over To Consistent Quality, Prof. Dr.-Ing. Georg Steinbichler." https://www.ptonline.com/articles/switch-over-to-consistent-quality.

[66] D. O. Kazmer, "Precision injection molding precision process control of injection molding," 2003.

[67] "G. Rønsch 2019, SPC i et multivariat system – en praktisk tilgang, presented at Få processen i kontrol med Statistisk Proceskontrol (SPC) conference at IDA." https://ida.dk/viden-og-netvaerk/videoer-fra-ida/kategorisider/videoer-produktion-og-industri/faa-processen-i-kontrol-med-statistisk-proceskontrol-spc. See video presentation from 1:38:00.

[68] D. C. Montgomery, "Statistical quality control: A modern introduction, 7th edition international student version," in *Statistical Quality Control*, Wiley, 2012.

[69] J. MacGregor and T. Kourti, "Statistical process control of multivariate processes," *Control Engineering Practice*, vol. 3, no. 3, pp. 403–414, 1995.

[70] B. De Ketelaere, M. Hubert, and E. Schmitt, "Overview of pca-based statistical process-monitoring methods for time-dependent, high-dimensional data," *Journal of Quality Technology*, vol. 47, pp. 318–335, 10 2015.

[71] T. Villegas Berbesi, M. Jesús, Fuente, and R. Miguel, "Principal component analysis for fault detection and diagnosis. experience with a pilot plant," *International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics - Proceedings*, 01 2010.

[72] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2019.

[73] "Librosa is a python package for music and audio analysis."

[74] S. Prasomphan, "Detecting human emotion via speech recognition by using speech spectrogram," in *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1–10, IEEE, 2015.

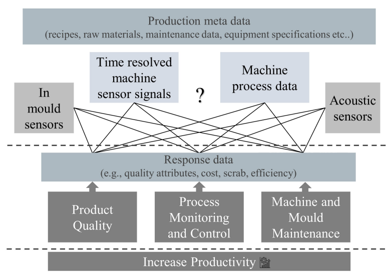
[75] V. Tiwari, "Mfcc and its applications in speaker recognition," *International journal on emerging technologies*, vol. 1, no. 1, pp. 19–22, 2010.

[76] "Fra BIG DATA til information, Datakilder, Transformation, Visualisering og modellering. Presented at SESAM by Georg Ørnskov Rønsch, 2016." https://sesam-world.com/_pdf/sesam-117/05-Fertin.pdf. Accessed: 2022-02-08.

Appendices

APPENDIX A

Paper D, Experiences with big data: Accounts from a data scientist's perspective



With the introduction of Industry 4.0, new focus, and hype, has been placed on data utilisation in manufacturing. Part of the hype is driven by an expectation that data utilisation is a simple solution to gain substantial productivity. That productivity gain can be achieved is likely. That this should be simple, is far from true. Many of the tasks related to efficient data utilisation are still not addressed, resulting in failed attempts to introduce analytic data applications. During the last few years, many new data analytical solutions have appeared, all focusing on making data analytics available for everyone. This is essential in making data science tools available for business and subject matter experts. These analytic solutions still mainly support data structured in tabular form, which is seldom the case in real-life applications. Therefore, it is necessary to combine business understanding, data engineering and data science skills for value creation to appear. Paper D addresses this with experience built through multiple data utilisation projects. This links paper D to RQ1 through reflections on challenges encountered when utilising data in a manufacturing context.

Research question 1 (RQ1)

How to best utilise data from injection moulding machines to increase the productivity of producing injection moulded elements?

The work is a collection of experiences and lessons learned from collaboration with numerous companies exploring and developing Big Data applications for production analytics.

A.1 Summary

The work is a collection of experiences obtained as a university group engaged with companies in exploring the use of data for production optimisation.

Currently, available data often lacks essential information to make it suitable for effective data utilisation. This can be data traceability and connection between the predictors and responses or completely missing predictors and response variables. This mainly reason for this is that the data collection has been initiated with the business objective being documentation and not production optimisation. Therefore, a recommendation is that data utilisation activities should be forward-looking, starting with the appropriated business objective.

More data analytic solutions become available for doing more or less automated data modelling, where the data have to be structured in a tabular format for it to work. Data from real industrial applications come in many different forms and structures. Data transformation is all about structuring and organising data for it to be suitable for meaningful data analytics. This takes experience in data science, but most importantly, it requires business insights and understanding, and therefore not something that can be automated in a software solution.

When engaging in data utilisation, it is crucial to recognise the current data maturity level present within the specific part of the company. Often digital readiness and data maturity differs within the companies, and we have experienced that the manufacturing areas of many companies do not have the required maturity level when initiating a data utilisation journey. This often is reflected in big ambitions without the needed insight to achieve the desired productivity gains. A sequential approach is presented to ensure that maturity develops as data utilisation progresses.

Achieved results

- Discussion of pitfalls and challenges in approaching data utilisation in a manufacturing context.

Contribution

- Collection of crucial insights and learnings gained through practical experience with data utilisation within different industries.
- Reflection on how to effectively engage with the utilisation of manufacturing data for process and production optimisation.

A.2 Paper D

Paper D

Experiences with Big Data: Accounts from a Data Scientist's Perspective

Murat Kulahci[1,2], Flavia Dalia Frumosu[1], Abdul Rauf Khan[1], Georg Ørnskov Rønsch[1], Max Peter Spooner[1]

[1]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark

[2]Department of Business Administration, Technology and Social Sciences, Luleå University of Technology, Luleå, Sweden

Abstract

Manufacturing has been rejuvenated by automation and digitalization. This has brought forth the new industrial era also called Industry 4.0. During the last few years we have collaborated with companies from various industries that have all been going through this transformation. Through these collaborations, we have collected numerous examples of (sometimes troublesome) experiences with Big Data applications of production analytics. These experiences reflect the current state of production data and the challenges it poses. Our goal in this paper is to share those experiences and lessons learned in dealing with practical issues from data acquisition to data management and finally to data analytics.

Keywords: Industry 4.0, Manufacturing, Digitalization, Big Data, Production Analytics

1. Introduction

A new industrial revolution is upon us. Aptly named Industry 4.0 (Lasi et al. (2014)), this fourth revolution is mainly driven by automation and information exchange with heavy reliance on digitalization, see Figure 1. Accordingly, many companies from various industries are all going through this transformation where the goal is to digitalize the information content in production and hence enhance knowledge about their processes. Usually, the expectation is that the abundant production data should contain information that can be used to further the goals of the company (e.g., reduce operating costs and/or increase profits). Moreover, the edge some companies have obtained through past movements such as Lean Production and Six Sigma may be winding down resulting in a search for new approaches.

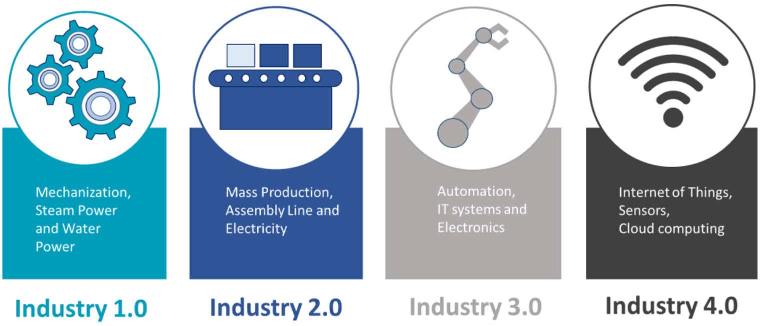


Figure 1. Industrial revolutions from the 18th century until the present

In their pursuit towards digitalization, many companies from a wide range of industries are now frantically scrambling to collect "more data" from their processes under the wishful thinking that it contains the "necessary information." It has however been our unfortunate and yet repeated observation that this is done in a frenzy with little consideration to key questions such as:

- What kind of problems do we actually want to solve?
- What kind of data is in fact needed and how can we collect such data efficiently?
- How can we handle such data properly and how can we make sure the extracted information is used most effectively?

Instead, this haste in data collection often generates more problems than it solves. As a response, academics and practitioners alike rush into yet another frenzy of dealing with these problems, which could very well be simple artifacts of poor planning in data collection or in problem solving. Over the years, we have had the privilege of working on many projects in collaboration with industry partners where the analysis of (big) production data has been the focus. These projects involved various industries ranging from high volume parts manufacturing as in injection molding and electronics as well as batch process manufacturing in food, pharmaceuticals and bio-based production. Granted our collaborations may not constitute a "big enough" sample for overreaching generalizations, but the small amount of variation in the common hurdles we needed to overcome leads us to believe that at least some of these issues are quite widespread. It should also be noted that throughout the paper, we restrict our attention to data analytics in manufacturing and hence use the term *production analytics* to summarize our efforts in product development and improvement as well as in process improvement, optimization, monitoring and control. Some of our current work in predictive maintenance as presented in Section 5.4, *New Opportunities*, would also fall under this umbrella. Since the experiences cited in this paper are mainly in manufacturing, we encourage readers from other fields of data analytics applications to draw their own parallels to our conclusions with caution and as they see fit.

In the rest of the paper, we present our observations regarding general expectations in the Big Data movement followed by a discussion of issues related to data management, content and analysis. We conclude our paper with the concept of digitalization readiness, as we have perceived it through our interactions with our industrial partners.

2. A pre-requisite: Matching expectations

Many of our production analytics collaborations involve two distinct parties: problem owner (industry) and problem solver (consultant/academia). Needless to say, both of these parties need to work together for a successful completion of any production analytics project. However, the increased complexity of production challenges has resulted in an ever-widening gap between academic output and industrial needs. The examples included in textbooks remain too sterile and new theoretical solutions developed by the academic research tend to get buried in journals that are not on the industry radar. Moreover, as in any other relationship, the success in a production analytics collaboration is closely tied to matching the respective expectations. That is, while close collaboration and commitment are preliminary conditions, we need to go further and establish early on a realistic mutual understanding of what is to be achieved.

The current inclination in industry, as they face an increasing pressure in highly competitive markets, is whenever possible to simply hand over the production data to the data analyst for the discovery of any hidden information. Another expectation from the industry is to obtain a fast and simple solution. We should at this point re-emphasize that this paper refers to our collaboration with industry as an external agent. There are companies with expanding in-house data science teams established towards achieving similar goals. Yet their journey may be plagued by problems that would be different from the ones we have encountered and therefore we refrain from elaborating on them further.

The seemingly insurmountable amount of production data and increased complexity of the problems combined with the reassuring promises from software companies and academics to deliver “the” answer through empirical evidence may lead the industry to assume a rather passive role. This should however, not be taken as a criticism since it is a natural reaction to ever evolving targets and the progression of digitalization and Internet of Things (IoT) by the production engineers who are not necessarily trained to address these new challenges. When we couple this with the increasing concerns of upper management believing that their competitors have a much better handle on digitalization, it is no wonder that we often hear companies proclaiming “We need AI!”. This however is seldom accompanied with a clear vision of its ultimate use and purpose. In our limited view, we may certainly be overlooking a potential “If you build it, he will come!” moment as in the movie “Field of Dreams” but we do nonetheless strongly believe that there are still plenty of opportunities for the use of real intelligence when it comes to manufacturing. This however, requires the industry to actively participate in data analytics efforts and not simply remain as an outside spectator expecting a data machinery to spit out an outcome that will alleviate all their problems. This includes not only providing useful input and feedback based on their process knowledge but also understanding and appreciating the capabilities and more importantly the shortcomings of the methods used therein.

The data analysts, on the other hand, tend to have little to no knowledge about the processes they work on. Yet they are usually conditioned to start any data analysis with defining the problem. This should not come as a surprise as most data analysts, similar to engineers, are trained in the basic scientific method in which setting up the hypothesis constitutes a crucial step. In the past, data analysts would often rely on the industry partner's expertise to help with this important juncture. As challenging as this may be at times, most engineers would have a reasonable handle on their process and could readily come up with suggestions. In case of predictive modeling, for example, it would not be too far-fetched to think that the engineers could provide a list of output(s) and potential inputs based on past experience and process knowledge. This would often be possible as that list would have usually consisted of a relatively limited number of process variables which were deemed important enough to collect the data for in the first place. In modern production however, the list of potential inputs and outputs has for almost any process expanded with the advances in sensor, data acquisition and storage technologies. It should be noted that this expansion has not necessarily been a direct result of a well thought out process in data gathering but rather a side-effect of a haste in collecting, often indiscriminately, everything we can get from the process. Unfortunately this results in stretching the process knowledge of the industry partners beyond their comfort zone and often prevents them from offering more concrete help in defining the problem to tackle. Hence to avoid an impasse, the data analysts have to adjust their expectations about the help they would traditionally require in defining the problem and seek to be part of that process based on empirical evidence. For example, if historical data is available, a preliminary analysis involving all potential variables, both as outputs and inputs, should be performed and observed correlations should be examined with the process experts for relevance and practicality. That in turn would help to define the problem and lead to a more focused study. In that sense, historical data can serve in setting up hypotheses. However, if the historical data does not reveal relevant information due to the many reasons we will discuss in the following sections, this initial study would still serve as the basis for what to do next in terms of what type of data is needed and how to collect it. In the next sections, we discuss many data related issues we have faced and provide, when possible, solutions we have adopted to overcome these issues.

3. Data Management

The promise of Big Data is one of the main characteristics of digitalization of modern production. Yet it often comes at a price. Before indulging in the content and the analysis of data, one has to deal with data management issues. Below, we discuss some of those issues. In many ways, our experiences and conclusions presented in this and subsequent sections follow the same path of data readiness levels suggested by Lawrence (2017), where the author proposes three data readiness levels concerning data accessibility, cleanliness and relevance.

3.1 Accessing the Data

Until recently, access to data would be achieved primarily through floppy disks, compact disks (CD's), memory sticks and attachments to emails. The sheer volume of production data from modern manufacturing processes renders these data transfer media inadequate.

Nowadays, we often resort to accessing the data directly at its source or at the very least at a virtual data warehouse, e.g., cloud. This however has proven to be quite challenging as well. Connection to cloud storage or to various databases with varying protocols requires different expertise than most data analysts have been trained for. Furthermore, the collaboration with industry is usually established through engineers and operators who themselves have varying levels of understanding of data acquisition systems. Therefore, a close collaboration with IT support in these companies becomes crucial for the success of the project. As IT departments often have other priorities, such as maintaining the security of production data, granting access to an external user usually creates more concerns on their part. Therefore, rather than granting full access to the data, the data analyst is given the data on a need-to-have basis. In this case, the engineer in the collaboration plays the role of an intermediary in providing access to the data. Hence, in a typical transaction, the data analyst would need to go through the engineer to receive the data. This two-step process often delays the data analysis greatly due to communication issues and prioritization mismatches. Ultimately, to avoid delays in data retrieval and analysis, the data analyst has to take the initiative to gain direct access to data. This brings back the original obstacle of the data analyst lacking the necessary background to cope with modern database systems and protocols for data transfer. Even in the case of a data analyst being well versed in these fields, security concerns remain an issue, especially when handling sensitive production data. These concerns are also valid in the use of cloud applications or setting up remote access for the data analyst. Establishing a proper and convenient data access protocol is a must before any serious attempt at data analysis.

In our applications so far, the most reliable option has been gaining access by Virtual Private Network (VPN) to the company's database through a portable computer, which is specifically configured by the company's IT support. This brings additional challenges such as the speed in data processing and security of the device. Yet it still remains to be the preferred solution in many circumstances.

It should also be noted that the IT department and the engineers have different expectations from production data as they have different aims and priorities. Engineers would like to make the best use of the data to achieve important production goals. The IT department, on the other hand, is mainly concerned with data management, which includes acquisition, transfer, storage, security and access. In that regard, what is convenient for one does not necessarily provide the same convenience for the other. One inadvertent outcome of rapid digitalization in most companies has been that the IT department has become the owner of the data with the abovementioned priorities. Hence, engineers much like data analysts will need to adapt to this new situation by making the utmost effort in learning more about data management to establish a common language with the IT department.

As for data analysts working with production data, the days of sitting in front of the computer waiting for the data to arrive are over. The distinction has been made between data analyst and data scientist in that the latter is expected to be more computer savvy and knowledgeable in data management. This transition needs to be made to remain a relevant player in production analytics efforts.

3.2 Merging Databases

As engineers and data analysts working with processes, we tend to focus on process and product data with little to no focus on more managerial issues such as operational and planning concerns. Hence, we often see our work as limited to making sense of production data. As relevant as this is, other types of data are also available and should be used in conjunction with process data to provide relevant and timely information for decision makers. A complete dataset consists of information from the company's Enterprise Resource Planning (ERP) system (Meer (2005)), Manufacturing Execution System (MES) (Meyer et al. (2009)) and historians containing process data (plant data), see Figure 2 (Liu et al. (2002)). These IT systems are often used as separate tools (and therefore not necessarily linked) and in many cases not designed for extraction of large amounts of data, which is mainly the case for ERP and MES. This makes it even more challenging to construct complete datasets encompassing all these systems. This is because of the current limited capacity for data extraction but also because the level of aggregation often differs in these systems. We may retract back to our old modus operandi and isolate our efforts to production. But our prospective solution as great as it may very well be, will most likely fall on deaf ears if it cannot be planned for properly or is not implementable at all.

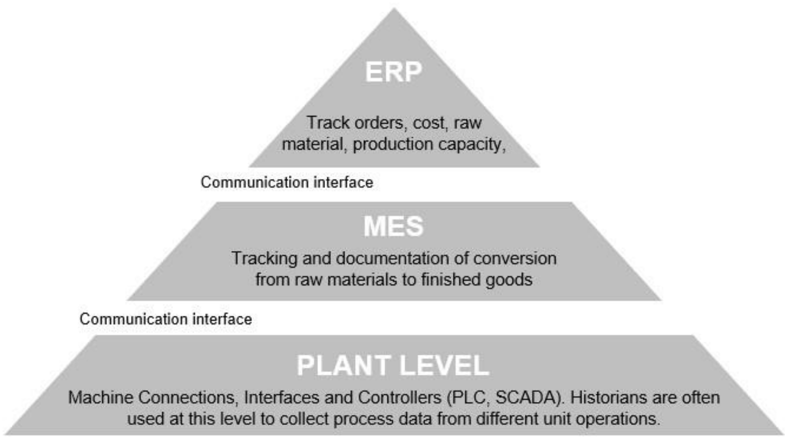


Figure 2. Illustration of a typical data hierarchy within a manufacturing company

Furthermore, besides the main data platforms described above, companies often use specialized systems for lab and quality measurements (e.g., Lab Information Management System (LIMS)) and local systems in R&D functions. This makes it even more challenging to construct the needed dataset. Therefore, with all its challenges, the urgency of connecting various datasets reflecting different components of daily operations is becoming more and more relevant.

3.3 Traceability

Many processes consist of several stages where raw material goes through a series of unit operations before the final product is obtained. When the aim is a comprehensive data analysis solution, which considers all these unit operations at once, traceability becomes a crucial issue. However, true cradle to grave traceability is hard to obtain, particularly in chemical processes. Often different production streams are combined and/or split to make the traceability almost impossible. In one of our projects, the raw material was going through a series of operations in batches before the final product is packaged and shipped.

Each of these operations is very tightly monitored and controlled through a series of sensors and controllers. This naturally gave the impression of the existence of traceability and availability of abundant process data. Upon further analysis, we realized that we actually had data for only a few batches from the beginning to the end. This was not because the data was not available. The data for all batches was simply collected locally for each operation without necessarily considering the preceding or following operations. The batches were tagged accordingly and it required an intimate knowledge of the process to match these tags for different operations. The data analyst working remotely lacked this knowledge and hence ended up with a significantly low number of batches with traceability, which clearly did not represent the reality. This problem was resolved in an accelerated manner by a second data analyst who was more familiar with the process and had direct access to key process engineers and IT personnel.

Another example we have encountered involved a company producing high precision metal parts in a milling process. The metal part undergoes several steps in different milling machines using different milling tools. The goal of the analysis was to develop a model for predicting tool wear and when tools should be changed. The dataset consisted of part information extracted from CAD, task information (type of tool, milling steps, duration, speed, etc.), machine tool measurements (length and diameter), and tool wear evaluations. When the first obstacles with data extraction were overcome, including consolidation of data from different local databases and the analysis started, it became clear that there were some major issues with traceability. The final evaluation of the tool wear was only done on a subset of the tools reducing the dataset dramatically. Also there were no ID numbers on the individual tools, making it impossible to link the results back to the tool usage. The key learning from the project was that there were missing tags to combine different data sources, key process variables were often not collected and the data structure was not suitable for the required analysis.

Traceability is of essence if the overall process is to be scrutinized. While many companies make great strides towards achieving it, the impression is that in many cases we are still a long way from tracking down a single product all the way back to the raw material. When starting a new project involving the entire production steps, we highly recommend establishing a flow diagram indicating each process step and availability of the data therein. Then the next step is to understand and establish (if none existing) the connections among these local data sources for assuring traceability. As self-evident as this recommendation may be, it has been our experience that this was overlooked in the mist of talks of Big Data, cloud computing, and new-fangled data analytics methods.

4. Data Content

4.1 Historical Data

It is not uncommon to hear in our early meetings with the companies that they have a vast amount of historical production data. However, in many cases this claim can be quite misleading. First of all, operating conditions in any unit operation let alone in the entire production can drastically change over time to make parts of the historical data incompatible with the rest. It should be noted that what we mean with drastic changes are changes beyond the unavoidable regime changes that processes usually go through, for example, a recent renovation of the plant, getting brand-new production equipment or permanently changing raw material with completely different characteristics as in the case of ecofriendly products. This may very well render previously obtained data using for instance the old production equipment incompatible with data obtained after the renovation.

Data collection schemes and measurement systems are also prime examples of such changes. In a particular study on railways, the geometric measurements of the tracks were regularly collected by a so-called measuring car. These measurements were then used to quantify the wear and tear that the tracks were experiencing in time and a degradation model was to be constructed for better scheduling of maintenance activities such as tamping of the tracks. These measurements would sometimes be taken by subcontractors that may change the measuring system in time. Furthermore, different subcontractors from different countries may have used different equipment with varying tolerances. This can once again render different parts of the historical data being incompatible for further analysis.

Even in the rare circumstances when the above-mentioned changes do not occur, there could still be issues rendering parts of the historical data irreconcilable. Historical data was often collected for documentation and control purposes, and not necessarily for analysis toward better understanding of the process or for its optimization. That is, old data may simply be used for example in case of a recall or to identify and isolate potential batches of products affected by a production problem. Since the process improvement would not necessarily be the primary goal in the collection and storage of historical data, it may correspondingly not contain valuable information to address serious production questions of today. This should by no means be a deterrent from the use of historical data but it should simply be taken as a word of caution when it comes to claims of "plenty of historical data" being available at the start of a project.

4.2 Multiple Production Sites

Over the last few years, one of the main activities of the IT departments has been building data historians. These are practically data warehouses where data from different unit operations and production sites is gathered, see Figure 3. The goal has been not only to provide easy access to data collected in multiple production sites but also to potentially generate opportunities for real time applications such as process surveillance and predictive maintenance. However, often those production sites are subjected to different operating conditions that may severely impair the possibility of combining data from different sites.

In one particular application, we worked on a fermentation process from three production sites. The idea was to combine the data to see the differences in key performance indicators, identify the causes for these differences and also perform process surveillance. However, these processes were tightly controlled with engineering control schemes, which differ from site to site; analogue and digital control schemes. Since the former is considered to be an old technology, we were then asked to focus only on the sites for which the digital controller is used. Furthermore, for one of the remaining two sites, it was deemed that the local crew was acting too differently from the main production site when it came to day-to-day operations and hence the large variation in key performance indicators could simply be attributed to this "cultural" difference. Considering the fact that the fermentation process was exposed to not only engineering control but also to manual control performed at the discretion of the skilled operators to avoid wasting any batch, it made more sense to initially focus on a single site with more relevant and somewhat more stable production characteristics.

In this case, one site was using a soon to be obsolete technology and it did not make sense to include the data coming from that process. For another, manual interference that was hard to predict with the available data limited our focus to production from one site only. In fact, manual control of the operations by experienced operators is not an uncommon occurrence even in modern production processes. The true digitalization should aim to minimize this and incorporate that knowledge gained based on years of experience through automated systems. Ultimately though, site-to-site differences are of course extremely valuable to extract in order to eventually minimize these differences and learn the impact of different operating conditions on the final product. In that regard, merging the datasets from different production sites will indeed be of importance but needs to be done with care.

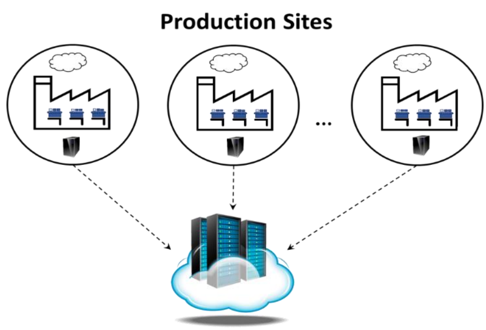


Figure 3. Establishing data historians (Computer clip arts are obtained from http://clipart-library.com/cloud-server-cliparts.html)

4.3 Multi-stage Processes

In many industrial processes consisting of multiple stages, the focus has historically been on unit operations sometimes taken care of by different, semi-autonomous groups within production. Each group would be responsible for its own unit operation and optimization would be performed locally with data collected and labeled accordingly. As a result, the data collection is seldom standardized, and connecting all these data becomes extremely challenging. This is also related to the traceability issue we discussed earlier. Another concern is that optimized individual unit operations do not necessarily translate into an optimized overall production. In the same vein, we have seen great interest from industry to pursue a general approach to data analysis and consider all stages of production at once. The aim in such a study is usually to discover the relationship of process variables from different stages of production with a final key performance indicator such as product quality. This could be achieved through observational or experimental data. The former often requires methods beyond standard statistical methods due to characteristics of data discussed below. However, it should be noted that the observational data could be used to obtain predictive models as in the case of machine learning methods (Hastie et al. (2009)). These can in turn be used for risk management, i.e., if you can predict accurately what is going to happen, you can take actions to hedge the risk in anticipation accordingly. However, if the goal is to inflict a desired change in key performance indicators through adjustments to the process variables, we need to uncover causal relationships. The time-tested approach to achieve this is through controlled experiments. The design and analysis of experiments for multi-stage processes can be elaborate and we currently need more methodological and applied work in this field (Tyssedal et al. (2011), Tyssedal and Kulahci (2015), Kulahci and Tyssedal (2016)).

Another concern related to considering all production stages at once is that the data collected in these different stages comes in great variety; numerical, categorical, text, etc. This so-called mixed data represents a significant hurdle for many analysis methods to overcome.

A somewhat similar circumstance happens when production occurs in multiple parallel lines with varying production equipment. We have had experience with processes involving injection molding where tens if not hundreds of molding machines in one production site is used on a continuous basis. The need to combine data from all these machines could for example, compensate for the scarcity of rare events that are to be predicted. We have been encountering this issue more and more frequently with the proliferation of Six Sigma processes for example, where low defect rates create unintended consequences in modeling as in classification (Khan et al. (2017a, 2017b), Khan et al. (2018)) as well as in statistical process control (Agresti and Coull (1998), Montgomery (2013), Wang (2009)). We will discuss this issue further in Section 5.2.

4.4 Data Transformation

From the time, data is measured by the sensors in production to the time that it reaches the data analyst, it would usually undergo several transformations. Firstly, the sensor itself will transform its input signals, depending on the sensor configuration, to produce the initial measurement, which may be taken by the sensor at unevenly spaced time intervals. Measurements from the sensor may be further manipulated upon storage in a data historian. Typically, the data is compressed in order to save storage space and increase retrieval speed (Thornhill et al. (2004)). A common compression method used by data historians is to only retain those measurements from a sensor that differ by some minimum amount compared to the previously stored measurement from that sensor. Finally, upon retrieval of data from the data historian, further manipulation can be made where the data historian itself can often by default fill in the previously discarded measurements according to the method selected by the user, and provide the data in a convenient format. This final manipulation is perhaps the most insidious, as interpolated data can be mistaken for actual observations measured during production. As a result, of these transformations, the data takes on an ephemeral quality and it can be a challenge for the data analyst to know what version of the data has been acquired.

We have found that the best approach in these situations is to insist on obtaining data from as close to the source as possible – which usually is in the form in which it is stored in the data historian. As tempting as it may be, it is misguided to rely on the retrieval modes of the data historian to provide a ready-to-use dataset, as important features of the data may be masked. For example, we have experienced cases where one variable was stored at the rate of less than 10 observations per hour, whilst another variable was stored at the rate of several thousand observations per hour. If we had not insisted on the data in its rawest format, but relied on the data historian's reconstructed version of the data where all variables were evenly spaced in time, we would have been unaware of this huge difference in sampling frequency, and unable to advise the company to reexamine the data storage parameters. Many analysis approaches to serially dependent data assume regularly spaced sampling and cannot handle irregular sampling frequencies, so some interpolation/imputation is required. By insisting on acquiring the raw data, the data analyst has at least a chance of understanding what is truly being measured, and has control over the interpolation method being used. Some multiresolution/multi-granularity methods have been proposed with the aim of dealing directly with data containing different granularities (Li et al. (2001), Reis (2019)) but the problem has otherwise been somewhat overlooked.

One of our projects involved injection molding of plastic components. These processes often operate at high production rates in a matter of a few seconds per "shot." They also show characteristics of batch processes for which each injection molding shot is a batch and process variables such as temperature and pressure in the mold follow expected profiles during the batch. However, due to physical constraints (e.g., difficulty of mounting sensors in the injection molding machine) and the high production rate, these profiles are seldom available. Instead, pre-determined features of the profiles as in maximum or steady state values may be available. Sometimes, the profiles are shown for each shot for the operator's instantaneous review but never to be stored afterwards. It was our first instinct to inquire about those profiles and ask to have access to that data accordingly. The company uses different equipment suppliers for the same product and only some allowed access to that data. Even then, a lot of IT work was needed to actually access those profiles. In the end, insisting in obtaining the entire profile data for the overall production or generating solutions based on the assumption of easy access to those profiles would ultimately require replacement and/or adjustment of production equipment. Hence, before venturing down that road, we needed to properly investigate the features that were readily accessible. Even if these features did not contain the right information, the costly adjustment of production equipment needed to be justified against the potential benefit of full access to production data. For this, we would like to once again refer back to the question we started with: how much and what type of data do we actually need? The systematic collection of data as we have advocated would suggest to tone down our inhibitions about the amount of data or lack thereof for that matter, gradually work with what is available and determine the next steps.

5. Data Analysis

5.1 Combining Process and Product Data

One of the key issues in process understanding, optimization and control is to connect process data with product characteristics. The process data is usually collected with the help of sensors and through automated data collection schemes. When companies claim to possess large amounts of data, they often refer to process data. Such data can be used in process monitoring, for example. Modeling of such data is done through so-called unsupervised learning methods where no response, e.g., product characteristics, is available.

Except for the case of 100% inspection, which has been an exception rather than the norm in quality assurance, product quality data is often more scarcely collected. This is due to issues such as cost of inspection, physical constraints and production rate. Fast production rates make it particularly difficult to obtain product quality for each product. When that happens, only a fraction of process data has the corresponding product quality characteristics. Depending on the magnitude of that fraction, this can result in cases where a majority of process data cannot be used for modeling the product characteristics, see Figure 4. In this case, semi-supervised methods where supervised and unsupervised data are combined in the same modeling approach can provide some relief (Frumosu and Kulahci (2018, 2019)). In these methods, the goal is to make use of unlabeled (unsupervised) data for which no output is available to improve, for example, the predictive model, which will otherwise be based solely on the labeled (supervised) data. In real life applications, the main concern is the similarity of the unlabeled and labeled process data to allow for merging these two for predictive modeling. Also, for fast production rates, the issue of serial dependence (autocorrelation) in process data becomes more evident and needs to be taken into consideration in many production analytics applications (Bisgaard and Kulahci (2011), De Ketelaere et al. (2015), Vanhatalo and Kulahci (2015a, 2015b), Vanhatalo et al. (2017)).

Eventually, we can aim for all labeled data, i.e., an established, direct connection between the product characteristics and corresponding process data. More sensorics applications are certainly needed to accomplish that. The advances in barcodes and QR codes for example will allow for tracing each product throughout production and even beyond. These codes can provide dynamic information, i.e., product specific information at the time of production or static information, e.g., ingredients in all pills with the same contents in tablet production in pharmaceutical industry. If the traceable product with encoded dynamic information can then be inspected automatically through for example, image analysis, the connection between the process data and the product characteristics can be established. We should once again emphasize that this will require a multi-disciplinary approach where engineers and IT personnel will work towards developing the hardware and software for tags based on the physical limitations of the process and products, sensorics experts will develop the right inspection scheme of an individual product and finally the data analyst will not only analyze the data but also support the efforts to encode the needed process information in the tags. In that sense, it is of utmost importance for the data analyst to be involved early in these efforts to avoid delays further down the line.

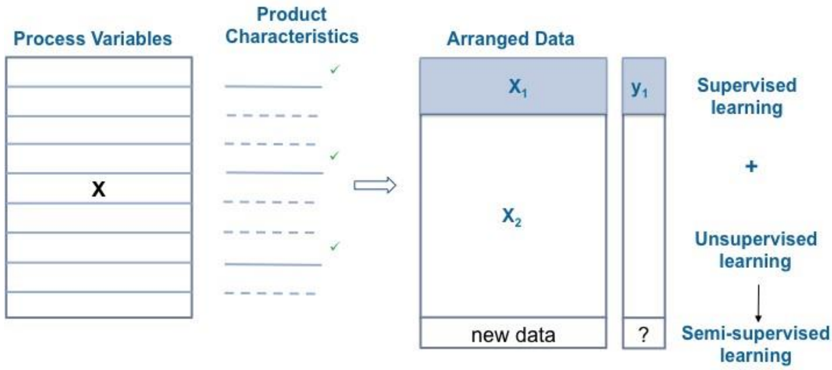


Figure 4. Semi-supervised learning where the original data (X) is split into supervised (X_1) and unsupervised (X_2) components, and a predictive model is established using both.

5.2 Lack of Specialty Data

Often the data available to the data analyst originates from routine production, where the quality characteristics of the end product must lie within narrow specification limits. Many manufacturing companies have gone through quality management schemes like Six Sigma and Lean Production and therefore, they are usually capable in producing products with desired specifications. Consequently, it is quite common to end up with a dataset containing very little variation in the measured production settings and quality characteristics. It has been on the other hand noted by data analysts that more interesting discoveries can be made when things go wrongly (Borne, 2014), so a lack of process and/or product failures can be an obstacle to new insights.

Similarly, a common starting goal in many projects is to develop a model that can predict end-product quality based on the measured production conditions at some point earlier in the process. A lack of variation in the observed product quality in the available data poses a major challenge in the identification of the factors that lead to good or bad quality. This challenge cannot necessarily be addressed by acquiring more data, if the new data is just “more of the same”. One approach to consider is to revise the goal and look for features in the data that do display promising variation. For example, in one project we ended up focusing on the production duration of each product as the variable to be predicted, after our prior attempts to predict product quality were unsuccessful due to lack of variation. Of course, it is vital to discuss the revision of goals with the industry partners to ensure the new goal is one that will provide value to the company.

Even in the case when enough variation is found, there can still be a lack of “bad processes” due to very tight control of the process particularly in industries such as in pharmaceutical and biomedical industries, where batch processes are being extensively used

(Croughan et al. (2015)). Each batch may involve a huge volume of raw material and take hours or even days to complete. It is, therefore, extremely costly to have a “bad” batch. Hence, these processes are very carefully controlled through both automatic control schemes and manual interventions by the operators. If in a process surveillance study the aim is not only to detect an out-of-control situation but also diagnose of the root cause, data capturing various fault scenarios is needed. In terms of modeling, this translates into a highly imbalanced division between “bad process” versus “good process” classes. Several methods have been proposed to alleviate the imbalance problem at both data and algorithm level (Haixiang et al. (2017)). At the data level, data sampling methods such as Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al. (2002)), under-sampling and over-sampling have been used with some success. For algorithms, ensemble methods, or algorithm modifications of traditional classifiers have been suggested (Haixiang et al. (2017)). Also, cost-sensitive learning methods have been proposed, which usually assume higher costs for the misclassification of the “bad process” samples (Haixiang et al. (2017)).

5.3 Large Data Processing

With increased accumulation of production data, one of the biggest challenges has become the allocation of enough computational resources to process it. Although new technologies, such as parallel computing and quantum computing, have revolutionized the whole field, memory capabilities are still limited. Most of the wellknown data analytics methods worked on the principal of in-memory processing. Computing frameworks such as Hadoop and Spark (Zaharia et al. (2010)) enable inmemory computation of large data streams and provide solutions to the problems prompted by the continuous streams of data (Agneeswaran (2014)). In terms of data storage, there is currently a transition towards NoSQL (“non SQL” or “nonrelational”) databases (Leavitt (2010)) as opposed to the traditionally structured relational databases. One of the key advantages of NoSQL databases is that they can handle large unstructured data efficiently.

This issue certainly ventures into the domain of IT folks whilst data analysts (or even data scientists) will most likely not be well-versed in this issue. But, the voracious appetite of digitalization is not likely to be satiated anytime soon and data management (or data wrangling for analytics purposes) will only get more complicated in time. “Not my problem” attitude to data management could no longer be a response for the data analyst/scientist. We should strive to be part of the solution for the reasons we discussed earlier. This requires at the very least being able to communicate with the IT folks and establish a common language early on in the process. Otherwise, we will risk falling out of the loop of data management process altogether.

5.4 New Opportunities

In some industries, such as in electronic component manufacturing and aerospace industry, inspection of each product is a legal requirement. But, now due to the affordable sensor technologies, it is becoming more common in other industries as well to inspect and report the quality characteristics of each product as we mentioned in Section 5.1. This information along with process data reinvigorates the possibilities to investigate new opportunities to optimize the production processes. In process monitoring, for example, the focus may shift to product characteristics as they contain process information anyway. This may simplify the monitoring schemes significantly. The process data and its connection to product characteristics can then be used for predictive purposes or when possible, to create a desired change in the process as in optimization and robustness studies. Furthermore, the reaction time to a fault can be significantly improved when we shift from sampling to 100% inspection. As we mentioned in the previous section, it has been our experience that despite all efforts in automation, manual interventions by the operators are alarmingly high. This is generally accepted as good practice relying on the years of experience of these operators. In modern manufacturing, this process knowledge needs to be digitalized and fed back to the process when needed. This will prevent unintended variation and lead to more standardized operations. Currently, not all these interventions are recorded making them impossible to digitalize. Action needs to be taken for proper collection of these manual interventions so that empirical models can be trained and put to use with this enhanced process understanding.

Another emerging field for using Big Data is in predictive maintenance (Alaswad and Xiang (2017), Peng et al. (2010), Wang (2002)). There have been primarily two approaches to maintenance employed by the companies: Time-based preventive maintenance and corrective maintenance. The former is very intuitive and often effective as planning and scheduling of the maintenance activities can be done on a regular basis. Yet the downside of it is that this may cause unnecessary or delayed maintenance if the scheduling does not reflect the actual wear and tear in the equipment/system. Similarly for the corrective maintenance, the damage would be done by the time maintenance is performed. In predictive maintenance (sometimes also referred to as condition based maintenance), maintenance is performed when deemed necessary by a predictive model reflecting the condition of the production equipment. While this is expected to reduce the maintenance cost, it also brings more challenges due to uncertainties in planning and scheduling of maintenance activities caused by prediction. We are currently involved in multiple projects both on delivering predictive models but also on planning and scheduling of maintenance activities. Production companies see tremendous value in predictive maintenance, particularly in cases where high throughput is the main goal and stoppages are of major concern. Similarly many engineering companies that produce the production equipment are interested in selling not only the equipment but also a service contract to provide the maintenance. The use of production data is the key in all of these efforts.

Advances in computer technology also had a very encouraging effect on obtaining simulation models of the production processes. These could be the key in one of the very important concerns in data analysis: correlation vs. causality. As we mentioned earlier, controlled experiments are the classical tool for unearthing causal relationships. We also indicated that there are many challenges when experimental work is to be performed on the process as a whole. But if a simulation model of the process exists, experiments can be run at a relatively lower cost without any disruption to the actual process. Once the crucial relationships are uncovered or the right settings of the process variables are established using the simulation model, confirmatory experiments can be run on the actual process. It should however be noted that simulation experiments differ in characteristics from their physical counterparts (Santner et al. (2003), Dehlendorff et al. (2011)). Nonetheless, the simulation models of the processes offer a great platform for process understanding and improvement.

6. Data Utilization Maturity – Digitalization Readiness

We have observed that many companies were not at the required level of maturity when it came to utilizing data within the manufacturing environment whereas the same companies were fairly good at utilizing data in other parts of the business, e.g., finance, marketing and R&D functions. There have been different illustrations and descriptions of data maturity, which have inspired the development of our interpretation of data maturity within a manufacturing environment as depicted in Figure 5.

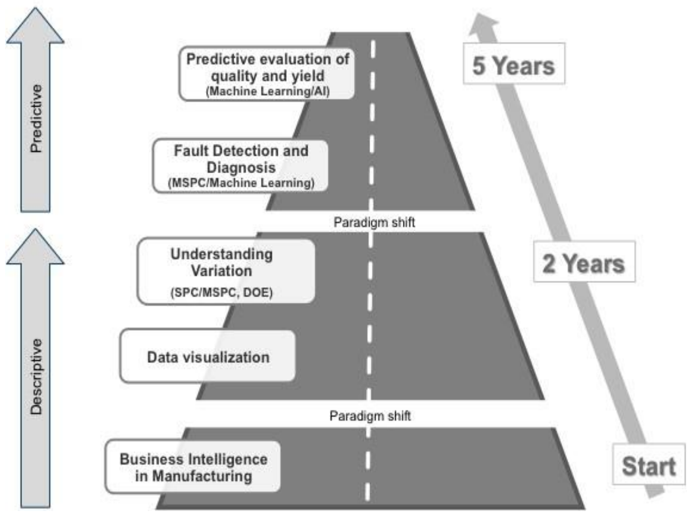


Figure 5. Evolution in Data Maturity and Data Utilization. The flags illustrate an estimated time of incubation and implementation.

Our assessment is that the jump many companies are trying to make from unexplored terrain, e.g., from using Business Intelligence techniques in manufacturing, to advanced analytics as in AI is considerably bigger than any company initially anticipates. In this attempt, there will be lessons to be learned related to data connectivity, data accessibility, data management, traceability, etc. In the worst case, these lessons will be so costly that they will hinder future data related activities. Our recommendation is, therefore, to follow a sequential approach as in the Data Maturity Road illustrated in Figure 5.

A natural first step is to connect to the manufacturing equipment and initiate data collection. Some of this data can be transformed into valuable information by just visualizing the data, e.g., in the form of simple time series plots for the relevant users (operators, line managers, specialists). This will also help to confirm the preexisting understanding of the process to a large extent. A natural next step is to establish the variation in the process through process monitoring. If this exceeds the expected level of variation, changes to the process are in order. This will be done through the relationships between process variables and key performance indicators such as product characteristics. Predictive models based on correlations can be initially obtained for further experimental work to discover/confirm causal relationships through which desired changes to the process will be made. It is important to acknowledge that going further on the data utilization journey will require establishing new competencies within the company. As one moves further on the data utilization road and starts relying on empirical models for insights, a new set of challenges will arise. Many companies immediately pursue AI for being able to optimize production. Often this seems more appealing than going for a sequential approach starting with understanding the process and hence understanding what the real needs are. However, we should always keep in mind that data and data analytics are simply means to an end and not the end itself. Therefore, a sequential approach to digitalization and use of data will most likely be more fruitful.

7. Conclusion

With the proliferation of automated data collection schemes, the amount of production data that is being generated is going to increase at an accelerated pace. However, currently available data often lacks the necessary information to understand, improve and optimize production. While some useful information can be obtained from historical data, its analysis usually reveals that new data should be collected. In that sense, Big Data activities should be forward looking and performed in a sequential manner. Moreover, the aim should not necessarily be to collect as much data as possible from production but rather to collect the data systematically through appropriate sampling strategies. These efforts can then be targeted towards solving specific production problems. We should reiterate that this could be a treacherous journey with plenty of deterrents along the way. Industry should have the patience and commitment to go through the processes while acquiring and developing the right skills and expertise. Data analyst/scientists should be open to new challenges and learning opportunities, and expand on their communication skills as today’s problems are truly multi-disciplinary and isolation will easily translate into irrelevance.

Acknowledgements

The authors would like to thank the industrial partners and MADE SPIR, MADE DIGITAL, BioPro2 and Innovation Fund Denmark for making this work possible.

References

Agneeswaran, V. S. (2014). "Big data analytics beyond Hadoop: real-time applications with storm, spark, and more Hadoop alternatives." FT Press.

Agresti, A. and B. A. Coull (1998). "Approximate is better than "exact" for interval estimation of binomial proportions," *The American Statistician*, 52, 119–126

Alaswad, S. and Y. Xiang (2017). "A review on condition-based maintenance optimization models for stochastically deteriorating system," *Reliability Engineering and System Safety*, 157, 54-63

Bisgaard, S. and M. Kulahci (2011). *Time Series Analysis and Forecasting by Example*, New York: Wiley

Borne, K. (2014). "Outlier Detection Gets a Makeover - Surprise Discovery in Scientific Big Data," *Statistics Views*

Capaci, F., E. Vanhatalo, M. Kulahci, and B. Bergquist (2019). "The Revised Tennessee Eastman Process Simulator as Testbed for SPM and DoE Methods," *Quality Engineering*, 31, 212-229

Chawla N.V., K. W. Bowyer, L.O. Hall and W.P. Kegelmeyer (2002). "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, 16, 321-357

Croughan, M. S., K. B. Konstantinov and C. Cooney (2015). "The future of industrial bioprocessing: Batch or continuous?" *Biotechnology and bioengineering*, 112, 648651.

De Ketelaere, B., M. Hubert and E. Schmitt (2015). "Overview of PCA-Based Statistical Process-Monitoring Methods for Time Dependent, High Dimensional Data", *Journal of Quality Technology*, 47, 318-335

Dehlendorff, C., M. Kulahci and K. K. Andersen (2011). "Designing Simulation Experiments with Controllable and Uncontrollable Factors for Applications in Health Care," *Journal of Royal Statistical Society, Series C*, 60, 31-49

Frumosu, F. D. and M. Kulahci (2018). "Big Data Analytics using Semi-supervised Learning Methods," *Quality and Reliability Engineering International*, 34, 1413-1423

Frumosu, F. D. and M. Kulahci (2019). "Outliers Detection Using an Iterative Strategy for Semi-Supervised Learning," *Quality and Reliability Engineering International*, 35, 1408-1423

Gajjar, S., M. Kulahci and A. Palazoglu (2016). "Use of Sparse Principal Component Analysis (SPCA) for Fault Detection," *Proceedings of the 11th IFAC Symposium on Dynamics and Control Process Systems*

Gajjar, S., M. Kulahci and A. Palazoglu (2017a). "Real Time Fault Detection and Diagnosis Using Sparse Principal Component Analysis," *Journal of Process Control*, 67, 112-128

Gajjar, S., M. Kulahci and A. Palazoglu (2017b). “Selection of Non-zero Loadings in Sparse Principal Component Analysis,” *Chemometrics and Intelligent Laboratory Systems,* 162, 160-171

Gao, H, S. Gajjar, M. Kulahci, Q. Zhu and A. Palazoglu (2016). “Process Knowledge Discovery Using Sparse Principal Component Analysis,” *Industrial and Engineering Chemistry Research,* 55, 12046-12059

Haixiang, G., L. Yijing, J. Shang, G. Mingyun, H. Yuanyue and G. Bing (2017). “Learning from class-imbalanced data: Review of methods and applications,” *Expert ystems With Applications*, 73, 220–239

Hastie, T., R. Tibshirani and J. Friedman (2009). *The Elements of Statistical Learning,* 2nd Ed., New York: Springer

Khan, A. R., H. Schiøler, M. Kulahci and T. S. Knudsen (2017). “Selection of Objective Function for Imbalanced Classification: An Industrial Case Study,” *Proceedings to 22nd Emerging Technologies and Factory Automation*, Limassol, Cyprus

Khan, A. R., H. Schiøler, T. Knudsen and M. Kulahci (2017). “Big Data Analytics for Industrial Process Control,” *Proceedings to 22nd Emerging Technologies and Factory Automation*, Limassol, Cyprus

Khan, A.R., H. Schiøler, M. Zaki and M. Kulahci (2018). “Rare-events classification: An approach based on Genetic Algorithm and Voronoi Tessellation,” *Proceedings of the 22nd Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Melbourne, Australia

Kulahci, M. and J. Tyssedal (2016). “Split Plot Designs for Multistage Experimentation,” *Journal of Applied Statistics*, 44(3), pp. 493-510

Lasi, H., H-G. Kemper, P. Fetteke and T. Feld (2014) “Industry 4.0,” *Business and Information Systems Engineering*, 6, 239-242

Lawrence, N. D. (2017) Data Readiness Levels, https://arxiv.org/abs/1705.02245 Leavitt, N. (2010). “Will NoSQL Databases Live Up to Their Promise?” *IEEE Computer*, 43, 12-14

Li D., S.L. Shah and T. Chen (2001). “Identification of fast-rate models from multirate data,” *International Journal of Control*, 74, 680-689

Liu, W., Chua, T. J., Larn, J., Wang, F. Y., & Yin, X. F. (2002). “APS, ERP and MES systems integration for semiconductor backend assembly,” *7th International Conference on Control, Automation, Robotics and Vision,* 3, 1403-1408

Meer, K.H. (2005). *Best Practices in ERP Software Applications.* iUniverse

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

Montgomery, D. C. (2013). *Statistical Quality Control - A Modern Introduction,* 7th edition. Wiley: New York

Peng, Y., M. Dong and M. J. Zuo, (2010). "Current status of machine prognostics in condition-based maintenance: a review," International Journal of Advanced Manufacturing Technology, 50, 297-313.

Reis, M.S. (2019). "Multiscale and Multi-Granularity Process Analytics: A Review," *Processes*, 7, 61-82

Santner, T. J., B. J. Williams and W. I. Notz (2003). *The Design and Analysis of Computer Experiments*. Springer: New York

Thornhill, N. F., M. S. Choudhury and S. L. Shah (2004). "The impact of compression on data-driven process analyses," *Journal of Process Control*, 14, 389398.

Tyssedal, J. and M. Kulahci (2015). "Experiments for Multi-Stage Processes," *Quality Technology and Quantitative Management*, 12, 13-28

Tyssedal, J., M. Kulahci and S. Bisgaard (2011). "Split-Plot Designs with Mirror Image Pairs as Subplots," *Journal of Statistical Planning and Inference*, 141, 36863696

Vanhatalo, E. and M. Kulahci (2015a). "The Effect of Autocorrelation on the Hotelling T^2 Control Chart," *Quality and Reliability Engineering International*, 31, 1779-1796

Vanhatalo, E. and M. Kulahci (2015b). "Impact of Autocorrelation on Principal Components and their Use in Statistical Process Control," *Quality and Reliability Engineering International*, 32, 1483-1500

Vanhatalo, E., M. Kulahci and B. Bergquist (2017). "On the Structure of Dynamic Principal Component Analysis Used in Statistical Process Monitoring," *Chemometrics and Intelligent Laboratory Systems*, 167, 1-11

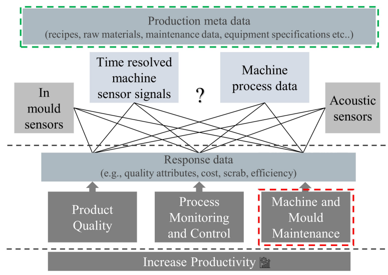
Wang, H. (2002). "A survey of maintenance policies of deteriorating systems," *European Journal of Operational Research*, 139, 469-489

Wang, H. (2009). "Comparison of p-control charts for low defective rate," *Computational Statistics and Data Analysis*, 53, 4210–4220

Zaharia, M., M. Chowdhury, M.J. Franklin, S. Shenker and I. Stoica (2010), "Spark: cluster computing with working sets," *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing*, Boston, USA

APPENDIX B

Paper E, Mould wear-out prediction in the plastic injection moulding industry: a case study



It has been normal practice in many industries to collect data for documentation and tracking of various processes. This covers documentation of goods received (e.g. raw materials and spare parts), production order information, bill of materials, product quality and maintenance records. This documentation is, in many situations, done digitally, making the information available for analytic purposes. Utilising these data is often more straightforward than initiating data collection from equipment. Therefore, it is highly relevant to consider these data sources when identifying potential applications in the initial phase of a data utilisation journey. Since the data have been collected for documentation purposes, there is a high risk that the information contained in the collected data wouldn't support the specific application. It is therefore essential to conduct an early data exploration. The work in paper E is an example of such data exploration and application development. The work is related to RQ1 by mitigating productivity loss caused by mould malfunction.

Research question 1 (RQ1)

How to best utilise data from injection moulding machines to increase the productivity of producing injection moulded elements?

The work is also related to RQ3 by investigating the use of available maintenance records for condition-based monitoring of injection moulds and by identifying the need for additional sensor data to obtain real-time condition-based maintenance.

Research question 3 (RQ3)

How to best do real-time condition-based maintenance on injection moulds?

B.1 Summary

The business objective addressed in paper E is linked to predictive maintenance and prediction of the current state of an injection mould. The potential value creation is achieved by predicting when a mould is to be declared worn-out and taken out of production. By indicating the worn-out time, a new mould can be ordered in due time, mitigating the risk of an idle period without a functioning production mould. In paper E, we utilise existing mould maintenance service data to perform a mould wear-out prediction to enable re-order of new moulds in due time. The work was motivated by detecting when an injection mould needed to be exchanged with a new mould. Failing to detect this might cause production stop (not having a functional mould) or production of defective products. The maintenance service data consist of both numeric and categorical data. Unsupervised random forest is used to overcome the challenges with the mixed data types. Using the Gini index, for variable importance, it is found that one of the mould characteristics made a meaningful clustering of the data. This clustering is used to split the data before further modelling. Using Kaplan-Meier survival curves, it is possible to calculate a probability that a given mould type will be declared worn-out at a given number of shots (moulding cycles). This is further developed into an early warning system using the likelihood that the mould is still running at a given number of shots and Hotelling T^2 statistics. This is combined into a visual dashboard that can be used for mould evaluation when performing mould maintenance. It can be evaluated if the mould is performing as moulds of the same type and if the mould is closing mould worn-out.

Achieved results

- Increased understanding of the information contained in the used maintenance data and how best to utilise this information.
- Interactive dashboard for exploration and monitoring of mould state.

Contribution

- Demonstration of survival curves used for mould wear-out prediction.
- Demonstrate effective use of mixed categorical and numeric data using an unsupervised random forest approach.

B.2 Paper E

Paper E

Mould wear-out prediction in the plastic injection moulding industry: A case study

Flavia Dalia Frumosu[a], Georg Ørnskov Rønsch[a], and Murat Kulahci[a,b]

[a]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark; [b]Department of Business Administration, Technology and Social Sciences, Luleå University of Technology, Luleå, Sweden

ABSTRACT
The current work addresses an industrial problem related to injection moulding manufacturing with focus on mould wear-out prediction. Real data sets are provided by an industrial partner that uses a multitude of moulds with different shapes and sizes in its production. An analysis of the data is presented and begins with clustering the moulds based on their characteristics and pre-chosen running settings. Using the results of the clustering, the mould wear-out is modelled using Kaplan-Meier survival curves. Furthermore, a random survival forest model is fitted for comparison and model performance is assessed. The main novelty of the case study is the implementation of mould wear-out prediction in real-time with the outcomes presented in terms of conditional survival curves including a proposed early warning system. For visualization and further industrial implementation, a R Shiny dashboard is developed and presented.

KEYWORDS
Industry 4.0; predictive maintenance; injection moulding; mixed data; reliability analysis; censored data; mould wear-out

Introduction

Like many industries, the plastic injection moulding industry is currently going through changes under the so-called Industry 4.0 (Lasi et al. 2014), which aims for higher degree of automation and digitalization. Companies are already embarked on the route to Industry 4.0 and even national policies were elaborated for a number of countries (Da Silva et al. 2020). With increasing availability of sensors and decreasing cost of data storage and computer power, this transformation often results in abundance of process data for both offline and online data analytics. For the latter, one specific area of application is in Condition-Based-Maintenance (CBM) (Peng, Dong, and Zuo 2010). CBM is a maintenance program in which maintenance decisions are based on information from condition monitoring and it consists of data acquisition, data processing and maintenance decision-making (Jardine, Lin, and Banjevic 2006). This is a significant departure from corrective maintenance (CM) in which the maintenance is performed after the failure happens, and from preventive maintenance where the maintenance is performed on regularly scheduled intervals irrespective of the condition of the production equipment (Ahmad and Kamaruddin 2012). Significant attention has been paid to CBM in the literature and more recently, the focus has been shifted towards predictive maintenance (PdM) (Li, Wang, and Wang 2017). According to Li, Wang, and He (2016), the goal of PdM is to reduce downtime and cost of maintenance through the monitoring of the equipment's working conditions as well as predicting equipment failure that allows the maintenance to be planned before the actual fault occurs. There is a strong connection between CBM and PdM, where CBM can be considered as PdM (Sai, Shcherbakov, and Tran 2019; Hashemian 2010) or CBM can be treated as an effective form of PdM (Amari, McLaughlin, and Pham 2006).

Recent work related to PdM within Industry 4.0 includes Li, Wang, and Wang (2017) who propose a system

CONTACT Flavia Dalia Frumosu, fdal@dtu.dk, Technical University of Denmark, Department of Applied Mathematics and Computer Science, Richard Petersens Plads, Building 324, DK - 2800 Kgs. Lyngby Denmark.

framework built on Industry 4.0 concepts that includes a fault analysis process and treatment used for predictive maintenance in machine centers. Cachada et al. (2018) propose an intelligent and predictive maintenance system along with its architecture that is aligned with Industry 4.0 principles. Haarman, Mulders, and Vassiliadis (2017) have introduced the concept of Predictive Maintenance 4.0 (PdM 4.0) which is about the prediction of future failures in assets and the selection of the most effective preventive measure through the means of advanced analytic techniques applied on big data.

The case study addressed in this article is from the plastic injection moulding industry with the ultimate aim of mould wear-out prediction. In this regard, it is related to a study of Remaining Useful Life (RUL) of an asset or system. RUL can be defined as the time between the current time until the end of the useful life, Si et al. (2011), where the emphasis is set on statistical methods for RUL estimation based on both directly and indirectly observed state processes. RUL estimation is one of the key issues in CBM as shown in Jardine, Lin, and Banjevic (2006), Peng, Dong, and Zuo (2010) as well as in Ahmad and Kamaruddin (2012). RUL estimation is an important pillar for predictive maintenance platforms (Aivaliotis, Georgoulias, and Chryssolouris 2019).

With the increasing availability of production data that also includes maintenance data, there is currently more focus on data driven approaches to maintenance planning and scheduling. Bukkapatnam et al. (2019) proposes Manufacturing System-wide Balanced Random Survival Forest (MBRSF), a nonparametric machine learning approach for long-term prognosis for breakdowns of production equipment. Alsina et al. (2018) compares machine learning methods for predicting component reliability. Ragab et al. (2016) presents a prognostic methodology based on Kaplan – Meier estimation for RUL.

The case study is based on several data sets provided by an industrial partner. The data that is subject to changes in time is denoted as time dependent and data that remains constant in time is denoted as time independent. Mould characteristics and pre-chosen running settings are examples of time independent data whereas production and maintenance data are time dependent. In this work, the mould pre-chosen running settings are defined as machine settings selected by the operator prior to the use of the mould in production. Furthermore, it is assumed that these settings are not changed during production.

Mould wear-out prediction is the main goal as it has a direct impact on product quality. Moreover, understanding the deterioration mechanisms of moulds is expected to facilitate effective maintenance planning and hence, reduce the cost of maintenance. The available data contains information on a limited number of worn-out moulds. The proposed solution and choice of model reflect this specific characteristic of the data. However, the authors also provide a data intensive modeling approach that can be implemented as data for more worn-out moulds becomes available.

The main research contributions of the article are in terms of prediction and monitoring of mould wear-out. A method for prediction of mould wear-out is developed based on conclusions drawn from the analysis of real data. Moreover, a monitoring strategy in the shape of an early warning system, which can be used by the industrial partner by means of a dashboard, is presented.

The first part of the study focuses on dimensionality reduction of the time independent data so that it is easier for practitioners to interpret and monitor the results. The second part of the study focuses on the prediction of mould wear and tear. Product quality data and more sensor data will only be available in the future. However, a prediction model using this additional data is also discussed. Throughout the study a special emphasis is put on visualization and interpretation. As the final product is meant for industrial use, the results are presented in the shape of a R Shiny (2019) dashboard. Finally, future directions for research and application are discussed. It should be noted that due to confidentiality reasons, the data and the results have been masked when considered necessary.

1. Industrial context

In this section, the technical background for the injection moulding process is provided as well as the description of various data sets used in this study.

Injection moulding process

The basic mechanism of an injection moulding machine is presented in Figure 1.

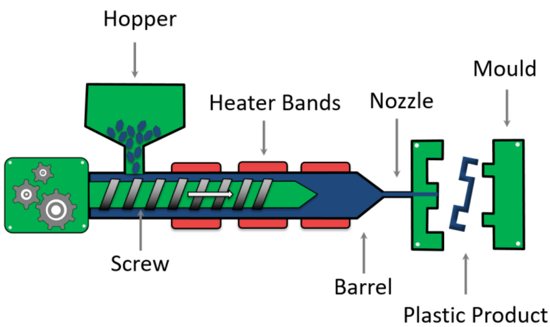


Figure 1. Injection moulding machine with major components.

The injection moulding production is discussed in terms of "shots" during which the moulding of a product (or a group of products depending on the mould characteristics) is performed. First, plastic pellets are fed into the hopper and the barrel. The plastic pellets in the barrel start to melt gradually due to the rotation of the screw that generates shear heat and due to external heat that is provided by the heater bands wrapped around the barrel. Next, as the screw rotates, the molten plastic is pushed into the nozzle which sends further the molten plastic to the mould. Usually, moulds have more than one cavity which means that more products are manufactured per shot. The plastic products from the mould are cooled while the mould remains closed via cooling channels placed inside the mould. After the plastic products solidify, the mould is opened and the plastic products are ejected from the mould, Figure 1.

The injection moulding process is a complex process during which various pressures, temperatures and speed-readings are collected and used for controlling the process through engineering control. Furthermore, the optimal settings of these variables are set by the operator based on prior experience. Hence, the process is controlled through both automated and manual interventions.

Data description

The industrial partner provided different types of real data sets related to the moulds used in production. The moulds are used for the manufacturing of plastic parts of different shapes and sizes and are usually with more than one cavity. The worn-out condition of the mould is defined internally by the industrial partner considering, for example, visual inspections. However, for confidentiality reasons, the exact definition is not presented in the article. The data sets are connected through the mould ID which is unique for each mould and are described in Table 1. All the data sets with the current status "Available" are analyzed or utilized in the article.

The mould characteristics include, for example, number of cavities and mould dimensions while the mould pre-chosen running settings are machine settings selected by the operator before the mould is used in production. Mould characteristics and pre-chosen running settings are not time dependent as these are assumed to be constant in time.

The maintenance data contains information regarding, for example, the number of cleanings, change of spare parts or other relevant maintenance information. Production data is related, for example, to the number of produced parts and stops in production. The maintenance and production data sets are time dependent as changes can occur over time.

The table also contains information about data sets that will become available in the future and can potentially have impact in determining the condition of the mould. Data sets containing both categorical data and numerical data are labeled as mixed data. The variables of the data sets in Table 1 are denoted as features, which is the machine learning terminology for variables. Furthermore, missing observations are present in some data sets as in the mould characteristics and running settings.

Table 1. Description of data sets.

Data set	Time dependence	Number Features	Data Type	Current status
Mould status (worn-out or running)	Dependent	2	numeric	Available
Mould characteristics	Independent	Around 20	mixed	Available
Mould pre-chosen running settings	Independent	Around 180	mixed	Available
Maintenance	Dependent	Around 10	mixed	Available
Production	Dependent	Around 15	numeric	Available
Environmental conditions	Dependent	N/A	N/A	Missing
Injection moulding machine sensors	Dependent	N/A	N/A	Missing
Metrology for products	Dependent	N/A	N/A	Missing

When determining the time for reaching a worn-out status, the number of shots is used as the unit of measurement. The number of shots provides a more realistic picture about the life of a mould compared to calendar time, as there are periods during which the mould is idle due to, for example, maintenance work. Following the terminology from survival analysis (Wang, Li, and Reddy 2019), the moulds that are currently still in use generate so-called right-censored data whereas the worn-out moulds generate uncensored. A depiction of the data is presented in Figure 2.

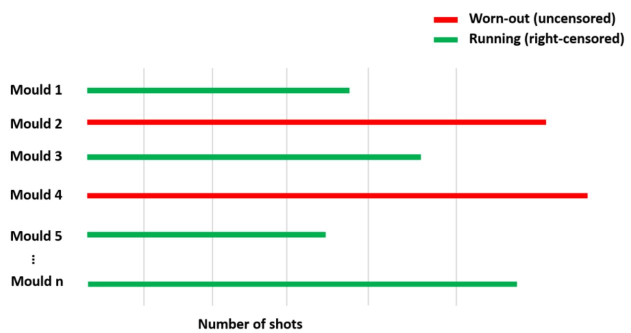


Figure 2. A depiction of worn-out versus running moulds.

Industrial problem

The primary goal of the case study is to assess the condition of the mould and predict when it will be worn-out. Another goal is to provide visual tools for the real time surveillance of the mould condition in production for the operators. It is important to note that the mould data includes a variety of different suppliers and production sites. Therefore, it is also crucial for the industrial partner to have an overview of the moulds performance at a global level, where consistency is key. Hence, the first step in the analysis is the empirical clustering of the moulds expecting the clusters to be formed by different characteristics and pre-chosen production settings of the moulds. This is followed by the modeling of the moulds' condition for each cluster. The methodology followed in the article is presented in the framework from Figure 3.

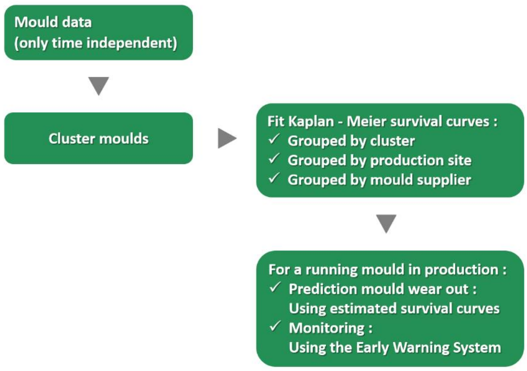


Figure 3. Framework describing the methodology followed in the article.

2. Theoretical background

Cluster analysis

Clustering is an unsupervised learning method which is used to group observations into clusters which are "similar" within the groups and "dissimilar" between groups (Ahmad and Khan 2019; Foss, Markatou, and Ray 2019). In our case, the data sets regarding the mould characteristics contain mixed data, which makes clustering challenging since applying direct mathematical operations such as summation or averaging used in obtaining similarity measures is not possible (Ahmad and Khan 2019). Mixed data is becoming more ubiquitous as modern production data often comes from increasingly heterogeneous sources (Foss, Markatou, and Ray 2019).

In general, clustering is performed using distance measures which quantify "(dis)similarity" between observations in the feature space (Foss, Markatou, and Ray 2019). For mixed data, a common distance measure is the so-called Gower distance (Gower 1971). After the distances between observations are computed through Gower or any other distance measure, clustering algorithms like k-means or hierarchical clustering (Hastie, Tibshirani, and Friedman 2013) can be applied.

In this study, an unsupervised random forest (Shi and Horvath 2006) is used for obtaining the distances between observations, which are called "proximities" (Breiman and Cutler 2019). The main advantages of the unsupervised random forest are its ability to deal with mixed data as well as allowing computation of variable importance, which can be of great value for interpreting the model outcome. In unsupervised random forest, for a given data set of N observations, the joint distribution of the variables is obtained and a synthetic data set is generated using this joint distribution. Often, the number of observations is kept the same for both the original and synthetic data sets. Each data set is labeled separately as for example 1 and 2, and this label is used as the response for the random forest model. Based on this model, an N x N similarity matrix is obtained through the frequency of any two observations, *i* and *j*, that are placed in the same terminal node. The variable importance can be computed using the Gini index decrease, i.e., the decrease in node impurity (Hastie, Tibshirani, and Friedman 2013). Each time a node split is made on a feature *m*, the Gini index for two descendant nodes becomes less than that of the parent node (Breiman and Cutler 2019). By adding up the Gini index decreases for each individual feature over all trees, the variable importance can be computed for each feature.

For assessing the number of clusters, Partition Around Medoids (PAM) (Kaufman and Rousseeuw 1987) along with the average silhouette width (Rousseeuw 1987) is applied. The PAM algorithm is very similar to K-means since both are partitional algorithms. The main difference is that K-means works with centroids while PAM works with medoids, which have the property that the average dissimilarity to all of the observations in the cluster is minimized. Every cluster can be represented by a silhouette (Rousseeuw 1987) that is built on the comparison of the clusters' tightness and separation. The average silhouette width can then be used to select a suitable number of clusters, with a high average silhouette width indicating a good clustering. There are various methods to assess the right number of clusters, however, for the purpose of the case study this approach is considered sufficient.

For clustering, only the results of PAM could be used, however, a visual interpretation is of interest and for this it was decided to use t-distributed Stochastic Neighbor Embedding (t-SNE) (Van Der Maaten and Hinton 2008). In t-SNE, the emphasis is put on modelling the dissimilar observations by large pairwise distances, while the similar observations are modelled by small pairwise distances. This approach was used as it provides a projection of the high-dimensional data into a two or three-dimensional map and thus, can be visualized in a plot. For the case study, a two-dimensional map was used.

Survival analysis

The results of the cluster analysis can be further used to check the worn-out and running moulds survival functions. As it was described in the Data description section and in Figure 2, the worn-out moulds can be regarded as uncensored data and the running moulds as right-censored data. Given this description, the life of the moulds can be presented in terms of a survival function (or curve) also called the reliability function. For the current article, the term survival curve is preferred.

If $T(\geq 0)$ is the number of shots until the mould reaches the worn-out state, then the survival function can be expressed as:

$$S(t) = P(T > t) \tag{1}$$

where, $S(t)$ is the probability that the mould is not worn-out by t number of shots.

Using the historical data, there are different ways of estimating the survival function in Equation (1). The Kaplan-Meier estimator (Kaplan and Meier 1958) is used:

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \tag{2}$$

where t_i is a number of shots when at least one event happened, d_i is the total number of worn-out events at t number of shots and n_i is the total number of moulds at risk at t number of shots.

Using the Kaplan-Meier estimator, the survival curve for each of the clusters (categories) from the cluster analysis can be estimated. For testing if there is a statistical significant difference between two or more survival curves, the G^{ρ} family of tests with $\rho = 0$ is used which is the log-rank or Mantel-Haenszel test (Harrington and Fleming 1982).

For a running mould belonging to a certain cluster, the survival curve can be displayed as a conditional survival curve (Zabor et al. 2013; Hieke et al. 2015). This is more intuitive than displaying the entire survival curve from zero number of shots. The confidence intervals can also be displayed. For this, the same approach as in Zabor et al. (2013) has been used for the case study.

As discussed earlier, the available data sets also contain production and maintenance information associated with the worn-out and running moulds. For building a model using this information as well, a random survival forest (Ishwaran et al. 2008) can be used for the case study. A random survival forest model is an ensemble learner based on the averaging of a tree base learner (Ishwaran et al. 2011). For the survival setting, the base-learner is formed from a binary survival tree and the ensemble is formed from a cumulative hazard function by averaging the Nelson-Aalen's cumulative hazard function for each tree (Ishwaran et al. 2011).

There are several advantages associated with a random survival forest model. As specified by Boström et al. (2018), a random survival forest model is robust for high-dimensional data and can accommodate high-level interactions between features. Other benefits inherited from random forests include the performance estimation using the Out-of-Bag (OOB) predictions, which can easily be implemented in parallel as well as variable importance. Furthermore, for survival analysis problems, random survival forest models do not require model assumptions such as proportional hazards, Ishwaran et al. (2008). In terms of the performance measure, the authors decided to use the c-index (also called as concordance or Harrell's index) (Harell et al. 1982). The c-index ranges from 0 to 1 and computes the proportion of concordant pair of observations among all pairs of observations in terms of survival time (Harrell, Lee, and Mark 1996).

Early warning system

Once the moulds are clustered, the next step in the case study is to provide an early warning system regarding the running mould status from cluster *c* at a given shot *t* ($S_t^{(c)}$) compared to historical data of the moulds from the same cluster. For this, the authors propose monitoring a percentage that is composed from the ratio of the worn-out moulds, which were still running at $S_t^{(c)}$ divided by the total number of worn-out moulds from that specific cluster. In this way, the operator can have a quick overview on how the running mould is behaving in relationship with previous moulds that were worn-out from the same cluster. If $S_i^{(c)}$ is defined as the number of shots at which the moulds in a certain cluster c were worn-out and taken off production, for $i = 1, \dots, n_c$, then that percentage can be expressed as:

$$p_{active} = \frac{\sum_{i=1}^{n_c} \delta_i}{n_c} \tag{3}$$

where, $\delta_i = \begin{cases} 1, & S_i^{(c)} \geq S_t^{(c)} \\ 0, & S_i^{(c)} < S_t^{(c)} \end{cases}$

Next steps involve the development of an early warning system using the maintenance and production data sets. The idea is that at $S_t^{(c)}$ the running mould should behave similarly in terms of maintenance and production with the worn-out moulds at $S_t^{(c)}$ that belong to the same category as the running mould.

For the online tracking, the authors propose a method for monitoring based on a Hotelling T^2 statistic (Montgomery 2012). For this purpose, at $S_t^{(c)}$, the average measurements is calculated for features of production and maintenance data of the worn-out moulds from first shot to $S_t^{(c)}$. This forms the basis to compare the running moulds at $S_t^{(c)}$. The data is denoted as W, which has the number of observations $m = \sum_{i=1}^{n_c} \delta_i$ and p features. The average is used since different features are collected at varying sampling frequencies. The same calculations are performed for a mould in use and the vector of size 1 x p is denoted as r. The Hotelling T^2 statistic is calculated as follows:

$$T^2 = (x - \bar{x})' K^{-1} (x - \bar{x}) \tag{4}$$

where, $\bar{x}$ is the sample mean vector and K is the covariance matrix obtained from W. This statistic gives the "scaled distance" of the current mould to the cluster in which it belongs in terms of its production and maintenance data. A threshold is then set, T_{UL} beyond which the current mould may no longer be considered belonging to that particular cluster depicted in W. The threshold is calculated using:

$$T_{UL} = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha,p,m-p} \tag{5}$$

Where, m is the number of samples, p is the number of variables/features and $F_{\alpha,p,m-p}$ is the F-distribution with p and m-p degrees of freedom at significance level α. This limit is obtained following the commonly used upper control limit of the Hotelling's T^2 control chart. Furthermore, the contribution of each feature to the T^2 statistic can also be calculated using:

$$d_j = T^2 - T^2_{(j)} \tag{6}$$

where, T^2 is the current value of the statistic while $T^2_{(j)}$ is the value of the statistic using all the features except for the j^{th} feature.

One potential problem is that there may not be enough data at $S_t^{(c)}$ to estimate the mean vector and covariance matrix in Equation (4). However, more data from each category is expected to be collected in the future and also if there are not enough data points for computing the Hotelling T^2 statistic then this means that not many worn-out moulds passed the $S_t^{(c)}$ number of shots.

3. Data analysis

Data processing

The first part of the analysis focuses on the cluster analysis and for this, the mould characteristics and pre-chosen running settings data sets are used. The data sets contained an abundance of missing values (NAs). Based on our discussions with the industrial partner, it was concluded that the missing observations were due to the specific characteristics of a mould and not due to measurement errors. For this, it was decided to encode the columns with NAs using a binary categorical variable with the categories, "missing" and "recorded".

The next step of the analysis involves the production and maintenance data sets. Here, it was necessary to combine two data sets originating from different sources and recorded at different frequencies. In order to overcome this challenge, the average over the time intervals is computed which will ensure the same sampling frequency for all features.

Clustering results

When the data sets on mould characteristics and pre-chosen running settings are merged, there are roughly 900 observations and more than 200 features. Using the unsupervised random forest model, the variable importance can be computed in terms of mean decrease Gini index. The first thirty features are presented in Figure 4 with "Mould Characteristic or Setting 182 #" being the most important feature.

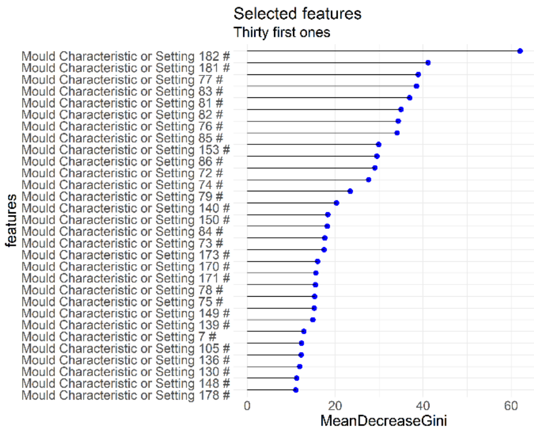


Figure 4. First thirty features given by the mean decrease in the Gini index.

The dissimilarity matrix for the moulds is obtained from the unsupervised random forest model. In Figure 5, using the dissimilarity matrix, the average silhouette width based on PAM is computed along with the t-SNE clustering using four clusters.

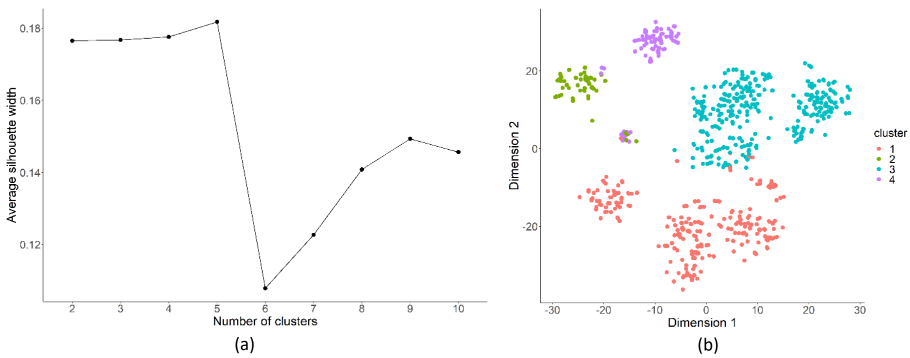


Figure 5. (a) Average silhouette width and (b) t-SNE clustering using four clusters.

Based on the average silhouette width in Figure 5a, it seems that the observations can be grouped in 5 clusters. Since the number of clusters of 2, 3 and 4 have average silhouette values close to the case with 5 clusters, t-SNE plots are also consulted and it is found that 4 clusters provide the best clustering of the data.

To simplify the interpretation of the clustering algorithm, solely the “Mould Characteristic or Setting 182 #” feature is considered as it has the highest importance value in the random forest model. In fact, this feature is further confirmed by the industrial partner as an important feature in production. As shown in Figure 6a, this feature alone separates the data very well with the exception of certain categories. Therefore, due to the practical interpretation and simplicity, for the next step of the analysis, the clustering given by the “Mould Characteristic or Setting 182 #” feature with the five categories is used as in Figure 6b.

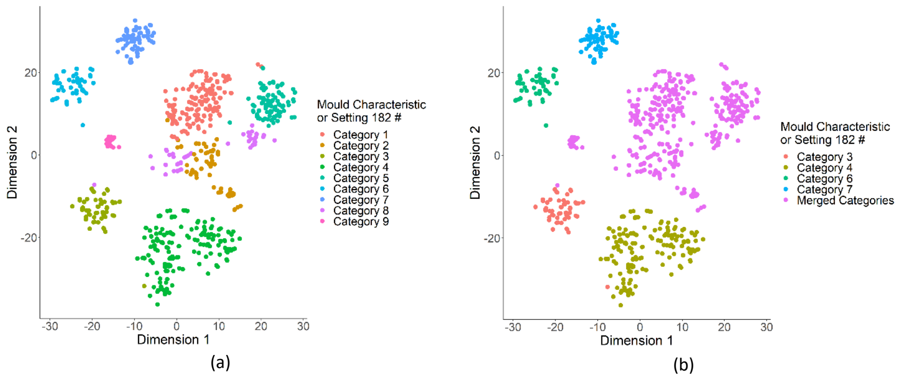


Figure 6. (a) t-SNE clustering using all categories and (b) t-SNE clustering using five selected categories.

Survival curves using the Kaplan-Meier estimator and random survival forest

In the available data set, only 10 % of the moulds are worn-out while the rest are still in use. Initially, the survival curves are estimated using the Kaplan-Meier estimator. Figure 7a shows a difference between the survival curves from different categories. The same conclusion is also reached after running the log rank test at a significance level of $\alpha = 0.05$. In Figure 7b, after also including the running moulds data, there are small differences in the survival curves that were based only on the worn-out data. It can be observed that more data is needed for the "Merged Categories" and "Category 7" for getting a better estimate of the survival curves.

The most important conclusion based on Figure 7 is that the clustering categories are beneficial for getting an overview of the mould's lifetime and can be used in production for getting an overview of the mould's state during production.

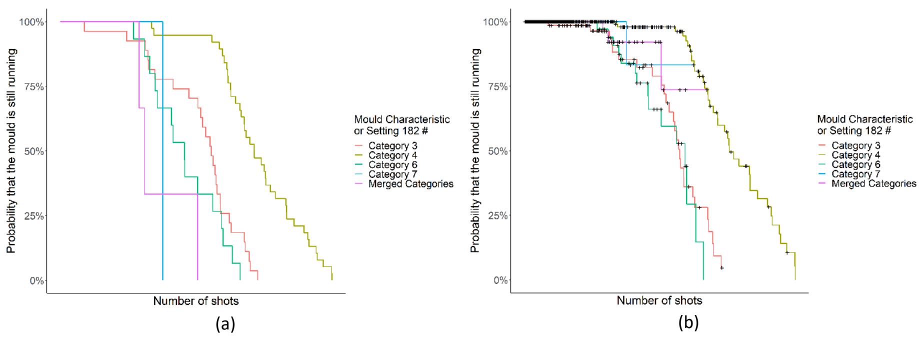


Figure 7. (a) Kaplan-Meier survival curve estimates using only worn-out data and (b) Kaplan-Meier survival curve estimates using worn-out and running data (censored data is marked with crosses).

What is important to mention at this stage is that during part production, a mould can be taken out for repair based on certain information that is unfortunately missing in the current case study. One example is given by the visual inspection of product quality, which is a good indicator when a mould is worn-out. However, this data set will only be available in the future, which means that a model using this information is plausible and needs to be considered. At present, the production and maintenance data sets are available and thus, only this data is incorporated into a random survival forest model, mostly for illustration purposes. The main point of interest is to compare the random survival forest model with the Kaplan-Meier estimated survival curves. A number of 10000 trees are used for fitting the random forest model and the worn-out data for training. After the model was trained, an OOB c-index of 0.85 was obtained, which indicates that the model is performing better than random. In Figure 8, the Kaplan-Meier survival curves are compared with the random survival forest survival curves from the worn-out data.

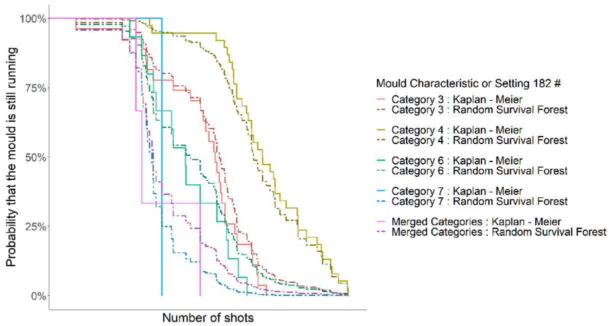


Figure 8. Comparison between Kaplan-Meier survival curve estimates and random survival forest on worn-out data.

Figure 8 shows that for "Category 3" and "Category 4", the curves are behaving similarly with a small difference towards the end of the life of "Category 3". For other categories however, the differences are more pronounced. On the other hand, the data in these categories is taking full advantage of the random survival forest model. Therefore, it was decided as a short term solution to use the Kaplan-Meier estimate of the survival curve and for the future, when more data is available, to use a random survival forest model. For a practical implementation in production, the authors are proposing a solution in the form of a dashboard using R Shiny (2019) which is described in detail in the next section.

4. R Shiny dashboard

The results presented above are incorporated into a dashboard such that it is easy for the industrial partner to monitor the status of the moulds. The R Shiny implementation is advantageous since it can easily be used on cloud or other data management platforms. The dashboard focuses on two different aspects. Firstly, the Kaplan-Meier survival curves are presented based on the historical data as well as using the running moulds data. Secondly, a way to monitor the running moulds in terms of conditional survival curves based on Kaplan-Meier estimates is presented. Furthermore, the early warning system for a running mould is also presented.

Historical data

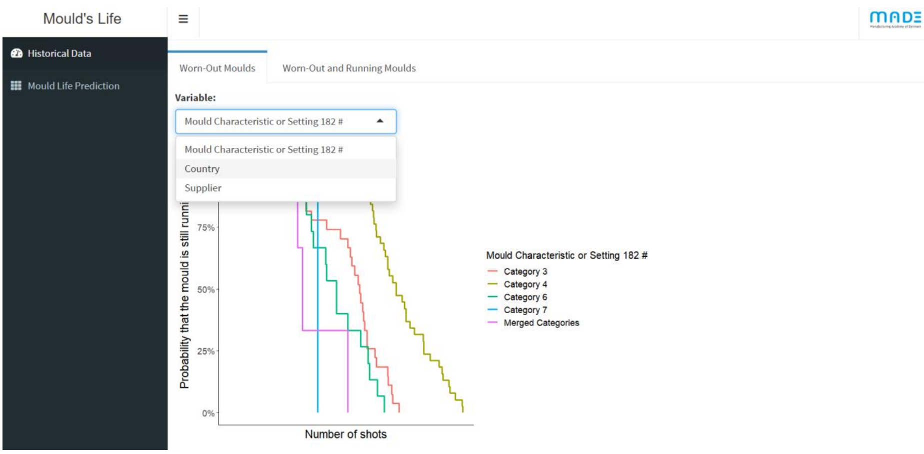


Figure 9. Snapshot from the Kaplan-Meier survival curve estimates for worn-out data.

In Figure 9, on the left panel in black, there are two options as a choice, namely "Historical Data" and "Mould Life Prediction". The "Historical Data" option is highlighted which further contains two tabs, i.e. "Worn-Out Moulds" and "Worn-Out and Running Moulds". The "Worn-Out and Running Moulds" contains the exact same plot as "Worn-Out Moulds" with the exception that the survival curves are built using all the available data, i.e., worn-out and running moulds data. The "Worn-Out Moulds" tab contains only the survival curves built on the worn-out moulds data. The "Variable:" field contains three different choices. The first choice is given by the five categories of "Mould Characteristic or Setting 182 #" which provides a quick overview over the historical worn-out moulds. Furthermore, since the industrial partner has different production sites, it was decided to provide survival curves computed using the Kaplan-Meier estimate for different countries. The same is valid for different mould suppliers. In this way, the industrial partner can have a convenient global overview over the worn-out moulds. Crucially, the curves can be updated in real-time. Thus, each time a new worn-out mould is stored in the database, the survival curves estimates can be updated instantaneously. Deciding when to update the database is also an important question but it is beyond the scope of this study. The option to also view the survival curves using the running moulds data was included so that the industrial partner can check if the running moulds are changing in some way the behavior of the survival curves in different regions.

Mould life prediction

In Figure 10, the "Mould Life Prediction" choice is highlighted. There are two tabs associated with this choice. The first tab "Using Worn-Out Moulds Data", contains the conditional survival curve (with the confidence interval) based on the category of the running mould using only the worn-out data. Moreover, this tab also contains the early warning system for the running mould. The second tab, "Using Worn-Out and Running Moulds Data" contains the conditional survival curve based on all the data which means worn-out and running data without the early warning system.

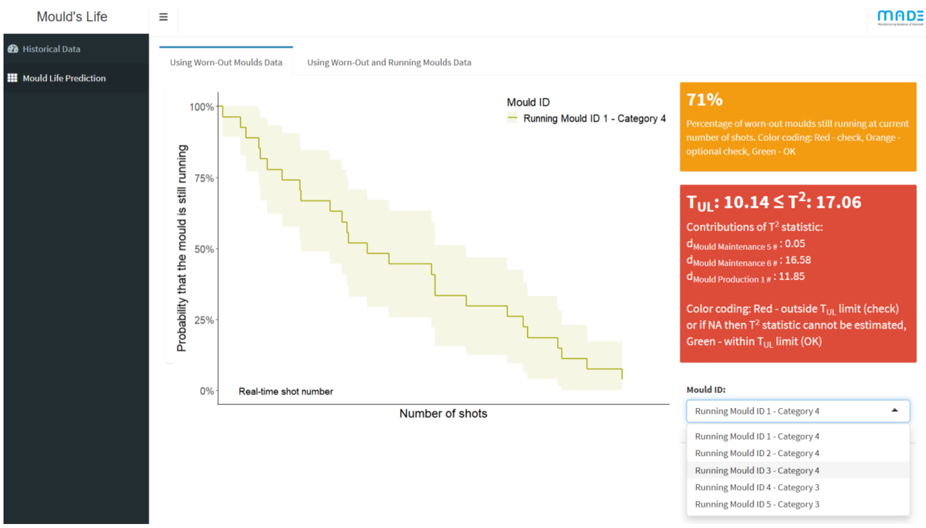


Figure 10. Snapshot from the early warning system along with the conditional survival curve for a running mould based on the Kaplan-Meier estimate.

In Figure 10, the conditional survival curve can be visualized along with the 95 % confidence interval for a running mould in "Category 4" at the corresponding running shot number. Furthermore, on the right side, two different boxes are presented. The first box presents the percentage of the active moulds from the worn-out moulds at that shot number. Here, there are 71.1 % of the worn-out moulds still running at this number of shots. It was decided to color the boxes according to the "risk" associated with the percentage of available historical worn-out moulds. For example, if the percentage lies between 0 and 20 % then a red color is used, whereas if the percentage lies between 20 and 80 % then the color becomes orange. If the percentage lies between 80 and 100 % then the box is colored green. These categories were chosen only for illustration purposes. In the industrial setting, these values should be adjusted by the operators' needs. Using the color coding, the industrial partner has an immediate overview of the status. The second box (in red) presents the Hotelling T^2 statistic along with the T_{UL} limit. Also, this box is color coded dynamically based on the results. If the box is green, the Hotelling T^2 statistic is within the limit whereas if the box is red, then the statistic lies outside the limit or the statistic cannot be estimated. In addition, the contributions for three features are also presented. These features were chosen based on some discussions with the industrial partner. The early warning system, can immediately point out if the running mould is behaving similarly in terms of production and maintenance with the historical data as the statistic is built on the average at that number of shots of the worn-out mould data.

As an upgrade for the "Mould Life Prediction" choice, one could use a random forest model instead of the Kaplan-Meier survival curves estimates, thus, incorporating also other production data. In terms of the early warning system, one could use a combined Hotelling T^2 statistic along with a Q chart with the corresponding contributions.

5. Conclusion and future outlook

In this article, a case study using industrial data from the plastic injection moulding industry is presented. The problem considered in the case study is the prediction of worn-out moulds with a focus on visualization and presentation in real-time through the means of an R Shiny dashboard. The analysis is divided into two parts. First, the authors focus on clustering using mixed data and then on worn-out mould prediction. Using the results of the cluster analysis, the worn-out moulds are presented in terms of survival curves using the Kaplan-Meier estimator. The cluster analysis shows that one feature in particular can characterize the moulds very well. Moreover, a comparison between the survival curves using the Kaplan-Meier estimate and a random survival forest model is presented.

In terms of real-time monitoring of the running moulds, a R Shiny dashboard has been implemented. One tab of the dashboard shows the survival curves estimates via the Kaplan-Meier estimator using both worn-out moulds as well as all available data including the data on the running moulds. The other tab focuses on presenting the conditional survival curve estimates for a running mould including an early warning system. This system is based on a percentage that indicates the proportion of the running worn-out moulds at that number of shots and a Hotelling T^2 statistics along with the production or maintenance feature contributions. To the best of our knowledge the early warning system has not been implemented for this type of problem before and also not in this manner.

As future work, the aim is to incorporate all available data into a random forest model for getting a better survival curve prediction. The data in this case study was unfortunately missing product metrology data, which it is considered important for assessing the mould wear and tear. Other matters that will be interesting to investigate are data management related issues. For example, how often should the training data used for model building be updated, and how much data should be included in the training data? In terms of the early warning system, if the p features are highly correlated, then the Hotelling T^2 and Q statistics (De Ketelaere, Hubert, and Schmitt 2015) using Principal Components Analysis (PCA) (Hastie, Tibshirani, and Friedman 2013) 2013) can be used. The contributions can also be computed for this case as presented in the work of Miller, Swanson, and Heckler (1998) and De Ketelaere, Hubert, and Schmitt (2015). Moreover, a combined statistic based on Hotelling T^2 and Q statistics (e.g. Frumosu and Kulahci 2019) could be used instead of just a Hotelling T^2 statistic such that more features could be used for monitoring along with the corresponding contributions presented as a bar chart.

Acknowledgments

This work has been carried out under MADE SPIR – Strategic Platform for Innovation and Research, Denmark and the authors are grateful for the given opportunity. The authors would like to thank Max Peter Spooner for proofreading the article and the industrial partner for providing the data sets and for the allocated time.

Disclosure statement

The authors of the article have not encountered any potential conflict of interest.

References

Ahmad, A., and S.S. Khan. 2019. "Survey of State-of-the-Art Mixed Data Clustering Algorithms." Ieee Access 7 (99): 31883 – 31902. doi:10.1109/ACCESS.2019.2903568.

Ahmad, R. and S. Kamaruddin. 2012. "An overview of time-based and condition-based maintenance in industrial application." Computers and Industrial Engineering 63 (1): 135-149. doi:10.1016/j.cie.2012.02.002.

Aivaliotis, P., K. Georgoulias, and G. Chryssolouris. 2019. "The use of Digital Twin for predictive maintenance in manufacturing." International Journal of Computer Integrated Manufacturing 32(11): 1067-1080. doi: 10.1080/0951192X.2019.1686173.

Alsina, E., M. Chica, K. Trawiński, and A. Regattieri. 2018. "On the use of machine learning methods to predict component reliability from data-driven industrial case studies." International Journal of Advanced Manufacturing Technology 94 (5-8): 2419-2433. doi:10.1007/s00170-017-1039-x.

Amari, S. V., L. McLaughlin, and H. Pham. 2006. “Cost-effective condition-based maintenance using Markov decision processes.” In RAMS'06 Annual Reliability and Maintainability Symposium, 464-469. IEEE. doi:10.1109/RAMS.2006.1677417

Boström, H., L. Asker, R. Gurung, I. Karlsson, T. Lindgren, P. Papapetrou. 2018. “Conformal prediction using random survival forests.” Proceedings - 16th Ieee International Conference on Machine Learning and Applications, Icmla 2017 2018: 812-817. doi:10.1109/ICMLA.2017.00-57.

Breiman L. 2001. “Random forests.” Machine Learning 45 (1):5-32. doi:10.1023/A:1010933404324.

Breiman , L., and A. Cutler. 2019. Random Forests. University of California, Berkley. Accessed June 21 2020. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm

Bukkapatnam, S.T.S., K. Afrin, D. Dave, and S.R.T. Kumara. 2019. ”Machine learning and AI for long-term fault prognosis in complex manufacturing systems.” Cirp Annals 68 (1): 459-462. doi:10.1016/j.cirp.2019.04.104.

Cachada, A., J. Barbosa, P. Leitão, C. A. S. Geraldes, L. Deusdado, J. Costa, C. Teixeira, J. Teixeira, A. H. J. Moreira, P. M. Moreira, and L. Romero. 2018. “Maintenance 4.0: Intelligent and predictive maintenance system architecture.” In 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), 139-146. IEEE. doi:10.1109/ETFA.2018.8502489.

Da Silva, V. L., J.L. Kovaleski, R.N. Pagani, J.D.M. Silva, and A. Corsi. 2020. “Implementation of Industry 4.0 concept in companies: Empirical evidences.” International Journal of Computer Integrated Manufacturing 33 (4): 325-342. doi: 10.1080/0951192X.2019.1699258.

De Ketelaere, B., M. Hubert, and E. Schmitt. 2015. “Overview of PCA-based statistical process-monitoring methods for time-dependent, high-dimensional data.” Journal of Quality Technology 47 (4): 318-335. doi:10.1080/00224065.2015.11918137.

Foss, A. H., M. Markatou, and B. Ray. 2019. “Distance metrics and clustering methods for mixed-type Data.” International Statistical Review 87 (1): 80-109. doi:10.1111/insr.12274.

Frumosu, F. D., and M. Kulahci. 2019. “Outliers detection using an iterative strategy for semi-supervised learning.” Quality and Reliability Engineering International 35 (5): 1408–1423. doi:10.1002/qre.2522.

Gower, J. C. 1971. “A General Coefficient of Similarity and Some of Its Properties.” Biometrics 27 (4): 857-871. doi:10.2307/2528823.

Harrell, F. E., K.L. Lee, and D.B. Mark. 1996. “Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors.” Statistics in Medicine 15 (4): 361-387.doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

Harrell, F.E., R. Califf, D. Pryor, K. Lee, and R. Rosati. 1982. “Evaluating the yield of medical tests.” Journal of the American Medical Association 247: 2543–2546. doi:10.1001/jama.1982.03320430047030.

Harrington, D. P., and T.R. Fleming. 1982. “A class of rank test procedures for censored survival data.” Biometrika 69 (3): 553-566. doi:10.1093/biomet/69.3.553.

Hashemian, H. M. 2010. “State-of-the-art predictive maintenance techniques.” IEEE Transactions on Instrumentation and measurement 60 (1): 226-236. doi:10.1109/TIM.2010.2047662.

Hastie, T., R. Tibshirani, and J. Friedman. 2013. The elements of statistical learning: Data mining, inference, and prediction. New York: Springer.

Hieke, S., M. Kleber, C. König, M. Engelhardt, and M. Schumacher. 2015. “Conditional survival: A useful concept to provide information on how prognosis evolves over time.” Clinical Cancer Research 21 (7): 1530-1536. doi:10.1158/1078-0432.CCR-14-2154.

Haarman, M., M. Mulders, and C. Vassiliadis. 2017. Predictive Maintenance 4.0 - Predict the unpredictable. PwC and Mainnovation. Accessed June 28 2020. https://www.pwc.nl/nl/assets/documents/pwc-predictive-maintenance-4-0.pdf

Ishwaran, H., U. Kogalur, X. Chen, and A. J. Minn. 2011. "Random survival forests for high-dimensional data. Statistical Analysis and Data Mining." 4 (1): 115-132. doi:10.1002/sam.10103.

Ishwaran, H., U.B. Kogalur, E.H. Blackstone and M.S. Lauer. 2008. "Random survival forests." Annals of Applied Statistics 2 (3): 841-860. doi:10.1214/08-AOAS169.

Jardine, A.K.S., D. Lin, and D. Banjevic. 2006. "A review on machinery diagnostics and prognostics implementing condition-based maintenance." Mechanical Systems and Signal Processing 20: 1483–1510. doi:10.1016/j.ymssp.2005.09.012.

Kaplan, E. L., and P. Meier. 1958. "Nonparametric Estimation from Incomplete Observations." Journal of the American Statistical Association 53 (282): 457-481. doi:10.1080/01621459.1958.10501452.

Kaufman, L., and P.J. Rousseeuw. 1987. "Clustering by means of medoids." Statistical Data Analysis Based on the L1-norm and Related Methods. First International Conference. 405-16.

Lasi, H., H-G. Kemper, P. Fetteke and T. Feld. 2014. "Industry 4.0." Business and Information Systems Engineering 6: 239–242. doi:10.1007/s12599-014-0334-4.

Li, Z., K. Wang, and Y. He. 2016. "Industry 4.0 - Potentials for predictive maintenance." Proceedings of the 6th International Workshop of Advanced Manufacturing and Automation, 42-46, Atlantis Press.

Li, Z., Y. Wang, and K.S. Wang. 2017. "Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario." Advances in Manufacturing 5(4): 377-387. doi: 10.1007/s40436-017-0203-8.

Miller, P., R. E. Swanson, and C. E. Heckler. 1998. "Contribution plots: a missing link in multivariate quality control." Applied Mathematics and Computer Science 8 (4): 775-792.

Montgomery, D. C. 2012. Introduction to Statistical Quality Control. 7th Edition. Wiley

Peng, Y., M. Dong, and M.J. Zuo. 2010. "Current status of machine prognostics in condition-based maintenance: A review." International Journal of Advanced Manufacturing Technology 50 (1-4): 297-313. doi:10.1007/s00170-009-2482-0.

R Shiny. 2019. Shiny: Easy web applications in R. RStudio Inc. Accessed 21 June 2020. http://shiny.rstudio.com/

Ragab, A., M. S. Ouali, S. Yacout, and H. Osman. 2016. "Remaining useful life prediction using prognostic methodology based on logical analysis of data and Kaplan - Meier estimation." Journal of Intelligent Manufacturing 27 (5): 943-958. doi: 10.1007/s10845-014-0926-3.

Rousseeuw, P. J. 1987. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." Journal of Computational and Applied Mathematics 20: 53-65.

Sai, V. C., M.V. Shcherbakov, and V.P. Tran. 2019. "Data-driven framework for predictive maintenance in Industry 4.0 concept." In Conference on Creativity in Intelligent Technologies and Data Science, 344-358. Springer, Cham. doi:10.1007/978-3-030-29743-5_28.

Shi, T., and S. Horvath. 2006. "Unsupervised learning with random forest predictors." Journal of Computational and Graphical Statistics 15 (1): 118-138. doi:10.1198/106186006X94072.

Si, X. S., W. Wang, C. H. Hu, and D. H. Zhou. 2011. "Remaining useful life estimation - A review on the statistical data driven approaches." European Journal of Operational Research 213 (1): 1-14. doi:10.1016/j.ejor.2010.11.018.

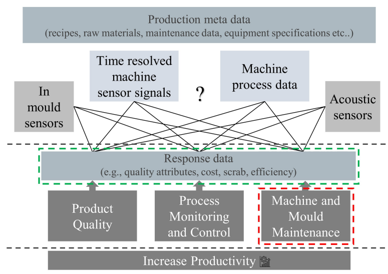
Van Der Maaten, L., and G. Hinton. 2008. "Visualizing data using t-SNE." Journal of Machine Learning Research 9: 2579-2625.

Wang, P., Y. Li, and C.K. Reddy. 2019. “Machine learning for survival analysis: A survey.” Acm Computing Surveys 51 (6): 110. doi:10.1145/3214306.

Zabor, E. C., M. Gonen, P.B. Chapman, and K. S. Panageas. 2013. “Dynamic Prognostication Using Conditional Survival Estimates.” Cancer 119 (20): 3589-3592. doi:10.1002/cncr.28273.

APPENDIX C

Paper F, Data-Driven Identification of Remaining Useful Life for Plastic Injection Moulds



Within injection moulding, the condition of the mould cavities and cores are critical to the quality of the produced element (both for element dimension and surface appearances). Visual inspection of the mould cavities and cores when the mould is maintained together with the daily visual inspection of the produced elements is effective in detecting surface problems in the mould. However, the same approach cannot be used to detect mould degradation affecting the element dimensions. To detect cavity and core degradation, measurements have to be done directly on the cavities and cores or alternatively on the produced elements. Since quality measures are already made on the produced elements, the natural next step would be to investigate if this data could be used to detect mould degradation (mould parts are not measured by default). However, variation in element dimension will also reflect process and material variations as demonstrated in paper A and B, which have to be taken into consideration. The work is linked to RQ1, addressing the loss of productivity by mitigating productivity loss caused by mould malfunction.

Research question 1 (RQ1)

How to best utilise data from injection moulding machines to increase the productivity of producing injection moulded elements?

As stated in the outline of future work in paper E, the product metrology data is expected to be essential for assessing mould wear and tear. Paper F addresses this by exploring metrology data to detect long-term degradation of injection moulds.

C.1 Summary

Metrology data from multiple identical moulds (producing the same element shape in the same material) was collected for this investigation. Twelve different metrology measures from quality inspections were included in the data. To enforce patterns in the metrology data related to mould degradation, the average values (samples collected in the first 2 million cycles) for individual mould/cavity/metrology measure is subtracted from the same mould/cavity/metrology measures making all values variate around zero. Subtracting the average of the initial samples for the individual mould also makes it possible to model across moulds since it will also subtract initial differences between the moulds. Plotting the scaled data for individual metrology measures as a function of the number of moulding cycles (used as a measure of time) reveals degradation patterns linked to an explainable physical phenomenon in the mould. Multivariate Statistical Process Control (MSPC) were tested for monitoring this degradation. To overcome problems with cross-correlation, the 12 metrology measures were first reduced to latent variables using PCA. Using T^2 and Q control charts, it was demonstrated that MSPC could be used for mould degradation monitoring. An alternative approach to MSPC could be a classification of the mould state (run-in, production and worn-out). To represent all mould states in the model and make it work for future moulds, data for multiple moulds needed to be included in the model. Since there is variation between the moulds and operating conditions, this has to be reflected by including additional data. The mould construction, production site, screw design, colour and moulding machine was added as categorical variables to reflect this variation. The categorical data were converted to dummy variables and reduced to ten latent variables, explaining 94% of the original variation and then combined with the raw metrology measures. Classification accuracy of 88% was achieved for correctly predicting a mould-worn out.

Achieved results

- Analysis of mould degradation pattern reflected in element metrology measures.
- Scaling approach to enable monitoring of samples consisting of elements from multiple moulds, cavities and different metrology measures.
- Unsupervised MSPC approach for monitoring the change in product quality for products with multiple quality attributes.

Contribution

- Comparison of unsupervised and supervised approach for detecting out-of-control behaviour for plastic injection moulds.
- The proposed solution can be utilised for supporting and scheduling mould maintenance and support mould worn-out evaluation.

C.2 Paper F

Paper F Data-Driven Identification of Remaining Useful Life for Plastic Injection Moulds

Till Böttjer[1(✉)], Georg Ørnskov Rønsch[4], Cláudio Gomes[1,2], Devarajan Ramanujan[2,3], Alexandros Iosifidis[1,2], and Peter Gorm Larsen[1,2]

[1] Department of Electrical and Computer Engineering, Aarhus University, Aarhus, Denmark
till.boettjer@ece.au.dk
[2] DIGIT Center, Aarhus University, Aarhus, Denmark
[3] Department of Mechanical and Production Engineering, Aarhus University, Aarhus, Denmark
[4] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark

Abstract. Throughout their useful life, plastic injection moulds operate in rapidly varying cyclic environments, and are prone to continual degradation. Quantifying the remaining useful life of moulds is a necessary step for minimizing unplanned downtime and part scrap, as well as scheduling preventive mould maintenance tasks such as cleaning and refurbishment. This paper presents a data-driven approach for identifying degradation progression and remaining useful life of moulds, using real-world production data. An industrial data set containing metrology measurements of a solidified plastic part, along with corresponding lifecycle data of 13 high production volume injection moulds, was analyzed. Multivariate Statistical Process Control techniques and *XGBoost* classification models were used for constructing data-driven models of mould degradation progression, and classifying mould state (early run-in, production, worn-out). Results show the *XGBoost* model developed using element metrology & relevant mould lifecycle data classifies worn-out moulds with an in-class accuracy of 88%. Lower in-class accuracy of 73% and 61% were achieved for the compared to mould-worn out less critical early run-in and production states respectively.

Keywords: Smart manufacturing · Injection moulding · Data-driven model · Machine learning · Predictive maintenance

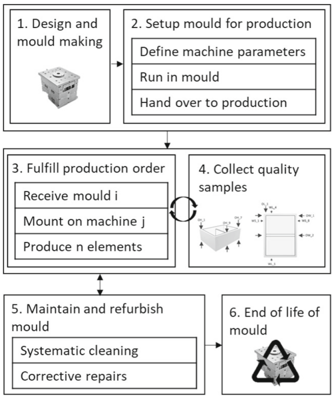
1 Introduction

High-volume manufacturing processes such as plastic injection moulding, require considerable upfront investment in tooling to develop a reliable moulding operation [1]. Consequently, manufacturers are keen to maximize tool productivity by eliminating unplanned maintenance, minimizing part scrap, and extending the useful life of their injection moulds. Previous research has developed physics-based models of wear mechanisms affecting functionality of the mould and the quality of the moulded element, and validated them using simulation and controlled experimental studies. Engelmann et al. [2] describe the progression of failure modes associated with specific components within a mould by optical inspection. Zabala et al. [3] develop a wear model quantifying the wear of the mould using a tribometer, erosion test, gravelometer and electro-chemical impedance spectroscopy with varying mould coatings and plastics raw material. Zhong et al. [4] compare the wear rate of three insert materials and determine the surface texture of the elements (using an electron microscope) as a function of injection cycles. However, these studies do not evaluate the applicability of the developed models to predict longevity of moulds in real-world production environments. To overcome these constraints, a variety of data-driven approaches and machine learning models have been proposed for estimating element quality (such as element weight, dimensions, etc.) as function of process and machine data. Schulze et al. [5] propose an automated workflow for developing predictive quality models for a plate specimen based on a variety of machine and process parameters. Ogorodnyk et al. [6] classify low and high quality elements on-line as a function of machine and process parameters. Frumosu et al. [7] present an industrial application to predict mould lifetime from a data set containing initial process settings and tool characteristics such as, layout, construction, and number of cavities. In spite of these studies, there is limited understanding on which models are applicable towards mould longevity prediction as the accuracy of such models are limited by the nature and resolution of data that can be effectively measured on the factory floor.

In this paper, we aim to address the above knowledge gap by systematically analyzing long-term mould life-cycle data collected from a real-world industrial case. Results from our study contribute to existing research on mould degradation prediction by, (i) presenting existing data-related and modeling-related challenges, and (ii) identifying opportunities for successful application of data-driven prediction models for injection mould degradation and classifying moulds based on remaining useful life.

2 Industrial Data Set

In this section, we introduce the typical life-cycle of an injection mould and relevant data collected at our industrial partner. Below we use the term "part" as reference to the physical components of a mould and "element" as reference to the solidified product from the moulding process. Our main goal is to use metrology samples that are commonly collected in industrial injection moulding for classifying three wear states of a mould in operation (early run-in, production, worn-out). Figure 1 displays the six life-cycle stages of a mould at our industrial partner (left) and lists relevant data, including type and life-cycle stage from which we collect it (right). The stages most relevant for our contribution are: (1) Design, create and assemble the mould for a specific element design; (2) Run tests to find the optimal operating point and hand-over the mould to production; (3) Fulfill an incoming production order; (4) Continuously monitor element quality



Stage	Data name	Data type
1-6	Mould ID	Categorical (13)
1-6	Moulding Date	Year (2011-2020)
1-6	Number of Cycles	Numeric (0-15e^6)
1	Construction	Categorical (5)
1	Layout	Categorical (3)
3	Order Number	Categorical (503)
3	Machine	Categorical (17)
3	Screw Design	Categorical (6)
3	Production Site	Categorical (4)
3	Color	Categorical (29)
4	Metrology point	Categorical (12)
4	Measurement value	Numeric (*)
3, 5-6	Maintenance data	Free text

Fig. 1. Illustration of the six mould lifecycle stages at our industrial partner (left) and description of the data set (right). *Stage* labels the origin of the data. *Data type* describes attributes of the data: Categorical types include number of unique values; Numeric types include the range. *The range of measurement values depends on the specific metrology point.

by collecting samples; (5) Conduct planned and unplanned maintenance; and (6) Remove the mould from production. We refer the reader to Kazmer [8] for more details on mould design, element approval and mould maintenance.

The data set contains historic measurements of the same moulded element geometry produced using 13 different injection moulds. The *Mould ID* uniquely identifies a mould throughout its lifetime. The data was collected in a period of nine years from 2011–2020. Instead of associating events with a date, we use the *Number of Cycles* to quantify the age of a mould as a function of usage in production independent of time spent in storage, workshop etc. Five mould constructions and three mould layouts were included in our analysis. The data set contains additional information on single production orders i.e. *Order Number* as unique label, machine type (hydraulic, electric, or hybrid), design of the injection screw and production site. The raw material used in this study was ABS mixed with one of 29 different color additives.

Metrology Data. The produced element is a small rectangular box, with a supporting rip at the center on the long side (to reduce warpage). Figure 2 illustrates the element geometry and metrology measuring points. In total 10 measuring points including four specific measures are collected (length, width, height and wall thickness). The element metrology is sampled every three weeks. A quality sample consists of a batch of elements from each cavity of the mould from a single injection cycle. In total, the data set contains 17,203 element metrology samples per measurement point. To preserve data confidentiality, all metrology measures are scaled according to the specification limits provided by the industrial partner between **+1** (upper specification limit) and **–1** (lower specification limit) with **0** indicating the target value. Thus, positive values indicate elements larger than target and negative values indicate elements smaller than target. Elements above/below **+1**/**−1** are outside of specifications and are thus rejected.

ID (Positions)	Measurement
OH_x (1,7,9)	Outside Height
OL_x (1)	Outside Length
OW_x (1,2)	Outside Width
WS_x (1,8)	Side wall thickness
WL_x (1,4)	End wall thickness
WS_tot (WS_1+WS_8)	Sum of wall thickness
WL_tot (WL_1+WL_4)	Sum of wall thickness

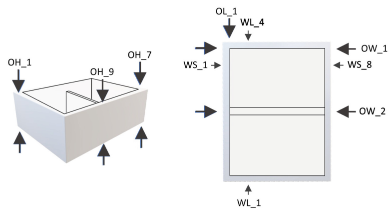


Fig. 2. Illustration of the moulded element and description of location of metrology measurements. WS_tot and WE_tot are calculated as the sum of the two wall thicknesses at the side positions and end positions respectively to reduce the impact of variations in the alignment of the two mould halves.

3 Exploratory Analysis of Maintenance and Metrology Data

A majority of the maintenance data is free text allowing detailed description of the problem/cause but interfering with an automated analysis. All free text entries are linked to one of 11 different root causes displayed in Fig. 3.

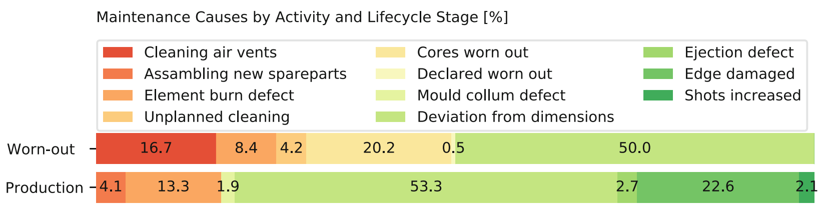


Fig. 3. Frequency percentage of 11 maintenance events by lifecycle stage. No maintenance was recorded in the early run-in period.

For both production and worn-out moulds, majority of maintenance are associated with deviation from dimensions of the elements (53.3% and 50.0% respectively). Followed by edge damage (22.6%) and burn marks (13.3%) for production moulds and cores worn-out (20.2%) and cleaning of air vents (16.7%) for worn out moulds. In the following analysis we focus on the factor with greatest impact i.e. deviation from dimensions.

Due to mould wear and degradation, we expect the different metrology measures to change over time. When introduced into production, injection moulds can have dissimilar dimensions due to the tolerances of upstream manufacturing processes. Consequently, initial element dimensions vary from mould to mould. Thus, to investigate mould degradation across moulds we align the metrology measurements by subtracting the average values for individual mould/cavity/metrology measure combination for the first 2.000.000 moulding cycles from the same mould/cavity/metrology measure combination. Our analysis indicates that for all moulds the height of the elements decreases with cycle count. Further, our analysis shows that the element wall thickness and length increases as function of cycle count. We show the change of element height (OH_7), wall thickness (WS_tot), outside width (OW_2) and (OL_1) for three illustrative moulds (Fig. 4) due to the large size of the entire data set. As shown, both element height and wall thickness (Fig. 4B and Fig. 4D) exhibit a consistent downward and upward trend respectively. The outside width (Fig. 4C) indicates a marginal positive slope resulting in wider elements. This can be the result of progressive compression of the mould parts in clamping direction reducing element height and abrasive material loses from cavity walls, and core surfaces increasing the wall thickness. While we expected a similar increasing trend for the element length due to abrasion of the cavity walls, Fig. 4A displays a steep positive slope for one mould, and a negative slope for the other two. Our analysis of metrology data shows the height of the elements and the wall thickness give the most consistent indicators of mould degradation.

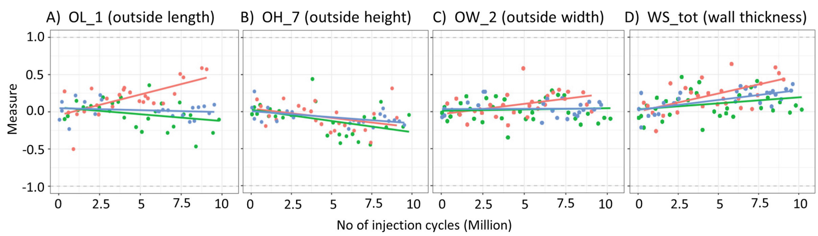


Fig. 4. Change in metrology measures by mould age (number of injection cycles) for three randomly selected moulds.

4 Monitoring of Mould Worn-Out Using MSPC

Section 3 shows specific element dimensions can be used as proxy for analyzing mould degradation. One of the most common methods for monitoring process quality (i.e. deviation in element dimensions) is using control charts [9]. Instead of monitoring the 12 metrology measures individually, we apply Multivariate Statistical Process Control (MSPC) with the quality measures reduced to latent variables using Principal Component Analysis (PCA) (see MacGregor et al. [10]) for systematically tracking element dimensions and thereby degradation of moulding parts in a single control chart. Based on the latent variables, we derive Hotelling T^2 for monitoring the matrix of the latent variables and Q-statistics for monitoring the residuals.

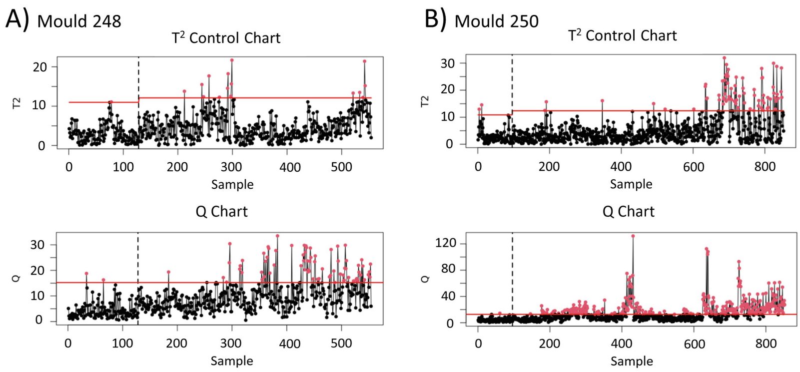


Fig. 5. MSPC for the two moulds declared worn-out due to deviations in dimensions. Samples from the initial 2.000.000 cycles were used to create the baseline model.

Even though element dimensions were the largest contributor to maintenance actions, the actual cause of disposal was only recorded for a small fraction of moulds in the data set. However, to verify the applicability of element quality measures for monitoring mould degradation, our analysis required moulds that were marked as worn-out due to element dimensions being out of specifications. Only two such moulds (Mould ID 248 & 250) were present in the data set. Therefore, we selected these moulds for demonstrating the applicability of MSPC and creating the relevant control charts. The Hotelling T^2 and Q statistic for these molds and control limits (the solid red lines) are shown in Fig. 5.

The control charts enabled monitoring the degradation of the moulds linked to elements being out of dimensions. Mould 248 was declared worn-out between sample 450–550. As show in Fig. 5A, the Hotelling T^2 indicates an out of control behaviour in this sample range. For mould 250 the T^2 chart in Fig. 5B implies fairly constant element dimensions (and correlation between dimensions) for the first 500 samples. From approximately sample 500, variation in T^2 is increasing. From sample 650 the first indication of worn-out is present as T^2 surpasses the control limit. The corresponding Q chart implies an out-of-control situation after sample 400 indicating a change in the model residuals. This can be due to changes in the part of the correlation structure not included in the selected latent variables. Situations like this have to be investigate before a more widely implementation of a MSPC solution. Further, for Mould ID 248 the control limit is exceeded between sample 250 and 300 on the T^2 chart shown in Fig. 5A. The free text in the maintenance data describes that shortly after sample 300 cavity inserts were change due to element dimension being outside of the specification limits. Using the T^2 chart an early indication of the insert worn-out could have been realized at sample 250. The Q chart in Fig. 5A shows that the model residual change after assembling of new insert parts. Thus, the underlying PCA model requires updating to account for variations introduced by changing the inserts. The above results from the two moulds indicate that MSPC is applicable for monitoring element dimension and that the variation seen in the T^2 charts can be linked to mould degradation.

5 Classification of Mould State

To extend the results in Sect. 4 and investigate mould degradation for all moulds using metrology measures, we develop a classification model with the wear states labelled early run-in, production and worn-out. Based on expert input from the industrial partner, we define the early run-in state as the first 30% of the collected samples, production as samples collected between 30–80%, and worn-out as the samples collected above 80% of the maximum cycle count for a mould. Partial Least Squares - Discriminant Analysis (PLS-DA) [11] and *XGBoost* [12] were tested on two different data sets, one being the 12 metrology measures (compare Fig. 2) and one being the 12 metrology measures combined with five selected categorical features (Construction, Production site, Screw, Colour and Machine). We choose PLS-DA since PCA captured the latent structure related to worn-out and XGBoost as it is a tree-based method performing well for a data-set with both numerical and categorical features. The categorical features were introduced in the data set by converting them to dummy variables and applying PCA for reducing them to ten latent variables (explaining 94% of the total variation). The XGBoost model using the metrology data combined with the latent representation of the five categorical variables shows the best performance among the four models i.e. PLS-DA with and without latent variables, XGBoost with and without latent variables. The results indicate that while the discrimination between early run-in and production achieves only a within-class accuracy of 73% and 61% respectively, the worn-out class achieves accuracy of 88%. In comparison, the XGBoost model excluding the latent variables leads an accuracy of 74% for the worn-out class. Further, the two PLS-DA models achieve an worn-out class accuracy of only 30% with and 35% without latent variables. Distinguishing between early run-in and production class is not critical, as detection of the worn-out state.

6 Discussion and Conclusion

This paper used real-world production data to develop data-driven models capable of detecting long-term degradation of plastic injection moulds. Analysis of metrology measurements from the industrial data set confirmed that dimensions of moulded elements changed systematically over time. This can be explained by degradation of moulding parts (compression of moulding parts resulting in reduced height of the elements, abrasion of material from the walls in the cavities and on the cores, resulting in longer and wider elements with thicker walls).

Based on these findings two different approaches were explored for monitoring degradation progression and detecting worn-out moulds (utilizing metrology data). Results show that an MSPC-based approach can be used for monitoring mould degradation and for supporting decisions related to change of inserts/cores and declaration of mould wear-out. Additionally, our results demonstrated that fairly accurate identification of remaining useful life of the mould (based on mould state classification) is possible using element metrology and a latent representation of key categorical variables. Both PLS-DA and XGBoost were tested as classifiers and XGBoost was found to be superior achieving a within-class accuracy of 88% for the worn-out class.

The results in this paper can be used for supporting and scheduling of mould maintenance. Further, they create a foundation for developing solutions for mould monitoring and decision support on when to perform a mould worn-out evaluation. We expect that collection of additional time series data such as machine and process data form the moulding machines can help to distinguish variation in the metrology measures due to material/process variations and degradation of mould parts.

References

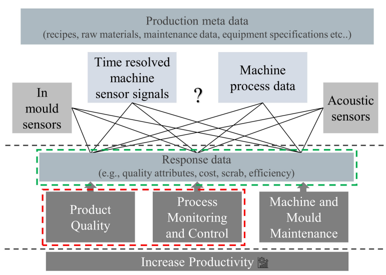
1. Ribeiro, I., Peças, P., Henriques, E.: Incorporating tool design into a comprehensive life cycle cost framework using the case of injection molding. J. Cleaner Prod. **53**, 297–309 (2013)
2. Engelmann, P., Hayden, K., Guichelaar, P., Dealey, R., Monfore, M.: Injection mold wear mechanisms and mold design. Plast. Eng. **57**, 40–46 (2001)
3. Zabala, B., et al.: Mechanism-based wear models for plastic injection moulds. Wear **440**, 203105 (2019)
4. Zhong, Z., Leong, M., Liu, X.: The wear rates and performance of three mold insert materials. Mater. Des. **32**(2), 643–648 (2011)
5. Schulze Struchtrup, A., Kvaktun, D., Schiffers, R.: A holistic approach to part quality prediction in injection molding based on machine learning. In: Hopmann, C., Dahlmann, R. (eds.) Advances in Polymer Processing 2020, pp. 137–149. Springer, Heidelberg (2020). https://doi.org/10.1007/978-3-662-60809-8_12
6. Ogorodnyk, O., Lyngstad, O.V., Larsen, M., Wang, K., Martinsen, K.: Application of machine learning methods for prediction of parts quality in thermoplastics injection molding. In: Wang, K., Wang, Y., Strandhagen, J.O., Yu, T. (eds.) IWAMA 2018. LNEE, vol. 484, pp. 237–244. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-2375-1_30
7. Frumosu, F.D., Rønsch, G.Ø., Kulahci, M.: Mould wear-out prediction in the plastic injection moulding industry: a case study. Int. J. Comput. Integr. Manuf. **33**(12), 1245–1258 (2020)
8. Kazmer, D.O.: Injection mold design engineering. Carl Hanser Verlag GmbH Co KG (2016)
9. Ge, Z., Song, Z.: Multivariate Statistical Process Control. Process Monitoring Methods and Applications, Springer, London (2013) https://doi.org/10.1007/978-1-4471-4513-4
10. MacGregor, J., Kourti, T.: Statistical process control of multivariate processes. Control Eng. Pract. **3**(3), 403–414 (1995)

11. Wold, S., Sjöström, M., Eriksson, L.: Pls-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. **58**(2), 109–130 (2001)
12. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery (2016)

APPENDIX D

MSPC-PCA used for monitoring of element quality

The work presented here is based on the presentation ”SPC in a multivariate system - a practical approach” [67], presented at ”Få processen i kontrol med Statistisk Proceskontrol (SPC)” conference at the Danish engineering association, IDA. The purpose of the presentation was to demonstrate the use of Multivariate Statistical Process Control (MSPC). The use case for this demonstration was taken from the industrial partner. The data used originates from a test of product stability, where injected moulded elements from one mould was sampled once each day for 24 days. Both the Product Quality and Process Monitoring application is framed in the illustration of the data utilisation framework to the right. This illustrates that rapid monitoring of element metrology could be used for process monitoring. The work can be linked to research question 1, by monitoring quality data to detect trends that potentially could lead to scraped elements and thereby lost productivity.



Research question 1 (RQ1)

How to best utilise data from injection moulding machines to increase the productivity of producing injection moulded elements?

D.1 Summary

The purpose of the conducted work is to demonstrate that rapid sampling of produced products can be used for quality and process monitoring. The scope was limited to metrology measures. The collected data consists of 24 samples collected over 24 days, thereby covering the usual gap between quality inspections (metrology samples are typically collected every third week). PCA was used for an initial exploration of the collected data. From this, it was evident that two of the collected samples had a changed correlation pattern compared to the rest of the samples. A root cause was found by inspecting the production log. Here it was found that there had been some issues with the cooling water supply in the period where the two extreme cycles were produced. The element dimensions driving this behaviour are linked to element width and caused by warpage of the element. Warpage is fault often caused by uneven cooling of the elements.

Based on this initial investigation, it was found that variation in element quality could be detected by using PCA. Groupings were identified and it was possible to link assignable causes to the observed groupings. One of the groupings seen was caused by mechanical changes in the mould (shift in alignment) and the other by a change in process conditions (change in mould temperature). All the observed variations are well within the specifications, demonstrating that PCA is sensitive enough to detect quality variations. That systematical variation was observed between the routine quality inspections indicates a potential gain by introducing a more frequent inspection. It is however not realistic to increase the inspection interval because of resources (both measuring equipment and human resources). The only way to implement an increased quality inspection would be to implement a soft-sensor approach using available data. Alternatively a PCA and MSPC approach could be implemented on quality data collected in the standard quality inspection scheme. In this case, the results would instead reflect mould degradation (as demonstrated in paper F), random process variations or shifts caused by changes in raw material supply.

An alternative use of MSPC-PCA on element quality could be when qualifying materials from a new vendor or a replacement material (e.g. a sustainable material). Here a baseline PCA could be made for element quality using the existing material, and element quality using the new material could be projected onto the model. If element quality for the new material exhibits the same correlations (in-control using MSPC-PCA) as the existing material, it would be approved, else it would have to be investigated what is causing the differences and if adjustments to the moulding process could be used to move the element quality in the right direction.

Achieved results

- Unsupervised MSPC-PCA approach for monitoring product quality changes for products with multiple quality attributes.

- Ability to detect out-of-control quality caused by changes in process conditions.

Contribution

- PCA is used to analyse quality data from products with multiple quality attributes.

- Demonstrating the use of MSPC-PCA to monitor products with multiple quality attributes.

D.2 Content from presentation

The example in the following is based on the same element (small rectangular box, see Figure 9 in paper A) used in paper A and B, where only a subset of metrology measures were used in paper A and B. In this example, a total of 43 metrology measures are used, which correspond to the measures used when approving a new mould and element. The elements are produced in a mould with 48 cavities and collected in 24 days, with one sample per day. During the 24 days, the mould was removed for the moulding machine, maintained and mounted in the machine again (before/after maintenance is represented with the groups in Figure D.2). All measures for all samples are well within specifications. The purpose of the conducted work is to demonstrate the use of MSPC-PCA (see section 4.3.4 for an introduction to MSPC-PCA) on quality measures.

D.2.1 Initial exploration of quality data

Initial data exploration is conducted to get a first impression of the variation in the collected quality data. From the correlation plot in Figure D.1 it can be seen that some of the metrology measures are highly correlated.

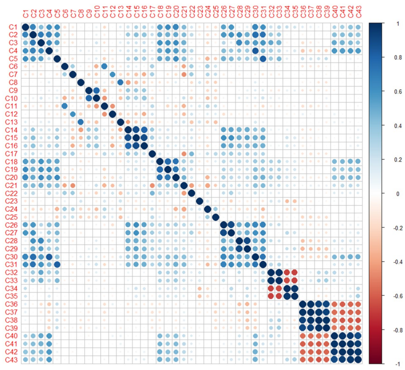


Figure D.1. Correlation among the 43 metrology measures.

One example is the eight measures in the lower right corner, which consists of four measures of the wall thickness on one side of the element and four measures on the opposite side. Measures on each side are highly positively correlated, whereas they are negatively correlated with the four measures on the opposite side (caused by alignment of the two mould halves).

With the number of quality measures and the observed correlation among these, MSPC based on PCA seems to be the appropriate choice for monitoring. An initial exploration using PCA is presented in the following. Figure D.2 shows the score plot of PC1 vs PC2 and PC1 vs PC3 since these combinations represent some interesting groupings. Group 1 represent samples collected before maintenance of the mould, and group 0, after the performed maintenance.

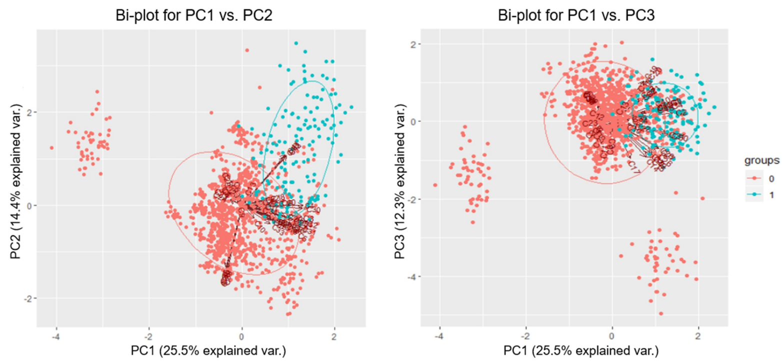


Figure D.2. Score plot of PC1 and PC2 to the left and PC1 and PC3 to the right. The plots are coloured according to before and (1) after maintenance (0).

Looking at the score and loading plot in Figure D.3 it can be seen that two groups of quality measures (C36-C39 and C40-C43) drive the grouping between group 0 and 1 (mainly captured by PC2).

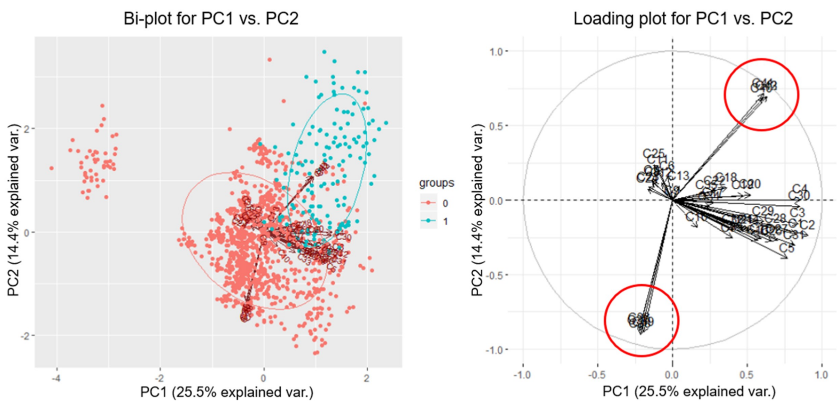


Figure D.3. Score plot of PC1 and PC2 (left) and the corresponding loading plot (right)

These eight measures are wall thickness on the sides of the element (four measures on each side). Inspecting the quality measures, it becomes clear that there is a shift in the wall thickness after maintenance. This is caused by a better alignment of the two mould halves (better centring of the cores in the cavities) after maintenance, making the wall thickness more even on the two sides of the element. This type of quality variation would most likely not be detected by a predictive model based on readily available process data and machine pressure profiles (linked to paper B). It could be interesting to investigate if it could be detected using acoustic emission from the mould.

Looking at the score and loading plot of PC1 and PC3 (Figure D.4), it is seen that two samples are placed outside the joint group of the remaining samples. The variables driving this behaviour are marked with red circles in the loading plot. All

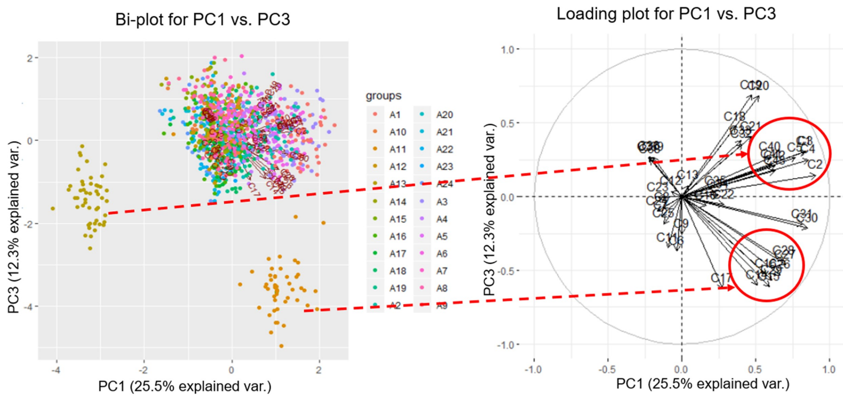


Figure D.4. Score plot of PC1 and PC3 (left) and the corresponding loading plot (right), where metrology measures impacting the grouping have been highlighted with red circles. Points are coloured according to sample number.

these dimensions are linked to element width, and the seen variation is caused by warpage of the elements. Warpage is often caused by uneven cooling of the elements. This was confirmed by inspecting the production log. Here it was found that there had been some issues with the cooling water supply in the period where the two extreme cycles were produced. Part of the cooling channels in the mould had been turned off, resulting in uneven cooling in the mould. This could have been detected by collecting and monitoring the cooling water flow in the mould. This is an option on some of the moulding machines at the industrial partner but not utilised.

D.2.2 MSPC-PCA on quality measures

To simulate a monitoring situation, data is rearranged so that the two extreme samples (in Figure D.4) are placed in phase II together with the three samples collected before these. A PCA model is then calculated on the phase I data (19 samples each with 48 elements), where 9 PCs include 78% of the variation. Control limits are calculated for T^2 and SPE, and the control charts are plotted in Figure D.5.

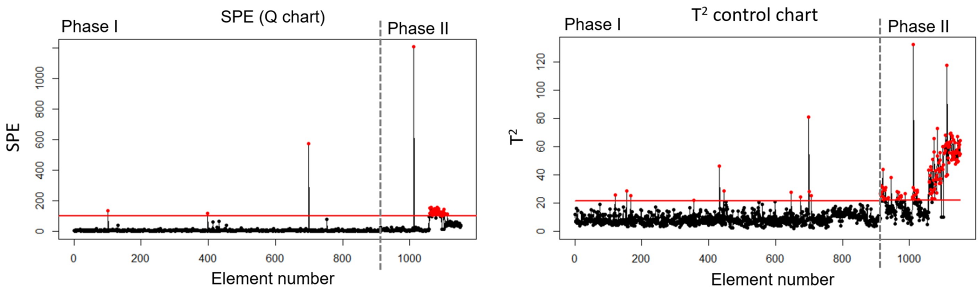


Figure D.5. Control chart for T^2 and SPE. Phase II data starts at element number 913.

One extreme element in phase I is present in the T^2 and SPE chart. This should be inspected, and the element should potentially be removed from the phase I model. Besides this, phase I is evaluated to be in control. Looking at the T^2 chart for phase II, the main part of the elements in the first three samples are in control (all are in control looking at the SPE chart). All elements in the last two samples are above the control limit in the T^2 chart and therefore evaluated as out of control. This demonstrates that the out-of-control situation caused by uneven cooling of the mould could be detected using MSPC-PCA.

D.2.3 Further use of MSPC-PCA on quality measures

In the above, the focus was on using MSPC-PCA to monitor a continuous collection of quality samples. Figure D.6 presents two alternative examples of using a PCA-based approach to utilise element metrology data.

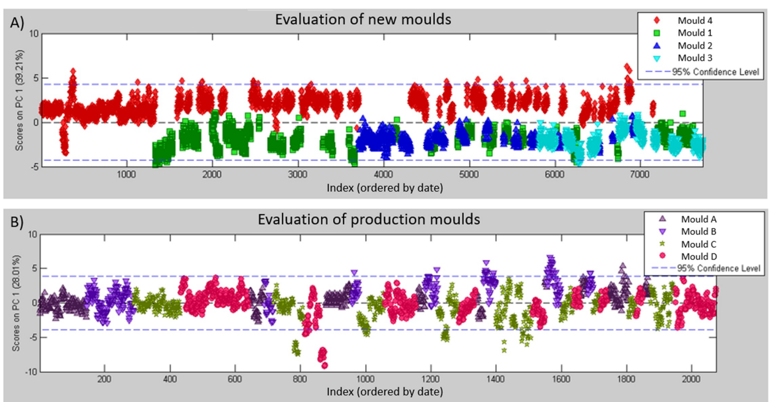


Figure D.6. Time-series plot of PC 1. A) demonstrates the use of PCA for evaluation of new moulds and B) the use of PCA to detect changes in quality samples over time.

As described in section 3.5, all new moulds at the industrial partner are subjected to a running-in process where the mould card is defined. The mould card is created so that all metrology measures individually are within specifications, hence neglecting the correlation between the metrology measures. Figure D.6 A) illustrates the use of PCA for comparison of similar moulds (producing the same element shape) in the initial mould qualification.

Mould 1, 2 and 3 represents three moulds that, after running-in, have been running in the production without any challenges, whereas mould 4 have had many issues with elements dimensions and functionalities. The data presented in D.6 A) is the score value of PC1 based on the metrology data generated as part of the running-in of the moulds. It is clear that moulds 1, 2 and 3 are overlapping, indicating high similarity, whereas mould 4 deviates from the remaining three moulds. Had this analysis been performed as part of running, it would have been evident that the quality produced with mould 4 exhibits a different correlation than historical moulds. This could have been investigated and potentially corrected before the mould was released to production. The same approach could be used to qualify materials from a new vendor or a replacement material (e.g. a sustainable material). Here a baseline PCA could have been made for element quality using the existing material, and element quality using the new material could be projected onto the model. Using an MSPC-PCA approach to get confidence limits for the comparison could be relevant.

Example B) in Figure D.6 uses PCA for detecting a change in element quality over time. These changes are mainly related to mould degradation and therefore closely related to the work conducted in paper F. Mould B is an example of a mould changing over time, where the PC1 value increases over time.