École Doctorale IAEM Lorraine

Scuola di Dottorato di Ricerca - DMMM
XXXII Ciclo

CONTRIBUTION À LA FORMALISATION D'INVARIANTS DE MODÉLISATION DE SYSTÈMES CYBER-PHYSIQUES, DIRIGÉS PAR LES DONNÉES

THÈSE EN COTUTELLE

Présentée et soutenue publiquement le 10 Juin 2020

Pour l'obtention du titre de

Doctorat de l'Université de Lorraine
En Automatique, Traitement du signal et des Images, et Génie Informatique

Dottore di Ricerca del Politecnico di Bari
In Ingegneria Meccanica, Matematica e Management

par

Concetta SEMERARO

Composition du jury

Président : Prof. Mariagrazia Dotoli, Professeur à l'École polytechnique de Bari, Italie
Rapporteurs :
Prof. Vincent Cheutet Professeur à INSA de Lyon, France
Prof. Gino Dini, Professeur à l'Université de Pise, Italie

Examinateurs :
Prof. Michele Dassisti, Professeur à l'École polytechnique de Bari, Italie (*Directeur de thèse*)
Prof. Hervé Panetto, Professeur à l'Université de Lorraine, France (*Directeur de thèse*)
Dr. Mario Lezoche, Maître de Conférences à l'Université de Lorraine, France (Co-*directeur de thèse*)
Prof. Mariagrazia Dotoli, Professeur à l'École polytechnique de Bari, Italie

Invités :
CFO Stefano Cafagna, CFO de Master Italy srl, Conversano, Italie
CEO Michele Loperfido, Directeur général de Master Italy srl, Conversano, Italie
Prof. Hind Bril-El Haouzi, Professeur à l'Université de Lorraine, France

DMMM Dipartimento di Meccanica, Matematica e Management (DMMM) – Politecnico di Bari
Centre de Recherche en Automatique de Nancy – UMR7039, Université de Lorraine, CNRS

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my supervisors Prof. Michele DASSISTI for his motivation, enthusiasm, humanity, generosity and foresight. Thank you for believing in me, always!
I would like to thank my supervisor Prof. Hervé PANETTO for the continuous support of my Ph.D. study and research, for his patience and immense experience. I would like to say thank you for giving me the co-mentorship opportunity.
I would like to thank my co-supervisor Mario LEZOCHE for the support and the help during my Ph.D. study.
I am very much grateful to the company Master Italy s.r.l for having funded my Ph.D. study. Without such a grant, this work would not have been realised. I would also like to thank the board of Master group, Michele LOPERFIDO, Maria Luisa LACATENA, Annacarla LOPERFIDO and Celeste VITTE.
I would like to thank my co-supervisor CFO Stefano CAFAGNA. Without your passionate participation and helping, the application of my studies in Master could not have been successfully conducted.
I am very thankful to Giuseppe LAPRESENTAZIONE for the transferred knowledge of die casting aluminium process.
I would like to thank the rest of my thesis committee. I would like to thank Prof. Vincent CHEUTET and Prof. Gini DINI, who agreed to be the official reviewers of this thesis, and the other members of the jury, Prof. Mariagrazia DOTOLI, Prof. Hind BRIL-EL HAOUZI, CEO Michele LOPERFIDO and CFO Stefano CAFAGNA.
I would like to thank my colleagues at Politecnico di Bari for the Ph.D. experience shared together and my colleagues at CRAN for the beautiful year lived in Nancy.
In addition, I would like to thank my best friends: Daniela, Mariabianca, Mary, Meri, Alessia, Tea, Ornella and Nicole for the support during my Ph.D. study.
I am very grateful to Saad for the continuous encouragement.
Finally, I express my profound love to my parents Francesco and Consiglia and to my sister Ileana. My achievements would not have been possible without you. Thank you!

PREFACE

This PhD thesis is the outcome of the co-mentorship between the Polytechnical University of Bari, University of Lorraine and the Master Italy s.r.l company.
Master Italy s.r.l has funded the PhD thesis. The involvement of Master Italy s.r.l, the project's industrial partner, not only helps reinforcing the comprehensive understanding of the context of the enterprises and their daily issues, but it also assists in building a realistic scenario for the validation of the PhD thesis' results.

CONTENTS

GLOSSARY

ACRONYMS AND VARIABLES	DEFINITION
C_1	Course of the plunger in the first phase
C_2	Course of the plunger in the second phase
CC	Multiplied course
CPPS	Cyber Physical Production System
CPS	Cyber Physical System
DT	Digital Twin
$c_{l,al,i}$	Latent heat of Aluminium
$c_{p,al,i}$	Specific heat of Aluminium
EA	Exergy analysis
e_i^{ph}	Physical exergy of material
Ex_{loss}	Exergy loss
$Ex_{M,in}$	Input exergy mass
$Ex_{M,in}$	Input exergy mass
$Ex_{M,out}$	Output exergy
$Ex_{Q,in}$	Input heat exergy
$Ex_{Q,out}$	Output heat exergy
$Ex_{W,in}$	Input exergy work
$Ex_{W,out}$	Output exergy work
$F_{inj,i}$	Injection force
FC	Clamping force
h_0	Specific enthalpy of environment
h_{al}	Specific enthalpy of Aluminium
IoS	Internet of Service
IoT	Internet of Things
LCA	Life cycle analysis
P_0	Environmental pressure
P_{a1}	Accumulator pressure in the first stage
P_{a2}	Accumulator pressure in the second stage
P_{aM}	Accumulator pressure in the multiplied stage
PF	Final pressure
PM	Multiplied pressure
PS	Specific pressure
$Q_{c,i}$	Energy for casting
$Q_{f,i}$	Energy for melting
$Q_{l,i}$	Energy for the transformation
s_0	Specific entropy of environment
s_{al}	Specific entropy of Aluminium
S_{inj}	Injection surface
SF	Smart Factory
SM	Thickness of the product
T_0	Environmental temperature
T_1	Time of injection in the first phase
T_2	Time of injection in the first phase
T_f	Melting temperature
T_{mo}	Mould temperature
T_s	Temperature of the ingots at the entry of the furnace
$T_{al,i}$	Injection temperature of Aluminium
T_l	Temperature loss during the injection stage

$\mathbf{TC}$	Cycle time
$\mathbf{V_1}$	Speed of the plunger in the first phase
$\mathbf{V_2}$	Speed of the plunger in the second phase
$\mathbf{V_{al,i}}$	Volume of Aluminium
$\mathbf{V_l}$	Volume of Aluminium Loss during the injection stage
$\boldsymbol{\rho_{al}}$	Density of Aluminium

GENERAL INTRODUCTION

GENERAL INTRODUCTION

1. RESEARCH CONTEXT AND MOTIVATION

Manufacturing enterprises are facing the need to align themselves to the new challenges of the market demand: one possible path is to endeavour the new information technologies. These are facing with an array of Industry of the Future challenges: “digital requirements” thus they need to be accurately assessed by sound analysis and deep understanding of the operational and technological criticalities in the manufacturing operations.

The Smart Factory (Zuehlke 2010) paradigm has born with the so-called "fourth industrial revolution" in manufacturing. It can be synthesized by integrating networks of physical components with control software that contribute to the “smartness” of the overall system, for better managing manufacturing processes. Smart systems typically are well to support decision making based on sound data. This is why smart systems include a variety of technological components, including sensors for signal acquisition, communication units for data transmission between components, control and management units for supporting decision making, and actuators for performing appropriate actions: the coordinated network of sensors is here called the sensing system. The collaborating computational capabilities are embedded in physical entities that are named Cyber Physical Systems (CPS). The Cyber Physical Systems (CPSs) have amplified the possibility to sense the world through a network of connected devices. The combination of intelligent systems and sensing systems forms a large-scale distributed cyber-physical system, that has enormous potentialities to bring smartness into many application areas. CPSs, on the other hand, suffer from a lack of modelling techniques for taking into account not only their technological parameters but also their high degree of information and functional inter-relationships.

Another concept associated with the cyber–physical integration is the Digital Twin (DT) (Koulamas and Kalogeras 2018). The DT refers to a “virtual” image of a system (a twin) synchronized with the real operating scenario, providing as much useful information as possible from this latter, to allow the human decision maker drawing sound decisions. This “digitalization” of the reality is an approach that enables the virtual replication of the factory to forecast and control real time physical processes. It may allow to connect the entire value chain (Ghobakhloo 2018) by merging sensor data acquired from the physical world into virtual or simulation-based models. Predicting the system behaviour is thus a matter of running these realistic and “up-to-date” models.

As the complexity of these systems continues growing, the challenge of developing integrated intelligent and sensing systems has surpassed the design complexity of their individual components. The main problem in the development of smart and sensing systems lies in the complexity of integrating and managing different components and technologies. A possible solution to this problem is to formalize the shared knowledge and define a modelling approach that helps to design systems regardless of the context. Shared knowledge representation is a branch of artificial intelligence that studies the way human reasoning occurs and defines symbols or languages (Roschelle and Teasley 1995) for this purpose.

In this context, the thesis aims firstly at providing an up-to-date picture of the state-of-the-art of main features and challenges about CPS and DT design and use, focussing on the different application domains and their related technologies. To answer the main questions on how a Digital Twin must be designed and implemented, the thesis funds its scientific basis on information, principles and hints derived from a systematic scientific literature review. The outcome is then a multi-perspective picture of the Digital Twin

paradigm, forming a framework emerging from scientific literature as well as the direct on-the-field experience of the authors.
The thesis aims to identify an approach to formalize data-driven invariant modelling constructs for improving the smartness of manufacturing processes and products, involving networked components. The idea behind data-driven invariant modelling constructs is to permit the re-use of predefined functional patterns for building digital models based on the specific application. The approach makes shared knowledge more easily reusable and it is the basis of some standardization efforts. The use also of Multi Relational Data Mining techniques, in the specific case of Relational Concept Analysis (Valtchev, Missaoui, and Godin 2004), allows the extraction of tacit knowledge embedded in the (big) data coming from the analysed processes.
The thesis proposes a series of modelling patterns (data-driven invariant modelling constructs) for the digital transformation of industrial production systems. A prototype for the analysis of a real industrial process on a production line at Master Italy s.r.l has been developed to experiment our knowledge extraction approach. The resulting tool can exploit existing knowledge and information from real systems to identify problems and to propose potential improvements.

2. RESEARCH QUESTIONS AND THE CONTRIBUTION OF THIS THESIS

The thesis defines the following research questions:

RQ1: What are the existing research works on digital twin: 'Which technologies need to be adopted or new ones explored?' 'How to design a Digital Twin?' 'How to implement a Digital Twin?'

One of the first steps in our research is to investigate and to define the digital twin paradigm in the context of Industry of the Future. We conducted a systematic literature review to identify the context, the applications, the functions in product life cycle, the possible architectures, and the components of a DT. Next, we intend to combine the different aspects of a Digital Twin and some related issues for developing our proposed digital twin paradigm. Hence, we need to link the different approaches related to the digital twin modelling for developing a common approach to detect data-driven invariants modelling constructs.

RQ2: How to develop data-driven invariant modelling constructs?

To answer to this question, we conducted a second literature review on model-based, data-driven and hybrid approaches to define the contribution positions. The idea is to identify and to formalize invariant modelling based on data analysis. The idea is to use, and especially re-use, predefined data-driven constructs for building digital models for different applications.

3. THE STRUCTURE OF THE MANUSCRIPT

The Chapter 1 gives an overview of the research context. First, we explore the basic definitions related to the concept of smart factory, the different technologies of the Industry of the Future and their application in SMEs. Second, we study the state of art of digital twin design and applications. Finally, research gaps are identified to demonstrate why it is necessary to develop an invariant approach based on data-driven model constructs.

In the **Chapter 2**, the comparative analysis aims at showing the evolution of model-based and data-driven approaches over the years for presenting the contribution positioning of the Ph.D. thesis. The idea of the thesis is to detect and to formalize data-driven invariant modelling constructs.

The Chapter 3 defines the approach developed to extract invariant modelling construct based on data analysis.

The Chapter 4 has the objective to prove the quality of the contribution through a real case study. Moreover, a Digital Twin prototype has been developed to demonstrate and to apply our data-driven invariant modelling constructs on a specific manufacturing process to build a digital twin.

Finally, a **Conclusion** discusses the research findings and presents some research perspectives derived from this work.

CHAPTER 1

DIGITAL TWIN PARADIGM: A SYSTEMATIC LITERATURE REVIEW

CHAPTER 1 – DIGITAL TWIN: A SYSTEMATIC LITERATURE REVIEW

INTRODUCTION

In this chapter, we outline the main concepts related to the Digital Twin. The present chapter concerns how to effectively design a digital twin to support the digital transformation of manufacturing enterprises domains. First, an overall dissertation to comprehend the digital transformation in SMEs, is presented in section 1.1. In section 1.2 a systematic literature review on Digital Twin is conducted to define context and application in 1.2.1, the implementations along product lifecycle 1.2.2, the functions in 1.2.3, the architectures in 1.2.4 and the components in 1.2.5. Therefore, an overview of the research gaps is presented in Section 1.3 to define the thesis objectives in 1.4.

1.1 DIGITAL TRANSFORMATION IN SMEs

In the past, due to the lack of information technologies, the physical space played the main role on controlling the production in shop-floors, leading to low efficiency, accuracy and transparency.
Until the 20th century, technologies such as computers, simulation tools, Internet, and wireless networks introduced a parallel virtual space to virtualize physical assets and to enable the cooperation with assets remotely. This has provided a possibility to conduct plans and operations more efficiently and effectively (Tao and Zhang 2017).
Nowadays, with the developments of new generation information technologies (New IT), such as the cloud computing, the Internet of Things (IoT), big data, and artificial intelligence (AI), the roles of the virtual space are becoming increasingly important. Therefore, the integration, interaction and cooperation between the physical and virtual spaces are more active than ever before and it will be an inevitable trend. This will create a new potential to improve the current situations and technologies in the fields of design, manufacturing, service, etc (Büchi, Cugno, and Castagnoli 2020).
Various countries are converging on this trend as the next industrial revolution and have proposed related national strategies, such as the "Industrie 4.0" in Germany; the "Advanced Manufacturing" or "Smart Manufacturing" in the United States; the "e-Factory" in Japan; the "Intelligent Manufacturing" in China; the "Industry of the Future" in France, the "Fabrica Intelligente" in Italy and more generally "The Factory of Future" in Europe (Drath and Horch 2014).
The term Industry 4.0 (I4.0) is referred to the fourth industrial revolution (Harris 1990). The first industrial revolution, which occurred at the end of seventeenth century, was driven by the advent of steam engines, waterpower and mechanization. The second industrial revolution was driven by the assembly lines, pioneered by Henry Ford who first officialised mass production almost a century ago. The third industrial revolution, which occurred in the 1970s, was driven using computer and automation in manufacturing processes. Industry 4.0 was announced at the Hannover Messe in 2011 and it is defined as: *"Networks of manufacturing resources (manufacturing machinery, robots, conveyor and warehousing systems and production facilities) that are autonomous, capable of controlling themselves in response to different situations, self-configuring, knowledge-based, sensor-equipped and spatially dispersed and that also incorporate the relevant planning and management systems"*(Kagermann et al. 2013).

In the United States, the Smart Manufacturing Leadership Coalition has worked on a new manufacturing paradigm, called **Advanced Manufacturing or Smart Manufacturing** (Yao et al. 2019) which is based on the integration of advanced new technologies such as IoT into the manufacturing area to improve produced goods and manufacturing processes.

The National Institute of Standards and Technology (NIST) defines the smart manufacturing as: *"Fully-integrated, collaborative manufacturing systems that respond in real time to meet changing demands and conditions in the factory, in the supply network, and in customer needs"*.

The research and development programs related to manufacturing in the U.S. focus on key technology, including IoT, big data, data analytics, CPS, system integration, sustainable manufacturing, and additive manufacturing to respond to the innovative manufacturing environment (Kang et al. 2016).

The **e-Factory** concept from Japan is achieving an advanced use of the industrial internet regarding both manufacturing control and data analytics, with the aim of effecting an optimization of productivity and energy conservation.

China is pushing forward its **Intelligent Manufacturing** initiative, which will drive manufacturing business by merging ICT, automation technology and manufacturing technology. The core of the idea behind Intelligent Manufacturing is to gain information from a ubiquitous measurement of sensor data in order to achieve automatic real-time processing as well as intelligent optimization decision-making.

According to the European Commission's (EC) vision (Horizon 2020 program), the **Factories of the Future (FoF)** is a new model of production systems and it should address transformable, networked and learning factories, depending on several drivers such as high-performance, extreme customisation, environmental-friendliness, superior efficiency of resources, eminent human potential and significant knowledge creation (EC 2013). FoF embrace ICT-based production systems with high quality manufacturing technologies and intelligent capabilities aimed to optimise the performances with a high degree of autonomy and adaptability (Jardim-Goncalves, Romero, and Grilo 2017), toward the creation of the collaborative factory of the future (CFoF) (Moghaddam and Nof 2017).

Although the strategies are proposed under different environments, their common objective is to capture the opportunity brought by the seamless integration and fusion of the physical and virtual spaces. Computing, communication and controlling capabilities need to be embedded in all types of objects in the physical environment. The systems that bridge the cyber-world with the physical world are referred to as cyber-physical systems (Rajkumar et al. 2010).

Cyber-physical system (CPS) (https://pages.nist.gov/cpspwg/) aims at embedding computing, communication and controlling capabilities (3C) into physical assets to converge the physical space with the virtual space (Monostori et al. 2016). In this way, the physical space can be integrated, monitored, controlled and coordinate by the virtual space in real-time and vice versa (Blume et al. 2014).

Internet of Thing (IoT) is an important driver for CPS. IoT enables the direct integration of the physical and cyber space to form a hybrid communication network where elements interact with each other in real-time. With the IoT data in all aspects of manufacturing can be collected and be converged continuously from the physical to the cyber (Zhu et al. 2010).

Then through the Internet of Services (IoS), internal and cross organizational services are offered and utilized by participants of the value chain. Internet of Services (IoS) is concerned with the systematically use of the internet for developing service business model (Becker et al., 2014).

The implementation of the IoT, IoS and CPS technologies in manufacturing systems enable the development of flexible manufacturing systems. Flexible manufacturing systems satisfy rapid changes in production volumes and customization. This type of manufacturing system is known as a smart factory.

Smart factory is defined as an intelligent production system which utilizes the integration of manufacturing and services. It integrates communication process, computing process, and control process to meet the industrial demands (Chen et al. 2017). CPS, IoT, IoS and Smart Factory are the four main pillars of the Industry of the Future paradigm and the related research and technological domains.

The Industry of the Future paradigm aims to construct a new way of organizing production by setting up smart factories (Strozzi et al. 2017), where everything is performed through the interaction and the integration of networked manufacturing systems, value networks and value chain (Osterrieder, Budde, and Friedli 2019) defined respectively as vertical, horizontal and end-to-end integration (Posada et al. 2015).

The Vertical integration denotes the integration of various Information Technology systems at different hierarchical levels (e.g. the actuator and sensor, control, production management, manufacturing and execution and corporate planning levels) (Pessl, Sorko, and Mayer 2017).

The Horizontal integration is related to the integration of the various IT systems used in the different stages of the manufacturing and business planning processes and involves an exchange of materials, energy and information both within a company (e.g. inbound logistics, production, outbound logistics, marketing) and between several different companies (value networks) (Weking et al. 2018).

The End-to-end digital integration refers to a holistic digital engineering view and proposes to close the gap between product design and manufacturing and the customer (Product Lifecycle Management) (Tsai et al. 2001), from the product design and development, through production planning, production engineering, production and associated services.

The potentialities of I4.0 paradigm are to ensure a better flexibility and scalability of manufacturing systems through information technologies and industrial automation (Dassisti and De Nicolò 2012), (Brettel et al. 2014). The design principles of the Industry of the Future paradigm, employed for design smart factories or for upgrading existing traditional factories to be smart, are here summarised as follows (Hermann, Pentek, and Otto 2016):

- Modularity is concerned with shifting from mass production models, rigid and inflexible systems toward an agile system (Mabkhot et al. 2018). Agile implementation processes provide autonomy and flexibility in smart factory implementation (Sjödin et al. 2018). Configure modular technology provides opportunities for continuous innovation to manage the complexity of digital systems (Sjödin et al. 2018).The factory will move from integral (with high production volume and low flexibility) to modular (with middle to high production volume and flexibility) (Yin, Stecke, and Li 2018) (Chen et al. 2017).
- Interoperability is the ability of all components of the smart factory including products and human resources, to share business information between manufacturing enterprises and customers (Panetto et al. 2016).
- Decentralization enables different components of the smart factory modules, material handling, products, etc.) to work independently and make decisions autonomously(Mabkhot et al. 2018). System elements (modules, material handling, products, etc.) will make decisions on their own, unsubordinated to a control unit. A decision will be made autonomously in real time without violating the overall organizational goal.
- Virtualization enables the virtual replication of the factory to monitor and simulate real time physical processes. It allows to create a digital twin of the entire value chain (Ghobakhloo 2018) by merging sensor data acquired from the physical world into virtual or simulation-based models.

- Service orientation refers to the concepts of Manufacturing-as-a-Service (MaaS) and Product-as-a-Service (PaaS). MaaS business model is the collective use of a networked manufacturing infrastructure to produce products and services (Mabkhot et al. 2018). In the PaaS business model, products are delivered as a service or virtualized experience. Factory-as-a-Service (FaaS) is an open manufacturing service that promotes personalized production and distributed manufacturing (Park et al. 2019).
- Real-time capability is the ability of the system to respond to changes on time, such as changes in customer requirements or the status of the internal production system (e.g., malfunctions and failures) (Ghobakhloo 2018).

The chance to implement “smartness” into the factory is made possible by the widespread diffusion of Information and Communication Technologies (ICT) (Nardello, Madsen, and Møller 2017). The key enabling technologies 4.0 are here below summarised (Rüßmann et al. 2015):

- Sensor and Actuator: the sensor is the most basic technology for collecting and controlling data real time (Bibby and Dehe 2018) and for .
- Big data analytics refer to a new generation of technologies and architectures that enable organizations to economically extract value through discovering, capturing and analysing very large volumes of a wide variety of data (LaValle et al., 2011). Big data analytics enables contemporary organizations to better gain value from the massive amounts of information they already have and identify what is likely to happen next and what actions should be taken to achieve the optimal results (L. Wang and Wang 2016).
- Blockchain enables transparent, secure, trustworthy and swift public or private solutions (Underwood, 2016) (A. G. Frank, Dalenogare, and Ayala 2019).
- Augmented reality (AR) has been regarded as a highly promising technology that allows for visualization of computer graphics placed in the real environment (Alcácer and Cruz-Machado 2019). AR is commonly used in the description, planning and real-time operation monitoring, fault diagnostic and recovery, and training related to industrial products and processes (Doshi et al., 2017; Khan et al., 2011).
- Virtual Reality (VR) provides an immersive, interactive, computer-mediated experience in which a person perceives an artificial environment by means of human-computer interaction equipment (Yew et al., 2016).
- Cloud Computing allows access to software and data storage in the cloud representation of the internet (Yao, Jin, and Zhang 2015).
- Collaborative Robots are machines developed for the purpose of performing specific tasks autonomously or by means of remote-control commands (L. D. Xu, Xu, and Li 2018).
- Additive manufacturing or 3D printing denotes the manufacturing technique in which parts are built by melting thin layers of powder and adding one layer of material, either plastic or metal, on top of another, based on the geometry suggested by Computer-Aided Design (CAD) (Esmaeilian et al., 2016) (Silva et al. 2019).
- Simulation and modelling techniques aim to simplify the design, realization and tests of manufacturing operations (Kocian et al., 2012). In the smart factories, simulation and modelling will be necessary for using real-time data to mirror the physical world in a virtual model, which can include machines, products and humans (Rüßmann et al., 2015).

The advances and the development of new technologies are largely responsible for the emerging popularity of the Industry of the Future concept and its potential use by SMEs.

Provided the need to encourage and support SMEs to adopt the I4.0 paradigm, it is necessary to clearly analyse the complexity to adopt I4.0 technologies and contemporarily its benefits for these kinds of companies. The true problems, for the smart factories' implementation, are related to the nature of the SMEs. These are often characterized by poorly formalized processes, by independent and/or legacy hardware and software systems and by smaller economical capabilities with respect to large companies. Often SMEs lack internal IT competences and the necessary technological knowledge. The critical success factor/challenge, for the successful implementation of Industry of the Future in SMEs (Sony and Naik 2019), is to understand the rationale behind these new technologies, and to define how, when and where these could be applied and why these are truly beneficial to the manufacturing industry.

Small and medium sized enterprises (SMEs) have the difficulty to be highly skilled in applications and technologies of Industry of the Future. This is caused by the fact that those companies usually do not have the possibility to invest in emerging technologies like an early adaptor, in order not to lose money by focusing on the wrong technologies (Faller and Feldmüller 2015).

Descending from these, a set of requirements were traced for the implementation of the I4.0 paradigm in SMEs. These needs were summarised into three main I4.0 solution requirements to meet the SMEs requests, and thus to promote their adoption (Dassisti, Panetto, et al. 2017):

- Minimal invasiveness: I4.0 solutions must rest on (and not replace) the existing systems, hardware and software (ERP, MES, SCADA, etc.).
- Turnkey: I4.0 solutions needs a minimal intervention of the end user at changing the use scenarios, i.e. They must embed the necessary knowledge for the different application classes.
- Extensibility: I4.0 solutions must be flexible for the subsequent interventions, so to support a gradual approach; i.e. They must ensure the possibility to reutilize all the components if we want to scale up the overall system.

The following performance main indicators measure the impact of investing in new technologies on SMEs: costs, quality, flexibility, productivity (Raymond 2005). Production systems are controlled using manufacturing execution systems (MES) (Kletti 2007). A MES is a technological solution that transforms data collected on the shop-floor into information to monitor the current conditions of the production system in real time. These technological advancements in sensing and connectivity have enabled SMEs to aggregate data from various measurement points (J. Lee et al. 2013) to achieve monitoring, control and optimization (Moeuf et al. 2018).

SME's need to become quickly suitable to the new market conditions. This has been proved to be highly dependent on the ability to develop I4.0 competences (Adolph, Tisch, and Metternich 2014). To develop employees' competencies, traditional teaching methods show limited effects (Cachay et al. 2012). The new factory needs the development of problem-solving skills (Schuh et al. 2015) and less physical strength.

The challenge is to facilitate the learning of new tasks and skills (Ruppert et al. 2018). Human actors in a future production scenario will need specific competencies to cope with the new challenges regarding technological and organizational developments (Nikolakis et al. 2018). Industry of the Future will lead to an increased automation of routine tasks which implies that workers will have to focus on creative, innovative and communicative tasks (Romero et al. 2016). It is necessary to stop production so that production never stops. It means equipping each machine or assembly lines with a system that support

every worker to be able to stop the production process at the first sign of some anomalous condition (Dalle Mura and Dini 2019).
To pursue this objective, learning factories are being developed to support education and training to let the industrial participants and employees discover Industry of the Future principles and technologies and apply them directly on problems in a real production (Dassisti and Semeraro 2018). The learning factory offers to the employees the possibility to analyse and to solve production problems and in this way, it is possible to:

- Collect and integrate internal knowledge
- Do some field experience and innovate (changing the status quo)
- Integrate external knowledge

A learning factory is a real learning environment which allows a balanced relationship in conveying specialized theoretical and analytical knowledge as well as hands-on experience (Seitz and Nyhuis 2015).
A learning factory is an organization dedicated to:

- Knowledge creation
- Knowledge collection
- Knowledge management

embedded in physical equipment and processes and embodied in people (Leonard-Barton 2000).
The main goals of learning factories in I4.0 are either technological and organizational innovation (if used for research), or an effective competency development (if used for education and training) (Abele et al. 2015) (Ogorodnyk, Granheim, and Holtskog 2016). Universality, mobility, modularity, scalability and compatibility were identified as the first order parameters for ideal classification of a learning factory (Wagner et al. 2012).
During the last years learning factories have already been used during several practical trainings on different topics within production processes (e.g. lean management and resource efficiency) (Erol et al. 2016). Vienna University has decided to build the Industry of the Future Pilot Factory. It will serve both as a research, a teaching and training platform regarding the future production (Baena et al. 2017). Learning factories can clearly have a positive influence on students' performances in comparison with traditional teaching. Cachay et al. in (Cachay et al. 2012) examined how the learning factory concept impacted students' performance in application. The EAFIT University defines a model to develop the learning factory 4.0 into an academic context (Simons, Abé, and Neser 2017). Many universities look at an existing production line as a starting point in order to implement a learning factory : Aalborg University (Nardello, Madsen, and Møller 2017), University of Applied Sciences Darmstadt (Küsters, Pra\s s, and Gloy 2017), RWTH Aachen University (Grä\s sler, Pöhler, and Pottebaum 2016). This are equipped with machines, materials and tools established to realize real physical products to support research projects about industry 4.0.
The IFA Learning Factory at Hannover University (Abele et al. 2015) defines how cyber-physical systems in combination with logistic models can improve planning, controlling and monitoring production. Other learning factories with focus on aspects of Cyber Physical Production Systems (CPPS) in learning factories are for example:

- CPPS learning factory at Paderborn University (Thiede, Juraschek, and Herrmann 2016).
- Learning factory at Braunschweig University (Rentzos et al. 2014).
- Teaching Factory at Patras University (Brenner and Hummel 2016).

The limits of the current learning factory concepts concerning (Tisch and Metternich 2017):

- Resources needed for learning factories.
- Mapping ability of issues in learning factories.
- Scalability of learning factory approaches.
- Mobility of learning factory approaches.
- Effectiveness of learning factories.

To overcome these limitations, the physical and virtual environments should be integrated (Wagner,2017). That offers new possibilities to transfer digitally created solutions to a real system for supporting employees in decision-making process (Terziyan, Gryshko, and Golovianko 2018).

The digital twin (DT) aims at creating high-fidelity virtual models for each physical entity to emulate their states and behaviours with abilities of evaluating, optimizing and predicting (Graessler and Poehler 2017). The physical and virtual spaces interact in a closed loop in sync. The DT can be developed for each phase of the product life cycle absolving different functions.

The digital twin in the design stage helps designers to configure and validate more quickly the future design. The DT can help decision maker to accurately interpret the market demands and the customer preferences. At the manufacturing phase, DT may enable the simulation and thus the decision maker to analyse the interactive behaviours among production factors by collecting data from order, design, purchase, production planning, manufacturing and product usage stage. The DT can help optimizing and evaluating in real time the production planning and the behaviour of the production process.

At the service stage, DT relies on real time state monitoring and virtual operations such as maintenance to predict the remaining life of components or products (H. Lee and Kim 2018).

The digital twin (DT) can be considered as a focused application of the CPS. CPS and DTs involve the integration between the physical and the cyber space. However, DTs focus more on virtual models, while CPS emphasize 3C capabilities (computing, communication and controlling). Sensors, actuators and data are considered as key elements in CPS, while models and data are the core elements in a DT (Tao et al. 2019). Sensors and actuators enable the interaction between the physical and cyber worlds and the data exchanging. Models play a significant role in DT to interpret the behaviour of the physical system (Tao et al. 2019). The next section presents a detailed state of art of digital twin.

1.2 THE DIGITAL TWIN PARADIGM: STATE OF ART OF RESEARCH AND TECHNICAL APPROACHES

The basic idea behind Digital Twin is a **high-fidelity virtual models** of the physical entities having the scope of **replicating and simulating** the states and behaviours of these latter's (Grieves 2014), (S. Jain, Shao, and Shin 2017), (Bao et al. 2018)**,** (Tao, Sui, et al. 2018), (Zhuang, Liu, and Xiong 2018).

The concept of using "twins" dated back to NASA's Apollo program, where two identical space vehicles were built to allow mirroring the conditions of the space vehicle during the mission. Professor Grieves at the University of Michigan firstly put forward the concept of 'Digital Twin' in Product Lifecycle Management (PLM) courses in 2003. At the present, the concept of "digital twin" refers to a holistic digital engineering view of the real system which may be applied to the product design and development, up to production planning, including the production engineering, production and associated services (Product Lifecycle Management).

The Digital Twin is defined, in fact, as a new paradigm in **simulation and modelling** (Rosen et al. 2015) because it extends the use of simulation modelling to all phases of the product **life cycle** (Garetti, Rosa, and

Terzi 2012), (Rodič 2017), (Glaessgen and Stargel 2012), (Ríos et al. 2015), (G. N. Schroeder et al. 2016), (H. Zhang et al. 2017), (Schleich et al. 2017), (Söderberg et al. 2017).
In the context of the Industry of the Future, **virtualization** is the design principle that enables the replication of the physical system into its "digital twin" throughout the entire value chain, by merging data into virtual models (Hermann, Pentek, and Otto 2016). **A closed loop** needs to be realized (Alam and El Saddik 2017), (Graessler and Poehler 2017), (Yun, Park, and Kim 2017), (Autiosalo 2018), (Nikolakis et al. 2018), (Leng et al. 2018) between the physical and virtual worlds through **real time data** connection (Ciavotta et al. 2017), (Schluse, Atorf, and Rossmann 2017) (Stark, Kind, and Neumeyer 2017) by collecting and analysing data (Weber et al. 2017) (H. Lee and Kim 2018), (Haag and Anderl 2018) to respond to changes over the time. Virtualization, in fact, is heavily dependent upon the real-time data management capability of the system (Moreno et al., 2017) (Dassisti, Panetto, et al. 2017).
The digital twin has been also defined as a **Virtual Knowledge System** (E. Tuegel 2012), (J. Lee et al. 2013), (Negri, Fumagalli, and Macchi 2017) (Asimov et al. 2018), (W. Luo et al. 2018), (Z. Liu, Meyendorf, and Mrad 2018), provided a DT aims to achieve transparency (J. Lee et al. 2013) and to overcome the information asymmetry among technology, process and people (Padovano et al. 2018).
The cycle Observe–Orient–Decide–Act (the O-O-D-A loop), was defined by Colonel John Boyd, a military strategist of the United States Air Force (Osinga,2007). The second O, orient, is the most important part of the O-O-D-A loop since it refers to the repository of our genetic heritage, cultural tradition, and previous experiences. The orientation, in the concept of the digital twin, is represented by the amount of data and the fidelity of the virtual model. Its shape the way the digital twin observes, decides and acts.
Different tools and technologies are available for developing high-fidelity virtual models (Schleich et al. 2017). They use different techniques, such as simulation and emulation, including distinct functionalities (McGregor 2002).
Simulation capabilities of a DT are provided by a design of its environment allowing to approximate off-line the behaviour of the real systems to represent how the system reacts (Law, Kelton, and Kelton 2000). It can be thought as a "static feature" of the DT.
On the other hand, the emulation refers to the capability of a DT to be synchronous with the real system, so as it behaves almost similarly to the actual behaviour of the physical system (Ayani, Ganebäck, and Ng 2018). Accordingly, this feature of DT can be thought as a "dynamic feature". An emulation model operates in a hardware-in-the-loop configuration to perform the same work of the physical system. It provides a closer replication with respect to the simulation model (C. G. Lee and Park 2014).
From the simulation point of view, the digital twin represents a new wave in modelling and simulation (Rosen et al. 2015). Simulations tools, such as Plant Simulation from SIEMENS© (Vachálek et al. 2017) is an example of how a DT can help understanding what may happens according to given situations of the real world.
From the emulation point of view, the digital twin duplicates and imitates the physical system in the virtual word. Emulation tools, such Simumatik3D® (Ayani, Ganebäck, and Ng 2018), can thus help to proactively understand what has to be done to react modifications in the real world.
Basically, either the static or the dynamic feature are a what-if analysis, but at different scales in the time domain.
According the literature review, a comprehensive definition of a digital twin proposed here can be:
<<An adaptive model that emulates the behaviour of a physical system getting real time data to update itself along its life cycle. The digital twin replicates the physical system to predict failures and opportunities

for changing, in order to prescribe real time actions for optimizing and/or mitigating unexpected events observing and evaluating the operating system profile>>.

Several studies have been devoted to analysing the DT concept, which results different as the context of application changes (aerospace, manufacturing, city management, etc). At the same time, in each context, the digital twins have their own specificity as function of the life-cycle phase of the product: namely design, manufacturing and service. As a result, each application of DT varies depending on a different perspective and needs accordingly.

As a summary of the wide literature review performed for this thesis, the digital twin paradigm is proposed as composed of the following parts:

1. Context and Application (Where to use DT?)
2. Product Life Cycle (When DT is developed?)
3. Functions (Why DT is used?)
4. Architecture (How to realize and to develop a DT?)
5. Components (What DT is made of?)

The digital twin paradigm is shown in Figure 1 to define the contexts, the phases of the product lifecycle (design, production and service), the possible functions of the Digital Twin for each phase, the architecture and the components. In the following paragraphs, each aspect is detailed.

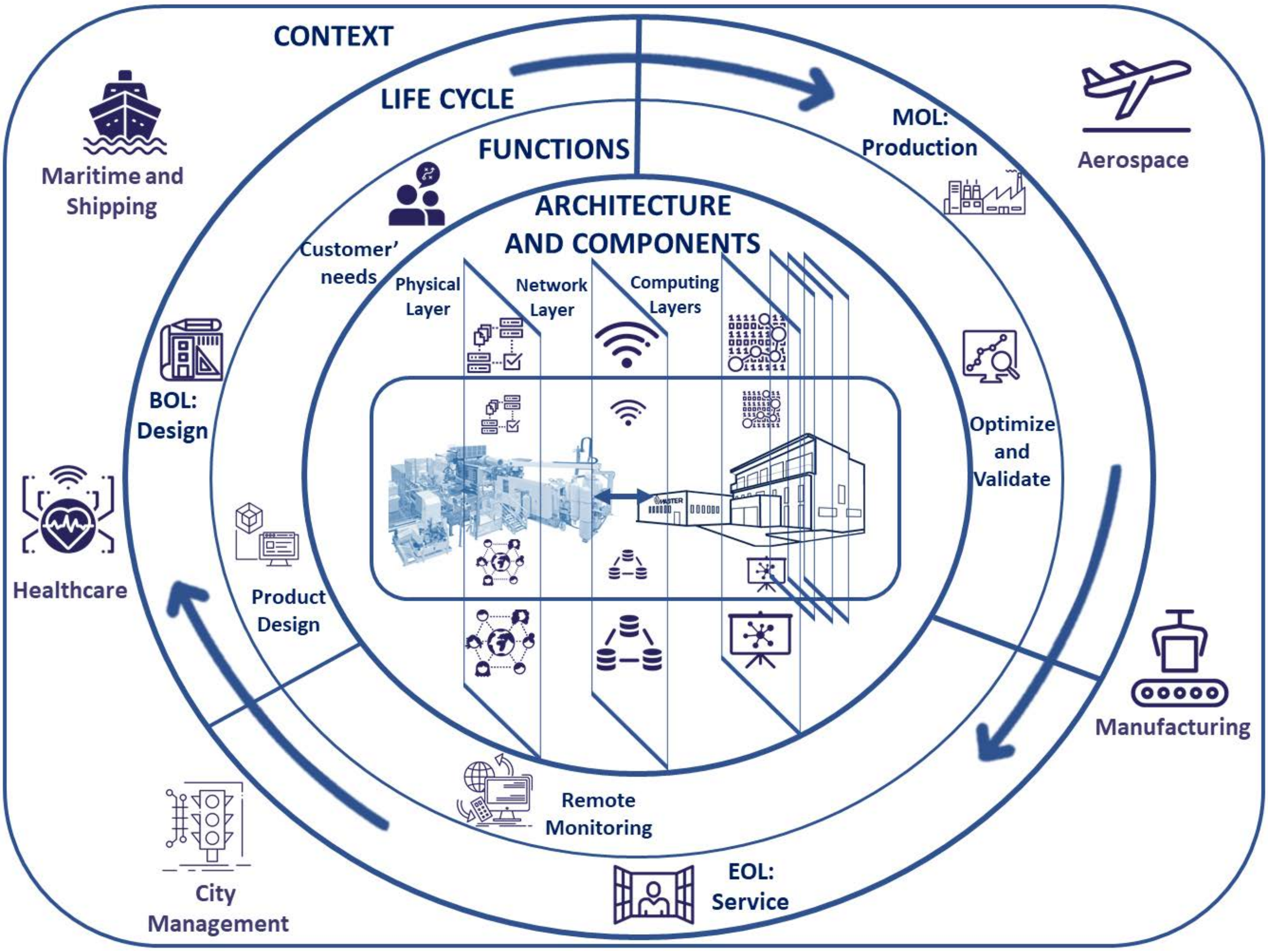


Figure 1: The Digital Twin Paradigm

1.2.1 CONTEXT AND APPLICATION (Where to use a DT?)

The DT has a series of benefits related to the application domain, such as providing visibility of business, accelerating product development, optimizing the operation process, and improving service quality (Barricelli, Casiraghi, and Fogli 2019). Principal application domains of the DT developed so far are: Healthcare; Maritime and Shipping; Manufacturing; City Management; Aerospace. According to each specific domain, the Digital twin can have different purposes, as follows:

- Digital Twins in Healthcare can be used for capturing and visualizing a hospital system in order to create a safe environment and to test the impact of potential changes on system performances. A digital twin can be used to predict the outcome of specific procedures. It can determine the better therapy option for a specific patient. In healthcare, a digital twin recording data of a person, combined with AI models, can provide answers for clinical problems (Bruynseels, Santoni de Sio, and van den Hoven 2018).
- Digital Twins in Maritime and Shipping are used as support for design. The design requires to invest significant amounts of time and money in preparing analytical models to perform simulations. The digital twin allows to visualize all key components, to perform analyses and calculations, and to improve the control of the effects of operation on the ship's structural and functional components (Arrichiello and Gualeni 2019).
- Digital Twin in Manufacturing involves different application based on the stages across the entire lifetime of a product, such as design, production, and maintenance (Dassisti, Giovannini, et al. 2017). The digital twin can support decision maker to predict an upcoming equipment failure, to inform an operator when an asset begins to show signs of non-optimal performance, to improve customer experience (Tao, Cheng, et al. 2018)
- Cities are areas of human settlement, with high population density, complete infrastructure, and buildings. Digital Twins, in City Management, improve the urban environment and people's quality of life. The digital twin can simulate people movements and emergency evacuations, modelling smart building, road traffic, air quality, infrastructure and circular urban economies. The benefits of modelling range from preventive maintenance to operational efficiencies and cost savings. The DTs improve services for citizens, and increase safety and security (Mohammadi and Taylor 2017).
- Aerospace companies have begun utilizing digital twins to accomplish the goal of reducing unplanned downtime for engines and other systems. Digital Twins in Aerospace may allow to receive advance warning and predictions, but also get a plan of actions based on simulated scenarios that consider the weather conditions, the performance of the asset, and several other variables (E. J. Tuegel et al. 2011). With the help of digital twins, it is possible to develop and to implement the predictive maintenance to increase platform operational availability and efficiency, extend its useful life cycle and reduce its life cycle cost. Moreover, DTs are capable of mitigating damage or degradation by activating self-healing mechanisms or by recommending changes in the mission profile to decrease loadings.

Digital Twins have attracted strong interests from industries too: GE Predix Platform, SIEMENS PLM, Microsoft Azure, IBM Watson, PTC Thing Worx, Aveva, SAP Leonardo Platform, Twin Thread, DNV-GL, Dassault 3D Experience, Sight Machine, Oracle Cloud. In their views, the DT finds application in the context previously explained, as shown in Figure 2, and it has a series of benefits, such as accelerating product development, optimizing the manufacturing processes, and improving customer's experience thorough the enhancement of service quality.

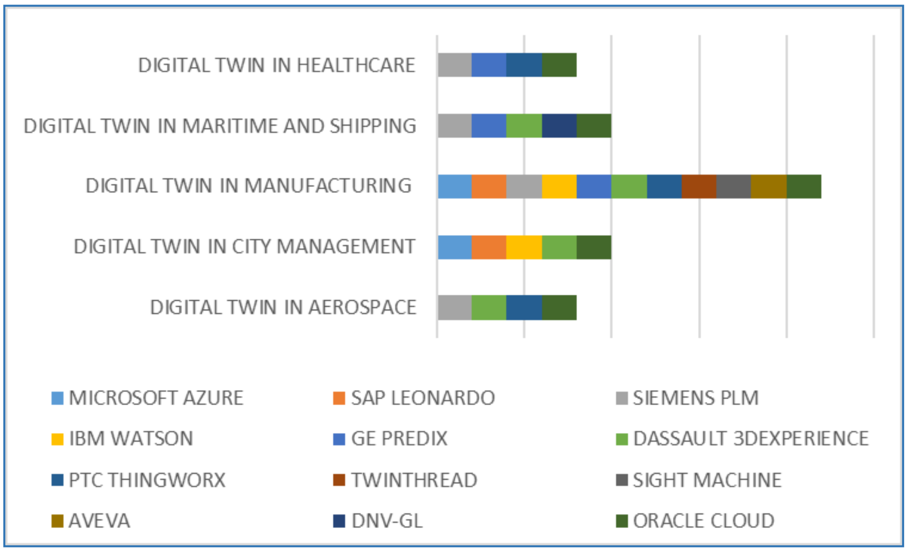


Figure 2: Fields of Digital Twin Application

1.2.2 THE DIGITAL TWIN ALONG THE PRODUCT-LIFE CYCLE (When a DT is developed?)

Digital twins provide an effective implementation for realising the end-to-end integration of a system or a service throughout the product lifecycle (Cheng et al. 2020). The end-to-end digital integration proposes to interconnect and to close the gap between product design, product development, production planning, production and associated services (Product Lifecycle Management) (Posada et al. 2015). In principle, out of the literature review, the digital twin in product life-cycle have been applied into the following phases (Tao, Cheng, et al. 2018), (Bao et al. 2018): Design phase; Production phase; Service phase.

1.2.2.1 THE DIGITAL TWIN IN THE DESIGN PHASE

Digital twins used in the design phase can guide the designers to iteratively adjust the customers' expectations and improve the design models, achieving personalized design.
The digital twin for a product design can be applied to the conceptual design, detailed design as well as virtual verification (Tao, Cheng, et al. 2018).
In the conceptual design stage, the digital twin serves to guide designers to formulate functional requirements (Tao, Sui, et al. 2018). It can make the communication between customers and designers more transparent and faster by using the real-time transmission data (Tao, Cheng, et al. 2018). In the detailed design phase, the digital twin enables simulation tests to ensure that the prototype can achieve the desired performance (Wärmefjord et al. 2017). In the virtual verification phase, the digital twin enables to simulate and predict the performance of the physical products based on virtual models (Damiani et al. 2018) (Bohlin et al. 2017). The digital twin in design stage is designed to generate digital design before the real execution (Q. Liu et al. 2018) (H. Zhang et al. 2017).

1.2.2.2 THE DIGITAL TWIN IN THE PRODUCTION PHASE

Digital twins in the production phase aim at real time monitoring and optimization and for predicting the future state of the physical twin, thus preventing downtime and failures (J. Lee et al. 2013).
(Leng et al. 2018) presents a Digital Twin for manufacturing cyber-physical system (MCPS). (Ding et al. 2019) introduces a DT-based Cyber-Physical Production System (DT-CPPS). MCPS is used for controlling the shop floor manufacturing while DT-CPPS for improving the flexibility, controllability and efficiency of shop floor manufacturing.
A digital twin for production control and optimization can analyse the online data collected from the physical line for searching the optimal solution to the physical line (Sun et al. 2017) or to complex product assembly shop-floors (Zhuang, Liu, and Xiong 2018). It can evaluate autonomously the production real-time (Vachálek et al. 2017) and optimize the resource allocation (H. Zhang, Zhang, and Yan 2018) autonomously (Rosen et al. 2015).
A Digital Twin reference model for rotating machinery fault diagnosis was developed in (J. Wang et al. 2018), defining the requirements for constructing the Digital Twin model. A digital twin for hydraulic supports (Xie et al. 2019) is built to simulate the actual hydraulic and to support diagnosis and degradation analysis. The digital twin finds application also in CNC machine tool (W. Luo et al. 2018) and in smart injection process (Liau, Lee, and Ryu 2018) to control the behaviours of the physical system in real-time. Based on these findings, the DT can optimize and elevate the production process to a higher level of effectiveness and flexibility.

1.2.2.3 THE DIGITAL TWIN IN THE SERVICE PHASE

The service phase refers to the phases after sale, including the product utilization and the maintenance (Tao, Cheng, et al. 2018). In the service phase, Digital Twins can provide value-added services with the support of physical simulation and data driven intelligence (Qi et al. 2018) (J. Wang et al. 2018). The Prognostics and health management (PHM) is crucial in the life-cycle monitoring of a product. The PHM is an engineering process of failure prevention, and predicting reliability and remaining useful lifetime (RUL) (Sutharssan et al. 2015). The digital twin (DT) improves the accuracy and efficiency of PHM (Tao, Zhang, et al. 2018) or it is utilized for improving the energy efficiency (M. Zhang, Zuo, and Tao 2018). (Abramovici, Göbel, and Savarino 2017) introduce a cloud-based Smart Product platform for the reconfiguration of Smart Products during the use phase using the concept of virtual product twins and an Internet of Things. The conceptual approach is prototypically demonstrated by considering a model environment for smart cars, which are temporarily reconfigured during their use phase.
The digital twin has been developed also for the waste electrical and electronic equipment recovery to support the manufacturing/remanufacturing operations (X. V. Wang and Wang 2018).

1.2.3 DIGITAL TWIN FUNCTIONS (Why DT is adopted?)

DTs may evolve over time or change their features according to an ideal life cycle (Macchi et al. 2018).
The following features/services have been assured to the DT so far based on the life-cycle stage previously referred:

1. Accelerating the product development speed: the digital twin can be used for designing products, testing them in real time situations, stipulating how the customer or the end user will use them and how the design will complement the product's environment. One of the biggest issues in the field of design is that it is hard to find customers who would appreciate design as much as the designers themselves do (Söderberg et al. 2017). Data from the real machine is loaded into the digital model to enable simulation and testing of ideas even before actual manufacturing starts. The digital twin can be used to plan, reconfigure the system in response to external changes.
2. Identifying customers' needs: performance and customer usage and preferences are reflected in the twin, and then feed into the product development process to increase the customer satisfaction and market share. Future products can be developed based upon the behaviour of existing products in the real world (Tao, Sui, et al. 2018).
3. Performance optimization and validation: the digital twin helps determining the optimal set of parameters and actions that can help maximizing some of the key performance, and providing forecasts for long-term planning (Vachálek et al. 2017). The digital twin can analyse performance data collected over time and under different conditions (Alcácer and Cruz-Machado 2019), reducing unplanned machine downtime, the amount of 'scrap' produced in each production line, and minimizing costly production quality faults.
4. Remote commissioning and diagnostics: the digital twin enables the remote visibility and diagnostic of the operations of interconnected systems such as manufacturing systems. This allows virtual monitoring systems and validation of the current status of production systems (i.e. energy monitoring and fault monitoring) (Qi et al. 2018). In addition, by obtaining the user's usage, the digital twin can upgrade personalized product functions (Cheng et al. 2020).

1.2.4 DIGITAL TWIN ARCHITECTURES (How to realize and to develop a DT?)

A general and standard architecture of a digital twin was first built by Grieves (Grieves 2014). (Stark, Kind, and Neumeyer 2017) characterize the DT as (1) an unique instance of the universal Digital Master model of an asset, (2) its individual Digital Shadow and (3) an intelligent linkage (algorithm, simulation model, correlation, etc.) of the two elements above (Kritzinger et al. 2018). An extended five-layer DT is proposed by (Tao, Zhang, et al. 2018) and it is composed by: (1) Physical entity (PE); (2) Virtual entity (VE); (3) Services (Ss) for PE and VE; (4) Data (DD); (5) Connection (CN) among PE, VE, Ss and DD. Compared to Grieves's architecture, data and services layers were added.
The five-layer DT architecture developed by (Ponomarev et al. 2017) presents: (1) cyber-physical layer; (2) primary processing/store data layer; (3) distributed computing and storage layer; (4) models and algorithms layer; (5) visualisation and user interfaces layer. This kind of architecture highlights the data storage, the distributed computing and management system as critical parts of the digital twin. An extended six-layer DT, is presented by (Redelinghuys, Basson, and Kruger 2019). The layers are: (1) physical devices; (2) local controllers; (3) local data repositories; (4) IoT gateway; (5) cloud-Based information repositories; (6) emulation and simulation. This structure is more focused on the transmission of data flow from the physical system (Layer 1 e 2) to the cloud (Layer 5).
Although there are various understandings of the DT architectures among researchers and industrial practitioners, models, data and connections always play the most important roles in the DT. From the computational perspective, the key functionality of digital twin is the combination of physics-based models

and data driven models to emulate assets accurately (Kaur, Mishra, and Maheshwari 2020). The IoT system carries out real-time data for the synchronization of the virtual twin with its corresponding physical twin with the capabilities of geometric assurance, anomaly detection, prediction, prescription and optimization.

1.2.5 DT COMPONENTS (What a DT is made of?)

According to the previous analysis of the architectures, several component types of the DT can be encompassed. Provided these perform according to different scopes and functionalities, the concept of "layer" can be used to group them, namely sequentially connected: Physical→ Network→Computing. In our view, each layer will thus be characterised by DT components (say, hardware or software technologies, models, information structures) with commonalities in their scope of use and interactions, having also complementary functionalities.

1.2.5.1 THE PHYSICAL LAYER

This layer includes the equipping of a physical system with sensors that record real-time states (e.g., vibration, force, torque, and speed) and working conditions (e.g., environment parameters, loads, and control orders) (Ruppert et al. 2018) of the physical systems. Some key elements for this layer are here listed: hardware, data type and Sources, application level.

1. The hardware shall be made up of sensors, actuators and embedded communication. The IoT system carries out real-time data acquisition through its smart gateway and edge computing devices (Gubbi et al. 2013). The technologies more used for the acquisition of data for a Digital Twin are:
   - RFID, that allows automatic identification and data capture using radio waves, a tag, and a reader (I. Lee and Lee 2015).
   - Wireless sensor networks (WSN) which consist of spatially distributed autonomous sensor-equipped devices to monitor physical or environmental conditions (Gubbi et al. 2013) (Tan and Wang 2010).
   - RFID Sensor networks, consisting of a very large number of nodes for monitoring and recording the physical conditions of the environment (Atzori, Iera, and Morabito 2010).

2. The data type and consequently the data sources can be classified in the following categories (Qi and Tao 2018):
   - Internet/Users Data from CRM, E-commerce platforms (e.g., Amazon) and social networking platforms (e.g., Twitter, Facebook, LinkedIn, and YouTube), to understand user preference, and behaviours.
   - Product data from computer-aided systems like CAD/CAM, CAE, etc.
   - Management data from manufacturing information systems such as MES, PDM, SCM, ERP, etc (W. Luo et al. 2018).
   - Manufacturing data including:
     - Operational data from manufacturing equipment such as product data, quality data, maintenance data (Dassisti, Semeraro, and Chimenti 2019).
     - Environmental data which affects the physical equipment operations, such as environmental pressure, ambient temperature, and moisture level (Cai et al. 2017).

Digital twins achieve data integration and data sharing between all production factors and information systems in line with the Vertical integration i.e. the integration of various Information Technology systems at different hierarchical levels.

3. The digital twin can be applied at different physical levels (Tao et al. 2019):
    - Unit Level: it is a minimum but independent individual, which cannot be further divided such as single piece of equipment (e.g., a machine tool or robot arm). It contains a basic closed loop of data between the physical and virtual spaces with the abilities of sensing and computing.
    - System Level: it is for example a production system such as a production line, a shop floor, or a factory. It is characterized by self-organization, self-configuration, and self-optimization.
    - System-of-Systems Level: it is characterized by enterprises collaborations. The application of the digital twin at the system-of systems level can achieve the horizontal integration. It refers to the exchange of information across supply chain such as resources management system, logistics, marketing, or intercompany value chains (Posada et al. 2015).

1.2.5.2 THE NETWORK LAYER

The role of the networking layer is to connect all components together for sharing data and information with other connected components (Da Xu, He, and Li 2014). This layer involves connections and interactions amongst physical elements and virtual space. Some key technologies belonging to this layer are: middleware; communication protocol analysis; communication protocol/interface conversion; wireless communication; Application Programming Interfaces (API).

1. The middleware is a software layer interposed between the technological and the application levels. The middleware architectures more used in the digital twin is the Service Oriented Architecture (SOA) approach. The adoption of the SOA principles allows for decomposing complex and monolithic systems into applications consisting of an ecosystem of simpler and well-defined components (Gubbi et al. 2013b).
2. The communication protocol analysis allows two or more entities in the DT to transmit information to each other. OPC Unified Architecture (OPC UA) and MTConnect protocols are the protocols more employed in digital twin applications to access to data and to transmit its real-time (Redelinghuys, Basson, and Kruger 2019).
3. The communication protocol/interface conversion transforms various communication protocols/ interfaces into a unit form. AutomationML is an open standard for a data format based on XML allowing the exchange of plant engineering information (Bao et al. 2018) (Drath et al. 2008). The AutomationML is used in digital twin to model attributes related to the digital twin. The goal is to interconnect the heterogeneous toolchain of digital manufacturing (Um, Weyer, and Quint 2017). It is used to exchange data between the Digital Twin and other systems and a methodology for communication and exchange of data (G. N. Schroeder et al. 2016) (Talkhestani et al. 2018);
4. The wireless communication can connect entities in the DT wirelessly, thus improving flexibility in data transmission.
5. The application Programming Interfaces (API) realize the communication between different software systems and models in the virtual space.

1.2.5.3 THE COMPUTING LAYER

This layer is the virtual layer and it is fundamental for computing and for decisional support and it is one of the most addressed in the last years either for the potential innovation or for the strong impact on decision support. It can be perceived as a set of "layers" interconnected, which includes data-driven models and analytics, physics-based models and application and user interface.

1.2.5.3.1 DATA DRIVEN MODELS AND ANALYTICS

The data layer includes all different types of data previously defined in physical layer (Uhlemann, Lehmann, and Steinhilper 2017). An important task is discovering information in data. Such task requires processing a large amount of data. Then the identified information is transformed into knowledge understood as information that has been confirmed and can be used to support decision-making (Paśko and Litwin 2019). Ackoff differentiates knowledge from data, information, and wisdom (Ackoff 1989) offers defining data, information, knowledge and wisdom (Figure 3) as follow (Rowley 2007):

- *Data are defined as symbols that represent properties of objects, events and their environment. They are the products of observation. But are of no use until they are in a useable (i.e. relevant) form. The difference between data and information is functional, not structural.*
- *Information is contained in descriptions, answers to questions that begin with such words as who, what, when and how many. Information systems generate, store, retrieve and process data. Information is inferred from data.*
- *Knowledge is know-how and it makes possible the transformation of information into instructions. Knowledge can be obtained either by transmission from another who has it, by instruction, or by extracting it from experience.*
- *Wisdom is the ability to increase effectiveness. Wisdom adds value, which requires the mental function that we call judgement.*

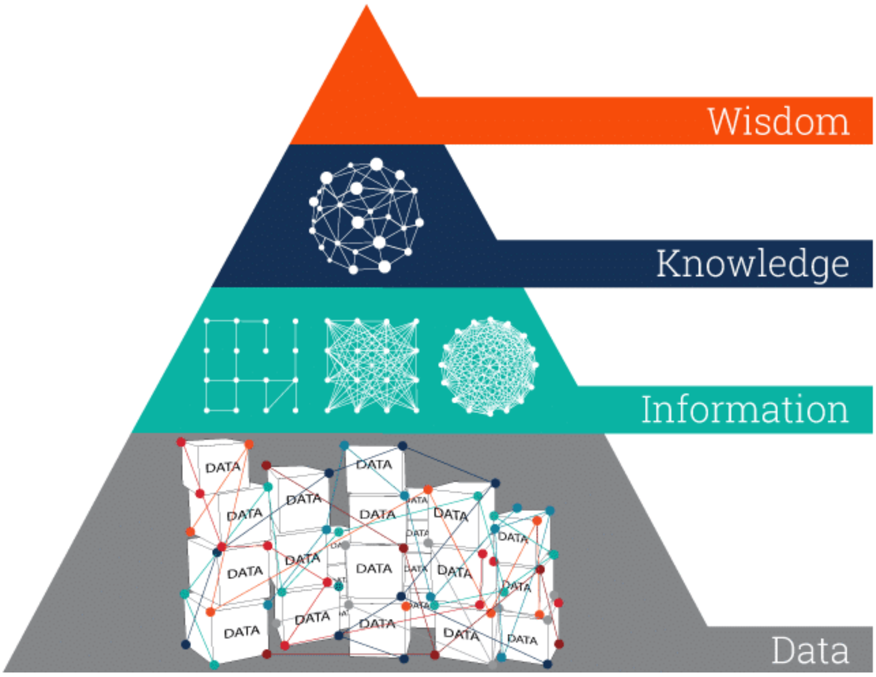


Figure 3: The progression from data to information, knowledge, and wisdom **(Ackoff 1989)**

The knowledge can be tacit and explicit (Nonaka 1994). The tacit knowledge is knowledge that cannot easily be expressed in words, such as cognitive skills, beliefs, mental models, and technical know-how. Explicit knowledge is knowledge that can be codified and captured in manuals, databases, presentations, models, and other media.

The data sub-layer has characteristics of heterogeneity of data and data sources, volume and speediness. According to the characteristics, data preparation and data analysis are the key aspects involved in the layer in analysis.

1. The data preparation process includes data selection, data cleaning, data modelling, data integration, and data transformation.
2. The data analysis includes all data-driven models such as machine learning, pattern recognition, and knowledge representation. Data-driven models are designed to extract knowledge from data (Y. Zhang et al. 2017). (J. Lee et al. 2014). The digital twin aims to integrate data across different domains (e.g. product, process and logistics) into virtual models (Kusiak 2018). The confidence level in the predictions depends on the available historical data (i.e. healthy and failure data) (Sutharssan et al. 2015). The data-driven models used in digital twins can be are:
   - Machine learning (Bishop 2016).
   - Neural Networks (G. P. Zhang 2000).
   - Deep Learning (Goodfellow, Bengio, and Courville 2016).

The machine learning refers the ability to give computers the possibility to learn without being explicitly programmed (Clarke, Fokoue, and Zhang 2009). It is classified in supervised, unsupervised (Sutharssan et al. 2015) and reinforcement (van Otterlo and Wiering 2012) learning (S. Jain, Shao, and Shin 2017). Machine learning techniques used in digital twin are supervised and unsupervised learning.

- The supervised learning develops model based on input and output data. The algorithms of this class consist of a target/outcome variable (or dependent variable) which is to be predicted by a given set of predictors (independent variables) (Tidriri et al. 2016). The application of supervised learning, across the digital twin, are for the prediction of system's failures (Asimov et al. 2018), or for prediction of the remaining useful life (RUL) of the physical twin (Z. Liu, Meyendorf, and Mrad 2018).
- The unsupervised learning, instead, discovers an internal representation from input data only (Sutharssan et al. 2015). It has no target or outcome variables to predict and estimate (Fahad et al. 2014). It enables to discover similar groups within data, based on clustering techniques (R. Xu and Wunsch 2005), (Grira, Crucianu, and Boujemaa 2004). In digital twin, these techniques are used for creating autonomously clusters for different working regimes to analyse machine conditions (J. Lee, Kao, and Yang 2014).

Artificial Neural Networks (ANN) and Deep Learning (DL) are computing systems that are inspired by human brain (Odom and Sharda 1990). The main scopes in using neural networks and deep learning in digital twin (J. Lee et al. 2013) are health assessment, performance prediction (R. Jain and Bhatnagar 2020), and fault diagnosis (Y. Xu et al. 2019). The scope is to make complex manufacturing increasingly autonomous.

1.2.5.3.2 PHYSIC-BASED MODELS

Belonging to the physic-based models' sub-layer, the key technologies are: model types, model features, and model interoperability.

1. The model types define the physics-based models and the functions of each model necessary to emulate the physical system. Physics-based models compare simulated results with known information, represented by mathematical models (Tidriri et al. 2016). A models represent a system in terms of logical and quantitative relationships that are then manipulated and changed to see how the model reacts, and thus how the system would react-if the mathematical model is a valid one (Law, Kelton, and Kelton 2000). The physics-based models are based on a set of different models to represent the structure, the behaviour and the interactions of a physical system (Tao, Zhang, et al. 2018) (Tidriri et al. 2016). The models can be summarised and classified in:
   - Quantitative model: use of static and dynamic relations among variables and process parameters to describe systems behaviour to carry out the detection in real-time (Venkatasubramanian, Rengaswamy, Yin, et al. 2003).
   - Qualitative model: use of static and dynamic relations among system variables and parameters for describing systems behaviour in qualitative terms (e.g. causalities or if-then rules) (Venkatasubramanian, Rengaswamy, and Kavuri 2003).

   Here below we present a list of the most studied and used models for digital twins:
   - A geometric model defining shapes, sizes, positions and assembling machine components is presented in (Tao, Zhang, et al. 2018) (Xie et al. 2019). It reflects the geometry, the kinematics, the logic and the interfaces of the real system (Ayani, Ganebäck, and Ng 2018).
   - A physical model is presented in (Tao, Zhang, et al. 2018) analysing the phenomena, such as deformation, cracking and corrosion. It simulates the physical properties and loads (Post, Groen, and Klaseboer 2009).

- A behaviour model describes the way the physical system is governed by driving factors (e.g. control orders) or disturbing factors (e.g. human interferences) (Tao, Zhang, et al. 2018) (Bao et al. 2018).
- A collaborative information model (Bao et al. 2018) defines how different components interact and simulates the collaborative behaviour among several assets.
- A decision-making model (Bao et al. 2018) makes the model capable of evaluating, reasoning, and validating. It consists of variable input, algorithms and a collection of constraints and rules. It includes rules of constraints, associations and deductions (Tao, Zhang, et al. 2018) and it stores and analyses the running status data, then it makes decisions using machine learning algorithm.

2. The model features involve the commonly features for modelling a digital twin. These are:
   - Scalability: it is the ability to provide an insight at different scales (from fine details to large systems) (Schleich et al. 2017) (Putnik et al. 2013).
   - Interoperability: it is the ability to convert, to combine, and to establish equivalence between different model representations (Schleich et al. 2017) (H. Zhang et al. 2017).
   - Fidelity: it is the ability to describe the closeness to the physical product (Schleich et al. 2017).
   - Dynamicity: it is the ability to reflect real time the physical process and modify autonomously itself if the physical system changes. This crucial issue concerns the convergence of the physical world with its digital counterpart (Weyer et al. 2016).
   - Modularity: it is the ability to integrate, to add, or to replace models (Guo et al. 2018) The idea behind this approach is to use and especially re-use predefined functional units (Semeraro, Lezoche, et al. 2019), that are systematically developed and logically interlinked for the configuration of a holistic manufacturing system (Stark, Kind, and Neumeyer 2017) (Negri et al. 2019).

3. The model interoperability is a critical aspect for the exchange of dynamic models and for Co-Simulation. Functional Mock-Up Interface (FMU) standard is commonly used in digital twin to solve this problem (Negri et al. 2019) (Schluse et al. 2018). FMU is an open standard for exchanging dynamical simulation models between different tools in a standardized format and for co-simulation (Blochwitz et al. 2011). FMI standard specifies two different kinds of FMUs:
   - Model Exchange (ME) – ME FMUs.
   - Co-Simulation (CS) – CS FMUs.

1.2.5.3.3 THE APPLICATION AND INTERFACE LAYER

In DT, applications export all the system's functionalities to the final user, enabling direct interactions between the digital twin and users. This layer exploits all the functionalities of the network and computing layers. In this layer, the DT can be combined with the virtual (VR) or augmented (AR) reality technologies to create interactive and immersive environments (G. Schroeder et al. 2016).

1.3 RESEARCH GAPS FOR THE DIGITAL TWIN

According to the analysis performed above, a generalisation of the Digital Twin can be figured out. A generic DT can be thought as consisting of several components organised into three main layers above recognised:

1. The physical layer, consisting of entities identified based on the stage of the product life cycle.
2. The network layer, connecting the physical domain to the virtual one. It shares data and information.
3. The computing layer, consisting of virtual entities emulating the corresponding real entities, including data-driven models and analytics, physic-based models, application and user.

The design criteria of the DT are not well assessed or even standardised. These are the more critical aspect. At the same time, the core capabilities provided by a generic digital twin are:

1. To Emulate: see, update, conceptualize.
2. To Think: compare, reason.
3. To Act: inform, decision support.

For the grace of this generalisation, we envision the following potential future research topics:

Research issue #1.

NEW BUSINESS OPPORTUNITIES

Most of digital twin applications refer to a single phase of a product life cycle, and they seem to be used for standard existing processes. There are, on the other hand, for instance, currently relatively few applications of digital twin for supporting network enterprises. The digital twin could potentially connect products, persons, machines, and enterprises within the virtual space. The digital twin can potentially help to integrate even the entire supply chain, throughout all phases of product life cycle. These potentialities may thus help to create or develop new scenarios or new business models for enterprise cooperation.

Research issue #2.

EVOLUTIONARY FEATURES (Adaptability, maintainability or flexibility over time).

The digital twin can be developed for different functions in each lifecycle phases based on the context or field of application. The Digital Twin should evolve synchronously with the real system along its whole life cycle. The DT is typically applied in contexts characterized by uncertainty and complexity, where the working conditions may vary depending on external and internal factors. The DT should have the ability to modify its initial configuration and to adapt itself to the current situation. The research issues should concern the criteria to design digital twins able to autonomously evolve, adapt, scale and/or reconfigure itself.

Research issue #3.

INTERACTION FUNCTIONALITIES

Existing Digital Twin applications are mainly developed for prediction purposes and used for decision-making support for human decision makers. The services assured by DT may vary depending not only on technologies available but also on its conception: say architecture, component interaction. Indeed, the unexplored field of research is the potentialities of synergy between components' features and functionalities, provided technologies are quite new and unexplored. The interaction between components is a critical design aspect. This interaction can modify the internal structures of the architecture in terms of properties or functionalities. The provided connections amongst components not necessarily assure an orchestrated functionality.

Research issue #4.
DT ARCHITECTURE
The DT consists of a set of models with complex structures and behaviour, that reflect the real-time operations of the physical system.
The lack of a reference architecture is almost well assessed (Lu et al. 2020). The lack of a univocal definition of a Digital and an univocal reference architecture leads to develop Digital Twin solutions using different technologies, interfaces, and communication protocols, models and data. Standard Digital Twin solutions should be developed to provide design criteria and design constraints where reference architectural aspects, reference information model and communication protocols are clearly defined.
There are applications where data-driven models and physic-based models are merged and others where these are used separately. It depends on the functionality for which the digital twin has been realized. It means that it is necessary to clarify the functions that the digital twin can fulfil. This may allow to identify the manufacturing components that shall be present in digital models and consequently, to define which types of models and which modelling techniques to use.
The interoperability and accuracy features are assured by an appropriate modelling, in order to provide an accurate virtual replica of the physical system and to achieve adequate real-time performances. At the same time, the modularization principle needs to be explored to improve the modelling efficiency: this would enable to improve the flexibility and reusability of digital twin towards different applications.

Research issue #5.
DT INVARIANT MODELLING
The product and production system are becoming increasingly complex, as the number of their components, frequency of market demand changes and need for related innovation increase. To manage the complexity, digital representations constitute a major challenge in improving the accuracy of existing and future simulation and emulation tools. A digital representation, hence, a digital twin, bridge the gap between the physical and virtual system improving and supporting the decision making (Estefan 2007).
The digital twin requires the building and the applying digital models representing the set of resources and processes knowledge. Modelling such digital copy of the physical system to perform real-time validation and optimization is quite complex and thus needs a big amount of data and some modelling patterns representing the operational semantics of the modelled elements.
However, it is hard to construct an accurate model by using traditional model-based approaches because of the complexity of the systems (J. Lee, Bagheri, and Kao 2015). On the other hand, recent advances in sensor technology (Dassisti, Panetto, et al. 2017) have enabled significant growth of data collection and analysis, leading researchers to focus on data-driven methods. Generally, the modelling action has a specific application type. For this reason, the core challenge of the digital transformation modelling is to create an invariant approach towards different applications.

1.4 SYNTHESIS AND THESIS OBJECTIVE

In this chapter, the state of art of digital transformation in SMEs and the digital twin have been explored. The state of art aims to demonstrate and to discuss the approaches developed to build a digital twin.
The development of accurate and repeatable models is identified as a challenge faced by digital twin research domain.
Based on the review results, the thesis objective is to address and to achieve the invariant modelling of manufacturing systems. The next chapter presents a detailed literature review of the existing approaches to present the contribution of the Ph.D. thesis.

CHAPTER 2

FORMALIZATION OF DATA-DRIVEN INVARIANT MODELLING CONSTRUCTS: CONTRIBUTION OF THE THESYS

CHAPTER 2 – FORMALIZATION OF DATA-DRIVEN INVARIANT MODELLING CONSTRUCTS: CONTRIBUTION OF THE THESIS

INTRODUCTION

To address the research gap related to the invariant modelling, the chapter 2 presents an overview of the existing approaches in literature with the aim of defining the contribution positioning of the thesis. We perform an exhaustive literature review in section 2.1 to identify and to describe model-based approaches in section 2.1.1, data-driven approaches in section 2.1.2, the hybrid approaches in section 2.1.3 and design pattern in section 2.1.4. Evaluating the research context and the identified limitation, the contribution positioning is defined in section 2.2.

2.1 STATE OF ART OF MODELLING APPROACHES

Manufacturing systems requires efficient approaches to model the complexity of industrial systems and their related performance requirements. Numerous methods and tools have been proposed and developed in the literature with the aim to model complex system. The methods and tools can be distinguish in three main approaches: model-based, data-driven, and hybrid approaches that involve models and data (Tobon-Mejia et al. 2012).

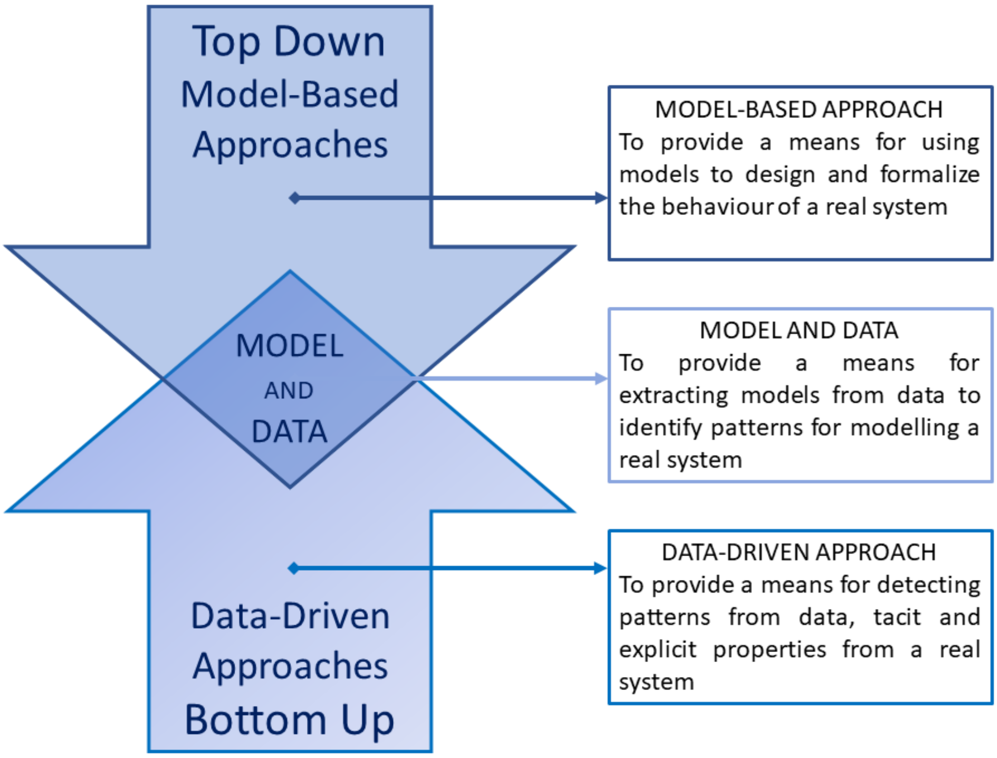


Figure 4: Summary of the Approaches

2.1.1 MODEL-BASED APPROACHES

A model is a representation of a system for a specific purpose (Konikow and Bredehoeft 1992). A system is defined as a collection of entities, e.g., people or machines, that act and interact together toward the accomplishment of some logical end (Schmidt 1970). The model is developed to understand and to formalize the system (Venkatasubramanian, Rengaswamy, and Kavuri 2003).

The use of models is the basis of model-driven architecture (MDA). The model-driven architecture (MDA) (J. Miller and J. Mukerji 2003) is an approach to design and implement software. MDA is developed by Object Management Group (OMG) in 2001 (https://www.omg.org/mda). MDA provides guidelines for structuring software specifications that are expressed as models. The MDA approach provides three types of models from three different viewpoints: the computation independent model (CIM), the platform independent model (PIM), the platform specific model (PSM).

A computation independent model (CIM) of a system describes the domain and the requirements of the system. A CIM consists of a model which captures information about the data of a system (informational viewpoint). It does not define the structure of systems.

A platform independent model (PIM) of a system describes the operation of the system independent of any platform. A PIM involves a model which captures information about the data of a system (informational viewpoint) and a model which captures information about the processing of a system (computational viewpoint), independent of any platform.

A platform specific model (PSM) of a system describes the operation of the system for specific platforms. A PSM includes a model which captures information about the data of a system (informational viewpoint), and a model which captures information about the processing of a system (computational viewpoint), based on a specific platform. Model are developed in Unified Modelling Language (UML). UML is a standard modelling language for visualizing, specifying, and documenting software systems (UML 2005).

Model-based systems engineering (MBSE) is the: *"formalized application of modelling to support system requirements, design, analysis, verification, and validation activities beginning in the conceptual design phase and continuing throughout development and later life cycle phases*" (International Council on Systems Engineering (INCOSE) 2007). MBSE is sponsored by the International Council on Systems Engineering (https://www.incose.org/) and the OMG Systems Engineering DSIG (https://www.omg.org/syseng/). MBSE became a standard practice thanks to the introduction of modelling standards such as SysML (ISO/IEC 19514 2017), UPDM (UPDM 2017), Modelica (https://www.modelica.org/). In practice, System Modelling Language (OMG SysML) finds large applications in MBSE (Madni, Madni, and Lucero 2019). SysML "is a general-purpose graphical modelling language for representing systems that may include combinations of hardware, software, data, people, *facilities, and natural objects"* (http://www.omgsysml.org/what-is-sysml.htm). The systems modelling language (SysML) refers to a subset of UML 2. The advantage of using SysML as modelling standard is the ability to describe the syntax and the semantic behind a system (Weilkiens 2011). SysML provides nine interrelated types of diagrams to reproduce and to describe a system defining the requirements, the structure and the behaviour for identifying the possible decisions or actions that need to be made.

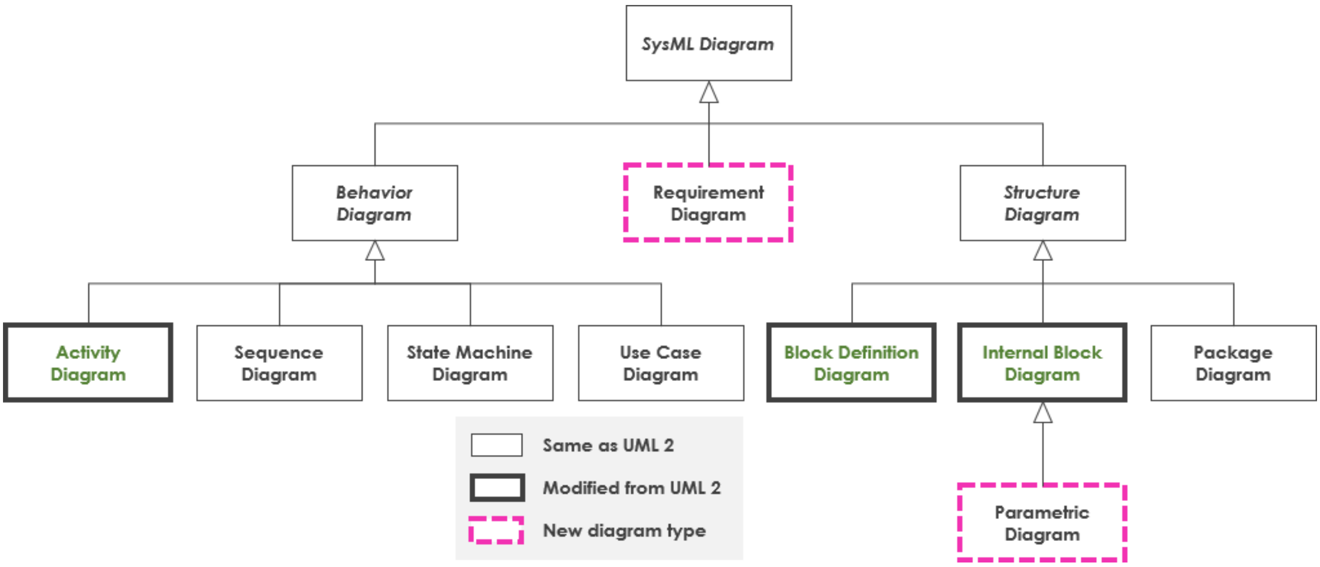


Figure 5: SysML Diagram (Friedenthal, Moore, and Steiner 2014)

SysML introduces two new types of diagrams compared to UML: the requirement diagram and the parametric diagram. Furthermore, it modifies three types of UML diagrams: the block definition diagram from the UML class diagram, the internal block diagram from the composite structure diagram of UML and the activity diagram. The last four diagrams (use case diagram, package diagram, sequence diagram, state machine diagram) are inherited from UML and reused without the addition of extensions.
The model-based approaches integrate and formalize the physical understanding into a set of models. The success of model-based approaches depends on the accuracy, the fidelity and the robustness of the model (Friedenthal, Moore, and Steiner 2014).

2.1.2 DATA-DRIVEN APPROACHES

In contrast to the model-based approaches, where a-priori knowledge about the process is needed, in data-driven approaches, the availability of large amount of historical process data is required (Venkatasubramanian, Rengaswamy, Kavuri, et al. 2003) to discover knowledge.
The process of discovering tacit and explicit knowledge, from a collection of data is called knowledge discovery in databases (KDD). KDD is the process of: *"using the database along with any required selection, pre-processing, subsampling, and transformations of it; to apply data mining methods (algorithms) to enumerate patterns from it; and to evaluate the products of data mining to identify the subset of the enumerated patterns deemed knowledge"* (U. M. Fayyad, Piatetsky-Shapiro, and Smyth 1996).
KDD is an inherently iterative process of selecting data, pre-processing it, transforming it into a workable form, data mining over it, and interpreting the results (U. M. Fayyad, Piatetsky-Shapiro, and Smyth 1996).
The knowledge discovery in databases (KDD) scopes are defined by the intended use of the system. There are two types of goals:

- Verification-Driven, known as deductive approach (top-down), where the system is limited to verifying the user's hypothesis.
- Discovery-Driven, known as inductive approach (bottom-up), where the system autonomously finds new patterns.

  The discovery objectives are split into:

- Predictive objectives (e.g., classification, regression, anomalies/outlier's detection), where the system finds patterns for predicting the future behaviour.
- Descriptive objectives (e.g., clustering, association rule discovery, sequential pattern discovery), where the system finds patterns from data (U. Fayyad, Piatetsky-Shapiro, and Smyth 1996).

Data mining is: "*a step in the KDD process consisting of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce an enumeration of patterns over the data"* (U. Fayyad, Piatetsky-Shapiro, and Smyth 1996).

Given a set of data (e.g., cases in a database), a pattern is defined as: "*an expression in some language describing a subset of the data or a model applicable to that subset. Given a set of facts (data) F, a language L, and some measure of certainty C, a pattern is defined as a statement S in L that describes relationships among a subset Fs of F with a certainty c, such that S is simpler (in some sense) than the enumeration of all facts in Fs"* (Frawley, Piatetsky-Shapiro, and Matheus 1992).

Data mining approaches are data-driven approaches designed to mine the hidden patterns and knowledge through the analysis of a huge amount of historical data (Y. Zhang et al. 2017). The patterns mining methods make use of the context data and they unveil the complex coupling relationships.

The data-driven approaches can be classified in supervised and unsupervised approaches (S. Jain, Shao, and Shin 2017) as explained below.

2.1.2.1 SUPERVISED DATA-DRIVEN APPROACHES

Supervised is the task of inferring a function from labelled training data. It aims to build a model that can correctly predict the output of an unseen instance by observing a set of labelled instances (Al-Sahaf et al. 2019). Supervised data mining attempts to explain the behaviour of the target as a function of a set of independent attributes or predictors widely (J. Zhang, Williams, and Wang 2018).

The principal supervised mining techniques are: classification and regression (Solomatine and Ostfeld 2008).

- Classification (Ahmed 2004) technique aims to predict target categories or classes. It is composed by a training dataset containing objects (input). Each object is evaluated based on a set of predictors. The Output (Target Class) is a model (classifier) that assigns a specific label (category) to each object based on the predictors (Gorunescu 2011). Classification algorithms find relationships between the values of the predictors and the label of the target (Mitra, Pal, and Mitra 2002).
- Regression (Ngai, Xiu, and Chau 2009) technique aims to predict a numeric value. It is composed by a training dataset containing objects (input). Each object is evaluated based on a set of predictors. The output (Target Class) is a model (classifier) that assigns a specific value to each object based on the predictors. In the model build (training) process, a regression algorithm estimates the value of the target as a function of the predictors.

The major strength of the supervised approaches is that these approaches do not require a high-level of domain knowledge (Y. Liu et al. 2019). The major weaknesses are the massive amounts of data needed for training a reliable model and the results and the performances depend on the quality of training data (Zhao et al. 2019).

2.1.2.2 UNSUPERVISED DATA-DRIVEN APPROACHES

The unsupervised learning (Sutharssan et al. 2015), instead, discovers an internal representation from input data only. The principal unsupervised mining techniques are: clustering, association rule mining, sequential patterns discovery and anomaly detection explained below.

- Clustering (Hansen and Jaumard 1997) technique aims to split a set of data into subsets (clusters) with similar characteristics. Clustering is used to find out groups of items that are similar (Yin et al. 2011). Accordingly, the clustering algorithms can be classified in:
    - Partitioning clustering (Fahad et al. 2014): given a set of n objects, a partitioning method constructs k un-nested partitions of the data, where each partition represents a cluster and $k \leq n$. It splits the data into k groups such that each group must contain at least one object.
    - Hierarchical clustering (Grira, Crucianu, and Boujemaa 2004): aims to obtain a hierarchy of nested clusters, called dendrogram. It shows how the data are grouped and which clusters are related to each other. The hierarchy of nested clusters can be agglomerative (bottom-up) or divisive (top-down). An agglomerative clustering starts with one object for each cluster and recursively combine in one cluster. A divisive clustering starts with all object in one cluster and recursively split in different clusters.
    - Density-clustering (Grira, Crucianu, and Boujemaa 2004): considers that clusters are dense sets of data items separated by less dense regions.
- Association (Agrawal, Imieliński, and Swami 1993) technique aims to detect the probability of the co-occurrence of items in a collection. The relationships between co-occurring items are expressed as association rules

  The generation of knowledge, through association rules, is promoted by the Formal Concept Analysis (FCA) (Poelmans et al. 2010) because it is particularly suited for exploration data (Valtchev, Missaoui, and Godin 2004). FCA is a mathematical theory oriented at applications in knowledge representation (Wille 2002). It provides. It provides tools to group the data and to discover formal patterns by representing it as a hierarchy of formal concepts organised in a semi ordered set named lattice (Ganter, Stumme, and Wille 2005).

  Indeed, a major part of the data is stored in relational databases. While most existing data mining approaches detect patterns in a single data table, multi-relational data mining (MRDM) approaches detect patterns that involve multiple tables (relations) from a relational database (Džeroski 2003). Relational Concept Analysis (RCA) is a multi-relational data mining approach. RCA enables to extract multi-relation association rules where each rule consists of several relations between entities.
- Sequential patterns discovery (Lin and Lee 2003) technique aims to identify associations or patterns over time. The goal is to model the states of a process generating the sequence trends and the report deviation over time.
- Anomaly detection (outlier/change/deviation detection) (Buczak and Guven 2015) technique aims to identify cases that are unusual within data that is seemingly homogeneous.

Based on the reviews, the major strength of the unsupervised data-driven approaches is the possibility to discover tacit knowledge from a set of data such as unknown faults and unknown operation patterns.

The major weaknesses are the massive amounts of data for training a reliable model and the results and the performances depend on the quality of training data (Tidriri et al. 2016). It is time consuming to find valuable knowledge from all the mined knowledge. Unsupervised data-driven approaches require a-priori knowledge model (Zhao et al. 2019).

2.1.3 HYBRID APPROACHES

Recently, there has been a growing interest in the integration of data-driven and model-based approaches in order to provide better results. In the following, a review of hybrid approaches is carried out. The types of hybrid approaches reported in the literature are listed as follows:

- DATA ∧ MODEL: *simultaneous combination*
  Use the data-driven and model-based approaches simultaneously for data acquisition and feature extraction. A combination of data sources can be used to generate hybrid diagnosis methods (Sheibat-Othman et al. 2014) which combines collected data during the system operations (field data) with data generated using the system physical model (simulation data) to address the lack of sufficient faulty data (S. Frank et al. 2016).
  Among feature extraction, data-driven and model-based are combined and compared to detect redundancy relation between known variables in the system such as parameters, process measurements, and inputs (Khorasgani et al. 2018).
- DATA → MODEL: *model-based is consequent to data-driven*
  Use a data-driven approach to infer a measurement model and a model-based approach to describe the behaviour of the system (Hanachi et al. 2019). The data-driven approach builds a mapping from the measurement to the system state. This enables to use a model-based approach to understand the system state (Goebel, Eklund, and Bonanni 2006). A second data-driven model has been introduced in (Liao and Köttig 2016) to reduce the uncertainty made by the model-based approach.
- MODEL → DATA: *data-driven is consequent to model-based*
  Use model-based approach to define the failure threshold (Li et al. 2019), and data-driven approach to calibrate the model to make the results more accurate and to narrow the uncertainty of model results (Huiguo Zhang, Kang, and Pecht 2009).

Despite the advantages and the potentials of hybrid approaches, the main obstacle is the lack of a generic framework for hybrid approach (Tidriri et al. 2016).

2.1.4 DESIGN PATTERN: OBJECT-ORIENTED PROGRAMMING

Most of the model-based and data-driven have been developed for specific application or equipment. A clear systematic way to model commonly problems does not exist. The need to develop time-tested solutions to recurring problems could be as closely related to objected-oriented design pattern domain. Object-oriented programming (OOP) is a programming language model that organizes software design around objects, rather than functions and logic. An object can be defined as a data field that has unique attributes and behaviour (Gamma, Helm, and Johnson 1995).
Designing object-oriented software involves the definition and the application of design patterns. A design pattern in OOP, differently than data-driven, is a general repeatable solution to a commonly occurring problem in software design (Gamma, Helm, and Johnson 1995). A pattern is defined as: *“a problem which occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over without ever doing it the same way twice"* (Alexander 1977).

The authors proposed 23 standard object-oriented design patterns classified into three sub-categories based on kind of problem they solve: creational, structural and behavioural (Gamma, Helm, and Johnson 1995). Creational patterns concern the process of object creation. Structural patterns deal with the composition of classes or objects. Behavioural patterns characterize the ways in which classes or objects interact.
A design pattern is described by four main elements: 1) its name, which concisely summarizes the objective of the pattern; 2) the problem it solves, which describes when to apply a pattern; 3) the structure of the general solution, which lists the participants of the pattern: elements, their relationships and roles they play; 4) consequences, which list the advantages, disadvantages, and compromises to be aware of when using the pattern. The following elements can be used to represent a design pattern more precisely: 1) intent, motivation, and applicability to understand the context; 2) structure, participants, and collaborations to make the application easier; 3) implementation, sample code, known uses, and related patterns to show the pattern and the correlation between patterns.
Design patterns are important also in other programming paradigms: 1) documenting the solution of a common problem in their paradigm; 2) solving common problem; and 3) automating some part of the program development by code generation.
Design patterns have been introduced for functional logic languages (Antoy and Hanus 2002). Five design patterns are described to solve some general and challenging problems in functional logic languages. Each design pattern presents these characteristics: name, intent, applicability, structure, consequences, known uses, and see also.
The design patterns have been built also as tools for communicating program construction expertise and propose a set of design patterns for functional strategic programming (Lämmel and Visser 2002). They produced a catalogue of 13 design patterns with these characteristics: name, category, intent, motivation, applicability, schema, description, sample code, consequences, and related patterns.
Design patterns find application in various areas of computer science and engineering such as software architecture (Bass, Clements, and Kazman 2003), building Common Object Request Broker Architecture (CORBA) applications (Mowbray, Ruh, and Soley 1997), real-time systems (Douglass 2003), distributed computing (Ramirez and Cheng 2010), and embedded network systems (Wahba, Hallstrom, and Soundarajan 2010).
Design patterns are useful in terms of transferring a solution among different implementations. Patterns are mostly used to generate some part of the code to eliminate low level implementation details (Budinsky et al. 1996). However, they also have some trade-offs and limitations. The detection of patterns and the update of new pattern is developed by experts. It is not an automatic process. The main consequence of using design patterns is trying to apply many of them across the same objects. It is not clear how many design patterns are enough in a project. There is also a risk of applying too many unnecessary design patterns.

2.2 CONTRIBUTION POSITIONING

In the above, a summary of methods for modelling systems have been presented. The approaches can be categorized into model-based, data-driven and hybrid approaches (J. Luo et al. 2003). The choice of the method to apply is typically based on the specific use case.
The model-based approaches rely on the use of models to simulate the systems behaviour in different operating conditions but these models are not easy to develop and keep updated during the system life-cycle (J. Luo et al. 2008).

The data-driven approaches allow to integrate parameters across different domains (e.g. product, process and logistics) into models that would be difficult to build with the traditional model-based approaches. The data-driven, in fact, aim at transforming the data into relevant information and into reliable behavioural models (Okoh et al. 2014) but the quality and scope of the data play a critical role (Kusiak 2018).
The main challenges of data-driven approaches involve the massive data sets and high dimensionality, the prior knowledge to understand pattern.
At the same time, a single approach cannot be adapted to all different applications because of the complexity and the variety that characterize manufacturing systems. Hybrid approaches (Tidriri et al. 2016) have been developed to cope with specific problems (X. Zhang and Hoo 2011). The hybrid approaches not define a common solution and framework to apply to different systems (Ghosh, Ng, and Srinivasan 2011).
For this reason, the objected-oriented design pattern domain is investigated. The design pattern (OOP) defines, develops and implements repeatable solution to a commonly occurring problem in software engineering.
The contribution positioning of the thesis is to make the modelling more structured and reliable. The key issue to be addressed is the invariance. The idea is to detect automatically from data invariant modelling constructs. Invariant modelling constructs need to be developed to describe/emulate a system independently from its context of application. Data-driven invariant modelling constructs can be used and in particular re-used to create digital models for different systems or processes (Semeraro, Lezoche, et al. 2019).

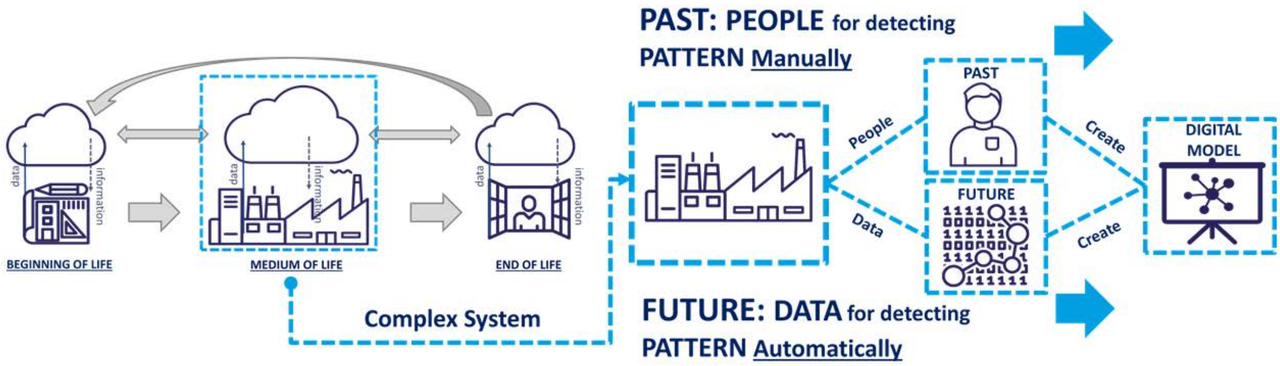


Figure 6: Research Context

The process to detect invariant modelling constructs is shown in Figure 7. The first aspect is to identify the systems to analyse (1), as well as the choice of products and the related manufacturing process. A model-based approach (2) is required to model the system describing the function, the structure and the behaviour. The model-based approach draws a detailed representation of system under consideration and it enables the selection of data to analyse (3). When the system and the data are selected, the data-driven approach is used (4) to detect and to automatically discover associations and relationships among data. The associations can describe recurrent behaviours of the system and it can codify tacit knowledge that can be used to better understand the behaviour of the system. The discovered association are analysed based on the model developed to extract knowledge (5) from data and to define the physical meaning of the associations. The extracted knowledge represents a data-driven invariant modelling construct. Data-driven invariant modelling constructs can be formalized (6) to be easier to understand and to analyse where and when these can be applied. Data-driven invariant modelling constructs enable to design the virtual model of a system for realising its digital twin (7). Data-driven invariant modelling constructs can be reused among systems or processes operating in a similar condition (8).

The next chapter presents the approach to detect data-driven invariant modelling constructs.

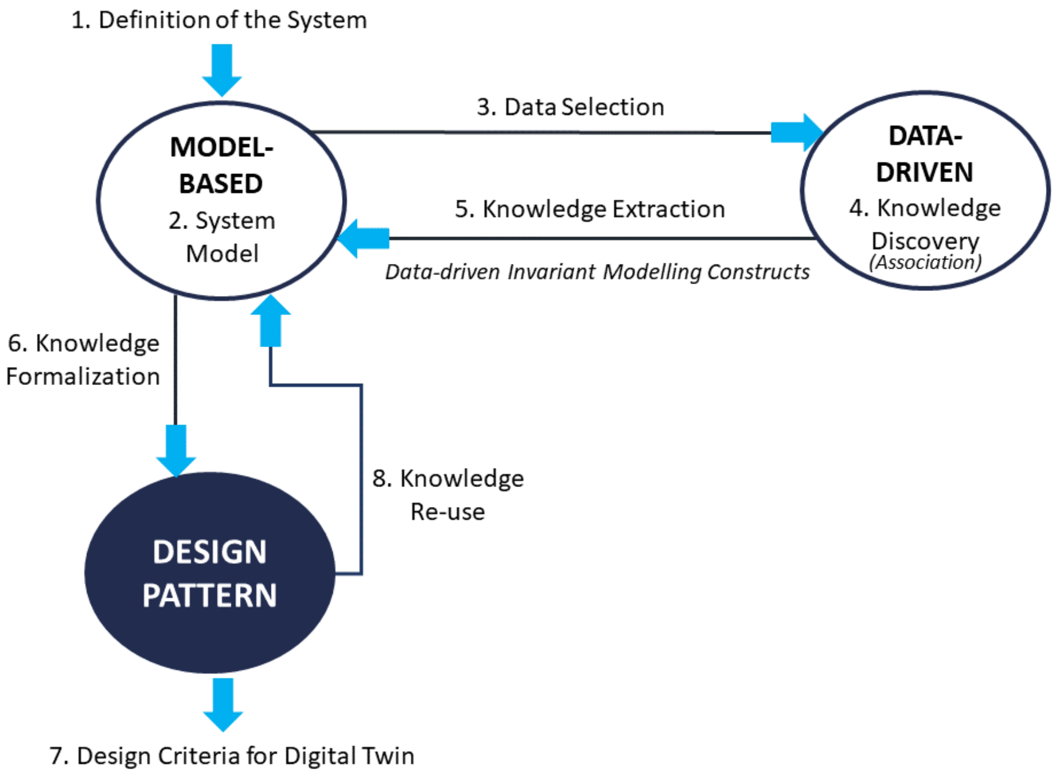


Figure 7: The Process for Detecting Data-driven Invariant Modelling Constructs

CHAPTER 3

THE DEFINITION OF THE APPROACH TO EXTRACT DATA-DRIVEN INVARIANT MODELLING CONSTRUCTS

CHAPTER 3 - THE DEFINITION OF THE APPROACH TO EXTRACT DATA-DRIVEN INVARIANT MODELLING CONSTRUCTS

INTRODUCTION

The chapter 3 aims at formalising an approach to detect and to formalize invariant modelling constructs based on the data analysis. Each single state is described in a dedicated section.

3.1 DEFINITION OF THE APPROACH

The approach is articulated in eight different stages as shown in Figure 8. The stages are: 1) Definition of the system; 2) System model; 3) Data selection; 4) Knowledge discovery; 5) Knowledge extraction; 6) Knowledge formalization; 7) Design criteria for building a digital twin; 8) Knowledge re-use.
The follows paragraphs explain each single stage in more detail.

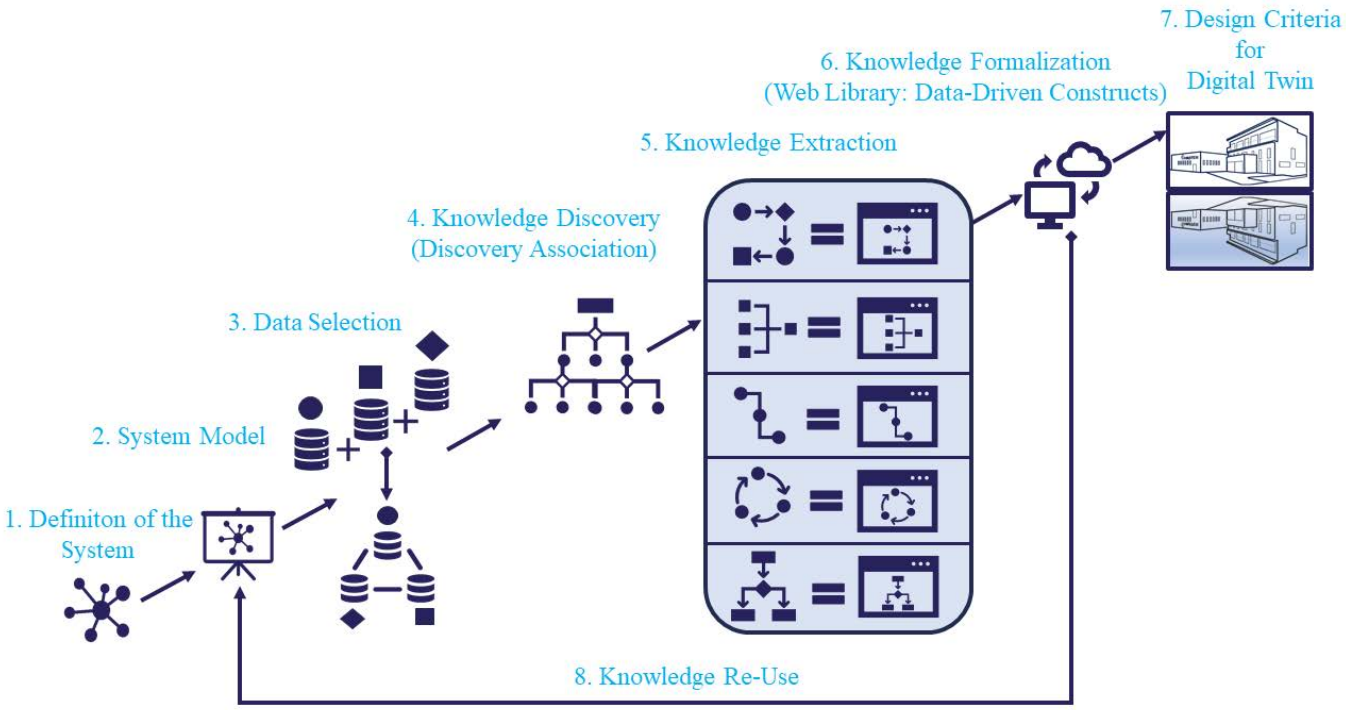


Figure 8: The Approach to Extract and to Formalize Data-driven Modelling Construct

3.1.1 DEFINITION OF THE SYSTEM

Provided the Industry of the Future paradigm relies on the use of sensors for the acquisition, processing and analysis of data, the most critical point is the definition and the selection of a complete set of parameters in order to effectively control manufacturing processes. For the definition of the system to analyse, two different approaches have been combined. The first is the life cycle analysis (LCA) because it assesses the

amount of resource consumed and relative emissions. The second one is the exergy analysis because it defines the information infrastructure.
Life cycle analysis (LCA) is an analytical approach used to quantify and to interpret the flows to-and-from the environment through the whole life cycle of a product, process or service but it suffers some limitations (UNI EN ISO 14040 2006). The first advantage is that it appreciates only quantities of elements flowing in the processes (say, energy, materials, etc.). The second is the dependence on standard databases referring to general or averaged figures, independently of the specific process analysed (Reap et al. 2008).
The exergy analysis (EA), instead, involve the use of state variables, concerning products and processes, to precisely and objectively quantify (both in physical and economical terms) the gap between the processes' efficiencies and their maximum achievable values, as a function of the physics of processes and the surrounding environment. Exergy analysis allows to identify the optimization opportunities and to recognize improvements and/or innovation paths for processes and products (Bakshi, Gutowski, and Sekulic 2011). Exergy is defined as *"maximum theoretical useful work obtainable as the system is brought into complete thermodynamic equilibrium with the thermodynamic environment while the system interacts with this environment only"* (Bakshi, Gutowski, and Sekulic 2011).
The exergy approach highlights the gap from the ideal condition allowing to identify and recognize the optimization opportunities based on the appropriate selection of the information infrastructure, and thus to recognize lack of symmetry in the system behaviour. The exergetic analysis is combined with the life cycle analysis (LCA) to define how to analyse a system. The stages of the combined approach are outlined and described below:

1. Identify the scenario of operation of systems under analysis, as well as the choice of products and the related manufacturing process by defining the thermodynamic model of the systems.
2. Split the system in different subsystems and draw a detailed representation of the operation of every subsystem under consideration.
3. Perform an exergy balance of each subsystem, providing a critical index based on the exergy loss Ex_{loss}.
4. Define the thermodynamic parameters critical to measure for each sub-system.

3.1.1.1 Identification of the system to analyse

This stage identifies the critical product and consequently the critical manufacturing process object of the analysis. The life cycle analysis defines the set of critical products and processes based on the functional unit (chosen as the single product). In International Standard Organization 1997, (UNI EN ISO 14040 2006), the functional unit is defined as *the measure of the performances of the functional outputs of a considered system. In the same standard the importance of the functional unit for the comparability of the LCA results is highlighted.*
The functional unit allows in this way to normalise the efficiency of processes, and thus to have a fast glance to the overall efficiency of the production process life cycle. LCA defines the boundaries of the system to be analysed. The boundaries of the system typically include the following elements: raw material production; semi-finished products preparation; additive production; internal and external logistic; product component production; waste management; water and energy consumption; emission in air (Owens 1997).

3.1.1.2 Application of the exergetic analysis

This stage enables to split the system identifies into different subsystems and to draw a detailed representation of the operations of every subsystem under consideration. Each subsystem is a thermodynamic system characterized by its corresponding input and output flows of mass and energy.
The exergetic view of manufacturing processes allows to recognise the critical sub-systems from the informational flow point of view. Exergetic analysis for manufacturing processes is, in fact, typically related to the quality of use of resources, with a deeper contextual view of the process (Dassisti, Semeraro, and Chimenti 2019).

3.1.1.3 Performing the exergy analysis of each subsystem to compute the exergy loss Ex_{loss}

The exergy approach provides the evaluation of the quality of energy usage, the identification and quantification of the inefficiency of the process through the measurement of the Exergy Loss. The exergetic analysis is performed by evaluating the exergy loss for each subsystem. Exergy approach consists of a model of a process to be controlled, whose inputs are the exergy flow of materials ($Ex_{M,in}$), the exergy flow of work ($Ex_{W,in}$), and the exergy flow of heat ($Ex_{Q,in}$) and the output are the exergy flow of materials ($Ex_{M,out}$), the exergy flow of work ($Ex_{W,out}$), and the exergy flow of heat ($Ex_{Q,out}$). The difference between the exergy values in input and the exergy values in output represents the exergy loss (Ex_{loss}).
The equation of exergetic balance is as follow (Moran, Shapiro, and Moran 2008):

$$Ex_{M,in} + Ex_{W,in} + Ex_{Q,in} = Ex_{M,out} + Ex_{W,out} + Ex_{Q,out} + Ex_{loss} \tag{1}$$

The optimization criterion in the exergetic analysis is to minimize the term Ex_{loss}, since the exergy loss is proportional to the generated entropy and this latter is responsible for the less-than-theoretical efficiency of the system. This criterion, in Industry of the Future implementations, may allow to select the critical systems as well as their critical parameters, from which the term Ex_{loss} and the exergetic efficiency depends (Dassisti, Semeraro, and Chimenti 2019). The equation adopted to compute Ex_{loss} is (Gutowski, Dahmus, and Thiriez 2006):

$$Ex_{loss} = Ex_{M,in} + Ex_{W,in} + Ex_{Q,in} - Ex_{M,out} - Ex_{W,out} - Ex_{Q,out} \tag{2}$$

3.1.1.4 Definition of the critical thermodynamic parameters to measure

The exergetic analysis combined with life cycle analysis aims to select the critical product and the critical process to analyse. A process is split in different subsystems to select the critical parameters. The key aspect is the classification of data for each subsystem.
Exergy analysis allows to identify and classify the parameters into four main groups:

- Main parameters already controlled.
- Main parameters not yet controlled.
- Derived parameters controlled.
- Non-controllable parameters.

Descending from the above parameters, it is possible to define 1) the set of parameters of each subsystem, 2) the parameters monitored and 3) the parameters to monitor. The functional view of the process provided by the exergetic analysis thus permit to recognise the sensorization path of the Industry of the Future paradigm (Dassisti, Siragusa, and Semeraro 2018).
The life cycle analysis combined with the exergetic analysis allows to define the system components to model, their contents and their relationships. The informational flows are typically tied to physical layout or logical operational flows.

3.1.2 SYSTEM MODEL

The purpose of system model is to show how system components, their contents (Value Properties, Behaviours, Constraints), and their relationships interact (Friedenthal, Moore, and Steiner 2014). The chosen modelling language is SysML.
The chosen tool is IBM rational rhapsody® designer for systems engineers (https://www.ibm.com/uk-en/marketplace/systems-design-rhapsody). It is a software that allows the creation of an environment thanks to which the model, designed with standardized rules, can be simulated.
In particular, the tool provides the analysis of the system requirements, the visual development of the model in order to capture the design graphically and, also, the simulation and model execution that are useful to validate the system behaviour.
The modelling stages to perform can be summarized as follow: 1) model requirements and use case; 2) model the subsystems in block definition diagrams; 3) model the relationships between subsystems in internal definition diagrams; 4) model constraints in parametric diagrams; 5) model the behaviour of the system in state machine diagrams.

3.1.2.1 Model requirements and use case

The requirement diagram (req) allows to represent project' requirements in textual form providing a detailed description of user's requirements directly within the model and allow to trace their relationship with other elements in the model. A requirement (notation: rectangle with «requirement» keyword) specifies a capability or condition that must (or should) be satisfied, a function that a system must perform, or a performance condition a system must achieve.
The requirements relationships include derive, satisfy, verify, refine, trace, and copy. These relationships provide a robust capability for managing requirements and supporting requirements traceability.
The requirements construct can be directly shown on use case diagrams. The use case diagram (uc) describes the functionality of a system in terms of how it is used to achieve the goals of its various users. A use case (notation: oval/ellipse) represents a system transaction with an external system user, called an Actor. A Use Case diagram shows communications among use cases and actors. Actors may represent persons, organizations, facilities, software systems, or hardware systems.
The inclusion relationship allows one use case, referred to as the base use case, to include the functionality of another use case, called the included use case, as part of its functionality when performed.
A use case can also extend another use case using the extension relationship. The extending use case is a fragment of functionality that is not considered part of the base use case functionality.

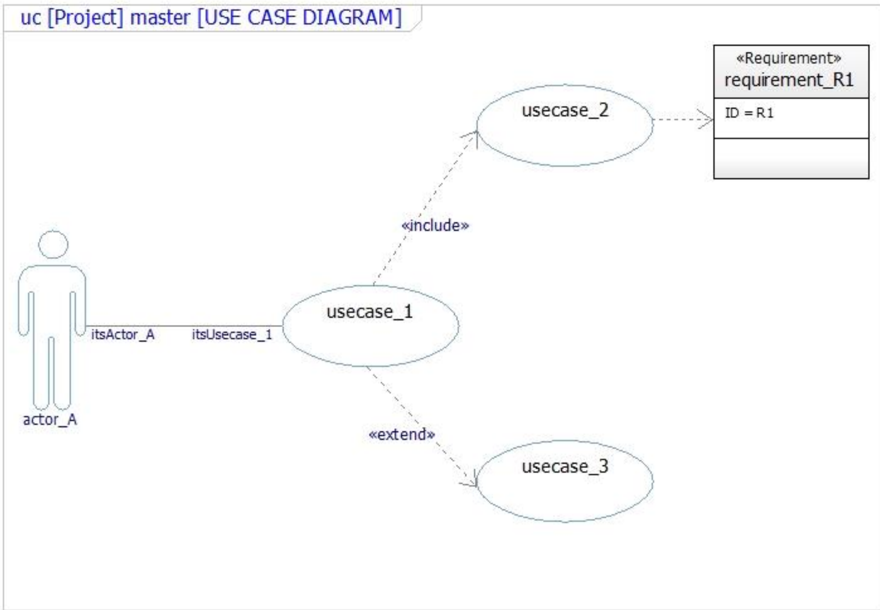


Figure 9: Use Case Diagram

3.1.2.2 Model the subsystems on block definition diagram

The block definition diagram (bdd) is used to define blocks and the relationships between them such as their hierarchical relationship. A block (notation: rectangle with keyword = «block») represents a modular structural unit that describes a system structure. It can define a type of a logical or conceptual entity, a physical entity (e.g., a system); a hardware, software, or data component. The subsystems defined in the previous stage can be modelled as blocks. Blocks have different structural and behavioural features. The structural features of a block are the properties. Properties describe a block's structural aspects in terms of its relationship to other blocks and its quantifiable characteristics. There are different forms of properties including those that represent parts, references, and values. Part properties are used to describe the composition hierarchy of a block and define a part in the context of its whole.

Value properties describe quantifiable physical, performance, and other characteristics of a block such as its weight or speed. The parameters describe in the previous stage are value properties. A value property is defined by a value type that describes its valid range of values, along with its quantity kind (e.g., length) and its units (e.g., feet or meters).

Behavioural features declare the set of services that characterize the blocks. The behavioural features of a block are operations and receptions. Operations describe synchronous interactions when the requester waits for the request to be handled. Receptions describe a synchronous behaviour when the requester can continue without waiting for a reply.

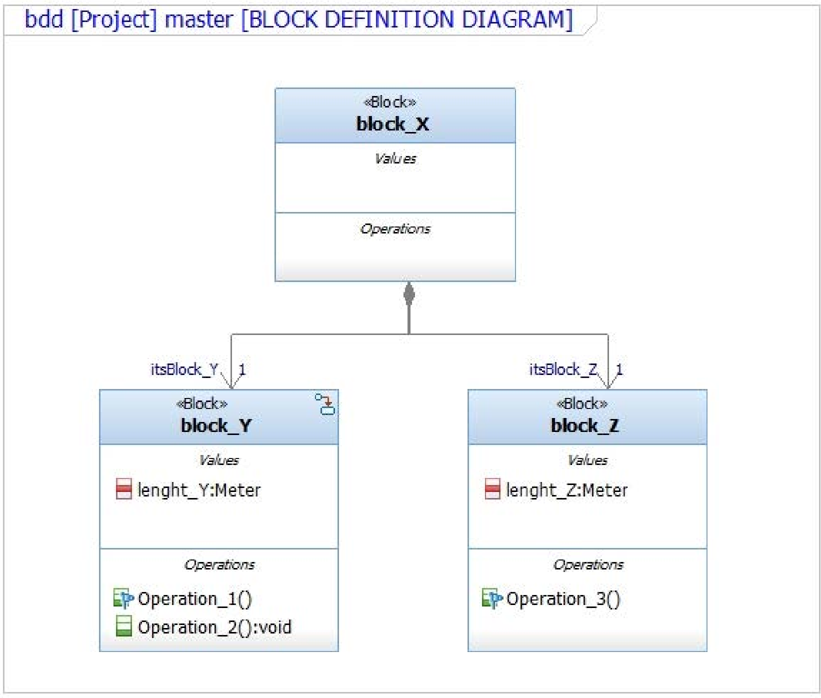


Figure 10: Block Definition Diagram

3.1.2.3 Model the relationships between subsystems on intern block diagram

The internal block diagram (ibd) is used to specify the internal structure of a single block. The ibd provides an internal vision of every single block contained in the bdd. The main elements are part properties, ports and item flows. A part property represents an internal structure of the block, while a port can be considered as an access point on the boundary of a block or between two parts. It is associated to the idea of a physical connection between two entities that communicate through an item flow. It is used to define the relationships and the connections between the subsystems defined with the exergetic analysis in terms of material and information flows. SysML has different types of ports: full port, proxy port and flow port. A flow port is used to describe an interaction point (or connection point) for items flowing in or out of a block. It is used to specify what input items can be received by the block and what output items can be sent by the block.

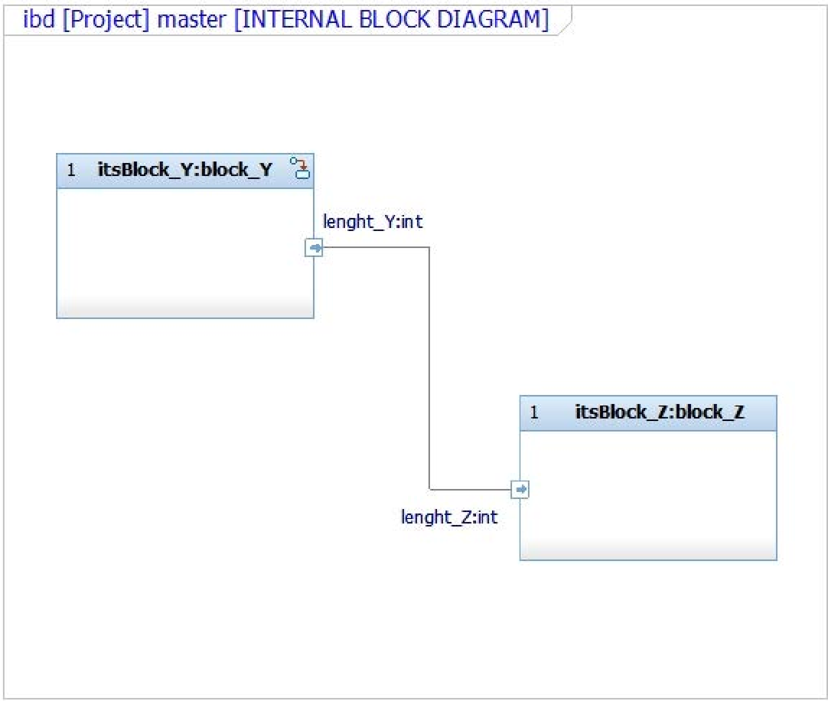


Figure 11: Internal Block Diagram

3.1.2.4 Model constraints on parametric diagram

SysML includes the concept of a constraint that can correspond to any mathematical or logical expression, including time-varying expressions and differential equations. SysML provides the ability to encapsulate the equations of the exergetic analysis in a constraint block. It can be reused and bound with other constraints to represent complex sets of equations. A constraint block defines a set of constraint parameters related to each other by a constraint expression. Parameters may have types, units, quantity kinds, and probability distributions.

The block definition diagram is used to define constraint blocks and their interrelationships. Constraint blocks can be defined in model libraries to facilitate specific types of analysis (performance, mass properties, thermal, etc.). Constraint properties are usages of constraint blocks.

The parametric diagram shows how constraint properties are connected by binding their parameters to one another and to the value properties of blocks. The binding connectors express equality between the values of the constraint parameters or value properties at their ends. In this way, constraint blocks can be used to constrain the values of block properties. The constraint blocks are different from the generic block described above, because they define only mathematical constructs and the parameters related to them.

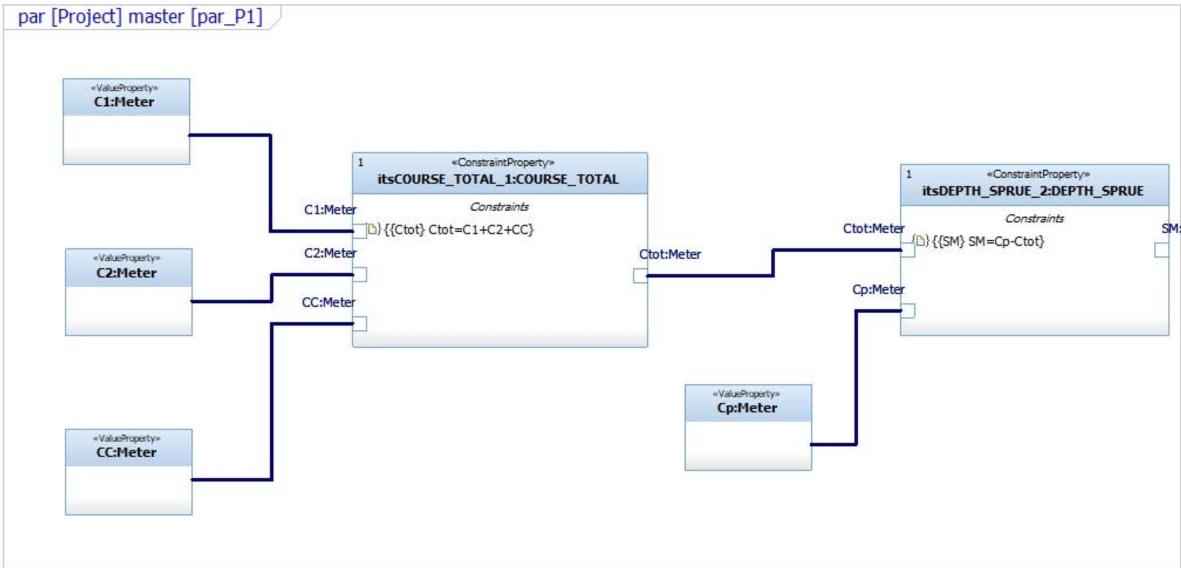


Figure 12: Parametric Diagram

3.1.2.5 Model the behaviour of the system on state machine diagram

The state machine diagram (stm) is a dynamic behavioural diagram that shows the sequences of States that an object or an interaction go through during its lifetime in response to Events (Triggers). A state (notation: rounded rectangle) represents a condition in the life of a block, and it is used to define which kind of change it undergoes, in response to the occurrence of an event or a guard condition, and what behaviours it performs. Change of state are defined by triggers and guards. The trigger indicates an event that can cause a transition from the initial state. The transition guard contains an expression (condition) that must evaluate to true for the transition to occur. The state machine diagram simulates how the states change based on internal or external events. Typically, state machines are used in SysML to describe the state-dependent behaviour of a block throughout its life cycle in terms of its states and the transitions between them.

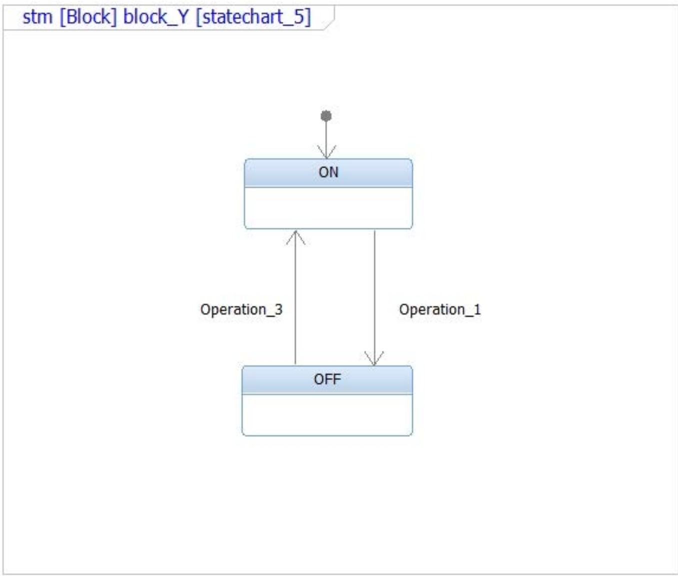


Figure 13: State Machine Diagram

3.1.3 DATA SELECTION

The previous steps enable to define which data select, their features and the system' performances (Dassisti, Siragusa, and Semeraro 2018). Data is an important element for monitoring and modelling complex systems. Data contains information useful for system optimization. The availability of knowledge about the processes, based on sound measurement systems (quite often with on-line setting) allow to clearly track the system features and system evolution. Those actions enable to collect and convert data in information, share the information acquired, formalize the knowledge, joint performance measurements and leverage the skills and the knowledge. The main idea here is that measurement is the fundamental of the knowledge model of a real system. Continued monitoring, data collection and analysis provide up-to-date information about the behaviours of the system in a continuous stream.

3.1.4 KNOWLEDGE DISCOVERY

The data selected in the previous stages are used for the knowledge discovery. Multi relation data mining (MRDM) and specifically the Relational Concept Analysis (RCA) approach is the data-driven approach selected for discovering association (Rouane-Hacene et al. 2013). RCA extends formal concept analysis (FCA) by enabling multi relational information (Rouane-Hacene et al. 2013).

In formal concept analysis (FCA), a formal context is a triple K = (G, M, I), where G and M are non-empty sets and I is a binary relation between G and M ($I \subseteq G \times M$). The formal context (G, M, I) of an input matrix of n rows and m columns consists of a set of objects defined as G = $\{g_1, \dots g_n\}$, a set of attributes defined as M=$\{m_1, \dots m_m\}$ and a binary relation I defined as $g_i, m_j \in$ I if and only if the intersection of i-th row and j-th column is not blank, i.e. object g_i has an attribute m_j (Škopljanac-Mačina and Blašković 2014).

For subsets $A \subseteq G$ of objects and subsets $B \subseteq M$ of attributes, two derivation operators (') are defined as follows:

- $A' = \{m \in M \mid (g,m) \in I \ \forall g \in A\}$, and dually
- $B' = \{g \in G \mid (g,m) \in I \ \forall m \in B\}$.

A pair (A, B) is a formal concept of a context (G, M, I) provided that:

$A \subseteq G$, $B \subseteq M$, $A' = B$, and $B' = A$.

Given a set of objects, a set of attributes, and defined the relations between objects and attributes, a formal concept represents a subset of objects sharing the same sub-set of attributes. A concept is constituted by two parts: its extension which consists of all objects belonging to the concept, and its intension which comprises all attributes shared by those objects. This understanding allows a formal discovering of associations among concepts and consequently recognizing which concepts are closely related based on the set of shared attributes (Williams and Simoff 2006). RCA extends FCA to the processing of multi-relational datasets, each provided with its own set of attributes, and relationships among those (Rouane-Hacene et al. 2013).

3.1.4.1 Data modelling

Data modelling is the process of creating a data model for the data stored in a Database. The data model illustrates the logical structure between databases by defining the entities (objects), their attributes, and showing the relationships between them.
The data selected in the previous step need to be organized in a data table as shown in Figure 14a. The data table presents the objects (G) on the rows and the attributes on the columns (M). The cross indicates that exists a relation between an object and an attribute (I). The relation between an object and an attribute is binary (exist/ does not exist).

3.1.4.2 Discovery association

RCA is applied to automatically discover association rules in data. RCA, using lattice creation algorithms, in the specific context using an improved version of "Inclose V" algorithm (Andrews 2009), converts automatically the data table into a concept lattice (Venter, Oosthuizen, and Roos 1997). Concept lattice in Figure 14b graphically portrays the underlying relationships between the objects and attributes for extracting the useful information (Wille 2002).

3.1.5 KNOWLEDGE EXTRACTION

The concept lattice, as mathematical abstraction of concept systems, support knowledge extraction. The concept lattice needs to be analysed to extract data-driven invariant modelling constructs (patterns). The data-driven constructs are evaluated with the SysML models as shown in Figure 14c to understand the meaning of their content. It aims to validate if a certain model (or hypothesis) is consistent with the available data or if it is necessary to collect new data. It means that it is possible to verify the comprehensibility, the usefulness, and the robustness of the model. At the same time, the data-driven constructs can detect tacit knowledge or discover associations to improve the model.
A data-driven invariant modelling construct contains the following sections:

- ID Pattern: unique name that helps in identifying the pattern.
- Pattern Name.
- Description: a description of the goal behind the pattern and the reason for using it.
- Example of association rule: an example of a concept for the pattern in analysis.
- Data View: a representation of the data for the pattern in analysis.
- SysML Model View: a graphical representation of the association between data. Block definition diagram or internal definition diagram or parametric or state chart diagram may be used for this purpose.
- Applicability: situations in which this pattern is usable.

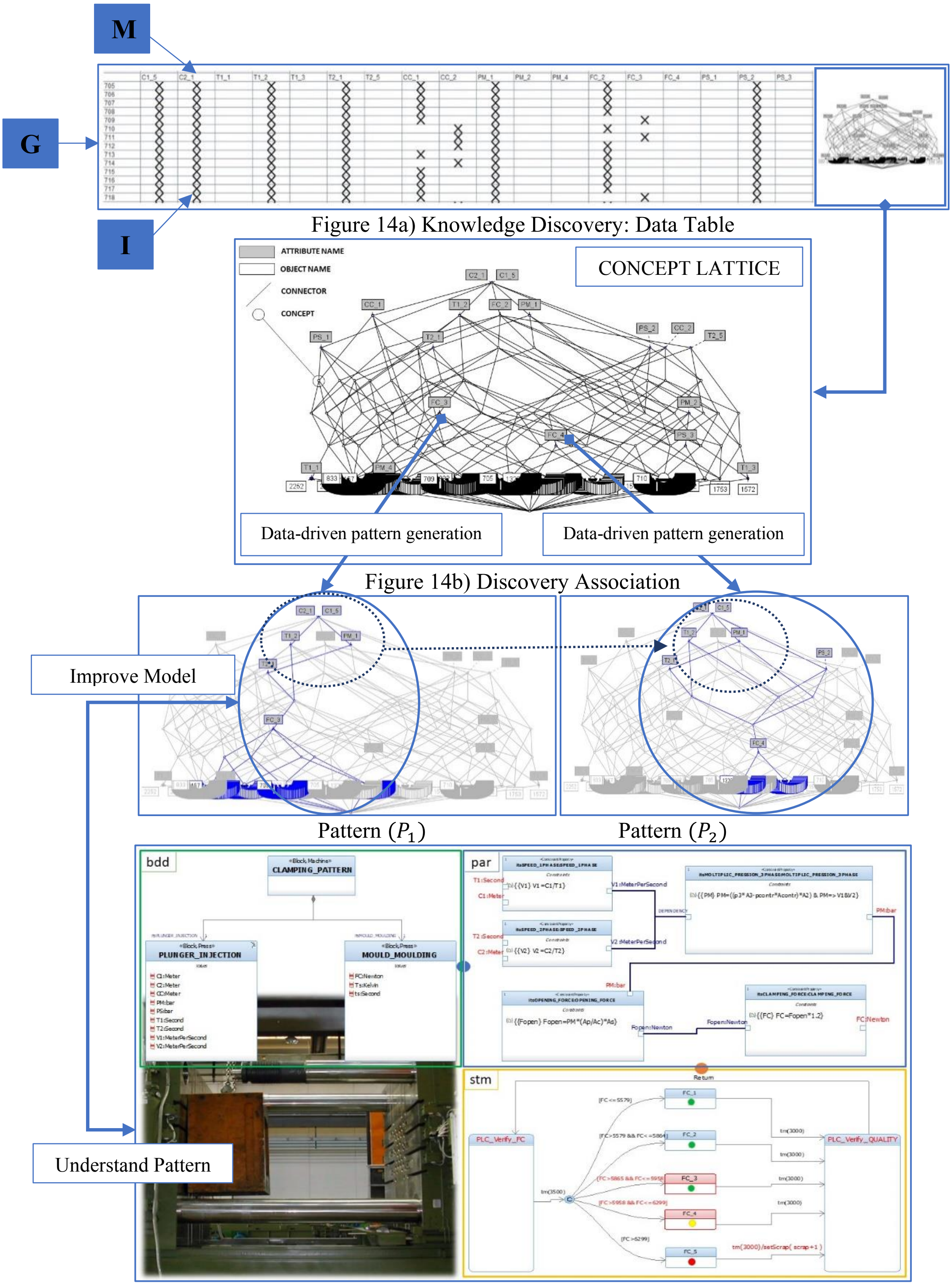


Figure 14a) Knowledge Discovery: Data Table

Figure 14b) Discovery Association

Figure 14c) SysML Model: Knowledge Extraction

Figure 14: From Knowledge Discovery to Knowledge Extraction (Semeraro, Lezoche, et al. 2019)

3.1.6 KNOWLEDGE FORMALIZATION: WEB LIBRARY OF DATA-DRIVEN CONSTRUCTS

The knowledge extracted need to be formalized to be understandable and accessible. The data warehouse system can be used for collecting, connecting, managing and analysing data from heterogeneous data sources. Structured Query Language (SQL) is the standard language that can be used for storing, manipulating and retrieving data stored in a relational database to: 1) Execute queries against database; 2) Create stored procedures in a database; 3) Create dashboards to represent the meaning of each data-driven construct. In this way, each pattern can be easily formalised, analysed and interpreted.

3.1.7 DESIGN CRITERIA FOR DIGITAL TWIN

The formalized data-driven constructs represent the design criteria for building simple and efficient end-user digital twins' interface to support the employees in decision making process. A data-driven constructs can include rules, constraints and deductions related to the production processes, such as the constraint of the processing capability of a certain equipment. These can be formalized in algorithms to make the DT able to judge, evaluate, optimize and/or predict.

3.1.8 KNOWLEDGE RE-USE

The formalized knowledge can be used directly into other systems for design digital twins or for other applications. The goal is to use the same patterns for modelling other systems. In this way, data-driven patterns can be combined, based on the specific application, to create easily dynamic models.

CHAPTER 4

CASE STUDY: MASTER ITALY DIGITAL TWIN

CHAPTER 4 - CASE STUDY: MASTER ITALY DIGITAL TWIN

INTRODUCTION

Master Italy s.r.l, is an Italian SME company that produces small hardware for civil window frames (https://www.masteritaly.com/). Master Italy realizes 97% of the added value of its in-house production, following the vertical integration strategy, pursued since its foundation.
Master Group realizes 97% of the added value of its in-house covering all phases previous to product selling: from the analysis of market's needs to design, prototyping and production of the goods: from the analysis of market's needs to design, prototyping and production of the goods. This choice, in the past, has been successful for the following reasons: a) reduced lead times, b) cost containment, c) accumulation of process know-how d) growth of the offer. Currently, however, market trends are identifying problems such as cost containment of products, introduction of new technologies, variability of market's demand, reduction in average order size, request for customized products.
In this context, the company has decided to strategically work on process and product innovation through the introduction or the development of technologies related to the Industry of the Future.
The Master's expected benefits in increasing the "smartness" of the factory are summarized as follow:

- To support people to make decisions quickly and effectively, to reduce errors, to work efficiently and ergonomically.
- To analyse the different production lines, in order to identify problems that, if compared, determine the exponential improvement in performance of the whole system
- To evaluate the possible actions to take, when the production processes are exposed to external events.
- To make decisions and make more accurate forecasts in terms of production and consumptions.
- To identify and quantify the resources that contribute to the increase in efficiency of the systems.
- To check and supervise the use of resources in the individual phases of the production process.
- To share and integrate the information among all members of the company.
- To optimize the business performance.

4.1 THE PRESENTATION OF THE CASE STUDY: CONTEXT ANALYSIS

Since 1986, Master Group has been designing accessories and components for doors and windows in aluminium, with a process made of research, investments, study of the aluminium's world, and through a deep attention to the quality of the materials, the research of advanced technologies (Master Group makes 97% of its own production's added value in house) to find new market's needs (domestic and international markets), and to customers' satisfaction and constant care (on time Delivery 95%). With a 54% foreign turnover and present in more than 58 countries worldwide, Master is nowadays a global brand, focused on: development of new international markets, attention to safety and quality of products, continuous improvement, and waste reduction according to lean manufacturing's principles.

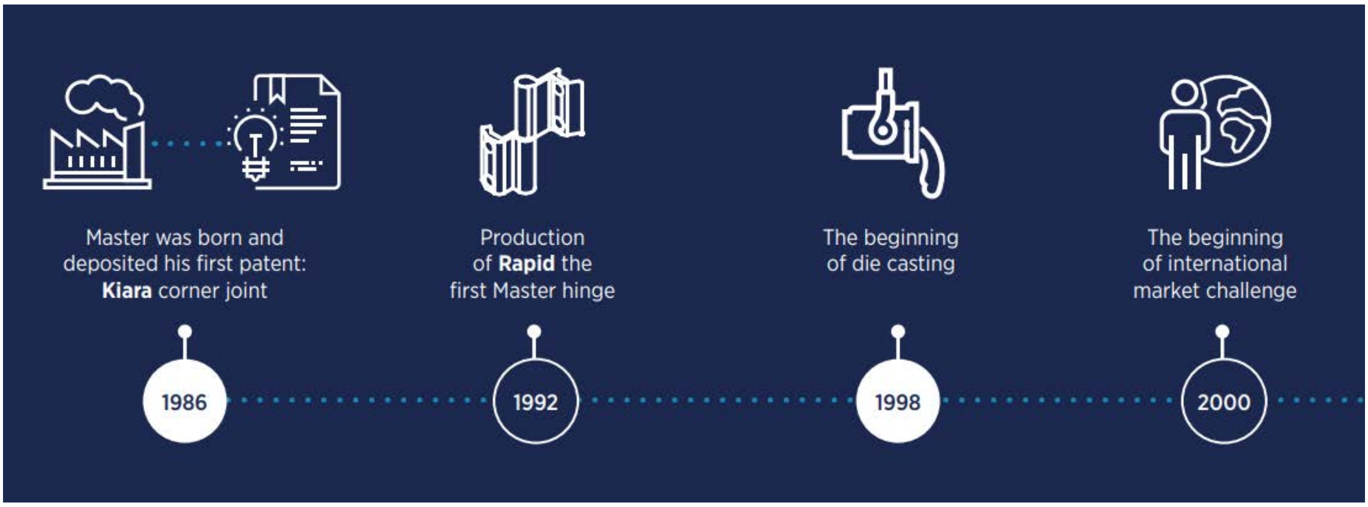


Figure 15: Master Profile I

In 2013, Master decided to implement a continuous improvement programme with the objective of guaranteeing customer satisfaction and managing the company's increasing complexity. The Master improvement programme is based on the lean thinking methodology, which aims to eliminate waste by simplifying processes, getting people involved and pushing for the creation of synchronised process flows. Activities are supported by enabling technologies: automation, integrated machines and systems that allow to maintain the right work conditions and analyse performances in real time. This is the basis to start increasingly challenging improvement initiatives. In 2018, five years after implementing the lean thinking methodology, the company has capitalised on the experience and has created the Master Italy Process System, a dynamic collection of the techniques and methods to be used in different operative environments: human resource development tools and best practices to inspire new projects. Its scope includes security and the environment, continuous improvement and innovation, digital transformation and skill development.

Figure 16: Master Profile II

Master Italy s.r.l production site is made of by 72 machines distributed over 15 processing departments, characterized by heterogeneity of competence and material: 1) Aluminium die casting, 2) Zama die casting,

3) Plastic moulding, 4) Cold steel moulding, 5) Painting, 6) Production and maintenance department for equipment and moulds 7) Drilling and threading, 8) Washing and tumbling, 9) Handle assembly, 10) Hinges assembly, 11) Tilt and Turn door assembly 12) Squadrons assembly to which are added 13) Maintenance department, 14) Internal logistics department and 15) Finished products departments.

The supply chain organization can be divided into 2 macro blocks. The first is the processing of raw materials (aluminium die-casting, zamak die-casting, plastic moulding, steel shearing, aluminium shearing, shot blasting, washing and tumbling, drilling and threading), characterised by a large production capacity, large production batches, a large number of machines and equipment. The second processing block are painting (and other surface finishes, carried out outdoors) and automatic and manual assemblies' lines, characterized by a great variability of production batches.

The continuous improvement for a company depends on the assets: organization, process and technologies. The Lean Manufacturing, process undertaken by the company since 2013, has made possible to structure the organization in the logic "value stream" structure categorising the manufacturing activities in value-added and not-value added. The absence of technological solutions on the production lines caused problems related to:

- Management of information (generation and storage)
  - Not timely detection of the information
  - Low accuracy of information (manual entering and transmission of info)
  - The fragmentation of information, coming from different sources (frequently on different supports), because lack of an integrated management system.
- Use of information:
  - Production info are not automatically updated.
  - No real time info of the production status is available.

This gives rise to the idea of automating the exchange and the tracking of information from the transformation of raw materials to finished goods. For this reason, the Sedapta MES platform (https://www.atomoshyla.com/it/) has been introduced in Master since 2017 to optimize, control and monitoring the manufacturing processes achieving:

- Real time processing data: MES collects the data gathered on the shop-floor, analyses data, and extracts the information necessary to provide an exhaustive picture real time of the current state of the process.
- Advanced scheduling: MES identifies the optimal sequence-planning considering the constraints of the process, like cycle time and setup, and the capacity of the workstations, and customers' requirements. The system also allocates resources like material and staff. The knowledge is shared and in this way all employee can plan and optimize the production. Thanks to this kind of technology, employees learn the technology and learn through the technology.
- Asset monitoring: The monitoring of production lines is performed in real-time and is displayed in graphic dashboards through mobile devices like tablets and smartphones.

A MES system does not develop systems capable of learning autonomously from their own behaviour in order to achieve predictive manufacturing system (PMS). The predictive manufacturing is defined as a system that *enables machinery and systems with "self-aware" capabilities with greater transparency for users to avoid potential issues* (J. Lee et al. 2013).

It requires the integration and the convergence between the cyber and the physical system i.e. the construction and the application of digital models to emulate, analyse and simulate the evolution of their real twins helping to prevent problems or to improve the performance through the real time analysis of data. The business strategy of Master for the Industry of the Future is to create digital twins able to estimate the actual condition of the processes, infer future fault events and even predict potential root cause of the problem.

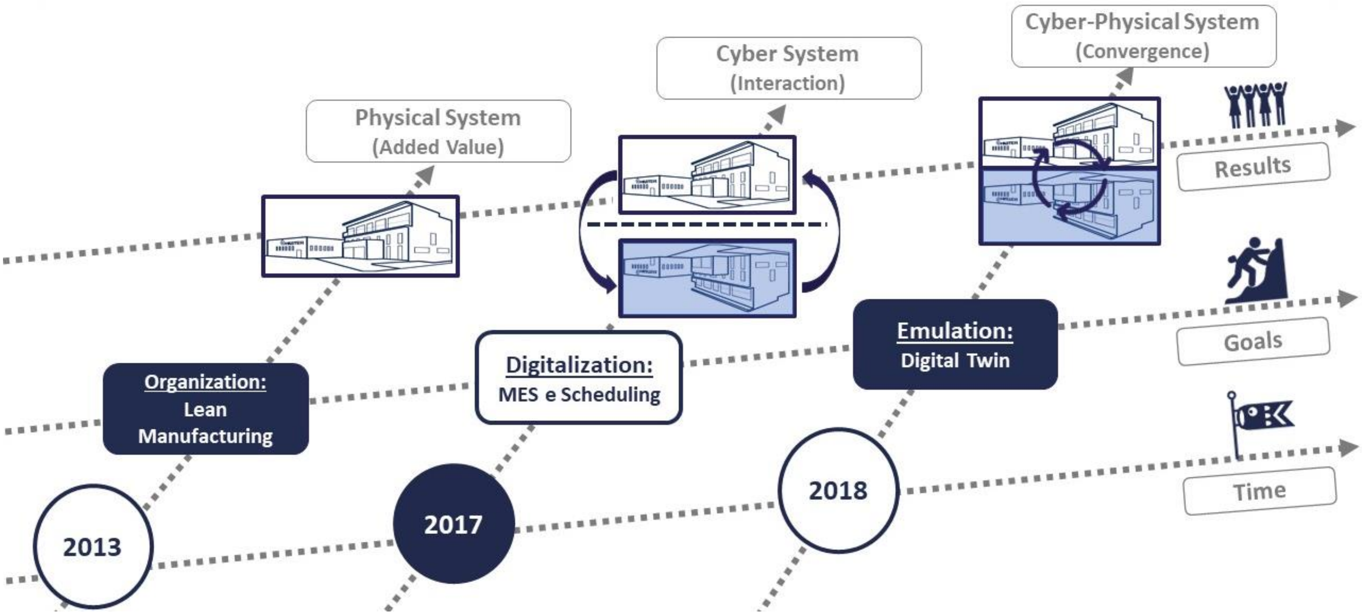


Figure 17: Master' Strategy

4.2 APPLICATION OF THE APPROACH TO THE INDUSTRIAL CASE TO DETECT DATA-DRIVEN INVARIANT MODELLING CONSTRUCTS

The approach discussed in the chapter 3 has been applied to design and to develop a digital twin, based on data-driven invariant modelling constructs, to:

- Simplify the programming and the control of production.
- Increase the productivity.
- Eliminate the errors during the operations.
- Reduce the operator's learning curve.

It is useful to prove the quality of the idea and meet the requirements above discussed.

4.2.1 DEFINITION OF THE SYSTEM

The implementation of the hybrid exergetic analysis-LCA approach aims to define the critical product and the critical system in Master and then to perform the exergetic analysis.

4.2.1.1 Identification of the system to analyse

The LCA analysis was drafted in Master with the aim of calculating the Global Warming Potential (GWP) over 100year of each product (Dassisti and Semeraro 2016). The objective of the LCA analysis is to

evaluate the amount of resources needed and the emissions produced to manufacture the various components. LCA was performed used the SIMAPRO® software using the Eco invent database.
The most important products of the company are hinges, steel corner, handles and tilt and turn used to aluminium windows. Thanks to the LCA assessment, the selected product, in Figure 18, is the steel corner since the GWP100 impact is: 0,282115 kg CO_2eq/pcs.
The different steel corner components undergo several mechanical processes: die casting aluminium, die casting zamak, varnishing, and assembly. Analysing each production process, the greatest contribution is given by the die casting aluminium process (1,1395 kg CO_2eq/pcs due to the methane gas consumption).

Figure 18: Steel Corner

Figure 19: Die Casting Aluminium Shop-floor

4.2.1.2 Application of the exergetic analysis

The LCA analysis performed allowed to measure the critical system (product and process) in terms of resource consumption and pollutions. This information was then used to provide input to the exergy analysis and split the selected system into different subsystems to identify the critical sub-system, evaluating the contribute of the exergy loss, and to assess their criticality the critical parameters to control.
Die casting aluminium is a manufacturing process in which molten metal is poured or forced into steel moulds. The moulds, also known as tools or dies, are created using steel and are specially designed for each project. The total cycle time is short, typically around 33-35 seconds (Dassisti, Semeraro, and Chimenti 2019). The process cycle of die casting aluminium consists of four main subsystems, which are explained below:

1. Melting: the aluminium enters at the solid state and exits at the molten state.
   Die casting requires that aluminium is heated well into its liquid phase for injection. The melting point of aluminium is: 680-700 °C (T_{al}). Once melted and taken up to proper temperature, the

aluminium is transferred to each die cast machine. Each die cast machine has its own holding furnace which maintains the molten aluminium at temperature waiting for use in the die cast machine.

2. Injection: the molten aluminium is injected into the mould, through a plunger.
The molten metal, which is maintained at a set temperature in the furnace, is next transferred into a chamber where it can be injected into the die. When a die cast machine is ready for its next cycle (die is closed ready for shot), an automated ladle takes a prescribed volume of molten aluminium (V_{al}) from the holding furnace and pours it into the mould. Once pouring is complete, the injection phases begin. The first injection phase is the slow phase (T_1) where the plunger moves forward at a low speed (V_1). After a prescribed distance (C_1), the plunger enters (C_2) an intermediate speed phase (V_2) where the speed is increased to fill the mould. Once this is complete (T_2), the machine enters a fast phase where speed is greatly increased to fill the part cavity with aluminium (CC).

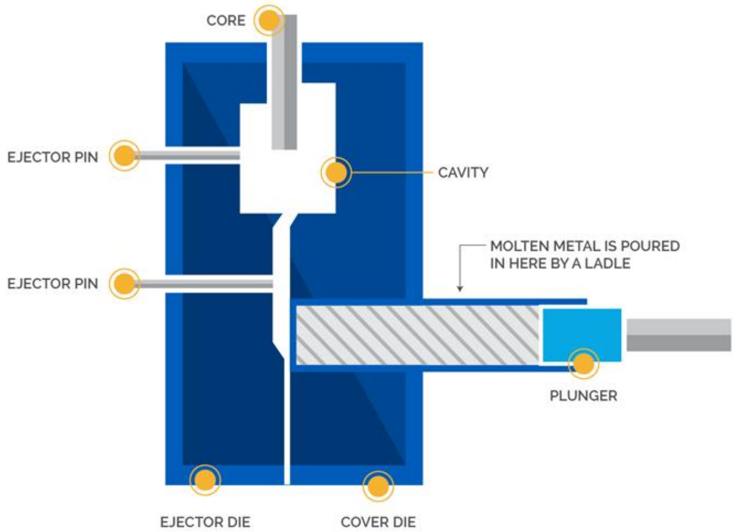


Figure 20: Injection stage (https://www.dynacast.com/en/specialty-die-casting/die-cast-process/cold-chamber-die-casting)

3. Cooling: the molten aluminium solidifies in the mould cavity.
After the part cavity is filled and the plunger has stopped moving, the hydraulic cylinder pushing the plunger is pressurized to a higher pressure (PM). This pressure holds the molten metal in the dies during solidification. When the entire cavity is filled and the molten metal solidifies, the final shape of the casting is formed. The die cannot be opened until the cooling time has elapsed and the casting is solidified. Clamping force (FC) must be applied to the die to keep it securely closed while the metal is injected. After a prescribed amount of time (TC), the die opens the ejector or moving half of the die.

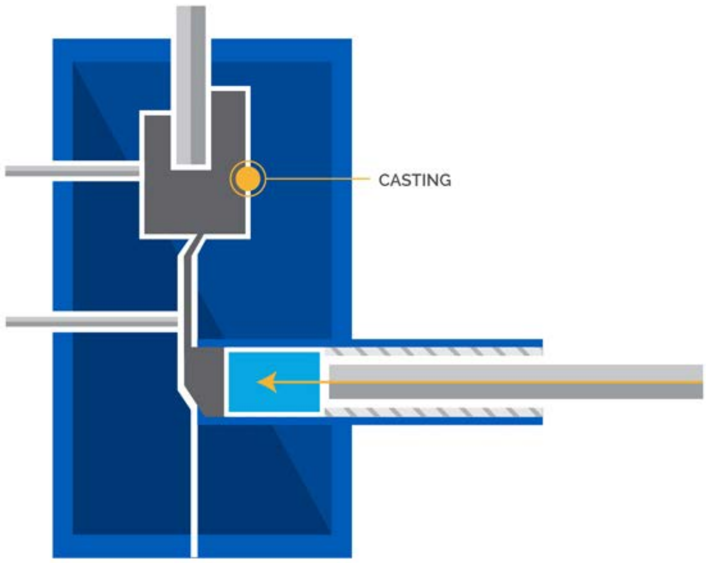


Figure 21: Cooling Stage (https://www.dynacast.com/en/specialty-die-casting/die-cast-process/cold-chamber-die-casting)

4. Extraction: an ejection mechanism pushes the product out of the mould cavity.
   Once the injection cycle is completed and the machine is fully open, the die cast pushed out and the thickness of the die cast (SM) is controlled to prevent quality defects. A die cast represents an injection cycle. The die cast contains 36 steel corners. It means that 36 steel corners are produced for each injection cycle every 33 seconds.

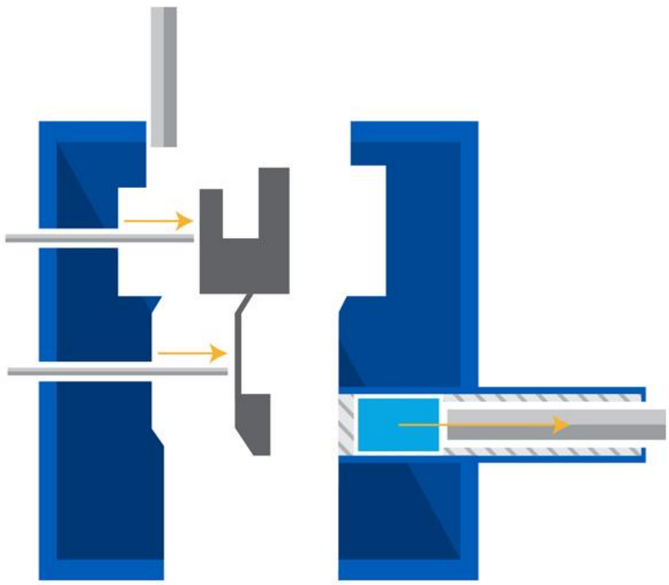


Figure 22: Extraction Stage (https://www.dynacast.com/en/specialty-die-casting/die-cast-process/cold-chamber-die-casting)

4.2.1.3 Performing the exergy analysis of each subsystem to compute the exergy loss Ex_{loss}

To perform exergy analysis production process, the system is intended as composed of several subsystems, where each subsystem is characterized by a corresponding input and output flows. For each subsystem the amount of exergy destroyed is reckoned with following equation:

$$Ex_{M,in} + Ex_{W,in} + \left(1 - \frac{T_0}{T_{in}}\right) Q_{in} = Ex_{M,out} + Ex_{W,out} + \left(1 - \frac{T_0}{T_{out}}\right) Q_{out} + Ex_{loss} \quad (3)$$

where the term $\left(1 - \frac{T_0}{T_{in,out}}\right) Q_{in/out}$ represents the exergy flow of input and output heat (Ex_Q).
The term exergy loss (Ex_{loss}) depends (see equation 3) on the variation between input and output of material exergy (Ex_M), work exergy (Ex_W), and heat exergy (Ex_Q) and it is responsible for the less-than-theoretical efficiency of the system (Semeraro, Panetto, et al. 2019).
For die casting aluminium, the exergy of material, work, and heat are reckoned in each subsystem (i) with the following equations:

$$Ex_{M,i} = e_i^{ph} = \rho_{al} \cdot V_{al,i} \cdot \left[(h_{al,i} - h_0) - T_0 \cdot (s_{al,i} - s_0)\right] \quad (4)$$

$$Ex_{W,i} = F_{inj,i} \cdot C_i = (P_i \cdot S_{inj}) \cdot C_i \quad \textit{where the stroke of the piston determines } C_i = V_i \cdot t_i \quad (5)$$

$$Q_i = Q_{f,i} + Q_{l,i} + Q_{c,i} = c_{s,al,i} \cdot \rho_{al} \cdot V_{al,i} \cdot (T_{f,i} - T_{s,i}) + c_{l,al,i} \cdot \rho_{al} \cdot V_{al} + c_{s,al,i} \cdot \rho_{al} \cdot V_{al} \cdot (T_{al,i} - T_{f,i}) \quad (6)$$

The application of the exergetic analysis shows that the cooling phase is the critical subsystem because the exergy loss is highest than other subsystems (Subsystem 3: 8.649 J) as shown in figures 23 and 24.

4.2.1.4 Definition of the critical thermodynamic parameters to measure

Analysing the exergetic analysis equations it is possible to define the critical thermodynamic parameters, identifying:

- Main parameters already controlled: $V_{al}, T_{al}, T_f, C_1, C_2, CC, T_1, T_2, PM, FC, SM, TC$
- Main parameters not yet controlled are: $V_l, T_s, T_l, T_0, P_0, T_{mo}, P_{a1}, P_{a2}, P_{aM}$
- Derived parameters controlled are: V_1, V_2, PS, PF
- Non-controllable parameters are: $\rho_{al}, c_{p,al}, c_{l,al}, h_0, h_{al}, s_0, s_{al}$

The process knowledge extracted can be modelled for collecting information for the control of the production process. The parameters identified for each sub-system; the thermodynamic model built resulted to be key informational elements for the correct control of the core business of the company.
The main controlled parameters are the parameters related to the volume and the temperature of the aluminium (V_{al}, T_{al}), the melting temperature of the aluminium (T_f), the course of the plunger in the first, second and multiplied phase of the injection stage (C_1, C_2, CC), the time of mould filling in the injection stage (T_1, T_2), the multiplied pressure (PM), the clamping force (FC), the thickness of the product (SM) and the cycle time (TC).
The main parameters to control are the volume of aluminium loss during the injection stage (V_l), the temperature of the solid ingots at the entry of the furnace (T_s), the temperature loss during the injection

stage (T_l), the environmental temperature and pressure (T_0, P_0), the temperature of the mould (T_{mo}), the pressure of the accumulator in the first, second and multiplied stage (P_{a1}, P_{a2}, P_{aM}).
The derived parameters controlled are the speed in the first and in the second phase of the injection stage (V_1, V_2) and the specific and final pressure (PS, PF).
The non-controllable parameters are the density of aluminium (ρ_{al}), specific heat and latent heat of aluminium ($c_{p,al}, c_{l,al}$), specific enthalpy of environment and aluminium (h_0, h_{al}), specific entropy of environment and aluminium (s_0, s_{al}).
The combined approach is a perfect way of structuring the process knowledge for assuring a correct transition toward the smartness at job shop level.

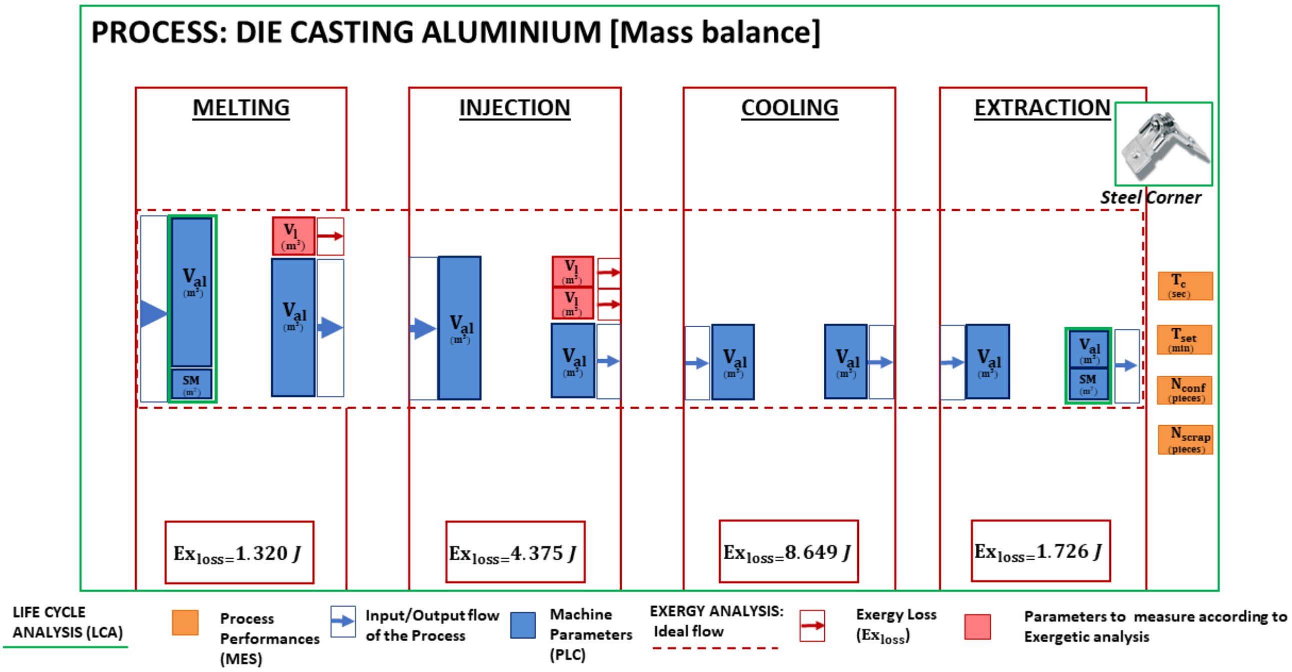


Figure 23: Mass Balance (Semeraro, Panetto, et al. 2019)

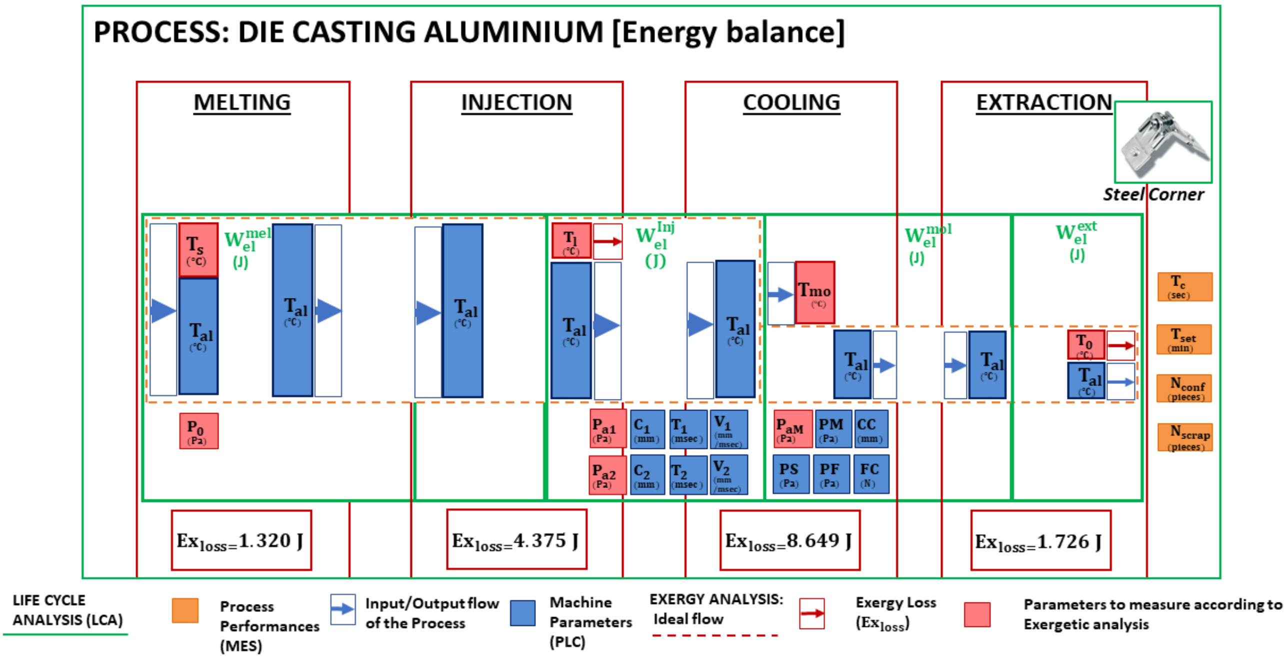


Figure 24: Energy Balance (Semeraro, Panetto, et al. 2019)

4.2.2 SYSTEM MODEL

SysML aims at formalising the following aspects of the system: 1) Structural composition, interconnection, and classification; 2) Constraints on the physical and performance properties; 3) Function-based and state-based behaviour; 4) Allocations between behaviour, structure, and constraints.

4.2.2.1 Model requirements and use case

The use case diagram in Figure 25 depicts some of the high-level functionality of the die casting process. The PLC and the MES are the software that currently control the production line as shown in Figure 23 and 24. PLC is an industrial digital computer which has been used for the control of technological parameters. Manufacturing Execution System (MES) is an information system which monitors the productivity and tracks machine down-time. ‘Monitor technological parameters’ and ‘Monitor machine down-time’ are the use cases, represented by the oval. PLC and MES are the actors, represented as the stick figure.

SysML provides the ability to specify inclusion or extension relationships between use cases. The ‘Correlate technological parameters to machine down-time’ use case include ‘monitor technological parameters’ and ‘monitor productivity’ use cases. The ‘Correlate technological parameters to machine down-time’ use case represents common functionality that is always performed when ‘monitor technological parameters’ and ‘monitor machine down-time’ are performed. The ‘monitor technological parameters’ and ‘monitor machine down-time’ are referred to as the base use cases, and ‘Correlate technological parameters to machine down-time’ is referred to as the included use case.

The requirements are often related to use cases. One approach is to use a refine relationship. An example is provided in Figure 25 where the requirement predictive manufacturing is refined by the use case ‘Correlate technological parameters to machine down-time’. The correlation between technological parameters and machine down-time allows to better understand the behaviour of the system for realising the predictive manufacturing.

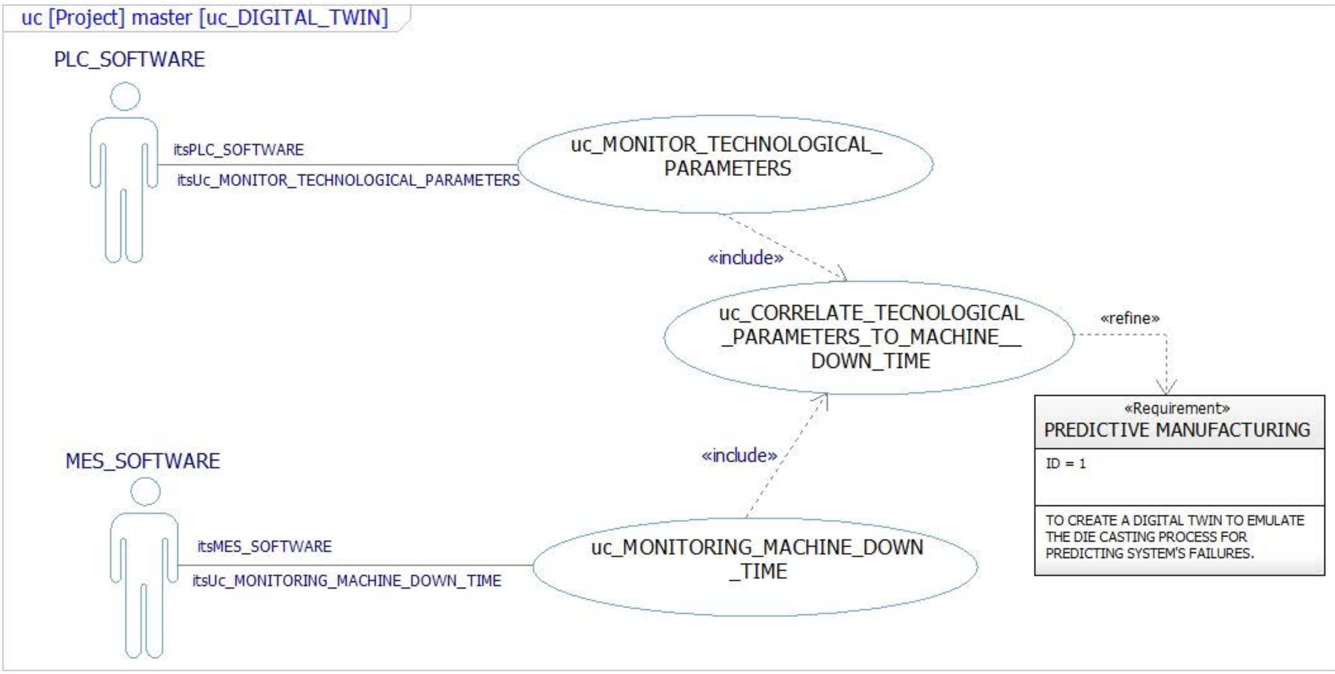


Figure 25: Use Case Diagram

4.2.2.2 Model subsystems on block definition diagram

SysML is applied to design and modelling the die casting aluminium process. The exergetic analysis split the die casting aluminium in four main subsystems: melting, injection, cooling and extraction.
The subsystems are modelled as blocks in the block definition diagram (bdd) as shown in Figure 26. The block definition diagram captures the relation between the blocks and their interrelationships. It is often used in systems modelling to depict multiple levels of the system hierarchy from the top-level domain or context block (e.g., machine F55) down to the blocks representing the subsystems of the process.
The die casting process (machine F55) is the top-level block in the block definition diagram. The block is composed of other blocks as indicated by the black diamond symbol and lines with the arrowhead pointing to the blocks that compose it. This whole-part relationship is called a composite association. The melting, injection, cooling and extraction subsystems are subclasses of die casting aluminium as indicated by the hollow triangle symbol.
A block is used to model entities and it can have a set of features. The features of the block are the value properties. The melting subsystem is a block that has three properties for aluminium temperature (T_{al}), melting temperature (T_f), and aluminium temperature at the solid state (T_s) with unit of Celsius.
Similarly, the injection subsystem is a block that has six properties for course of the plunger in the first phase (C_1) and in the second phase (C_2) of the injection stage, the time of injection in the first (T_1) and in the second phase (T_2) and the speed in the first (V_1) and in the second phase (V_2). Each property has the respective unit of measurement.
The cooling subsystem is a block that has five properties for multiplied course (CC), multiplied pressure (PM), specific pressure (PS), final pressure (PF), clamping force (FC).

The extraction subsystem is a block that has four properties for thickness of product (SM), cycle time (TC), conformed ($N_{Conformed}$) and discarded quantity (N_{scrap}).

These properties are used along with other properties to support the analysis of constraints. The block definition diagram presents constraint blocks to model the equations and the associated parameters needed to analyse the subsystems as defined with the exergetic analysis.

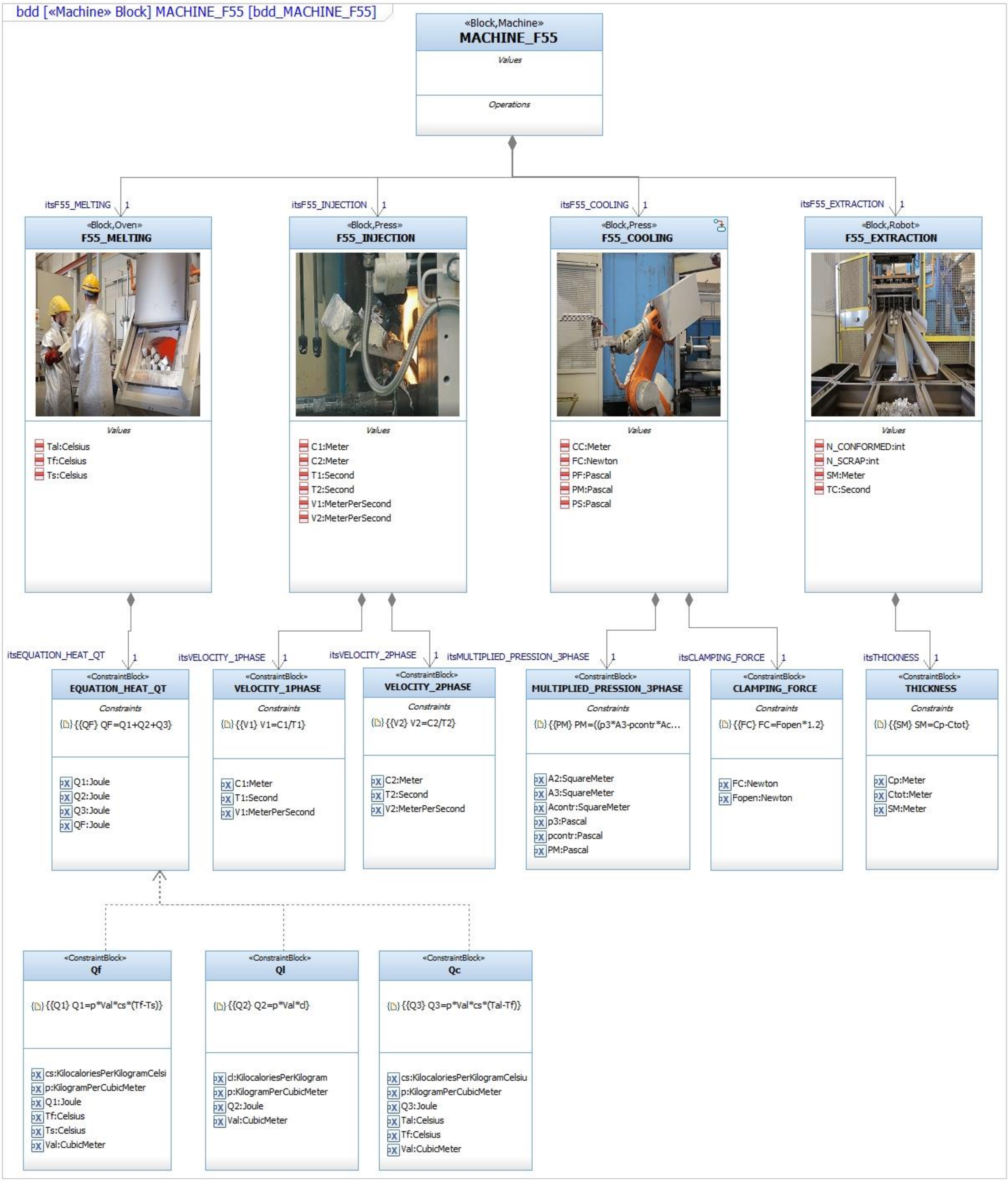


Figure 26: Bdd Diagram of Process

The operating parameters and the performances of die casting process are controlled by programmable logic controller (PLC) and manufacturing execution system (MES). The block definition diagram in Figure 27 shows which parameters (value properties) are monitor by the programmable logic controller (PLC) and which are monitor by the manufacturing execution system (MES).

The programmable logic controller (PLC) monitors and records run-time the following operating parameters: the univocal ID of the injection cycle (MEASSETID), the name of the article to product (ARTICLECODE), the name of the machine (RESOURCENAME), data and time for each MEASSETID (TIMESTAMPLOCAL), the number of injection cycle for a single production order (N_INJECTION).

The PLC monitors and records run-time the following technological parameters: the course of the plunger in the first phase (C_1) and in the second phase (C_2) of the injection stage, the time of injection in the first (T_1) and in the second phase (T_2), the speed in the first (V_1) and in the second phase (V_2) of the injection, the multiplied course (CC), the multiplied pressure (PM), the specific pressure (PS), the final pressure (PF), the clamping force (FC), the thickness of product (SM) and the cycle time (TC).

The PLC automatically start and stop the process, generates alarms if a machine malfunction, based on pre-programmed values of each parameters. Each parameter has a minimum value (MIN) and a maximum value (MAX) that are listed as value properties, in the block PLC, as follow: MIN_C1, MIN_T1, MIN_V1, MIN_C2, MIN_T2, MIN_V2, MIN_CC, MIN_PM, MIN_PF, MIN_PS, MIN_FC, MIN_SM, MIN_TC, MAX_C1, MAX_T1, MAX_V1, MAX_C2, MAX_T2, MAX_V2, MAX_CC, MAX_PM, MAX_PF, MAX_PS, MAX_FC, MAX_SM, MAX_TC.

MES system enables the control of multiple elements of the production process (e.g. inputs, personnel, machines and support services) to monitor the production output and performances. The parameters listed as value properties, in the block MES, are: univocal ID (ID_RKD), day of production (DAY_SHIFT), day and time (DATA_TIME), possible state of the machine (STATE), possible activities of the machine (ACTIVITIES), name of the machine (RESOURCE), name of the employee (WORKER), id of the production order (PRODUCTION_ORDER), name of articles to product (ARTICLE), duration time of a setup (SETUP_TIME), duration time of production (UPTIME), duration time of maintenance down-time (MAINTENANCE_TIME), duration of a downtime (DOWNTIME), quantity of conformed products (CONFORMED_QUANTITY), quantity of discarded products (DISCARDED_QUANTITY), number of injection cycles during the production time (N_MOULDED), univocal id of a machine down-time (ID_ MACHINE_DOWN_TIME), description of a possible machine down-time (DESCRIPTION_ MACHINE_DOWN_TIME), shift (morning/afternoon/evening) (N_SHIFT), number of products for each injection cycle (N_PIECES).

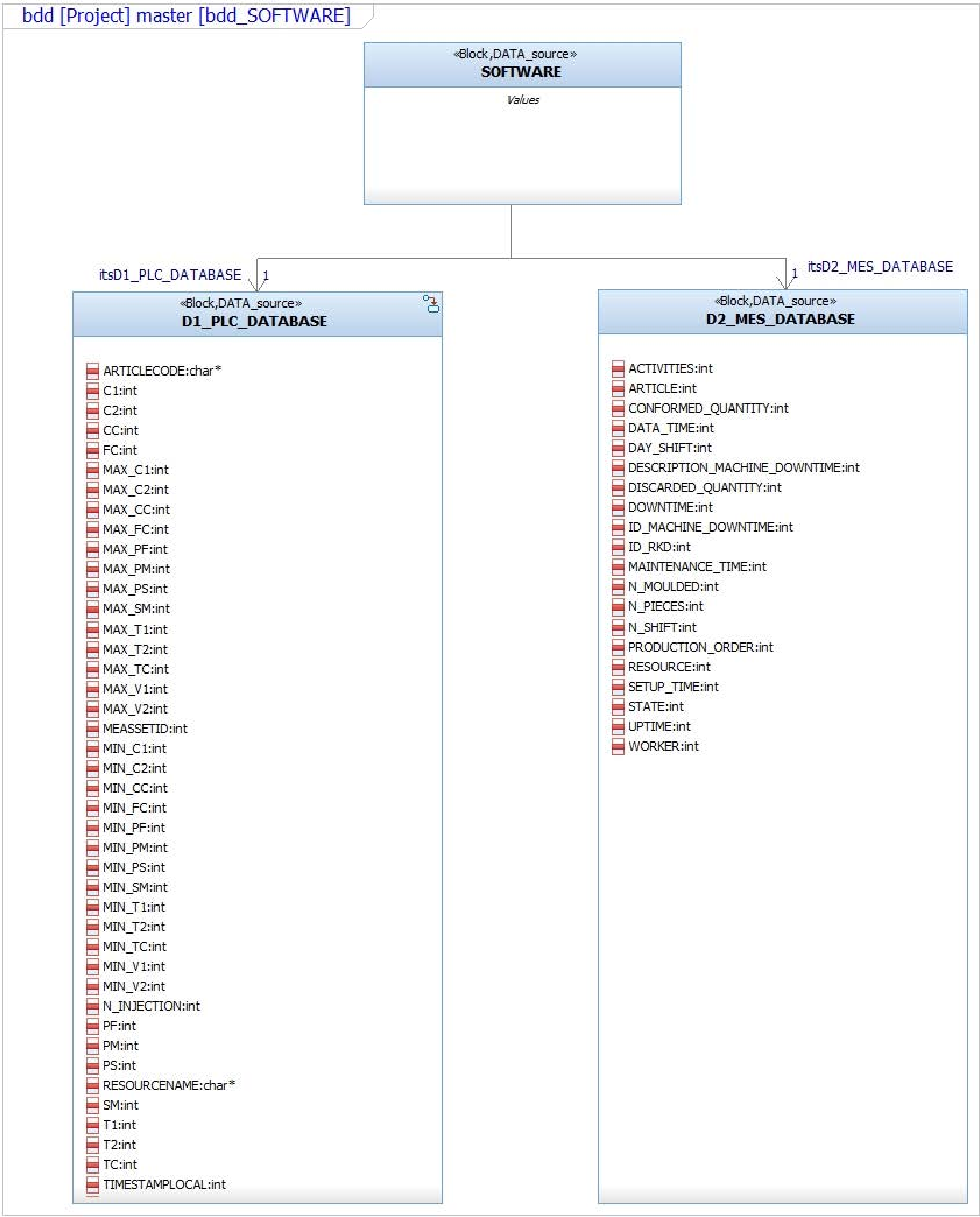


Figure 27: Bdd Diagram of Data Source

4.2.2.3 Model the relationships between subsystems on intern block diagram

The internal block diagram in Figure 28 shows how the different parts of the block definition diagrams interact.

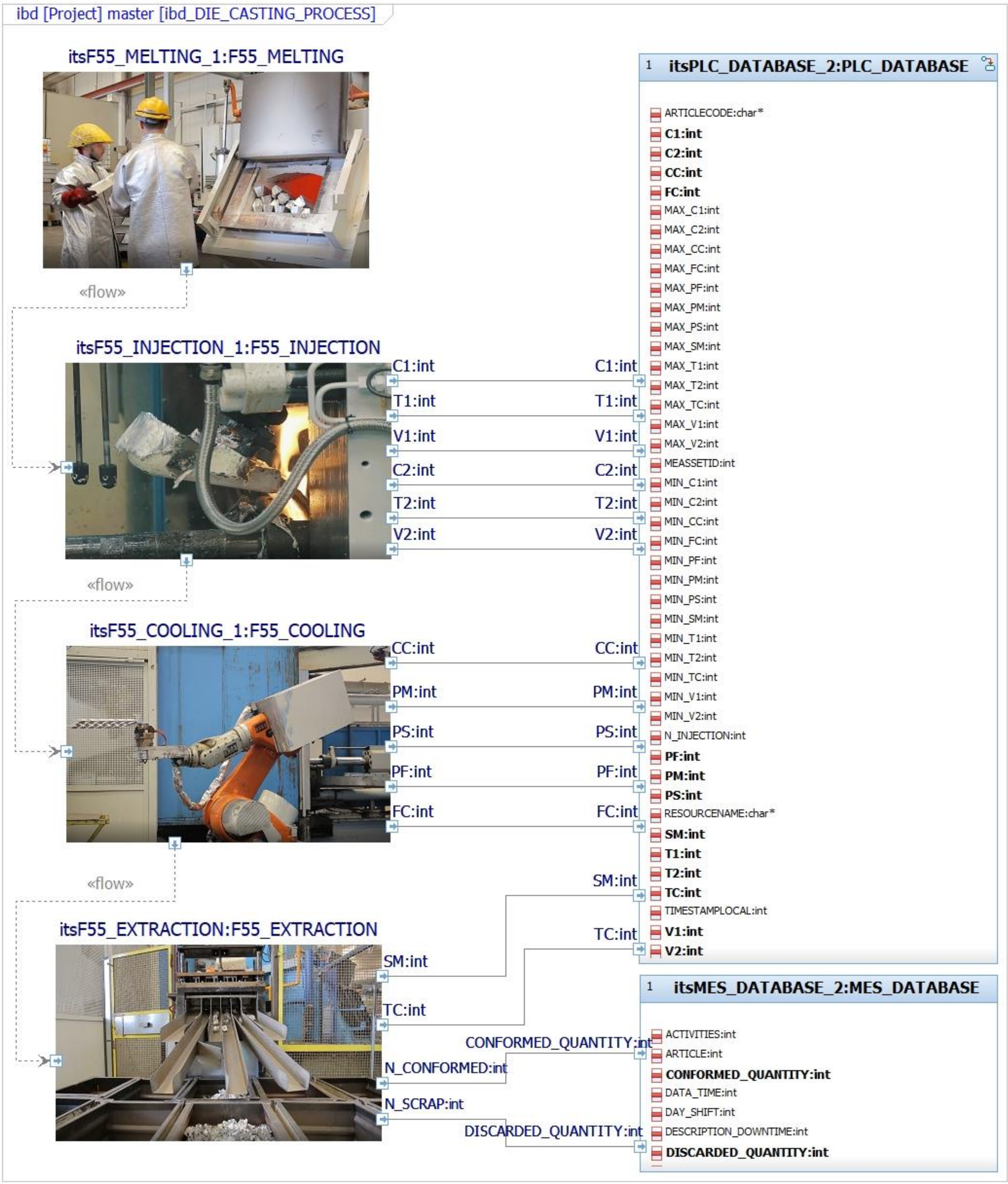


Figure 28: Ibd Diagram

The flow ports are represented as the small squares on the boundary of the parts and specify interfaces with other parts. Connectors are represented as lines between the ports and define how parts connect to one another. The internal block definition diagram shows the interactions between subsystems and between a subsystem and a data source (PLC and MES).

The system starts out with a data acquisition system using appropriate sensor to record signals such as course, pressure, speed, temperature etc. The programmable logic controller (PLC) receives technological data from connected sensors or input devices. It processes the data, and triggers outputs based on pre-programmed parameters (MIN and MAX values for each parameter) to monitor the injection, cooling and extraction phases. The manufacturing execution system (MES) receives production data to monitor the performance, the productivity of the process and the down time.

4.2.2.4 Model constraints on parametric diagram

The parametric diagram in Figure 29 shows a network of constraints (equations). Each constraint is a usage of a constraint block defined in the block definition diagram in Figure 26. The constraint parameters of the equation are shown as small rectangles flush with the inside boundary of the constraint. A parameter in one equation can be bound to a parameter in another equation by a binding connector. In this example, only a few of the die casting equations are shown. The parametric diagram and related modelling information can be provided to the appropriate simulation and/or analysis tools to support execution (e.g. Simulink). This engineering analysis is used to perform sensitivity analysis and determine the property values that are required to satisfy the correct behaviour of the process.

For example, the constraint: $V_2 = \frac{C_2}{T_2}$ is modelled as shown in Figure 29. The constraint represents the rapport between the course of the plunger in the second phase and the time of the injection to calculate the speed of the plunger in the second phase.

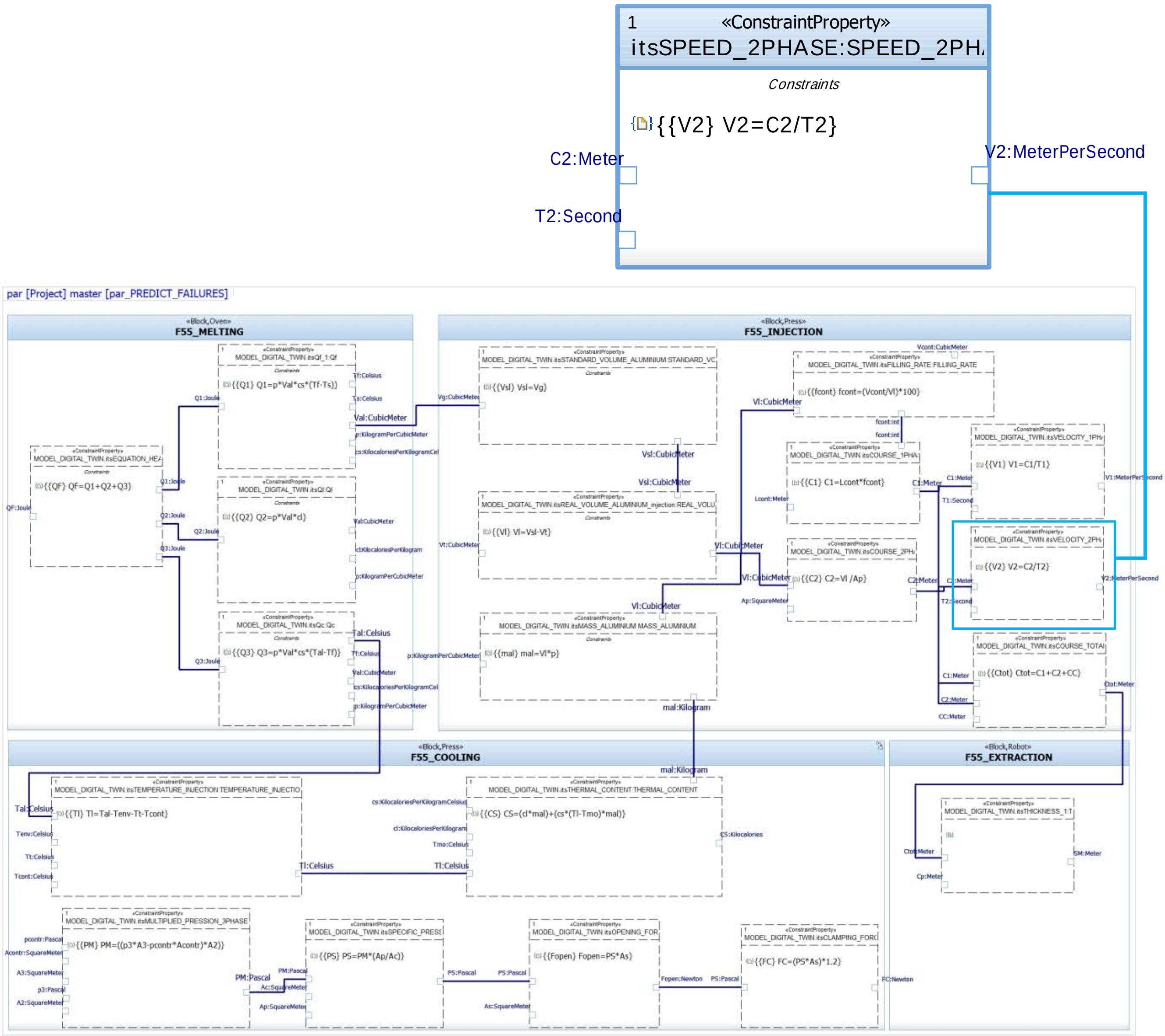


Figure 29: Parametric Diagram

4.2.2.5 Model the behaviour of the system on state machine diagram

The PLC processes the data, and triggers outputs based on the setting values of each technological parameter as shown in Figure 27. The process can assume the following operating conditions:

- RECASTING (red): the parameter is out of range and it causes the recasting of the injection cycle (≈ 36 products). V_2, PM, FC, SM, TC are the critical parameters because If V_2, PM, FC, SM, TC are equal to or below the minimum value or greater than or equal to the maximum value there is the recasting of the injection cycle. A recasting is recognized as scrap (DISCARDED_QUANTITY) by the MES.
- NOT CONFORMED (red): the parameter is out of range, but it does not generate the recasting of the injection cycle.
- CHECK (yellow): the parameter is closes out of range.
- CONFORMED (green): the parameter operates into the defined range.

The lower limit and an upper limit define five different state for each parameter as shown in Table 1:

Table 1: Parameters of PLC Database

PARAMETER	DATA TYPES	RANGE	STATE (S)	S
C1- Course First Phase	NUMBER(10)	C1_1	NOT CONFORMED	
C1- Course First Phase	NUMBER(10)	C1_2	CHECK	
C1- Course First Phase	NUMBER(10)	C1_3	CONFORMED	
C1- Course First Phase	NUMBER(10)	C1_4	CHECK	
C1- Course First Phase	NUMBER(10)	C1_5	NOT CONFORMED	
T1- Time First Phase	NUMBER(10)	T1_1	NOT CONFORMED	
T1- Time First Phase	NUMBER(10)	T1_2	CHECK	
T1- Time First Phase	NUMBER(10)	T1_3	CONFORMED	
T1- Time First Phase	NUMBER(10)	T1_4	CHECK	
T1- Time First Phase	NUMBER(10)	T1_5	NOT CONFORMED	
V1- Speed First Phase	NUMBER(10)	V1_1	NOT CONFORMED	
V1- Speed First Phase	NUMBER(10)	V1_2	CHECK	
V1- Speed First Phase	NUMBER(10)	V1_3	CONFORMED	
V1- Speed First Phase	NUMBER(10)	V1_4	CHECK	
V1- Speed First Phase	NUMBER(10)	V1_5	NOT CONFORMED	
C2- Course Second Phase	NUMBER(10)	C2_1	NOT CONFORMED	
C2- Course Second Phase	NUMBER(10)	C2_2	CHECK	
C2- Course Second Phase	NUMBER(10)	C2_3	CONFORMED	
C2- Course Second Phase	NUMBER(10)	C2_4	CHECK	
C2- Course Second Phase	NUMBER(10)	C2_5	NOT CONFORMED	
T2- Time Second Phase	NUMBER(10)	T2_1	NOT CONFORMED	
T2- Time Second Phase	NUMBER(10)	T2_2	CHECK	
T2- Time Second Phase	NUMBER(10)	T2_3	CONFORMED	
T2- Time Second Phase	NUMBER(10)	T2_4	CHECK	
T2- Time Second Phase	NUMBER(10)	T2_5	NOT CONFORMED	
V2- Speed Second Phase	NUMBER(10)	V2_1	RECASTING	
V2- Speed Second Phase	NUMBER(10)	V2_2	CHECK	
V2- Speed Second Phase	NUMBER(10)	V2_3	CONFORMED	
V2- Speed Second Phase	NUMBER(10)	V2_4	CHECK	
V2- Speed Second Phase	NUMBER(10)	V2_5	RECASTING	
CC- Multiplied Course	NUMBER(10)	CC_1	NOT CONFORMED	
CC- Multiplied Course	NUMBER(10)	CC_2	CHECK	
CC- Multiplied Course	NUMBER(10)	CC_3	CONFORMED	
CC- Multiplied Course	NUMBER(10)	CC_4	CHECK	
CC- Multiplied Course	NUMBER(10)	CC_5	NOT CONFORMED	
PM- Multiplied Pression	NUMBER(10)	PM_1	RECASTING	
PM- Multiplied Pression	NUMBER(10)	PM_2	CHECK	
PM- Multiplied Pression	NUMBER(10)	PM_3	CONFORMED	
PM- Multiplied Pression	NUMBER(10)	PM_4	CHECK	
PM- Multiplied Pression	NUMBER(10)	PM_5	RECASTING	

PS- Specific Pression	NUMBER(10)	PS_1	NOT CONFORMED	
PS- Specific Pression	NUMBER(10)	PS_2	CHECK	
PS- Specific Pression	NUMBER(10)	PS_3	CONFORMED	
PS- Specific Pression	NUMBER(10)	PS_4	CHECK	
PS- Specific Pression	NUMBER(10)	PS_5	NOT CONFORMED	
PF- Final Pression	NUMBER(10)	PF_1	NOT CONFORMED	
PF- Final Pression	NUMBER(10)	PF_2	CHECK	
PF- Final Pression	NUMBER(10)	PF_3	CONFORMED	
PF- Final Pression	NUMBER(10)	PF_4	CHECK	
PF- Final Pression	NUMBER(10)	PF_5	NOT CONFORMED	
FC- Clamping Force	NUMBER(10)	FC_1	RECASTING	
FC- Clamping Force	NUMBER(10)	FC_2	CHECK	
FC- Clamping Force	NUMBER(10)	FC_3	CONFORMED	
FC- Clamping Force	NUMBER(10)	FC_4	CHECK	
FC- Clamping Force	NUMBER(10)	FC_5	RECASTING	
SM- Thickness	NUMBER(10)	SM_1	RECASTING	
SM- Thickness	NUMBER(10)	SM_2	CHECK	
SM- Thickness	NUMBER(10)	SM_3	CONFORMED	
SM- Thickness	NUMBER(10)	SM_4	CHECK	
SM- Thickness	NUMBER(10)	SM_5	RECASTING	
TC- Cycle Time	NUMBER(10)	TC_1	RECASTING	
TC- Cycle Time	NUMBER(10)	TC_2	CHECK	
TC- Cycle Time	NUMBER(10)	TC_3	CONFORMED	
TC- Cycle Time	NUMBER(10)	TC_4	CHECK	
TC- Cycle Time	NUMBER(10)	TC_5	RECASTING	

The state chart diagram in Figure 30 represents the state-based behaviour of the die casting aluminium. The state machine diagram shows the possible state of each parameter, and the events that can trigger a transition between the states.

The process is initially in the 'OFF' state. When it is ready to be evaluated, the 'EvOn' event triggers a transition from the 'ON' state to 'READ' state as shown in Figure 30a. Upon entry to the 'READ' state, a time event triggers a transition to 'VERIFY_C1' state in a time interval of 3 seconds. Once the process has entered in 'VERIFY_C1' state, it immediately transitions to the neutral state. A time event of 2 seconds triggers a transition to the condition connectors. It split a single segment into several branches. Branches are labelled with guard conditions that are evaluated contemporary to determine which branch satisfy the condition. The connector evaluates if the value C1=150 activates the state C1_1 or C1_2 or C1_3 or C1_4 or C1_5 (Figure 30a). A time event of 3 seconds triggers a transition to the 'VERIFY_T1' state. The logic is the same as the one just presented. The guard conditions for the 'VERIFY_V2' state in Figure 30b are evaluate based on the constraint $\left(V_2 = \frac{C_2}{T_2}\right)$ modelled in the parametric diagram in Figure 29. If the state V2_1 or V2_5 are activates, the number of DISCARDED_QUANTITY is calculated as follow:

tm(1000)/setDISCARDED_QUANTITY(DISCARDED_QUANTITY +1.

When also the 'VERIFY_SM' state is evaluated as shown in Figure 30c, the presence of a fault or a machine down-time should be detected. It requires the knowledge of all possible correlation between technological parameters with machine down-time for predicting all possible faults or breakdowns. For this reason, data-driven invariant modelling constructs will be detected in knowledge discovery and in knowledge extraction stages.

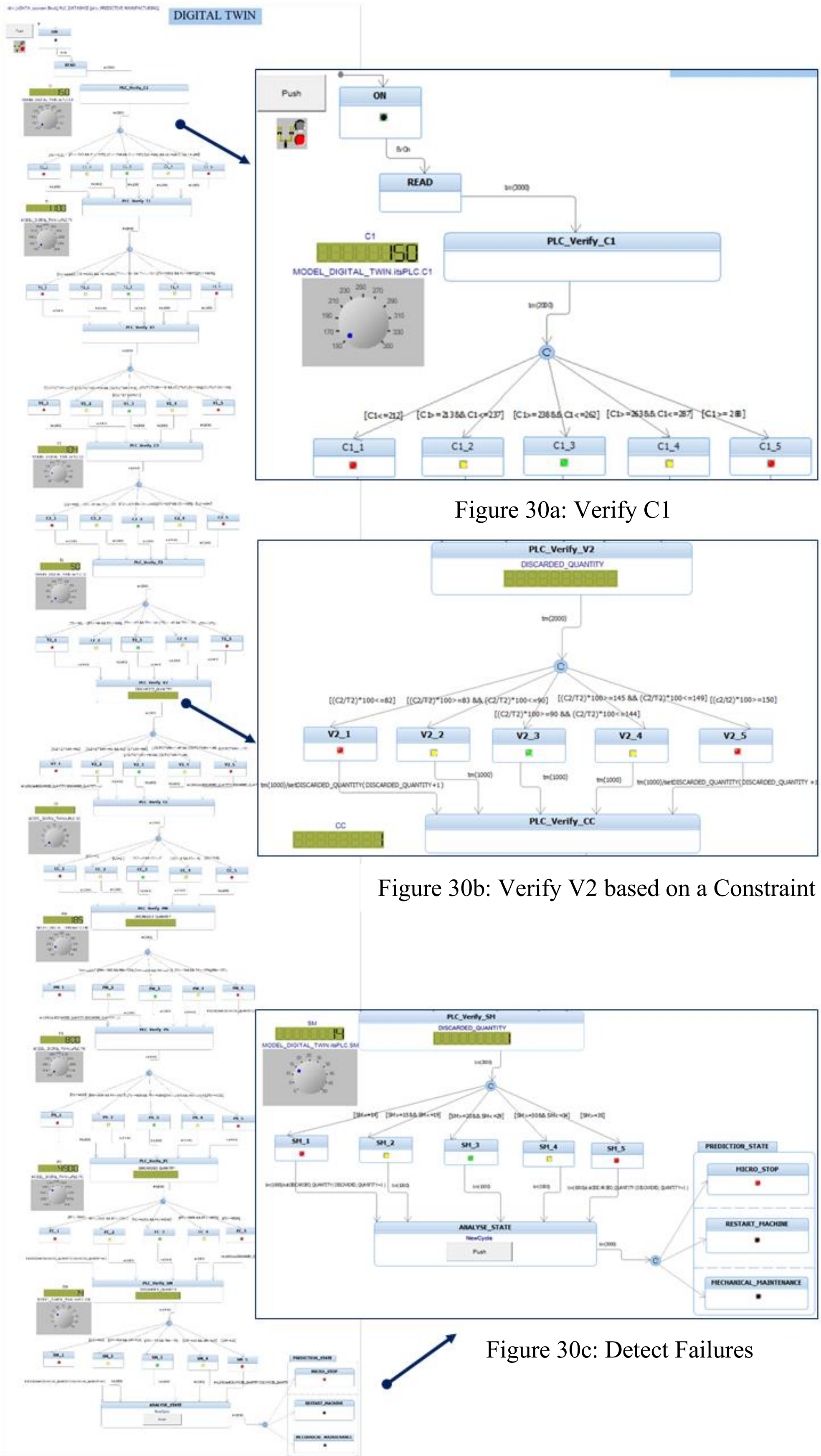


Figure 30a: Verify C1

Figure 30b: Verify V2 based on a Constraint

Figure 30c: Detect Failures

Figure 30: State Chart Diagram

4.2.3 DATA SELECTION

The previous paragraph shows the components of the die casting process and it defines which data need to be analysed to pursue the predictive manufacturing requirement. The data sources are the PLC (D1) and MES (D2).
PLC data need to be gathered and combined with MES data to define all possible and existing correlations between technological parameters and machine down-time. This aims to achieve improvements of efficiency of die casting process (reduction of downtime, reduction of waste, optimization of maintenance).
MES is a fundamental support tool for improving the production management. The system can detect all events that occur in a factory environment, to calculate the statistical parameters associated with specific problems and to monitor the performance of the plant in real time. All possible machine down-time (DESCRIPTION_ MACHINE_DOWN_TIME) are tracked, descripted and listed as follow:

- Thickness SM<=14 indicates a quality problem of the product.
- Thickness SM=15 indicates a quality problem of the product.
- Thickness SM=16 indicates a possible quality problem of the product.
- Micro stop for SM_LimLow indicates a down-time less the five minutes generated by the parameter SM.
- Component problem indicates that a breakdown of a component of the machine (plunger, mould etc).
- Electrical maintenance indicates a down-time for electrical breakdown.
- Mechanical maintenance indicates a down-time for mechanical breakdown.
- Recasting start machine for micro stop (after SM_LimLow) indicates the recasting of the injection cycle after the restart of the machine due to the micro-stop generated by the parameter SM.
- Recasting for SM_LimUpp indicates the recasting of the injection cycle after the restart of the machine due to a problem of the SM parameter.
- Recasting start machine for generic micro stop indicates the recasting of the injection cycle after the restart of the machine due to a generic micro-stop.
- Recasting start machine after cleaning of the equipment indicates the recasting of the injection cycle after the restart of the machine due to the cleaning of the equipment.
- Recasting start machine after auto-maintenance indicates the recasting of the injection cycle after the restart of the machine due to the regular maintenance activities.
- Recasting start machine after component problem indicates the recasting of the injection cycle after the restart of the machine due to component breakdown.
- Recasting start machine after electrical maintenance indicates the recasting of the injection cycle after the restart of the machine due to electrical breakdown.
- Recasting start machine after mechanical maintenance indicates the recasting of the injection cycle after the restart of the machine due to mechanical breakdown.
- Recasting start machine after operator absence indicates the recasting of the injection cycle after the restart of the machine due to the absence of the operator during the production.
- Recasting start machine after not planned indicates the recasting of the injection cycle after the restart of the machine due to a stop not planned.

- Recasting start machine after setup indicates the recasting of the injection cycle after the restart of the machine due to the setup of the machine.
- Recasting start machine after mechanical control indicates the recasting of the injection cycle after the restart of the machine due to a mechanical control.

Another database has been created to define a set of features to describe each down-time problem (D3). The machine down-time are described by 20 features divided into five categories:

F1. The Problem's nature
- Low cooling of final piece
- Low solidification withdrawal
- Standard cooling of final piece
- Standard solidification withdrawal
- High cooling of final piece
- High solidification withdrawal

F2. Problem' typology:
- Quality problem
- Mould problem
- Plunger problem
- Mechanical component problem
- PLC problem
- Electrical subsystem problem

F3. Duration downtime:
- Downtime (< 5min)
- Downtime (>15min)
- Downtime (>60min)

F4. Discarded Quantity:
- Recasting (=2 mould casting ≈ 72 pieces)
- Recasting (>=4 mould casting ≈ 144 pieces)

F5. Economic Impact:
- High downtime cost
- Medium downtime cost
- Low downtime cost

4.2.4 KNOWLEDGE DISCOVERY

The data selected are data from: PLC (D1), MES (D2). The PLC and MES databases cover the last 18 months of production. RCA is applied to find out automatically associations between technological parameter, machine down-time and features' machine down-time.

4.2.4.1 Data modelling

The first step is to analyse the relationships between dataset. The datasets have many to many relationships i.e. one instances of a dataset is associated with more than one instances of another dataset. The symbol *, in Figure 31, denotes the multiplicity many to many. For example, one technological parameter of PLC dataset can generate more than one machine down-time defined in MES dataset. At the same time, one machine down-time can be caused by more than one technological parameter and it can be described by more than one feature problem. The same feature can describe more than one problem.

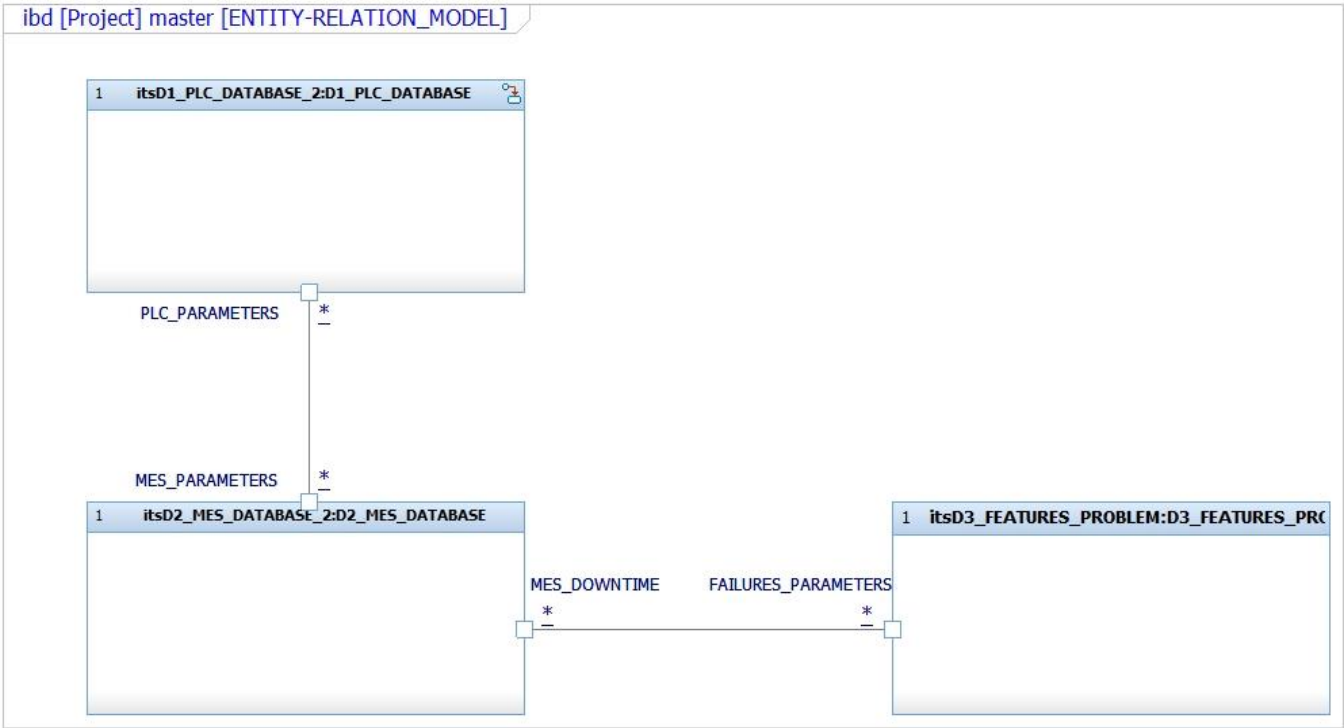


Figure 31: Multi Relation Database

It is necessary to define the objects and the attributes for each data set. The original PLC data set (D1) presents the injection cycles (MEASSETID) as objects (G) and technological parameters, defined in Figure 27, as attributes (M).

The original MES data set (D2) presents the data (DATA_TIME) as objects (G) and the value properties, defined in Figure 27, as attributes (M).

The data sets have been transformed to perform RCA analysis according to the logic defined in Figure 32. The MES data table (D2) has the same objects of PLC (D1) data table (OBi_{D1}) to find out the associations between technological parameters and downtime problems.

The attributes of MES (ATi_{D2}) data table (D2) are the objects of the feature machine down-time data table (D3) to describe clearly the discovered association between technological parameters and downtime problems.

D1(PLC)	$AT1_{D1}$	$AT2_{D1}$	$AT3_{D1}$	ATi_{D1}
$OB1_{D1}$	X		X	
$OB2_{D1}$		X		
$OB3_{D1}$			X	X
OBi_{D1}	X		X	X

D2(MES)	$AT1_{D2}$	$AT2_{D2}$	$AT3_{D2}$	ATi_{D2}
$OB1_{D1}$		X		X
$OB2_{D1}$	X			X
$OB3_{D1}$	X			
OBi_{D1}		X	X	

D3(FM)	$AT1_{D3}$	$AT2_{D3}$	$AT3_{D3}$	ATi_{D3}
$AT1_{D2}$		X	X	
$AT2_{D2}$	X			
$AT3_{D2}$		X		X
ATi_{D2}			X	

Figure 32: Data Modelling

The data table of PLC, shown in Table 2, has 57.779 objects (G) and 67 attributes (M). The objects (OBi_{D1}) are the injection cycle. A row is generated each 30 seconds. Each injection cycle is evaluated based on the following attributes (ATi_{D1}):

- A_M, B_P, C_N to define the shift (A: morning; B: afternoon; C: night).
- 10A003, 10A024, 10A026, 10A035, 10A043, 10A048 to define the article to product.
- C1_1, C1_2, C1_3, C1_4, C1_5, C2_1, C2_2, C2_3, C2_4, C2_5, T1_1, T1_2,T1_3, T1_4, T1_5, V1_1, V1_2, V1_3, V1_4, V1_5, T2_1, T2_2, T2_3, T2_4, T2_5, V2_1,V2_2, V2_3, V2_4, V2_5, CC_1, CC_2, CC_3, CC_4, CC_5, PM_1, PM_2, PM_3, PM_4, PM_5, FC_1, FC_2, FC_3, FC_4, FC_5, PS_1, PS_2, PS_3, PS_4, PS_5, SM_1, SM_2, SM_3, SM_4, SM_5 to define the technological parameters and their corresponding range.
- OK, KO, CHECK to define the quality state of the injection cycle.

Table 2: PLC Data Transformation

D1	A_M	B_P	C_N	10A003	10A024	10A026	10A035	10A043	C1_1	C1_2	C1_3	C1_4	…
489	X	.	.	X	.	.	.	.	X	.	.	.	.
490	X	.	.	X	.	.	.	.	X	.	.	.	.
491	X	.	.	X	.	.	.	.	X	.	.	.	.
492	.	X	.	X	.	.	.	.	X	.	.	.	.
493	.	X	.	X	.	.	.	.	.	X	.	.	.
494	.	.	X	X	.	.	.	.	.	.	X	.	.
…	.	.	.	.	.	.	.	.	.	.	.	.	.

The data table of MES, shown in Table 3, has 57.779 objects (G) and 19 attributes (M). The objects (OBi_{D1}) are the injection cycles The attributes (ATi_{D2}) are the machine downtime problems listed in the paragraph 4.2.3.

Table 3: MES Data Transformation

D2	THICKNESS <=14	THICKNESS =15	THICKNESS =16	RECASTING FOR SM_LimUpp	MICROSTOP FOR SM_LimLow	…
489	.	.	X	.	.	.
490	.	.	X	.	.	.
491	.	.	X	.	.	.
492	.	X	.	.	.	.
493	.	X	.	.	.	.
494	X	.	.	.	.	.
…	.	.	.	.	.	.

The data table of Features of machine down-time, shown in Table 4, has 19 objects (G) and 20 attributes (M). The objects (ATi_{D2}) are the machine downtime problems and the attributes (ATi_{D3}) are the features listed in the paragraph 4.2.3.

Table 4: FEATURE' MACHINE DOWN TIME Data Transformation

D3	LOW COOLING OF FINAL PIECE	LOW SOLIDIFICATION WITHDRAWAL	STANDARD COOLING OF FINAL PIECE	STANDARD SOLIDIFICATION WITHDRAWAL	HIGH COOLING OF FINAL PIECE	…
THICKNESS <=14	X	.	.	X	.	.
THICKNESS =15	X	.	.	X	.	.
THICKNESS =16	X	.	.	X	.	.
…	.	.	.	.	.	.

4.2.4.2 Discovery association

RCA detects 184.353 concepts (association rules). A cluster logic has been introduced to analyse all associations rules (concepts). The clusters have been defined based on all possible combination between technological parameters and machine down time as follow:

- C1 IF technological parameters THEN technological parameters
- C2 IF technological parameters and machine down time THEN technological parameters
- C3 IF technological parameters THEN machine down time
- C4 IF machine down time THEN technological parameters
- C5 IF machine down time THEN technological parameters and machine down time
- C6 IF machine down time THEN machine down time
- C7 IF technological parameters and machine down time THEN problems
- C8 IF technological parameters THEN technological parameters and machine down time
- C9 IF technological parameters and machine down time THEN technological parameters and machine down time

The concepts can be split into two groups:

- A1 includes all concepts belonging to the cluster C1.
  It means that the concepts contain all association rules between technological parameters.
- A2 includes all concepts belonging to the cluster C2, C3, C4, C5, C6, C7, C8, C9.

It means that the concepts contain all association rules between technological parameters, machine down time and features.

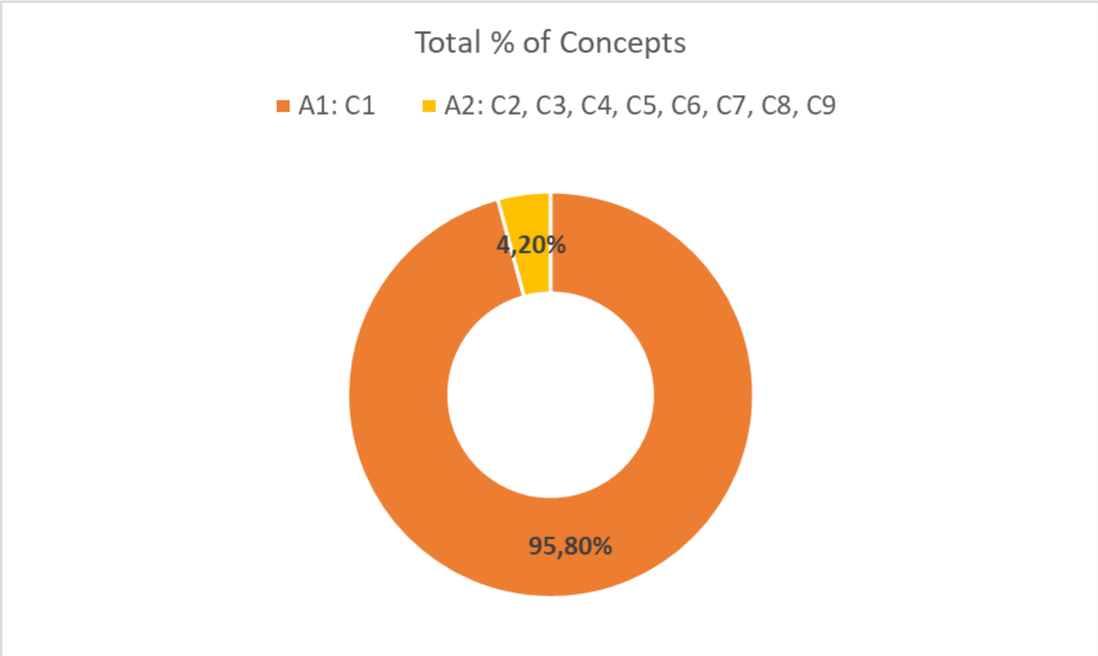


Figure 33: Total % of Concepts

4.2.5 KNOWLEDGE EXTRACTION

The group A2 has been selected to detect data-driven invariant modelling constructs based a data analysis. A2 includes 7.743 concepts. The number of concepts for each machine down time is shown in Figure 34. Quality and mechanical problems present the highest number of concepts to investigate.

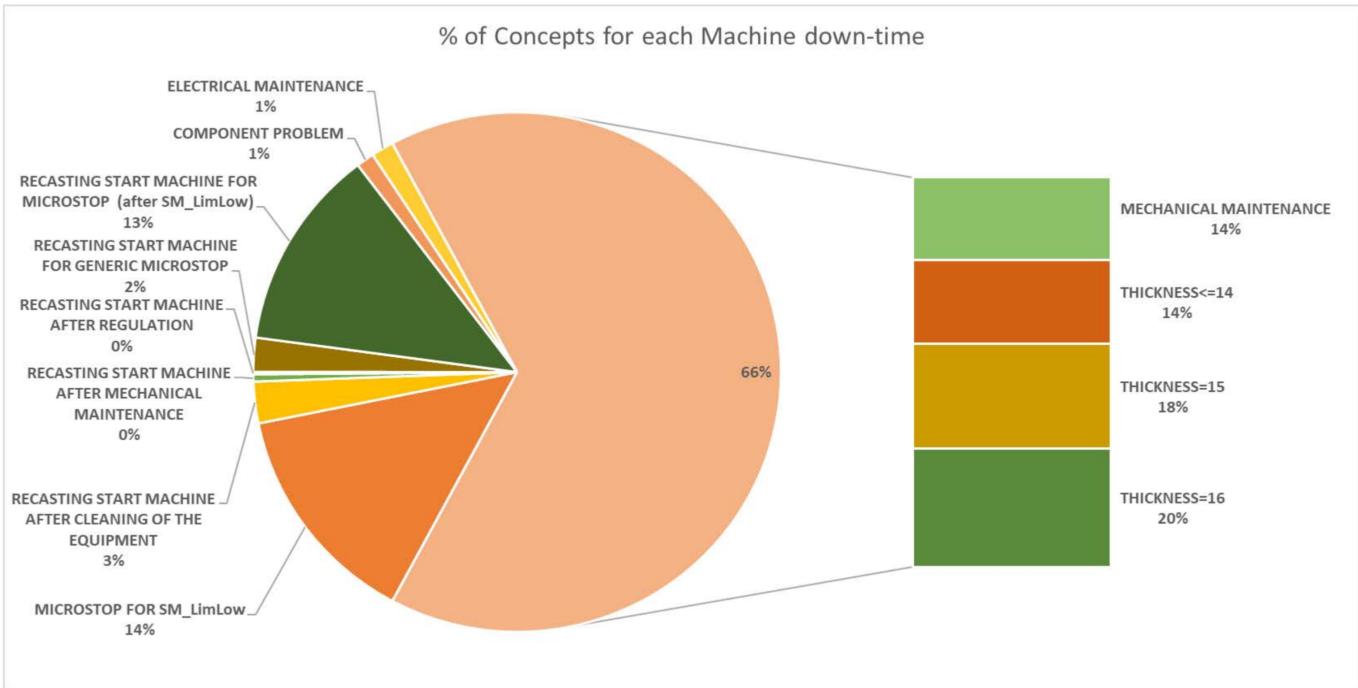


Figure 34: % of Concepts for each Machine Down-time

The group A2 presents all clusters except the cluster C1 as shown in Figure 35. The clusters C2, C8 and C9 need to be investigated further because these have the highest number of concepts.

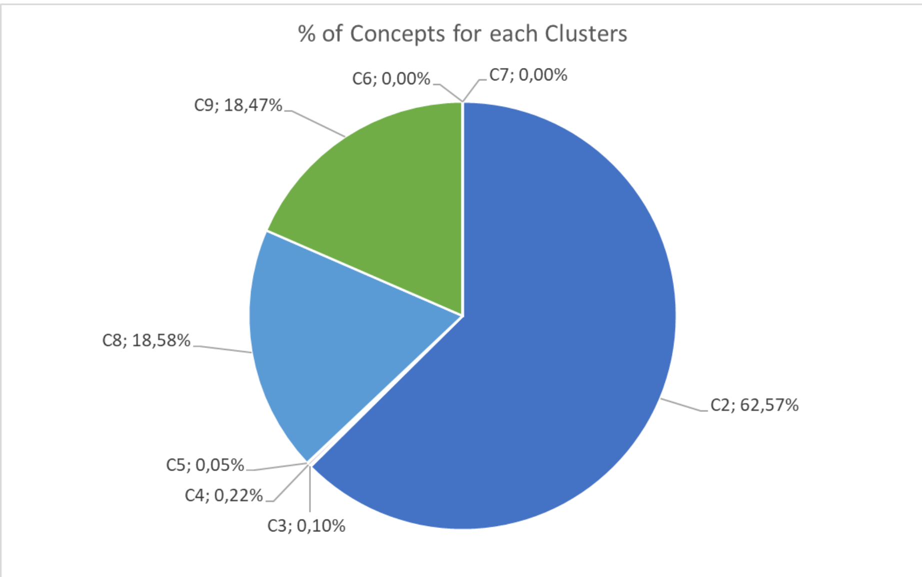


Figure 35: % of Concepts for each Clusters of the group A2

The clusters C2, C8, C9 have been analysed in relation to each machine down time as explained in Figure 36. Recasting problems, thickness =15 and thickness =16 have concepts belonging only to cluster C2. Component problem, Electrical maintenance, Mechanical maintenance, Recasting start machine for micro stop (after SM_LimLow) have concepts belonging to cluster C2 and C8. Thickness <=14 and Micro stop for SM_LimLow have concepts belonging to cluster C2, C8, C9. This is clearly summarised also in figure 37.

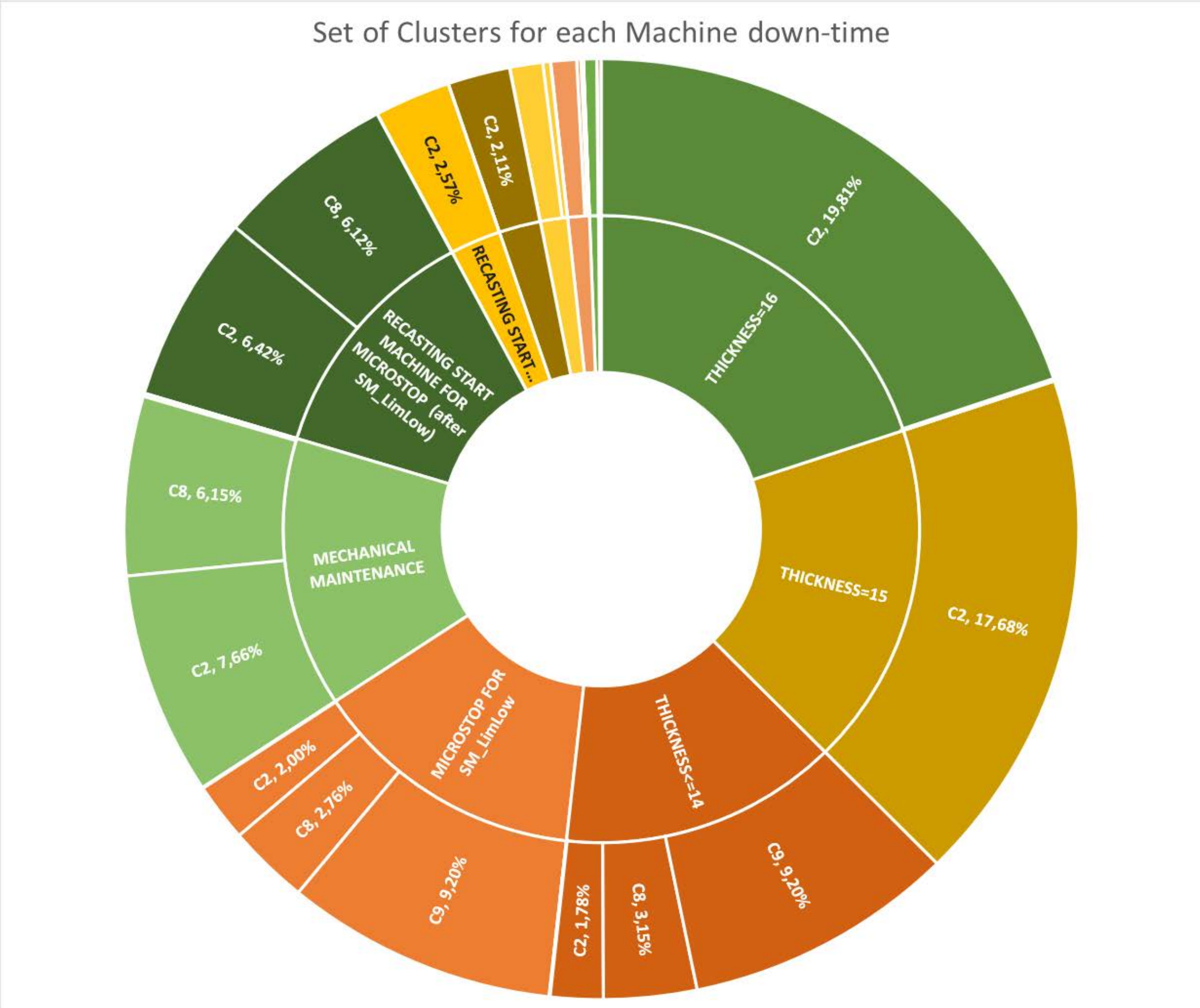


Figure 36: Clusters for each Machine down-time

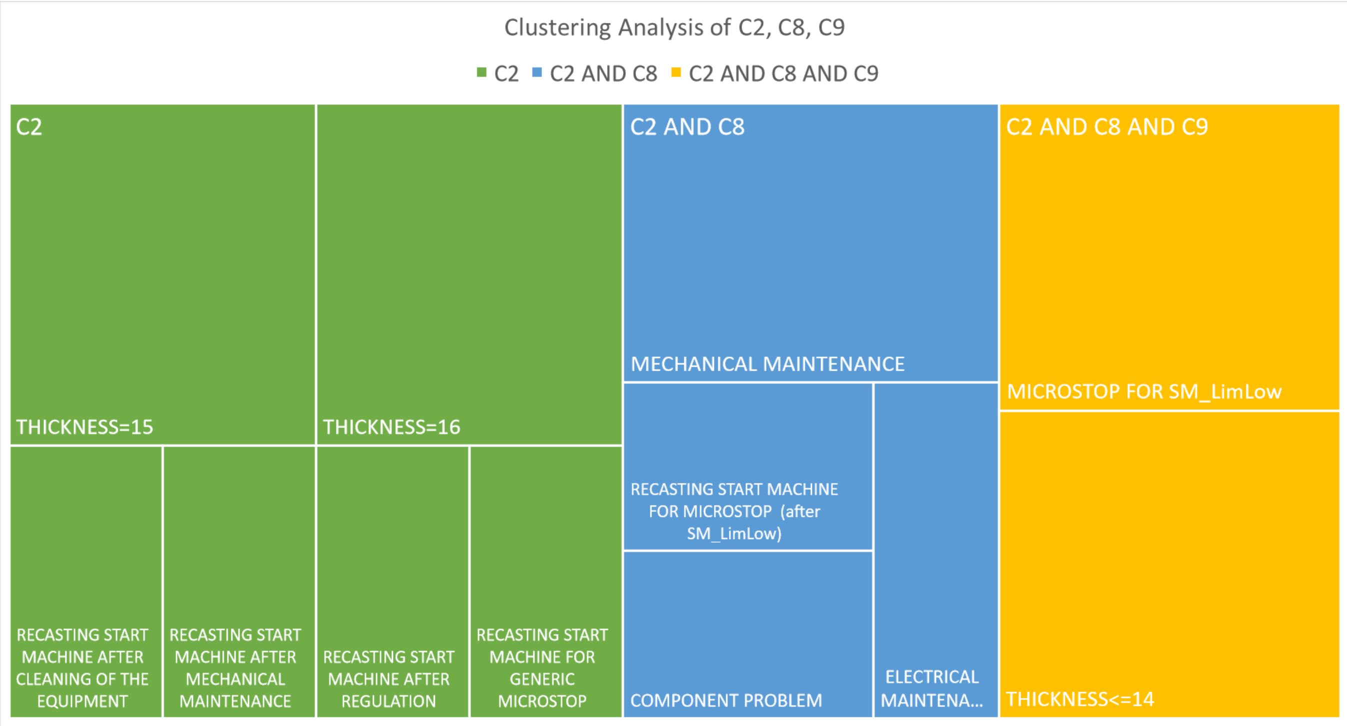


Figure 37: Cluster Analysis of C2, C8, C9

An example of concept (association rule) for each down time is presented below:

Table 5: Examples of Association Rule for each Machine down-time

DOWN-TIME	ID Cluster	CONCEPTS: ASSOCIATION RULES
THICKNESS<=16	C2	*76665 : A_M, PM_3, FC_4, Ex gen pbs : (**THICKNESS=16,** STANDARD COOLING OF FINAL PIECE , STANDARD SOLIDIFICATION WITHDRAWAL) → 10A035, C1_4, C2_3, T1_5, V1_4, CC_3, SM_2, OK [0.000484605, 1]*
THICKNESS<=15	C2	*76388 : FC_2, OK, Ex gen pbs : (**THICKNESS=15,** STANDARD COOLING OF FINAL PIECE , STANDARD SOLIDIFICATION WITHDRAWAL) → C1_5, C2_1, V1_5, T2_1, CC_2, SM_2 [0.0223611, 1]*
THICKNESS<=14	C9	*14537 : B_P, PM_1, Ex gen pbs : (MICROSTOP FOR SM_LimLow, LOW COOLING OF FINAL PIECE , LOW SOLIDIFICATION WITHDRAWAL , MEDIUM COST), Ex gen pbs : (LOW COOLING OF FINAL PIECE , LOW SOLIDIFICATION WITHDRAWAL , DOWN TIME (< 5MIN), MEDIUM COST) → 10A026, C1_5, C2_1, V1_5, T2_1, CC_2, FC_2, SM_1, KO, CHECK, Ex gen pbs : (**THICKNESS<=14,** LOW COOLING OF FINAL PIECE , LOW SOLIDIFICATION WITHDRAWAL , QUALITY DEFECT, DOWN TIME (< 5MIN), MEDIUM COST)[0.000934596, 1]*
MICROSTOP FOR SM_LimLow	C9	*14674 : T1_1, SM_1, Ex gen pbs : (THICKNESS<=14, LOW COOLING OF FINAL PIECE , LOW SOLIDIFICATION WITHDRAWAL , QUALITY DEFECT, DOWN TIME (< 5MIN), MEDIUM COST) → C1_5, C2_1, V1_5, T2_1, CC_2, KO, CHECK, Ex gen pbs : (**MICROSTOP FOR SM_LimLow,** LOW COOLING OF FINAL PIECE , LOW SOLIDIFICATION WITHDRAWAL , MEDIUM COST)[0.00347877, 1]*
RECASTING START MACHINE	C2	*78613 : PS_5, PS_5, Ex gen pbs : (LOW COOLING OF FINAL PIECE , LOW SOLIDIFICATION WITHDRAWAL , DOWN TIME (< 5MIN), MEDIUM COST) → V2_5, T2_1, KO,*

FOR MICRO-STOP (after SM_LimLow)		*Ex gen pbs : (**RECASTING START MACHINE FOR MICROSTOP (after SM_LimLow),** LOW COOLING OF FINAL PIECE , LOW SOLIDIFICATION WITHDRAWAL , DOWN TIME (< 5MIN), RECASTING(=2 MOULD CASTING), MEDIUM COST)[0.00257879, 1]*
RECASTING START MACHINE AFTER CLEANING OF THE EQUIPMENT	C2	*78952 : TC_5, T2_5, PS_1, SM_5, Ex gen pbs : (**RECASTING START MACHINE AFTER CLEANING OF THE EQUIPMENT,** DOWN TIME (>15MIN), RECASTING(>=4 MOULD CASTING), MEDIUM COST) →* *PM_1, V2_1, KO [0.000484605, 1]*
RECASTING START MACHINE AFTER MECHANICAL MAINTENANCE	C2	*79056 : V2_5, T2_1, CHECK, Ex gen pbs : (**RECASTING START MACHINE AFTER MECHANICAL MAINTENANCE,** DOWN TIME (>60MIN), RECASTING(>=4 MOULD CASTING), MEDIUM COST) →* *PM_5, PS_1, TC_2, KO [0.000501912,1]*
RECASTING START MACHINE AFTER MECHANICAL CONTROL	C2	*79085 : T2_1, Ex gen pbs : (**RECASTING START MACHINE AFTER MECHANICAL CONTROL,** DOWN TIME (>15MIN), RECASTING(>=4 MOULD CASTING), HIGH COST) →* *10A003, 10A024, 10A043, TC_5, SM_5 KO [0.000536527, 1]*
RECASTING START MACHINE FOR GENERIC MICROSTOP	C2	*78755 : 10A035, SM_5, Ex gen pbs : (**RECASTING START MACHINE FOR GENERIC MICROSTOP ,** DOWN TIME (< 5MIN), RECASTING(=2 MOULD CASTING), MEDIUM COST) →* *PM_5, PS_5, KO [0.000484605, 1]*
COMPONENT PROBLEM	C2	*79153 : PM_4, Ex gen pbs : (**COMPONENT PROBLEM,** MOULD PROBLEM, PLUNGER PROBLEM, MECHANICAL COMPONENT PROBLEM, DOWN TIME (>60MIN), HIGH COST) →* *A_M, 10A026, FC_2, CHECK [0.00837675, 1]*
ELECTRICAL MAINTENANCE	C2	*79185 : FC_4, Ex gen pbs : (**ELECTRICAL MAINTENANCE**, PLC PROBLEM, ELECTRICAL SUBSYSTEM PROBLEM, DOWN TIME (>60MIN), HIGH COST) →* *A_M, PM_2, CHECK [0.000484605, 1]*
MECHANICAL MAINTENANCE	C8	*9302 : PM_2, Ex gen pbs : (HIGH COST) →* *A_M, 10A035, FC_4, CHECK* *Ex gen pbs : (**MECHANICAL MAINTENANCE,** MOULD PROBLEM, PLUNGER PROBLEM, MECHANICAL COMPONENT PROBLEM, DOWN TIME (>60MIN), HIGH COST)[0.000501912, 1]*

The analysis of cluster has been merged with the analysis of the technological parameters to identify data-driven invariant modelling constructs. All machine down time can be split into three groups based on the correlation between technological parameters:

- **P1** The down time Thickness =16, Thickness =15, Thickness <=14, Micro stop for SM_LimLow share the following technological parameters: C_1, C_2, CC, SM_1, SM_2 as shown in Figure 40.
- **P2** All problems of recasting have in common the following parameters: SM_5, V_2_1, V_2_5, T_2_1, T_2_5, PS_1, PS_5, PM_1, PM_5, TC_5.
- **P3** The machine down time related to component problem, electrical maintenance and mechanical maintenance have the parameters PM_2, PM_4, FC_2, FC_4.

P1, P2, P3 are data-driven invariant modelling constructs. P1 covers the 65,75% of concepts clustered in the group A2, P2 covers the 17,85% and P3 the 16,40% of concepts in A2 as presented in Figure 39.

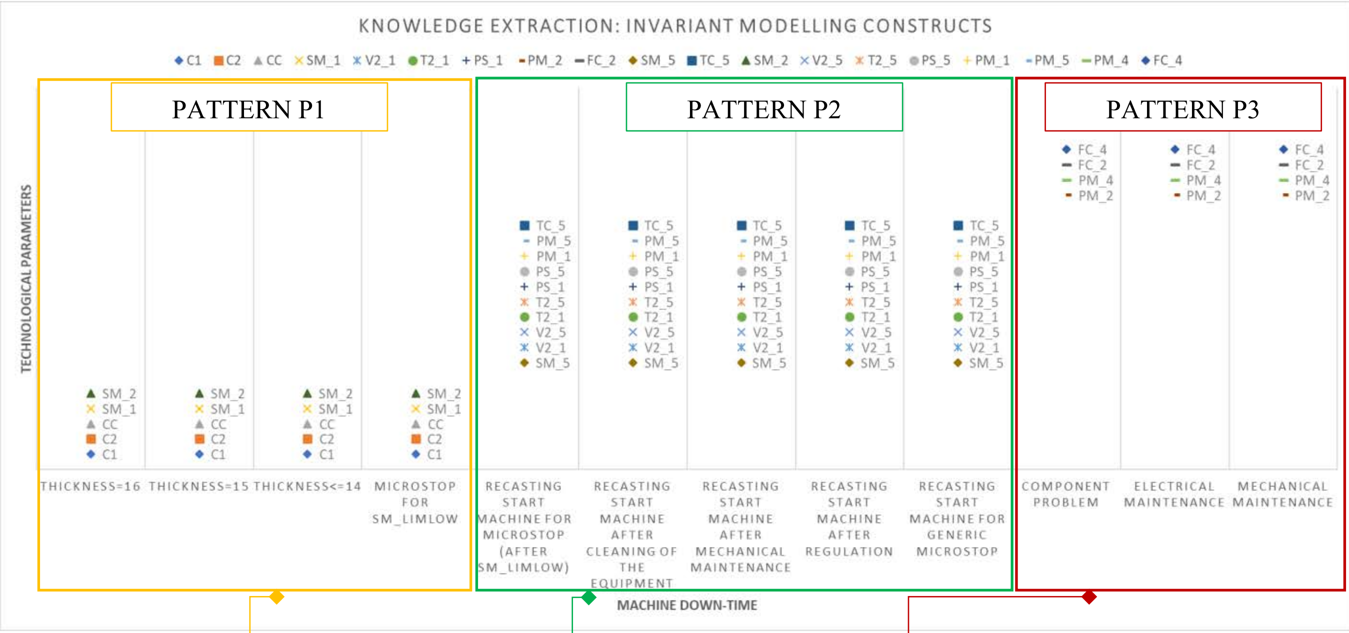


Figure 38: Invariant Modelling Constructs

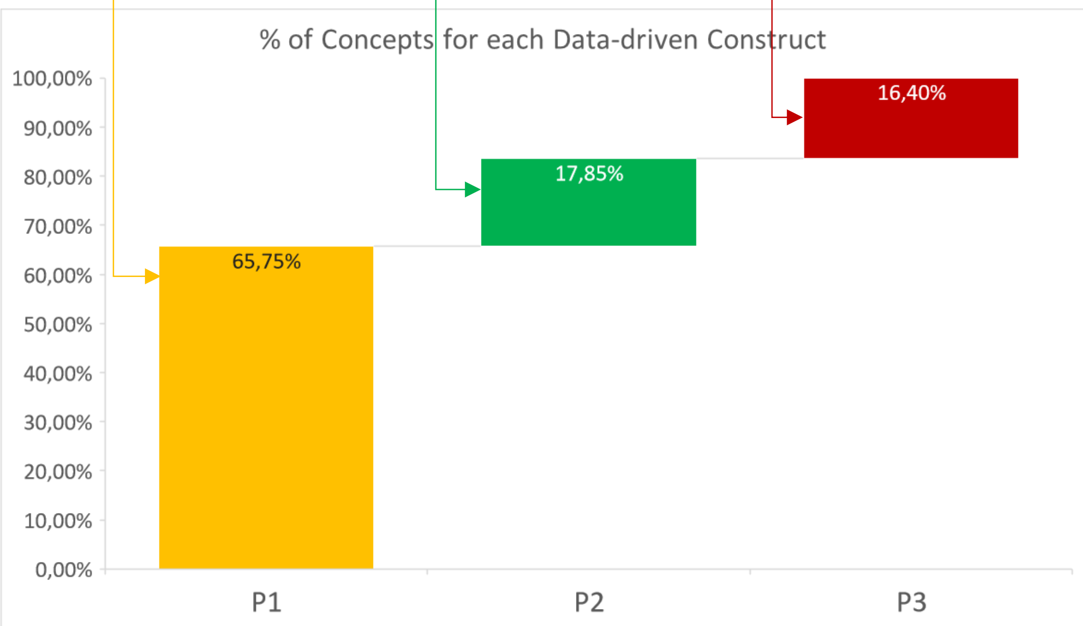


Figure 39: % of Concept grouped in A2 for each Data-driven Construct

The physical mining of each invariant modelling constructs has been defined analysing the model discussed in the paragraph 4.2.2. A list of some data-driven invariant modelling constructs is presented in Table 6. A data-driven invariant modelling construct contains is described by 1) ID Pattern; 2) Pattern Name; 3) Description; 4) An example of association rule; 5) Data View; 6) SysML Model View; 7) Applicability.

Table 6: The List of Data-driven Invariant Modelling Constructs

P1: MOULD FILLING PATTERN

- **ID: P1**
- **PATTER NAME:** MOULD FILLING PATTERN
- **DESCRIPTION:** the pattern aims to show and to describe the correlations between the parameters C1, C2, CC. The pattern represents the course of the plunger for the filling of the mould. This impacts on the quality of the product (SM).

P1 EXAMPLE OF ASSOCIATION RULE:

14675 : T1_1, SM_1 →
C1_5, C2_1, V1_5, T2_1, CC_2, KO, CHECK, Ex gen pbs : (MICROSTOP FOR SM_LimLow, , MEDIUM COST)
[0.00347877, 1]

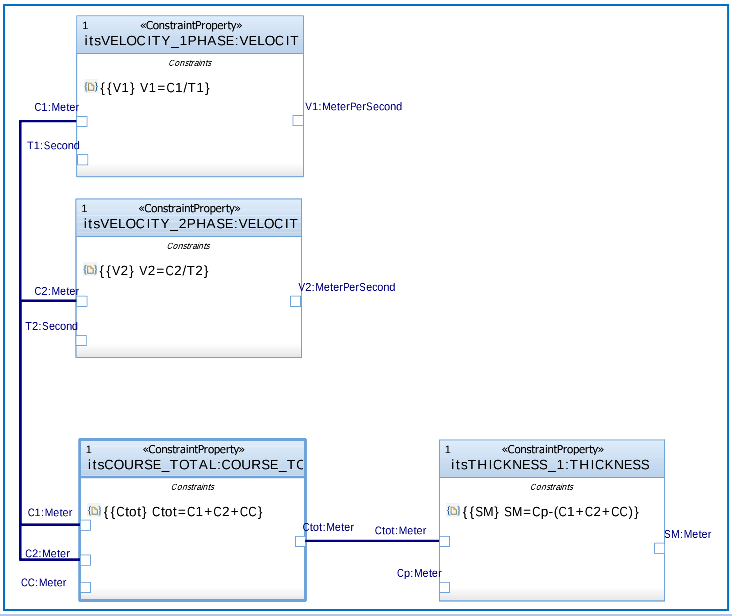
P1 DATA VIEW:

The set of data includes:

- Plunger course C_1 (mm) in the first phase of the injection stage
- Plunger course C_2 (mm) in the second phase of the injection stage
- Plunger course CC (mm) in the multiplied phase of the injection stage
- Product thickness SM (mm²)

C1	C2	CC	SM
295	81	3	17
295	79	3	17
297	79	4	16
297	79	3	17
297	79	3	17
296	81	3	16
297	80	3	16
297	80	3	16

P1 SYSML MODEL VIEW:



P1 APPLICABILITY: Use mould filling pattern for:

- Describing and preventing problems of micro-stop.
- Describing and preventing problems during filling processes.
- Describing the behaviour of the plunger course.
- Describing the behaviour of an injection system.

P2: MACHINE RESTART PATTERN

- **ID: P2**
- **PATTER NAME:** MACHINE RESTART PATTERN
- **DESCRIPTION:** the pattern aims to show and to describe the correlations between V2, T2, PM, PS, SM, TC. The pattern denotes the restart of the machine after a machine down time. The restart of a machine generates a recasting of the product.

P2 EXAMPLE OF ASSOCIATION RULE:

78953 : TC_5, T2_5, PS_1, SM_5, Ex gen pbs : (RECASTING START MACHINE AFTER CLEANING OF THE EQUIPMENT) → PM_1, V2_1, KO
[0.000484605, 1]

P2 DATA VIEW

The set of data includes:

- Plunger speed $\mathbf{V_2}$ (m/sec) in the second phase of the injection stage.
- Plunger time $\mathbf{T_2}$ (msec) in the second phase of the injection stage.
- Multiplied pressure PM (Pa) in the multiplied phase of the injection stage
- Specific pressure PS (Pa)
- Product thickness SM (mm²)
- Cycle time TC (sec)

V2	T2	PM	PS	SM	TC
35	175	215	928	20	33
37	180	229	946	20	33
37	167	231	993	20	33
122	59	223	928	19	33
120	60	224	933	19	33
127	55	223	946	19	33

P2 SYSML MODEL VIEW:

P2 APPLICABILITY: Use machine restart pattern when:

- Describing and preventing problems of recasting.
- Describing the restart of a machine.

P3 CLAMPING SYSTEM PATTERN

- **ID: P3**
- **PATTER NAME:** CLAMPING SYSTEM PATTERN
- **DESCRIPTION:** the pattern aims to show and to describe the correlations between PM, PS, and FC. The pattern identifies the cooling stage of the die casting aluminium process. The multiplied pressure are the clamping force are the parameters for monitoring the state of the clamping system. The values of the clamping force FC and multiplied pressure PM can generate mechanical problem.

P3 EXAMPLE OF ASSOCIATION RULE:

9302 : PM_2, Ex gen pbs : (HIGH COST) →
A_M, 10A035, FC_4, CHECK Ex gen pbs : (MECHANICAL MAINTENANCE, MOULD PROBLEM, PLUNGER PROBLEM, MECHANICAL COMPONENT PROBLEM, DOWN TIME (>60MIN), HIGH COST)
[0.000501912, 1]

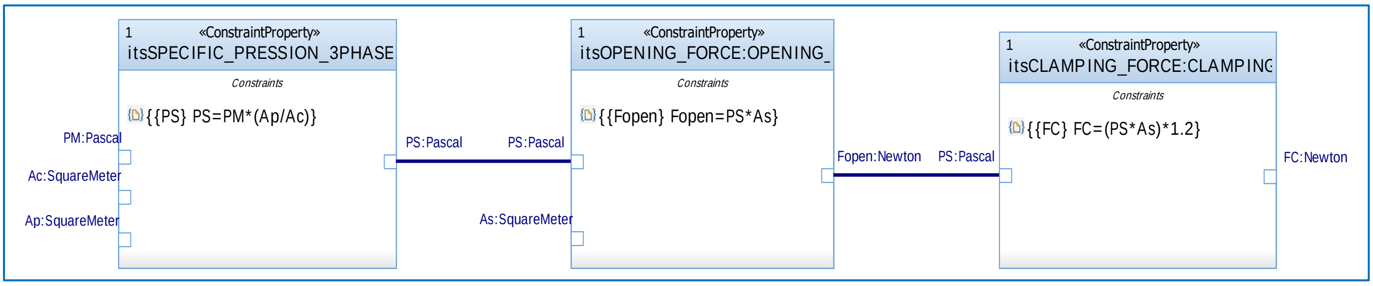
P3 DATA VIEW

The set of data includes:

- Clamping Force FC (kN)
- Multiplied Pressure PM (Pa)

PM	FC
217	5780
217	5737
218	5780
218	5748
218	5780
218	5791
217	5759
217	5758

P3 SYSML MODEL VIEW



P3 APPLICABILITY: Use clamping system pattern for:

- Describing the behaviour of a clamping system (e.g. the clamping of a mould).
- Describing and preventing mechanical breakdown.

P4 HYDRAULIC PRESSURE PATTERN

- **ID: P4**
- **PATTER NAME:** HYDRAULIC PRESSURE PATTERN
- **DESCRIPTION:** the pattern aims to describe a hydraulic system.

P4 EXAMPLE OF ASSOCIATION RULE:

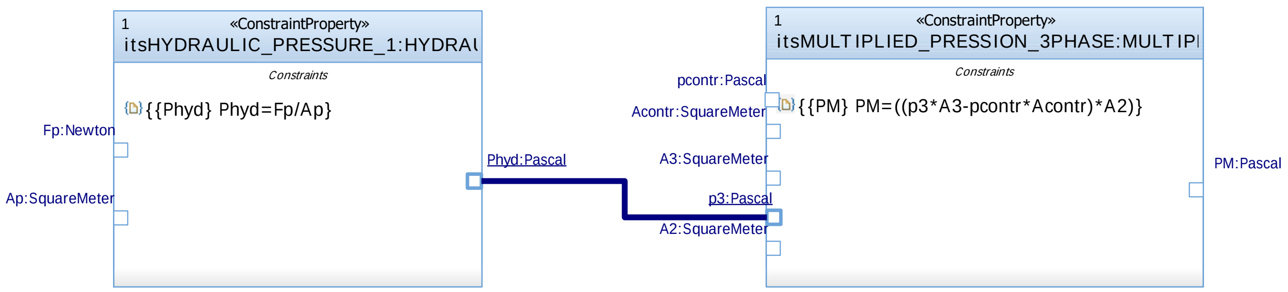
22706 : PS_1, PF_2→
PM_2
[0.447654, 1]

303 : PS_3, PF_4→
PM_4
[0.5000, 1]

P4 DATA VIEW

PM	PS	PF
217	933	215
218	937	216
218	933	215
217	933	215
217	933	215
218	937	216

P4 SYSML MODEL VIEW



P4 APPLICABILITY: Use clamping system pattern for:

- Describing a hydraulic system (e.g. the clamping of a mould).
- Describing hydraulic pressure.

P5 STATE OF A PARAMETER PATTERN

- **ID: P5**
- **PATTER NAME:** STATE OF A PARAMETER PATTERN
- **DESCRIPTION:** the pattern aims to evaluate the state of a parameter based on the values that it assumes.

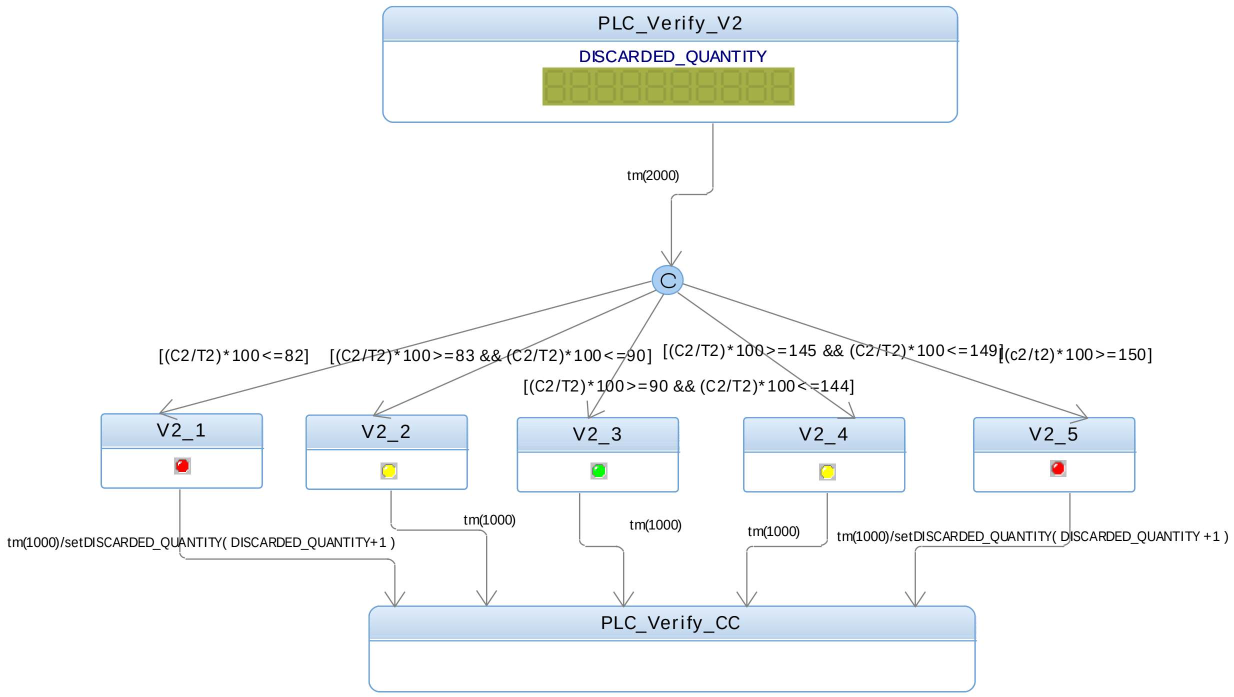
P5 EXAMPLE OF ASSOCIATION RULE:

*723 : V2_3 → **OK*** *[0.630783, 1]*	*724 : V2_4 → **CHECK*** *[0.64000, 1]*	*725 : V2_5 → **KO*** *[0.30000, 1]*

P5 DATA VIEW

V2	STATE
130	OK
127	OK
127	OK
125	CHECK
125	CHECK
117	KO

P5 SYSML MODEL VIEW



P5 APPLICABILITY: Use the state of a parameter pattern for:

- Describing and understanding the state of a parameter (*e.g. check, ok, ko, conformed, not conformed).*
- Simulating the state behaviour.

P6 STATE OF A MACHINE PATTERN

- **ID: P6**
- **PATTER NAME:** STATE OF A MACHINE PATTERN
- **DESCRIPTION:** the pattern aims to evaluate the state of a machine based on the correlations of a set of parameters (*e.g. check, ok, ko, conformed, not conformed, state 1/2/3).*

P6 EXAMPLE OF ASSOCIATION RULE:

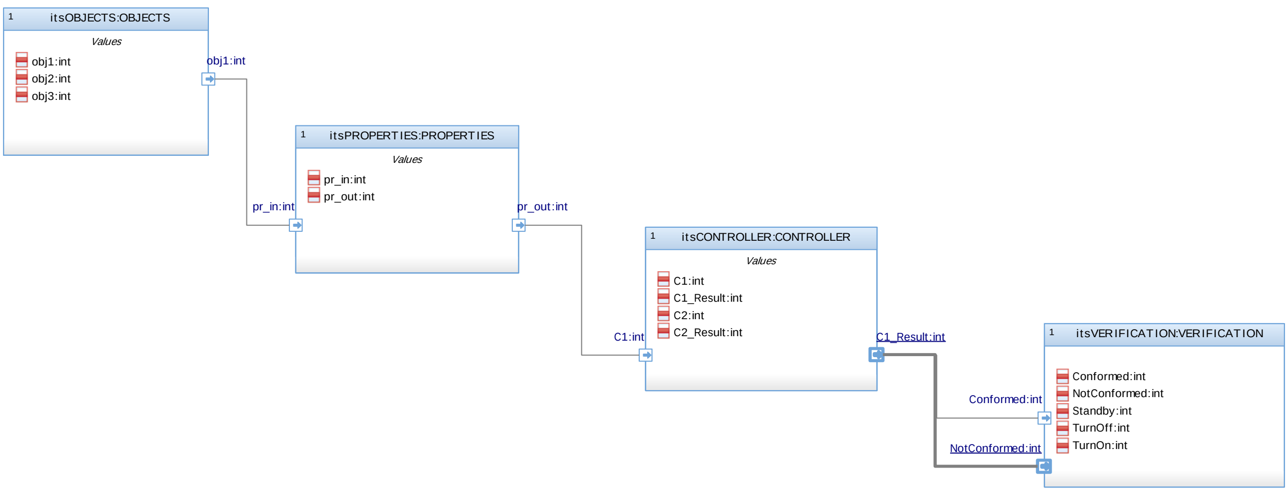
59452 : V1_5, T2_1, CC_2, PS_2, SM_3 →
C1_5, C2_1, CHECK
[0.40449, 1]

1103 : 10A026, CC_2 →
C1_5, C2_1, V1_5, OK
[0.405286, 1]

P6 DATA VIEW

C1	C2	CC	T1	T2	V1	V2	PM	PS	PF	FC	SM	TC	STATE MACHINE
295	81	3	1241	69	23	117	217	933	215	5780	17	33	OK
295	79	3	1251	63	23	125	217	928	214	5737	17	33	OK
297	79	4	1252	62	23	127	218	937	216	5780	16	34	OK
297	79	3	1246	62	23	127	218	928	214	5748	17	32	OK
297	79	3	1245	63	23	125	218	933	215	5780	17	33	OK
296	81	3	1245	66	23	123	218	933	215	5791	16	33	OK
297	80	3	1250	62	23	129	217	933	215	5759	16	33	OK
297	80	3	1245	62	23	129	217	933	215	5758	16	33	CHECK

P6 SYSML MODEL VIEW



P6 APPLICABILITY: Use evaluate the state of a machine pattern for:

- Describing and understanding the state of a machine.
- Simulating different states of the machine based on the same set of values.

4.2.6 KNOWLEDGE FORMALIZATION: WEB LIBRARY OF DATA-DRIVEN CONSTRUCTS

A Web platform based on SQL language in Oracle database has been created as shown in Figure 40 to join the PLC and MES datasets for formalising the invariant modelling constructs above described and for creating analytical reports for the employees. A data warehouse has been created to store and upload all data from PLC and MES datasets.

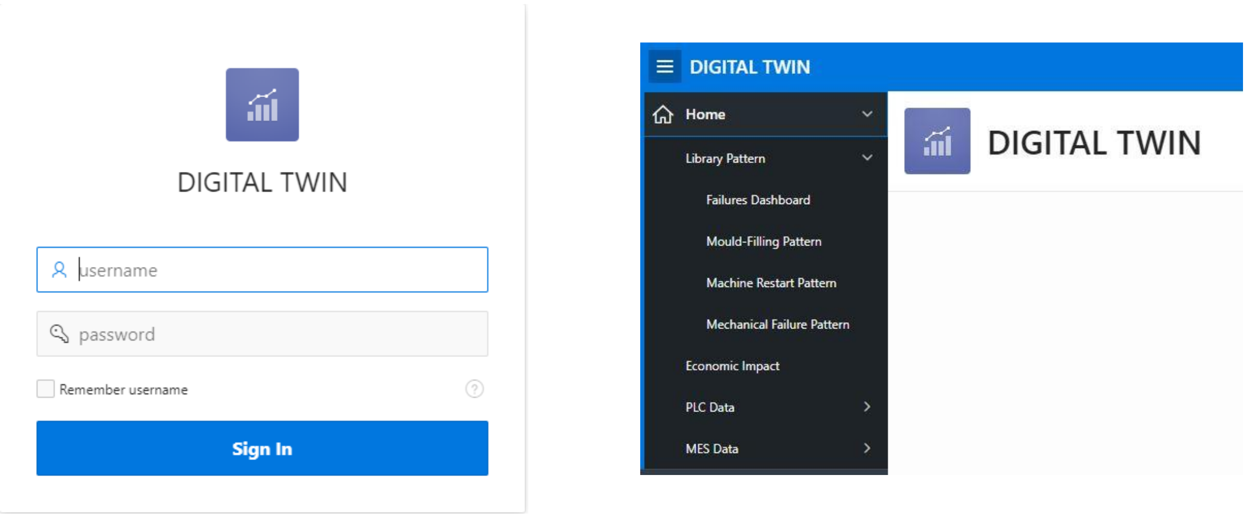


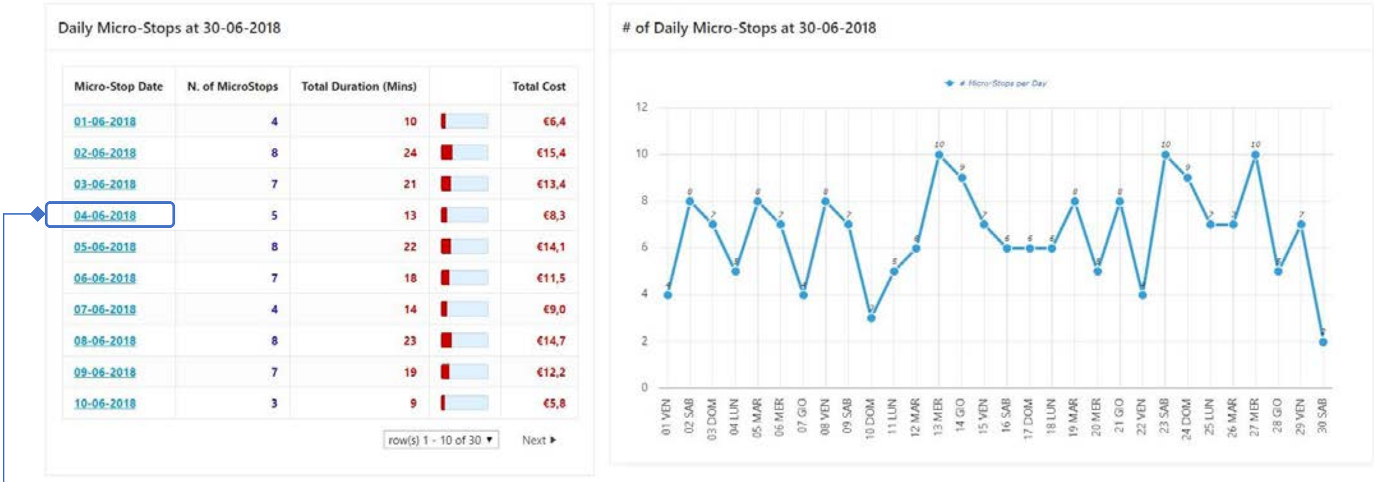
Figure 40: Web Platform for Data-driven constructs

4.2.6.1 MOULD FILLING PATTERN (P1)

The pattern P1 aims to describe and to predict a machine downtime called: micro-stop. It is generated by the correlation between the following parameters:

- Plunger course C_1 (mm) in the first phase of the injection stage
- Plunger course C_2 (mm) in the second phase of the injection stage
- Plunger course CC (mm) in the multiplied phase of the injection stage
- Product thickness SM (mm²)

The dashboard daily micro-stop in Figure 41 has been created to understand the frequency of this problem over time, the duration and the economic impact of each occurrence.
It is possible to click on each occurrence to comprehend the degradation trends of SM over time, as shown in Figure 42. The red line indicates the threshold value of SM (SM Min). As explained in table 1, if SM is equal to or below SM Min there is the recasting of the injection cycle (≈ 36 products) during the production. The dashboard shows that a micro stop is always preceded by a recasting. The requirement is to predict SM before the threshold value. It is necessary to discover the relationship between SM, C1, C2 and CC. In particular, the dashboard Pattern Values in Figure 43 demonstrates that C1 and C2 are inversely proportional as a function of SM. All possible combination between C1, C2, CC and SM need to be explored to define how to detect all possible behaviour of the system.



Daily Micro-Stops at 30-06-2018

Micro-Stop Date	N. of MicroStops	Total Duration (Mins)		Total Cost
01-06-2018	4	10		€6,4
02-06-2018	8	24		€15,4
03-06-2018	7	21		€13,4
04-06-2018	5	13		€8,3
05-06-2018	8	22		€14,1
06-06-2018	7	18		€11,5
07-06-2018	4	14		€9,0
08-06-2018	8	23		€14,7
09-06-2018	7	19		€12,2
10-06-2018	3	9		€5,8

Figure 41: Daily Micro-stop

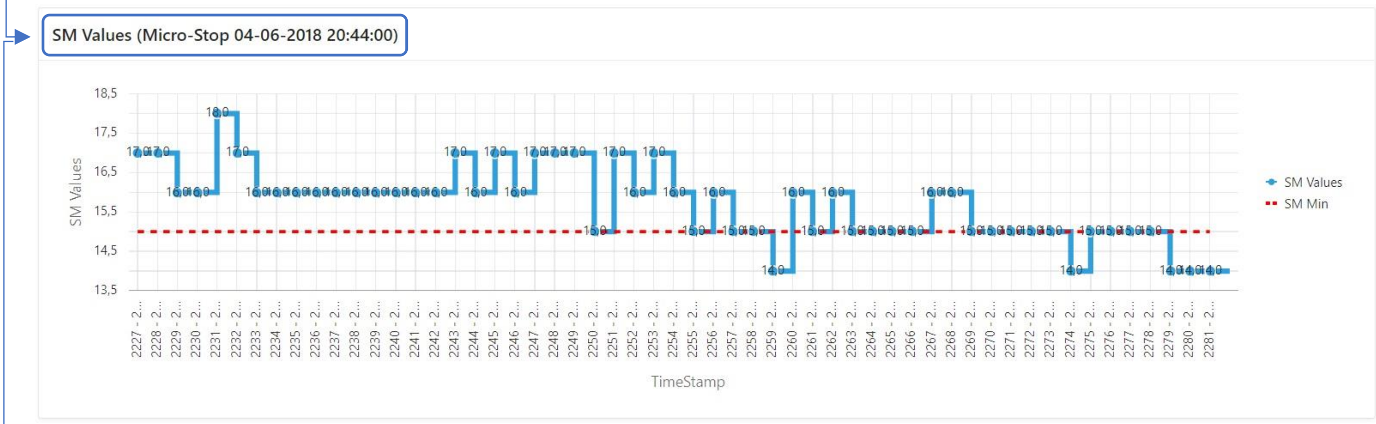


Figure 42: SM Values

Figure 43: Pattern Values

The result is to split all correlations into three categories as shown in Figure 44:

- The green area detects all possible combination values between C1, C2 and CC when SM is greater than the upper Limit. These combinations define when the system is functioning properly.

- The red area detects all possible combination between values C1, C2 and CC when SM is equal to or below SM Min. These combinations indicate a failure of the system generating problem of thickness SM<=LimLow and consequent micro stop of the machine.
- The yellow area detects all possible combination values between C1, C2 and CC when SM is in check ranges. It aims at discovering all combination for predicting all possible occurrences of Micro-Stop and quality problems.

The outcome is the possibility to predict possible quality problems and micro stop of the machine.

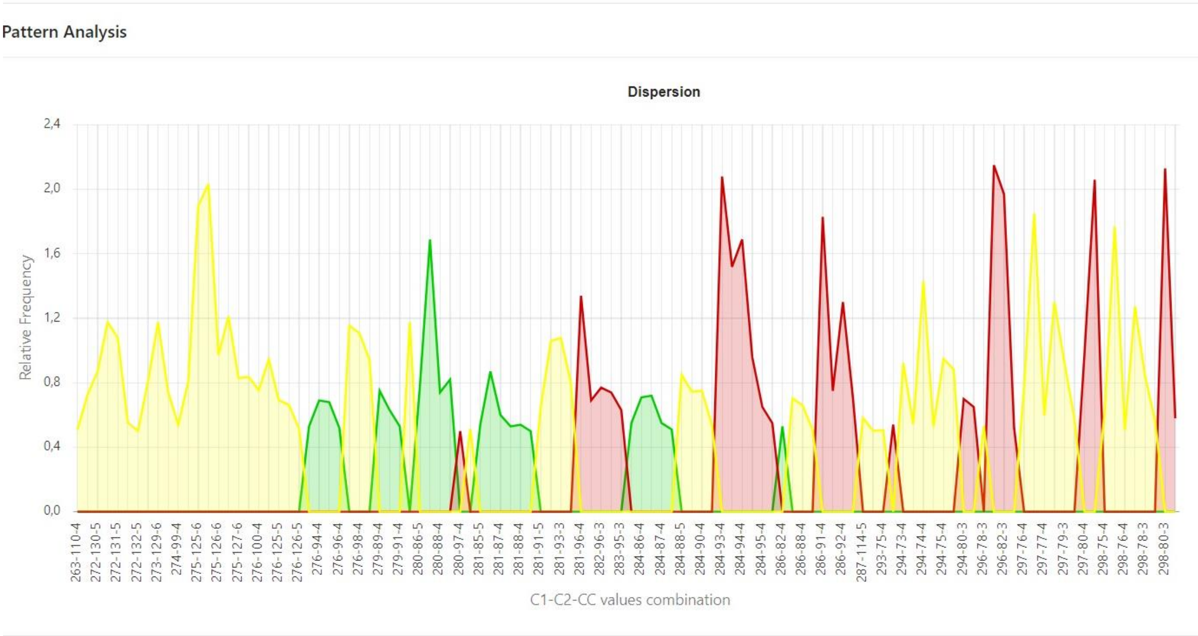


Figure 44: Pattern Analysis for Prediction

4.2.6.2 MACHINE RESTART PATTERN (P2)

The pattern P2 aims to describe a typical restart behaviour of the machine. The machine restart has the previous parameters out of range regardless of the type of machine downtime.
The pattern detects the correlation between the following parameters:

- Plunger speed V_2 (m/sec) in the second phase of the injection stage.
- Plunger time T_2 (msec) in the second phase of the injection stage.
- Multiplied pressure PM (Pa) in the multiplied phase of the injection stage
- Specific pressure PS (Pa)
- Product thickness SM (mm^2)
- Cycle time TC (sec)

The dashboard in Figure 45 has been created to demonstrate that whenever a downtime occurs, the restart of the machine presents one or more than one parameter, above listed, out of the threshold value. The parameters V_2, T_2, PM, PS could present values greater than or below threshold value while SM an TC only values greater than threshold value. It is possible to click on each downtime problem (cause code) as shown in Figure 46 to analyse which parameter causes it frequently and which value assumes.

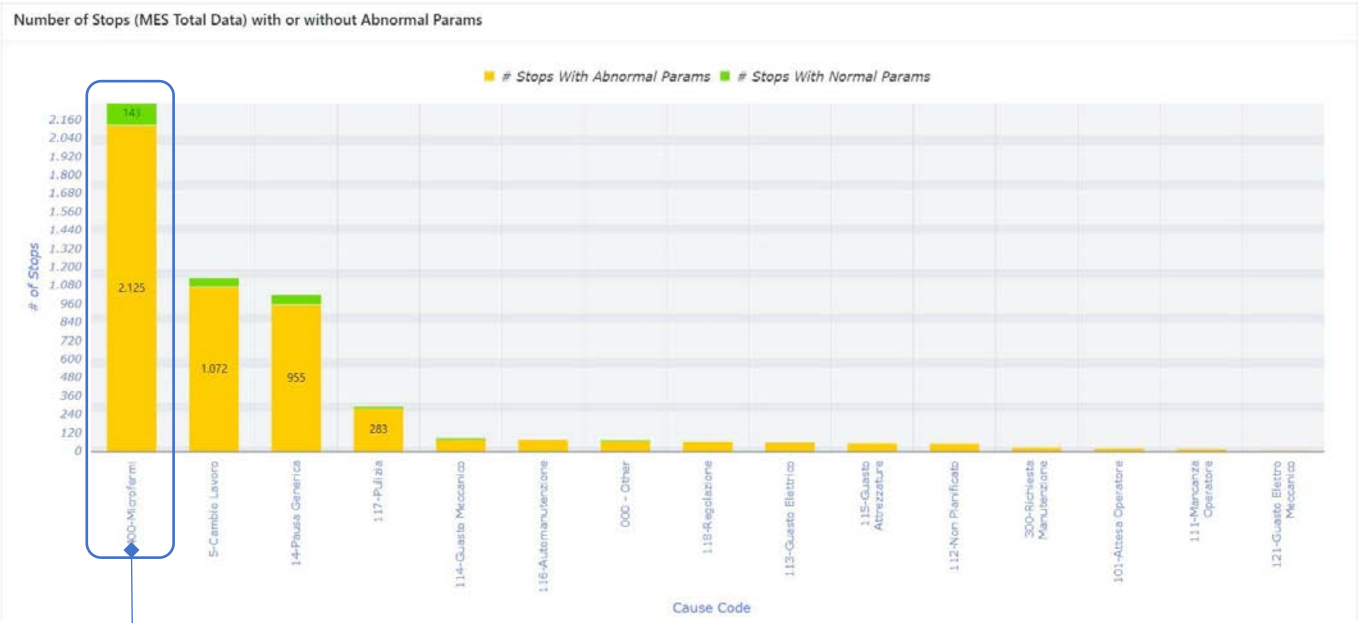


Figure 45: Number of Stops Machine

Pattern Analysis 400-Microfermi

V2	T2	PM	PS	SM	TC	# Stops per Pattern	Pattern % Share	% Graph
X	-	-	-	-	-	755	33,3%	
-	X	-	-	-	-	618	27,2%	
-	X	-	X	-	-	513	22,6%	
-	-	-	-	-	-	143	6,3%	
X	X	-	-	-	-	139	6,1%	
X	X	X	-	-	-	52	2,3%	
X	X	-	X	-	-	13	0,6%	
-	-	-	X	-	-	10	0,4%	
-	X	-	X	X	-	4	0,2%	
X	X	X	X	X	-	3	0,1%	
X	X	-	X	X	-	3	0,1%	
-	-	X	X	-	-	2	0,1%	
X	X	-	-	X	-	2	0,1%	
-	X	X	-	-	-	2	0,1%	
X	X	X	X	-	-	2	0,1%	

Pattern Detail 400-Microfermi

row(s) 1 - 15 of 755 Next

Stop DateTime	V2	T2	PM	PS	SM	TC
2018-06-15 14:53:00	64	115	244	894	22	266
2018-06-23 07:26:00	114	94	251	1067	24	88
2018-06-23 08:25:00	117	98	248	1054	17	312
2018-06-23 08:35:00	116	95	228	976	21	216
2018-06-23 10:31:00	116	91	240	1024	24	128
2018-06-23 10:53:00	115	94	246	1054	24	122
2018-06-23 11:52:00	115	95	247	1054	22	210
2018-06-23 17:18:00	118	92	235	1006	22	127
2018-06-23 19:57:00	118	90	234	1002	24	133
2018-06-23 21:42:00	115	91	227	980	26	245
2018-06-23 23:54:00	114	95	225	976	24	178
2018-06-24 02:34:00	116	92	225	972	24	203
2018-06-24 04:16:00	120	94	254	1059	17	219
2018-06-24 04:54:00	115	91	231	993	25	194
2018-06-24 05:03:00	120	91	251	1063	21	150

Figure 46: Restart Machine Pattern Analysis

4.2.6.3 CLAMPING SYSTEM PATTERN (P3) APPLIED TO MECHANICAL FAILURE

The pattern P3 aims to predict the mechanical breakdown. The pattern detects the correlation between the following two parameters:

- Clamping Force FC (kN)
- Multiplied Pressure PM (Pa)

The dashboard percentage of mechanical failure in Figure 47 has been created to comprehend the number of occurrences of mechanical maintenance for each month.

The yellow dashed lines represent the thresholds values. Each intended line represents the trend of each mechanical maintenance downtime occurrence over time. The result is that FC is inversely proportional to PM. It means that if an occurrence presents values of FC or PM above or below the line it is possible to detect possible incoming problems of maintenance.

In the first scenario in Figure 48, the first four occurrences of the month of June 2018 have been investigated. PM operate below the lower range than FC is in range.

In the second scenario in Figure 49, the last four occurrences of the month of June 2018 have been investigated to demonstrate that in this case the situation is inverse. FC operate above the upper range than PM is in range.

The outcome is the possibility to predict failure of the system based on the evaluation of the correlations between FC and PM.

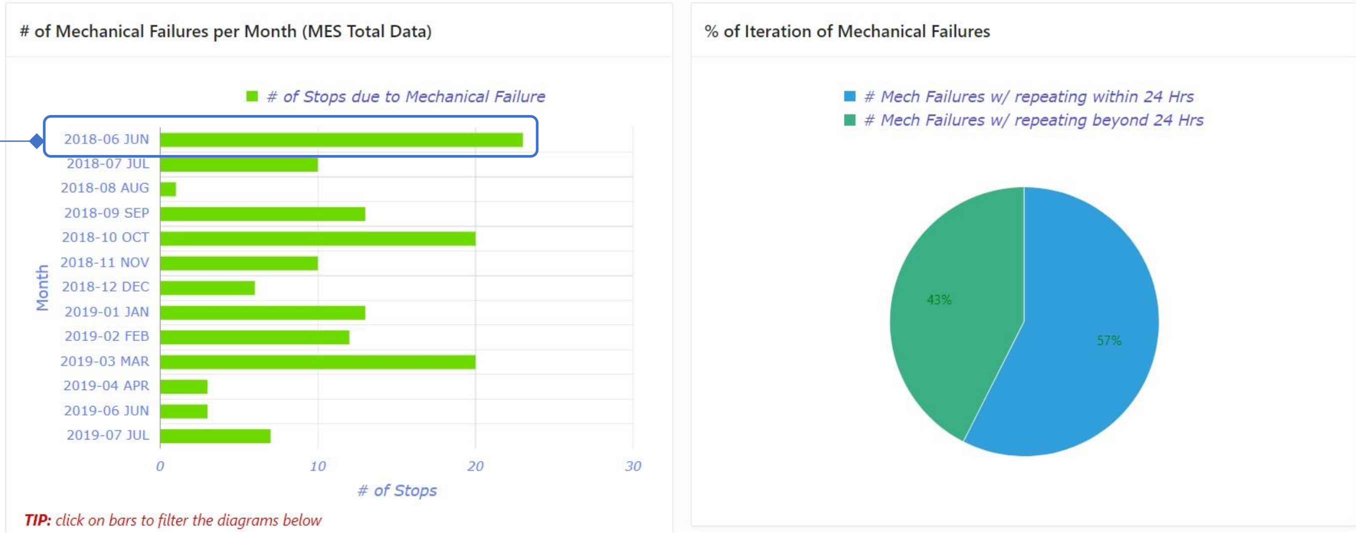


Figure 47: Mechanical Failures

The first scenario

Figure 48: Scenario I: FC and PM Trends over Time

The second scenario

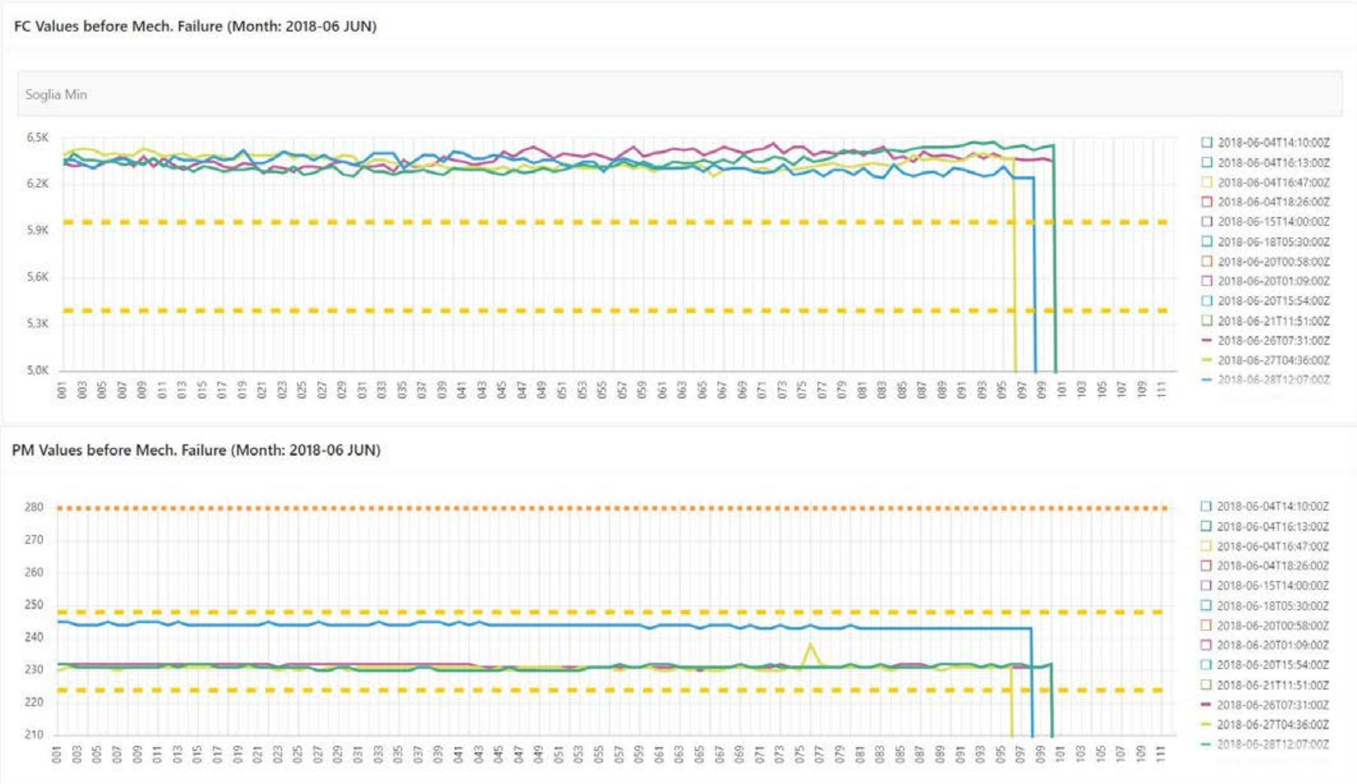


Figure 49: Scenario II: FC and PM Trends over Time

4.2.7 DESIGN CRITERIA FOR DIGITAL TWIN

The patterns described (P1, P2, P3) above have been used to design the digital twin of the die casting process in order to predict micro-stop and mechanical problems as shown in Figure 50.
The solutions here proposed consists of a realistic production environment but “augmented” with intrinsic technological knowledge. With our approach, the physical settings interact with the digital space, according to specific properties and rules, to understand:

- The behaviour of the process.
- The correlations between technological parameters.
- The correlations between parameters and effects like quality defects, maintenances problems etc.

The digital twin has been designed to support the employees in decision-making process to:

- Identify autonomously the several quality problems of the components, compared to the standards (dimensions, tolerances, finishes, quantity).
- Alert operators through proper alarm systems about abnormal or out of tolerance situations.
- Analyse and correlate the symptoms and causes of failures and defects in production.
- Support the choice of corrective actions to eliminate the detected failures and defects.

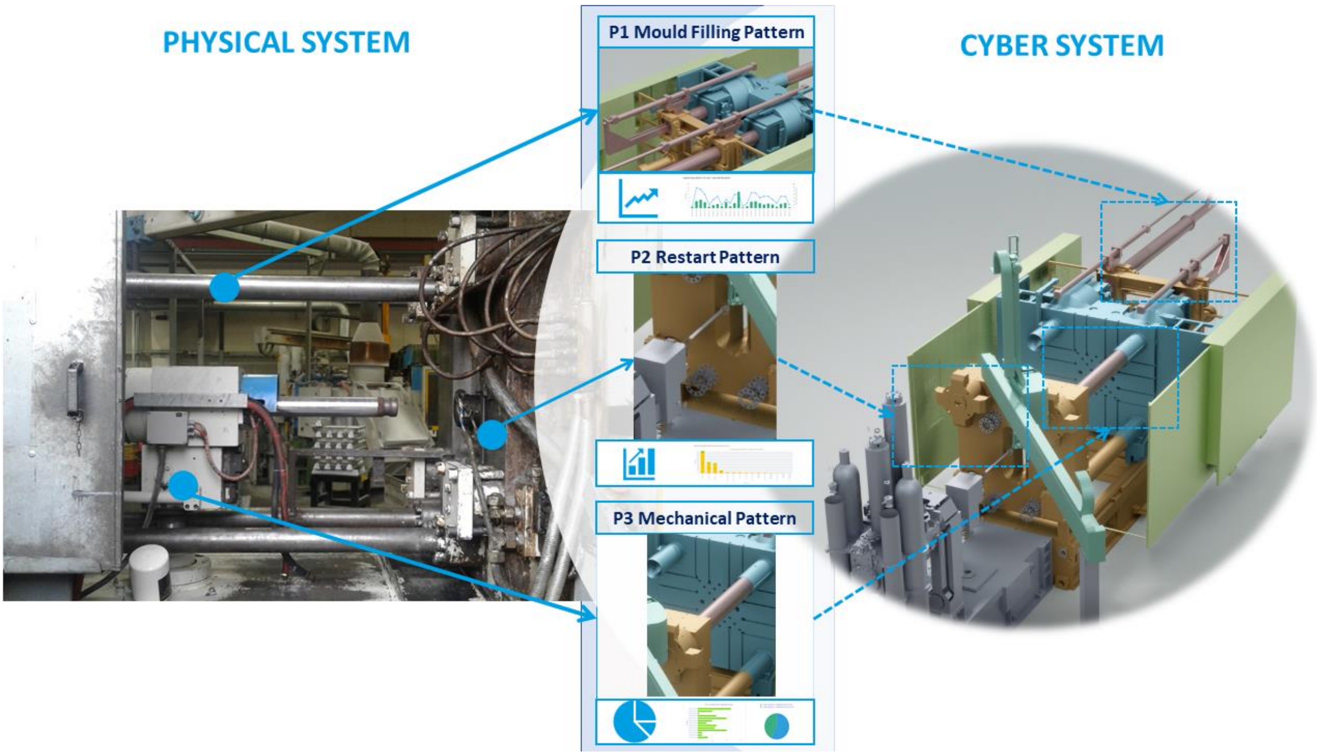


Figure 50: Digital Twin

The gathered data from the shop floor level are aggregated, analysed and interpreted in the digital environment according to the conditions listed in table 7. The conditions are defined and formalized based on the patterns showed and analysed in the platform. A condition is associated to a possible state that a digital twin should detect and recognize.

Table 7: DT Conditions

STATE	VERIFY STATE	CONDITIONS	DECISION-SUPPORT (ACTION TO DO)	ID PATTERN
0	"Machine Works without Problems"	-	-	-
1	"Prediction Micro-Stop"	SM_LimLow<=SM<=SM_LimUpp	Clean the machine otherwise recasting of 5 die cast in 5 minutes	Mould Filling Pattern (P1)
2	"Recasting for Micro-Stop"	V2<=LimLow OR V2=>LimUpp OR T2<=LimLow 0R T2>= LimUpp PM<=LimLow OR PM>=LimUpp OR PS<=LimLow OR PS >=LimUpp SM>=LimUpp OR TC>=LimUpp	The machine was just restarted. Recasting product in progress	Machine Restart Pattern (P2)
3	"Prediction Mechanical Maintenance"	LimLow<=PM<=LimUpp OR LimLow<=PM<=LimUpp OR	Perform mechanical maintenance on the mould otherwise	Clamping System Pattern (P3)

		LimLow<=FC<=LimUpp OR LimLow<=FC<=LimUpp	machine downtime in one hour (estimated time of intervention 1.5h)	
4	"Recasting for Micro-Stop"	SM<=LimLow	Recasting in progress for micro-stop	Mould Filling Pattern (P1)
5	"Stop Machine for Mechanical Maintenance"	FC<=LimLow OR FC>=LimUpp	Stop machine for mechanical maintenance	Clamping System Pattern (P3)

The conditions and states are encapsulated in an algorithm presented below to make the digital twin able to judge, evaluate, optimize and/or predict:

→Verify State "Prediction Micro-Stop"

```
IF cur_stringa.SM BETWEEN LimLow AND LimUpp THEN v_stato_rkd := 1; END IF;
```

→Verify State "Re-start Machine"

```
IF v_stato_rkd = 0 AND
     (cur_stringa.V2 <= LimLow OR cur_stringa.V2 >= LimUpp
       OR cur_stringa.T2 <= LimLow OR cur_stringa.T2 >= LimUpp
      OR cur_stringa.PM <= LimLow OR cur_stringa.PM >= LimUpp
     OR cur_stringa.PS <= LimLow OR cur_stringa.PS >= LimUpp
      OR cur_stringa.SM >= LimUpp
     OR cur_stringa.TC >= LimUpp
      )
THEN v_stato_rkd := 2;
END IF;
```

→Verify State "Prediction Mechanical Maintenance"

```
IF v_stato_rkd = 0 AND cur_stringa.TIMESTAMPLOCAL >= TO_DATE('01-03-2019','DD-MM-YYYY')
AND
   (cur_stringa.PM BETWEEN LimLow AND LimUpp)
   OR
   (cur_stringa.FC BETWEEN LimLow AND LimUpp) OR (cur_stringa.FC BETWEEN LimLow AND
  LimUpp)
  )
THEN
v_stato_rkd := 3;
END IF;

IF v_stato_rkd = 0 AND cur_stringa.TIMESTAMPLOCAL < TO_DATE('01-03-2019','DD-MM-YYYY') AND
  ( (cur_stringa.PM LimLow AND LimUpp) OR
 (cur_stringa.FC BETWEEN LimLow AND LimUpp) OR (cur_stringa.FC BETWEEN LimLow AND
LimUpp)
  )
THEN
```

```
v_stato_rkd := 3;
END IF;
```

→Verify "Recasting for Micro-Stop"

```
IF v_stato_rkd = 0 AND
 ( cur_stringa.SM <= LimLow
 )
THEN
v_stato_rkd := 4;
END IF;
```

→Verify State "Stop Machine for Mechanical Maintenance"

```
IF v_stato_rkd = 0 AND
 ( cur_stringa.FC <= LimLow OR cur_stringa.FC >= LimUpp
 )
THEN
v_stato_rkd := 5;
END IF;
```

The digital twin receives and reads a set of data, emulates the behaviour of the machine and recognizes the possible failures. An example of the potentials and the results has been presented in figure 51. The digital twin detects different states based on the same set of parameters as shown in pattern P6. In this way the digital twin predicts possible problems and prescribes actions based on the knowledge extracted, verified and formalized above (Table 7).

Stato Record	Meassetid ↓	Timestamplocal
2	1030030	07-09-2019 18:46:00
4	1030029	07-09-2019 18:43:00
1	1030028	07-09-2019 18:42:30
1	1030027	07-09-2019 18:42:00
0	1030026	07-09-2019 18:41:30
0	1030025	07-09-2019 18:41:00
0	1030024	07-09-2019 18:40:00
0	1030023	07-09-2019 18:39:30
0	1030022	07-09-2019 18:39:00

Figure 51: Digital Twin Emulation

Reading and evaluating the operating profile shown in Figure 52, the digital twin is able to recognize the state of the machine in the following sequence:

1. 07-09-2019 18:39:30→RECOGNIZE STATE 0: The digital twin does not detect any problem as shown in Figure 54.

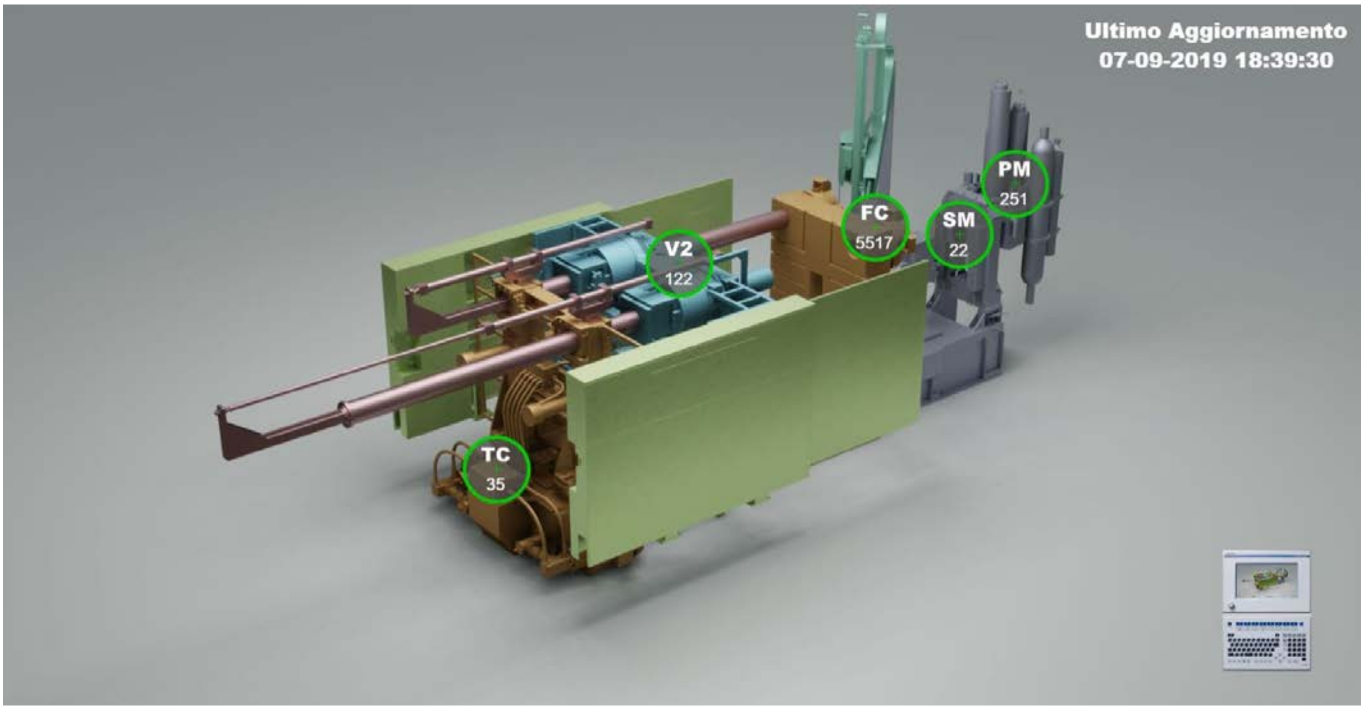


Figure 52: State 0

2. 07-09-2019 18:42:00→RECOGNIZE STATE 1: The digital twin predicts that a problem of micro-stop is going to happen in 5 minutes based on the values of C1, C2, CC and SM. The DT prescribes to the employees which action performs as shown in Figure 53.

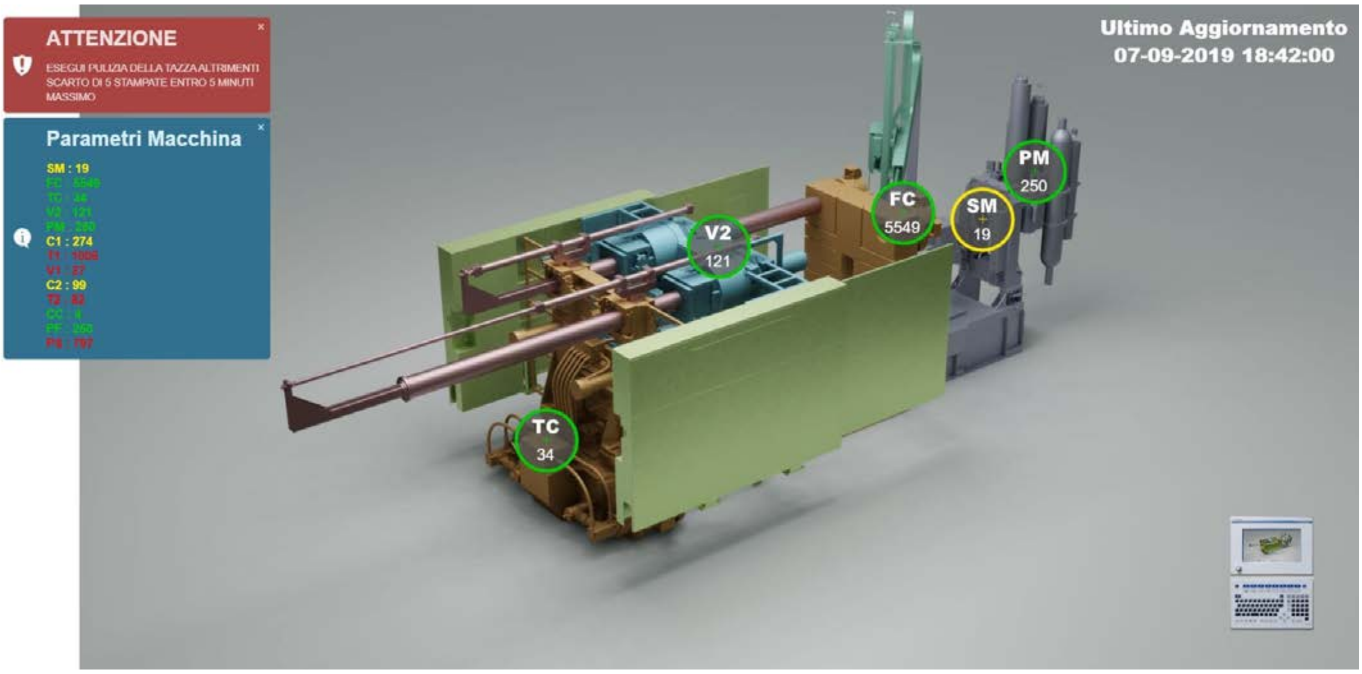


Figure 53: State 1

3. 07-09-2019 18:43:00→RECOGNIZE STATE 4: The digital twin informs the employees that a problem of micro-stop occurred because they did not perform the action prescribed in the state 1, as shown in Figure 54.

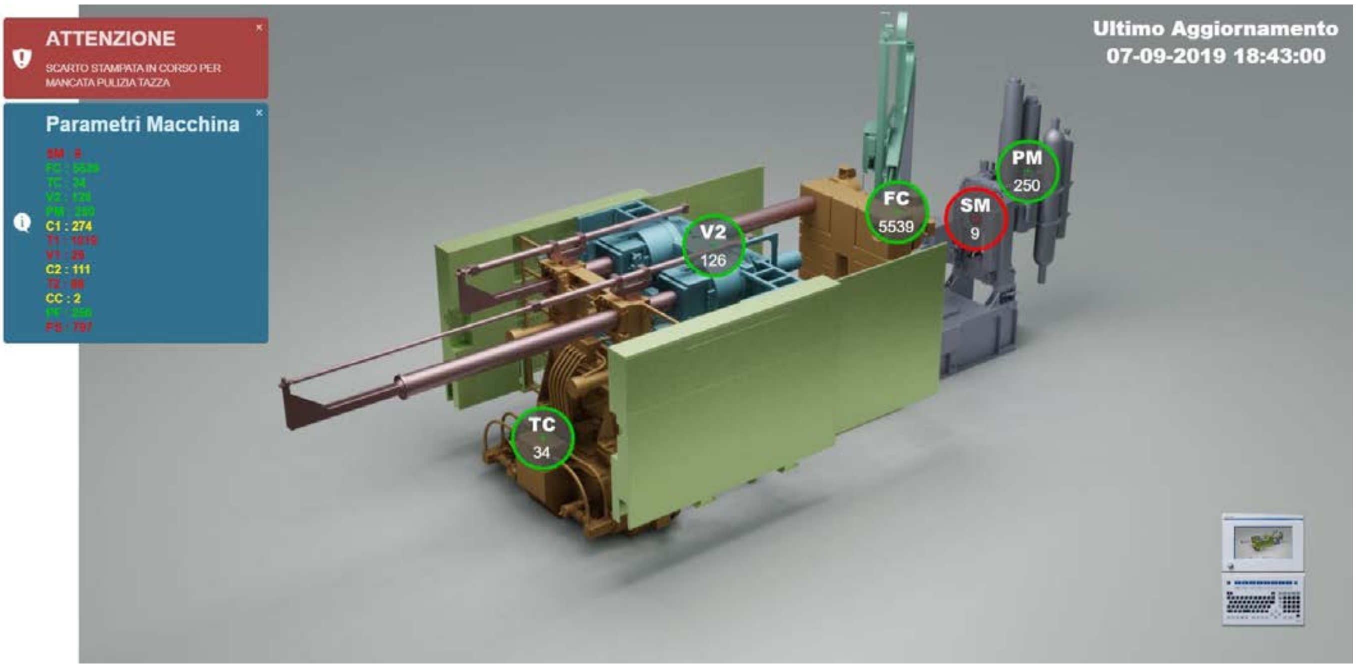


Figure 54: State 4

4. 07-09-2019 18:46:00→RECOGNIZE STATE 2: The digital twin detects in Figure 55 the restart machine after 3 minutes downtime due to the value of V2 and TC.

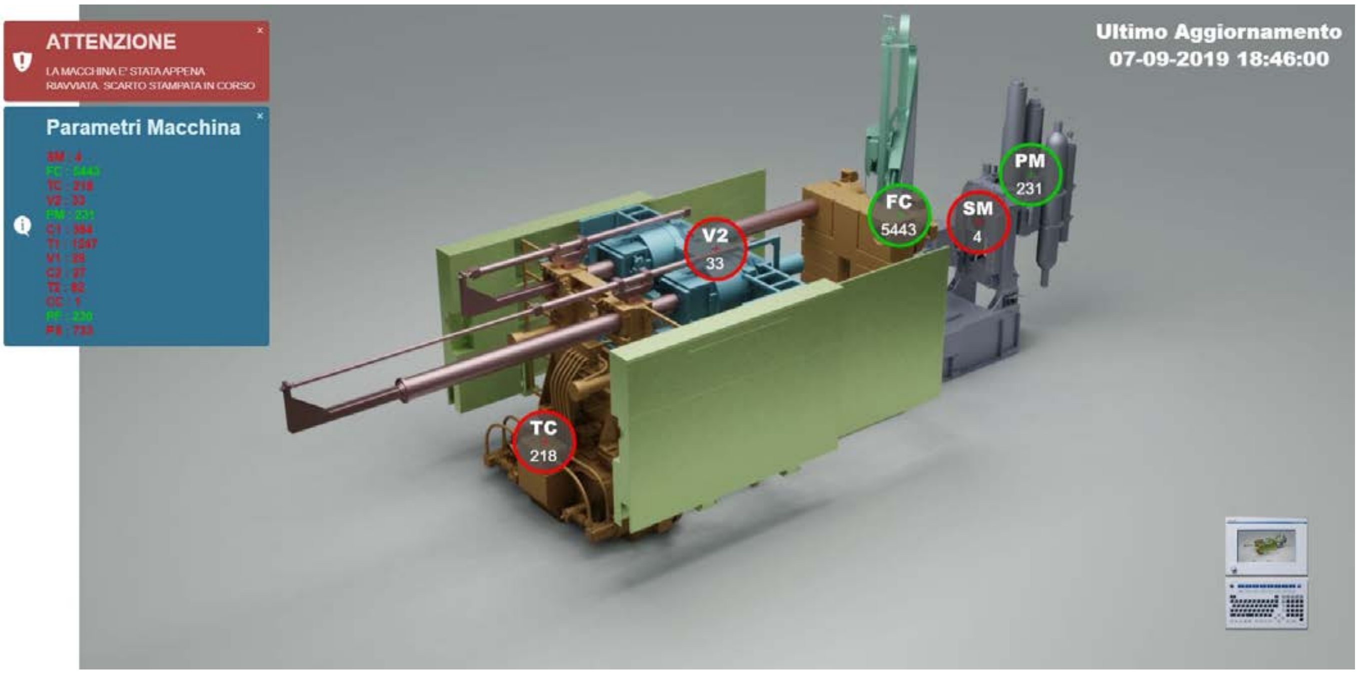


Figure 55: State 2

4.2.8 KNOWLEDGE RE-USE

The formalized knowledge can be used for design digital twin of other manufacturing processes of the company. The goal is to apply the same data-driven pattern constructs in other domains e.g. customer's need, design product and monitoring product.

CONCLUSION

INTRODUCTION

This thesis contributes to define and to present how to construct data-driven invariant modelling constructs for digital modelling transformation.
In this chapter, we will firstly revisit the research problem, the state-of-the-art limitations as well as an overview of our contribution. After, we will identify directions for future research.

1. SUMMARY OF THE THESIS

In chapter 1 the digital transformation in SMEs has been analysed. In particular, the main problem related to the implementation of Industry of the future in SMEs is to understand the new technologies, and to define how, when and where these could be applied. The digital twin has been reviewed to solve this issue.
The digital twin is defined in this thesis as: << *An adaptive model that emulates the behaviour of a physical system getting real time data to update itself along its life cycle. The digital twin replicates the physical system to predict failures and opportunities for changing, in order to prescribe real time actions for optimizing and/or mitigating unexpected events observing and evaluating the operating system profile>>.*
The state of art of Digital Twin has led to define the research gaps. The invariant modelling is the research issue to address. A literature review on model-based, data-driven and hybrid approaches have been performed in Chapter 2. It allowed us to identify the approaches regarding model-based and data-driven approaches. Based on the review results, one significant limitation is that modelling digital copy of the physical system is quite complex and generally the modelling action has a specific application type.
For this reason, the core challenge is to create data-driven invariant modelling constructs to use and re-use towards different applications. The research is the formalization and the standardization of invariant modelling constructs for modelling different system.
The result is the formalization of an approach in Chapter 3 for discovering data-driven constructs. It involves the combination of data-driven and model-based approaches with design pattern to define and identify data-driven invariant modelling. The iterative approach consists of eight different stages:

1. Definition of the system to analysis with exergy analysis.
2. System Model through SysML.
3. Data Selection.
4. Knowledge Discovery through relation concept analysis.
5. Knowledge Extraction.
6. Knowledge Formalization.
7. Design Criteria.
8. Knowledge Re-use.

Finally, in Chapter 4, we evaluated our contribution trough a case study: Master Italy s.r.l. The die casting aluminium process is analysed based on the result of exergy analysis. The die casting process is modelled in SysML to describe the requirements, the structure and the behaviour. Relation concept analysis is used to detect automatically explicit and tacit associations in data. This has allowed to formalise data-driven invariant modelling constructs. Three patterns have been used to develop a web platform. A digital twin of the process has been realized using a set of data-driven constructs discovered.

2. PERSPECTIVES FOR FUTURE RESEARCH WORK

The work presented in this thesis induces other research paths that can be considered in the future:

- The first research direction is to enrich the pattern' semantics to create a comprehensive library of formalized data-driven patterns. It means to select different production lines to apply the same patterns for designing models for different applications and scopes. Other sets of data need to be selected such as logistic data, product data, customers' data to extract new data-driven constructs. In this way it is possible to create a consistent pattern library. A pattern library is a collection of patterns that describe a recurrent problem, behaviour or situation. In this way, data-driven patterns can be combined to create easily virtual models.

- The state of art demonstrates that model-based or data-driven approaches are more analysed than hybrid approaches. Therefore, the second research direction is to propose new contributions for the combination of model and data approaches for improving the modelling of systems.

- The third research direction is to explore the semantic and the syntax of patterns to better formalize the concept of the invariance. What is the invariance? How to define the invariance? How to recognize the invariance in patterns? These are some of the questions that could guide the research.

- The fourth research direction is to automatize the approach in order to recognize automatically new data-driven modelling constructs by using machine learning algorithm or neural networks. It means to set criteria or common elements to perform the recognition.

ANNEXE - CASE STUDY

Table 8: Cluster Definition

PROBLEM	CLUSTER ID	CLUSTER_LABEL	NUMBER OF CONCEPT
---	C1	IF PARAMETER THEN PARAMETER	176.610
MICROSTOP FOR SM_LimLow	C8	IF PARAMETER THEN PROBLEM AND PARAMETER	214
	C2	IF PROBLEM AND PARAMETER THEN PARAMETER	155
	C9	IF PROBLEM AND PARAMETER THEN PROBLEM AND PARAMETER	712
	C4	IF PROBLEM THEN PARAMETER	2
	C5	IF PROBLEM THEN PROBLEM AND PARAMETER	1
RECASTING START MACHINE AFTER CLEANING OF THE EQUIPMENT	C2	IF PROBLEM AND PARAMETER THEN PARAMETER	199
	C4	IF PROBLEM THEN PARAMETER	1
RECASTING START MACHINE AFTER MECHANICAL MAINTENANCE	C2	IF PROBLEM AND PARAMETER THEN PARAMETER	34
	C4	IF PROBLEM THEN PARAMETER	1
RECASTING START MACHINE AFTER MECHANICAL CONTROL	C2	IF PROBLEM AND PARAMETER THEN PARAMETER	10
	C4	IF PROBLEM THEN PARAMETER	1
RECASTING START MACHINE FOR GENERIC MICROSTOP	C2	IF PROBLEM AND PARAMETER THEN PARAMETER	163
	C4	IF PROBLEM THEN PARAMETER	1
RECASTING START MACHINE FOR MICROSTOP (after SM_LimLow)	C8	IF PARAMETER THEN PROBLEM AND PARAMETER	474
	C2	IF PROBLEM AND PARAMETER THEN PARAMETER	497
	C4	IF PROBLEM THEN PARAMETER	1
RUL_COMPONENT PROBLEM	C8	IF PARAMETER THEN PROBLEM AND PARAMETER	11
	C2	IF PROBLEM AND PARAMETER THEN PARAMETER	67
	C9	IF PROBLEM AND PARAMETER THEN PROBLEM AND PARAMETER	3

	C4	IF PROBLEM THEN PARAMETER	3
	C5	IF PROBLEM THEN PROBLEM AND PARAMETER	1
RUL_ELECTRICAL MAINTENANCE	C8	IF PARAMETER THEN PROBLEM AND PARAMETER	20
	C2	IF PROBLEM AND PARAMETER THEN PARAMETER	86
	C4	IF PROBLEM THEN PARAMETER	1
RUL_MECHANICAL MAINTENANCE	C3	IF PARAMETER THEN PROBLEM	8
	C8	IF PARAMETER THEN PROBLEM AND PARAMETER	476
	C2	IF PROBLEM AND PARAMETER THEN PARAMETER	593
	C4	IF PROBLEM THEN PARAMETER	1
THICKNESS<=14	C8	IF PARAMETER THEN PROBLEM AND PARAMETER	244
	C2	IF PROBLEM AND PARAMETER THEN PARAMETER	138
	C9	IF PROBLEM AND PARAMETER THEN PROBLEM AND PARAMETER	712
	C4	IF PROBLEM THEN PARAMETER	2
	C5	IF PROBLEM THEN PROBLEM AND PARAMETER	1
THICKNESS=15	C2	IF PROBLEM AND PARAMETER THEN PARAMETER	1.369
THICKNESS=16	C2	IF PROBLEM AND PARAMETER THEN PARAMETER	1.534
	C9	IF PROBLEM AND PARAMETER THEN PROBLEM AND PARAMETER	3
	C4	IF PROBLEM THEN PARAMETER	3
	C5	IF PROBLEM THEN PROBLEM AND PARAMETER	1

1. AN EXAMPLE OF SQL QUERY FOR MOULD FILLING PATTERN (P1)

Daily Micro-Stops Detail (Figure 43):

```
WITH DATI_MICROFERMI
AS
(
SELECT DISTINCT
      DATA_ORA,
      TO_CHAR(DATA_ORA,'DD-MM-YYYY HH24:MI:SS') DATAORA
  FROM MES_DATI
 WHERE CAUSALE = :P121_CAUSALE
   ORDER BY 1
),
DATI_MESE
AS
(
SELECT MAX(m.DATA_ORA) DATA_ORA_FG,
      mf.DATA_ORA DATA_ORA_MF,
      (mf.DATA_ORA-MAX(m.DATA_ORA))*(24*60) DURATA_FERMO_MIN
  FROM MES_DATI m
 INNER JOIN DATI_MICROFERMI mf ON mf.DATA_ORA >= m.DATA_ORA
 WHERE CAUSALE = 99
 AND TRUNC(m.DATA_ORA) BETWEEN TRUNC(TO_DATE(:P121_MESE,'DD-MM-YYYY'),'MONTH')-1 AND
TRUNC(LAST_DAY(TO_DATE(:P121_MESE,'DD-MM-YYYY')))
HAVING (mf.DATA_ORA-MAX(m.DATA_ORA))*(24*60)>=1
 GROUP BY
      mf.DATA_ORA
)
SELECT TO_CHAR(DATA_ORA_FG,'DD-MM-YYYY HH24:MI:SS') DATA_ORA_FG,
      TO_CHAR(DATA_ORA_MF,'DD-MM-YYYY HH24:MI:SS') DATA_ORA_MF,
      LAG(DATA_ORA_FG, 1, TRUNC(DATA_ORA_FG)) OVER (ORDER BY DATA_ORA_FG) AS
DATA_ORA_FERMO_PREC,
      (DATA_ORA_FG-(LAG(DATA_ORA_FG, 1, TRUNC(DATA_ORA_FG)) OVER (ORDER BY
DATA_ORA_FG)))*24*60 AS MIN_DA_FERMO_PREC,
      CASE WHEN (DATA_ORA_FG-(LAG(DATA_ORA_FG, 1, TRUNC(DATA_ORA_FG)) OVER (ORDER BY
DATA_ORA_FG)))*24*60>30 THEN 30
         ELSE (DATA_ORA_FG-(LAG(DATA_ORA_FG, 1, TRUNC(DATA_ORA_FG)) OVER (ORDER BY
DATA_ORA_FG)))*24*60
      END AS DELTA_MINUTI,
      DURATA_FERMO_MIN
  FROM DATI_MESE
 WHERE TRUNC(DATA_ORA_FG) = TO_DATE(:P121_GIORNO,'DD-MM-YYYY')
```

2. AN EXAMPLE OF SQL QUERY FOR MACHINE RESTART PATTERN (P2)

Query PLC RESTART (Figure 47):

```
WITH DATI_FERMI
AS
(
SELECT DISTINCT
      DATA_ORA DATA_ORA_FERMO,
      TO_CHAR(DATA_ORA,'DD-MM-YYYY HH24:MI:SS') DATAORA_FERMO,
      CAUSALE,
      DESCRIZIONE_SOSPENSIONE
  FROM MES_DATI
 WHERE CAUSALE IN ('112','113','114','115')
   AND TRUNC(DATA_ORA) >= TO_DATE('04-06-2018','dd-mm-yyyy')
   ORDER BY 1
),
DATI_PLC_GUASTO
AS
(
SELECT plc.MEASSETID,
      MIN(plc.MEASSETID) OVER (PARTITION BY df.DATA_ORA_FERMO) MIN_ID,
      plc.TIMESTAMPLOCAL,
      TO_CHAR(plc.TIMESTAMPLOCAL,'dd-mm-yyyy hh24:mi:ss') ORA_PLC,
      TO_CHAR(df.DATA_ORA_FERMO,'dd-mm-yyyy hh24:mi:ss') AS DATAORA_GUASTO,
      df.CAUSALE,
      df.DESCRIZIONE_SOSPENSIONE,
      plc.C1,
      plc.MIN_C1,
      plc.MAX_C1,
      plc.T1,
      plc.MIN_T1,
      plc.MAX_T1,
      plc.V1,
      plc.MIN_V1,
      plc.MAX_V1,
      plc.C2,
      plc.MIN_C2,
      plc.MAX_C2,
      plc.T2,
      plc.MIN_T2,
      plc.MAX_T2,
      plc.V2,
      plc.MIN_V2,
      plc.MAX_V2,
      plc.CC,
      plc.MIN_CC,
      plc.MAX_CC,
      plc.PM,
      plc.MIN_PM,
      plc.MAX_PM,
      plc.PF,
```

```
      plc.MIN_PF,
      plc.MAX_PF,
      plc.PS,
      plc.MIN_PS,
      plc.MAX_PS,
      plc.FC,
      plc.MIN_FC,
      plc.MAX_FC,
      plc.SM,
      plc.MIN_SM,
      plc.MAX_SM,
      plc.TC,
      plc.MIN_TC,
      plc.MAX_TC
 FROM PLC_DATI plc
INNER JOIN DATI_FERMI df ON plc.TIMESTAMPLOCAL BETWEEN (df.DATA_ORA_FERMO)-1/(24*60) AND
(df.DATA_ORA_FERMO)+1/(24*60)
)
(
SELECT g.*,
       FROM DATI_PLC_GUASTO g
 WHERE MEASSETID = MIN_ID
);
```

3. AN EXAMPLE OF SQL QUERY FOR CLAMPING SYSTEM PATTERN (P3)

# of Mechanical Failures per Month (MES Total Data) (Figure 49):

```
WITH DATI AS
(
SELECT DISTINCT
      DATA_ORA,
      TO_CHAR(DATA_ORA,'DD-MM-YYYY HH24:MI:SS') DATAORA,
      TO_CHAR(DATA_ORA,'DD','NLS_DATE_LANGUAGE=ITALIAN') GIORNO,
      TO_CHAR(DATA_ORA,'YYYY-MM MON','NLS_DATE_LANGUAGE=ENGLISH') MESE,
      TO_CHAR(LAST_DAY(DATA_ORA),'DD-MM-YYYY') DATAFINE_MESE
 FROM MES_DATI
 WHERE CAUSALE = 114
ORDER BY DATA_ORA
)
SELECT DATAFINE_MESE,
      MESE,
      COUNT(*) N_EVENTI
 FROM DATI
 GROUP BY
      DATAFINE_MESE,
      MESE
 ORDER BY MESE
```

% of Iteration of Mechanical Failures:

```
 WITH DATI AS
(
SELECT DISTINCT
      DATA_ORA,
      TO_CHAR(DATA_ORA,'DD-MM-YYYY HH24:MI:SS') DATAORA,
      TO_CHAR(DATA_ORA,'DD','NLS_DATE_LANGUAGE=ITALIAN') GIORNO,
      TO_CHAR(DATA_ORA,'YYYY-MM MONTH') MESE
  FROM MES_DATI
 WHERE CAUSALE = 114
ORDER BY DATA_ORA
),
DATI_ELAB
AS
(
SELECT DATA_ORA,
      TO_CHAR(DATA_ORA,'DD-MM-YYYY HH24:MI:SS') DATAORA_CHR,
      CASE WHEN (((LEAD(DATA_ORA, 1, DATA_ORA) OVER (ORDER BY DATA_ORA)) - DATA_ORA)*24*60)
BETWEEN 1 AND 1440 THEN 1 ELSE 0 END AS FLG_FERMO_1GG
  FROM DATI
)
SELECT LABEL,
      N_CASI
  FROM
(
SELECT '# Mech Failures w/ repeating within 24 Hrs' AS LABEL,
      COUNT(*) N_CASI
  FROM DATI_ELAB
 WHERE FLG_FERMO_1GG = 1
UNION ALL
SELECT '# Mech Failures w/ repeating beyond 24 Hrs' AS LABEL,
      COUNT(*) N_CASI
  FROM DATI_ELAB
 WHERE FLG_FERMO_1GG = 0
)
```

LIST OF FIGURES

LIST OF TABLES

BIBLIOGRAPHY

Abele, Eberhard, Joachim Metternich, Michael Tisch, George Chryssolouris, Wilfried Sihn, Hoda ElMaraghy, Vera Hummel, and Fabian Ranz. 2015. 'Learning Factories for Research, Education, and Training'. *Procedia CiRp* 32: 1–6.

Abramovici, Michael, Jens Christian Göbel, and Philipp Savarino. 2017. 'Reconfiguration of Smart Products during Their Use Phase Based on Virtual Product Twins'. *CIRP Annals* 66 (1): 165–168.

Ackoff, Russell L. 1989. 'From Data to Wisdom'. *Journal of Applied Systems Analysis* 16 (1): 3–9.

Adolph, Siri, Michael Tisch, and Joachim Metternich. 2014. 'Challenges and Approaches to Competency Development for Future Production'. *Journal of International Scientific Publications–Educational Alternatives* 12: 1001–10.

Agrawal, Rakesh, Tomasz Imieliński, and Arun Swami. 1993. 'Mining Association Rules between Sets of Items in Large Databases'. In *Acm Sigmod Record*, 22:207–216. ACM.

Ahmed, Syed Riaz. 2004. 'Applications of Data Mining in Retail Business'. In *International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004.*, 2:455–459. IEEE.

Alam, Kazi Masudul, and Abdulmotaleb El Saddik. 2017. 'C2PS: A Digital Twin Architecture Reference Model for the Cloud-Based Cyber-Physical Systems'. *IEEE Access* 5: 2050–2062.

Alcácer, V., and V. Cruz-Machado. 2019. 'Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems'. *Engineering Science and Technology, an International Journal*.

Alexander, Christopher. 1977. *A Pattern Language: Towns, Buildings, Construction*. Oxford university press.

Al-Sahaf, Harith, Ying Bi, Qi Chen, Andrew Lensen, Yi Mei, Yanan Sun, Binh Tran, Bing Xue, and Mengjie Zhang. 2019. 'A Survey on Evolutionary Machine Learning'. *Journal of the Royal Society of New Zealand* 0 (0): 1–24. https://doi.org/10.1080/03036758.2019.1609052.

Andrews, Simon. 2009. 'In-Close, a Fast Algorithm for Computing Formal Concepts'.

Antoy, Sergio, and Michael Hanus. 2002. 'Functional Logic Design Patterns'. In *International Symposium on Functional and Logic Programming*, 67–87. Springer.

Arrichiello, Vincenzo, and Paola Gualeni. 2019. 'Systems Engineering and Digital Twin: A Vision for the Future of Cruise Ships Design, Production and Operations'. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 1–8.

Asimov, R. M., S. V. Chernoshey, I. Kruse, and V. S. Osipovich. 2018. 'Digital Twin in the Analysis of a Big Data'.

Atzori, Luigi, Antonio Iera, and Giacomo Morabito. 2010. 'The Internet of Things: A Survey'. *Computer Networks* 54 (15): 2787–2805.

Autiosalo, J. 2018. 'Platform for Industrial Internet and Digital Twin Focused Education, Research, and Innovation: Ilmatar the Overhead Crane'. In *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, 241–44. https://doi.org/10.1109/WF-IoT.2018.8355217.

Ayani, M., M. Ganebäck, and Amos H. C. Ng. 2018. 'Digital Twin: Applying Emulation for Machine Reconditioning'. *Procedia CIRP*, 51st CIRP Conference on Manufacturing Systems, 72 (January): 243–48. https://doi.org/10.1016/j.procir.2018.03.139.

Baena, Felipe, Alvaro Guarin, Julian Mora, Joel Sauza, and Sebastian Retat. 2017. 'Learning Factory: The Path to Industry 4.0'. *Procedia Manufacturing* 9: 73–80.

Bakshi, Bhavik R., Timothy G. Gutowski, and Dusan P. Sekulic. 2011. *Thermodynamics and the Destruction of Resources*. Cambridge University Press.

Bao, Jinsong, Dongsheng Guo, Jie Li, and Jie Zhang. 2018. 'The Modelling and Operations for the Digital Twin in the Context of Manufacturing'. *Enterprise Information Systems* 0 (0): 1–23. https://doi.org/10.1080/17517575.2018.1526324.

Barricelli, Barbara Rita, Elena Casiraghi, and Daniela Fogli. 2019. 'A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications'. *IEEE Access* 7: 167653–167671.

Bass, Len, Paul Clements, and Rick Kazman. 2003. *Software Architecture in Practice*. Addison-Wesley Professional.

Bibby, Lee, and Benjamin Dehe. 2018. 'Defining and Assessing Industry 4.0 Maturity Levels – Case of the Defence Sector'. *Production Planning & Control* 29 (12): 1030–43. https://doi.org/10.1080/09537287.2018.1503355.

Bishop, C.M. 2016. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer New York. https://books.google.fr/books?id=kOXDtAEACAAJ.

Blochwitz, Torsten, Martin Otter, Martin Arnold, Constanze Bausch, H. Elmqvist, A. Junghanns, J. Mauß, M. Monteiro, T. Neidhold, and Dietmar Neumerkel. 2011. 'The Functional Mockup Interface for Tool Independent Exchange of Simulation Models'. In *Proceedings of the 8th International Modelica Conference; March 20th-22nd; Technical Univeristy; Dresden; Germany*, 105–114. Linköping University Electronic Press.

Blume, Marco, Nils Koch, Jahanzaib Imtiaz, Holger Flatt, Jürgen Jasperneite, Miriam Schleipen, Olaf Sauer, and Steffen Dosch. 2014. 'An OPC-UA Based Approach for Dynamic-Configuration of Security Credentials and Integrating a Vendor Independent Digital Product Memory'. In *Jahreskolloquium Kommunikation in Der Automation (KommA 2014)*.

Bohlin, R., J. Hagmar, K. Bengtsson, L. Lindkvist, J.S. Carlson, and R. Söderberg. 2017. 'Data Flow and Communication Framework Supporting Digital Twin for Geometry Assurance'. In . Vol. 2. https://doi.org/10.1115/IMECE2017-71405.

Brenner, Beate, and Vera Hummel. 2016. ‘A Seamless Convergence of the Digital and Physical Factory Aiming in Personalized Product Emergence Process (PPEP) for Smart Products within ESB Logistics Learning Factory at Reutlingen University’. *Procedia CIRP* 54: 227–232.

Brettel, Malte, Niklas Friederichsen, Michael Keller, and Marius Rosenberg. 2014. ‘How Virtualization, Decentralization and Network Building Change the Manufacturing Landscape: An Industry 4.0 Perspective’. *International Journal of Mechanical, Industrial Science and Engineering* 8 (1): 37–44.

Bruynseels, Koen, Filippo Santoni de Sio, and Jeroen van den Hoven. 2018. ‘Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm’. *Frontiers in Genetics* 9: 31.

Büchi, Giacomo, Monica Cugno, and Rebecca Castagnoli. 2020. ‘Smart Factory Performance and Industry 4.0’. *Technological Forecasting and Social Change* 150 (January): 119790. https://doi.org/10.1016/j.techfore.2019.119790.

Buczak, Anna L., and Erhan Guven. 2015. ‘A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection’. *IEEE Communications Surveys & Tutorials* 18 (2): 1153–1176.

Budinsky, Frank J., Marilyn A. Finnie, John M. Vlissides, and Patsy S. Yu. 1996. ‘Automatic Code Generation from Design Patterns’. *IBM Systems Journal* 35 (2): 151–171.

Cachay, Jan, Jan Wennemer, Eberhard Abele, and Ralf Tenberg. 2012. ‘Study on Action-Oriented Learning with a Learning Factory Approach’. *Procedia-Social and Behavioral Sciences* 55: 1144–1153.

Cai, Yi, Binil Starly, Paul Cohen, and Yuan-Shin Lee. 2017. ‘Sensor Data and Information Fusion to Construct Digital-Twins Virtual Machine Tools for Cyber-Physical Manufacturing’. *Procedia Manufacturing* 10: 1031–1042.

Chen, Baotong, Jiafu Wan, Lei Shu, Peng Li, Mithun Mukherjee, and Boxing Yin. 2017. ‘Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges’. *IEEE Access* 6: 6505–6519.

Cheng, Jiangfeng, He Zhang, Fei Tao, and Chia-Feng Juang. 2020. ‘DT-II: Digital Twin Enhanced Industrial Internet Reference Framework towards Smart Manufacturing’. *Robotics and Computer-Integrated Manufacturing* 62: 101881.

Ciavotta, Michele, Marino Alge, Silvia Menato, Diego Rovere, and Paolo Pedrazzoli. 2017. ‘A Microservice-Based Middleware for the Digital Factory’. *Procedia Manufacturing* 11: 931–938.

Clarke, Bertrand, Ernest Fokoue, and Hao Helen Zhang. 2009. *Principles and Theory for Data Mining and Machine Learning*. Springer Science & Business Media.

Da Xu, Li, Wu He, and Shancang Li. 2014. ‘Internet of Things in Industries: A Survey’. *IEEE Transactions on Industrial Informatics* 10 (4): 2233–2243.

Dalle Mura, Michela, and Gino Dini. 2019. ‘Designing Assembly Lines with Humans and Collaborative Robots: A Genetic Approach’. *CIRP Annals* 68 (1): 1–4.

Damiani, Lorenzo, Melissa Demartini, Piero Giribone, Margherita Maggiani, Roberto Revetria, and Flavio Tonelli. 2018. ‘Simulation and Digital Twin Based Design of a Production Line: A Case Study.”’. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*. Vol. 2.

Dassisti, Michele, and Michele De Nicolò. 2012. ‘Enterprise Integration and Economical Crisis for Mass Craftsmanship: A Case Study of an Italian Furniture Company’. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, 113–123. Springer.

Dassisti, Michele, Antonio Giovannini, Pasquale Merla, Michela Chimienti, and Hervé Panetto. 2017. ‘Hybrid Production-System Control-Architecture for Smart Manufacturing’. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, 5–15. Springer.

Dassisti, Michele, Herve Panetto, Mario Lezoche, Pasquale Merla, Concetta Semeraro, Antonio Giovannini, and Michela Chimienti. 2017. ‘Industry 4.0 Paradigm: The Viewpoint of the Small and Medium Enterprises’. In *7th International Conference on Information Society and Technology, ICIST 2017*, 1:50–54.

Dassisti, Michele, and Concetta Semeraro. 2016. ‘Exergetic Control Charts (Variability Analysis in a Real Injection-Moulding Industrial Application)’. *Procedia CIRP* 41: 264–269.

Dassisti, Michele, Concetta Semeraro, and Michela Chimenti. 2019. ‘Hybrid Exergetic Analysis-LCA Approach and the Industry 4.0 Paradigm: Assessing Manufacturing Sustainability in an Italian SME’. *Procedia Manufacturing* 33: 655–662.

Dassisti, Michele, Noemi Siragusa, and Concetta Semeraro. 2018. ‘Exergetic Model as a Guideline for Implementing the Smart-Factory Paradigm in Small Medium Enterprises: The Brovedani Case’. *Procedia CIRP* 67: 534–539.

Dassisti, and Concetta Semeraro. 2018. ‘Smart Sustainable Manufacturing: A New Laboratory-Factory Concept to Test Industry 4.0 Principles’.

Ding, Kai, Felix T. S. Chan, Xudong Zhang, Guanghui Zhou, and Fuqiang Zhang. 2019. ‘Defining a Digital Twin-Based Cyber-Physical Production System for Autonomous Manufacturing in Smart Shop Floors’. *International Journal of Production Research* 0 (0): 1–20. https://doi.org/10.1080/00207543.2019.1566661.

Douglass, Bruce Powel. 2003. *Real-Time Design Patterns: Robust Scalable Architecture for Real-Time Systems*. Addison-Wesley Professional.

Drath, Rainer, and Alexander Horch. 2014. ‘Industrie 4.0: Hit or Hype?[Industry Forum]’. *IEEE Industrial Electronics Magazine* 8 (2): 56–58.

Drath, Rainer, Arndt Luder, Jorn Peschke, and Lorenz Hundt. 2008. ‘AutomationML-the Glue for Seamless Automation Engineering’. In *2008 IEEE International Conference on Emerging Technologies and Factory Automation*, 616–623. IEEE.

Džeroski, Sašo. 2003. 'Multi-Relational Data Mining: An Introduction'. *ACM SIGKDD Explorations Newsletter* 5 (1): 1–16.

EC. 2013. *Factories of the Future Multi-Annual Roadmap for the Contractual PPP under Horizon 2020*. Luxembourg: Publications Office of the European Union.

Erol, Selim, Andreas Jäger, Philipp Hold, Karl Ott, and Wilfried Sihn. 2016. 'Tangible Industry 4.0: A Scenario-Based Approach to Learning for the Future of Production'. *Procedia CIRP* 54: 13–18.

Estefan, Jeff A. 2007. 'Survey of Model-Based Systems Engineering (MBSE) Methodologies'. *Incose MBSE Focus Group* 25 (8): 1–12.

Fahad, Adil, Najlaa Alshatri, Zahir Tari, Abdullah Alamri, Ibrahim Khalil, Albert Y. Zomaya, Sebti Foufou, and Abdelaziz Bouras. 2014. 'A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis'. *IEEE Transactions on Emerging Topics in Computing* 2 (3): 267–279.

Faller, Clemens, and Dorothee Feldmüller. 2015. 'Industry 4.0 Learning Factory for Regional SMEs'. *Procedia CIRP* 32: 88–91.

Fayyad, Usama M., Gregory Piatetsky-Shapiro, and Padhraic Smyth. 1996. 'Knowledge Discovery and Data Mining: Towards a Unifying Framework.' In *KDD*, 96:82–88.

Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. 1996. 'From Data Mining to Knowledge Discovery in Databases'. *AI Magazine* 17 (3): 37–37.

Frank, Alejandro Germán, Lucas Santos Dalenogare, and Néstor Fabián Ayala. 2019. 'Industry 4.0 Technologies: Implementation Patterns in Manufacturing Companies'. *International Journal of Production Economics* 210 (April): 15–26. https://doi.org/10.1016/j.ijpe.2019.01.004.

Frank, Stephen, Michael Heaney, Xin Jin, Joseph Robertson, Howard Cheung, Ryan Elmore, and Gregor Henze. 2016. 'Hybrid Model-Based and Data-Driven Fault Detection and Diagnostics for Commercial Buildings'. National Renewable Energy Lab.(NREL), Golden, CO (United States).

Frawley, William J., Gregory Piatetsky-Shapiro, and Christopher J. Matheus. 1992. 'Knowledge Discovery in Databases: An Overview'. *AI Magazine* 13 (3): 57–57.

Friedenthal, Sanford, Alan Moore, and Rick Steiner. 2014. *A Practical Guide to SysML: The Systems Modeling Language*. Morgan Kaufmann.

Gamma, Erich, Richard Helm, and Ralph E. Johnson. 1995. *Design patterns: elements of reusable object-oriented software*. Reading (Mass.), Etats-Unis d'Amérique, Canada, France.

Ganter, Bernhard, Gerd Stumme, and Rudolf Wille. 2005. *Formal Concept Analysis: Foundations and Applications*. Vol. 3626. springer.

Garetti, Marco, Paolo Rosa, and Sergio Terzi. 2012. 'Life Cycle Simulation for the Design of Product–Service Systems'. *Computers in Industry* 63 (4): 361–369.

Ghobakhloo, Morteza. 2018. 'The Future of Manufacturing Industry: A Strategic Roadmap toward Industry 4.0'. *Journal of Manufacturing Technology Management* 29 (6): 910–936.

Ghosh, Kaushik, Yew Seng Ng, and Rajagopalan Srinivasan. 2011. 'Evaluation of Decision Fusion Strategies for Effective Collaboration among Heterogeneous Fault Diagnostic Methods'. *Computers & Chemical Engineering* 35 (2): 342–355.

Glaessgen, Edward, and David Stargel. 2012. 'The Digital Twin Paradigm for Future NASA and US Air Force Vehicles'. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA*, 1818.

Goebel, Kai, Neil Eklund, and Pierino Bonanni. 2006. 'Fusing Competing Prediction Algorithms for Prognostics'. In *2006 IEEE Aerospace Conference*, 10–pp. IEEE.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT press.

Gorunescu, Florin. 2011. *Data Mining: Concepts, Models and Techniques*. Vol. 12. Springer Science & Business Media.

Graessler, I., and A. Poehler. 2017. 'Integration of a Digital Twin as Human Representation in a Scheduling Procedure of a Cyber-Physical Production System'. In *Industrial Engineering and Engineering Management (IEEM), 2017 IEEE International Conference On*, 289–293. IEEE.

Grä\s sler, Iris, Alexander Pöhler, and Jens Pottebaum. 2016. 'Creation of a Learning Factory for Cyber Physical Production Systems'. *Procedia CIRP* 54: 107–112.

Grieves, Michael. 2014. 'Digital Twin: Manufacturing Excellence through Virtual Factory Replication'. *White Paper*.

Grira, Nizar, Michel Crucianu, and Nozha Boujemaa. 2004. 'Unsupervised and Semi-Supervised Clustering: A Brief Survey'. *A Review of Machine Learning Techniques for Processing Multimedia Content* 1: 9–16.

Gubbi, Jayavardhana, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. 2013. 'Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions'. *Future Generation Computer Systems* 29 (7): 1645–1660.

Guo, Jiapeng, Ning Zhao, Lin Sun, and Saipeng Zhang. 2018. 'Modular Based Flexible Digital Twin for Factory Design'. *Journal of Ambient Intelligence and Humanized Computing*, 1–12.

Gutowski, Timothy, Jeffrey Dahmus, and Alex Thiriez. 2006. 'Electrical Energy Requirements for Manufacturing Processes'. In *13th CIRP International Conference on Life Cycle Engineering*, 31:623–638.

Haag, Sebastian, and Reiner Anderl. 2018. 'Digital Twin – Proof of Concept'. *Manufacturing Letters*, Industry 4.0 and Smart Manufacturing, 15 (January): 64–66. https://doi.org/10.1016/j.mfglet.2018.02.006.

Hanachi, Houman, Wennian Yu, Il Yong Kim, Jie Liu, and Chris K. Mechefske. 2019. ‘Hybrid Data-Driven Physics-Based Model Fusion Framework for Tool Wear Prediction’. *The International Journal of Advanced Manufacturing Technology* 101 (9–12): 2861–2872.

Hansen, Pierre, and Brigitte Jaumard. 1997. ‘Cluster Analysis and Mathematical Programming’. *Mathematical Programming* 79 (1–3): 191–215.

Harris, Stephen. 1990. *Industry 4.0: The Next Industrial Revolution*. London: The Engineer.

Hermann, Mario, Tobias Pentek, and Boris Otto. 2016. ‘Design Principles for Industrie 4.0 Scenarios’. In *System Sciences (HICSS), 2016 49th Hawaii International Conference On*, 3928–3937. IEEE.

International Council on Systems Engineering (INCOSE). 2007. ‘International Council on Systems Engineering (INCOSE), Systems Engineering Vision 2020, Version 2.03, TP-2004-004-02, September 2007.’ 2007. http://www.omg.org/spec/OCL/.

ISO/IEC 19514. 2017. ‘ISO/IEC 19514:2017’. ISO. 2017. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/52/65231.html.

J. Miller, and J. Mukerji. 2003. ‘J. Miller and J. Mukerji. MDA Guide Version 1.0.1. Technical Report, Object Management Group (OMG),2003.’ 2003. https://www.omg.org/news/meetings/workshops/UML_2003_Manual/00-2_MDA_Guide_v1.0.1.pdf.

Jain, R., and R. Bhatnagar. 2020. ‘Applications of Machine Learning in Cyber Security - A Review and a Conceptual Framework for a University Setup’. *Advances in Intelligent Systems and Computing* 921: 599–608. https://doi.org/10.1007/978-3-030-14118-9_60.

Jain, Sanjay, Guodong Shao, and Seung-Jun Shin. 2017. ‘Manufacturing Data Analytics Using a Virtual Factory Representation’. *International Journal of Production Research* 55 (18): 5450–5464.

Jardim-Goncalves, Ricardo, David Romero, and Antonio Grilo. 2017. ‘Factories of the Future: Challenges and Leading Innovations in Intelligent Manufacturing’. *International Journal of Computer Integrated Manufacturing* 30 (1): 4–14. https://doi.org/10.1080/0951192X.2016.1258120.

Kagermann, Henning, Johannes Helbig, Ariane Hellinger, and Wolfgang Wahlster. 2013. *Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Securing the Future of German Manufacturing Industry; Final Report of the Industrie 4.0 Working Group*. Forschungsunion.

Kang, Hyoung Seok, Ju Yeon Lee, SangSu Choi, Hyun Kim, Jun Hee Park, Ji Yeon Son, Bo Hyun Kim, and Sang Do Noh. 2016. ‘Smart Manufacturing: Past Research, Present Findings, and Future Directions’. *International Journal of Precision Engineering and Manufacturing-Green Technology* 3 (1): 111–128.

Kaur, Maninder Jeet, Ved P. Mishra, and Piyush Maheshwari. 2020. ‘The Convergence of Digital Twin, IoT, and Machine Learning: Transforming Data into Action’. In *Digital Twin Technologies and Smart Cities*, edited by Maryam Farsi, Alireza Daneshkhah, Amin Hosseinian-Far, and Hamid Jahankhani, 3–17. Internet of Things. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-18732-3_1.

Khorasgani, Hamed, Ahmed Farahat, Kosta Ristovski, Chetan Gupta, and Gautam Biswas. 2018. ‘A Framework for Unifying Model-Based and Data-Driven Fault Diagnosis’. In *PHM Society Conference*. Vol. 10.

Kletti, Jürgen. 2007. *Manufacturing Execution System-MES*. Springer.

Konikow, Leonard F., and John D. Bredehoeft. 1992. ‘Ground-Water Models Cannot Be Validated’. *Advances in Water Resources* 15 (1): 75–83.

Koulamas, Christos, and Athanasios Kalogeras. 2018. ‘Cyber-Physical Systems and Digital Twins in the Industrial Internet of Things [Cyber-Physical Systems]’. *Computer* 51 (11): 95–98.

Kritzinger, Werner, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. 2018. ‘Digital Twin in Manufacturing: A Categorical Literature Review and Classification’. *IFAC-PapersOnLine* 51 (11): 1016–1022.

Kusiak, Andrew. 2018. ‘Smart Manufacturing’. *International Journal of Production Research* 56 (1–2): 508–517.

Küsters, Dennis, Nicolina Pra\s s, and Yves-Simon Gloy. 2017. ‘Textile Learning Factory 4.0–Preparing Germany’s Textile Industry for the Digital Future’. *Procedia Manufacturing* 9: 214–221.

Lämmel, Ralf, and Joost Visser. 2002. ‘Design Patterns for Functional Strategic Programming’. In *Proceedings of the 2002 ACM SIGPLAN Workshop on Rule-Based Programming*, 1–14.

Law, Averill M., W. David Kelton, and W. David Kelton. 2000. *Simulation Modeling and Analysis*. Vol. 3. McGraw-Hill New York.

Lee, Chi G., and Sang C. Park. 2014. ‘Survey on the Virtual Commissioning of Manufacturing Systems’. *Journal of Computational Design and Engineering* 1 (3): 213–222.

Lee, H., and T. Kim. 2018. ‘Smart Factory Use Case Model Based on Digital Twin’. *ICIC Express Letters, Part B: Applications* 9 (9): 931–36. https://doi.org/10.24507/icicelb.09.09.931.

Lee, In, and Kyoochun Lee. 2015. ‘The Internet of Things (IoT): Applications, Investments, and Challenges for Enterprises’. *Business Horizons* 58 (4): 431–440.

Lee, Jay, Behrad Bagheri, and Hung-An Kao. 2015. ‘A Cyber-Physical Systems Architecture for Industry 4.0-Based Manufacturing Systems’. *Manufacturing Letters* 3: 18–23.

Lee, Jay, Hung-An Kao, and Shanhu Yang. 2014. ‘Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment’. *Procedia Cirp* 16: 3–8.

Lee, Jay, Edzel Lapira, Shanhu Yang, and Ann Kao. 2013. 'Predictive Manufacturing System-Trends of next-Generation Production Systems'. *IFAC Proceedings Volumes* 46 (7): 150–156.

Lee, Jay, Fangji Wu, Wenyu Zhao, Masoud Ghaffari, Linxia Liao, and David Siegel. 2014. 'Prognostics and Health Management Design for Rotary Machinery Systems—Reviews, Methodology and Applications'. *Mechanical Systems and Signal Processing* 42 (1–2): 314–334.

Leng, Jiewu, Hao Zhang, Douxi Yan, Qiang Liu, Xin Chen, and Ding Zhang. 2018. 'Digital Twin-Driven Manufacturing Cyber-Physical System for Parallel Controlling of Smart Workshop'. *Journal of Ambient Intelligence and Humanized Computing*, 1–12.

Leonard-Barton, Dorothy. 2000. 'The Factory as a Learning Laboratory'. In *Strategic Learning in a Knowledge Economy*, 91–118. Elsevier.

Li, Congbo, Lingling Li, Ying Tang, Yantao Zhu, and Li Li. 2019. 'A Comprehensive Approach to Parameters Optimization of Energy-Aware CNC Milling'. *Journal of Intelligent Manufacturing* 30 (1): 123–138.

Liao, Linxia, and Felix Köttig. 2016. 'A Hybrid Framework Combining Data-Driven and Model-Based Methods for System Remaining Useful Life Prediction'. *Applied Soft Computing* 44: 191–199.

Liau, Y., H. Lee, and K. Ryu. 2018. 'Digital Twin Concept for Smart Injection Molding'. In . Vol. 324. https://doi.org/10.1088/1757-899X/324/1/012077.

Lin, Ming-Yen, and Suh-Yin Lee. 2003. 'Improving the Efficiency of Interactive Sequential Pattern Mining by Incremental Pattern Discovery'. In *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of The*, 8–pp. IEEE.

Liu, Qiang, Hao Zhang, Jiewu Leng, and Xin Chen. 2018. 'Digital Twin-Driven Rapid Individualised Designing of Automated Flow-Shop Manufacturing System'. *International Journal of Production Research*, 1–17.

Liu, Yinling, Tao Wang, Haiqing Zhang, and Vincent Cheutet. 2019. 'Simulation-Based Fuzzy-Rough Nearest Neighbour Fault Classification and Prediction for Aircraft Maintenance'. *Journal of Simulation*, 1–15.

Liu, Z., N. Meyendorf, and N. Mrad. 2018. 'The Role of Data Fusion in Predictive Maintenance Using Digital Twin'. In . Vol. 1949. https://doi.org/10.1063/1.5031520.

Lu, Yuqian, Chao Liu, I. Kevin, Kai Wang, Huiyue Huang, and Xun Xu. 2020. 'Digital Twin-Driven Smart Manufacturing: Connotation, Reference Model, Applications and Research Issues'. *Robotics and Computer-Integrated Manufacturing* 61: 101837.

Luo, Jianhui, Madhavi Namburu, Krishna Pattipati, Liu Qiao, Masayuki Kawamoto, and SACS Chigusa. 2003. 'Model-Based Prognostic Techniques [Maintenance Applications]'. In *Proceedings AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference.*, 330–340. Ieee.

Luo, Jianhui, Krishna R. Pattipati, Liu Qiao, and Shunsuke Chigusa. 2008. 'Model-Based Prognostic Techniques Applied to a Suspension System'. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 38 (5): 1156–1168.

Luo, W., T. Hu, W. Zhu, and F. Tao. 2018. 'Digital Twin Modeling Method for CNC Machine Tool'. In *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, 1–4. https://doi.org/10.1109/ICNSC.2018.8361285.

Mabkhot, Mohammed, Abdulrahman Al-Ahmari, Bashir Salah, and Hisham Alkhalefah. 2018. 'Requirements of the Smart Factory System: A Survey and Perspective'. *Machines* 6 (2): 23.

Macchi, Marco, Irene Roda, Elisa Negri, and Luca Fumagalli. 2018. 'Exploring the Role of Digital Twin for Asset Lifecycle Management'. *IFAC-PapersOnLine* 51 (11): 790–795.

Madni, Azad M., Carla C. Madni, and Scott D. Lucero. 2019. 'Leveraging Digital Twin Technology in Model-Based Systems Engineering'. *Systems* 7 (1): 7.

McGregor, Ian. 2002. 'Equipment Interface: The Relationship between Simulation and Emulation'. In *Proceedings of the 34th Conference on Winter Simulation: Exploring New Frontiers*, 1683–1688. Winter Simulation Conference.

Mitra, Sushmita, Sankar K. Pal, and Pabitra Mitra. 2002. 'Data Mining in Soft Computing Framework: A Survey'. *IEEE Transactions on Neural Networks* 13 (1): 3–14.

Moeuf, Alexandre, Robert Pellerin, Samir Lamouri, Simon Tamayo-Giraldo, and Rodolphe Barbaray. 2018. 'The Industrial Management of SMEs in the Era of Industry 4.0'. *International Journal of Production Research* 56 (3): 1118–36. https://doi.org/10.1080/00207543.2017.1372647.

Moghaddam, Mohsen, and Shimon Y. Nof. 2017. 'The Collaborative Factory of the Future'. *International Journal of Computer Integrated Manufacturing* 30 (1): 23–43. https://doi.org/10.1080/0951192X.2015.1066034.

Mohammadi, Neda, and John E. Taylor. 2017. 'Smart City Digital Twins'. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–5. IEEE.

Monostori, László, Botond Kádár, T. Bauernhansl, S. Kondoh, S. Kumara, G. Reinhart, O. Sauer, G. Schuh, W. Sihn, and K. Ueda. 2016. 'Cyber-Physical Systems in Manufacturing'. *Cirp Annals* 65 (2): 621–641.

Moran, Michael J., Howard N. Shapiro, and Michael J. Moran. 2008. 'Table A-6 Properties of Saturated Water (Solid-Vapor): Temperature Table'. *Appendices to Accompany Fundamentals of Engineering Thermodynamics. Hoboken, NJ: Wiley* 826.

Mowbray, Thomas J., William A. Ruh, and Richard M. Soley. 1997. *Inside CORBA: Distributed Object Standards and Applications*. Vol. 130. Addison-Wesley Reading, Massachusetts.

Nardello, Marco, Ole Madsen, and Charles Møller. 2017. 'The Smart Production Laboratory: A Learning Factory for Industry 4.0 Concepts'. In *CEUR Workshop Proceedings*, 1898:5.

Negri, Elisa, Luca Fumagalli, Chiara Cimino, and Marco Macchi. 2019. 'FMU-Supported Simulation for CPS Digital Twin'. *Procedia Manufacturing*, 7th International conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV2018), 28 (January): 201–6. https://doi.org/10.1016/j.promfg.2018.12.033.

Negri, Elisa, Luca Fumagalli, and Marco Macchi. 2017. 'A Review of the Roles of Digital Twin in CPS-Based Production Systems'. *Procedia Manufacturing*, 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy, 11 (January): 939–48. https://doi.org/10.1016/j.promfg.2017.07.198.

Ngai, Eric WT, Li Xiu, and Dorothy CK Chau. 2009. 'Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification'. *Expert Systems with Applications* 36 (2): 2592–2602.

Nikolakis, Nikolaos, Kosmas Alexopoulos, Evangelos Xanthakis, and George Chryssolouris. 2018. 'The Digital Twin Implementation for Linking the Virtual Representation of Human-Based Production Tasks to Their Physical Counterpart in the Factory-Floor'. *International Journal of Computer Integrated Manufacturing*, 1–12.

Nonaka, Ikujiro. 1994. 'A Dynamic Theory of Organizational Knowledge Creation'. *Organization Science* 5 (1): 14–37.

Odom, Marcus D., and Ramesh Sharda. 1990. 'A Neural Network Model for Bankruptcy Prediction'. In *1990 IJCNN International Joint Conference on Neural Networks*, 163–168. IEEE.

Ogorodnyk, Olga, Malin Victoria Granheim, and Halvor Holtskog. 2016. 'Preconditions for Learning Factory A Case Study'. *Procedia CIRP* 54: 35–40.

Okoh, Caxton, Rajkumar Roy, Jorn Mehnen, and L. Redding. 2014. 'Overview of Remaining Useful Life Prediction Techniques in Through-Life Engineering Services'. *Procedia CIRP* 16: 158–163.

Osinga, Frans PB. 2007. *Science, Strategy and War: The Strategic Theory of John Boyd*. Routledge.

Osterrieder, Philipp, Lukas Budde, and Thomas Friedli. 2019. 'The Smart Factory as a Key Construct of Industry 4.0: A Systematic Literature Review'. *International Journal of Production Economics*, August. https://doi.org/10.1016/j.ijpe.2019.08.011.

Otterlo, Martijn van, and Marco Wiering. 2012. 'Reinforcement Learning and Markov Decision Processes'. In *Reinforcement Learning*, 3–42. Springer.

Owens, J. W. 1997. 'Life-Cycle Assessment: Constraints on Moving from Inventory to Impact Assessment'. *Journal of Industrial Ecology* 1 (1): 37–49.

Padovano, Antonio, Francesco Longo, Letizia Nicoletti, and Giovanni Mirabelli. 2018. 'A Digital Twin Based Service Oriented Application for a 4.0 Knowledge Navigation in the Smart Factory'. *IFAC-PapersOnLine* 51 (11): 631–636.

Panetto, Hervé, Milan Zdravkovic, Ricardo Jardim-Goncalves, David Romero, Joe Cecil, and István Mezgár. 2016. 'New Perspectives for the Future Interoperable Enterprise Systems'. *Computers in Industry* 79: 47–63.

Park, Kyu Tae, Young Wook Nam, Hyeon Seung Lee, Sung Ju Im, Sang Do Noh, Ji Yeon Son, and Hyun Kim. 2019. 'Design and Implementation of a Digital Twin Application for a Connected Micro Smart Factory'. *International Journal of Computer Integrated Manufacturing* 32 (6): 596–614. https://doi.org/10.1080/0951192X.2019.1599439.

Paśko, \Lukasz, and Pawe\l Litwin. 2019. 'Methods of Data Mining for Quality Assurance in Glassworks'. In *Working Conference on Virtual Enterprises*, 185–192. Springer.

Pessl, Ernst, Sabrina Romina Sorko, and Barbara Mayer. 2017. 'Roadmap Industry 4.0–Implementation Guideline for Enterprises'. *International Journal of Science, Technology and Society* 5 (6): 193–202.

Poelmans, Jonas, Paul Elzinga, Stijn Viaene, and Guido Dedene. 2010. 'Formal Concept Analysis in Knowledge Discovery: A Survey'. In *International Conference on Conceptual Structures*, 139–153. Springer.

Ponomarev, K., N. Kudryashov, N. Popelnukha, and V. Potekhin. 2017. 'Main Principals and Issues of Digital Twin Development for Complex Technological Processes'. In , 523–28. https://doi.org/10.2507/28th.daaam.proceedings.074.

Posada, Jorge, Carlos Toro, Iñigo Barandiaran, David Oyarzun, Didier Stricker, Raffaele de Amicis, Eduardo B. Pinto, Peter Eisert, Jürgen Döllner, and Ivan Vallarino. 2015. 'Visual Computing as a Key Enabling Technology for Industrie 4.0 and Industrial Internet'. *IEEE Computer Graphics and Applications* 35 (2): 26–40.

Post, Jan, Manso Groen, and Gerrit Klaseboer. 2009. 'PHYSICAL MODEL BASED DIGITAL TWINS IN MANUFACTURING PROCESSES'. *Optics Letters* 34 (13): 1982–4.

Putnik, G., A. Sluga, H. ElMaraghy, R. Teti, Y. Koren, T. Tolio, and B. Hon. 2013. 'Scalability in Manufacturing Systems Design and Operation: State-of-the-Art and Future Developments Roadmap'. *CIRP Annals* 62 (2): 751–774.

Qi, Qinglin, and Fei Tao. 2018. 'Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison'. *IEEE Access* 6: 3585–3593.

Qi, Qinglin, Fei Tao, Ying Zuo, and Dongming Zhao. 2018. 'Digital Twin Service towards Smart Manufacturing'. *Procedia CIRP*, 51st CIRP Conference on Manufacturing Systems, 72 (January): 237–42. https://doi.org/10.1016/j.procir.2018.03.103.

Rajkumar, Ragunathan, Insup Lee, Lui Sha, and John Stankovic. 2010. 'Cyber-Physical Systems: The next Computing Revolution'. In *Design Automation Conference*, 731–736. IEEE.

Ramirez, Andres J., and Betty HC Cheng. 2010. 'Design Patterns for Developing Dynamically Adaptive Systems'. In *Proceedings of the 2010 ICSE Workshop on Software Engineering for Adaptive and Self-Managing Systems*, 49–58.

Raymond, Louis. 2005. 'Operations Management and Advanced Manufacturing Technologies in SMEs'. *Journal of Manufacturing Technology Management*.

Reap, John, Felipe Roman, Scott Duncan, and Bert Bras. 2008. 'A Survey of Unresolved Problems in Life Cycle Assessment'. *The International Journal of Life Cycle Assessment* 13 (5): 374.

Redelinghuys, A., A. Basson, and K. Kruger. 2019. 'A Six-Layer Digital Twin Architecture for a Manufacturing Cell'. *Studies in Computational Intelligence* 803: 412–23. https://doi.org/10.1007/978-3-030-03003-2_32.

Rentzos, L., M. Doukas, D. Mavrikios, D. Mourtzis, and G. Chryssolouris. 2014. 'Integrating Manufacturing Education with Industrial Practice Using Teaching Factory Paradigm: A Construction Equipment Application'. *Procedia CiRP* 17: 189–194.

Ríos, José, Juan Carlos Hernández, Manuel Oliva, and Fernando Mas. 2015. 'Product Avatar as Digital Counterpart of a Physical Individual Product: Literature Review and Implications in an Aircraft.' In *ISPE CE*, 657–666.

Rodič, B. 2017. 'Industry 4.0 and the New Simulation Modelling Paradigm'. *Organizacija* 50 (3): 193–207. https://doi.org/10.1515/orga-2017-0017.

Romero, David, Peter Bernus, Ovidiu Noran, Johan Stahre, and \AAsa Fast-Berglund. 2016. 'The Operator 4.0: Human Cyber-Physical Systems & Adaptive Automation towards Human-Automation Symbiosis Work Systems'. In *IFIP International Conference on Advances in Production Management Systems*, 677–686. Springer.

Roschelle, Jeremy, and Stephanie D. Teasley. 1995. 'The Construction of Shared Knowledge in Collaborative Problem Solving'. In *Computer Supported Collaborative Learning*, 69–97. Springer.

Rosen, Roland, Georg von Wichert, George Lo, and Kurt D. Bettenhausen. 2015. 'About the Importance of Autonomy and Digital Twins for the Future of Manufacturing'. *IFAC-PapersOnLine* 48 (3): 567–572.

Rouane-Hacene, Mohamed, Marianne Huchard, Amedeo Napoli, and Petko Valtchev. 2013. 'Relational Concept Analysis: Mining Concept Lattices from Multi-Relational Data'. *Annals of Mathematics and Artificial Intelligence* 67 (1): 81–108.

Rowley, Jennifer. 2007. 'The Wisdom Hierarchy: Representations of the DIKW Hierarchy'. *Journal of Information Science* 33 (2): 163–180.

Ruppert, Tamás, Szilárd Jaskó, Tibor Holczinger, and János Abonyi. 2018. 'Enabling Technologies for Operator 4.0: A Survey'. *Applied Sciences* 8 (9): 1650.

Rüßmann, Michael, Markus Lorenz, Philipp Gerbert, Manuela Waldner, Jan Justus, Pascal Engel, and Michael Harnisch. 2015. 'Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries'. *Boston Consulting Group* 9 (1): 54–89.

Schleich, Benjamin, Nabil ANWER, Luc Mathieu, and Sandro Wartzack. 2017. 'Shaping the Digital Twin for Design and Production Engineering'. *CIRP Annals - Manufacturing Technology* 66 (1). https://doi.org/10.1016/j.cirp.2017.04.040.

Schluse, Michael, Linus Atorf, and Juergen Rossmann. 2017. 'Experimentable Digital Twins for Model-Based Systems Engineering and Simulation-Based Development'. In *Systems Conference (SysCon), 2017 Annual IEEE International*, 1–8. IEEE.

Schluse, Michael, Marc Priggemeyer, Linus Atorf, and Juergen Rossmann. 2018. 'Experimentable Digital Twins—Streamlining Simulation-Based Systems Engineering for Industry 4.0'. *IEEE Transactions on Industrial Informatics* 14 (4): 1722–1731.

Schmidt, J. William. 1970. *Simulation and analysis of industrial systems*. Homewood (Ill.), Etats-Unis d'Amérique: R. D. Irwin.

Schroeder, Greyce N., Charles Steinmetz, Carlos E. Pereira, and Danubia B. Espindola. 2016. 'Digital Twin Data Modeling with AutomationML and a Communication Methodology for Data Exchange'. *IFAC-PapersOnLine* 49 (30): 12–17.

Schroeder, Greyce, Charles Steinmetz, Carlos Eduardo Pereira, Ivan Muller, Natanael Garcia, Danubia Espindola, and Ricardo Rodrigues. 2016. 'Visualising the Digital Twin Using Web Services and Augmented Reality'. In *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)*, 522–527. IEEE.

Schuh, Günther, Thomas Gartzen, Timon Rodenhauser, and Alexander Marks. 2015. 'Promoting Work-Based Learning through Industry 4.0'. *Procedia CIRP* 32: 82–87.

Seitz, Kai-Frederic, and Peter Nyhuis. 2015. 'Cyber-Physical Production Systems Combined with Logistic Models–a Learning Factory Concept for an Improved Production Planning and Control'. *Procedia CIRP* 32: 92–97.

Semeraro, Concetta, Mario Lezoche, Hervé Panetto, Michele Dassisti, and Stefano Cafagna. 2019. 'Data-Driven Pattern-Based Constructs Definition for the Digital Transformation Modelling of Collaborative Networked Manufacturing Enterprises'. In *Working Conference on Virtual Enterprises*, 507–515. Springer.

Semeraro, Concetta, Hervé Panetto, Mario Lezoche, Michele Dassisti, and Stefano Cafagna. 2019. 'A Monitoring Strategy for Industry 4.0: Master Italy Srl Case Study'. *INSIGHT-International Council on Systems Engineering (INCOSE)* 22 (4): 20–22.

Sheibat-Othman, Nida, Nassim Laouti, Jean-Pierre Valour, and Sami Othman. 2014. 'Support Vector Machines Combined to Observers for Fault Diagnosis in Chemical Reactors'. *The Canadian Journal of Chemical Engineering* 92 (4): 685–695.

Silva, Vander Luiz Da, João Luiz Kovaleski, Regina Negri Pagani, Jaqueline De Matos Silva, and Alana Corsi. 2019. 'Implementation of Industry 4.0 Concept in Companies: Empirical Evidences'. *International Journal of Computer Integrated Manufacturing* 0 (0): 1–18. https://doi.org/10.1080/0951192X.2019.1699258.

Simons, Stephan, Patrick Abé, and Stephan Neser. 2017. 'Learning in the AutFab–the Fully Automated Industrie 4.0 Learning Factory of the University of Applied Sciences Darmstadt'. *Procedia Manufacturing* 9: 81–88.

Sjödin, David R., Vinit Parida, Markus Leksell, and Aleksandar Petrovic. 2018. 'Smart Factory Implementation and Process Innovation'. *Research-Technology Management* 61 (5): 22–31. https://doi.org/10.1080/08956308.2018.1471277.

Škopljanac-Mačina, Frano, and Bruno Blašković. 2014. 'Formal Concept Analysis–Overview and Applications'. *Procedia Engineering* 69: 1258–1267.

Söderberg, Rikard, Kristina Wärmefjord, Johan S. Carlson, and Lars Lindkvist. 2017. 'Toward a Digital Twin for Real-Time Geometry Assurance in Individualized Production'. *CIRP Annals* 66 (1): 137–140.

Solomatine, Dimitri P., and Avi Ostfeld. 2008. 'Data-Driven Modelling: Some Past Experiences and New Approaches'. *Journal of Hydroinformatics* 10 (1): 3–22. https://doi.org/10.2166/hydro.2008.015.

Sony, Michael, and Subhash Naik. 2019. 'Critical Factors for the Successful Implementation of Industry 4.0: A Review and Future Research Direction'. *Production Planning & Control* 0 (0): 1–17. https://doi.org/10.1080/09537287.2019.1691278.

Stark, Rainer, Simon Kind, and Sebastian Neumeyer. 2017. 'Innovations in Digital Modelling for next Generation Manufacturing System Design'. *CIRP Annals* 66 (1): 169–172.

Strozzi, Fernanda, Claudia Colicchia, Alessandro Creazza, and Carlo Noè. 2017. 'Literature Review on the "Smart Factory" Concept Using Bibliometric Tools'. *International Journal of Production Research* 55 (22): 6572–91. https://doi.org/10.1080/00207543.2017.1326643.

Sun, Heqing, Cheng Li, Xinyu Fang, and Hao Gu. 2017. 'Optimized Throughput Improvement of Assembly Flow Line with Digital Twin Online Analytics'. In *Robotics and Biomimetics (ROBIO), 2017 IEEE International Conference On*, 1833–1837. IEEE.

Sutharssan, Thamo, Stoyan Stoyanov, Chris Bailey, and Chunyan Yin. 2015. 'Prognostic and Health Management for Engineering Systems: A Review of the Data-Driven Approach and Algorithms'. *The Journal of Engineering* 2015 (7): 215–222.

Talkhestani, B.A., N. Jazdi, W. Schlögl, and M. Weyrich. 2018. 'A Concept in Synchronization of Virtual Production System with Real Factory Based on Anchor-Point Method'. In , 67:13–17. https://doi.org/10.1016/j.procir.2017.12.168.

Tan, Lu, and Neng Wang. 2010. 'Future Internet: The Internet of Things'. In *2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, 5:V5–376. IEEE.

Tao, Fei, Jiangfeng Cheng, Qinglin Qi, Meng Zhang, He Zhang, and Fangyuan Sui. 2018. 'Digital Twin-Driven Product Design, Manufacturing and Service with Big Data'. *The International Journal of Advanced Manufacturing Technology* 94 (9–12): 3563–3576.

Tao, Fei, Qinglin Qi, Lihui Wang, and A. Y. C. Nee. 2019. 'Digital Twins and Cyber-Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison'. *Engineering* 5 (4): 653–661.

Tao, Fei, Fangyuan Sui, Ang Liu, Qinglin Qi, Meng Zhang, Boyang Song, Zirong Guo, Stephen C.-Y. Lu, and A. Y. C. Nee. 2018. 'Digital Twin-Driven Product Design Framework'. *International Journal of Production Research*, 1–19.

Tao, Fei, and Meng Zhang. 2017. 'Digital Twin Shop-Floor: A New Shop-Floor Paradigm towards Smart Manufacturing'. *Ieee Access* 5: 20418–20427.

Tao, Fei, Meng Zhang, Yushan Liu, and A. Y. C. Nee. 2018. 'Digital Twin Driven Prognostics and Health Management for Complex Equipment'. *CIRP Annals*.

Terziyan, Vagan, Svitlana Gryshko, and Mariia Golovianko. 2018. 'Patented Intelligence: Cloning Human Decision Models for Industry 4.0'. *Journal of Manufacturing Systems*.

Thiede, Sebastian, Max Juraschek, and Christoph Herrmann. 2016. 'Implementing Cyber-Physical Production Systems in Learning Factories'. *Procedia CIRP* 54: 7–12.

Tidriri, Khaoula, Nizar Chatti, Sylvain Verron, and Téodor Tiplica. 2016. 'Bridging Data-Driven and Model-Based Approaches for Process Fault Diagnosis and Health Monitoring: A Review of Researches and Future Challenges'. *Annual Reviews in Control* 42: 63–81.

Tisch, M., and J. Metternich. 2017. 'Potentials and Limits of Learning Factories in Research, Innovation Transfer, Education, and Training'. *Procedia Manufacturing* 9: 89–96.

Tobon-Mejia, Diego Alejandro, Kamal Medjaher, Noureddine Zerhouni, and Gerard Tripot. 2012. 'A Data-Driven Failure Prognostics Method Based on Mixture of Gaussians Hidden Markov Models'. *IEEE Transactions on Reliability* 61 (2): 491–503.

Tsai, Wei-Tek, Xiaoying Bai, Ray Paul, Weiguang Shao, and Vishal Agarwal. 2001. 'End-to-End Integration Testing Design'. In *25th Annual International Computer Software and Applications Conference. COMPSAC 2001*, 166–171. IEEE.

Tuegel, Eric. 2012. 'The Airframe Digital Twin: Some Challenges to Realization'. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA*, 1812.

Tuegel, Eric J., Anthony R. Ingraffea, Thomas G. Eason, and S. Michael Spottswood. 2011. ‘Reengineering Aircraft Structural Life Prediction Using a Digital Twin’. *International Journal of Aerospace Engineering* 2011.

Uhlemann, Thomas H.-J., Christian Lehmann, and Rolf Steinhilper. 2017. ‘The Digital Twin: Realizing the Cyber-Physical Production System for Industry 4.0’. *Procedia Cirp* 61: 335–340.

Um, Jumyung, Stephan Weyer, and Fabian Quint. 2017. ‘Plug-and-Simulate within Modular Assembly Line Enabled by Digital Twins and the Use of AutomationML’. *IFAC-PapersOnLine* 50 (1): 15904–15909.

UML. 2005. ‘About the Unified Modeling Language Specification Version 2.0’. 2005. https://www.omg.org/spec/UML/2.0/.

UNI EN ISO 14040. 2006. ‘UNI EN ISO 14040:2006’. 2006.

UPDM. 2017. ‘About the Unified Profile for DoDAF and MODAF Specification Version 2.1.1’. 2017. https://www.omg.org/spec/UPDM/.

Vachálek, Ján, Lukás Bartalskỳ, Oliver Rovnỳ, Dana Šišmišová, Martin Morháč, and Milan Lokšík. 2017. ‘The Digital Twin of an Industrial Production Line within the Industry 4.0 Concept’. In *Process Control (PC), 2017 21st International Conference On*, 258–262. IEEE.

Valtchev, Petko, Rokia Missaoui, and Robert Godin. 2004. ‘Formal Concept Analysis for Knowledge Discovery and Data Mining: The New Challenges’. In *International Conference on Formal Concept Analysis*, 352–371. Springer.

Venkatasubramanian, Venkat, Raghunathan Rengaswamy, and Surya N. Kavuri. 2003. ‘A Review of Process Fault Detection and Diagnosis: Part II: Qualitative Models and Search Strategies’. *Computers & Chemical Engineering* 27 (3): 313–326.

Venkatasubramanian, Venkat, Raghunathan Rengaswamy, Surya N. Kavuri, and Kewen Yin. 2003. ‘A Review of Process Fault Detection and Diagnosis: Part III: Process History Based Methods’. *Computers & Chemical Engineering* 27 (3): 327–346.

Venkatasubramanian, Venkat, Raghunathan Rengaswamy, Kewen Yin, and Surya N. Kavuri. 2003. ‘A Review of Process Fault Detection and Diagnosis: Part I: Quantitative Model-Based Methods’. *Computers & Chemical Engineering* 27 (3): 293–311.

Venter, F. J., G. D. Oosthuizen, and J. D. Roos. 1997. ‘Knowledge Discovery in Databases Using Lattices’. *Expert Systems with Applications*, Selected Papers from the PACES/SPICIS’97 Conference, 13 (4): 259–64. https://doi.org/10.1016/S0957-4174(97)00047-X.

Wagner, Ulf, Tarek AlGeddawy, Hoda ElMaraghy, and E. MŸller. 2012. ‘The State-of-the-Art and Prospects of Learning Factories’. *Procedia CiRP* 3: 109–114.

Wahba, Sally K., Jason O. Hallstrom, and Neelam Soundarajan. 2010. ‘Initiating a Design Pattern Catalog for Embedded Network Systems’. In *Proceedings of the Tenth ACM International Conference on Embedded Software*, 249–258.

Wang, Jinjiang, Lunkuan Ye, Robert X. Gao, Chen Li, and Laibin Zhang. 2018. ‘Digital Twin for Rotating Machinery Fault Diagnosis in Smart Manufacturing’. *International Journal of Production Research* 0 (0): 1–15. https://doi.org/10.1080/00207543.2018.1552032.

Wang, Lidong, and Guanghui Wang. 2016. ‘Big Data in Cyber-Physical Systems, Digital Manufacturing and Industry 4.0’. *International Journal of Engineering and Manufacturing (IJEM)* 6 (4): 1–8.

Wang, Xi Vincent, and Lihui Wang. 2018. ‘Digital Twin-Based WEEE Recycling, Recovery and Remanufacturing in the Background of Industry 4.0’. *International Journal of Production Research*, 1–11.

Wärmefjord, Kristina, Rikard Söderberg, Lars Lindkvist, Björn Lindau, and Johan S. Carlson. 2017. ‘Inspection Data to Support a Digital Twin for Geometry Assurance’. In *ASME 2017 International Mechanical Engineering Congress and Exposition*, V002T02A101–V002T02A101. American Society of Mechanical Engineers.

Weber, Christian, Jan Königsberger, Laura Kassner, and Bernhard Mitschang. 2017. ‘M2DDM–a Maturity Model for Data-Driven Manufacturing’. *Procedia CIRP* 63: 173–178.

Weilkiens, Tim. 2011. *Systems Engineering with SysML/UML: Modeling, Analysis, Design*. Elsevier.

Weking, Jörg, Maria Stöcker, Marek Kowalkiewicz, Markus Böhm, and Helmut Krcmar. 2018. ‘Archetypes for Industry 4.0 Business Model Innovations’.

Weyer, Stephan, Torben Meyer, Moritz Ohmer, Dominic Gorecky, and Detlef Zühlke. 2016. ‘Future Modeling and Simulation of CPS-Based Factories: An Example from the Automotive Industry’. *IFAC-PapersOnLine* 49 (31): 97–102.

Wille, Rudolf. 2002. ‘Why Can Concept Lattices Support Knowledge Discovery in Databases?’ *Journal of Experimental & Theoretical Artificial Intelligence* 14 (2–3): 81–92. https://doi.org/10.1080/09528130210164161.

Williams, Graham J., and Simeon J. Simoff. 2006. *Data Mining: Theory, Methodology, Techniques, and Applications*. Vol. 3755. Springer.

Xie, Jiacheng, Xuewen Wang, Zhaojian Yang, and Shangqing Hao. 2019. ‘Virtual Monitoring Method for Hydraulic Supports Based on Digital Twin Theory’. *Mining Technology* 0 (0): 1–11. https://doi.org/10.1080/25726668.2019.1569367.

Xu, Li Da, Eric L. Xu, and Ling Li. 2018. ‘Industry 4.0: State of the Art and Future Trends’. *International Journal of Production Research* 56 (8): 2941–62. https://doi.org/10.1080/00207543.2018.1444806.

Xu, Rui, and Donald C. Wunsch. 2005. ‘Survey of Clustering Algorithms’.

Xu, Yan, Yanming Sun, Xiaolong Liu, and Yonghua Zheng. 2019. ‘A Digital-Twin-Assisted Fault Diagnosis Using Deep Transfer Learning’. *IEEE Access* 7: 19990–19999.

Yao, Xifan, Hong Jin, and Jie Zhang. 2015. 'Towards a Wisdom Manufacturing Vision'. *International Journal of Computer Integrated Manufacturing* 28 (12): 1291–1312. https://doi.org/10.1080/0951192X.2014.972462.

Yao, Xifan, Jiajun Zhou, Yingzi Lin, Yun Li, Hongnian Yu, and Ying Liu. 2019. 'Smart Manufacturing Based on Cyber-Physical Systems and Beyond'. *Journal of Intelligent Manufacturing* 30 (8): 2805–2817.

Yin, Yong, Ikou Kaku, Jiafu Tang, and JianMing Zhu. 2011. *Data Mining: Concepts, Methods and Applications in Management and Engineering Design*. Springer Science & Business Media.

Yin, Yong, Kathryn E. Stecke, and Dongni Li. 2018. 'The Evolution of Production Systems from Industry 2.0 through Industry 4.0'. *International Journal of Production Research* 56 (1–2): 848–861.

Yun, Seongjin, Jun-Hong Park, and Won-Tae Kim. 2017. 'Data-Centric Middleware Based Digital Twin Platform for Dependable Cyber-Physical Systems'. In *Ubiquitous and Future Networks (ICUFN), 2017 Ninth International Conference On*, 922–926. IEEE.

Zhang, Guoqiang Peter. 2000. 'Neural Networks for Classification: A Survey'. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 30 (4): 451–462.

Zhang, H., Q. Liu, X. Chen, D. Zhang, and J. Leng. 2017. 'A Digital Twin-Based Approach for Designing and Multi-Objective Optimization of Hollow Glass Production Line'. *IEEE Access* 5: 26901–11. https://doi.org/10.1109/ACCESS.2017.2766453.

Zhang, H., G. Zhang, and Q. Yan. 2018. 'Dynamic Resource Allocation Optimization for Digital Twin-Driven Smart Shopfloor'. In *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, 1–5. https://doi.org/10.1109/ICNSC.2018.8361283.

Zhang, Huiguo, Rui Kang, and Michael Pecht. 2009. 'A Hybrid Prognostics and Health Management Approach for Condition-Based Maintenance'. In *2009 IEEE International Conference on Industrial Engineering and Engineering Management*, 1165–1169. IEEE.

Zhang, Junlin, Samuel Oluwarotimi Williams, and Haoxiang Wang. 2018. 'Intelligent Computing System Based on Pattern Recognition and Data Mining Algorithms'. *Sustainable Computing: Informatics and Systems* 20 (December): 192–202. https://doi.org/10.1016/j.suscom.2017.10.010.

Zhang, M., Y. Zuo, and F. Tao. 2018. 'Equipment Energy Consumption Management in Digital Twin Shop-Floor: A Framework and Potential Applications'. In *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, 1–5. https://doi.org/10.1109/ICNSC.2018.8361272.

Zhang, Xi, and Karlene A. Hoo. 2011. 'Effective Fault Detection and Isolation Using Bond Graph-Based Domain Decomposition'. *Computers & Chemical Engineering* 35 (1): 132–148.

Zhang, Yingfeng, Shan Ren, Yang Liu, Tomohiko Sakao, and Donald Huisingh. 2017. 'A Framework for Big Data Driven Product Lifecycle Management'. *Journal of Cleaner Production* 159: 229–240.

Zhao, Yang, Chaobo Zhang, Yiwen Zhang, Zihao Wang, and Junyang Li. 2019. 'A Review of Data Mining Technologies in Building Energy Systems: Load Prediction, Pattern Identification, Fault Detection and Diagnosis'. *Energy and Built Environment*, November. https://doi.org/10.1016/j.enbenv.2019.11.003.

Zhu, Qian, Ruicong Wang, Qi Chen, Yan Liu, and Weijun Qin. 2010. 'Iot Gateway: Bridgingwireless Sensor Networks into Internet of Things'. In *2010 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*, 347–352. Ieee.

Zhuang, Cunbo, Jianhua Liu, and Hui Xiong. 2018. 'Digital Twin-Based Smart Production Management and Control Framework for the Complex Product Assembly Shop-Floor'. *The International Journal of Advanced Manufacturing Technology* 96 (1–4): 1149–1163.

Zuehlke, Detlef. 2010. 'SmartFactory—Towards a Factory-of-Things'. *Annual Reviews in Control* 34 (1): 129–138.

RÉSUMÉ

Titre : Contribution à la Formalisation d'invariants de Modélisation de Systèmes Cyber-Physiques, dirigés par les Données

Abstract: La transformation numérique des entreprises manufacturières collaboratives en réseau nécessite la construction et l'application de modèles numériques représentant l'ensemble des ressources et des connaissances sur les processus. La modélisation d'une telle copie numérique du système physique pour effectuer une validation et une optimisation en temps réel est assez complexe et nécessite donc une grande quantité de données et quelques modèles de modélisation représentant la sémantique opérationnelle des éléments modélisés. En règle générale, l'action de modélisation a un type d'application localisé et spécifique. Fort de ce constat, le principal défi de la modélisation de la transformation numérique est de créer une approche invariante, à savoir un modèle décomposable et recomposable vers différentes applications. La thèse vise à identifier et formaliser des éléments de modélisation contribuant à construire des modèles informationnels et fonctionnels pour améliorer la durabilité des processus de fabrication et des produits, basés sur des composants en réseau. Ces éléments formels permettront alors de représenter la connaissance et sa relation profonde avec les processus de fabrication. Ils rendent ainsi les connaissances partagées plus facilement réutilisables et sont à la base des efforts de normalisation.

Mots-clefs : Modèles technologiques, usine intelligente

Le paradigme Smart Factory représente la « quatrième révolution industrielle » dans le domaine de l'industrie manufacturière qui peut être synthétisée dans des réseaux intégrant des composants physiques et des logiciels de contrôle et l'amélioration des procédés de fabrication. Les systèmes intelligents comprennent généralement différents composants, notamment des capteurs pour l'acquisition de signaux, des unités de communication pour la transmission de données entre les composants, des unités de contrôle et de gestion pour la prise de décision et des actionneurs pour effectuer les actions appropriées. Ces dernières années, l'émergence des systèmes cyber-physiques
(CPS) a amplifié la capacité de détecter le monde à travers un réseau d'appareils connectés utilisant l'infrastructure de réseau existante. Le regroupement des systèmes intelligents et des systèmes de détection formant un système cyber-physique distribué à grande échelle présente un énorme potentiel pour amener des systèmes intelligents dans de nombreux domaines d'application. Cependant, ils souffrent d'un manque de techniques de modélisation prenant en compte non seulement leurs paramètres technologiques mais aussi

leur fort degré d'information et d'intercorrélations fonctionnelles. Au fur et à mesure que la complexité de ces systèmes continue de croître, le défi du développement de systèmes intelligents et de détection intégrés a dépassé la complexité de conception de leurs composants individuels. Le problème principal du développement de systèmes intelligents et de détection réside dans la complexité d'intégrer et de gérer ces différents composants, technologies et objectifs à travers un large spectre. Il est alors nécessaire de formaliser les connaissances partagées pour définir une méthode de modélisation qui aide à analyser une nouvelle forme de systèmes intelligents (smart) et de détection dans une perspective durable. La représentation du savoir partagé est une branche de l'intelligence artificielle qui étudie la manière dont le raisonnement humain se produit et définit des symboles ou des langues. Cette représentation permet la formalisation de la connaissance pour la rendre compréhensible aux machines, alignées sur des modèles de référence. Dans ce contexte, le travail de thèse vise à identifier et formaliser des éléments de modélisation contribuant à construire des modèles informationnels et fonctionnels pour améliorer la durabilité des processus de fabrication et des produits, basés sur des composants en réseau. Premièrement ces éléments formels, concrétisés à travers la modélisation des systèmes et des procédures en langage SYSML, permettent de représenter la connaissance et sa relation profonde avec les processus de fabrication. Ils rendent ainsi les connaissances partagées plus facilement réutilisables et sont à la base des efforts de normalisation. L'utilisation aussi des techniques de Multi Relational Data Mining, dans le cas spécifique de Relational Concept Analysis, ont permis l'extraction de la connaissance tacite incluse dans les données issues des processus analysés. La thèse propose une série de patrons de modélisation pour la transformation numérique des entreprises de production industrielle. Ce cadre d'évaluation a été expérimenté sur une étude de cas réelle impliquant une véritable entreprise basée en Italie et en réseau du Politecnico de Bari et analysée de manière critique.
Un prototype d'analyse des processus industriels sur une ligne de production comme modèle d'extraction de connaissance a été développé. L'outil résultant peut exploiter les connaissances sur existantes et les informations provenant de systèmes évalués pour identifier les problèmes et proposer des améliorations potentielles.

La thèse définit les questions de recherche suivantes :

RQ1 : Quels sont les travaux de recherche existants sur le jumelage numérique : "Quelles technologies doivent être adoptées ou quelles nouvelles technologies doivent être explorées", "Comment concevoir un digital twin", "Comment mettre en œuvre un digital twin" ?

L'une des premières étapes de notre recherche consiste à étudier et à définir le paradigme du digital twin dans le contexte de l'industrie du futur. Nous avons effectué une analyse documentaire systématique pour identifier le contexte, les applications, les fonctions dans le cycle de vie du produit, les architectures possibles et les composants d'un DT. Ensuite, nous avons l'intention de combiner les différents aspects d'un Digital Twin et certaines questions connexes pour développer notre proposition de paradigme de jumeau numérique. Par conséquent, nous devons relier les différentes approches liées à la modélisation du jumeau numérique pour développer une approche commune afin de détecter les constructions de modélisation des invariants guidés par les données.

RQ2 : Comment développer des modèles invariants basés sur des données ?

Pour répondre à cette question, nous avons effectué une deuxième analyse documentaire sur les approches basées sur des modèles, guidées par les données et hybrides afin de définir les positions de contribution. L'idée est d'identifier et de formaliser la modélisation invariante basée sur l'analyse des données. L'idée est d'utiliser, et surtout de réutiliser, des constructions prédéfinies basées sur des données pour construire des modèles numériques pour différentes applications.

Le chapitre 1 donne un aperçu du contexte de la recherche. Tout d'abord, nous explorons les définitions de base liées au concept d'usine intelligente, les différentes technologies de l'industrie du futur et leur application. Ensuite, nous étudions l'état de l'art de la conception et des applications du digital twin. Enfin, nous identifions les lacunes de la recherche afin de démontrer pourquoi il est nécessaire de développer une approche invariante basée sur des constructions de modèles basés sur des données.

Dans le chapitre 2, l'analyse comparative vise à montrer l'évolution des approches basées sur les modèles et les données au fil des ans pour présenter le positionnement de la contribution de la thèse de doctorat. L'idée de la thèse est de détecter et de formaliser les constructions de modèles invariants basés sur des données.

Le chapitre 3 définit l'approche développée pour extraire des constructions de modélisation invariantes basées sur l'analyse de données.

Le chapitre 4 a pour objectif de prouver la qualité de la contribution à travers une étude de cas réel. De plus, un prototype de Digital Twin a été développé pour démontrer et appliquer nos constructions de modélisation invariante basée sur des données sur un processus de fabrication spécifique pour construire un jumeau numérique.

Enfin, une conclusion présente les résultats de la recherche et quelques perspectives de recherche dérivées de ce travail.

Dans le chapitre 1, nous exposons les principaux concepts liés au Digital Twin. Le présent chapitre concerne la manière de concevoir efficacement un jumeau numérique pour soutenir la transformation numérique des domaines des entreprises manufacturières. Tout d'abord, un mémoire général pour comprendre la transformation numérique dans les PME, est présenté dans la section 1.1. Dans la section 1.2, une analyse documentaire systématique sur le Digital Twin est réalisée pour définir le contexte et l'application en 1.2.1, les mises en œuvre tout au long du cycle de vie du produit en 1.2.2, les fonctions en 1.2.3, les architectures en 1.2.4 et les composants en 1.2.5. Par conséquent, un aperçu des lacunes en matière de recherche est présenté dans la section 1.3 pour définir les objectifs de la thèse en 1.4.

Selon l'analyse effectuée, une généralisation du Digital Twin peut être envisagée. On peut penser qu'un DT générique est constitué de plusieurs composants organisés en trois couches principales reconnues ci-dessus:

1. La couche physique, constituée d'entités identifiées en fonction de l'étape du cycle de vie du produit.
2. La couche réseau, qui relie le domaine physique au domaine virtuel. Elle permet de partager des données et des informations.
3. La couche informatique, constituée d'entités virtuelles émulant les entités réelles correspondantes, y compris les modèles et les analyses basés sur les données, les modèles basés sur la physique, l'application et l'utilisateur.

Les critères de conception du DT ne sont pas bien évalués, ni même standardisés. Ce sont les aspects les plus critiques.

En même temps, les capacités de base fournies par un jumeau numérique générique sont :

1. Emuler : voir, mettre à jour, conceptualiser.
2. Penser : comparer, raisonner.
3. Agir : informer, aider à la décision.

Pour la grâce de cette généralisation, nous envisageons les futurs sujets de recherche potentiels suivants :
Question de recherche n° 1.
NOUVELLES OPPORTUNITÉS COMMERCIALES

La plupart des applications jumelles numériques se réfèrent à une seule phase du cycle de vie d'un produit, et elles semblent être utilisées pour des processus standard existants. D'autre part, il y a, par exemple, relativement peu d'applications de jumelage numérique pour soutenir les entreprises de réseau. Le jumelage numérique pourrait potentiellement connecter des produits, des personnes, des machines et des entreprises dans l'espace virtuel. Le jumeau numérique peut potentiellement aider à intégrer même la chaîne d'approvisionnement entière, à travers toutes les phases du cycle de vie des produits. Ces potentialités peuvent donc aider à créer ou à développer de nouveaux scénarios ou de nouveaux modèles commerciaux pour la coopération entre entreprises.

Question de recherche n° 2.
CARACTÉRISTIQUES ÉVOLUTIONNELLES (adaptabilité, maintenabilité ou flexibilité dans le temps).
Le jumeau numérique peut être développé pour différentes fonctions dans chaque phase du cycle de vie en fonction du contexte ou du domaine d'application. Le Digital Twin doit évoluer de manière synchrone avec le système réel tout au long de son cycle de vie. Le DT est généralement appliqué dans des contextes caractérisés par l'incertitude et la complexité, où les conditions de travail peuvent varier en fonction de facteurs externes et internes. Le DT doit avoir la capacité de modifier sa configuration initiale et de s'adapter à la situation actuelle. Les questions de recherche devraient porter sur les critères de conception de jumeaux numériques capables d'évoluer, de s'adapter, de s'échelonner et/ou de se reconfigurer de manière autonome.

Question de recherche n°3.
FONCTIONNALITÉS D'INTERACTION
Les applications Digital Twin existantes sont principalement développées à des fins de prédiction et utilisées pour aider les décideurs humains à prendre des décisions. Les services assurés par DT peuvent varier en fonction non seulement des technologies disponibles mais aussi de sa conception : par exemple l'architecture, l'interaction des composants. En effet, le domaine inexploré de la recherche est celui des potentialités de synergie entre les caractéristiques et les fonctionnalités des composants, à condition que les technologies soient tout à fait nouvelles et inexplorées. L'interaction entre les composants est un aspect critique de la conception. Cette interaction peut modifier les structures internes de l'architecture en termes de propriétés ou de fonctionnalités. Les connexions prévues entre les composants n'assurent pas nécessairement une fonctionnalité orchestrée.

Question de recherche n°4.
ARCHITECTURE DT

Le DT consiste en un ensemble de modèles aux structures et comportements complexes, qui reflètent les opérations en temps réel du système physique.

L'absence d'une architecture de référence est presque bien évaluée (Lu et al. 2020). L'absence d'une définition univoque d'une architecture de référence numérique et univoque conduit à développer des solutions Digital Twin utilisant différentes technologies, interfaces et protocoles de communication, modèles et données. Les solutions Digital Twin standard doivent être développées pour fournir des critères de conception et des contraintes de conception où les aspects architecturaux de référence, le modèle d'information de référence et les protocoles de communication sont clairement définis.

Il existe des applications où les modèles basés sur les données et les modèles basés sur la physique sont fusionnés et d'autres où ils sont utilisés séparément. Cela dépend de la fonctionnalité pour laquelle le jumeau numérique a été réalisé. Cela signifie qu'il est nécessaire de clarifier les fonctions que le jumeau numérique peut remplir. Cela peut permettre d'identifier les composants de fabrication qui doivent être présents dans les modèles numériques et, par conséquent, de définir les types de modèles et les techniques de modélisation à utiliser.

Les caractéristiques d'interopérabilité et de précision sont assurées par une modélisation appropriée, afin de fournir une réplique virtuelle précise du système physique et d'obtenir des performances adéquates en temps réel. En même temps, le principe de modularisation doit être exploré pour améliorer l'efficacité de la modélisation : cela permettrait d'améliorer la flexibilité et la réutilisabilité des jumeaux numériques vers différentes applications.

Numéro de recherche n°5.

MODÉLISATION INVARIANTE DU DIGITAL TWIN

Le produit et le système de production deviennent de plus en plus complexes, à mesure que le nombre de leurs composants, la fréquence des changements de la demande du marché et le besoin d'innovation connexe augmentent. Pour gérer cette complexité, les représentations numériques constituent un défi majeur dans l'amélioration de la précision des outils de simulation et d'émulation existants et futurs. Une représentation numérique, donc un jumeau numérique, comble le fossé entre le système physique et virtuel en améliorant et en soutenant la prise de décision (Estefan 2007).

Le jumeau numérique nécessite la construction et l'application de modèles numériques représentant l'ensemble des ressources et des connaissances sur les processus. La modélisation d'une telle copie numérique du système physique pour effectuer une validation et une optimisation en temps réel est assez complexe et nécessite donc une grande quantité de données et quelques modèles de modélisation représentant la sémantique opérationnelle des éléments modélisés.

Cependant, il est difficile de construire un modèle précis en utilisant les approches traditionnelles basées sur des modèles en raison de la complexité des systèmes (J. Lee, Bagheri et Kao 2015). D'autre part, les récents progrès de la technologie des capteurs (Dassisti, Panetto, et al. 2017) ont permis une croissance significative de la collecte et de l'analyse des données, ce qui a conduit les chercheurs à se concentrer sur les méthodes basées sur les données. En général, l'action de modélisation a un type d'application spécifique. Pour cette raison, le principal défi de la modélisation de la transformation numérique consiste à créer une approche invariante pour différentes applications.

Dans ce chapitre, l'état de l'art de la transformation numérique dans les PME et le jumeau numérique ont été explorés. L'état de l'art vise à démontrer et à discuter des approches développées pour construire un jumeau numérique.
Le développement de modèles précis et reproductibles est considéré comme un défi pour le domaine de recherche du jumelage numérique.
Sur la base des résultats de l'examen, l'objectif de la thèse est d'aborder et de réaliser la modélisation invariante des systèmes de fabrication.
Afin de combler le manque de recherche lié à la modélisation invariante, le chapitre 2 présente un aperçu des approches existantes dans la littérature dans le but de définir le positionnement de la contribution de la thèse. Nous effectuons une revue exhaustive de la littérature dans la section 2.1 afin d'identifier et de décrire les approches basées sur des modèles dans la section 2.1.1, les approches basées sur des données dans la section 2.1.2, les approches hybrides dans la section 2.1.3 et le modèle de conception dans la section 2.1.4. En évaluant le contexte de la recherche et la limite identifiée, le positionnement de la contribution est défini à la section 2.2.

Dans le chapitre 2, un résumé des méthodes de modélisation des systèmes a été présenté. Les approches peuvent être classées en trois catégories : les approches fondées sur des modèles, les approches fondées sur des données et les approches hybrides (J. Luo et al. 2003). Le choix de la méthode à appliquer est généralement basé sur le cas d'utilisation spécifique.
Les approches basées sur les modèles reposent sur l'utilisation de modèles pour simuler le comportement des systèmes dans différentes conditions d'exploitation, mais ces modèles ne sont pas faciles à développer et à maintenir à jour pendant le cycle de vie du système (J. Luo et al. 2008).
Les approches fondées sur les données permettent d'intégrer des paramètres dans différents domaines (par exemple, le produit, le processus et la logistique) dans des modèles qui seraient difficiles à construire avec les approches traditionnelles fondées sur les modèles. En fait, elles visent à transformer les données en

informations pertinentes et en modèles comportementaux fiables (Okoh et al. 2014), mais la qualité et la portée des données jouent un rôle essentiel (Kusiak 2018).

Les principaux défis des approches fondées sur les données concernent les ensembles de données massifs et la haute dimensionnalité, c'est-à-dire les connaissances préalables nécessaires pour comprendre les modèles.

En même temps, une approche unique ne peut pas être adaptée à toutes les applications différentes en raison de la complexité et de la variété qui caractérisent les systèmes de fabrication. Des approches hybrides (Tidriri et al. 2016) ont été développées pour faire face à des problèmes spécifiques (X. Zhang et Hoo 2011). Les approches hybrides ne définissent pas une solution et un cadre communs à appliquer aux différents systèmes (Ghosh, Ng, et Srinivasan 2011).

Pour cette raison, le domaine des modèles de conception orientés objet est étudié. Le modèle de conception (OOP) définit, développe et met en œuvre une solution répétable à un problème courant dans le domaine du génie logiciel.

Le positionnement de la contribution de la thèse consiste à rendre la modélisation plus structurée et plus fiable. La question clé à traiter est celle de l'invariance. L'idée est de détecter automatiquement à partir des données les constructions de modélisation invariantes. Les constructions de modélisation invariantes doivent être développées pour décrire/émuler un système indépendamment de son contexte d'application. Les concepts de modélisation invariante basés sur des données peuvent être utilisés et notamment réutilisés pour créer des modèles numériques pour différents systèmes ou processus (Semeraro, Lezoche, et al. 2019).

The chapter 3 aims at formalising an approach to detect and to formalize invariant modelling constructs based on the data analysis. Each single state is described in a dedicated section.

The approach is articulated in eight different stages as shown in Figure 8. The stages are: 1) Definition of the system; 2) System model; 3) Data selection; 4) Knowledge discovery; 5) Knowledge extraction; 6) Knowledge formalization; 7) Design criteria for building a digital twin; 8) Knowledge re-use.

Le chapitre 3 vise à formaliser une approche permettant de détecter et de formaliser des constructions de modélisation invariantes basées sur l'analyse des données. Chaque état est décrit dans une section dédiée.

L'approche est articulée en huit étapes différentes. Ces étapes sont les suivantes :

1. Définition du système à analyser avec l'analyse énergétique.

Pour la définition du système à analyser, deux approches différentes ont été combinées. La première est l'analyse du cycle de vie car elle évalue la quantité de ressources consommées et les émissions relatives. La seconde est l'analyse énergétique, car elle définit l'infrastructure de l'information.

Les étapes de l'approche combinée sont présentées et décrites ci-dessous :

- Identifier le scénario de fonctionnement des systèmes analysés, ainsi que le choix des produits et le processus de fabrication associé en définissant le modèle thermodynamique des systèmes.
- Diviser le système en différents sous-systèmes et dessiner une représentation détaillée du fonctionnement de chaque sous-système considéré.
- Effectuer un bilan énergétique de chaque sous-système, en fournissant un indice critique basé sur la perte d'énergie.
- Définir les paramètres thermodynamiques critiques à mesurer pour chaque sous-système.

2. Modèle du système par SysML.

L'objectif du modèle de système est de montrer comment les composants du système, leur contenu (propriétés de valeur, comportements, contraintes) et leurs relations interagissent (Friedenthal, Moore et Steiner 2014). Le langage de modélisation choisi est le SysML.
L'outil choisi est IBM rational rhapsody® designer for systems engineers (https://www.ibm.com/uk-en/marketplace/systems-design-rhapsody). Il s'agit d'un logiciel qui permet de créer un environnement grâce auquel le modèle, conçu avec des règles standardisées, peut être simulé.
En particulier, l'outil fournit l'analyse des exigences du système, le développement visuel du modèle afin de capturer la conception graphiquement et, également, la simulation et l'exécution du modèle qui sont utiles pour valider le comportement du système.
Les étapes de modélisation à réaliser peuvent être résumées comme suit : 1) modéliser les exigences et le cas d'utilisation ; 2) modéliser les sous-systèmes dans des diagrammes de définition de blocs ; 3) modéliser les relations entre les sous-systèmes dans des diagrammes de définition internes ; 4) modéliser les contraintes dans des diagrammes paramétriques ; 5) modéliser le comportement du système dans des diagrammes de machines d'état.

3. Sélection des données.

Les étapes précédentes permettent de définir les données sélectionnées, leurs caractéristiques et les performances du système (Dassisti, Siragusa et Semeraro 2018). Les données sont un élément important pour la surveillance et la modélisation de systèmes complexes. Les données contiennent des informations utiles pour l'optimisation du système. La disponibilité de connaissances sur les processus, basées sur des systèmes de mesure solides (très souvent avec un réglage en ligne) permet de suivre clairement les caractéristiques du système et son évolution. Ces actions permettent de collecter et de convertir les données en informations, de partager les informations acquises, de formaliser les connaissances, de mesurer conjointement les performances et d'exploiter les compétences et les connaissances. L'idée principale ici est que la mesure est le fondement du modèle de connaissance d'un système réel. La surveillance continue, la collecte et l'analyse des données fournissent des informations actualisées sur les comportements du système dans un flux continu.

4. Découverte de connaissances par l'analyse des concepts de relations.

Les données sélectionnées dans les étapes précédentes sont utilisées pour la découverte des connaissances. Le data mining multi relationnel (MRDM) et plus particulièrement l'approche de l'analyse de concepts relationnels (RCA) est l'approche basée sur les données qui a été sélectionnée pour l'association de découverte (Rouane-Hacene et al. 2013). La RCA étend l'analyse formelle des concepts (FCA) en permettant l'information multi relationnelle (Rouane-Hacene et al. 2013).

Étant donné un ensemble d'objets, un ensemble d'attributs, et défini les relations entre les objets et les attributs, un concept formel représente un sous-ensemble d'objets partageant le même sous-ensemble d'attributs. Un concept est constitué de deux parties : son extension qui consiste en tous les objets appartenant au concept, et son intention qui comprend tous les attributs partagés par ces objets. Cette compréhension permet de découvrir formellement des associations entre les concepts et donc de reconnaître quels concepts sont étroitement liés en fonction de l'ensemble des attributs partagés (Williams et Simoff 2006). La RCA étend la FCA au traitement d'ensembles de données multi-relationnelles, chacun doté de son propre ensemble d'attributs, et aux relations entre ceux-ci (Rouane-Hacene et al. 2013).

5. Extraction des connaissances.

Un modèle invariant basé sur des données contient les sections suivantes :

- ID Pattern : nom unique qui aide à identifier le modèle.
- Nom du modèle.
- Description : une description de l'objectif du modèle et de la raison de son utilisation.
- Exemple de règle d'association : un exemple de concept pour le modèle dans l'analyse.
- Vue des données : une représentation des données pour le modèle en analyse.
- Vue du modèle SysML : une représentation graphique de l'association entre les données. Un diagramme de définition de bloc ou un diagramme de définition interne ou un diagramme paramétrique ou d'état peut être utilisé à cette fin.
- Applicabilité : situations dans lesquelles ce modèle est utilisable.

6. Formalisation de la connaissance.

Les connaissances extraites doivent être formalisées pour être compréhensibles et accessibles. Le système d'entrepôt de données peut être utilisé pour la collecte, la connexion, la gestion et l'analyse de données provenant de sources hétérogènes. Le langage SQL (Structured Query Language) est le langage standard qui peut être utilisé pour le stockage, la manipulation et la récupération de données stockées dans une base de données relationnelle : 1) Exécuter des requêtes sur une base de données ; 2) Créer des procédures stockées dans une base de données ; 3) Créer des tableaux de bord pour représenter la signification de chaque construction guidée par les données. De cette façon, chaque construction peut être facilement formalisée, analysée et interprétée.

7. Critères de conception.

Les constructions formalisées basées sur des données représentent les critères de conception pour la construction d'une interface numérique simple et efficace pour l'utilisateur final afin d'aider les employés dans le processus de prise de décision. Une construction guidée par les données peut inclure des règles, des contraintes et des déductions liées aux processus de production, telles que la contrainte de la capacité de traitement d'un certain équipement. Elles peuvent être formalisées dans des algorithmes pour que le DT puisse juger, évaluer, optimiser et/ou prévoir.

8. Réutilisation des connaissances.

Les connaissances formalisées peuvent être utilisées directement dans d'autres systèmes pour concevoir des jumeaux numériques ou pour d'autres applications. L'objectif est d'utiliser les mêmes modèles pour la modélisation d'autres systèmes. De cette façon, des modèles basés sur des données peuvent être combinés, en fonction de l'application spécifique, pour créer facilement des modèles dynamiques.

Le chapitre 4 présente une case d'étudie et l'application de l'approche développé.
Master Italy s.r.l, est une PME italienne qui produit de petites ferrures pour les châssis de fenêtres civils (https://www.masteritaly.com/). Master Italy réalise 97% de la valeur ajoutée de sa production interne, suivant la stratégie d'intégration verticale, poursuivie depuis sa fondation.
Le groupe Master réalise 97% de la valeur ajoutée de sa production interne en couvrant toutes les phases précédant la vente du produit : de l'analyse des besoins du marché à la conception, au prototypage et à la production des marchandises. Ce choix, dans le passé, a été couronné de succès pour les raisons suivantes : a) réduction des délais, b) maîtrise des coûts, c) accumulation du savoir-faire en matière de processus, d) croissance de l'offre. Actuellement, cependant, les tendances du marché identifient des problèmes tels que

la maîtrise des coûts des produits, l'introduction de nouvelles technologies, la variabilité de la demande du marché, la réduction de la taille moyenne des commandes, la demande de produits personnalisés.

Dans ce contexte, l'entreprise a décidé de travailler stratégiquement sur l'innovation des processus et des produits par l'introduction ou le développement de technologies liées à l'industrie du futur.

Les avantages escomptés du Master en ce qui concerne l'augmentation de l'"intelligence" de l'usine sont résumés comme suit :

- Aider les gens à prendre des décisions rapidement et efficacement, à réduire les erreurs, à travailler de manière efficace et ergonomique.
- Analyser les différentes lignes de production, afin d'identifier les problèmes qui, s'ils sont comparés, déterminent l'amélioration exponentielle des performances de l'ensemble du système
- Evaluer les actions possibles à entreprendre, lorsque les processus de production sont exposés à des événements extérieurs.
- Prendre des décisions et faire des prévisions plus précises en termes de production et de consommation.
- Identifier et quantifier les ressources qui contribuent à l'augmentation de l'efficacité des systèmes.
- Contrôler et superviser l'utilisation des ressources dans les différentes phases du processus de production.
- Partager et intégrer les informations entre tous les membres de l'entreprise.
- Optimiser les performances de l'entreprise.

L'approche décrite au chapitre 3 a été appliquée à la conception et au développement d'un jumeau numérique, basé sur des modèles invariants basés sur des données :

- Simplifier la programmation et le contrôle de la production.
- Simplifier la programmation et le contrôle de la production. Augmenter la productivité.
- Éliminer les erreurs lors des opérations.
- Réduire la courbe d'apprentissage de l'opérateur.

Il est utile de prouver la qualité de l'idée et de répondre aux exigences mentionnées ci-dessus.

L'analyse LCA a été rédigée en Master. L'objectif de l'analyse LCA est d'évaluer la quantité de ressources nécessaires et les émissions produites pour fabriquer les différents composants. L'LCA a été réalisée à l'aide du logiciel SIMAPRO® en utilisant la base de données Eco invent.

Les produits les plus importants de l'entreprise sont les charnières, les coins en acier, les poignées et les oscillo-battants utilisés pour les fenêtres en aluminium. Grâce à l'évaluation LCA, le produit sélectionné, est le coin en acier puisque l'impact du PRP100 est : 0,282115 kg CO_2eq/pcs.

Les différents composants de l'angle en acier sont soumis à plusieurs processus mécaniques : moulage sous pression de l'aluminium, moulage sous pression du zamak, vernissage et assemblage. En analysant chaque processus de production, la plus grande contribution est donnée par le processus de moulage sous pression de l'aluminium (1 1395 kg CO_2eq/pcs en raison de la consommation de méthane).
Es connaissances formalisées peuvent être utilisées directement dans d'autres systèmes pour la conception de jumeaux numériques ou pour d'autres applications. L'objectif est d'utiliser les mêmes modèles pour la modélisation d'autres systèmes. De cette façon, des modèles basés sur des données peuvent être combinés, en fonction de l'application spécifique, pour créer facilement des modèles dynamiques.

Le SysML vise à formaliser les aspects suivants du système : 1) Composition structurelle, interconnexion et classification ; 2) Contraintes sur les propriétés physiques et les performances ; 3) Comportement basé sur les fonctions et les états ; 4) Allocations entre comportement, structure et contraintes.

Les données sélectionnées sont des données provenant de : PLC, MES. Les bases de données PLC et MES couvrent les 18 derniers mois de production.
La RCA est appliquée pour trouver automatiquement les associations entre le paramètre technologique, le temps d'arrêt de la machine et le temps d'arrêt de la machine des caractéristiques.
Des exemples des patterns découvertes sont dans les tables :

P1: MOULD FILLING PATTERN

- **ID: P1**
- **NOM DU MODÈLE : MOULD FILLING PATTERN**
- **DESCRIPTION :** le modèle vise à montrer et à décrire les corrélations entre les paramètres C1, C2, CC. Le modèle représente la course du piston pour le remplissage du moule. Cela a un impact sur la qualité du produit (SM).

P1 EXEMPLE DE REGLE D'ASSOCIATION :

14675 : T1_1, SM_1 →
C1_5, C2_1, V1_5, T2_1, CC_2, KO, CHECK, Ex gen pbs : (MICROSTOP FOR SM_LimLow, , MEDIUM COST)
[0.00347877, 1]

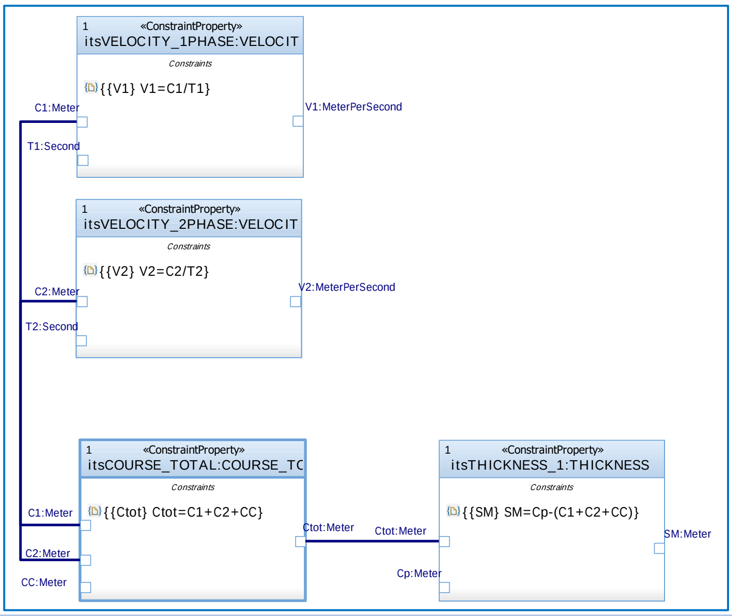
P1 VUE DE DONNÉES :

L'ensemble des données comprend :

- Parcours du piston C1 (mm) dans la première phase de l'injection
- Course du piston C2 (mm) dans la deuxième phase de l'injection
- Course du piston CC (mm) dans la phase multipliée de l'injection
- Épaisseur du produit SM (mm²)

C1	C2	CC	SM
295	81	3	17
295	79	3	17
297	79	4	16
297	79	3	17
297	79	3	17
296	81	3	16
297	80	3	16
297	80	3	16

P1 VUE DU MODÈLE SYSML :



P1 APPLICABILITÉ : Utilisez le modèle de remplissage de moule pour :

- Décrire et prévenir les problèmes de micro-arrêt.
- Décrire et prévenir les problèmes lors des processus de remplissage.
- Décrire le comportement du cours du piston.
- Décrire le comportement d'un système d'injection.

P2: MACHINE RESTART PATTERN

- **ID: P2**
- **NOM DU MODÈLE :** MACHINE RESTART PATTERN
- **DESCRIPTION :** le modèle vise à montrer et à décrire les corrélations entre V2, T2, PM, PS, SM, TC. Le modèle indique le redémarrage de la machine après un temps d'arrêt. Le redémarrage d'une machine génère une refonte du produit.

P2 EXEMPLE DE REGLE D'ASSOCIATION :

78953 : TC_5, T2_5, PS_1, SM_5, Ex gen pbs : (RECASTING START MACHINE AFTER CLEANING OF THE EQUIPMENT)
→ PM_1, V2_1, KO
[0.000484605, 1]

P2 VUE DE DONNÉES :

L'ensemble des données comprend

- La vitesse du piston V_2 (m/sec) dans la deuxième phase de l'injection.
- Le temps de piston T_2 (msec) dans la deuxième phase de l'étape d'injection.
- Pression multipliée PM (Pa) dans la phase multipliée de l'étape d'injection
- Pression spécifique PS (Pa)
- Épaisseur du produit SM (mm²)
- Temps de cycle TC (sec)

V2	T2	PM	PS	SM	TC
35	175	215	928	20	33
37	180	229	946	20	33
37	167	231	993	20	33
122	59	223	928	19	33
120	60	224	933	19	33
127	55	223	946	19	33

P2 VUE DU MODÈLE SYSML :

P2 APPLICABILITÉ : Utilisez le modèle de redémarrage de la machine lorsque :

- Décrire et prévenir les problèmes de refonte.
- Décrire le redémarrage d'une machine.

P3 CLAMPING SYSTEM PATTERN

- **ID: P3**
- **NOM DU MODÈLE :** CLAMPING SYSTEM PATTERN
- **DESCRIPTION :** Le modèle vise à montrer et à décrire les corrélations entre les PM, les PS et les FC. Le modèle identifie l'étape de refroidissement du processus de moulage de l'aluminium sous pression. Les pressions multipliées sont la force de serrage et les paramètres de contrôle de l'état du système de serrage. Les valeurs de la force de serrage FC et de la pression multipliée PM peuvent générer des problèmes mécaniques.

P3 EXEMPLE DE REGLE D'ASSOCIATION :

9302 : PM_2, Ex gen pbs : (HIGH COST) →
A_M, 10A035, FC_4, CHECK Ex gen pbs : (MECHANICAL MAINTENANCE, MOULD PROBLEM, PLUNGER PROBLEM, MECHANICAL COMPONENT PROBLEM, DOWN TIME (>60MIN), HIGH COST)
[0.000501912, 1]

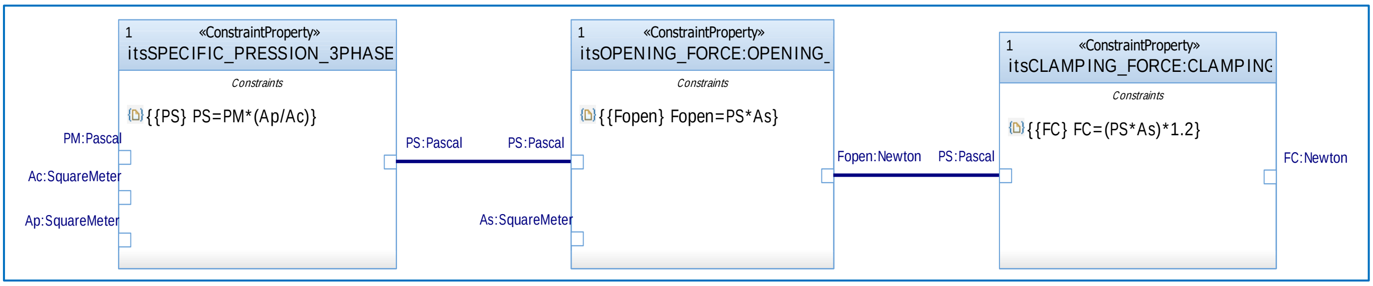
P3 VUE DE DONNÉES :

L'ensemble des données comprend :

- Force de serrage FC (kN)
- Pression multipliée PM (Pa)

PM	FC
217	5780
217	5737
218	5780
218	5748
218	5780
218	5791
217	5759
217	5758

P3 VUE DU MODÈLE SYSML :



P3 APPLICABILITÉ : Utilisez le modèle de système de serrage pour :

- Décrire le comportement d'un système de serrage (par exemple, le serrage d'un moule).
- Décrire et prévenir les pannes mécaniques.

Les patterns décrits (P1, P2, P3) ci-dessus ont été utilisés pour concevoir le jumeau numérique du processus de moulage sous pression afin de prévoir les micro-blocs et les problèmes mécaniques, comme le montre la figure 50.

Les solutions proposées ici consistent en un environnement de production réaliste, mais "enrichi" de connaissances technologiques intrinsèques. Avec notre approche, les paramètres physiques interagissent avec l'espace numérique, selon des propriétés et des règles spécifiques, à comprendre :

- Le comportement du processus.
- Les corrélations entre les paramètres technologiques.
- Les corrélations entre les paramètres et les effets tels que les défauts de qualité, les problèmes de maintenance, etc.

Le jumeau numérique a été conçu pour aider les employés dans le processus de prise de décision :

- Identifier de manière autonome les différents problèmes de qualité des composants, par rapport aux normes (dimensions, tolérances, finitions, quantité).
- Alerter les opérateurs par des systèmes d'alarme appropriés en cas de situations anormales ou hors tolérance.
- Analyser et corréler les symptômes et les causes des défaillances et des défauts de production.
- Soutenir le choix des actions correctives pour éliminer les défaillances et défauts détectés.

Le jumeau numérique reçoit et lit un ensemble de données, émule le comportement de la machine et reconnaît les éventuelles défaillances.

Les connaissances formalisées peuvent être utilisées pour la conception numérique jumelle d'autres procédés de fabrication de l'entreprise. L'objectif est d'appliquer les mêmes modèles de conception basés sur les données dans d'autres domaines, par exemple les besoins des clients, la conception des produits et le suivi des produits.

Cette thèse contribue à définir et à présenter la manière de construire des modèles invariants guidés par les données pour la transformation de la modélisation numérique.

Au chapitre 1, la transformation numérique dans les PME a été analysée. En particulier, le principal problème lié à la mise en œuvre de l'industrie du futur dans les PME est de comprendre les nouvelles technologies, et de définir comment, quand et où elles pourraient être appliquées. Le jumelage numérique a été examiné pour résoudre ce problème.

Dans cette thèse, le jumeau numérique est défini comme suit : << Un modèle adaptatif qui émule le comportement d'un système physique en obtenant des données en temps réel pour se mettre à jour tout au long de son cycle de vie. Le jumeau numérique reproduit le système physique pour prédire les défaillances et les possibilités de changement, afin de prescrire des actions en temps réel pour optimiser et/ou atténuer les événements inattendus en observant et en évaluant le profil du système d'exploitation>>.

L'état de l'art du Digital Twin a permis de définir les lacunes de la recherche. La modélisation invariante est la question de recherche à aborder. Une revue de la littérature sur les approches basées sur des modèles, guidées par les données et hybrides a été réalisée au chapitre 2. Elle nous a permis d'identifier les approches concernant les approches basées sur les modèles et les approches guidées par les données. D'après les résultats de l'analyse, une limitation importante est que la modélisation de la copie numérique du système physique est assez complexe et que l'action de modélisation a généralement un type d'application spécifique.

Pour cette raison, le principal défi consiste à créer des constructions de modélisation invariantes basées sur des données à utiliser et à réutiliser pour différentes applications. La recherche consiste à formaliser et à normaliser des constructions de modélisation invariante pour modéliser différents systèmes.

Le résultat est la formalisation d'une approche dans le chapitre 3 pour découvrir des constructions basées sur des données. Elle implique la combinaison d'approches basées sur les données et les modèles avec des modèles de conception pour définir et identifier la modélisation invariante basée sur les données. L'approche itérative comprend huit étapes différentes :

1. Définition du système à analyser avec l'analyse énergétique.
2. Modèle du système par SysML.
3. Sélection des données.
4. Découverte de connaissances par l'analyse des concepts de relations.
5. Extraction des connaissances.
6. Formalisation de la connaissance.
7. Critères de conception.
8. Réutilisation des connaissances.

Enfin, au chapitre 4, nous avons évalué notre contribution au moyen d'une étude de cas : Master Italy s.r.l. Le procédé de moulage sous pression de l'aluminium est analysé sur la base du résultat d'une analyse énergétique. Le processus de moulage sous pression est modélisé en SysML pour décrire les exigences, la structure et le comportement. L'analyse du concept de relation est utilisée pour détecter automatiquement les associations explicites et tacites dans les données. Cela a permis de formaliser des constructions de modélisation invariantes basées sur des données. Trois modèles ont été utilisés pour développer une plate-forme web. Un jumeau numérique du processus a été réalisé à l'aide d'un ensemble de constructions guidées par les données découvertes.

Le travail présenté dans cette thèse induit d'autres pistes de recherche qui peuvent être envisagées dans le futur :

- Le premier axe de recherche consiste à enrichir la sémantique du modèle pour créer une bibliothèque complète de modèles formalisés basés sur des données. Cela signifie sélectionner différentes lignes de production pour appliquer les mêmes modèles afin de concevoir des modèles pour différentes applications et portées. D'autres ensembles de données doivent être sélectionnés, tels que les données logistiques, les données sur les produits, les données sur les clients, afin d'extraire de nouveaux modèles basés sur des données.
  De cette façon, il est possible de créer une bibliothèque de modèles cohérente. Une bibliothèque de modèles est une collection de modèles qui décrivent un problème, un comportement ou une situation récurrente. De cette façon, les modèles basés sur des données peuvent être combinés pour créer facilement des modèles virtuels.

- L'état de l'art démontre que les approches basées sur des modèles ou guidées par les données sont plus analysées que les approches hybrides. Par conséquent, la deuxième orientation de la recherche consiste à proposer de nouvelles contributions pour la combinaison des approches fondées sur des modèles et des données afin d'améliorer la modélisation des systèmes.

- Le troisième axe de recherche consiste à explorer la sémantique et la syntaxe des modèles afin de mieux formaliser le concept d'invariance. Qu'est-ce que l'invariance ? Comment définir l'invariance ? Comment reconnaître l'invariance dans les modèles ? Ce sont quelques-unes des questions qui pourraient guider la recherche.

- La quatrième direction de recherche consiste à automatiser l'approche afin de reconnaître automatiquement les nouvelles constructions de modélisation pilotées par les données en utilisant des algorithmes d'apprentissage automatique ou des réseaux de neurones. Cela signifie qu'il faut fixer des critères ou des éléments communs pour effectuer la reconnaissance.