Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 136 (2025) 468–473

www.elsevier.com/locate/procedia

35th CIRP Design 2025

Process Mining For Supporting Data Processing And Process Design In Production And Logistics

Maxi Grobis[a], Steffen Ihlenfeldt[a,b]

[a]Fraunhofer Institute for Machine Tools and Forming Technology IWU, Reichenhainer Str. 88, 09126 Chemnitz,, Germany
[b]Institute of Mechatronic Engineering, Technische Universität Dresden, Helmholtzstraße 7a, 01069 Dresden, Germany

* Corresponding author. *E-mail address:* maxi.grobis@iwu.fraunhofer.de

Abstract

The increasing individualisation of production and logistics elevates the complexity of material flows, as diverse process steps, along with their specific characteristics and interdependencies, must be effectively integrated. Managing these material flow processes demands extensive and accurate data. Process Mining, a technique originally developed in business process analysis, enables the use of existing data to make physical processes transparent in the digital world. This paper examines the application of Process Mining in production and logistics, with a particular focus on how it can impact data processing and support established tools like simulation through more accurate process design in modelling.

Keywords: Process Mining; Simulation; Design method; Production; Modelling

1. Introduction

Digital transformation and Industry 4.0 present companies with challenges such as shortened product life cycles, customer-oriented production, and increasing cost pressure [1,2]. By integrating processes and technologies and leveraging data in a targeted manner, workflows can be optimized and efficiency enhanced, enabling crucial competitive advantages [3].

In the context of Industry 4.0, simulation has established itself as a recognized tool for production and logistics systems to effectively manage the growing complexity of modern production environments [4]. In particular, material flow simulation enables the testing of various scenarios and helps identify potential issues at an early stage, thereby preventing follow-up costs [4,5]. Additionally, simulation enhances flexibility in responding to unforeseen changes and supports innovative approaches to increasing efficiency and agility within a dynamic and competitive business environment [4, 6]. The complexity involved in conducting simulation studies has been thoroughly described and discussed in various scientific publications over the years. From a forward-looking perspective, key challenges in the modelling and simulation of complex production systems include reducing the time required to carry out simulation studies and integrating simulation into physical production as essential aspects [7,8,9,10]. A study by Müller-Sommer [11] highlights that the individual tasks of data preparation, model development (system analysis and model formalization), and verification and validation alone account for approximately 55% of the total effort in simulation projects.

The increasing digitalisation leads to a significantly larger collection of data, which in turn increases higher requirements for connectivity and the use of appropriate tools for efficient data processing, as well as for subsequent modelling to represent manufacturing systems [12]. This is precisely where Process Mining can serve as a supportive, innovative method. By acting as an advanced analytical discipline, Process Mining automatically extracts and analyses real process data from IT systems as event logs, facilitating modelling and

10.1016/j.procir.2025.08.081

enabling targeted analysis and improvement of business processes [13]. Similar to simulation, Process Mining is based on process-oriented insights; although traditionally used primarily in business processes, it is increasingly applied in production and logistics [14].

Process Mining is used to derive targeted insights from manufacturing systems but is often confined to small, specific areas providing supplementary information. This paper illustrates how (1) Process Mining can serve as a valuable tool for the development of simulation models and process design. It specifically examines (2) the types of input data required to transform these data into simulation-relevant information, thereby facilitating effective knowledge extraction into a simulation model (data perspective). Additionally, it explores (3) whether Process Mining can address the challenges associated with data preparation, and model development (process perspective), with consideration given to the operational realities of corporate IT systems. By linking Process Mining and simulation, an analysis-driven process design is enabled, utilising real process data to derive well-founded design solutions. While Process Mining identifies process structures and deviations, simulation allows for the targeted testing and optimization of alternative process configurations. This integrated approach combines process analysis with proactive process design while simultaneously supporting the development and engineering of manufacturing systems, machines, and products through the insights generated.

2. Background

2.1. Data-driven simulation in production and logistics

With the increasing digitalisation in Industry 4.0, research is increasingly focusing on improving the creation of simulation models. Various approaches aim to reduce the complexity of individual phases of the simulation process. This process, which is divided into seven standardized steps according to VDI 3633, includes goal definition, system analysis, data acquisition and preparation, model formalization, implementation, as well as concluding experiments and analyses [5]. Milde and Reinhart [10] demonstrate that 28% of the effort in the process goes into data management and 35% into model development and generation. Their approach utilises tracking and tracing data to minimize data collection and automate model development. Further studies focus on optimizing specific simulation phases [8,15,16,17].

In [15], the authors present an extensive literature review on the automation of input data for discrete-event simulation (DES) in the manufacturing sector. A key issue in industrial DES applications, according to their analysis, is the high time and cost effort required for the collection and integration of input data, with data preparation accounting for up to 40% of the project duration. The authors classify existing approaches into three categories: frameworks for DES data input, standards for data transfer, and automation methods. Skoogh and Johansson [17] previously conducted an in-depth analysis of this phase, further delineating it into data acquisition, data processing, and data interfacing. They proposed methods comprising both manual and semi-automated approaches aimed at enhancing process efficiency. Bergmann [8] expanded on this approach by developing a CMSD-based framework for the automated generation and adaptation of simulation models, wherein ensuring data quality and accurately capturing dynamic changes represent significant challenges. Son and Wysk [18] introduced a methodology for the automated generation of simulation models, based on real-time control and leveraging existing resource and control models. However, this methodology does not encompass the essential phase of data acquisition and preparation, which remains a prerequisite for model generation.

2.2. Process Mining as a tool for data analysis and process modeling

Process Mining has established itself as an advanced analytical discipline by enabling detailed analysis of business processes through the extraction of valuable insights from existing datasets [13,19]. Process Mining leverages digital traces recorded in an organisation's software systems to identify inefficient process flows and perform in-depth analyses of business processes (Discovery), reveal deviations from the intended process (Conformance), and identify and implement optimization opportunities (Enhancement), thereby enhancing data-driven decision-making [20,21]. Recent studies highlight that the use of Process Mining can, among other benefits, improve process efficiency and reduce operational costs [22]. Originally developed for traditional business processes, Process Mining is increasingly applied in production and logistics to improve process efficiency and transparency [13,23]. For instance, Birk et al. examined the application of Process Mining in complex, multi-stage manufacturing processes and developed a method to integrate heterogeneous data sources into a unified event log. Their work illustrates how Process Mining can uncover bottlenecks and optimization potentials, while also addressing challenges in representing and analysing these processes [24]. In [25], a study explores how changes in production systems can be detected using Process Mining techniques. It outlines a method for analysing raw sensor data to calculate relevant system metrics that make changes in the production line visible.

Object-centric Process Mining (OCPM) extends traditional Process Mining by analysing interconnected objects and their interactions [26]. Particularly in production and logistics, this approach enhances the modelling of complex dependencies between materials, resources, and orders. Van der Aalst [27] emphasizes that OCPM captures multiple object types simultaneously, whereas conventional approaches typically focus on individual entities, such as production orders. As a result, OCPM enables a more realistic analysis of complex processes [27].

Process Mining extracts and analyses large volumes of real-time process data, identifies patterns, and supports the development and validation of simulation models, thereby enabling precise, data-driven process design and optimization in production and logistics. Özkul et al. [28] integrate Process Mining with DES to optimize decision-making processes based on data. Similarly, Ortega et al. [29] employ DES to model coal flow in an open-pit mine, utilising Process Mining to capture real process data and enable an accurate simulation of the production process.

3. Process Mining for Simulation – An Approach

Integrating Process Mining into material flow simulations creates synergies for both methods and enhances data-driven modelling. Specific requirements for data and process models are considered, proposing OCPM as a promising method during these phases for developing simulation-based models. Figure 1 illustrates the schematic integration of Process Mining into the creation of discrete-event simulations. Process mining begins with the collection and preprocessing of event data recorded in the various information systems of an enterprise, such as Enterprise Resource Planning (ERP) or Manufacturing Execution Systems (MES). These data serve as the foundation for applying the Process Mining technique known as discovery. A process model generated through Process Mining can serve as a data and information model, as it provides a detailed representation of the actual workflows within a system.

Although Process Mining provides a process-oriented perspective, an object-centric viewpoint is crucial for simulation-based analyses to adequately capture the interactions and relationships between entities and resources. While traditional Process Mining focuses unidimensionally on a single entity (e.g., an order), OCPM allows for a multidimensional analysis by simultaneously examining multiple objects and their interconnections. The goal is to establish a direct link and interoperability between the process-oriented and object-centric perspectives, thereby promoting a holistic examination of dynamic interactions in the simulation results.

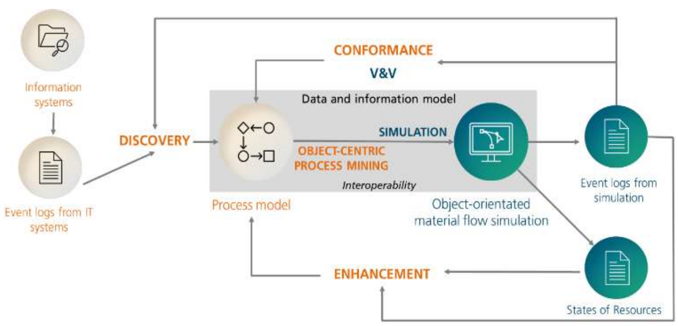

Fig. 1. Approach to the use of Process Mining for simulation.

By utilising the data and information obtained from Process Mining, a simulation model can be developed based on the current actual situation and historical process data. Furthermore, Process Mining inherently generates the process flow, thereby revealing connections and dependencies between processes. Process Mining serves as a powerful tool for analysing and structuring large volumes of data by not only evaluating isolated data points but also uncovering the underlying process interdependencies. This approach supports the verification and validation by enabling a conformity check through the comparison of simulated results with the derived as-is processes. Simulation serves as a tool within analysis-driven design, providing forward-looking insights and enabling flexible adaptation to changing conditions. When bottlenecks or inefficiencies in the current processes are identified, they can be redesigned and optimized using the simulation model in conjunction with input data and insights derived from Process Mining. This approach allows for the prior testing of process designs and system configurations within the manufacturing system before implementation. By simulating various scenarios and evaluating their impacts, informed decision-making is facilitated, ensuring that changes are assessed and refined before being introduced into real production environments. Furthermore, event logs with detailed information on processes, resources, and their states can be generated from the simulation. These event logs facilitate the development of a new reference model. Here, conformance checking, a central technique of Process Mining, is employed to compare observed workflows with the modelled processes and identify any discrepancies. Additionally, the technique of enhancement describes the optimization of the model based on deviations and inefficient practices. This feedback process enables the integration of insights gained from simulation into a target process model, thereby enhancing its explanatory power and adaptability. From the target process model and the resulting event logs, various processes can be derived at different levels. This approach allows for the design and in-depth refinement of both high-level processes, such as the actual material flow, and detailed machine processes, including their states. The bidirectional interaction between simulation and Process Mining fosters a data-driven design approach and supports the continuous improvement of models and processes.

4. Requirements for the approach from a production and logistics perspective

Based on the described approach, the specific requirements from the perspective of production and logistics are systematically outlined. The integration of internal enterprise IT systems establishes the framework for data-related requirements in process modelling. Modern production and logistics systems typically consist of heterogeneous IT systems distributed across various levels of the automation pyramid, ranging from the field level to the enterprise level (see Figure 2). This diversity generates large, heterogeneous data sets, necessitating harmonisation to ensure a consistent process representation. Seamless integration therefore requires the alignment of data formats, semantics, and communication standards.

The development of discrete-event and object-oriented simulation models necessitates a precise definition of objects, particularly resources and their interactions. Key requirements include the representation of resource and material flow dependencies to accurately capture the dynamics of production and logistics processes. Manufacturing data, work plans, and bills of materials are essential, as they define structures, processes, and processing steps, forming the foundation for accurate model development.

4.1. Data perspective

The successful implementation of the described approach requires the availability of consistent data. The quality of the simulation models is critically dependent on the completeness and level of detail of the underlying data, enabling the accurate representation of real processes and interactions and facilitating reliable forecasting. The VDI Guideline 3633 provides a robust foundation by clearly defining and structuring the essential data categories for material flow simulation. These categories encompass organisational data, system load data, simulation result data, and technical data to ensure the systematic collection and utilisation of the necessary data [5].

The development of a simulation model using Process Mining also requires a structured data basis that accurately represents relevant production data. The Entity Relationship Model (ERM) illustrated in Figure 3 provides the structural foundation for capturing production data within an event-driven and modular simulation model. Key entities such as production order, product, bill of materials, production plan, process and subprocesses, resource, material and measurement data are interrelated, thereby depicting the classical structure of data flows as well as material and information flows.

The following section outlines how data should be structured within IT systems from an ERM perspective using Process Mining to enable the development of a simulation model. Dividing the automation pyramid into a business level and a manufacturing level (see Figure 2) helps define the specific requirements for data collection and processing, particularly in terms of identifying the transition point at which event logs are no longer recorded.

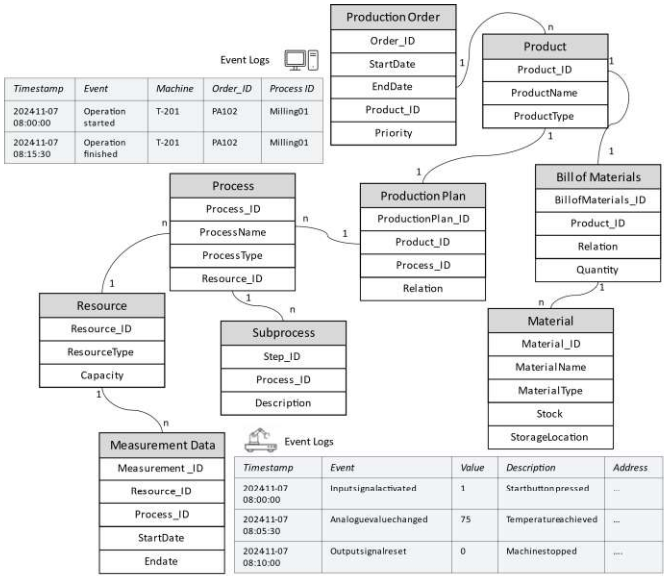

Fig. 3. ERM of the data in a classic material flow simulation model.

The business level encompasses management and planning layers, where company-wide control tasks are performed. At this level, event logs primarily serve to capture business processes and contain information on production orders, such as order ID, product type, quantity, and scheduling. For effective data processing, these logs must include not only timestamps but also additional attributes to ensure the linkage between entities and events. For instance, a single order may comprise multiple products, representing a one-to-many (1:n) relationship. In logistics, event logs record transport orders, inventory movements, and warehouse information to facilitate material flow control. At the MES level, production processes are recorded in detail, with event logs containing specific data on work steps, machine assignments, and resource utilisation. These logs facilitate the representation of material movements and resource allocations, as well as the modelling of material flow structures. However, they often lack the necessary level

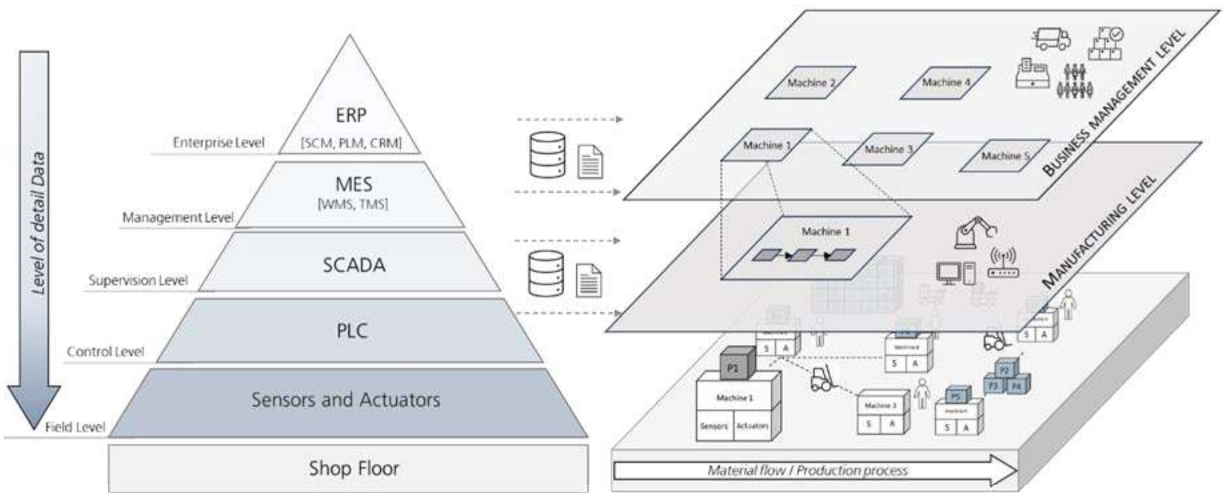

Fig. 2. Overview of information systems and data from the perspective of the automation pyramid.

of detail for an accurate depiction of material flows, making it essential to incorporate additional data from operational levels. The manufacturing level encompasses the control, process control system, and field levels, where physical production processes are monitored and controlled. Event data provide detailed insights into resource activities and machine states by capturing process control events such as status changes, alarms, and adjustments of process variables. These data are crucial for modelling dynamic system behaviours, such as queuing processes, in simulations. However, many of these data are not stored as conventional IT event logs but rather utilised as real-time raw data for control tasks. To ensure precise modelling, machine data must be temporally synchronized with the corresponding processes.

For a comprehensive process reconstruction through Process Mining, event logs from all levels must be integrated and harmonised. Since products are often not directly assigned to a production order, the linkage is established via shared attributes across multiple logs, enabling an object-oriented perspective. The quality and granularity of event data determine the level of detail in simulation models. Start and end times of work operations provide processing times as input parameters for resources. Additionally, the extent to which resources are consistently equipped with sensors and actuators is crucial. Small and medium-sized enterprises (SMEs) often focus on machines, leading to inadequate capture of transport times, waiting times, and inventory levels. In such cases, mathematical methods such as regression analysis are required. Moreover, the combination of multi-level event data enables the identification of process variants, bottlenecks, and dependencies for simulation purposes.

4.2. Process perspective

For simulation-based analyses, adopting an object-oriented perspective is essential, enabling the detailed tracking of products across machines and work operations while capturing interactions and dependencies between products, machines, and other resources. A practical example from production and logistics is an order that includes a main product, which is manufactured in advance, and a purchased component, which is assembled at a later stage. The purchased component is not produced internally but stored in an intermediate warehouse and retrieved in parallel with the order release to ensure its timely availability at the assembly machine. Petri nets are well suited for modelling such processes, as they provide a formal representation of discrete events and state transitions in DES.

A Petri net is a triple $P = (S, T, F)$, where:

- S is a finite set of **places**,
- T is a finite set of **transitions**,
- $F \subseteq (S \times T) \cup (T \times S)$ is a **relation**.

Places represent system states, while transitions denote their changes or activities. For an object-oriented perspective, event logs must be analysed separately by object type, with distinct Petri nets constructed for each type. These nets must then be semantically aligned and consistently integrated. Moreover, object-oriented simulation requires an additional perspective, specifically that of resources.

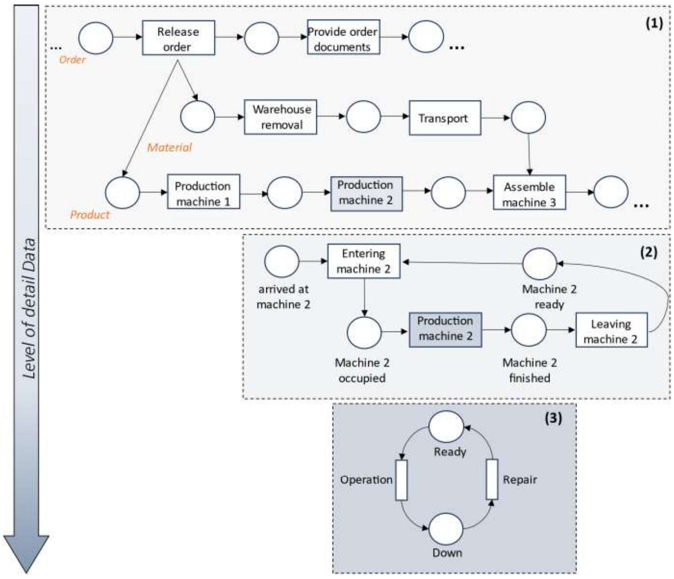

Fig. 4. Petri nets from an object-centric perspective regarding the various information systems.

Figure 4 presents three distinct process models from different system levels. The first Petri net (1) considers the production order as the central object at the business level, while the other models represent the process from the perspective of the main product and the purchased component, which are merged during assembly. For material flow simulation, products must be assignable to their respective production orders. In the simulation, the production order would be created within a source module, generating the required products. Additionally, information on the assembly relationship is necessary, which can be derived through resources. Petri net (1) represents processes at the ERP level, providing a high-level depiction of process sequences. Since data at this level are aggregated, it offers only an abstract overview. To enhance model accuracy, detailed process information from different system levels must be integrated. The Petri net in (2) models processes monitored at the MES level, such as the status of Machine 2 from entry to exit within the work area. Furthermore, the lower representation (3) at the SCADA/PLC level provides information about the operational state of the machine. This enables a more precise depiction of machine behaviour and supports the modelling of resource statuses within the simulation. The main challenge lies in integrating these different levels of detail across the automation pyramid while ensuring semantic consistency. This is crucial, as ERP, MES, and other systems employ different structuring principles and levels of abstraction.

5. Conclusion and Outlook

In summary, this study presents an approach demonstrating how Process Mining can be leveraged as a tool for developing simulation models and designing processes in production and logistics. Specific requirements for data processing and process modelling were identified, and implementation strategies were discussed to efficiently extract simulation-relevant information. Key challenges highlighted include the harmonisation of data and models, ensuring semantic consistency, integrating different levels of process data granularity, and incorporating object-oriented process models to enable a consistent and realistic simulation. Initial approaches, such as the integration of OCPM, show promising potential for enhancing the efficiency of automated model generation for simulations.

The proposed approach follows the principle of analysis-driven design by utilising data-based insights from Process Mining for the iterative improvement of simulation models and process configurations. This concept is currently being further developed in a case study to evaluate its practical applicability and effectiveness. In addition to the identified requirements, challenges also arise regarding data availability and quality, as well as the practical applicability of the approach across various production and logistics environments. The insights gained aim to identify methodological limitations and reveal potential areas for optimisation. Future research will focus particularly on validating the proposed methodology through a concrete application case and assessing its transferability to other use cases, with the results to be presented in subsequent publications.

References

[1] Klemichen A, VDI Zentrum Ressourceneffizienz. Leitfaden Design für Industrie 4.0, Ressourcen schonen durch digitale Technologien. Düsseldorf: VDI Technologiezentrum GmbH; 2024.

[2] Arndt H. Logistikmanagement. Wiesbaden: Springer Gabler; 2015.

[3] Bertschek I, Clement R, Buhr D, Hirsch-Kreinsen H, Falck O, Heimisch A, Jacob-Puchalska A, Mazat A. Industrie 4.0: Digitale Wirtschaft - Herausforderung Und Chance Für Unternehmen Und Arbeitswelt. In: München: Ifo Schnelldienst, vol. 68, no. 10; 2015, p. 3-18.

[4] Gutenschwager K, Rabe M, Spieckermann S, Wenzel S. Simulation in Produktion und Logistik: Grundlagen und Anwendungen. 2017 ed. Berlin, Heidelberg: Springer Berlin Heidelberg; 2017.

[5] Verein Deutscher Ingenieure e.V. VDI 3633 Blatt 1: Simulation von Logistik-, Materialfluss und Produktionssystemen, Grundlagen. Düsseldorf ; 2014.

[6] Atzori L, Iera A, Morabito G. The Internet of Things: A survey. Computer Networks 2010; 54(15), p. 2787–2805.

[7] Popovics G, Pfeiffer A, Kádár B, Vén Z, Kemény L, Monostori L. Automatic Simulation Model Generation Based on PLC Codes and MES Stored Data. Procedia CIRP 2012; 3(1), p. 67-72.

[8] Bergmann S. Automatische Generierung adaptiver und lernfähiger Modelle. In: Eymann T, editor. Tagungsband zum Doctoral Consortium der WI 2011, Zürich: Bayreuther Arbeitspapiere zur Wirtschaftsinformatik; 2011.p. 9-16.

[9] Hönig J, Schnierle M, Scheifele C, Spielmann T, Münster C, Roth A, Röck S, Verl A. Mixed-Reality-in-the-Loop Simulation: Von Produktionssystemen zur Aus- und Weiterbildung. Atp-Edition : Automatisierungstechnische Praxis 2021; Vol.63 (6-7).

[10] Milde M., Reinhart G. Automated model development and parametrization of material flow simulations. Proceedings of the Winter Simulation Conference 2019; p. 2166–2177.

[11] Müller-Sommer H, Strassburger S. Methoden zur Plausibilisierung von Eingangsdaten für Belieferungssimulationen. in Logistik-Planungssystemen der Digitalen Fabrik. In: Zülch G., Stock P. editors. Proceedings der 14. ASIM-Fachtagung Simulation in Produktion und Logistik. Karlsruhe: KIT Scientific Publishing; 2010. p.61-68.

[12] Lidong W, Guanghui W. Big Data in Cyber-Physical Systems, Digital Manufacturing and Industry 4.0. International Journal of Engineering and Manufacturing 2016; 6(4), p. 1–8.

[13] Van der Aalst WMP. Process Mining - Data Science in Action. Berlin, Heidelberg: Springer Berlin / Heidelberg; 2016.

[14] Rubin V, Günther CW, van der Aalst WMP., Kindler E, van Dongen B F, Schäfer W, Pfahl D, Raffo DM, Wang Q. Process Mining Framework for Software Processes. In: Lecture Notes in Computer Science (Vol. 4470). Germany: Springer Berlin / Heidelberg 2007; p. 169–181.

[15] Barlas P, Heavey C. Automation of input data to discrete event simulation for manufacturing: A review. International Journal of Modeling, Simulation and Scientific Computing 2016; 7(1).

[16] Robertson N, Perera T. Automated data collection for simulation?. Simulation Practice and Theory 2002; 9(6), p.349-364.

[17] Skoogh A, Johansson B, Stahre J. Automated input data management: evaluation of a concept for reduced time consumption in discrete event simulation. Simulation (San Diego, Calif.) 2012; 88(11), p. 1279–1293.

[18] Son YJ, Wysk RA. Automatic simulation model generation for simulation-based, real-time shop floor control. Computers in Industry 2001; 45(3), p. 291–308.

[19] Zerbino P, Stefanini A, Aloini D. Process Science in Action: A Literature Review on Process Mining in Business Management. Technological Forecasting & Social Change 2021; 172.

[20] Marti N, Fischer DA, Kerpedzhiev GD, Goel K, Leemans SJJ, Röglinger M, van der Aaalst WMP, Dumas M, La Rosa M, Wynn MT. Opportunities and Challenges for Process Mining in Organizations: Results of a Delphi Study. Business & Information Systems Engineering 2021; 63(5), p. 511–527.

[21] Van Der Aalst WMP. Process mining: Overview and opportunities. ACM Transactions on Management Information Systems 2012; 3(2), p. 1–17.

[22] Bolt A, de Leoni M, van der Aalst WMP. Scientific workflows for process mining: building blocks, scenarios, and implementation. International Journal on Software Tools for Technology Transfer 2016; 18(6), p. 607–628.

[23] Becker T, Intoyoad W. Context Aware Process Mining in Logistics. Procedia CIRP 2017; 63, p. 557-562.

[24] Birk A, Wilhelm Y, Dreher S, Flack C, Reimann P, Gröger C. A Real-World Application of Process Mining for Data-Driven Analysis of Multi-Level Interlinked Manufacturing Processes. Procedia CIRP 2021; 104, p. 417–422.

[25] Chiò E, Alfieri A, Pastore E. Change-point visualization and variation analysis in a simple production line: a process mining application in manufacturing. Procedia CIRP 2021; 99, p. 573–579.

[26] Van der Aalst WMP, Salaün G, Ölveczky PC. Object-Centric Process Mining: Dealing with Divergence and Convergence in Event Data. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).Switzerland: Springer International Publishing AG 2019;. p. 3-25.

[27] Van der Aalst WMP. Object-Centric Process Mining: Unraveling the Fabric of Real Processes. Mathematics (Basel) 2023; 11(12), 2691.

[28] Özkul F, Sutherland R, Wenzel S. Simulation-enhanced Action-oriented Process Mining in Production and Logistics. ARGESIM Report 47 2024; p. 193-201.

[29] Díaz Rodríguez JL, García Mogollón JM, Parra Ortega C. Modeling and simulation of coal flow in a mine using discrete events. Prospectiva (Barranquilla. Online) 2024; Vol.22 (2).