POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Knowledge-based Model Enhancement through Conformance Checking techniques in Manufacturing Systems

Tesi di Laurea Magistrale in
Mechanical Engineering - Ingegneria Meccanica

Author: **Spilotros Lorenzo - Urbinati Daniele**

Student ID: 10579978 - 10619007
Advisor: Prof. Tullio Antonio Maria Tolio
Co-advisors: Dr. Maria Chiara Magnanini
Academic Year: 2021-22

Abstract

The work is set in a context in which digitization has become a driving force for change and improvement in many areas as a result of the fourth industrial revolution. The study aims to develop a new methodology for extracting information from the manufacturing systems based on a comparison between the nominal behavior of the system and the real one. Process Mining techniques, such as Conformance Checking and Model Enhancement, can be used to make such a comparison. The existing methodology coupled these techniques with Process Discovery, a technique that is advantageous when applied in contexts in which the process model is not known a priori. Since manufacturing processes are known a priori, the work proposes to replace Process Discovery with a system modeling phase based on the study of available process knowledge. The knowledge is translated into a Petri Net model representing the manufacturing system and it is used to highlight deviations with the actual process. In this study, it is intended to demonstrate that the developed methodology can avoid classical problems presented by the existing methodology by better extracting information from process data. This demonstration is developed by studying the results obtained applying the methodology to a case study and through an experiment aimed at comparing the two methodologies.

Keywords: Industry 4.0, Process Mining, Conformance Checking, Manufacturing Systems, What-if Analysis, Knowledge Extraction, Petri Net.

Abstract in lingua italiana

Il lavoro si inserisce in un contesto in cui, a seguito della quarta rivoluzione industriale, la digitalizzazione è diventata una forza motrice di cambiamento e miglioramento in molti ambiti. Lo studio vuole sviluppare una nuova metodologia di estrazione delle informazioni dai sistemi manifatturieri, basata sul confronto tra il comportamento nominale del sistema e quello reale. Per effettuare tale confronto è possibile sfruttare tecniche di Process Mining come il Conformance Checking ed il Model Enhancement. La procedura attuale accoppia queste tecniche al Process Discovery, tecnica che risulta vantaggiosa se applicata in contesti in cui il modello di processo non è noto a priori (metodologia esistente). Essendo i processi manifatturieri noti a priori, lo studio propone di sostituire il Process Discovery con una fase di modellazione del sistema basato sullo studio della conoscenza aziendale disponibile. Tale conoscenza viene tradotta in un modello di rete di Petri che rappresenta il sistema di produzione e viene usata per evidenziare deviazioni tra il processo reale e la conoscenza nominale del processo. In questo studio si vuole dimostrare che la metodologia sviluppata può evitare problematiche classiche che la metodologia esistente presenta, migliorando l'estrazione di informazioni dai dati di processo. Tale dimostrazione viene sviluppata studiando i risultati ottenuti dall'applicazione della metodologia ad un caso studio e attraverso un esperimento volto al confronto delle due metodologia.

Parole chiave: Industria 4.0, Process Mining, Conformance Checking, Sistemi di Produzione, Analisi di Scenari, Estrazione della Conoscenza, Petri Net.

Contents

1 | Introduction

The third industrial revolution, which is based on widespread digitization, provides the basis for a fundamental paradigm shift, the fourth industrial revolution or Industry 4.0. Compared with the third revolution, a further step has been taken toward improvement and efficiency from various points of view by advancing digitization within factories and introducing Internet technologies. The increasing digitization has enabled the recording of a large amount of process data. In fact, digitization is becoming a driving force for new process analysis methods and technologies such as simulations or digital twins.

The starting point for applying Industry 4.0 technique in a production system is the Manufacturing Execution System (MES). MES stands in the middle between Cyber-Physical Systems (CPS) and Cyber-Physical Production Systems (CPPSs), as it can flow vertically and horizontally in the process integrating all systems initially separated by stage, from supply chain to production. Therefore, MES is the basis for data collection and integration in manufacturing systems.

In this context, Process Mining, which was born in 1999 with the aim of understanding business processes, was also developed in manufacturing. Process Mining techniques, that are summarized by Process Discovery, Conformance Checking and Model Enhancement, use data from MES to improve processes and better understand the real behavior of products on the shop floor. Applying Process Mining to the manufacturing domain, it is possible to understand whether processes are in line with expectations. If not, Process Mining techniques will provide all the information in an orderly and comprehensive visualization to determine where and how to improve processes. Starting from logs, Process

Mining techniques allow processes to be discovered, monitored and improved by extracting relevant knowledge.

This study proposes the development of a novel methodology for knowledge-based model enhancement through Conformance Checking techniques in manufacturing systems. The idea arises from the fact that Process Discovery in the manufacturing systems can be considered unnecessary and redundant since every company has structured and available knowledge of its production process. Production processes appear to be already known a priori. Therefore, it is no longer necessary to discover processes from the logs. However, the model construction can be based on translating the abundant knowledge already available and structured into an appropriate model.

Furthermore, the novel methodology is validated using data collected over 65 months by a leading company in innovative technologies and complete lifecycle solutions for the marine and energy markets. Lastly, to objectively compare the two methodologies, a comparison is proposed based on the results of an experiment conducted with 8 working groups performing the two methodologies in parallel.

The workflow of the paper is organized as follows: first, the context in which the work is developed is explained through the Context in Chapter 2 and State of Art in Chapter 3. Next, in Chapter 4, an explanation of both the existing and proposed methodology is given, focusing on their drawbacks and advantages. Then, the proposed methodology is applied to a case study (Chapter 5). The work is further enriched by an experiment, in Chapter 6, in which the models were quantitatively evaluated through 8 working groups. Finally, the Conclusions (7) summarize critically the related work.

2 | Context

It is amply demonstrated how industrial revolutions have changed life on our planet and brought human evolution to another level, just think of the demographic increase due to better welfare reported in Figure 2.1.

The first industrial revolution (1784) replaced manual labor with water and steam power resulting in revolutionary mechanical production facilities. The second industrial revolution (1870) introduces the development of mass production and usage of electrical power electronics, making many products accessible to a poorer portion of the population by increasing the quality of life. A century later, the third industrial revolution (1969) changed the way manufacturing was developed by introducing automated production and the development of IT technologies. Forty years later, society is living through a new revolution characterized by the internet, big data, networks, and connectivity: Industry 4.0 (I4.0).

The differences between the first, second, and third revolutions are various. The third revolution brought considerable maturity in technology, process, and production. Between the third and the fourth revolution, a second step was taken towards improvement and efficiency from different points of view: from increasingly going towards the customer's needs to the reduction of raw material consumption.

Researchers started talking about I4.0 in 2011. After 11 years, only 35% [98] of European companies are adopting I4.0 solutions in those areas: production, IT, product development, human resource, logistic, maintenance, supply chain, finance, sales, and marketing. Some of the causes of the poor involvement of the industry in implementing I4.0 solutions are low responsiveness of the sector to absorb innovation, inability to see benefits, and to

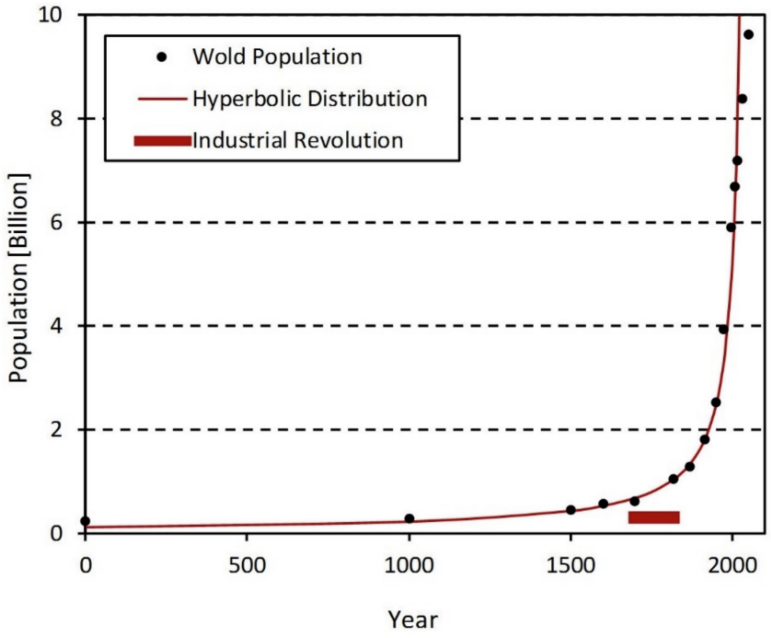


Figure 2.1: Growth of the world population [71].

be open to sharing data with external partner companies.

However, since 2018, companies have been interested in I4.0 solutions, mainly due to the increasing global trade that humanity is experiencing. Over the years, the volume of goods traded between different countries and trade routes has increased significantly, as shown in Figure 2.2, leading local producers to compete with companies working in contexts completely different from their local ones. Firms in low-development areas can count on low labor costs, while those in highly developed areas on more advanced technological processes.

This gap has been further accentuated in recent years due to the lower cost of access to new technologies. Thanks to this, emerging reality has improved their technological processes and become even more competitive, pushing more developed companies to review their business methods, for example, lowering production costs or adding more value to the finished products. Thus, Industry 4.0 started to be seen as a way to exploit more advantages from existing businesses.

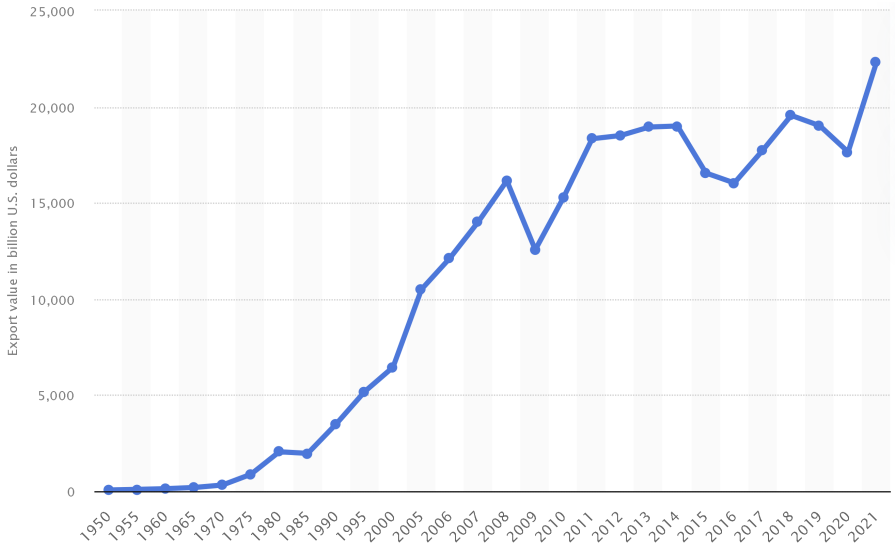


Figure 2.2: Trends in global export value of trade in goods from 1950 to 2021 [3].

2.1. Industry 4.0

Industry 4.0 is transforming how businesses produce, enhance, and distribute their goods. The Internet of Things (IoT), cloud computing, analytics, Artificial Intelligence (AI), and machine learning are among the cutting-edge technology that manufacturers are incorporating into their manufacturing processes.

Companies are moving from being simple production sites to smart factories: manufacturing facilities with cutting-edge sensors, embedded software, and robotics that gather data, analyze them, and help with decision-making. A higher value is created when shop floor data are combined with data from enterprise systems like ERP, MES, supply chain, customer service and others to create new levels of visibility and insight from previously siloed information. These digital technologies lead to increased automation, predictive maintenance, optimization of process improvements and, above all, a new level of efficiencies and responsiveness to customers not previously possible. Industry 4.0 concepts and technologies can be applied across all industrial companies, including discrete and process manufacturing.

It is not possible to talk about I4.0 tools, but it is possible to highlight some principles

or practices that digitalization and integration made possible [1]:

- **The Internet of Things** (IoT): it is a key component of smart factories. Machines on the factory floor are equipped with sensors with an IP address that allows them to connect to other Web-enabled devices or local servers for processing. The mechanization and connectivity make those connection possible and generate large amounts of valuable data to be collected, analyzed, and exchanged. In the last years, there is a tendency to move machines from OT to IT networks to have data available directly in cloud solutions, which allows heavier computational analysis, such as neural network computation in big database centers.

- **Cloud computing** Data centres normally have much greater computing power than company local servers, this allow to process vast amount of data ande evaluated more quickly and affordably, but also in a more sustainable way. For small and medium-sized manufacturers who can appropriately assess their demands and scale as their firm grows, cloud computing can significantly lower startup costs.

- **AI and machine learning** allow manufacturing companies to take full advantage of the volume of information generated not just on the factory floor, but across their business units, and even from partners and third-party sources. AI and machine learning can produce insights that give operations and business processes visibility, predictability, and automation. Company can do machine learning-based predictive maintenance using data gathered from these assets, increasing up-time and efficiency.

- **Edge computing**: the demands of real-time production operations mean that some data analysis must be done at the "edge", namely where the data is created. This local data computation reduces the amount of time between the production of data and the need for a response. For instance, the device may need to be used to take action in close to real time when a safety or quality concern is discovered. Depending on how reliable the network is, it may take too long to transport data

from the manufacturing floor to the enterprise cloud and back. Data stays close to its source when edge computing is used, lowering security threats.

- **Cybersecurity**: Manufacturing companies have not always considered the importance of cybersecurity or cyber-physical systems. However, the same connectivity of operational equipment in the factory or field (OT) which enables more efficient manufacturing processes, also exposes new entry paths for malicious attacks and malware. When undergoing a digital transformation to Industry 4.0, it is essential to consider a cybersecurity approach that encompasses IT and OT equipment.

- **Digital twin**:the digital transformation offered by Industry 4.0 has allowed manufacturers to create digital twins, that are virtual replicas of processes, production lines, factories and supply chains. Data is collected from IoT sensors, gadgets, PLCs, and other internet-connected devices to construct a digital twin. They are tools that manufacturers can employ to create new goods, streamline workflows, and boost production. For instance, manufacturers can test changes to the process to find ways to reduce downtime or increase capacity by modeling a production process.

All those tools offer an unprecedented opportunity to improve business, based on their decentralized vision, autonomous networks, and smart products. This is the direction that manufacturing industries need to take to achieve intelligence, resource efficiency, and high performance.

Data collection is the real common thread in all those tools: data are the starting point upon which each analysis starts. Data extraction is not trivial and, because of this, it is the first area in which company need to be good is data collection (garbage in, garbage out): the way in which data were collected in the last decades needs to be revised.

2.2. The Crucial Role of Data in manufacturing system

During the third revolution, most of the collected data were used only for direct feedback control in real time and for forensic purposes. Nowadays, with I4.0, the scope of data collection has drastically changed. State of art of Industrial analytics is able to process data and exploit important knowledge compared to previous data usage. Transparency about operations of the equipment, materials used, facility logistics, and even the human operators is made possible by data generation. Data analytics, namely, the use of statistical tools and machine learning techniques to identify specific data properties and patterns, is what makes this openness possible. Machine learning, a branch of data analytics, is being applied in various industrial applications, including process flow optimization, internal defect reduction, warranty claim reduction, predictive maintenance, and test time reduction [47]. The general goal of using analytics in manufacturing is to improve productivity by reducing costs without compromising quality. This, in turn, makes the manufacturing process efficient. It is possible to see a recurrent need for analytics among five categories:

1. Reducing test time and calibration. This includes predicting test results and calibration parameters. Collecting and analyzing data from different machines allows keeping machining and setup time updated.

2. Improving quality. This means the reduction of producing scrap (bad parts) costs by identifying the root cause for scrap and self-optimizing the assembly line.

3. Reducing warranty cost. This includes using quality testing and process data to predict field failures, as well as cross value-stream analysis.

4. Improving yield. This includes conducting benchmark analyses across production lines and facilities, enhancing first-pass yield, and identifying the root causes of performance bottlenecks, such as Overall Equipment Effectiveness (OEE) or cycle

times for products that pass quality inspections after a single iteration.

5. Performing predictive maintenance. This includes analyzing machine health, identifying the top causes of failure, and predicting component failures to avoid unscheduled machine downtimes.

Besides these categories, there are multiple other advantages that a company may benefit from by performing analytics. All services that support production, such as supply chain, warehouse and business management, take important advantages from data analysis.

In addition to the highlighted gains, there is a large indirect benefit to structure knowledge acquisition. In industry it is usually difficult to have different departments of the same company aligned and constantly up to date with the latest data. Analyzing data forces problems to be solved and, usually, tools that collect data are managed by different departments that need to cooperate and to be aligned together. A lot of research validates how structured procedures are needed to enable continuous improvement.

Even if industry is starting to adopt analytical tools, with millions of parts being produced on manufacturing lines and with thousands of processes and quality measurements collected for each of them, improving existing processes is not trivial.

2.3. Automation Pyramid and Information Systems

Focusing on production systems, the main question that should be addressed is: *how can data be collected in order to be useful for further analysis*? For this scope, the concept of Automation Pyramid shown in Figure 2.3 was developed in the 1980's, and it has been used for decades by companies as a guide for organizing data collection and interaction among different levels.

At the **field level** (0) there are sensors and actuators that are responsible for collecting production data and executing commands. The **control level** (1) is responsible for

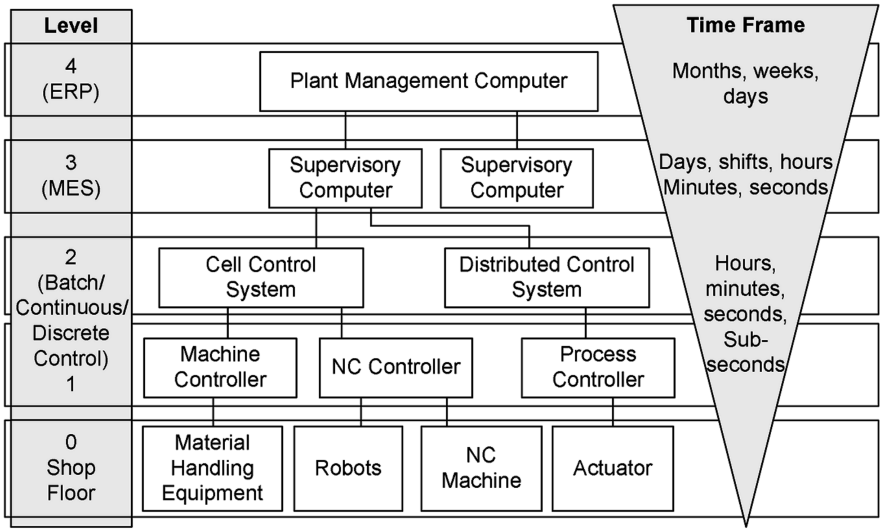


Figure 2.3: Automation Pyramid [101].

controlling and manipulating all devices that are at the field layer. At the **supervisory level** (2) there are SCADA systems for supervisory control and data acquisition. At the **Planning level** (3) there is the Manufacturing Execution System (MES) that monitors the entire manufacturing process in a plant or factory from raw materials to finished products. The top of the pyramid is what it is called **management** (4) level. This level uses the companies integrated management system, which is known as Enterprise Resource Planning (ERP), to manage and control all operations.

However, after the new techniques that industrial analytic brought, it is not yet possible to visualize data fully separated between layers. Because of this, the automation pyramid is now strongly discussed. Enabling data interaction between different layers allows a better overview of the system.

Having a broad overview is critical to interpret and solve problems. Personnel who lack of it is not inclined to think outside the box and tends to be less able to solve problems. The importance of having a big picture of the context was already known by Sophocles in 440 BC: “Let not your first thought be your only thought, think if there cannot be some other way. Surely, to think your own the only wisdom, and yours the only word, the only will, betrays a shallow spirit, an empty heart”.

Before the last decades, within manufacturing companies, the privilege to have an overview of the process was in the hand of very few people. It is fascinating to see how industry took so much time to reach the integration between the layers of the pyramid. Renee Guzlas has been fantastic in emphasizing the overview concept (Figure 2.4):

Figure 2.4: The blind men and the elephant. Poem by John Godfrey Saxe (Cartoon originally copyrighted by the authors; G. Renee Guzlas).

Several voices suggest a flattening of the pyramid to allow the flexibility and interconnectivity which Industry 4.0 aims at. A proposal for change comes from **??**, who shown how the implementation of a Cyber-Physical System (CPS) and Cyber-Physical Production Systems (CPPSs) would be the answer for the interaction of each layer. Figure **??** show an example of automation pyramid of the future.

CPSs are physical objects with embedded software and computing power and will incorporate self management capabilities after the introduction of Industry 4.0 to the plant floor. On the other side, CPPSs represent the production facilities, which leverage on different software enhanced machines, capacities and configuration options. CPS is capable of autonomously exchanging information, triggering actions and controlling each other independently, thus allowing the shop floor to become a market place of capacity.

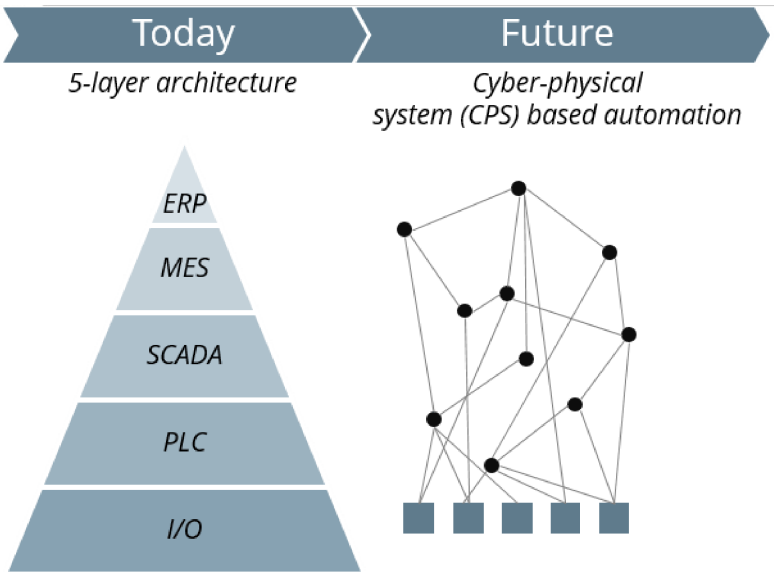


Figure 2.5: Automation pyramid of the future [68].

Smart material and products (in CPS) are service consumers and smart equipment and plants (in CPPS) are service providers. The combination of CPS and CPPS is likely to trigger significant changes in manufacturing production and to control towards completely decentralized systems.

At the heart of CPS and CPPS there is the Manufacturing Execution System (MES), which is likely to play an essential role in the manufacturing enterprise's information technology landscape as it sits at the critical point where revenue-generating products come into being.
MES provides a strong foundation around which manufacturers can build the I4.0 application. Instead of separate systems at each stage of production and supply chain, MES can flow vertically and horizontally in the process.

As for connectivity, MES will require different apps, which will later be able to control equipment and eventually open the doors to augmented reality scenarios. MES of the future must also leverage cloud computing and advanced analytics. MES 4.0 will be a completely new generation of systems, which must be able to cope with all these unaccustomed challenges and shall allow companies that adopt it to gain a solid, yet flexible,

infrastructure for the big and long transformation that I4.0 actually is. Figure 2.6 show which are the principal tasks of MES.

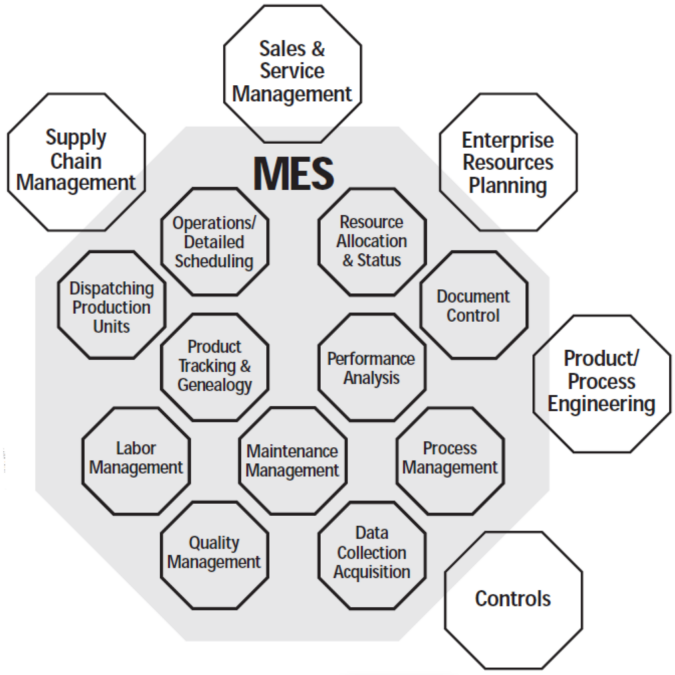


Figure 2.6: MES Functional Model [50]

At the state of art, MES is and is going to be the base for data collection and integration in manufacturing systems. Because of it, the current research of knowledge extraction focuses on data coming from MES.

3 State of Art

3.1. Event Logs in Production Systems

As described in Section 2.3, MES is the most important source of data companies have in manufacturing systems. It provides up-to-minute mission-critical information about production activities across the shop floor and supply chain via communications networks, such as local area networks. MES accomplishes this task by guiding, initiating, responding to, and reporting on plant activities in real time, by using current and accurate data [99].

Each MES system has its own structure, because the characteristics of manufacturing process drive the MES implementation process. Basically, each solution is customized and requires ad hoc implementation: MES has to be designed for the plant-specific conditions. This represents a problem since MES deployment team has the flexibility to implement custom solution, taking the risk, on the one hand, of mismapping the processes in a wrong way, and, on the other hand, of missing the requirements due to the lack of structured methods. Starting from this condition, it is important to highlight that the owner of MES system needs to know in advance which are the analysis that need to be performed with data coming from the system. The phase of needs identification is often underestimated during MES implementation due to the fact that knowledge related to those technologies is still not spread among enterprises. Summarizing, there are different factors that drive the MES data structure:

1. Final usage of data: information needs of all operational layers. All information needed to correctly model the desired application are collected and analyzed.

2. Availability of infrastructure.

After having a clear understanding of those requirements, it is possible to implement a conceptual design by means of the model structure of the process. In academic literature, there is a flexibility on the selection of the type of model to use, but the preferred one is the Business Process Modeling Notation (BPMN) diagram [33]. This notation makes setting up the association between the process and databases that is going to collect data. The Figure 3.1 gives an example of BPMN diagram of a manufacturing company.

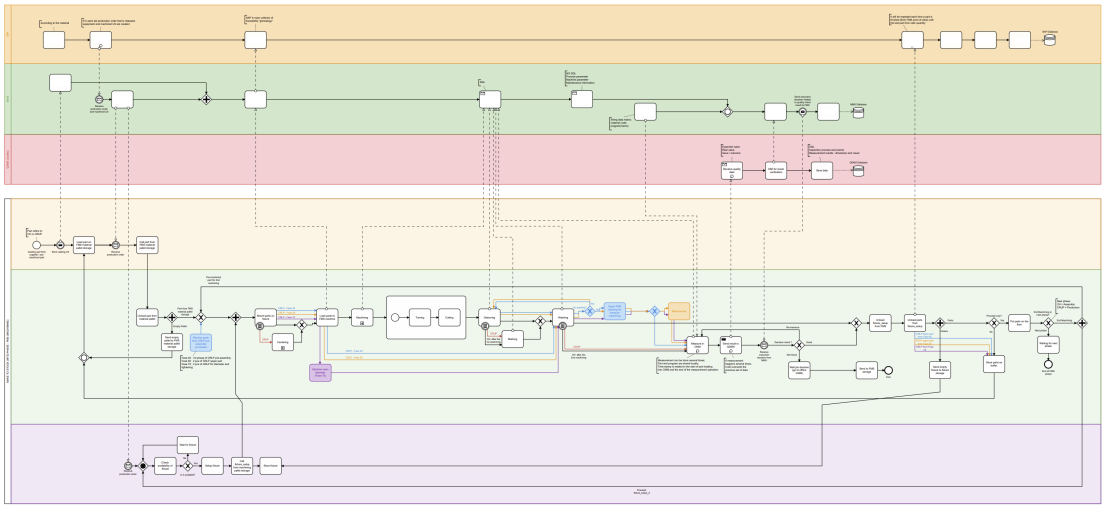


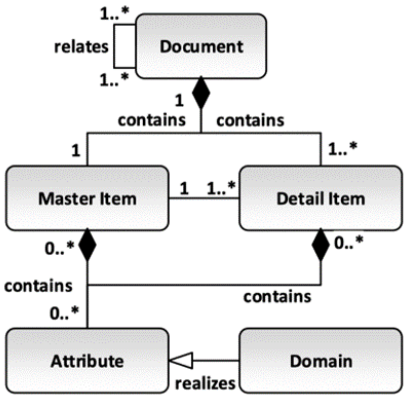
Figure 3.1: BPMN with ERP, MES and quality layer.

After having set the analytics to be performed with the data, as well as having collected the information about the infrastructure and the final data storage from BPMN, it is possible to build the conceptual model, namely the Entity-Relationship (ER) model, which will be used to describe the conceptual schema.

At this point of the review, it is necessary to introduce how relational database for information systems works. Systems like ERP and MES are Object-Centric, this means that they generate and store data in an object-centric manner, i.e. transactions update a database (often relational) storing information about objects (that are instances of a class). Those data consist of record stored in different tables (objects) that are related by cardinalities. The logical design is usually built by UML class diagram. In Figure 3.2 an

example of relationship between different objects is shown. Each relation is named and it has its own cardinality, defined as follow:

- 1:1 An object in the first class can only be related to one object from the second class, and vice versa.
- 1:n An object in the first class can be related to several objects in the second, but objects from the second class can only be related to one object from the first.
- m:n An object in the first class can be related to many objects of the second class, and vice versa.



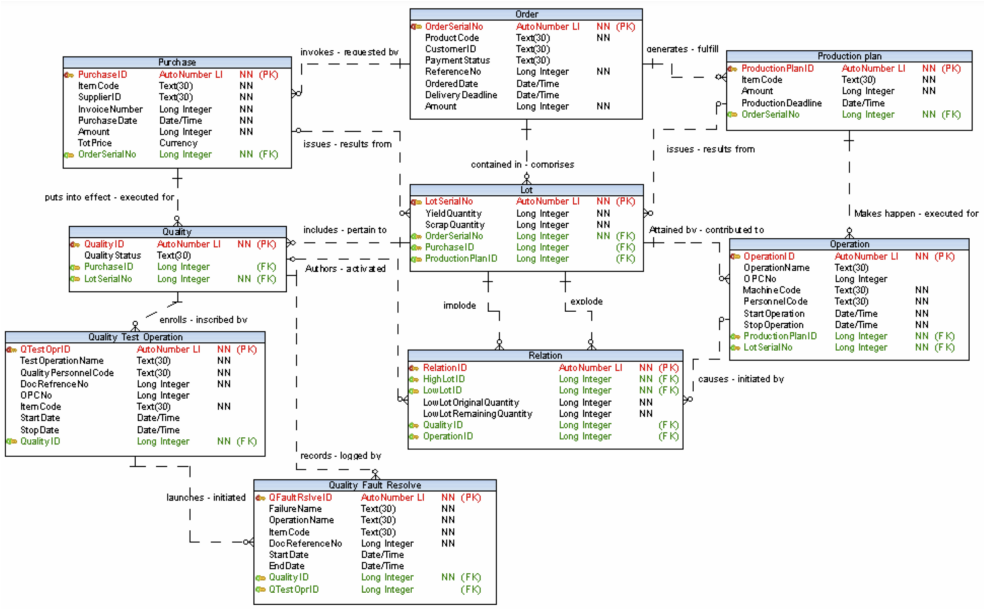


Figure 3.2: Example of UML class diagram of a data model [45].

Then, all MES data are stored in tables like the one in Figure 3.2. In order to be used for enhancement purposes, an extraction is needed. What is crucial to mention now is that it is determinant to know which are the relations between objects, not only the data

contained in the class. In order to take into account this information, analysts model relations through UML diagram and they use this knowledge to develop query in SQL language for data extraction.

However, this is a old-fashioned way to perform the extraction. Indeed, in the case there are changes in the connection between classes, analysts need to modify the query in SQL every time and keep UML up to date. In addition, it is not trivial to extract at posteriori the relations between different objects. In other words, the data perspective can be described, but the more powerful information (e.g., cardinality constraints) used in Entity-Relationship (ER) models [20], Object-Role Models (ORM) [84] or UML class models [32] are not employed at all. As a result, data and control-flow need to be described in separate diagrams.

This type of extraction, in jargon, analysts say that is “flattening” data, meaning that it is not exporting many-to-many relations or multiple case notions. A flattened event log is considered a particular view on the whole data set and it ruins the completeness of the original data set.

With the aim of performing process mining practices, the current state of the art on the extraction of logs in information systems (MES, ERP) proposes two different approaches. The first one is very common due to its simplicity, it is analysed in Chapter 3.1.1. The second one may be the future state for these kinds of extractions, it is analysed in Chapter 3.1.2.

3.1.1. XES

The statement "garbage in, garbage out" is always very powerful, in this context the sentence gives us the understanding of the importance to extract good data from MES system, that will be the bases for further analytics. The easiest way to extract data from databases is to perform queries through SQL language and export data in .CSV format. Usually this format allow data to be read by all software, and contain on the row the event,

and on the column all needed information related to the event (case identifier, activity name, timestamp and optional attributes like resource or cost). Such extraction is like a particular view on the whole data set and it ruins the completeness of the original data set. Those logs are called "flattened event log" and are needed in order to eliminate problem of convergence (one event is related to multiple cases) and divergence (independent, repeated executions of a group of activities within a single case) that we can have if we analyze data that have many-to-many relationships or multiple case notions [94]. This may lead to the replication of events and thus misleading results (e.g., duplicated events are counted twice). It may also lead to loops in process models which are not really loops (but concurrency at the sub-instance level).
This problem is a plague for analyst that do process mining techniques (especially for business application), but fortunately industry use Unique Identifier Item (UII) that can completely avoid the generation of this problem. Even if a product is produced multiple time, each part can be tracked as unique thanks to the UII tag.

After the extraction of .CSV, can be possible to perform process mining technique, but this would require high computational analysis. Because of this historically different format have been developed for this scope, the most effective is the .XES file. The XES log format which stands for eXtensible Event Stream (www.xes-standard.org) is the an exchange format for discovery techniques. In general, an XES log consists of a collection of traces. A trace describes the life-cycle of a particular case (i.e., a process instance) in terms of the activities executed. The most used software for converting file from .CSV to .XES is ProM [97].

3.1.2. OCBC

For the sake of completeness, is necessary to discuss how data analytics performed on extraction from other information system, like ERP, mostly fail to deal with one-to-many and many-to-many relationships between those data objects. To solve these issues, object-

centric approaches become promising, where objects are the central notion, and one event may refer to multiple objects. In particular, along this direction, the Object-Centric Event Logs (OCEL) standard [4] has been proposed recently. Basically, the metamodel of OCEL format capture all the information about events and objects involved in the events [2].

The meta-model for the specification of OCEL is shown in Figure 3.3 as a UML class diagram. The log, event, and object classes define the high-level structure of logs. The description for each class is the following:

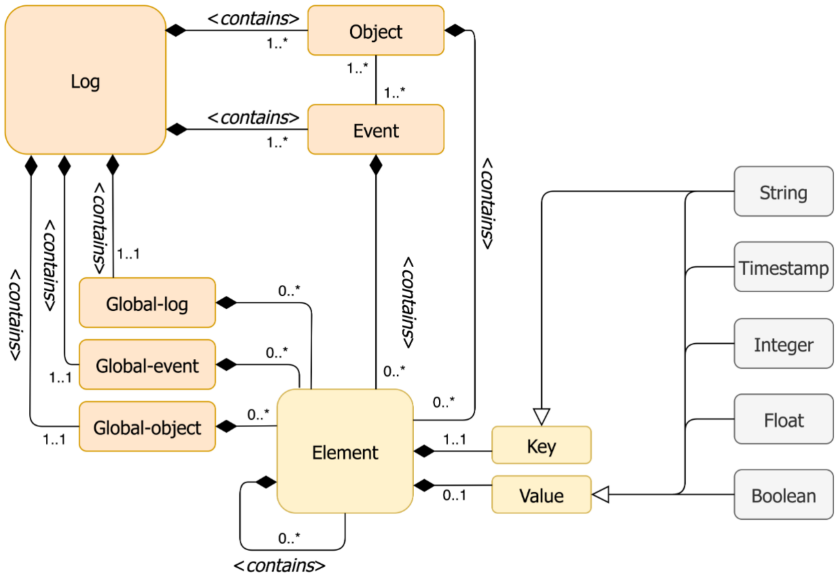


Figure 3.3: Example of UML class diagram of OCEL [4]

- Log: The log class contains sets of events and objects. A log contains global elements such as global log, global event, and global object elements. First, a global log element contains the version, attribute names, and object types that compose the log. Second, a global event element specifies some default values for the elements of events. Finally, a global object element specifies some default values for the elements of objects.

- Event: An event represents an execution record of an underlying process. It associates multiple elements (e.g., an identifier, an activity, a timestamp, and relevant objects) and possibly optional features.

- Object: An object indicates the information of an object instance in the process. It contains required (e.g., type) and optional (e.g., color and size) elements.

OCBC models are appealing because they faithfully describe the relationship between behavior and data and are able to capture all information in a single integrated diagram. However, OCBC models tend to be too complex and the corresponding discovery and conformance checking techniques are not very scalable. The best way to avoid noise during process mining processes is still to use a case identifier.

3.2. Petri Net

Petri Nets (PNs) are both graphical and mathematical techniques of modeling and analysis of process information [38]. From the graphical point of view, they are used in the design of the systems allowing their simple visualization. Instead, mathematically, they allow the drafting of equations that capture the behavior of systems.

Petri Net are based on nodes and arcs and tokens [70]. Nodes are divided into two types: places, represented as circles, and transitions, represented by rectangles, as shown in Figure 3.4. Arcs are the representations of relations that can be from places to transition and from transition to places, as shown in Figure 3.5. Tokens, that are considered the fundamental actors of Petri Net, are represented by black dots; in a Petri Net system they define objects such as resources, people and parts.

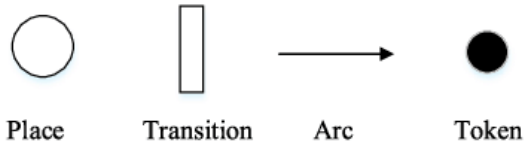


Figure 3.4: Elements of a Petri Nets.

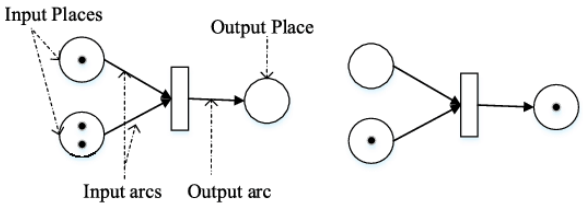


Figure 3.5: Example of simple Petri Nets.

Historically, Petri Nets have been first developed by Carl Adam Petri in 1962 and their

application spanned many fields. The main fields of use are distributed software systems [66] [106] [7] [18], distributed database systems [34] [63] [82], concurrent and parallel programs [103] [54] [75], industrial control or flexible manufacturing systems [22] [28] [108], logic inference [6] [56] and decision models [85] [64]. In addition, performance evaluation and communication protocols are considered successful application areas.
There are several reasons, in addition to their simplicity of representation, for which Petri Nets have spread to so many areas. The main advantages of Petri Nets are [31]:

1. They allow the representation of nondeterminism, so they are considered a sequence of discrete events whose order is one of many possible: in Petri Nets there is no consciousness of the flow of time, but this dimension is controlled through the sequence of events translated into a non-interleaving partial-order relation.

2. The same paradigm is used to represent both the system and its properties, so it provides a representation of the system's independence and dependencies.

3. Petri Nets allow the modeling of a system in a hierarchical manner, so a system can be modeled in different levels of depth or abstraction without the need to change the modeling formalism (rules according to which a model is built).

On the other hand, when modeling very complex systems Petri Net has shown some weaknesses [31], such as:

1. Tokens are of one type and they represent either information or the flow of control: there is no simultaneous representation of various players in a system, such as information, parts, resources, etc.

2. Likewise, only one type of place exists.

3. They do not allow the possibility of restricting the flow of tokens within the network: an oversized schema is usually needed to represent complex precedence constraints and conditions among system processes.

All these aspects, over time, have led to focus on the search for new methods that allow to increase the modeling power of Petri Nets aiming to reduce their size (complex systems involve large models). Given the large size of Petri Net models for industrial and business applications, an extended version of the classical Petri Net model have been developed: the High-level Petri Nets (HPNs) [87].

There are different type of HPNs that have been developed to model and analyze in a better way a variety of systems in application domains, ranging from logistics to office automation. All those different version of Petri Net are described in the following.

In many systems there is the needing to model the same process multiple time in the same net, i.e. by having the same subnet multiple time, this has led to the extension of individual tokens, that allow all identical components of a system to be modeled only once. Examples of this extension are the *Predicate/Transition nets* (PrT-nets) [30] and the most well-known *Colored Petri Nets* (CPNs). In particular this last model associates to the tokens, or to the places or to the transition itself, a colour that allows the distinction in classes of the elements. In this manner the colour can describe the properties of the object modelled by means of the token. For example, Drakaki et al. in [25] proposed a method based on Colored Petri Nets to model inventory management in a multi-stage serial supply chain, under normal operating conditions and in the presence of disruptions, for both traditional and information-sharing configurations. In this case, the various colours of tokens correspond to different types of supply chain elements such as product, material, information and financial elements.

Subsequently other methods have focused on the opportunity to have a modified semantics of arcs, places and transitions. This allows to increase the descriptive power of modeling with Petri Nets to cover more areas of interest. However, this type of extension does not affect the computational power of the Petri Nets, but only the compactness. The downside is the lack of direct analysis techniques.

Another type of High-level Petri Nets is the *Hierarchical High-level Petri Nets* (HHPNs)

[13]. The introduction of this category is dictated by the need to have a clear visualization and separation of all the components and the parts of a system, in addition to the need to facilitate the formalization process and the possibility of having an instrument with high repeatability. HHPNs allow the connection between components by joining transitions, places and arcs facilitating process analysis while retaining properties, but tightly mates the net components.

Finally, *Fuzzy Petri Nets* [15] are a specific application of High-level Petri Nets to represent uncertainty in operations and to approximate conditions in different areas such as robotics and flexible manufacturing. Train schedules are classic examples of uncertain information, and in fact, as explained in [67], Fuzzy Petri Nets can be used to simulate train traffic and thus estimate train delays.

3.3. Process Mining

As discussed in the Section 2, the digital universe is exploded and, over the years, the digital universe and the real world are aligning. This alignment makes the recording and analysis of events possible.
Therefore, the current challenge is to harness the data efficiently and effectively: identifying bottlenecks, predicting problems, and providing suggestions are some examples in which the data can be integrated with the real world.
The scope of Process Mining techniques is the use of event logs in a significant manner.

The starting point of Process Mining is the event logs. As explained in Section 3.1, data are collected in any sectors, from education to finance, in manufacturing and even in the healthcare sector [77] [12] [65] [73]. Starting from logs, Process Mining techniques allow processes to be discovered, monitored and improved by extracting relevant knowledge. These techniques allow managers to understand whether their processes are in line with expectations and, if not, they provide all the information in an orderly and comprehensive

visualization to figure out where and how to act to improve processes. The Institute of Electrical and Electronics Engineers (IEEE), in 2011 published the Process Mining Manifesto [89], that promotes Process Mining techniques in companies to study and redesign business operations.

The increasing interest in Process Mining is due both to the large amount of available logs and to the need to be increasingly competitive on the marketplace and thus to continuously improve business processes. It is applied in different fields and the distribution of publications by field of interest is shown in Figure 3.6.

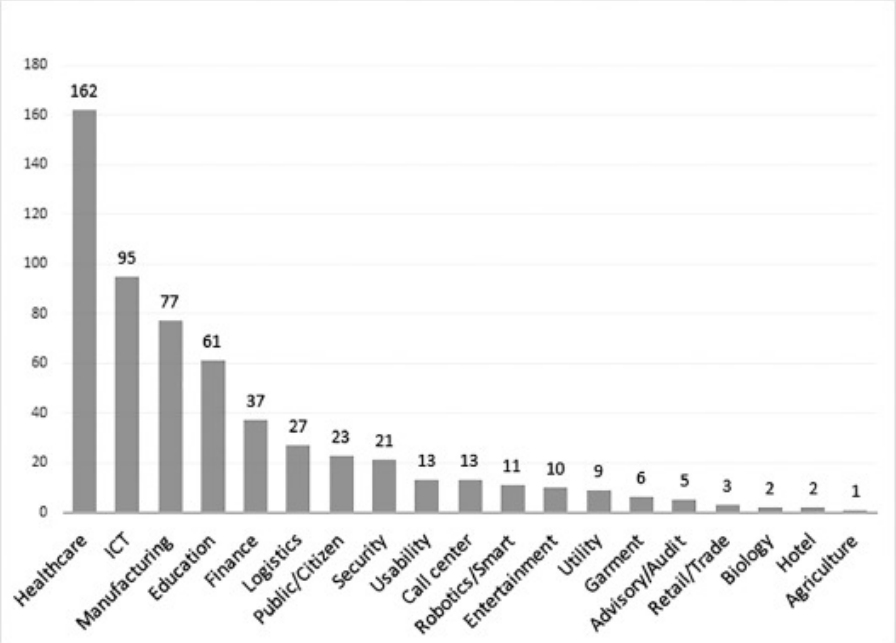


Figure 3.6: Number of papers on Process Mining by application domain [24].

As can be seen from the figure, the articles in the literature do not uniformly cover all the fields of application of Process Mining. In particular, one of the most studied sectors is the healthcare sector, that includes hospitals and clinical pathways. The second most studied sector, far behind the first one in terms of number of papers published, is the IT sector, with a focus on software development and on the study of maintenance. The literature also focuses on industrial and manufacturing cases, as well as education and finance. A general overview of the Process Mining applications in the various sectors is

explained as follows:

1. Healtcare: this area of interest covers the treatment of patients, the primary processes of an hospital and the clinical pathway of patients. Many studies point out that the characteristics of the models in this field are very different from the ones used in other fields of interest: healtcare is characterised by high security and private information, as well as the multidisciplinary nature of the examinations [61] [69] [76], the treatments required to people's care and the high variability due to several reasons such as diseases and treatments performed or biological interactions [74]. Process Mining is one of the techniques whose results show the greatest benefits in this field [72] [37]; in fact, numerous studies are focused on discovering processes to be compared with clinical guidelines to identify possible improvements.

2. ICT: this area of interest covers software development, IT operational services and telecommunication companies. Failures and change management was studied for compliance with the IT Infrastructure Library (ITIL) [8]. Another case of application is the reduction of patterns by improving the comprehensibility of models through a trace clustering approach [42]. Process Mining is also able to improve the degree of maturity of a software [52].

3. Manufacturing: this area of interest covers all industrial activities and the sector of most interest is automotive sector. Because of the increasing presence of information systems, such as ERP and MES, several papers are focused on the use of Process Mining in the manufacturing field. The first studies were conducted by Ho and Lau [36], aimed to improve a discrete production process to suggest enhancements from a huge amount of digital data. This approach led to greater flexibility and greater ability to aid decision-making. Later on, the focus shifted from administrative processes to industrial equipment [80]. Finally, the attention switched to industrial maintenance. In particular, Process Mining is used to extract execution rules of maintenance activities such as time intervals [81].

4. Education: this area of interest covers all activities related to the improvement of educational processes such as the best learning path according to the student profile or the analysis of trends related to online learning or of student interactions. Therefore, it is possible to reconstruct the complete educational process by extracting this knowledge from educational information systems.

5. Financial: this area of interest covers activities related to banking, insurance related to the processes of investment, payment and transfer of money, analysis and risk reduction. A practical example is the study presented in [48] that simulates the management of insurance claims for more accurate forecasting.

Process Mining is a relatively new discipline that lies between Business Process Management (BPM) and Data Mining, focusing on process analysis using event logs.
Business Process Management, an evolution of Workflow Management (WM), is a structured and systematic approach of process analysis focusing on operations and management roles in addition to work management.
Instead, the main difference between Data Mining and Process Mining is the different focus. In particular, Data Mining stems from the need to discover patterns from a large amount of data and focuses on storage and processing; in contrast, other applications involving the time series of events is not sufficiently considered. So if Data Mining is data-based, Process Mining is flow-based and therefore process-based. So the goal of Process Mining is to study processes from data by bridging the gap between Business Process Management and Data Mining.

Process Mining is divided into three macro areas shown in Figure 3.7:

1. Process Discovery [95]: this area encompasses all those techniques that allow a model to be built from the logs without external influences.

2. Conformance Checking [95]: this area encompasses all those techniques for monitoring and comparing a model with the logs to find if nominal processes are confirmed

in practice identifying deviation from expected behaviour.

3. Model Enhancement: this area encompasses all those techniques that allow the model to be improved and extended. These techniques allow the manager to improve the existing process.

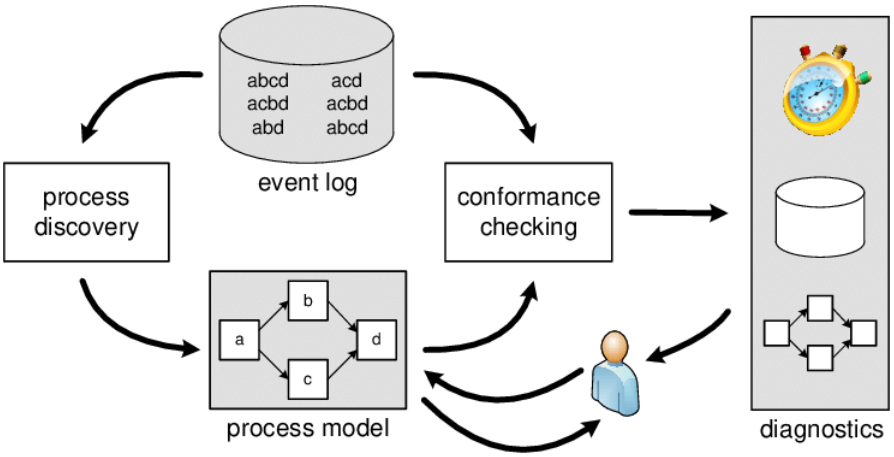


Figure 3.7: Positioning Process Mining techniques.

Therefore, Process Mining allows the identification of inefficiencies in a process, such as bottlenecks, that leads to greater innovation and quality as well as cost reduction and process efficiency increment. The main challenges Process Mining has to face with are listed in the following.

1. Data Quality [102]: the application of Process Mining requires integration and cleansing of data that are distributed across various data sources.

2. Concept Drift [83]: processes are dynamics, so they can change during analysis.

In the following, Section 3.3.1, Section 3.3.2 and Section 3.3.3, the three areas in which Process Mining is divided will be explained in details with the aim of answering to some questions. What is their purpose? What benefits do they bring to the process? Which algorithms or software are used for their implementation?

3.3.1. Process Discovery

The main purpose of Process Discovery is to build a model from data, i.e. event logs. As explained in Section 3.1, event logs are constituted by various attributes such as resource, timestamp, parts, activity, etc. These information are the starting point for Process Discovery techniques to build a process model based on common behaviors in the data. Therefore, the challenge of Process Discovery is to discover common behaviors and translate them into a model.

The logs, however, contain infrequent behaviors, patterns or traces that are much more subtle than the main and common behavior, so the problem that arises is whether or not these secondary behaviors should be included in the construction of the model [60].

There are many methods of representation (models) to describe a process depending on the techniques used for discovery. The best known representations are:

1. Directly-Follows Graph (DFGs): in this representation each node represents an activity and each arc represents a relationship between several activities [92].

2. Business Process Model and Notation (BPMN): this representation allows the construction of compact models. In addition, BPMN is appealing to both process mining and business users since it can be simply integrated into a BPMN diagram by combining control flow perspective with data perspective [43].

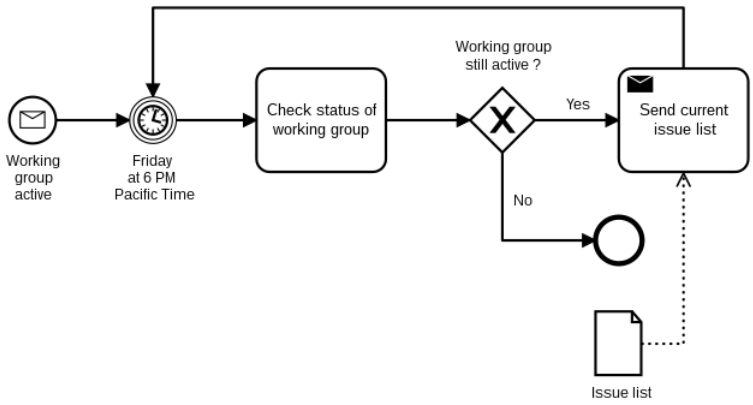


Figure 3.8: Example of a Business Process Model and Notation for a process with a normal flow.

3. Petri Nets: this methodology is described in detail in Section 3.2 and it allows a higher level representation by showing different types of transformations between activities.

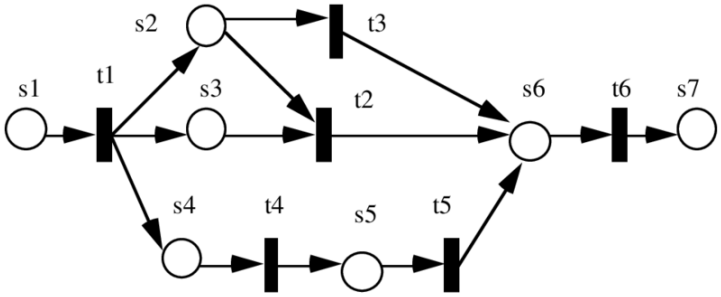


Figure 3.9: An Example of Petri Nets Graph.

Process Discovery aims to shed light on control-flow of a process and many techniques and algorithms have been developed and proposed in the literature, such as:

1. Alpha Algorithm [90]: this algorithm allows the construction of a Petri Net representing the input event logs. It is not a mining technique but it provides a good introduction to the topic and it can be used for Process Discovery.

2. Heuristic Mining [100]: it is similar to the representation techniques called Causal

Nets, which take into account the frequency of events and their sequence. According to the Causal Nets, infrequent events should not be considered in model construction.

3. Genetic Process Mining [62]: this approach mimics the natural evolution process of biological systems. Genetic Algorithms test possible solutions in the search space by combining them through the mutation process. It is not a deterministic approach and depends on randomization.

4. Inductive Mining [51]: this approach is considered the most used Process Discovery technique due to its flexibility and scalability, in fact it allows different variations from the basic approach.

One of the most important challenges that arises from Process Discovery is that the model discovered through Process Discovery techniques described in this section does not coincide with the real process and usually it is more complex. This aspect is of paramount importance from the perspective of a manufacturing systems since it will not allow a quick and logical transition between the digital model and the actual process.

3.3.2. Conformance Checking

As explained in Section 3.3, Process Mining is that set of techniques that links Data Mining and BPMN and deals with the discovery, monitoring, and improvement of processes. This Section will explain the techniques that enable Conformance Checking, one of the three areas into which Process Mining is divided.

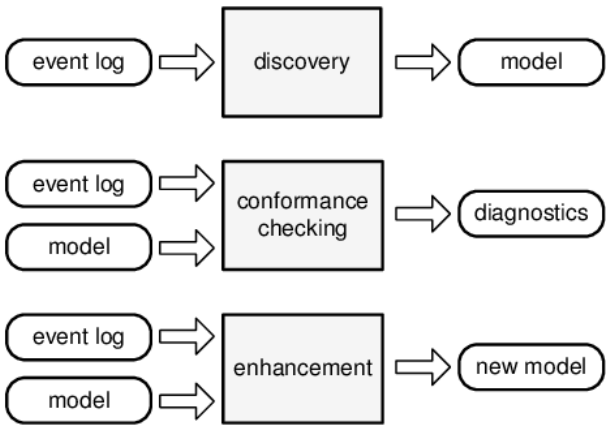


Figure 3.10: The three basic types of process mining explained in terms of input and output: Discovery, Conformance Checking and Enhancement.

Conformance Checking compares the model of a process with its event logs. As input it needs the event logs and the process model, as showed in Figure 3.10, usually constructed through Process Discovery techniques explained in Section 3.3.1. The final output will be diagnostic information that seeks to capture similarities and deviations, i.e. differences between the model and the data.

The main Conformance Checking techniques are based on process models with graphical representations and in particular Petri Net. Important relevance has the digital data, in fact there are minimum requirements for log content without which Conformance Checking cannot operate, this is due to the fact that the logs have to match the process models. The minimum requirements are the so-called ID, that is a unique identifier, a label and a timestamp (date and time) [39]. In addition, the notation with which the process model input is constructed also important since it must follow the specifications of a particular modeling language such as Unified Modelling Language (UML) activity diagrams, Business Process Model and Notation (BPMN) and Petri nets [26].
Once all the characteristics of the input elements have been determined, the algorithm for comparing model and event logs must be chosen or developed. In particular, there are two approaches in the literature:

1. Log Replay [16]: these algorithms interpret logs and then retrace each trace, event by event, on the model. An example of log replay algorithm is the token-based log replay [78]: each time the model reaches a deadlock, a token is generated; model compliance is determined based on the sum of the redundant and generated tokens.

2. Trace Alignment [88]: they also express deviations and similarities directly at the event level. In particular, it is possible to obtain various information about log violations from the model but also about the occurrences of the events in question.

To compare the models with their respective logs, the literature explains four different dimensions/metrics of quality: Fitness, Simplicity, Precision and Generalization.
In the numbered list below they are explained with the help of Figure 3.11.

In particular, Figure 3.11 shows in the left side four models (M1, M2, M3, M4) constructed through Process Discovery techniques. The models are very different from each other since they are defined with different algorithms or input parameters. On the other side, in the right part it shows an event log L with 1,391 traces divided as in the graph.
Let's explain in details the four quality dimension:

1. Fitness: the fitness value ranges from 0 to 1, extremes included. It indicates how many traces are repeatable in the process model. The fitness values computed on the models shown in Figure 3.11 are as follows: fitness(L, M1)=1, fitness(L, M2)=0.8, fitness(L, M3)=1 and fitness(L, M4)=1.

   However the fitness quality dimension individually does not help to understand the Conformance Checking in each levels, in fact as can be seen with closer analysis, the value of fitness of model M2 is 80% although only 33% of the logs can be repeated from start to finish. Such a high value is explained by the fact that almost all event logs start and end for the right activity captured by the model (activity a as starting point and activity h as end point).

2. Simplicity: this quality dimension is closely related to Occam's razor philosophy:

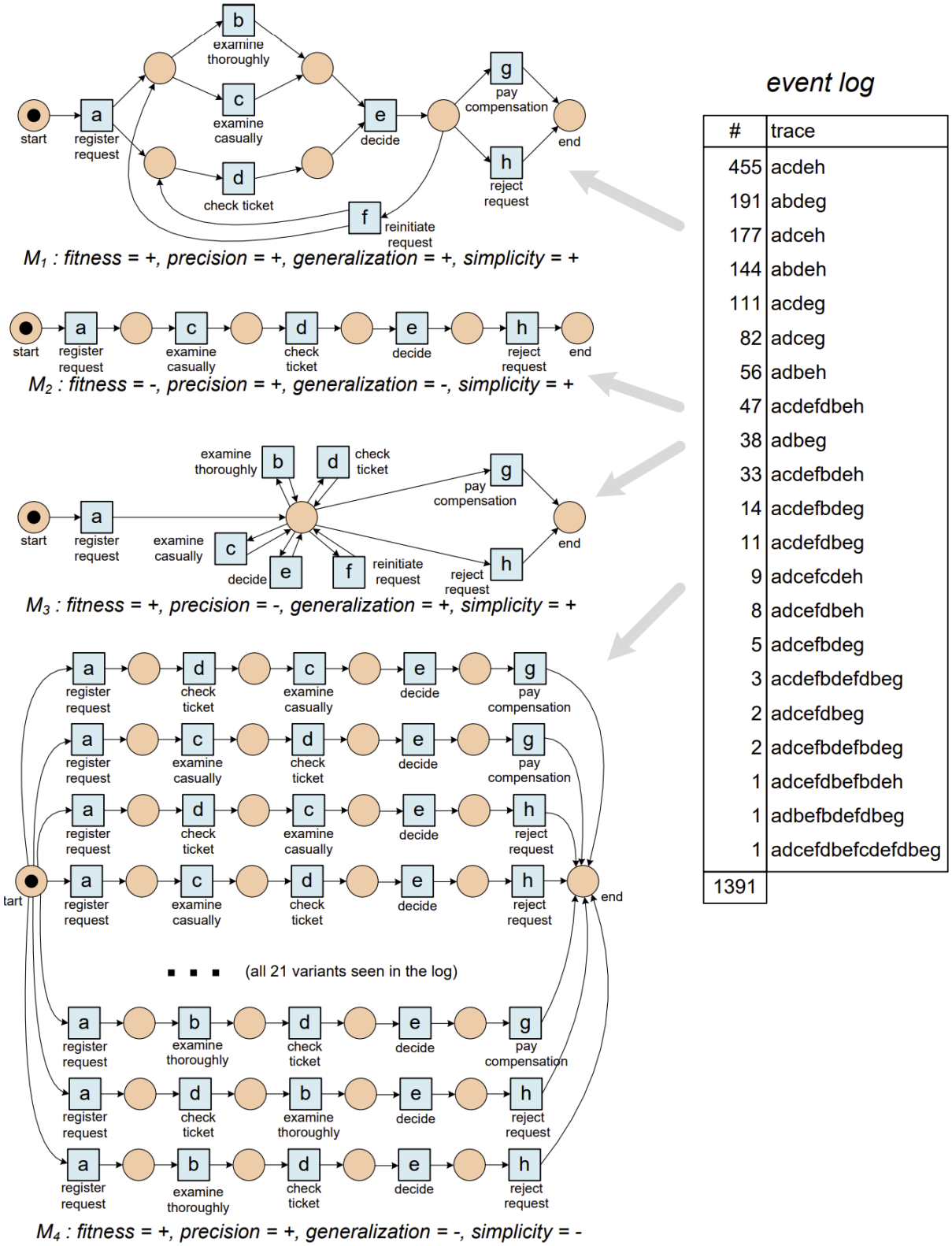


Figure 3.11: Four model process (M1, M2, M3, M4) and an event log L.

"it is not necessary to increase the number of entities needed to explain something.". This leads to choosing the simplest model that explains event logs as the best model. Thus, model M4 turns out to be the least suitable because the first model (M1) is visually simpler.

Model complexity can be studied in several ways: from simpler methodologies, such as the number of nodes and arcs, to more sophisticated methodologies that take into account the "structural" complexity of the model or "entropy" of the model.

3. Precision: as explained above, the fitness value alone does not help to understand Conformance Checking at each levels, so, to answer the question "how well does the model represent the behavior of the event logs?", the precision metric is introduced. Like the fitness value, this metric is also a value between 0 and 1 inclusive extremes (if all behaviors allowed by the model are actually observed, then precision(L, M)=1) and takes into account what is superfluous (activities and connections).

   A practical example is as follows: the fitness value of model M3 is 100%, but it represents very different behavior and relationships from the intrinsic logs. In fact, the accuracy value of model M3 is 41% showing how this model is not suitable to explain logs since it has "under-fitting" problem. The "under-fitting" is intrinsically related to models such as the third that overgeneralize a behavior.

   The precision value of the other models is: precision(L,M1)=0.97, precision(L,M2)=1 and precision(L,M4)=1.

4. Generalization: the generalization dimension makes it possible to explain whether a process model merely shows the observable behavior of digital data or represents it generically. As can be seen from Figure 3.11, model M4 does not generalize, limiting itself to encoding every possible trace in the event log L.

   Generalization helps to understand and identify the problem of "over-fitting", in fact this problem is related to the construction of a very specific model, which aims to explain the particular sample under consideration (event logs L in case of Figure

3.11) by making the process model lose generality.

These four dimensions described above are not all used in the same way. The fitness metric is the most widely used; in fact, it indicates "how well an observed behaviour fits the defined process modell". In contrast, the other three metrics explained above are mostly used to study the quality of the process model constructed through discovery.

3.3.3. Model Enhancement

The third area of Process Mining is Model Enhancement.
This area of Process Mining depends on the two first defined in Section 3.3.1 and in Section 3.3.2 and it is inserted at a third time level: firstly, the process model is discovered through Process Discovery techniques; then the similarities and divergences between the observed and modeled behaviour are studied through Conformance Checking; finally, if the process model is not conform to reality, it is necessary to extend or correct, it means to enhance, the model itself to better capture the behaviour of the event logs.

The purpose of Model Enhancement is thus to improve an existing process model using the information contained in the logs also by means the diagnostics performed through Conformance Checking. For example, intrinsic information from log timestamps can be used to directly show bottlenecks.
Since improvement is greatly influenced by the business and process as well as the situation under consideration, the definition of Model Enhancement of processes is of crucial importance. "The extension or improvement of an existing process model using the actual process information recorded in an event log" [5] is the most frequently cited definition in the literature [104].
In addition, the dependence of improvements with respect to the process and the business leads to a lack of automated techniques, as opposed to Process Discovery and Conformance Checking techniques described in the aforementioned sections.

The literature referring to Model Enhancement focuses on two different aspects [104]:

process model re-design and process model repair. Regarding process model repair, the literature mainly focuses on control-flow perspective since it is the backbone of the model and it is of primary importance over other perspectives, such as work distribution.

The main reasons why model repair is used are three [29]:

1. The processes evolve over time, because worker competence improves and thus different real cases are handled differently, but also because formal or informal procedures change. Evolution of the processes leads models to become obsolete over time.

2. Improving the Conformance Checking, in particular the four quality metric: fitness, precision, simplicity and generalization.

3. Customize the initial model in order to describe more accurately the processes and to be more in line with the business type.

The concept of model repair is based on the identification of sub-processes needed to repair a model: sequences of non-reproducible events that can be traced back to the same location are grouped into a sublog; a subprocess is built that reproduces the sublog using a Process Discovery technique; finally, the subprocess is integrated at the point into the model where the deficiency was discovered.
Iterating this process results in a repaired model capable of describing the event log.
The method described is reported in [29] and the article shows how this technique requires less modification than an iterative use of Process Discovery adopted if Conformance Checking highlights non conformity between discovered model and event logs.

On the other hand, the extension of the discovered model aims to integrate new perspectives: if the first two steps, together with the repair Model Enhancement, focus on the control-flow perspective, the purpose of the extension is to discover knowledge in order to reflect on possible improvements to the process itself [35].

The different perspectives that are studied in literature are listed below:

1. Organizational Perspective [14]: it focuses on the resource information hidden in the event log. This perspective aims to create a classified structure of the people involved in the process by role through a Social Network Analysis.

2. Temporal perspective [9]: it focuses on the timing and frequency of events. This perspective allows, through timestamps, to study bottlenecks and also monitor resource utilization, for example the saturation of machinery.

3. Case perspective [27]: this focuses on the properties of cases (process instances) such as the path or resources involved.

Important now is to outline what tools are used for model enhancement. The most commonly used tools are ProM framework, for most parts of the case studies, and Disco. To a lesser extent, other tools used are SQL, DpiL Miner, and Weka.

3.4. What-if Analysis

Process Mining is a powerful tool that allows, through the analysis of historical data, to map production processes with the aim of improving them. Process Mining is thus backward-oriented.
Production processes are subject to frequent changes, such as a sudden increase in demand or delays from suppliers. In addition, more than any other processes, they are also subject to internal changes such as the replacement of a machinery with a better performing one, or re-configurations or changes in routing policies.

To deal with these changes, the ability to make predictions with analytical tools is a great advantage in taking important decisions. For this reason, managers need a tool that looks forward [93] answering questions such as "What if..?". The ability to look forward has to incorporate the knowledge gained about the process by Process Mining in order to enable What-if Analysis confident and close to reality.

This section explains some state-of-the-art techniques for performing forward-looking

analyses. The key publications concerning simulation are typically focused either on statistical aspects [46] or on a specific simulation language, such as Arena [44].
In this chapter, two different approaches will be considered: Discret Event Simulation and the Queuing Network. After an initial overview of both of them, one typology of Queuing Network, named Jackson Network, will be explained specifically.

Discrete Event simulation (DES) is one of the best known techniques for simulating manufacturing processes [91]. It generates events based on rules defined by the simulation model. Events occur at a fixed instant of time and allow a change of state of the system; new states allow the generation of new events. A simulation describes one of the several ways in which the model can be reproduced.
By means of DES key performance indicators (KPIs), such as the waiting times of parts in the model, can be computed. These KPIs will support managers to make decisions about the process. The main limitation of DES is the huge amount of time required to achieve a good simulation model in addition to the fact that the interpretation of the results is not easy to interpret. In addition, the level of detail required by this type of simulation does not allow its use for long-term forecasting.

For this reason, despite the great ability of some simulations to capture the behavior of production systems, the application of detailed simulations such as DES in real life is limited. As can be seen in Figure 3.12, there are other approaches to model manufacturing systems that lead to more or less realistic results depending on the initial level of abstraction. The most widely used models are Building Block and Decomposition Model [58], and the Queuing Network Model.

In the following we will focus on the Queuing Network Model. It is an evaluative model, meaning that, unlike generative models, it does not provide to the user an "optimal solution", but it evaluates a given set of decisions by providing performance measures. The study of Queuing Networks was pioneered by applications in the telephone industry [11], but Queuing Networks owe their fame to two crucial works conducted by Jackson

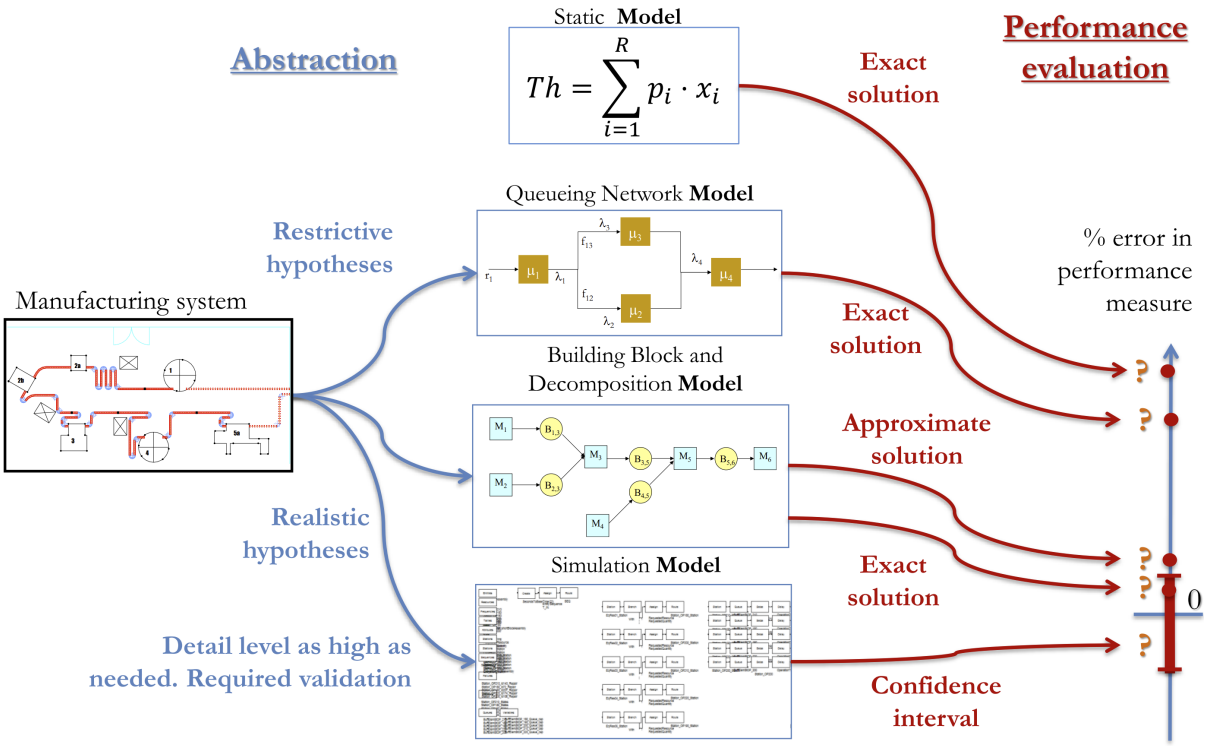


Figure 3.12: Different approaches to model manufacturing systems [86].

[40], [41]. Subsequently, Queueing Networks depopulated in many areas such as computer science, telecommunications, and FMS.

Queuing networks are divided into two macroareas: Open Model and Closed Model. Open Networks are related to those systems where customers can freely enter and exit the system. In a Closed Network the total number of jobs in the network is a constant N. There are no external arrivals and no jobs ever leave the network. In manufacturing there are at least two practical situations where Closed Networks are required to model a system:

- In many automated systems each job has to be mounted on a pallet throughout its circulation within the system. Usually the number of pallets is a given constant, say N.
- If it is important to keep the Work In Progress in the system constant, it is possible to adapt a policy where as soon as a job completes all its processing requirements

and leaves the network, a new job is immediately released into the network. In this way the production rate can be maintained at a desirable constant level.

There is a possibility of mixing these two types of Queuing Networks to build a Semi-Open Queuing Network. A Semi-Open Network is an open network in which a maximum number of K jobs can be accommodated.

Unfortunately, the exact analysis for Open Queuing Networks with a finite number of servers was only possible for networks with the following characteristics [10]:

1. Exponential service time distributions.
2. Service requirements at each station are independent of the product family. If the service times are allowed to depend on the product family, then exact analysis is possible with a preemptive resume, last-come-first-served discipline.
3. Priority discipline at each queue is independent of the product family.
4. Arrival process to the network is a Poisson process.

The main reason why Queuing Network is widely used to simulate manufacturing systems are listed in the following.

1. It allows to take into consideration the interactions between the machines.
2. It is a computationally efficient model to test different alternative solutions.
3. It follows the system point of view and not the jobs related performance measures.
4. Throughput and Work In Progress can be used as performance measures.

An example of Queuing Network is reported in Figure 3.13.
First consider a network consisting of M nodes. Each node represents a machining center. In each workcenter parts are being processed or are waiting in a queue. The state of the system is completely defined when the number of parts in each node is known. To define the network, the routing matrix $[r_{ij}]$ should be introduced. Each element of this matrix

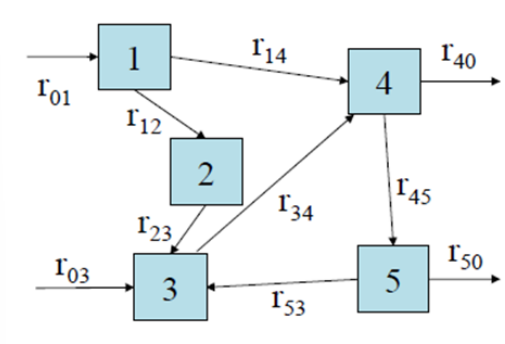


Figure 3.13: An example of Queuing Network.

represents the probability that a job leaving node i will go to node j. It can be interpreted as the proportion of jobs leaving node i that next visit node j.

The convention adopted is that node 0 represents everything outside the network.

This paper focuses on a particular class of Queuing Networks, called Jackson Networks, that is well suited to study of manufacturing systems. Jackson Network will be explained in more details in the Section 3.4.1.

3.4.1. Jackson Networks

Concerning the Jackson Network, it is necessary to introduce the traffic equations.
Let λ_i be the overall arrival rate at node i, including both external arrivals and internal transitions, the traffic equations can be written as:

$$\lambda_i = \alpha \cdot r_{0i} + \sum_{j=1}^{N} \lambda_j \cdot r_{ji} \tag{3.1}$$

being $i = 1, 2, .., M$.

It is also important to assume the processing times at each node to be exponentially distributed with a mean dependent on the number of jobs in the node. Also:

- Assuming that there are x_i jobs at node i, the processing rate is state dependent: $\mu_i(x_i)$.

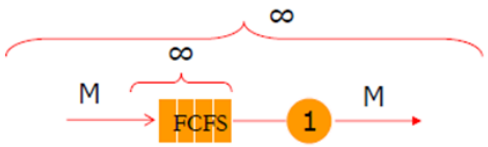


Figure 3.14: A scheme of a M/M/1 queue.

- It can be assumed that $\mu_i(0) = 0$ and $\mu_i(x_i) > 0$ for all $x_i > 0$, $(i = 1, 2, \ldots, M)$

The main hypothesis that have to be taken into account when a Jackson Network - Open Model is used, are:

1. In an open network, jobs arrive from outside following a Poisson process with rate $\alpha > 0$.

2. Each arrival is independently routed to node j with probability $r_{0j} > 0$ (equivalently this can be viewed as each node having an external Poisson stream of job arrivals with rate $\alpha \cdot r_{0j}$).

$$\sum_{j=1}^{M} r_{0j} = 1 \tag{3.2}$$

3. Upon service completion at node i, a job may leave the network with probability r_{0j}.

4. Service discipline at the queues is FIFO.

The Jackson Network is based on further mathematical consideration aimed to demonstrate that in equilibrium, the M nodes of the network are independent, each following the distribution of a birth-death queue. Birth-death processes are particular types of Markov Chain where the transitions can happen only among adjacent states. Being a birth dead queue, each machine of the line can be modeled as a M/M/1 queue model, Kendall's notation. The scheme of the M/M/1 queue is reported in the Figure 3.14.

M/M/1 means that arrival process (arrival rate, λ) and service process (service rate, μ) are exponentially distributed; there is only one server and one queue with infinite capacity;

the number of clients that can enter into the system is infinite; and finally the dispatching policy is FCFS.

The ratio between the arrival rate and the maximum rate (capacity) at which the system can perform its work is called utilization factor ρ. For a M/M/1 queue $\rho = \frac{\lambda}{\mu}$.

The possibility to model each machine of the system by means of a M/M/1 queue model, allows an easy calculation of the main performance indicators. If the saturation $\rho < 1$, the average number of customers in the system can be computed:

$$n = \frac{\rho}{1-\rho} \tag{3.3}$$

Moreover, the average number of customers in the server (n_s) and in the queue (n_q) can be computed as follows:

$$n_s = \rho \tag{3.4}$$

$$n_q = \frac{\rho^2}{1-\rho} \tag{3.5}$$

Finally, applying the Little's theorem it is possible to compute the Flow Time (T), the Service Time (T_s) and the Waiting Time (T_q).

$$\lambda T = n \longrightarrow T = \frac{1}{\mu - \lambda} \tag{3.6}$$

$$\lambda T_s = n_s \longrightarrow T_s = \frac{1}{\mu} \tag{3.7}$$

$$\lambda T_q = n_q \longrightarrow T_q = \frac{\rho}{\mu \cdot (1-\rho)} \tag{3.8}$$

4 | Models and Metodologies

Process Mining techniques are used in a multitude of fields, as shown in Figure 3.6. This work focuses on applying Process Mining techniques in the production system domain by integrating them in a new way.

Starting from logs, Process Mining techniques enable the discovery, monitoring, and improvement of processes in any field of use by extracting relevant information. Process Mining is divided into three areas: Process Discovery, Conformance Checking, and Model Enhancement. Most papers focus on the application to specific cases of individual areas of Process Mining; others explain the integration between two different areas. Few papers attempt to explain the methodology starting with Process Discovery and ending with Model Enhancement. Van der Aalst, a full professor at RWTH Aachen University, is one of the pioneers of Process Mining techniques, including the manufacturing domain application. Following a detailed study of the current literature regarding Process Mining techniques, the existing methodology was summarized and explained in detail in the next section (Section 4.1).

Process Discovery is a fundamental tool when the flow of instances or the possible implications of decisions or actions are unknown, which is the case in almost all business applications.
An example is the healthcare domain, where the implications of some clinical treatments are not known and depend on individual diseases or biological interactions of patients and, therefore, studying historical data allows the discovery of standard unknown clinical pathways. Also the insurance domain processes unknown information; in fact, the appli-

cation of Process Mining techniques allows risk minimization in dependence on previous decisions. The information technology and social networking or education domain also follow this logic line: the process to be discovered is not known a priori, and Process Mining techniques allow it to be precisely discovered and improved.

Manufacturing systems, however, do not follow the same rules as the aforementioned fields. In fact, Process Discovery may be considered unnecessary and redundant in manufacturing: this is the idea behind the methodology developed in this work whose aim is to enhance knowledge-based model through Conformance Checking techniques.

The reason why Process Discovery is redundant is justified by the fact that every company has structured and available knowledge of its production process. All information is collected and stored in different sources, such as ERP, MES, BPMN Diagram, senior workers, as well as technical knowledge about the technologies used. The production process appears to be already known a priori.

Therefore, it is no longer necessary to discover the process from the logs, since all the information needed to build the model are already present within the company and thus it would be "discovering something that is already known". Consequently, model building may not be based on logs, but it may be based on a translation of the abundant knowledge already available and structured into an appropriate model.

This chapter is divided as explained in the following. The first section, Section 4.1, explains the current method of integrating Process Mining techniques and it highlights the main technical problems of the existing methodology applied to manufacturing systems. Finally, the Section 4.2, will explain the novel methodology and how it responds to the problems mentioned in the previous section.

4.1. Existing Methodology

This section discusses the current methodology for integrating Process Mining techniques in different application domains. Furthermore, the main criticalities of this method applied to the manufacturing domain are highlighted.

Following an in-depth study of the current literature concerning Process Mining techniques, the following methodology can be summarized. The Figure 4.1 shows how the various areas of Process Mining interact with each other, starting from the data in order to study real processes and propose improvement solutions.

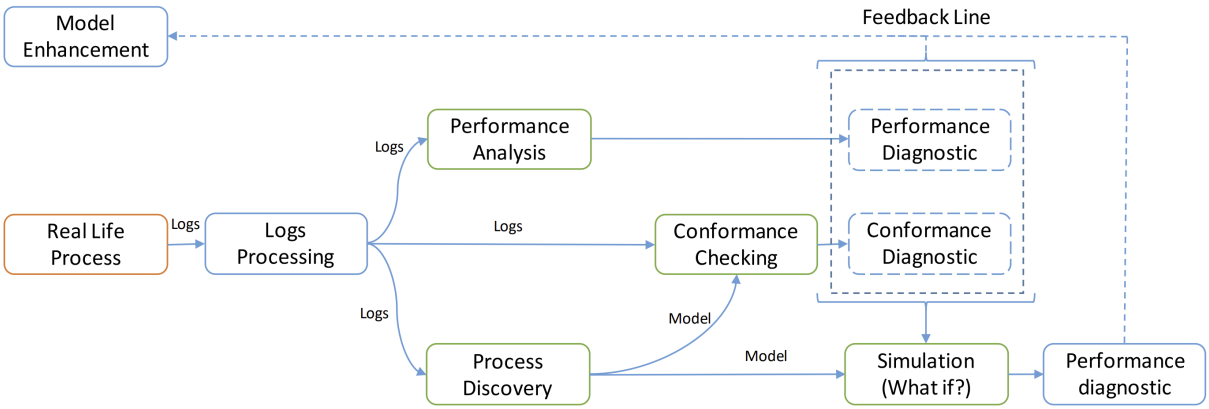


Figure 4.1: Existing Methodology integrating Process Mining techniques.

The first step in the existing methodology, after the acquisition of event logs, is Data Processing. Its purpose is to process the raw data in order to obtain a data-set containing the necessary information for Process Mining operations. The data thus obtained are used to calculate the main key performance indicators, such as OEE if applied to manufacturing domain.

The main Process Mining technique is *Process Discovery*. The purpose of Process Discovery is, starting from the logs and without additional information, to discover the process through the construction of a model, usually returned in the form of a Petri Net, Process Tree Diagram or BPMN Diagram. Numerous tools and algorithms are available to

discover the model from event logs. The choice of the algorithm is based on the field of application and on the type of the collected data. Once the algorithm has been chosen, it is necessary to set a number of parameters, such as the frequency of events to be considered, their dependency, how much importance to give to loops, the noise threshold to be considered in the data and many others. The choice of initial parameters greatly influences the model discovered by means of Process Discovery techniques. An example are the Petri Nets shown in Figure 4.2 and Figure 4.3. By taking a simple data-set processed with the Inductive Miner algorithm and changing only the value of the initial parameter *"noise threshold"*, it is possible to obtain very different models.

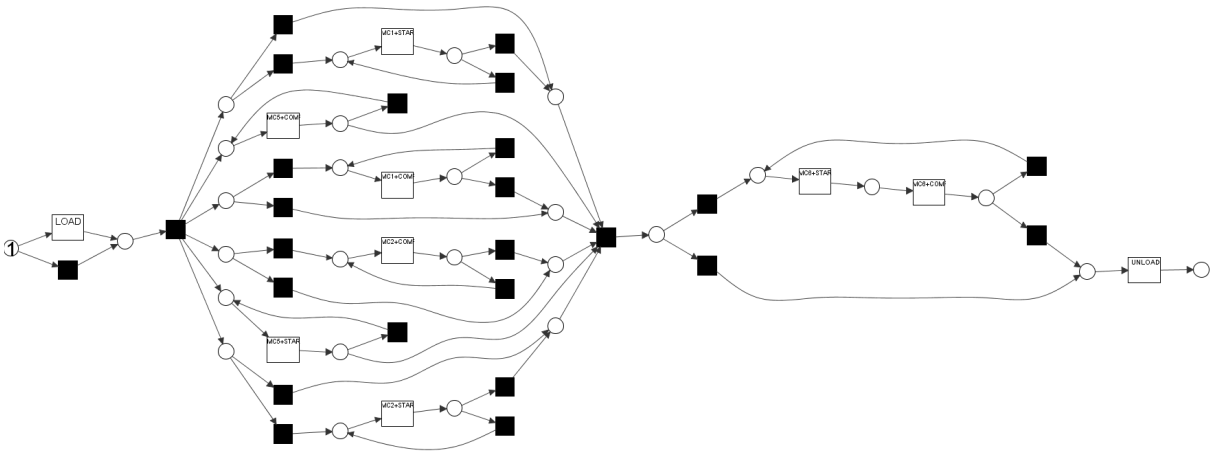


Figure 4.2: Petri Net mined by means of Inductive Miner with noise threshold equal to 10%.

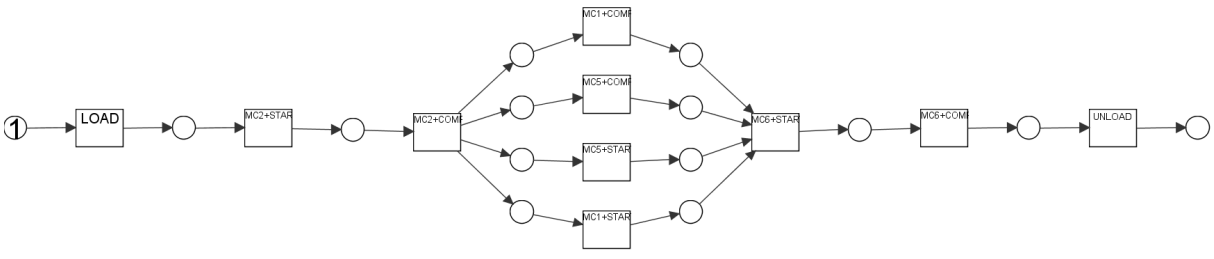


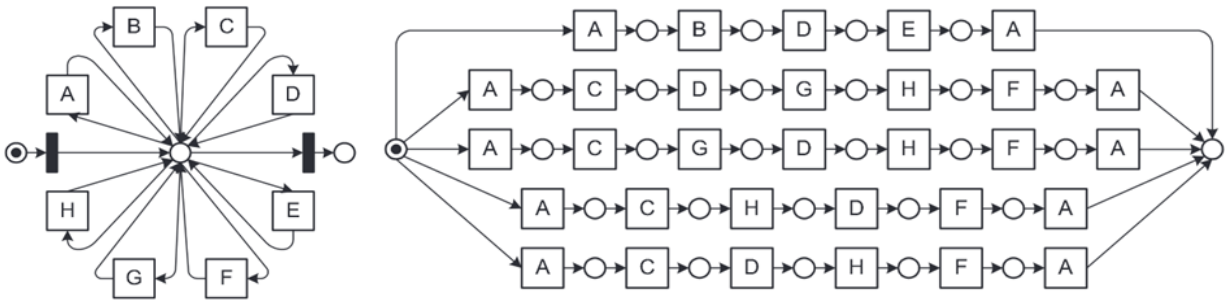
Figure 4.3: Petri Net mined by means of Inductive Miner with noise threshold equal to 70%.

The models obtained are not wrong model, they are merely a different representation of the information contained in the event logs. The choice of the parameters, therefore,

influences how the algorithm returns the model. Since the model discovered will be the starting point for all the next steps, the tuning of the parameters influences affects all subsequent results.

The tuning of the parameters is based on the experience of the operators; furthermore, it is influenced by several aspects, such as the type of collected data, the complexity of the system under analysis, and the quality and quantity of the data. This underlines how Process Discovery is difficult to reproduce as the input data changes. Consequently, it is an iterative process and based on a trial-and-error procedures.

Having generated the model, it is possible to perform the *Conformance Checking*. Conformance Checking is a Process Mining technique that, as explained in 3.3.2, returns the divergence between the data and the discovered model. In addition, Conformance Checking returns certain metrics whose purpose is to estimate the quality with which the model discovered through Process Discovery describes the actual data. The four quality performance indicators are Fitness, Precision, Simplicity and Generalisation; in Section 3.3.2 is reported an explanation of the metrics. Depending on its final use, the model have to respect a trade-off between these quality indicators. For example, models shown in Figure 4.4 are generated from the same data-set, but they explain the data in a different way having different value of the aforementioned metrics.



(a) (b)

Figure 4.4: Both the models are characterized by a fitness of 100%, but the model (b) has a higher precision than the model (a).

The model built by Process Discovery greatly influences the performance metrics; being the model influenced by the tuning parameters, those will also impact the performance metrics. It is difficult to know a priori the values of the tuning parameter that return a suitable trade-off between the four described quality indicators. For this reason, Process Mining becomes a complex and iterative technique.

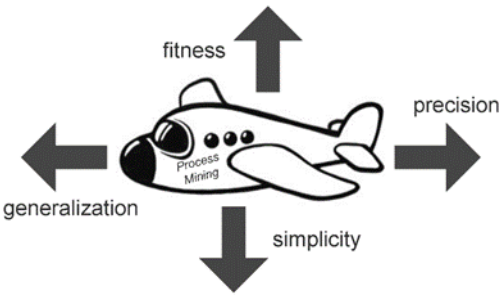


Figure 4.5: Right trade-off is not trivial [96].

The purpose of the Conformance Checking is to identify discrepancies between a model and event logs. In all fields of application in which starting knowledge of the process under analysis is not available, it is not possible to validate the discovered model. However, in industrial domain applications of Process Mining, it is possible to compare the model used in Conformance Checking with the company's knowledge of that process.
In particular, the aim is to use a model that highlights the differences between corporate knowledge and the behaviour of the production system. The focus of Conformance Checking techniques have to be precisely to highlight such divergences in order to investigate their root causes.

At that point, the deviations discovered with the actual methodology are those between the model constructed through Process Discovery, and the real data; and not between the knowledge available in the company and the data collected from the floor shop. The four metrics returned by the Conformance Checking focus on how well the constructed model explains the behaviour of the data. For example, the Fitness indicates how many traces are repeatable in the model, but it does not have the ability to distinguish between the parts processed correctly and those that are processed in the wrong sequence.
Therefore, in the case in which a model different from the one that summarises corporate knowledge is being used, the four dimension of quality do not explain whether the model represents the real behaviour of the process under study. An example of this is-

sue is reported in the Figure 4.6. The model, discovered by Process Discovery, shows the passage of different parts within the production system, but allows parts that pass through MC4+START not to pass through MC4+COMPLETE. This is a mapping error of the production system, since all the parts that start machining in one machine have to necessarily finish it in the same machine. This model does not allow the detection of the described error.

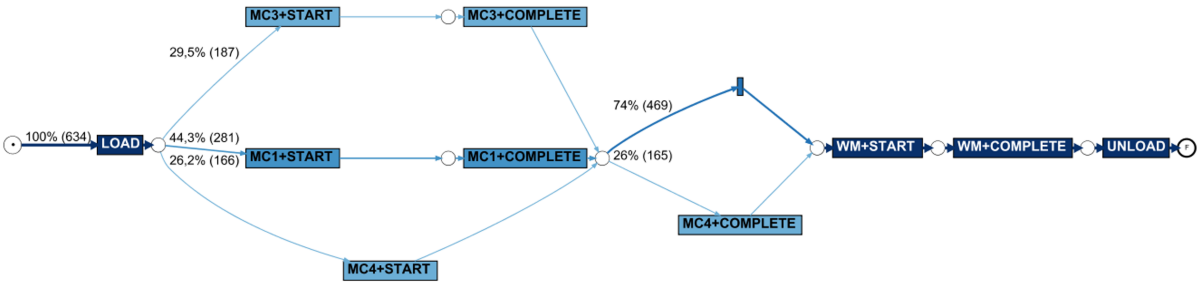


Figure 4.6: Petri Net mined with trace errors.

Conformance Checking returns several possible deviations from the initial model. Once it has been established that the highlighted deviations need to be incorporated into the model, *Model Enhancement* can be performed. It is divided into two categories:

1. Redesign of the model through Process Discovery. This again leads to the problems mentioned above and does not allow the extraction of new knowledge in order to increase awareness of the process.

2. Repair the previously discovered model. Repairing the model avoids Process Discovery problems but limits knowledge extraction to the discovered model. When a change in a model is performed, like after an update, new knowledge can be discovered starting from the updated model, so the knowledge update practice needs to be iterative. To modify the model, there are Petri Net editing tools that allow to add or remove subnets to the model. Technically, editing the Petri Net is not tricky. However, operationally, it isn't very easy to identify where the editing needs to be done since the models discovered by Process Discovery are easy interpretations.

The main purpose of Process Discovery and Conformance Checking is to learn about a process; on the other hand, the goal of Model Enhancement is to integrate possible deviations into the current model. The manufacturing world needs to integrate processes already known a priori with the only relevant information discovered through Process Mining techniques. Not all the extracted information have to change the model: if an item has performed a wrong route, it is important to identify and study it, but the model should remain unchanged.

Finally, it is possible to simulate what the performance would be if something in the process under investigation changed. This type of analysis is named *What-if Analysis* and it is applicable in several domains, especially manufacturing. In literature, several methods can be used, but they all lack of a real integration with the previous Process Mining phases. Currently, it is necessary to collect all the knowledge extracted in the previous steps and rebuild a model on a different platform. This makes the process laborious. In the case in which the only source of knowledge for generating the simulation is the Process Discovery, it is not possible to be certain of the truthfulness of the simulation. As emphasised by the Figure 4.6, the models may contain errors and allow tokens to travel wrong routes.

4.2. Proposed Methodology

The proposal arose from the needing to better adapt Process Mining techniques to the manufacturing systems. This domain has peculiarities that make it more easily adaptable to Process Mining practices because it allows a better integration of all the tools compared to other fields. In addition, the need to integrate and update knowledge from different areas of the company, through a Feedback Line, is of paramount importance. It will be demonstrated how criticalities related to Process mining are avoided and how the proposed methodology can enrich the knowledge of the company and its employees.

The proposed methodology is designed to be general, and therefore can be applied to different production realities. The application requirements are:

1. The processes of the production system have to consist of several operations, tracked by an information system. In practice, it is necessary be able to reconstruct the real path of the part within the production system. This implies that process-based production systems, such as blast furnaces and refining, cannot benefit from the method.

2. Each part within the system must be traced with a unique identifier number. This avoids convergence and divergence problems during conformance checking.

3. In order to carry out what if analyses, it is necessary to have information on the average processing time of products. This implies that the production monitoring system, which is generally the MES, must have a clock-on and clock-off function in work orders.

4. Last but not least, information on how processes are to be executed nominally must be present in the company. Generally, this information can be found in various company documents, such as BPMN Diagram or Process Flow Diagram.

Corporate knowledge is distributed in different sources, such as information systems (ERP, MES, etc.), the BPMN Diagram and the Product Lifecycle Management, but also in workers or managers. Therefore, knowledge is fragmented, and each source has its level of knowledge. For example, the MES system allows the extraction of cycle times of processes, and the real routing of parts in the system. Workers, on the other hand, hold knowledge derived from experience especially in relation to all those non-nominal operations that are performed to simplify processes and that are not reported in the documents, such as the availability of a machine to process only a particular part for reasons of convenience. Since knowledge is fragmented, there is no system or model that groups it all together in a clear and easy-to-view manner, and there is no system that allows this knowledge to be

updated globally. Each knowledge system has its own inherent method of updating.

In addition, all managerial systems that track processes are tied to nominal aspects, but, in practice, actual processes may not be in line with nominal knowledge. For instance, nominal cycle time is not always respected; this difference is due to many reason, such as downtimes, interaction between various resources working simultaneously on a single process or unknown factors.

What has been written has led to the need to structure business knowledge, and also to structure the method by which it is mapped. Having a system that conveys all corporate knowledge related to the process into a model allows both the alignment of so-called nominal knowledge with real one, and an overall integration of the various fragments of knowledge spread across the various sources.

Importantly, this approach opens the door to forecasting practices or decisions based on hypothetical scenarios: starting from incomplete, fragmented knowledge, that is not in line with real information, it is not possible to do scenario analysis based on a model not fully consistent with the real production system; instead, having a model that represent the real behaviour of the system allows the extraction of all the information needed to study performance and scenarios that are as realistic as possible.

The proposed methodology seeks to improve on the current method of approaching the problem, described in Section 4.1, by means of an a priori implementation of the knowledge hidden in the various systems and almost silent. The methodology is shown in Figure 4.7 and it is described in the following.

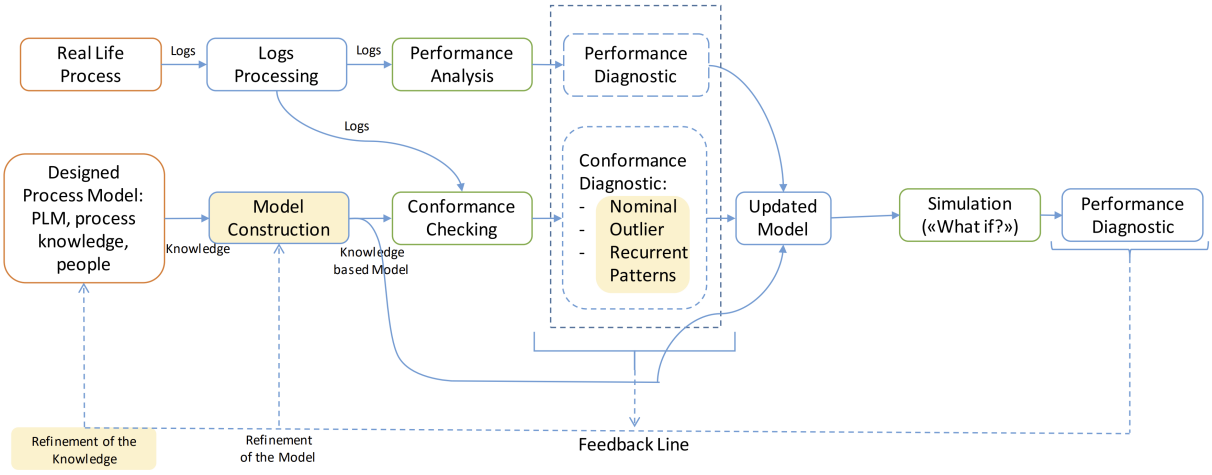


Figure 4.7: Proposed Methodology.

As can be immediately observed from the block diagram, the proposed approach eliminates Process Discovery techniques. Process Discovery involves several problems related to the choice of the algorithm to implement, the tuning of the input parameters, and the fact that the model does not always visually describe the production system, making the extraction of relevant information difficulty. In manufacturing systems, Process Discovery can be replaced by a priori *Model Construction* based on the fragmented knowledge present in all the systems described above. In particular in Figure 4.7 the novelties of the proposed methodology with respect to the existing one are highlighted.

The constructed model, preferably in the form of a Petri Net, is the crux of the methodology. The model have to be built on the basis of all the company's knowledge and have to visually describe the actual process. Therefore, studying all the sources that hold the fragmented knowledge is the first step in the proposed methodology. All essential data, such as number of machines present, type of parts produced, cycle times and routing of parts, have to be extracted and studied in detail. Once all the information has been collected and studied, the model is built in the form of a Petri Net. This approach thus avoids trial-and-error model discovery from event logs. It also allows for a more detailed overview of the company and its processes.

Event logs still play a key role. As in the existing methodology, they must be processed, filtered, and converted into XES files before they enter input to Conformance Checking. Therefore a parallel activity to be performed with the Model Construction is the *Logs Processing*. From those data it is possible to compute all metrics used by the company to track production performance and also it is possible to perform Conformance Checking.

The next step of the proposed methodology is the *Conformance Checking*. The inputs are the logs and the knowledge-based model. The Conformance Checking techniques assess the similarity between the actual behavior in the event log and the behavior described by the a priori constructed model. Since the input model is descriptive of the nominal production plan, the Conformance Checking will identify anything that diverges from it, such as rework not described in the BPMN Diagram or predisposition of some machinery to process only certain parts and much more. Thereby integrating Conformance Checking, it allows to identify real deviations in the production plan and not deviations from the model. The output analysis of the Conformance Checking is called *Conformance Diagnostic*.

In addition, it is possible, by means of alignments, to classify more quickly the *Nominal Flow*, which will turn out to be consistent with the model, the *Outliers*, meaning some parts that deviates from the nominal path as an exceptional case, and *Recurrent Patterns*, meaning paths that are not in line with the constructed model, but executed by a large number of items.

The Recurrent Patterns are of fundamental importance. They in fact represent a gap, a lack in the initial knowledge on which the model was built. Recurrent Patterns are to be considered not only as paths other than nominal, such as a rework, but also cycle times far from nominal values or not aligning in the sequence of operations or machinery.

All information not explained by the model are essential. The Recurrent Patterns are the main source of knowledge upgrades and, consequently, of *Model Enhancement*. Instead, Outliers will be able to be identified and then studied individually. What the current

methodology defines as error and attempts to suppress with ever-improving Process Mining techniques, in the current methodology is a resource that allows for improvements on all levels of the business.

The Conformance Diagnostic enables the next step: the *Refinement of the Model* and the *Update of the Knowledge*. These steps are related to the Feedback Line and they allow, once the lack of knowledge is easily identified, to update both the model built a priori and the information about the system. Every change of the model is called *Model Enhancement*. The proposed methodology allows to discover in a easier way practices nominally not known and that, applying the existent methodology, remained hidden among the Process Discovery techniques. For example, it can be discovered from the model that a part produced with a processing time of 8 hours is actually processed on two phases of 4-hour cycle time; another case may be related to information that allows for knowledge enrichment, but which will not change the model lest it lose its generality, as for example the coupling machine-product type. The model already built will also be updated faster and more quickly than if a new Process Discovery or repair Model Enhancement technique is applied. The model thus built will gain "experience", meaning that all updates will be performed on the same model enriching its process knowledge capability.

Filling knowledge gap brings with it many advantages, chief among them the possibility of making predictions about the production system. As explained in Section 3.4, those scenarios are called *What-if Analysis*. Having a model updated with the correct information coming from the system allows to build a second model able to study how the system behaves if input conditions change. What-if Analysis allow the application of this methodology at a strategic level, allowing for the study of improvements in the medium or long term.

Therefore, the proposed methodology is useful in the following:

1. Extract information related to how many pieces were produced as planned.

2. Identify deviations, and their causes, of the parts from the production plan.
3. extract system performance indicators such as lead time, saturation of the machines, and up-to-date process parameters, such as cycle time and part routing.
4. Use the up-to-date process parameters to perform What-if Analysis.

5 | Case Study

In this chapter, the proposed methodology explained in the previous chapter is applied to a real case. In order to validate the proposed methodology, it was chosen to use data collected by a manufacturing company that performs machining operations on product types that are mature for that company, meaning products with known and monitored processes over time. Between all the different types of manufacturing systems, shown in Figure 5.1, it was chosen to analyze a Flexible Manufacturing System (FMS). The FMS is in the middle of the graph in terms of production capacity and flexibility; this allow the production system to have good production capacity and high flexibility that allow abundance of data and variability in the technological process.

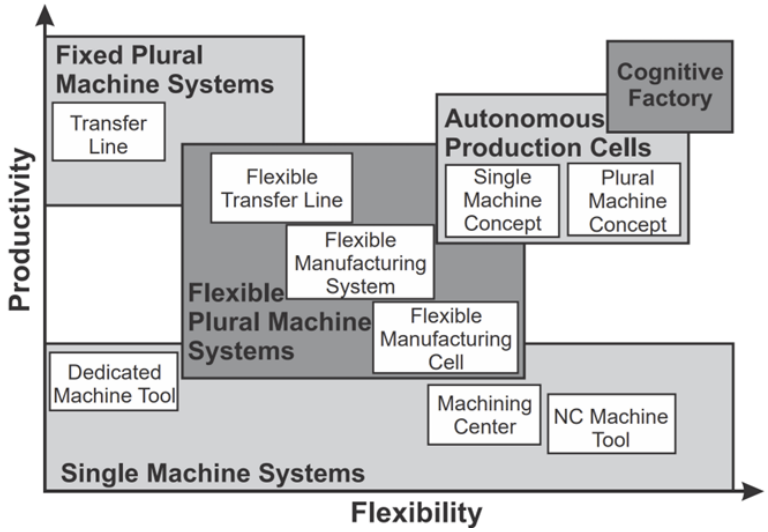


Figure 5.1: Productivity and flexibility of FMSs of different levels of complexity [55].

The company that provided the data is a global leader in innovative technologies and complete lifecycle solutions for the marine and energy markets. By emphasizing sus-

tainable innovation, data analytics, and overall efficiency, this company maximizes the environmental and economic performances of its customers' vessels and power plants. The company provided several sources of information, including data extraction from the MES database. Therefore, the methodology assumption reported in Section 4.2, are verified. The extraction was the basis for the analysis that will be shown in this chapter.
The data available were collected over 65 months of production, and they include 18,200 manufactured parts performing 500,000 different events. In addition, nominal cycles of the produced parts and the BPMN diagram were provided as an additional source of knowledge, as well as company layouts and technical drawings of the parts. All those information was collected, studied, and synthesized for model construction.

The production mix of the examined plant includes 3 product types: PROD1, PROD2, and PROD3. All products are divided into two different weight classes, named X and Y. Summarizing, the whole groups of products are six, subdivided into three types and, for each type, two weight classes.

The production of the products is divided into three main processes: machining, pre-assembly, and assembly.
The machining processes are performed inside the Flexible Manufacturing System (FMS). They start when the ERP system releases a production order allowing the casting part from the supplier or pre-machined part. The production order enters the FMS and ends only once the parts are unloaded from the fixture after being measured by the Coordinate Measuring Machine.
The pre-assembly process starts when the ERP releases a production order and ends after an ultrasonic test process. The assembly process joins all different parts to the final good, which is ready to be installed on the final component.

In this case study, the focus will be on the machining processes. In detail, the castings of the three products are delivered from the supplier to the production site. Once the castings arrive in the factory, they are loaded on the FMS material pallet storage, and the

Manufacturing Management System (MMS) reads their casting UII. When the company receives an order, the MMS executes the production order and releases the Machined ID. The system calls the casting from FMS material pallet storage, unloads it, and then mounts it on a fixture that has been set. Once the part is mounted on the fixture, it is loaded to the FMS, allowing different operations into the six machines (MC1, MC2, MC3, MC4, MC5, MC6). Finally, the part can be unloaded from the fixture: the finished product returns to the FMS material pallet storage, while the fixture is ready to receive a new part to be machined.

In this context, the MES system is responsible for collecting data from the FMS; those data contain information like timestamps, ID, operations, and other product statuses like quality control results. Therefore, the following work is based on the data collected from the MES.

5.1. Data Processing

The following section is dedicated to processing event logs extracted from the FMS. This operation is called Data Processing. It is a step in which the data are standardized, filtered, and selected according to the needs of the analytical processes that will be conducted in this study. The data to be processed are extracted from the MES database by means of queries. MES systems collect and generate data based on standards. Usually, this prevents loss of data and corruption. Because of this, it was not necessarily a big data processing, so Excel has been used to filter and organize data. Since the fields extracted from the MES were not known in advance, a tool such as Excel made it possible to quickly view the data format and the different fields, such as the machinery connected to the MES. Only some of the machines in the FMS have a connection to the MES.

The available data contain a wide variety of information such as ERP order number, registration date, registration time, part ID, part name, fixture ID, type of fixture, order

code, and a lot of other details. All this information is not essential for Process Mining, thus only the valuable information for the purpose of the study is taken into account. In particular, the information preserved from the data are:

- temporal information: date and time;
- unique identification code: Machined ID;
- type of machined part;
- type of operation performed;
- machinery on which the part was machined;
- status of the machinery on which the part is machined.

The raw data are listed in Table 5.1, while the Table 5.2 shows the data filtered out of unnecessary information.

ERP Order Number	Date	Time	Source Device	ID Fixture	Fixture Description	ID Part	Part Description	Number Operation	Production Order	Machine Status	Machined UII
A1	09/02/2021	06:44:00	LOAD	ABCD1	PROD2	LMNO098765	PROD2 X	70	8394857	L	FI48000Z1#20J#8394857#5#B#S
A2	09/02/2021	11:50:00	MC2	ABCD2	PROD3	LMNO098766	PROD2 X	70	8394857	S	FI48000Z1#20J#8394857#5#B#S
A3	09/02/2021	12:00:00	MC2	ABCD3	PROD4	LMNO098767	PROD2 X	70	8394857	C	FI48000Z1#20J#8394857#5#B#S
A4	09/02/2021	14:13:00	MC6	ABCD4	PROD5	LMNO098768	PROD2 X	70	8394857	S	FI48000Z1#20J#8394857#5#B#S
A5	09/02/2021	14:39:00	MC6	ABCD5	PROD6	LMNO098769	PROD2 X	70	8394857	C	FI48000Z1#20J#8394857#5#B#S
A6	09/02/2021	15:09:00	UNLOAD	ABCD6	PROD7	LMNO098770	PROD2 X	70	8394857	UL	FI48000Z1#20J#8394857#5#B#S

Table 5.1: Raw data coming from MES extraction.

Date	Time	Source Device	Part Description	Number Operation	Machine Status	Machined UII
09/02/2021	06:44:00	LOAD	PROD2 X	70	L	FI48000Z1#20J#8394857#5#B#S
09/02/2021	11:50:00	MC2	PROD2 X	70	S	FI48000Z1#20J#8394857#5#B#S
09/02/2021	12:00:00	MC2	PROD2 X	70	C	FI48000Z1#20J#8394857#5#B#S
09/02/2021	14:13:00	MC6	PROD2 X	70	S	FI48000Z1#20J#8394857#5#B#S
09/02/2021	14:39:00	MC6	PROD2 X	70	C	FI48000Z1#20J#8394857#5#B#S
09/02/2021	15:09:00	UNLOAD	PROD2 X	70	UL	FI48000Z1#20J#8394857#5#B#S

Table 5.2: Data processed.

Further steps are necessary to streamline and standardize the information. In particular, the information related to the date and time of execution has been merged to give the information a single format (date-time).

Different reasoning should be addressed to the status of the machinery. This information began to be collected towards the beginning of 2020, at the same time as a change in

data collection. From the beginning of 2020, the logs related to machining operations "double up," showing both the Clock On information (part entry) and the Clock Off information (part exit), and this has made necessary the introduction of an additional column explaining the start and end of the operation: the machining status. At this point two considerations should be made:

1. It was decided to consider only available data collected in the last two years. Over these years, 213,000 events have been recorded, divided into 6,729 parts produced. This choice was dictated by the need to have the most detailed information possible and, in particular, to be able to make reasoning related to the cycle times of the machines. In fact, by recording the entry and exit of the workpiece in the machinery, it is possible, by subtraction, to derive the time required to process the workpiece.

2. To streamline the amount of data available and to facilitate Conformance Checking, it was decided to combine the information related to the machine on which a part is processed, and its status. Therefore, instead of having two separate columns with duplicate information that must necessarily go ahead together in order to be completed, it is possible to obtain a single column structured as follows: machine MC1 starting (machining status S) the operation will become MC1+START, while machining MC1 ending (machining status C) will become MC1+COMPLETE.

According to the described considerations, the entire event log will look as follows:

Timestamp	Source Device	Part Description	Number Operation	Machined UII
09/02/2021 06:44	LOAD	PROD2 X	70	FI48000Z1#20J#8394857#5#B#S
09/02/2021 11:50	MC2+START	PROD2 X	70	FI48000Z1#20J#8394857#5#B#S
09/02/2021 12:00	MC2+COMPLETE	PROD2 X	70	FI48000Z1#20J#8394857#5#B#S
09/02/2021 14:13	MC6+START	PROD2 X	70	FI48000Z1#20J#8394857#5#B#S
09/02/2021 14:39	MC6+COMPLETE	PROD2 X	70	FI48000Z1#20J#8394857#5#B#S
09/02/2021 15:09	UNLOAD	PROD2 X	70	FI48000Z1#20J#8394857#5#B#S

Table 5.3: Final data used for Conformance Checking.

Data are now ready to be used for Process Mining techniques.

5.2. Knowledge Study

Knowledge is the basis of the proposed methodology. It is essential to build the model used in Conformance Checking and comprehensively understand the Conformance Diagnostic. At this stage of the work, all the information available from the company has been collected and summarized. The following sources have been analyzed:

1. Plant Layout: it is possible to see in which place parts are physically stored, the machinery in which they are processed, and the subdivision of departments. In this case, the FMS area is well separated from the other departments, and it is connected directly to the warehouse.

2. BPMN Diagram: Business Process Modeling and Notation (BPMN) is the global standard for business process modeling, a fundamental part of business process management. BPMN Diagrams enable the visualization of business processes in order to facilitate workflows. In this way, it has been possible to derive most of the information about the work processes carried out and how data are saved in different information systems. Figure 5.2 is the BPMN used as a reference in the case studied.

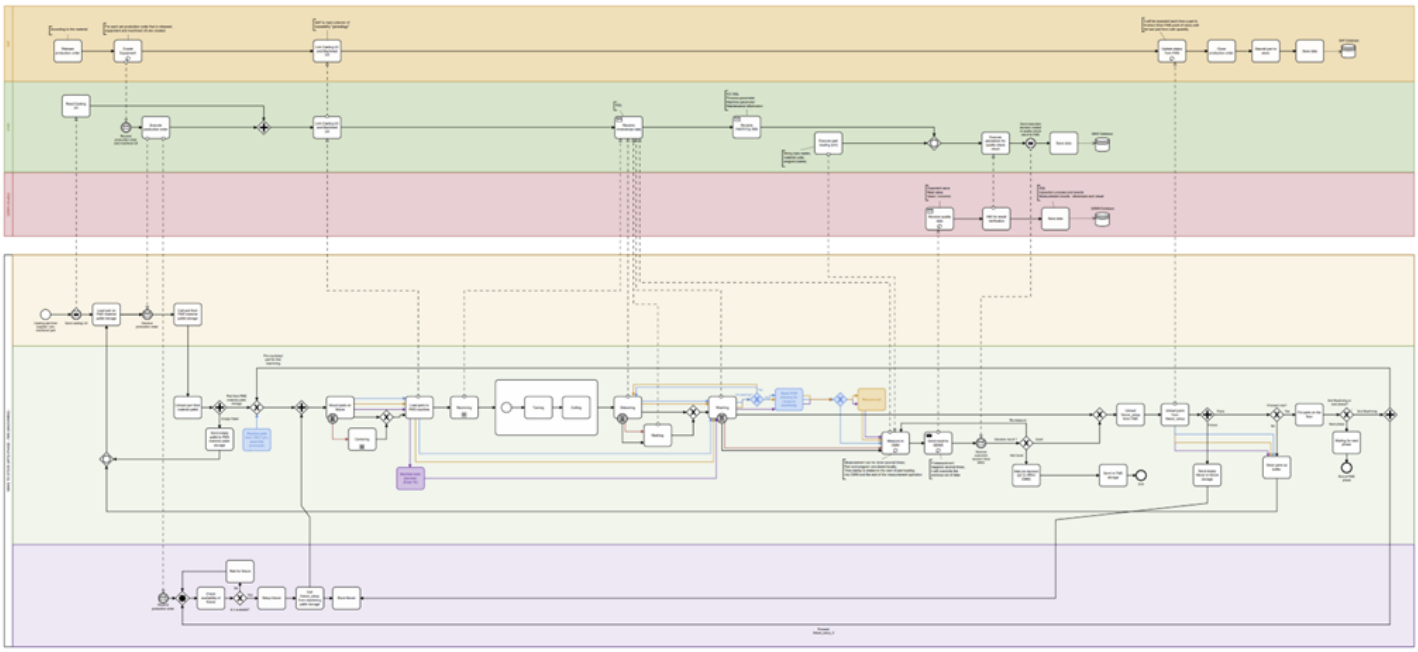


Figure 5.2: BPMN Diagram of the production system.

3. Process Flow Diagram: a flow diagram is a diagram that describes the steps of a process and their sequence. In this case, it describes all the technological sequences needed to complete a product. This means that a process diagram is available for each part processed in the FMS. This diagram has been fundamental in understanding the nominal path of each product. For reasons of privacy, the Process Flow Diagrams that the company made available have not been reported in this document.

4. Dotted Chart: it is a widely adopted tool that allows the interactive visualization of several perspectives of an event log and, therefore, it is a Process Mining analysis technique. For results evaluations, it is used to get insights on how to process data further. Because of this, data related to products in different operations have been analyzed.

After studying the BPMN and the nominal routing, it has been possible to find that each product type performs different operations. An operation can be defined as a sequence of technological processes aimed to obtain predefined features. In order to obtain the

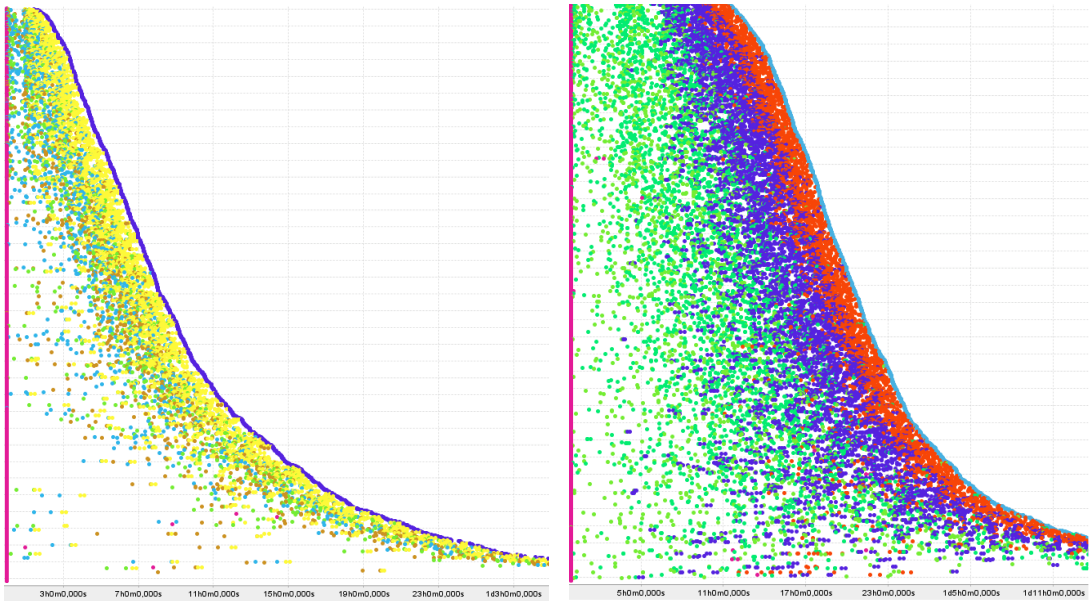
finished product, a series of operations have to take place on the raw part. The following table (Table 5.6) summarises the operations performed on the products under consideration:

Part	**Operation**				
PROD1 X,Y	10	20			
PROD2 X,Y	10	20	40	50	70
PROD3 X,Y	10 i	10 ii			

Table 5.4: Performed operations according to product type.

To construct a production system model valuable for Process Mining, it is not enough to know the technological process of the individual stages; it is also necessary to study the processing times. This is due to the fact that processing times may vary depending on the operation performed. In order to study the processing times, the Dotted Chart divided by operation of each part has been constructed.

An example of a Dotted Chart constructed for PROD1 X, focusing on the operations 10 and 20, is shown in Figure 5.3. Each event is represented by a coloured dot according to the machine connected to that event. The x-axis and y-axis indicate respectively the time and part ID (obscured for privacy reasons). There are several ways to order the events in the Dotted Chart. In this case, events are sorted according to the duration of the event. The pink dots, in correspondence of time t_0 on the left side of Figures 5.3a and 5.3b, and blue dots on the right side of the Figure 5.3a and light blue dots on the right side of the Figure 5.3b, represent respectively the start and end of the operations.



(a) Operation 10. (b) Operation 20.

Figure 5.3: Dotted Chart of the PROD1 X.

The study of the Dotted Chart shows that the cycle time changes depending on the operation performed and the type of part machined. In fact the distribution of machining time is clearly different depending on the operation performed, this concept is highlighted by the different widths of the two Dotted Chart. Thus, the cycle time of a machining process in an operation is usually different from another machining process that is performed in another operation, even if it is done in the same machine.

In practice, machining the PROD1 X in operation 10, Figure 5.3a, takes a different amount of time that machining the same product on the same machine during a different operation (operation 20), Figure 5.3b.

For a better understanding of how data are distributed over the time period considered, a second Dotted Chart has been constructed, shown in Figure 5.4. Here on the y-axis the are the products (PROD1 X, PROD1 Y, etc), and on the x-axis the time dimension.

It is evident how the production mix varies over time, and it is possible, at the bottom of the Dotted Chart, to view an indicator showing machine utilization as time changes.

Figure 5.4: Dotted Chart of 2020 production year.

It is important to emphasise that BPMN Diagram, Process Flow Diagram and Dotted Chart from MES are related to different levels and types of information. The BPMN is related to the flow and to the sequence of parts and information from the beginning to the end of the business process, showing high level information related to the machining area (cutting or washing) and not to the individual machine or to the type of operation. The nominal routing shifts the focus to the individual products, explaining the individual flow, taking care to explain the type of operation (first a roughing cycle, then a hole machining cycle). Lastly, the MES focuses on collecting the most critical machinery data of the process.

5.3. Model Construction

Once all the information necessary to fully understand the production process have been studied, the model can be constructed in the form of a Petri Net. This method of representing production processes was explained in the Section 3.2 and was performed by

means of the open source software WoPeD (Workflow Petrinet Designer). The Petri Net software has been chosen for simplicity of model construction after an in-depth study of the different tools available that support the import and export of PNLM (Petri Net Markup Language) as specified in the official standard. In addition, WoPeD can import PNML files from the main Process Mining software used, ProM.
It is important to emphasise that the choice of the software is not binding for the results: it is possible to choose any other tool that allows the construction and the transfer of a Petri Net to Process Mining tools.

After analysing the Process Flow Charts and the data collected by the MES, a mismatch has been found between the nominal routing and the data available in the form of event logs: some technological processes shown in the flow charts are not tracked by the MES. Consequently, these process steps will not be considered in the construction of the model.

It emerged from the Process Flow Chart that the routes of a part are the same for each operation. A route is the sequence of machinign processes that a part may have to be complete. The available routes are listed below:

1. <LOAD, MC1, MC5, MC6, UNLOAD>
2. <LOAD, MC2, MC5, MC6, UNLOAD>
3. <LOAD, MC3, MC5, MC6, UNLOAD>
4. <LOAD, MC4, MC5, MC6, UNLOAD>

This allows the same sub-net to be used for all the operations. The sub-net constructed is shown in the Figure 5.5.

Since the production of a part is the succession of several operations, that can be modelled with the same sub-net, it is possible to model the entire production cycle by inserting a loop in the model in Figure 5.5. The result obtained is the following Petri Net:

Using the latter model to perform Process Mining operations is not optimal. The ProM software allows the execution of automatic statistics, such as calculation of lead time,

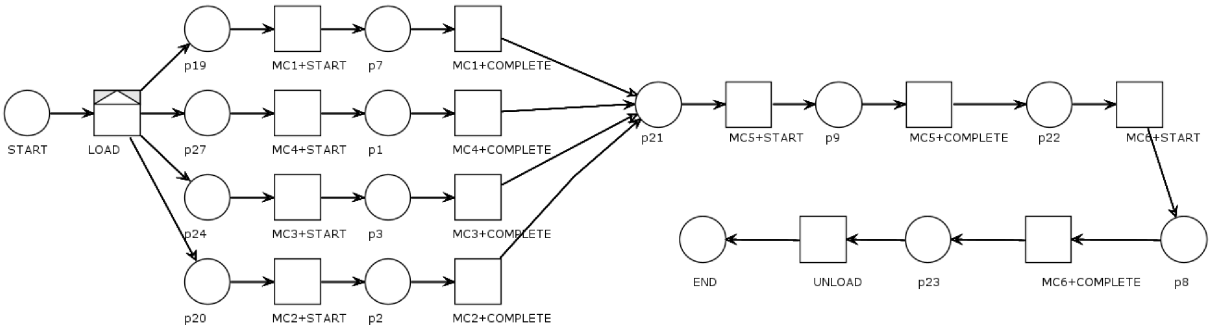


Figure 5.5: Petri Net model of the single operation in the production cycle of a part.

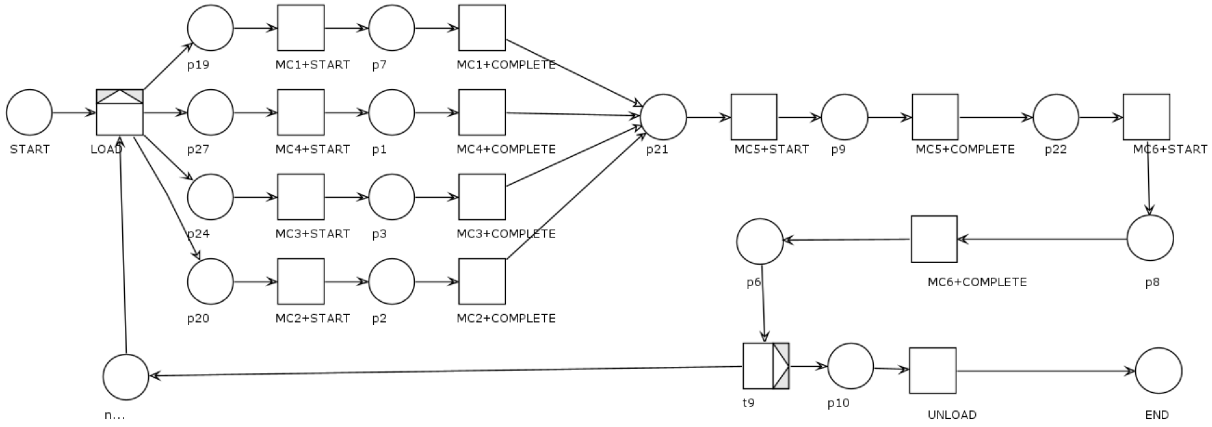


Figure 5.6: Petri Net Model of the production cycle of a part.

cycle time, waiting time, distribution of parts in the machines, number of total parts. All these information are needed to perform the What-if Analysis. If only one model were used for the entire production cycle, such as the one in Figure 5.6, the calculation of the statistics would be the result of averaging the statistics of each operation. As pointed out above, as all the information differ per operation, the average value is not functional for the analysis.

In order to not lose timing information, the use of a loop model as shown in Figure 5.6 is not the optimal way to proceed. The model used is a redundant model built from as many sub-nets, shown in Figure 5.5, as the number of operations performed by the part under examination.

In particular, following the Table 5.6, three complete models will be constructed. The first, related to PROD1, consists of two sub-nets since the operations performed are two (operation 10 and operation 20); the second one, related to PROD2, consists of five sub-nets since the operations it undergoes are five (operation 10, 20, 40, 50 and 70); the third model, related to PROD3, consists of two identical sub-nets since this product performs twice the operation 10. In this way, the method is flexible and it gives a complete view of the process, allowing ad-hoc analysis for each product. Figure 5.7 represents the Petri Net model for the product PROD1.

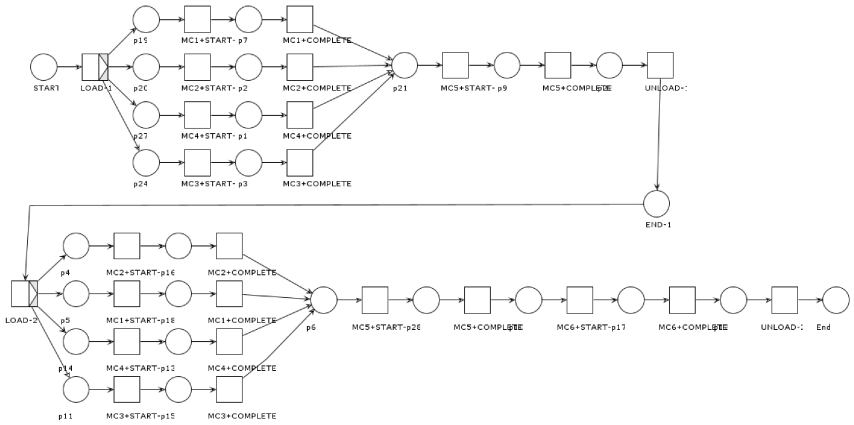


Figure 5.7: Petri Net model for PROD1.

A relevant observation is a possibility of directly translating BPMN into Petri Net through special tools. This possibility is very interesting from a business point of view since companies widely adopt the BPMN language due to its simplicity of construction. Furthermore, it contains more information, not only related to the workflow of the parts but also to how and where the information is collected in the process, how they are processed, and how the pallets are managed once the part is unloaded.

This step is conceptually very interesting and requires in-depth studies, but it was not implemented in this specific case as it is outside the scope of this paper. ProM offers a plug-in to convert a BPMN model into Petri Nets automatically.

5.4. Conformance Checking and Knowledge Extraction

After the Model Construction phase, the next step of the proposed methodology is Conformance Checking. The goal of this step is to discover misalignments between the knowledge-based model and the event logs, i.e. the actual data recorded during the process. This area of Process Mining, integrated into the methodology, allows the initial knowledge of the company to be enriched. In fact, the identification of a mismatch between the constructed model and the logs means a mismatch between the knowledge of the process and the way in which the production is executed. Therefore, this section encompasses both the explanation of the Conformance Checking and the identification of different and new information from the initial business knowledge.

The software used is ProM. ProM is a free and extensible framework that supports many Process Mining techniques by means of plug-ins. As explained for the Petri Net software WoPeD, the choice of this tool is not limiting: any tool that allows the use of Process Mining techniques can be used. The software chosen is ProM because, as shown in [105], in addition to being the most widely used tool in the literature, followed by Disco and Celonis, it is a software that allows the integration between all the analysis techniques of the proposed methodology.
In fact, ProM allows the import of Petri Net, the conversion between CSV and XES files format, and Conformance Checking via alignments by means of the "*Multi-perspective Process Explorer*" plug-in. Furthermore, it allows the extraction of all the information needed to build a model for What-if Analysis.

Among the different plug-ins available on ProM, the "*Multi-perspective Process Explorer*" has been chosen for its unique features compared to the others available on the platform [59]. The main features are the integration of existing discovery, conformance checking and performance analysis techniques, integrated filtering based on process attributes and

trace variants.

Given the abundance of available data, this study focuses on the year 2020 by analysing 1,111 machined IDs, resulting in 110,000 events in the event logs. Conformance Checking can be performed over periods from few days up to years. If a short period is analysed, it is difficult to understand whether the divergences in the traces are recurrent patterns, or exceptions to the period under examination. It is therefore useful to have a more meaningful sample of the production system. However, it should not be made too large as the model can risk the opposite problem, namely that of making the reading of the traces too complicated due to the large quantity. In the present case, the choice of a one-year period allows a right trade-off.

As explained in Section 3.3.2, the necessary inputs for the Conformance Checking are process model and event logs. The construction of the model was explained in the previous section (Section 5.3).
The event logs have been created by extracting data related both to the individual parts (PROD1, PROD2, PROD3) and to the year under review. Then, the files containing the extracted data are converted into XES format using the ProM plug-in *"Convert CSV to XES"*. At this point, the MPE plug-in can be used to perform the Conformance Checking.

The Conformance Checking is performed for all the parts (PROD1 X, PROD1 Y, PROD2 X, etc), but for reasons of synthesis it was decided to report the Conformance Diagnostic and knowledge extraction only for the product PROD1 X. This product has been selected since it is representative of the divergences studied in the other products. The Petri Net in the Figure 5.8 shows the Conformance Checking of the selected product performed by means of the Plug-in MPE.

As it can be noticed from the Figure 5.8, the Conformance Checking allows the display of part routing and transition times between one event (MC1+STARTI-10) and the next one (MC1+STARTI-20). Part routing can be visualised using different criteria, the one

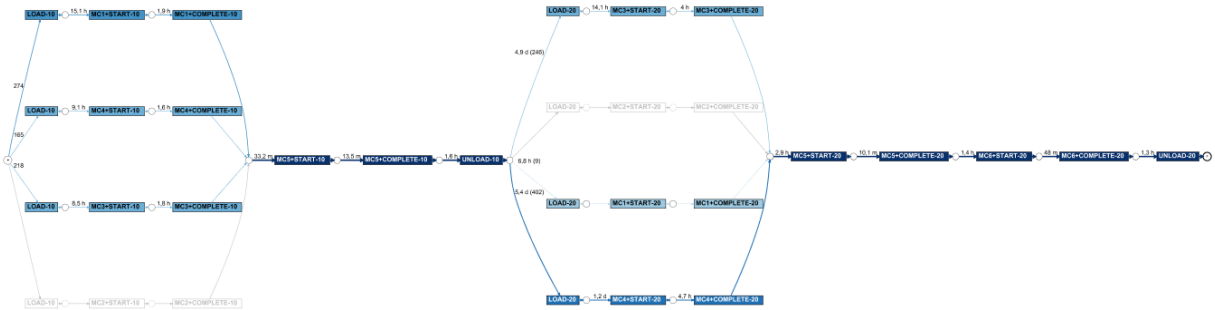


Figure 5.8: Conformance Checking of the product PROD1 X.

selected is the frequency, i.e. the number of parts following a route. This visualisation makes it possible to know both the quantity of parts produced in the selected time frame and to calculate the percentage of parts that perform a transition. In particular, between the 657 parts that perform operation 10, 274 (41.7%) are executed on the MC1 machine in 1.9 hours, 165 (25.1%) on the MC3 machine with a cycle time of 1.8 hours and 218 (33.2%) are executed on the MC4 machine in 1.6 hours.

This information is crucial when compared with initial knowledge. In fact, it can be seen that operation 10 of the product under investigation is not performed on the MC2 machine; on the other hand, for what concerns the next operation, operation 20, a predisposition of the MC3 and MC4 machines to perform the product PROD1 X is evident. By studying the Conformance Diagnostic of the other products, a clear coupling between the various machines and the product type is noted. The Table 5.5 shows this machinery-product type coupling.

	MC1	MC2	MC3	MC4
PROD1 OP10	✓		✓	✓
PROD1 OP20			✓	✓
PROD2 ALL OP	✓	✓		
PROD3 ALL OP	✓	✓		

Table 5.5: Machinery-product type coupling

Figure 5.9 shows the Conformance Checking focused on operation 10 of product PROD2 X; it shows the PROD2 coupling with the machinery MC1 and MC2.

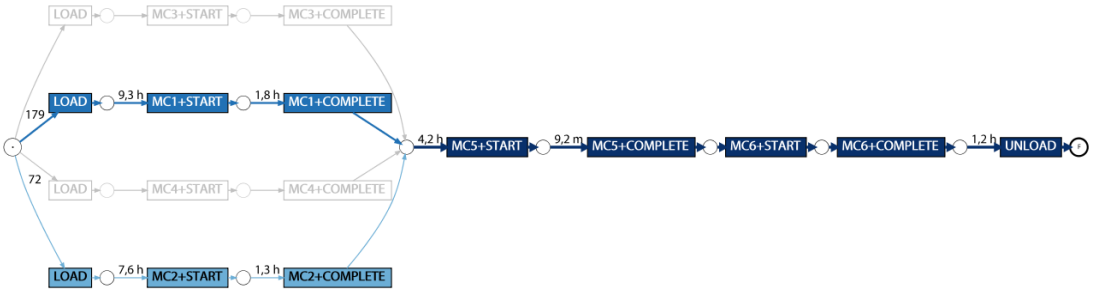


Figure 5.9: Conformance Checking of the product PROD2 X related to Operation 10.

This first analysis already allows to extract absent information in the initial knowledge sources. In fact, there is a clear distinction between machines that are used and type of part produced. At this point, it would suffice to ask process engineers why this distinction is made. It is possible that there is a conscience about which machines perform better certain operations in different parts. This deep dig has many benefits; first of all, it allows one to have a clear idea of what is going on in production; it also establishes a control mechanism aimed at finding out why specific actions are being done. "It's always been done this way" is a problem that companies interface with on a daily basis.

This knowledge upgrade will not change the process model. The choice of coupling product and machinery is not binding, such as the sequence of operations to obtain a finished product, and it can be changed in the future in favour of, perhaps, a balanced routing of machinery. Therefore, by not making this change to the process model, any such future changes can be recognized from the event logs.

In addition, information on machine routing and cycle times are easily extracted from Figures 5.8 and 5.9. These values can be compared with the standard process parameters that the engineering department must keep up-to-date. It is also possible to calculate indicators such as OEE, saturations and others. All this information are essential to

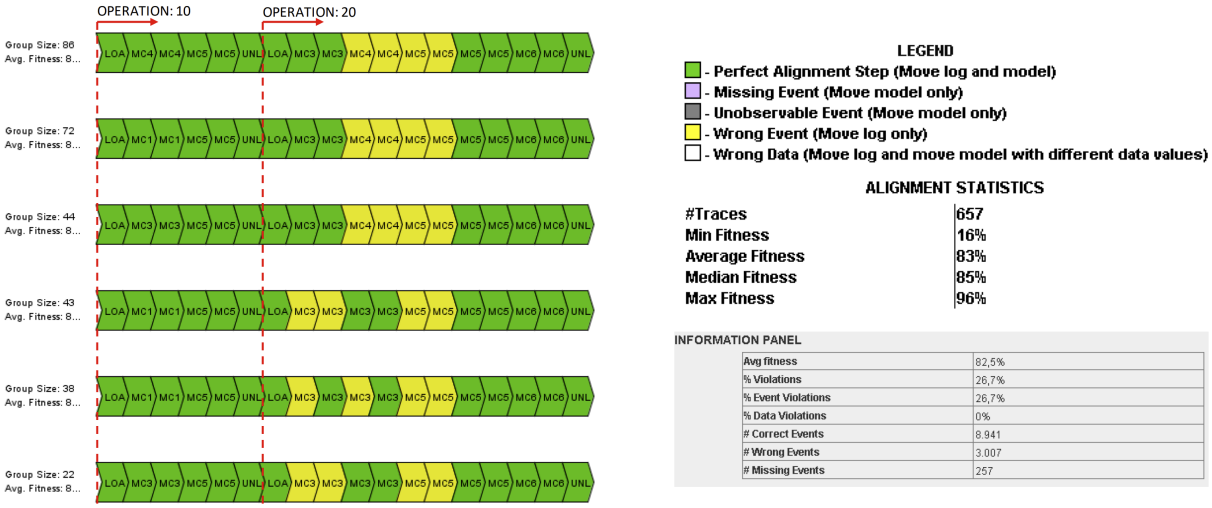


Figure 5.10: Trace View of the product PROD1 X.

study up-to-date company performance and to be able to better study the production process.

At this point, the study can focus on part traces. A trace is a sequence of events and, as explained in Section 5.3, traces are limited in number and are mined by log. This means that different datasets can have different traces, and only the production can affect traces.

The task of this second analysis is to uncover divergences between the process model and the logs at the trace level; the question to be answered becomes: "do the parts produced follow the routes studied, or are there cases not mapped by the knowledge sources?"
In order to answer this question, the display option *"Toggle Trace View"* available in the MPE plug-in is used. This display mode groups individual parts with an identical trace, i.e. parts that follow the same route will be grouped together; it also allows groups of parts to be sorted according to a selected criterion, which in this case is the Fitness value. Each group contains the parts that have the same trace. By selecting each group, it is possible to study the trace of each individual product belonging to that group, so a more detailed study of individual parts is possible if necessary. The latter type of analysis is interesting if finding out the sequence of machines of a part that has defects or problems during customer use. The Trace View of ProM is shown in the following Figure 5.10.

A legend helps to understand the traces. Green events are events in which the alignment between model and log is perfect; purple events are the so-called Missing Events, events in the model that are not found in the logs; finally, Wrong Events, events that took place in the logs and that are not available in the model, are represented in yellow.

Studying the traces allows further understanding of the process. Indeed, they identify the sequences of events performed by the parts and encapsulated in the logs. A mismatch in the traces, identified by a yellow or purple event, results in part behaviour not mapped by the basic company's knowledge.
Going into the specific case of the product PROD1 X, it can be seen that in most traces there is a misalignment during operation 20. Let will study it in detail.

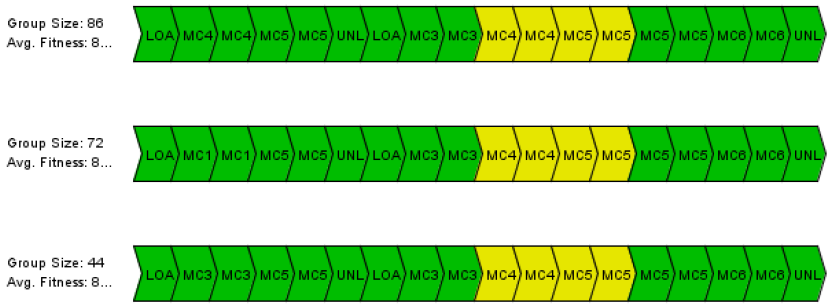


Figure 5.11: Detail of the Trace View of the product PROD1 X.

During operation 20 the parts pass twice over the machines MC4 and MC5. This passage on several machines is not mapped in any company document, so an investigation will be necessary to understand the reason for this divergence.
In case an assignable cause is found, the model may be modified so that the divergence is not considered an error. In case the divergence is not beneficial and it is a production error, the model should not be updated.

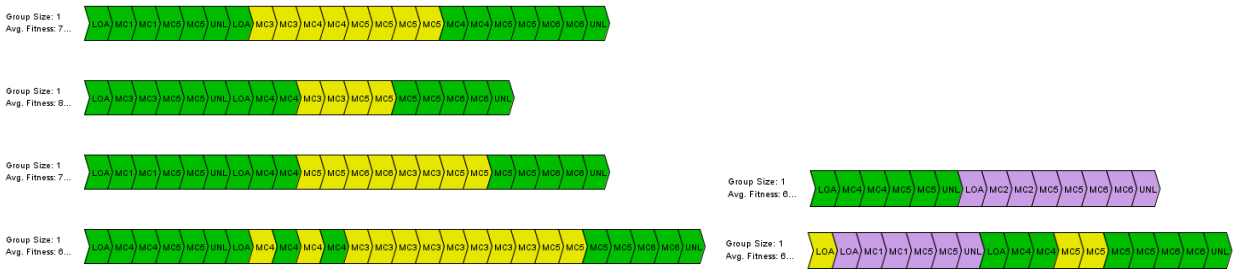
It is important, as in the case of the previous analysis focusing on timing and routing, that this operation is not automatic, but that there is a study between reality and what is discovered by means of the Conformance Checking. It is crucial to interrogate the workers on the reason why this double processing is taking place.

Since the data provider could not be questioned about the reason for this double processing, only hypotheses could be made. The most plausible hypotheses are a rework due to features not respected, such as the tolerance; or the choice of splitting a machining with a cycle time too high into two shorter machining operations to obtain the same result; finally, the last hypothesis is related to the fact that an operation previously carried out on another machine, perhaps not connected to the MES, was moved to the MC3 and MC4 machines, whose data are instead tracked by the MES.
The improvement of the model is the next step in the proposed methodology and will be explained in detail in the next section.

The Conformance Checking, as explained in the 3.3.2, returns the values of some dimensions related to the quality of the model. In particular, the fitness and precision values are given, as well as the number of correct, wrong and missing events, the violation rates and the number of detected traces.
These values will be successively used for the comparison between the old and the proposed methodology, explained in details in the following chapter (Chapter 6). For reasons of completeness KPI parameters are here explained:

1. Average Fitness: it is the most important quality dimension to understand how good the production is to perform processes as designed.
2. Correct Events: it is the number of correct event.
3. Wrong Events: it is the number of wrong event.
4. Missing Events: it is the number of missing event.

Figure 5.11 shows traces sorting them by fitness. For the sake of completeness a trace with low fitness is analyzed in Figure 5.12a.



(a) Detail of Wrong Events.

(b) Detail of Missing Events.

Figure 5.12: Trace of some parts with low fitness.

In this case, it is possible to see multiple rework done at the end of operation 10 and during operation 20. In order to investigate the cause, it is necessary to analyze the quality related to these operations.

Another case is shown in the Figure 5.12b, in which the first trace is missing operation 20, and the second trace is missing operation 10. This is due to missing data related to those parts, which is caused by the fact that the FMS can perform operations on different days. Being our analysis based on data from a particular period of time, some parts started to be machined before the period, and others ended after the period; those operations will not be tracked.

5.5. Model Enhancement

In the previous chapter, the presence of re-work has been highlighted. In this section, the model is corrected by means of the knowledge extracted from Conformance Checking. Since the knowledge extraction process is iterative, at least a second Conformance Checking has to be performed following the model update. The mentioned iteration process is necessary since, by modifying the model, it is possible to find new divergences that were not revealed by the previous Conformance Diagnostic, and thus can bring new possibilities for improvement.

A Petri Net modification tool can be used to update the model. In this case study, the

software WoPeD for the creation of the model is used to upgrade the model. The analysis of the trace views in the previous chapter allows the precise identification of the re-work position within the technological process. In order to introduce the re-work into the model, it is possible to simply change it by adding arcs, transitions and places.

Doing this step, it is important to remember how the construction of loops in the Petri Net affects the possible routes. In the case under consideration, it would be possible to model the re-work as in Figure 5.13, but this would introduce parts that perform re-work more than twice to not be highlighted as incorrect. In addition, in the event in which the machining times of the re-work were different to those performed by the machine in the previous process, this type of modelling would not return the actual machining statistics.

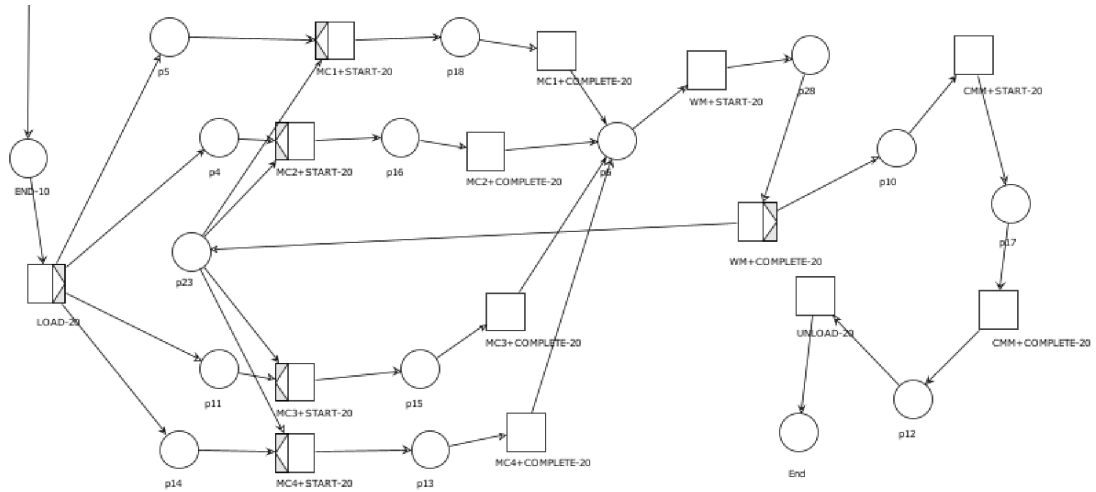


Figure 5.13: Petri Net model with loop rework.

The correct method to integrate the re-work is duplicate the machines, so that the study of timing, routing and possible future changes are tracked. Therefore, the corrected model is the one shown in Figure 5.14.

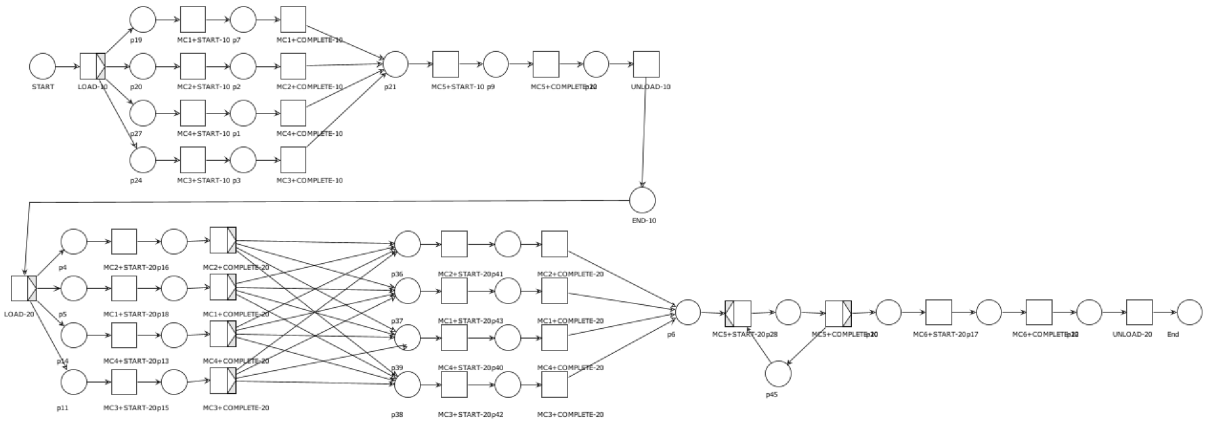


Figure 5.14: Petri Net used to model rework avoiding loops.

Once the model is updated, Conformance Checking can be performed with the same event logs used in the previous analysis and the new model corrected by knowledge extraction. The procedure carried out in the previous section (Section 5.4) is iterated. The results obtained are as follows.

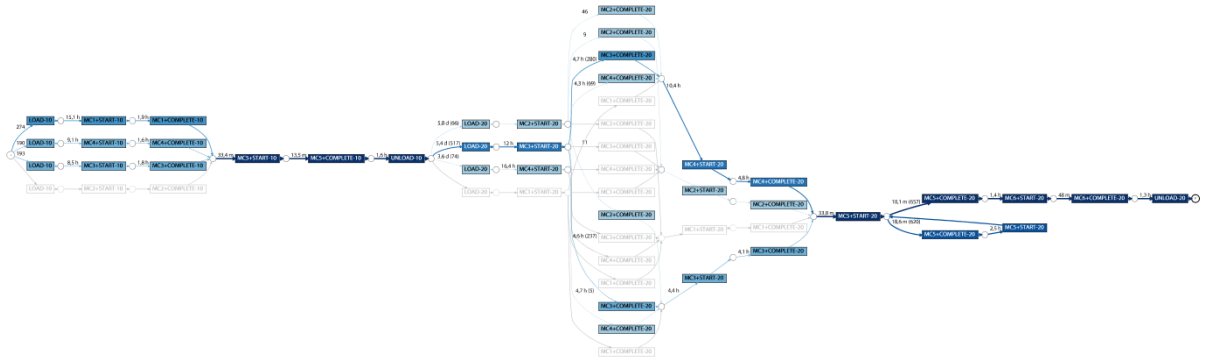


Figure 5.15: Conformance Checking of the product PROD1 X with the updated model.

The Conformance Diagnostic confirms the previously discovered machine-product coupling. The timing and routing for operation 10 turned out to be very similar to the previously discovered, whereas the timing and routing for operation 20 are much different. The difference is due to the fact that the new Conformance Checking is more accurate and closer to reality than the previous one. At this point, the relevant information for studying the performance and for future analyses, such as What-if, are extracted.

Once the first analysis regarding routing and timing is complete, the traces can be studied via *"Toggle Trace View"* visualisation. The ProM trace view is shown in the Figure 5.16:

Figure 5.16: Trace View with model enhanced of the product PROD1 X.

Due to the model update, the majority of the traces of the logs are aligned with the constructed model. Some Wrong and Missing Events are present. They can be studied punctually by taking a single case under consideration at a time. In the following some example are reported.

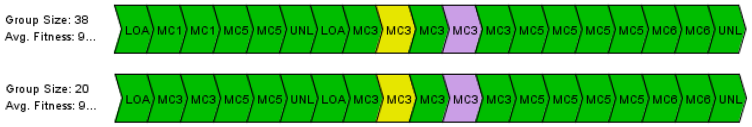


Figure 5.17: Detail of the Trace View of the product PROD1 X.

Taking these two groups of traces as an example (Figure 5.17), it is possible to study specifically the trace followed by the two groups. In particular, there is a double recording of the beginning of the processing and the end of the processing. This type of error is related to the sensor or event recording. Having identified this error, it is possible to trace the machined ID of the parts that are part of this group and then correct the problem if

no assignable cause can be associated with it.

The two groups in Figure 5.18, on the other hand, show double machining in machine MC5 and MC6 prior to the already known re-work. It is then possible to identify the parts that performed these traces and study the specific case in detail.

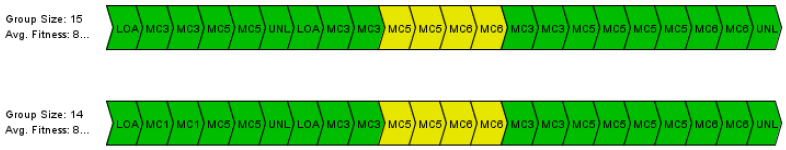


Figure 5.18: Detail of the Trace View of the product PROD1 X, Wrong Events.

A key aspect of this way of proceeding is the possibility to focus on missing events. They are in fact more significant: if a critical operation or a measurement on important features is not performed on a item, it is possible to identify the item that miss the event quickly even if it belongs to a group consisting of only a few items. Figure 5.19 is an example of what has just been explained:

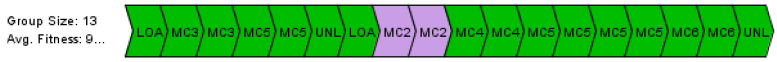


Figure 5.19: Detail of the Trace View of the product PROD1 X, Missing Events.

The model constructed in this way no longer needs to be modified and can be used for Conformance Checking with new data to discover new divergences or to refine parameters such as cycle times, routing or coupling-machine part.

5.6. What-if Analysis

In this last section of the case study, the Jackson Network was used to carry out scenario analyses. The Jackson Network is a particular class of Queuing Networks, suitable to study manufacturing systems, as explained in Section 3.4.1. It, applied to manufacturing systems, allows the computation of key performance indicators such as Lead Time, Waiting Time and saturations of machinery. The Jackson Network was implemented in `Matlab`, with data covering the first quarter of the year 2020, i.e., the time frame between January and March. During this period, as shown by the Dotted Chart 5.4, all the various types of products are processed and for this reason it is a period during which the system is at regime.

According to the theory, it is possible to model the system as a Jackson network – since queue networks are said to be open when customers can enter the network from outside and customers can leave the network after being served/processed. The main hypotheses are:

1. In an open network, jobs arrive from outside following a Poisson process with rate $\alpha > 0$.
2. Each arrival is independently routed to node j with probability $r_{0j} > 0$ (equivalently this can be viewed as each node having an external Poisson stream of job arrivals with rate $\alpha \cdot r_{0j}$).
$$\sum_{j=1}^{M} r_{0j} = 1 \tag{5.1}$$
3. Upon service completion at node i, a job may leave the network with probability r_{0j}.
4. Service discipline at the queues is FIFO.

As explained in section 3.4.1, it is possible to assume that in equilibrium the nodes of the network are independent, each following the distribution of a birth death queue. For this

reason, each machine in the line was patterned as an M/M/1 queue model.

We discarded other models like M/M/m queue or the M/M/1/K queue. The first is surely wrong since we do not have m servers, but each node represents only one machine; the second could be useful since the hypothesis of finite queue capacity is more realistic, but it would lead to an over complex model which is developed on the assumption of the value of K. So is used M/M/1 model since it perfectly fits the data.

Because there are multiple products involved and because each product has different characteristics such as the sequence of operations, the Jackson Network Open Model was constructed as a Multi-Class Network.
$G = 16$ different types of jobs were considered. The division is driven by the type of part (3 different products PROD1, PROD2, PROD3), the weight class of each type of product (2 weight classes named X and Y) and the number of operations performed, shown in Figure 5.6.

Part	**Operation**				
PROD1 X,Y	10	20			
PROD2 X,Y	10	20	40	50	70
PROD3 X,Y	10 i	10 ii			

Table 5.6: Performed operations according to product type.

This subdivision is necessary since the cycle time and waiting time vary as the operation or weight class change: for example, operation 10 of product PROD2 of weight class X has a different cycle time than operation 10 of product PROD2 of weight class Y.

Applying the Jackson Network Multi-Class Model, with G different type of jobs, it is possible to solve the traffic equations separately for each type of jobs g.
Let's define p_{ig} the percentage of jobs of type g among those produced by the machine i. Since all types of jobs must be aggregated into a single class, which averages the behavior of the various jobs, we obtain the following equations:

$$p_{ig} = \frac{\lambda_{ig}}{\sum_{g=1}^{G} \lambda_{ig}} \tag{5.2}$$

$$\lambda_i = \sum_{g=1}^{G} \lambda_{ig} \tag{5.3}$$

$$r_{ij} = \sum_{g=1}^{G} p_{ig} r_{ijg} \tag{5.4}$$

$$\mu_i^{-1}(x_i) = \sum_{g=1}^{G} p_{ig} \mu_{ig}^{-1}(x_{ig}) \tag{5.5}$$

Therefore, for each type of jobs, the model parameters must be estimated: part arrival rate, matrix of probability and cycle time of each machine. Building the model identical to the production system, it is well suited for extracting these parameters quickly, as can be seen from the figure 5.20.

Figure 5.20: Multi-perspective Process Explorer of the product PROD1 X, operation 10.

An example of values extracted from the Conformance Diagnostic in Figure 5.20, and used as input to the Jackson Network are shown in Figure 5.7.

130 u	MC1	MC2	MC3	MC4	MC5	MC6	UNLOAD
LOAD	37.0%		36.0%	27.0%			
MC1					100.0%		
MC2							
MC3					100.0%		
MC4					100.0%		
MC5							100.0%
MC6							

	Cycle Time [h]
MC1	1.7
MC2	
MC3	1.4
MC4	1.8
MC5	0.25
MC6	

Table 5.7: Exported data from the Multi-perspective Process Explorer of the PROD1 X Operation 10 to the Jackson Network.

5.6.1. Verification of the assumptions

The first Open Network assumption to be checked is that the processing time of each machine is exponentially distributed. To test this hypothesis, the histogram of the processing time and arrival time of each machine for each type of jobs is performed. For brevity, only one of the histograms performed is shown.

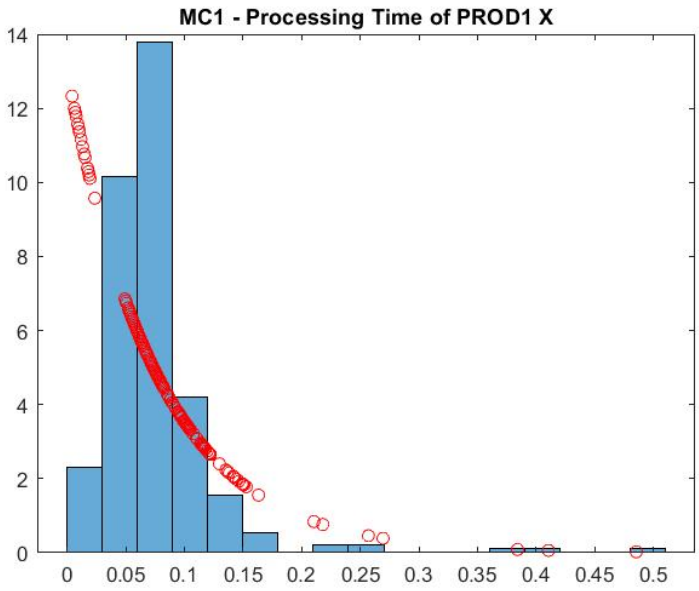


Figure 5.21: Histogram of the Processing Time product PROD1 related to machine MC1. CV is equal to 0.6421.

As shown by the histogram 5.21, the processing time of the machine MC1 related to

the product PROD1 X is not properly distributed as an exponential, in fact the value of Coefficient of Variation (CV), the ratio between standard deviation and mean, is not around 1, as it should be for a perfect exponential distribution, but, in that specific case, it is $CV = 0.6421$.

Many product-form distributions depend on the service time distributions only through their means and hence, general service times can easily be incorporated into Queuing Network models. For instance, even if the service times are not exponential, applying Jackson's Theorem as if they were exponential will often yield the correct product-form distribution.

Insensitivity results still typically require Poisson arrivals, but allowing for general service time distributions offers considerable modeling flexibility.

Instead, the same logic is used to verify the hypothesis that the interarrival time at each machine is exponentially distributed. The histogram **??** is an example of the distribution of the interarrival time.

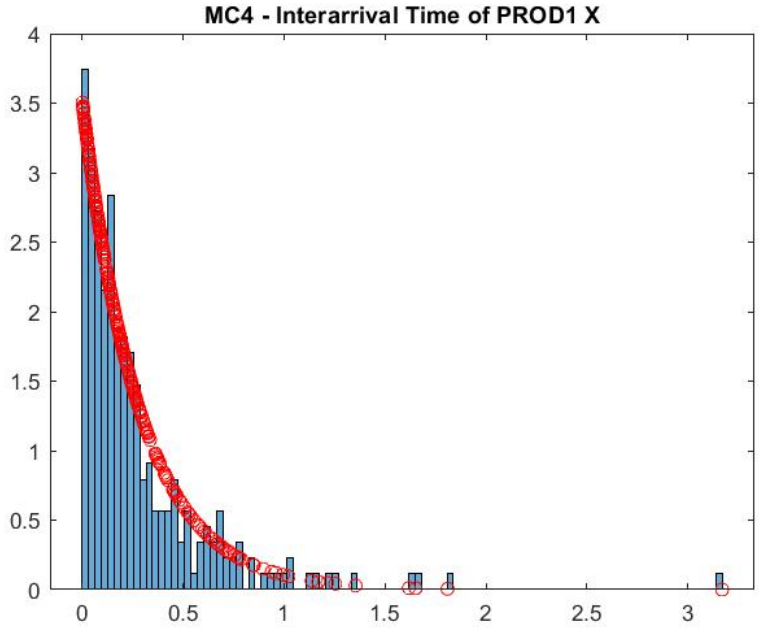


Figure 5.22: Histogram of the Interarrival Time product PROD1 related to machine MC4. CV is equal to 1.2268

Also in this case, the CV of the distribution is calculated. The value obtained is $CV = 1.2268$. It is reasonable to assume that the interarrival time is exponentially distributed.

In general, considering all the machines, the exponential distribution fits most of the interarrival times quite well, while it fits the processing times less well. Since the processing time is not perfectly exponential we expect that all the indicators estimated using M/M/1 queue will be influenced since this model requires exponential times. It must be noticed that in real situations it is quite rare to find perfect exponential distributed times but we know, from state of art, that this model is commonly used and accepted.

Therefore, it was decided to postpone all conclusions to the validation phase where the measured performance will be compared with that estimated by the model to assess the gap between them.

5.6.2. Model Validation

In this section, the values of the saturation, Lead Time and Waiting Time parameters were computed from the event logs and later compared with those computed by the model.

In order to estimate the Lead Time of the system the difference between the time in which the part enters in the system and the time in which the part exit from the system (UNLOAD - LOAD of a single type of jobs) is computed and then averaged. The average Lead Time obtained is 16.42 hours.
The machine saturation ρ is computed as the ratio between production time needed to process all type of jobs and the available time of the plant.
Finally, the Waiting Time is computed as the ratio between the sum of the Waiting Time extracted by the Conformance Checking of ProM of each type of jobs and of each machine, and the total number of parts processed. The results obtained are the following:

Machines	Saturation
MC1	78.94%
MC2	75.99%
MC3	69.30%
MC4	94.60%
MC5	27.33%
MC6	29.36%

Lead Time	16.42 h
Waiting Time	13.43 h

Table 5.8: Performance indicators computed from logs.

Also note that the saturations is always $\rho < 1$; since the saturation of a machine cannot be greater than 1 otherwise the system is not feasible.

In order to validate the model, the results computed from the logs are compared with the results obtained by means of Jackson Network. The results obtained by the Jackson Network model are reported in Figure 5.9.

Machines	Saturation
MC1	77.66%
MC2	74.89%
MC3	67.68%
MC4	92.42%
MC5	26.78%
MC6	28.92%

Lead Time	17.40 h
Waiting Time	14.43 h

Table 5.9: Performance indicators computed from the model.

Is it possible to notice that the four machines that are in parallel are not too unbalanced (MC1, MC2, MC3, MC4). The most saturated machine is machine MC3 with $\rho = 92.42\%$ and $n = 11.26$ parts in queue.

The results obtained by the model and the values computed from the data are reported in the following tables.

Machines	Saturation		% difference
MC1	78.94%	77.66%	1.62%
MC2	75.99%	74.89%	1.45%
MC3	69.30%	67.68%	2.34%
MC4	94.60%	92.42%	2.31%
MC5	27.33%	26.78%	2.02%
MC6	29.36%	28.92%	1.48%

Lead Time	16.42 h	17.40 h	5.97%
Waiting Time	13.43 h	14.43 h	7.45%
	logs	**model**	**% diff**

Table 5.10: Comparison of model and log performance indicators.

Lead time, saturations, and Waiting Time can be now compared between the data and the model. The difference between the model and the data can be due to the fact that the processing and interarrival times are not perfectly exponentially distributed. The difference of the Lead Time is 6.1% and therefore acceptable.
Notice that the ranking of the machines is almost always maintained in the area: the most saturated machine in the area is the same both in data and model, same for the less saturated one.

The comparison in Figure 5.10, confirms the validity of the model built through Jackson Network. This means that it is employable to perform What-if Analysis. It can answer the question "what if...?" and it support managers to better reflect on possible changes of any kind, such as changing routing policy by aiming for flow balancing across the various machines or thinking about adopting a COWIP policy. Therefore, the simulation can be used as a support tool for decision making.

5.6.3. Examples of What-if Analysis

This section is dedicated to some What-if Analysis on strategic production scenarios. The implemented method makes possible to predict system performance by changing some parameters such as cycle time of machinery or routing of products. Since the assumptions for applying the Jackson Network have already been verified, answering "what if..?" questions requires changing the input data of the Jackson Network. In this way, system performance, such as Lead Time, Waiting Time and saturation of the machines,

can be predicted and the Jackson Network can then be used as a tool for decision making. In the following, three different scenarios will be discussed.

1. The first scenario proposed is related to a routing change. Specifically, since the MC3 machine is the one with the lowest saturation value, what if the parts of product PROD1 machined by the machine MC1 are moved to the machine MC4? Figure 5.11 shows the comparison between the current Key Performance Indicator values and those computed by the model related to the scenario SCENARIO1.

Machines	Saturation		% difference
MC1	77.66%	69.45%	-10.57%
MC2	74.89%	74.89%	0.00%
MC3	67.68%	74.87%	10.62%
MC4	92.42%	92.42%	0.00%
MC5	26.78%	26.78%	0.00%
MC6	28.92%	28.92%	0.00%
	Actual	**Scenario 1**	

Lead Time	17.40 h	17.14 h	1.49%
Waiting Time	14.43 h	14.19 h	1.66%
	Actual	**Scenario 1**	**% diff**

Table 5.11: Comparison of actual and modified performance indicators, first scenario.

2. The second scenario proposed is related to the product PROD2 X. In particular, since it is the less-produced one, it is interested to study what if it is not processed. Figure 5.12 shows the comparison between the current Key Performance Indicator values and those computed by the model related to the scenario SCENARIO2.

Machines	Saturation		% difference
MC1	77.66%	64.24%	-17.28%
MC2	74.89%	65.56%	-12.46%
MC3	67.68%	67.68%	0.00%
MC4	92.42%	92.42%	0.00%
MC5	26.78%	23.15%	-13.55%
MC6	28.92%	24.46%	-15.42%
	Actual	**Scenario 2**	

Lead Time	17.40 h	16.95 h	2.53%
Waiting Time	14.43 h	13.89 h	3.74%
	Actual	**Scenario 2**	**% diff**

Table 5.12: Comparison of actual and modified performance indicators, second scenario.

3. The third and last scenario proposed is related to the cycle time of the machine

MC4 and specifically it is focused on the operation 20. The machine MC4 is the one with the highest saturation value and it is probably the bottleneck of the system. Therefore, it can be interested to study what if the processing time of machine MC4 performing operation 20 decreases by the 15%. Figure 5.13 shows the comparison between the current Key Performance Indicator values and those computed by the model related to the scenario SCENARIO3.

Machines	Saturation		% difference
MC1	77.66%	77.66%	0.00%
MC2	74.89%	74.89%	0.00%
MC3	67.68%	67.68%	0.00%
MC4	92.42%	79.53%	-13.95%
MC5	26.78%	26.78%	0.00%
MC6	28.92%	28.92%	0.00%
	Actual	**Scenario 3**	

Lead Time	17.40 h	10.67 h	38.68%
Waiting Time	14.43 h	7.81 h	45.88%
	Actual	**Scenario 3**	**% diff**

Table 5.13: Comparison of actual and modified performance indicators, third scenario.

As can be noticed, the three scenarios are related to different strategic decisions. Specifically, the first and second scenario are of easy implementation business being connected to routing decisions; on the other hand, the third scenario is of more complex implementation since it involves the reduction of the processing time of machine MC4 performing operation 20. The complexity of the latter practice is due to the fact that applying it requires making more radical change decisions. However, the study of the performances points out that the latter practice is the most advantageous compared to the other two. In fact, the average Lead Time and Waiting Time of parts decrease by 38.68These hypothetical scenarios show how the Jackson Network can be used as a tool for decision making. In fact, any kind of change can be simulated through Jackson Network to evaluate the goodness of the decision by studying the variations directly on the system performances.

6 | Comparison of Existing and Proposed Methodology

The purpose of this section is to compare the proposed methodology, Knowledge-based Model Enhancement, with the existing methodology, Process Mining. The section is divided into two parts; the first part compares the results obtained in Chapter 5 with the results obtained by extracting knowledge from a model constructed through Process Discovery. The second part, instead, quantitatively highlights the differences between the two methodologies, by analyzing the results from an experiment conducted with 8 working groups performing the methodologies in parallel.

In order to perform the comparison, the case study is here summarized: After an initial in-depth study of company knowledge and a Data Processing phase, it was possible to construct the model of the production process in the form of Petri Net. This model, referring to product PROD1, is summarized in Figure 5.7. The next step has been Conformance Checking through a plug-in available on the ProM framework. The output of Conformance Checking is called Conformance Diagnostic and it allows comparison between the a priori constructed model and the available event logs. Conformance Diagnostic revealed a misalignment in the traces, identified by a yellow or purple event, which results in partial behavior not mapped by the nominal knowledge of the company.

In the specific case of the PROD1 X product, misalignments are present in most traces during operation 20, as shown in Figure 5.10. The constructed model fits the data with a Fitness value of 82.5%, this value can be increased by trying to find the causes of the

double processing performed in operation 20.
Since the data provider could not be questioned about the reason for this double processing, only hypotheses could be made. The most plausible hypotheses are a rework due to features not respected, such as the tolerance; or the choice of splitting a machining with a cycle time too high into two shorter machining operations to obtain the same result; finally, the last hypothesis is related to the fact that an operation previously carried out on another machine, perhaps not connected to the MES, was moved to the MC3 and MC4 machines, whose data are instead tracked by the MES.

The improvement of the model is the next step in the proposed methodology. Therefore, the model is enhanced. The correct method to integrate the rework is duplicate the machines, in order to track timing information, the routing and possible future changes in a better way. The model for PROD1 X is the one shown in figure 5.14.

At this point, the new model will be the input of a second Conformance Checking, his trace view is reported in figure 5.16. Due to the model update, most of the traces of the logs are aligned with the constructed model. Some Wrong and Missing Events are present. They can be studied punctually by taking a single case under consideration at a time. It is possible to verify that Fitness is increased to 94.3%.

What results would be obtained by applying the *existing methodology* to the case study? The existing methodology emphasizes Process Discovery through event logs without additional information. As explained in Chapter 4, this way of proceeding shows some criticalities in the manufacturing systems.

Tuning several parameters is necessary for model generation. The Inductive Miner algorithm offers a simplified tuning, acting only on the noise threshold. A noise threshold allowing the same Fitness value as the model in figure 5.16 is set to compare the results obtained in the previous chapter. The noise threshold ranges from 0 to 1 and acts as a high-pass filter. It governs the degree of freedom related to infrequent events: the lower its value, the more infrequent events will be considered in the model construction. There-

fore, with a noise threshold value of 0 all events will be considered; instead, with a noise threshold value of 1 no infrequent events will be considered. The model shown in Figure 6.1 was built with a threshold of 0.18.

Figure 6.1: Model of PROD1 X discovered by means of Inductive Miner algorithm with a noise threshold equal to 0.18.

It is important to note that the model returned by Process Discovery does not allow a clear mapping of the actual process. In fact, if studied in detail, mapping errors such as the one explained in Figure 4.6 are present. In addition, the sequence of machining stations related to operation 20 also appears to be nonlinear. Therefore, it is necessary to compare this model with the event logs to better study the divergences. The result of Conformance Checking performed with the latter model is shown in Figure 6.2.

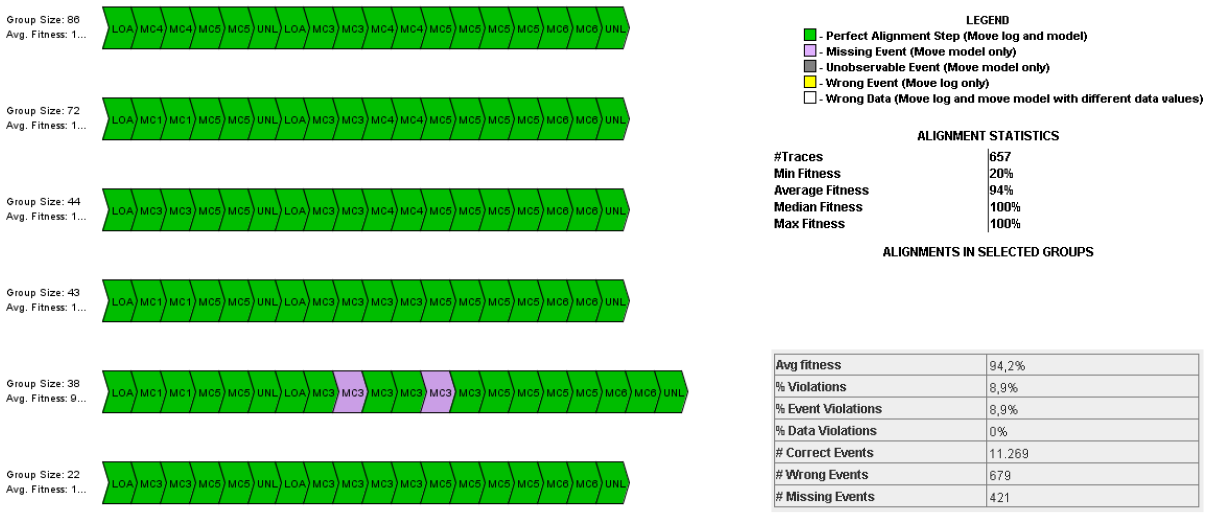


Figure 6.2: Trace View of the model in Figure 6.1.

Traces in Figure 6.2 do not report the rework identified in Section 5.4. What information can be extracted from this last analysis? It was previously reported that all traces in Figure 5.10 show wrong and missing events, which the company should study. The traces in Figure 6.2 do not show any wrong events, and it is difficult to compare these results

with the nominal process, even though the Fitness of the two models is the same. This is a further example of how the proposed methodology is an entirely different approach to enhance the processes with respect to the existing methodology. In fact, it allows to point out divergences that can occur between the nominal process consisting in company knowledge and the production system data. These divergences are the primary source of knowledge updates.

6.1. Experimentation

In order to quantitatively compare how the two methodologies approach the extraction of information in a different manner, the concept behind the Value Stream Map, i.e. the distinction between time spent for value-added and non-value-added activities, is applied. In particular, an experiment is carried out with 8 teams running the two methodologies in parallel. During the experiment, the groups were asked to keep track of the time spent performing the various steps of the methodologies. In this way, it was subsequently possible to construct a map, similar to the Value Stream Map, identifying the activities that added value for the ultimate goal of the methodology, the knowledge extraction, with their respective timeframes. Knowledge extraction is translated into directed questions equal for all the working groups.

The experimentation is developed as follows: 16 engineering students from Politecnico di Milano are divided into groups composed by two people, for a total of 8 groups. Four groups have been dedicated to the execution of the existing methodology, the remaining four to the execution of the proposed methodology. The groups are created in such a way that the results will be as objective as possible and independent of external factors such as previous experience in the field. In fact, each participant had to fill in a pre-test in order to map personal knowledge in the field of Process Mining. The results of the pre-test are used in order to match people with different prior experience, so without advantaged groups, and thus without unbalanced results. In addition, people with no experience of

working in the same team have been paired. The different groups are asked to apply the two methodologies to a simplified, ad-hoc constructed case study. WoPeD software was used to generate the data for the case study. Through the simulation of a production system, 1,952 events, relating to the production of 250 parts in 7 machines, have been generated. The Figure 6.3 shows the Petri Net model used to generate data. It is intended to emphasize that the questions the groups answered aimed to identify known deviations contained in the data.

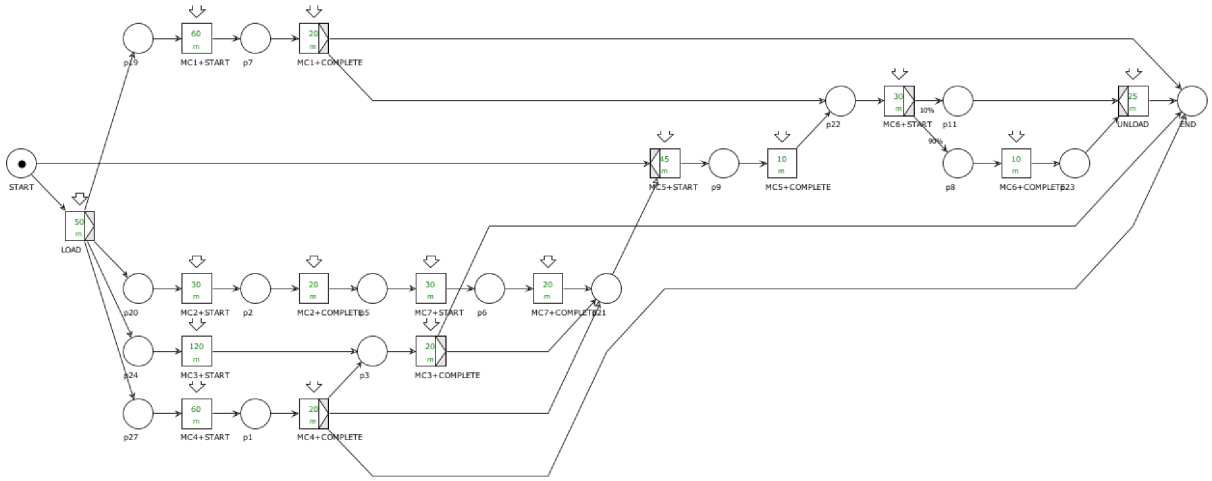


Figure 6.3: Petri Net with noise used to generate event logs.

All groups started the experiment by studying the case study documentation, containing information such as BPMN Diagram, cycle times and company scheduling policies. Afterwards, they were asked to perform different steps depending on the methodology:

- Existing methodology: once the algorithm has been chosen, a phase of selecting the input parameters was necessary to obtain a suitable process model. Subsequently, Conformance Checking was performed to extract information.
- Proposed methodology: Conformance Checking was performed from a given model, which described the nominal knowledge. Finally, again, the Conformance Checking study allowed the extraction of information.

The aim of both groups was to extract information in order to answer targeted questions

about the case under investigation, which were the same for both groups and were intended to simulate the extraction of knowledge from a real system. Once all questions were answered, the group finished the experiment. During the execution, each group had to keep track of the time spent on the different steps. The tasks were clustered as follows:

1. Knowledge study: time dedicated to the study of case study documentation (adding value time).
2. Choice of the algorithm and tuning of the parameters required to build the model (non adding value time).
3. Study of the Petri Net model (non adding value time).
4. Study of the model after the Conformance Checking (adding value time).
5. Study of the traces (adding value time).

The results obtained were summarised by means of a parameter called *accuracy*. This parameter indicates how complete the answer to the questions is: a value of 100% indicates a completely correct answer, a value of 50% indicates a half complete answer. Figure 6.4 shows, dividing by methodology, the average accuracy of each answer. For sake of simplicity, the questions have not been reported as they are not of interest for comparison. Furthermore, Figure 6.5 shows the time taken by the different groups in performing the steps of the methodologies:

	Question 1	Question 2	Question 3	Question 4	Question 5
Existing methodology	100%	71.50%	100%	75%	50%
Proposed methodology	89.50%	92.50%	100%	100%	100%

Figure 6.4: Accuracy of answers, divided by methodology and questions.

Existing methodology

	Task	1	2	3	4	5	6	7	8	9	10	11	12	13
Group 1	5											x	x	
	4								x	x	x			
	3					x	x	x						
	2		x	x	x									
	1	x												
Group 2	5												x	x
	4										x	x		
	3								x	x				
	2				x	x	x	x						
	1	x	x	x										
Group 3	5								x	x				
	4										x	x		
	3					x	x	x						
	2			x	x									
	1	x	x											
Group 4	5								x	x	x			
	4							x				x	x	
	3					x	x							
	2		x	x	x									
	1	x												

Proposed methodology

	Task	1	2	3	4	5	6	7	8	9	10	11	12	13
Group 5	5				x		x	x						
	4		x	x										
	3					x								
	2													
	1	x												
Group 6	5					x	x	x	x					
	4				x	x								
	3		x											
	2													
	1	x												
Group 7	5					x	x	x						
	4				x	x								
	3		x											
	2													
	1	x												
Group 8	5							x	x					
	4				x	x								
	3			x										
	2													
	1	x	x											

Figure 6.5: Time distribution of each group, each column correspond to 5 minutes of work.

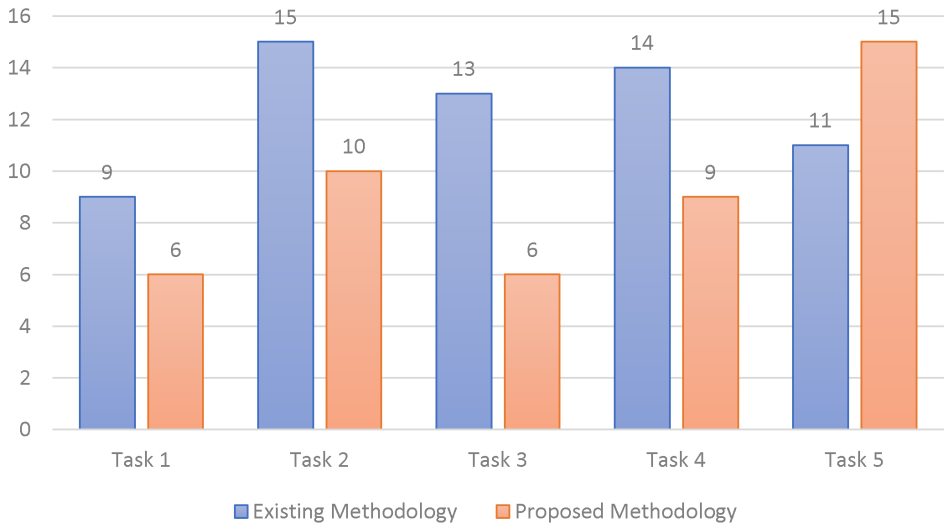


Figure 6.6: Bar graph of time spent in minutes on each task.

First of all, it is possible to see how accuracy in answers is generally higher for the proposed methodology. This is due to the fact that the Conformance Checking done with

the knowledge-based model returns information more clearly, allowing a more complete overview of the process. Subsequently, Figure 6.6 highlight how the average time taken to perform the different tasks is different: on average, the groups performing the existing methodology took 61 minutes, compared to 36 minutes for the proposed methodology.

Note that in the proposed methodology, the Petri Net of the knowledge-based model is provided and it is not constructed by the groups. This is because the construction, although trivial, requires the use of a software that can be complicated when first used. The average time for constructing the Petri Net for the case under consideration is reasonably assumed to be around 10 minutes.

At this point, it is possible to build the Value Stream Map of the two methodologies; the map in Figure 6.7 summarizes the steps executed during the experimentation, and classifies them according to whether or not they add value to corporate knowledge.

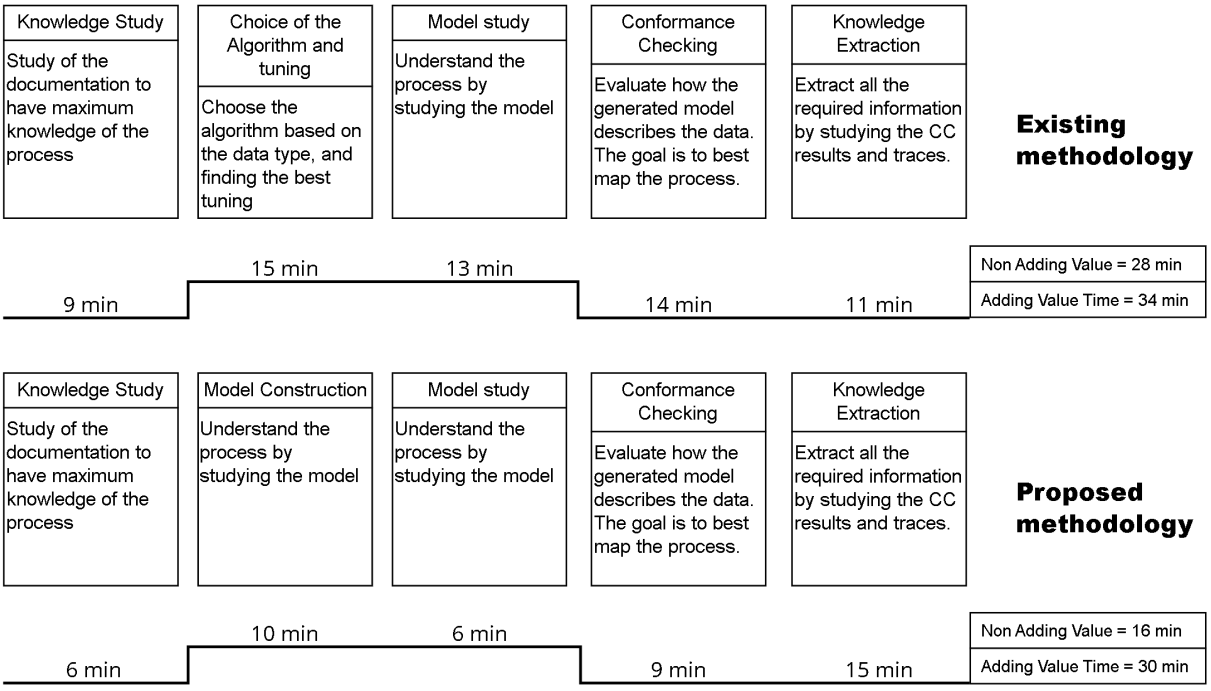


Figure 6.7: Value Stream Map of the two methodology.

By analysing the Figure 6.7, it can be noticed how both methods have similar adding value times, but different non-adding value times. According to the same adding value

time, Figure 6.4 shows how the proposed methodology has a more detailed and complete information extraction capability. It can be seen that in the proposed methodology, less time was spent on the study of knowledge and on the study of the model. This is because, as the model reflects the information available, there are no discrepancies between the model and the nominal information. These conclusions should be added to the previous advantages discussed in Section 4.2, such as having a model that can be used to perform What-if Analysis.

In conclusion, the comprehensiveness of the answers from the students who worked on the different methodologies also greatly supported the thesis that starting from a knowledge-based model of the process leads to a greater ability to extract relevant information. Indeed, groups that worked on the proposed methodology provided generally more complete and detailed answers than groups that worked on the existing methodology. Moreover, during the experimentation, the groups dedicated to the existing methodology needed more support for the execution of the various steps and, in particular, for the step concerning the tuning of the algorithm parameters.

7 | Conclusions and Further Improvements

The increasing digitization has enabled recording of a vast amount of process data in many fields, especially in production systems. Data collection allows the integration of Process Mining techniques into production systems, whose purpose is to discover, monitor, and improve processes. The application of Process Mining to production systems has highlights some criticalities like the unnecessary of Process Discovery since every company has structured and available knowledge. In fact, Process Discovery is fundamental when the flow of instances is unknown, which is the case of almost all business applications, but it is not the case of the manufacturing systems.

This study developed a novel methodology for knowledge-based model enhancement through Conformance Checking techniques in manufacturing systems. The knowledge-based approach makes it possible to avoid the main criticalities of Process Mining and offers more significant opportunities for studying process data, allowing the company to align production and the engineering department continuously.

To validate the work, the existing and the proposed methodology have been applied to a real case study and through an experiment conducted by 8 working teams. The case study shows the application of the proposed methodology to a real industry. The results obtained prove that the a priori model construction enables the same fitness results as the old methodology, but the information extraction is more detailed and comprehensive. The case study highlights two other beneficial aspects. The first one is the rigorous iden-

tification of all the information not explained by the model, such as Recurrent Patterns that are the primary source of knowledge upgrades and, consequently, of Model Enhancement; instead, Outliers will be able to be identified and then studied individually. The second advantage is the possibility of integrating the methodology with forward-looking techniques allowing the application of the new methodology at a strategic level.

The main results obtained from the experiment confirm those obtained from the case study; in fact, the new approach requires less time and it allows more detailed and complete information to be extracted. Furthermore, the experiment allowed the construction of a map identifying the tasks that add value to the ultimate goal of the methodology, i.e. knowledge extraction, with their respective timeframes. It shows that the percentage of time spent on value-adding tasks is higher in the proposed methodology (65%) with respect to the existing one (55%).

In conclusion, the knowledge-based model enhancement through Conformance Checking will provide significant benefits in the manufacturing systems allowing to highlight the divergences between the nominal process and data from the production system resulting in continuous alignment between system digital model and real physical system. The proposed methodology could further be improved by taking care of different aspects. Particularly, it would be interesting to:

1. Automatically translate BPMN into Petri nets through ad-hoc tools. This possibility is very interesting from a business point of view, as companies widely adopt the BPMN language due to its simplicity of construction.

2. Automatic integration of extracted knowledge into business documentation, such as flow charts or BPMN.

3. Automatic integration of the Petri Net model with tools for performing what-if analyses. The work shows how the Petri Net model used for knowledge extraction can be directly translated and simulated via Jackson Network. This integration

could be automated.

Bibliography

[1] Guideline industrie 4.0. guiding principles for the implementation of industrie 4.0 in small and medium sized businesses. https://www.vdma.org/documents/34570/0/Guide%20to%20Industrie%204.0.pdf/1ca94350-1631-465a-e73c-47cd1d5b412d.

[2] Ocel standard. http://ocel-standard.org/).

[3] Trends in global export value of trade in goods from 1950 to 2021. https://www.statista.com/statistics/264682/worldwide-export-volume-in-the-trade-since-1950/.

[4] OCEL: A standard for object-centric event logs.

[5] W. Aalst. *Process Mining: Data Science in Action*. 01 2016. ISBN 9783662498507. doi: 10.1007/978-3-662-49851-4.

[6] A. Albu. Logical inference modeled by petri nets. pages 137–140, 05 2016. doi: 10.1109/SACI.2016.7507358.

[7] P. Azema, G. Juanole, E. Sanchis, and M. Montbernard. Specification and verification of distributed systems using prolog interpreted petri nets. In *Proceedings of the 7th International Conference on Software Engineering*, ICSE '84, page 510–518. IEEE Press, 1984. ISBN 0818605286.

[8] T. Baier, J. Mendling, and M. Weske. Bridging abstraction layers in process mining. *Information Systems*, 46:123–139, 2014. ISSN 0306-4379. doi: https://doi.org/10.

1016/j.is.2014.04.004. URL https://www.sciencedirect.com/science/article/pii/S0306437914000714.

[9] N. P. Ballambettu, M. A. Suresh, and R. Bose. Analyzing process variants to understand differences in key performance indices. In *International Conference on Advanced Information Systems Engineering*, pages 298–313. Springer, 2017.

[10] F. Baskett, K. M. Chandy, R. R. Muntz, and F. G. Palacios. Open, closed, and mixed networks of queues with different classes of customers. *Journal of the ACM (JACM)*, 22(2):248–260, 1975.

[11] G. R. Bitran and S. Dasu. A review of open queueing network models of manufacturing systems. *Queueing systems*, 12(1):95–133, 1992.

[12] M. Borkowski, W. Fdhila, M. Nardelli, S. Rinderle-Ma, and S. Schulte. Event-based failure prediction in distributed business processes. *Information Systems*, 81:220–235, 2019. ISSN 0306-4379. doi: https://doi.org/10.1016/j.is.2017.12.005. URL https://www.sciencedirect.com/science/article/pii/S0306437917300030.

[13] P. Buchholz. Hierarchical high level petri nets for complex system analysis. In R. Valette, editor, *Application and Theory of Petri Nets 1994*, pages 119–138, Berlin, Heidelberg, 1994. Springer Berlin Heidelberg. ISBN 978-3-540-48462-2.

[14] A. Burattin, A. Sperduti, and M. Veluscek. Business models enhancement through discovery of roles. In *2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pages 103–110. IEEE, 2013.

[15] J. Cardoso, R. Valette, and D. Dubois. Fuzzy petri nets: An overview. *IFAC Proceedings Volumes*, 29(1):4866–4871, 1996. ISSN 1474-6670. doi: https://doi.org/10.1016/S1474-6670(17)58451-7. URL https://www.sciencedirect.com/science/article/pii/S1474667017584517. 13th World Congress of IFAC, 1996, San Francisco USA, 30 June - 5 July.

[16] J. Carmona, B. van Dongen, A. Solti, and M. Weidlich. Conformance checking. *Switzerland: Springer.[Google Scholar]*, 2018.

[17] L. Cattaneo, L. Fumagalli, M. Macchi, and E. Negri. Clarifying data analytics concepts for industrial engineering. 51(11):820–825. ISSN 24058963. doi: 10.1016/j.ifacol.2018.08.440. URL `https://linkinghub.elsevier.com/retrieve/pii/S2405896318315672`.

[18] C. Chang, Y.-F. Chang, C.-C. Song, and M. Aoyama. Integral: Petri-net approach to distributed software development. *Information and Software Technology*, 31(10):535–545, 1989. ISSN 0950-5849. doi: https://doi.org/10.1016/0950-5849(89)90175-4. URL `https://www.sciencedirect.com/science/article/pii/0950584989901754`.

[19] M.-S. Chen, J. Han, and P. Yu. Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6):866–883, 1996. doi: 10.1109/69.553155.

[20] P. P. S. Chen. The entity-relationship model—toward a unified view of data.

[21] C. CHOI, C. KIM, and C. KIM. Towards sustainable environmental policy and management in the fourth industrial revolution: Evidence from big data analytics.

[22] M. Chu Zhou, F. DiCesare, and D. Rudolph. Control of a flexible manufacturing system using petri nets. *IFAC Proceedings Volumes*, 23(8, Part 5):47–52, 1990. ISSN 1474-6670. doi: https://doi.org/10.1016/S1474-6670(17)51710-3. URL `https://www.sciencedirect.com/science/article/pii/S1474667017517103`. 11th IFAC World Congress on Automatic Control, Tallinn, 1990 - Volume 5, Tallinn, Finland.

[23] G. Clark. The industrial revolution. page 67.

[24] C. dos Santos Garcia, A. Meincheim, E. R. Faria Junior, M. R. Dallagassa, D. M. V. Sato, D. R. Carvalho, E. A. P. Santos, and E. E. Scalabrin. Process mining

techniques and applications – a systematic mapping study. *Expert Systems with Applications*, 133:260–295, 2019. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2019.05.003. URL `https://www.sciencedirect.com/science/article/pii/S0957417419303161`.

[25] M. Drakaki and P. Tzionas. A colored petri net-based modeling method for supply chain inventory management. *Simulation*, 98(3):257–271, 2022.

[26] S. Dunzer, M. Stierle, M. Matzner, and S. Baier. Conformance checking: a state-of-the-art literature review. In *Proceedings of the 11th international conference on subject-oriented business process management*, pages 1–10, 2019.

[27] M. L. v. Eck, N. Sidorova, and W. M. van der Aalst. Discovering and exploring state-based models for multi-perspective processes. In *International Conference on Business Process Management*, pages 142–157. Springer, 2016.

[28] J. Ezpeleta, J. Colom, and J. Martinez. A petri net based deadlock prevention policy for flexible manufacturing systems. *IEEE Transactions on Robotics and Automation*, 11(2):173–184, 1995. doi: 10.1109/70.370500.

[29] D. Fahland and W. M. van Der Aalst. Model repair—aligning process models to reality. *Information Systems*, 47:220–243, 2015.

[30] H. J. Genrich and K. Lautenbach. The analysis of distributed systems by means of predicate/transition-nets. In G. Kahn, editor, *Semantics of Concurrent Computation*, pages 123–146, Berlin, Heidelberg, 1979. Springer Berlin Heidelberg. ISBN 978-3-540-35163-4.

[31] V. C. Gerogiannis, A. D. Kameas, and P. E. Pintelas. Comparative study and categorization of high-level petri nets. *Journal of Systems and Software*, 43(2): 133–160, 1998.

[32] O. M. Group. Omg unified modeling language 2.5. omg.

[33] M. Hain, H. Moutachaouik, A. Zakrani, and A. Enaanai. A new approach to MES system deployment.

[34] Y. Han, C. Jiang, and X. Luo. A study of concurrency control in web-based distributed real-time database system using extended time petri nets. In *7th International Symposium on Parallel Architectures, Algorithms and Networks, 2004. Proceedings.*, pages 67–72, 2004. doi: 10.1109/ISPAN.2004.1300459.

[35] B. N. A. Hidayat, A. P. Kurniati, et al. Process model extension using heuristics miner:(case study: Incident management of volvo it belgium). In *2016 International Conference on Computational Intelligence and Cybernetics*, pages 73–78. IEEE, 2016.

[36] G. T. S. Ho and H. C. W. Lau. Development of an olap-fuzzy based process mining system for quality improvement. In Z. Shi, K. Shimohara, and D. Feng, editors, *Intelligent Information Processing III*, pages 243–258, Boston, MA, 2007. Springer US. ISBN 978-0-387-44641-7.

[37] Z. Huang, W. Dong, P. Bath, L. Ji, and H. Duan. On mining latent treatment patterns from electronic medical records. *Data mining and knowledge discovery*, 29 (4):914–949, 2015.

[38] G. S. Hura, H. Singh, and N. K. Nanda. Some design aspects of databases through petri net modeling. *IEEE Transactions on Software Engineering*, SE-12(4):505–510, 1986. doi: 10.1109/TSE.1986.6312897.

[39] V. Huser. Process mining: Discovery, conformance and enhancement of business processes, 2012.

[40] J. R. Jackson. Networks of waiting lines. *Operations research*, 5(4):518–521, 1957.

[41] J. R. Jackson. Jobshop-like queueing systems. *Management science*, 10(1):131–142, 1963.

[42] P. Juneja, D. Kundra, and A. Sureka. Anvaya: An algorithm and case-study on improving the goodness of software process models generated by mining event-log data in issue tracking systems. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, volume 1, pages 53–62, 2016. doi: 10.1109/COMPSAC.2016.64.

[43] A. Kalenkova, W. Aalst, I. Lomazova, and V. Rubin. Process mining using bpmn: Relating event logs and process models // process mining using bpmn. relating event logs and process models. *Software and Systems Modeling*, pages 1–30, 01 2015.

[44] W. D. Kelton. *Simulation with ARENA*. McGraw-hill, 2002.

[45] M. Khabbazi, N. Ismail, M. Y. Ismail, and S. Mousavi. Data modeling of traceability information for manufacturing control system.

[46] J. P. Kleijnen and W. van Groenendaal. *Simulation: a statistical perspective*. John Wiley & Sons, Inc., 1992.

[47] P. Lade, R. Ghosh, and S. Srinivasan. Manufacturing analytics and industrial internet of things. 32(3):74–79. ISSN 1541-1672. doi: 10.1109/MIS.2017.49. URL `http://ieeexplore.ieee.org/document/7933925/`.

[48] G. T. Lakshmanan, D. Shamsi, Y. N. Doganata, M. Unuvar, and R. Khalaf. A markov prediction model for data-driven semi-structured business processes. *Knowledge and Information Systems*, 42(1):97–126, 2015.

[49] S. Lee, S. J. Nam, and J.-K. Lee. Real-time data acquisition system and HMI for MES. .

[50] S. Lee, S. J. Nam, and J.-K. Lee. Real-time data acquisition system and HMI for MES. .

[51] S. J. J. Leemans, D. Fahland, and W. M. van der Aalst. Discovering block-structured process models from event logs - a constructive approach. In *Petri Nets*, 2013.

[52] A. M. Lemos, C. C. Sabino, R. M. Lima, and C. A. Oliveira. Using process mining in software development process management: A case study. In *2011 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1181–1186. IEEE, 2011.

[53] M. Li, L. Liu, L. Yin, and Y. Zhu. A process mining based approach to knowledge maintenance. *Information Systems Frontiers*, 13(3):371–380, 2011.

[54] G. Liu, M. Zhou, and C. Jiang. Petri net models and collaborativeness for parallel processes with resource sharing and message passing. *ACM Trans. Embed. Comput. Syst.*, 16(4), may 2017. ISSN 1539-9087. doi: 10.1145/2810001. URL `https://doi.org/10.1145/2810001`.

[55] D. Lukic, A. Antic, S. Borojevic, M. Jocanovic, and I. Kuric. Evaluation of the technological effects of application of the FMS elements.

[56] J. Luo, K. Tan, H. Luo, and M. Zhou. Inference approach based on petri nets. *Information Sciences*, 547:1008–1024, 2021. ISSN 0020-0255. doi: https://doi.org/10.1016/j.ins.2020.09.023. URL `https://www.sciencedirect.com/science/article/pii/S0020025520309324`.

[57] F. J. S. Magallanes. Technologies of industry 4.0 and the impact on the new generation of manufacturing execution system: A systematic literature review. page 79.

[58] M. C. Magnanini and T. A. Tolio. Performance evaluation of asynchronous two-stage manufacturing lines fabricating discrete parts. *CIRP Journal of Manufacturing Science and Technology*, 33:488–505, 2021.

[59] F. Mannhardt, M. de Leoni, and H. Reijers. The multi-perspective process explorer. volume 1418, 08 2015.

[60] F. Mannhardt, M. de Leoni, H. A. Reijers, and W. M. P. van der Aalst. Data-driven process discovery - revealing conditional infrequent behavior from event logs. In

E. Dubois and K. Pohl, editors, *Advanced Information Systems Engineering*, pages 545–560, Cham, 2017. Springer International Publishing. ISBN 978-3-319-59536-8.

[61] R. S. Mans, W. M. Van der Aalst, and R. J. Vanwersch. *Process mining in healthcare: evaluating and exploiting operational healthcare processes*. Springer, 2015.

[62] A. Medeiros, A. Weijters, and W. Aalst. Genetic process mining: An experimental evaluation. *Data Mining and Knowledge Discovery*, 14:245–304, 04 2007. doi: 10.1007/s10618-006-0061-7.

[63] J. Medina, X. Li, J. Corona Armenta, M. Montufar-Benítez, M. Oscar, and A. Pérez-Rojas. A petri net model for an active database simulator. pages 431–437, 01 2008.

[64] A. Mehrez, M. Muzumdar, W. Acar, and G. Weinroth. A petri net model view of decision making: an operational management analysis. *Omega*, 23(1):63–78, 1995. ISSN 0305-0483. doi: https://doi.org/10.1016/0305-0483(94)00049-G. URL `https://www.sciencedirect.com/science/article/pii/030504839400049G`.

[65] A. Meincheim, C. d. S. Garcia, J. C. Nievola, and E. E. Scalabrin. Combining process mining with trace clustering: Manufacturing shop floor process - an applied case. In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 498–505, 2017. doi: 10.1109/ICTAI.2017.00082.

[66] L. Mekly and S. Yau. Software design representation using abstract process networks. *IEEE Transactions on Software Engineering*, SE-6(5):420–435, 1980. doi: 10.1109/TSE.1980.230490.

[67] S. Milinković, M. Marković, S. Vesković, M. Ivić, and N. Pavlović. A fuzzy petri net model to estimate train delays. *Simulation Modelling Practice and Theory*, 33:144–157, 2013. ISSN 1569-190X. doi: https://doi.org/10.1016/j.simpat.2012.12.005. URL `https://www.sciencedirect.com/science/article/pii/S1569190X12001700`. EUROSIM 2010.

[68] M. Misita and D. D. Milanovic. MANAGING ECONOMIC GROWTH ON THE BASIS OF NATIONAL PRODUCT QUALITY IN THE CONDITIONS OF INDUSTRY 4.0.

[69] J. Munoz-Gama and I. Echizen. Insuring sensitive processes through process mining. In *2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing*, pages 447–454, 2012. doi: 10.1109/UIC-ATC.2012.83.

[70] T. Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580, 1989.

[71] R. W. Nielsen. Explaining the mechanism of growth in the past two million years – vol. i. page 86.

[72] F. Paster and E. Helm. From ihe audit trails to xes event logs facilitating process mining. *Studies in health technology and informatics*, 210:40–4, 05 2015. doi: 10.3233/978-1-61499-512-8-40.

[73] S. A. Priyambada, E. Mahendrawathi, and B. N. Yahya. Curriculum assessment of higher educational institution using aggregate profile clustering. *Procedia Computer Science*, 124:264–273, 2017. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2017.12.155. URL `https://www.sciencedirect.com/science/article/pii/S187705091732923X`. 4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia.

[74] M. Prodel, V. Augusto, X. Xie, B. Jouaneton, and L. Lamarsalle. Discovery of patient pathways from a national hospital database using process mining and integer linear programming. In *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 1409–1414, 2015. doi: 10.1109/CoASE.2015.7294295.

[75] M. Rawson and M. Rawson. Petri nets for concurrent programming, 2022. URL https://arxiv.org/abs/2208.02900.

[76] Á. Rebuge and D. R. Ferreira. Business process analysis in healthcare environments: A methodology based on process mining. *Information systems*, 37(2):99–116, 2012.

[77] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro. Process mining in healthcare: A literature review. *Journal of Biomedical Informatics*, 61:224–236, 2016. ISSN 1532-0464. doi: https://doi.org/10.1016/j.jbi.2016.04.007. URL https://www.sciencedirect.com/science/article/pii/S1532046416300296.

[78] A. Rozinat and W. M. Van der Aalst. Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1):64–95, 2008.

[79] A. Rozinat, I. S. M. de Jong, C. W. GÜnther, and W. M. P. v. der Aalst. Process mining applied to the test process of wafer scanners in asml. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 39(4):474–479, 2009. doi: 10.1109/TSMCC.2009.2014169.

[80] A. Rozinat, I. Jong, C. Gunther, and W. Aalst. Process mining applied to the test process of wafer scanners in asml. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 39:474 – 479, 08 2009. doi: 10.1109/TSMCC.2009.2014169.

[81] E. Ruschel, E. A. P. Santos, and E. de Freitas Rocha Loures. Mining shop-floor data for preventive maintenance management: Integrating probabilistic and predictive models. *Procedia Manufacturing*, 11:1127–1134, 2017. ISSN 2351-9789. doi: https://doi.org/10.1016/j.promfg.2017.07.234. URL https://www.sciencedirect.com/science/article/pii/S2351978917304420. 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy.

[82] B. B. Sarkar and N. Chaki. Modeling analysis of transaction management for

distributed database environment using petri nets. In *2009 World Congress on Nature Biologically Inspired Computing (NaBIC)*, pages 918–923, 2009. doi: 10.1109/NABIC.2009.5393869.

[83] D. M. V. Sato, S. C. De Freitas, J. P. Barddal, and E. E. Scalabrin. A survey on concept drift in process mining. *ACM Computing Surveys (CSUR)*, 54(9):1–38, 2021.

[84] T. M. T. Halpin. Information modeling and relational databases.

[85] D. Tabak and A. H. Levis. Petri net representation of decision models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-15(6):812–818, 1985. doi: 10.1109/TSMC.1985.6313468.

[86] T. A. M. Tolio and M. C. Magnanini. Manufacturing systems engineering course lecture slides. 2021-2022.

[87] W. van der Aalst. Putting high-level petri nets to work in industry. *Computers in Industry*, 25(1):45–54, 1994. ISSN 0166-3615. doi: https://doi.org/10.1016/0166-3615(94)90031-0. URL `https://www.sciencedirect.com/science/article/pii/0166361594900310`.

[88] W. Van Der Aalst. Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems (TMIS)*, 3(2):1–17, 2012.

[89] W. van der Aalst and et al. Process mining manifesto. In F. Daniel, K. Barkaoui, and S. Dustdar, editors, *Business Process Management Workshops*, pages 169–194, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-28108-2.

[90] W. van der Aalst, T. Weijters, and L. Maruster. Workflow mining: discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128–1142, 2004. doi: 10.1109/TKDE.2004.47.

[91] W. M. van der Aalst. Process mining and simulation: A match made in heaven! In *SummerSim*, pages 4–1, 2018.

[92] W. M. van der Aalst. A practitioner's guide to process mining: Limitations of the directly-follows graph. *Procedia Computer Science*, 164:321–328, 2019. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2019.12.189. URL `https://www.sciencedirect.com/science/article/pii/S1877050919322367`. CENTERIS 2019 - International Conference on ENTERprise Information Systems / ProjMAN 2019 - International Conference on Project MANagement / HCist 2019 - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN/HCist 2019.

[93] W. M. van der Aalst, T. Brockhoff, A. F. Ghahfarokhi, M. Pourbafrani, M. S. Uysal, and S. J. v. Zelst. Removing operational friction using process mining: challenges provided by the internet of production (iop). In *International Conference on Data Management Technologies and Applications*, pages 1–31. Springer, 2020.

[94] W. M. P. van der Aalst. Object-centric process mining: Dealing with divergence and convergence in event data.

[95] W. M. P. van der Aalst. Distributed process discovery and conformance checking. In J. de Lara and A. Zisman, editors, *Fundamental Approaches to Software Engineering*, pages 1–25, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-28872-2.

[96] W. M. P. van der Aalst, T. Brockhoff, A. F. Ghahfarokhi, M. Pourbafrani, M. S. Uysal, and S. J. van Zelst. Removing operational friction using process mining: Challenges provided by the internet of production (IoP).

[97] H. M. W. Verbeek, J. C. A. M. Buijs, B. F. van Dongen, and W. M. P. van der Aalst. XES, XESame, and ProM 6.

[98] J. Váně, F. Kalvas, and J. Basl. Engineering companies and their readi-

ness for industry 4.0. 70(5):1072–1091. ISSN 1741-0401. doi: 10.1108/IJPPM-06-2020-0318. URL https://www.emerald.com/insight/content/doi/10.1108/IJPPM-06-2020-0318/full/html.

[99] S. Waschull, J. C. Wortmann, and J. A. C. Bokhorst. Manufacturing execution systems: The next level of automated control or of shop-floor support? In *Advances in Production Management Systems. Smart Manufacturing for Industry 4.0*, pages 386–393. Springer International Publishing.

[100] A. Weijters, W. Aalst, and A. Medeiros. *Process Mining with the Heuristics Miner-algorithm*, volume 166. 01 2006.

[101] M. Witsch and B. Vogel-Heuser. Towards a formal specification framework for manufacturing execution systems.

[102] M. T. Wynn and S. Sadiq. Responsible process mining - a data quality perspective. In T. Hildebrandt, B. F. van Dongen, M. Röglinger, and J. Mendling, editors, *Business Process Management*, pages 10–15, Cham, 2019. Springer International Publishing. ISBN 978-3-030-26619-6.

[103] W. Yao and X. He. Mapping petri nets to concurrent programs in cc++. *Information and Software Technology*, 39(7):485–495, 1997. ISSN 0950-5849. doi: https://doi.org/10.1016/S0950-5849(97)00006-2. URL https://www.sciencedirect.com/science/article/pii/S0950584997000062.

[104] F. Yasmin, F. Bukhsh, and P. Silva. Process enhancement in process mining: A literature review. 12 2018.

[105] F. A. Yasmin, F. A. Bukhsh, and P. D. A. Silva. Process enhancement in process mining: A literature review. In *CEUR workshop proceedings*, volume 2270, pages 65–72. Rheinisch Westfälische Technische Hochschule, 2018.

[106] S. Yau and M. Caglayan. Distributed software system design representation using

modified petri nets. *IEEE Transactions on Software Engineering*, SE-9(6):733–745, 1983. doi: 10.1109/TSE.1983.235581.

[107] B. Zhou, S. Wang, and L. Xi. Data model design for manufacturing execution system. 16. ISSN 1741-038X.

[108] M. Zhou and K. Venkatesh. *Modeling, Simulation, and Control of Flexible Manufacturing Systems*. WORLD SCIENTIFIC, 1999. doi: 10.1142/3376. URL https://www.worldscientific.com/doi/abs/10.1142/3376.

A | Appendix A

The main hypothesis of Open Network to be checked is that interarrival time and processing time of each machine are exponentially distributed. To test these hypothesis, the histogram of the arrival time and processing time of each machine for each type of jobs is performed. In the following only the histograms related to the product PROD1 X are reported for synthesis.

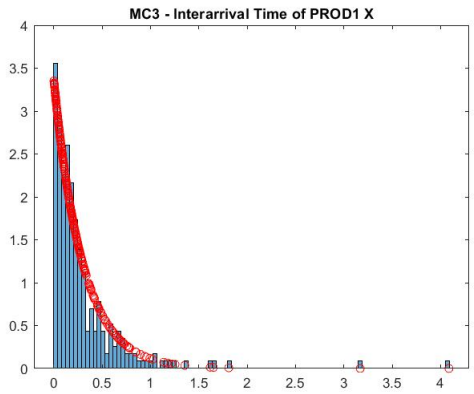


Figure A.1: Histogram of the Interarrival Time product PROD1 X related to machine MC3. CV is equal to 1.4.

As can be seen from the reported histogram the hypothesis of exponential distribution of interarrival time and exponential time is respected.

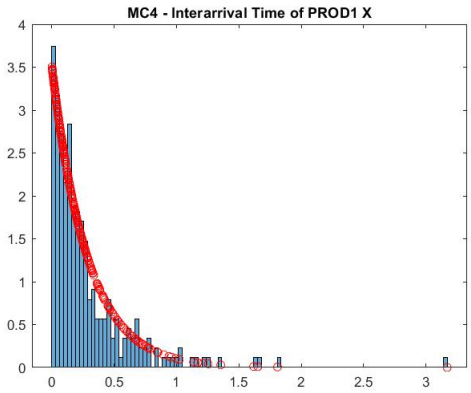


Figure A.2: Histogram of the Interarrival Time product PROD1 X related to machine MC4. CV is equal to 1.2268.

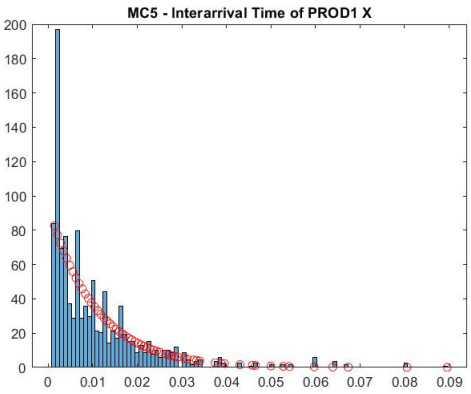


Figure A.3: Histogram of the Interarrival Time product PROD1 X related to machine MC5. CV is equal to 1.065.

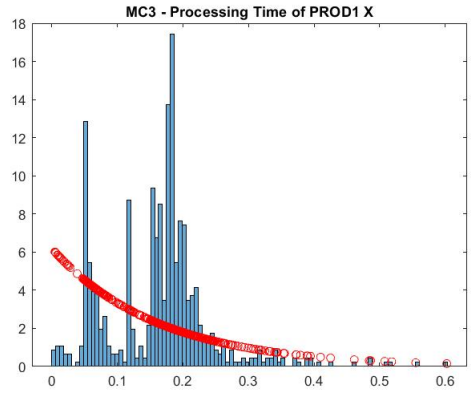


Figure A.4: Histogram of the Processing Time product PROD1 X related to machine MC3. CV is equal to 0.5.

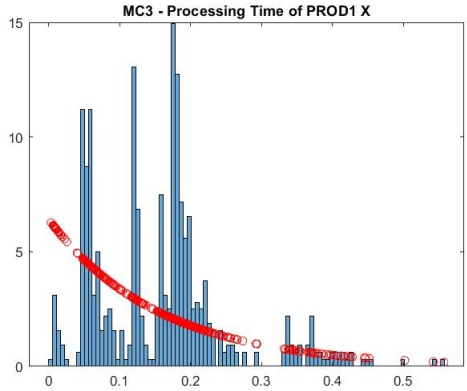


Figure A.5: Histogram of the Processing Time product PROD1 X related to machine MC1. CV is equal to 0.6061.

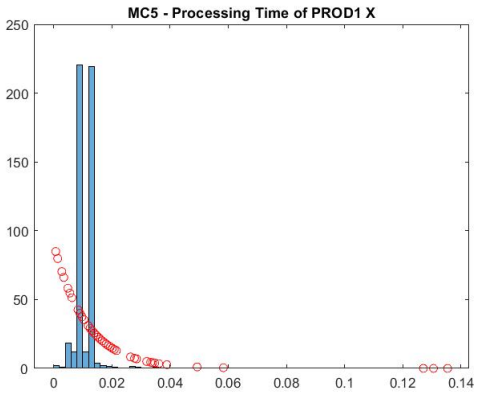


Figure A.6: Histogram of the Processing Time product PROD1 X related to machine MC5. CV is equal to 0.6015.

List of Figures

List of Tables

Acknowledgements

First of all, we would like to thank our advisor Prof. Tullio Tolio and our co-advisor, Dr. Maria Chiara Magnanini of the Department of mechanical engineering Politecnico di Milano, who was always ready to give us the right guidance at every stage in the creation of the paper. Thanks to you, we have increased our knowledge and competence. Special gratitude goes to the 16 students who actively participated in the experiment by spending their precious time with us.