Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

PROCESS MINING AS CATALYST OF DIGITAL TWINS FOR PRODUCTION SYSTEMS: CHALLENGES AND RESEARCH OPPORTUNITIES

Giovanni Lugaresi[1]

[1]Department of Mechanical Engineering, KU Leuven, Leuven, BELGIUM

ABSTRACT

The advantages in terms of productivity from recent investments toward higher automation come along with an escalating complexity in manufacturing systems. This prompts for leveraging digital support assistants, notably digital twins, for production planning and control tasks. Process mining has emerged as a valuable tool in the realm of digital twinning as it proved its efficacy in tasks such as model generation, trace profiling, and performance evaluation. However, several challenges persist in different methodological and application areas, presenting valuable opportunities for both academia and industry. This work aims to shed light on a selection of topics uniting the fields of process mining and digital twins for manufacturing systems, with a reflection on current challenges and research opportunities that can be seized in the near future.

1 INTRODUCTION

The landscape of manufacturing systems recently witnessed a paradigm shift marked by a surge in investments directed towards augmenting automation levels. Beside the notable improvements in productivity (Rao et al. 2008), they also introduced a layer of complexity that necessitates the integration of advanced digital support systems. To enhance production efficiency, digital twins have been deployed in various processes such as production scheduling and internal logistics (Qi et al. 2018). A digital twin represents a virtual counterpart of a physical object or system, serving diverse purposes such as simulating its behavior, forecasting its performance, and refining its design and operation. In production planning and control phases, digital twins can be based on digital representations of the physical counterparts that manifest as discrete-event simulation models. Indeed, augmenting these models with real-time data streams enables production planners to assess solutions optimal for the current system state (Tavakoli et al. 2008). For a successful digital twin implementation, the digital representations of production systems must accurately mirror the system's current configuration, possibly at any given time (Cardin and Castagna 2012). However, constructing digital models for each production system remains a challenging and time-consuming task, especially for large enterprises that cope with a large number of systems such as automakers (Wozniak and Clements 2015).

A *model generator* is a tool for translating the structure and behavior of a physical system into code, enabling a computer to perform the tasks typically done by a human modeler (Ng et al. 2011). In the manufacturing context, model generation involves identifying the characteristics of individual resources and their logical relationships, such as spatial constraints, temporal statistics, and scheduling policies (Biesinger et al. 2019). Hence, the deployment of a model generator along the life cycle of a production line can ensure that any addition, removal, or modification of a resource's behavior is promptly reflected in the digital model (Santillan Martinez et al. 2018). Process mining (Van Der Aalst 2016) has surfaced as a powerful and versatile collection of model generation techniques that contribute substantially as enablers of digital twins. Process mining applications extend across various facets, including model generation, trace profiling, and performance evaluation. While the integration of process mining and digital twinning

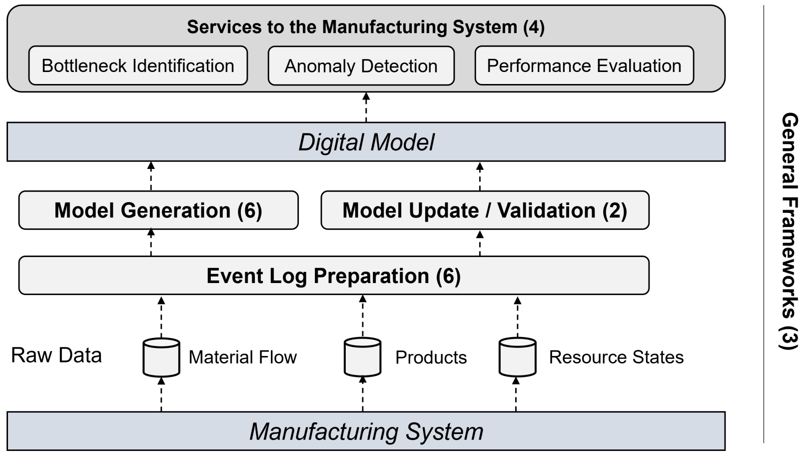

Figure 1: Schematic view of the main topics discussed in the literature section (the numbers refer to the number of papers found).

gains momentum, several research challenges persist, spanning methodological intricacies over diverse application domains.

This paper aims to present unexplored avenues at the intersection of process mining and industrial engineering, identifying research challenges that deserve attention from both academic and industrial perspectives. The discussion wishes to unravel the opportunities and frontiers that await exploration in the near future, fostering collaborative efforts for future research. The remainder of the paper is organized as follows. Section 3 lists the limitations of the existing approaches. Section 4 outlines promising research directions. Section 5 concludes the paper with the author's remarks and suggested case studies.

2 STATE OF THE ART

For this work, a literature review has been done based on the following query: *"digital twin" AND ("manufactur*" OR "produc*") AND "process mining"*. This query has been lastly excecuted on 2024-03-05 on the Scopus database and returned 36 results (Scopus 2024). The papers have been selected based on the following exclusion criteria: (1) the paper is accessible and written in English; (2) the paper is about manufacturing system engineering topics; (3) the work focuses on the production operations (i.e., business processes are excluded in this work). As a result, 21 papers have been included in this analysis. The papers have been classified with respect to their main scope: (1) *general frameworks* refer to works that propose a conceptual or practical overview on how process mining can be included in a digital twin architecture; (2) *data preparation* includes the works that focus on building datasets that can be effectively exploited for relevant operations in a digital twin structure; (3) *model generation and tuning* collects the works focusing on model building phases and on techniques to control the level of detail of the generated models; (4) *model validation and update* are works focusing on the physical-to-digital alignment enabled by the inclusion of process mining techniques; finally, works focusing on (5) *services to the production systems* exploit process mining to provide useful responses for decision-making. Figure 1 schematically represents the topics in the literature based on how they relate to the main digital twin services. The following sections elaborate on each of the main topics, presenting the main contributions that are also summarized in Table 1.

2.1 General Frameworks

Francis et al. (2021) suggested a versatile data-driven framework for the automated development of digital twins tailored for smart factories. The proposed framework introduced elements such as a data-driven methodology integrating machine learning and process mining techniques, as well as continuous model enhancement and validation processes. Friederich et al. (2022) introduced a comprehensive data-driven framework for the automated creation of simulation models as the foundation for digital twins in smart factories. The objective is to reduce the requirement for specialized expertise in extracting corresponding simulation models. The data-driven methodology leverages machine learning and process mining techniques, alongside continual model refinement and validation. Similarly, Yadav et al. (2023) proposed a comprehensive framework aimed at automating the generation of simulation models driven by data. The authors proposed to add event labeling, an activity conducted in a semi-automated manner, involving manual intervention by stakeholders or experts who demonstrate expertise in the factory operations: relevant events are identified and categorized to ensure accurate labeling. Then, machine learning models are employed to autonomously identify additional events, using the labeled data as a foundation.

2.2 Data Preparation

Bayomie et al. (2022) presented a technique for generating a tailored event log from sensor data, suggesting various encodings for the sensor data designed to align with the interests of process analysts. An experiment conducted with real-world data demonstrated the efficacy of the approach.

Mayr et al. (2022) explored how process mining methodologies can be used to uncover and track process control flow if paired with data-driven digital twins. The authors introduced the so called self-discovering and event-log-aware digital control-flow twins to connect process mining, digital twins, and process data, as well as to address the abstraction disparity by producing discrete event logs supported by descriptive statistics of recurring process-state features through unsupervised clustering techniques. The proposed approach concentrates on the lack of time-stamped event log data, requiring to extract recurring patterns from process-state data with minimal input prerequisites in an unsupervised setting, and without prior knowledge of event duration or their occurrence time.

Tan et al. (2023) introduced a data assimilation technique to manage structural shifts within the system. The method aims to automatically generate a digital model concurrently with the data assimilation procedure. The method involves the continuous generation of new models and the selection of models based on their optimal fitness. Events trigger the automatic creation of corresponding process models for each data sample. Throughout the data assimilation process, system performance is monitored and aggregated.

Telli et al. (2023) created records by transmitting a process representation through fault and repair process event logs. The authors circumvent the loss of context inherent in multi-perspective analysis, despite incorporating other perspective data through multi-modal analysis. Multi-modal models are developed by integrating additional perspective data into the analysis through root cause analysis, aided by metrics, detection of novel behaviors, and process refinement with new models.

2.3 Model Generation and Tuning

Lugaresi and Matta (2021a) demonstrated the implementation of a digital model generation technique exploiting a test case using a lab-scale model of a manufacturing system consisting of six stations. The investigation focuses on the transitional phase of the online model generation process. The availability of an adequate number of data points from production event logs is indicated as essential element for accurate model building. Lugaresi and Matta (2021b) proposed a graph model generation technique to produce a digital twin skeleton able to be translated into a discrete-event simulation model. Material flow data from the event logs are used to retrieve the system parameters and feed the generated simulation model. Also, a technique to control the granularity of the generated models is proposed. Lugaresi and Matta (2023) extended the proposed technique for data-driven model building for manufacturing systems with assembly operations. The proposed method takes into account the merging point of material flows that are characterizing these systems. The generated models are directed graphs enriched with production system parameters that can be translated into discrete-event simulation models.

Novák and Vyskočil (2022) suggested using a formal representation of production resources and their corresponding operations, which can be transformed into an executable digital twin that can then be enriched with operation duration times acquired through process mining techniques, resulting in accurate simulation models of the production system.

de Oliveira et al. (2023) assessed the combination of process mining, simulation, and multi-criteria decision making using event logs of a manufacturing system. The study constructs the process model through process mining, which is then used to drive simulations in Python for analyzing production scenarios, then Multi-Criteria Decision Making approach is employed to evaluate the scenarios and determine the most suitable solution considering the current constraints of the production process.

2.4 Model Validation and Update

Takakura et al. (2023) introduced a novel metric based on maximum log-likelihood to assess how accurately the graph retains the actual process-flow information. Additionally, an evaluation metric is proposed to combine both log-likelihood and model dimensions. Empirical testing was carried out using real steel-making process data to evaluate the effectiveness of the proposed maximum log-likelihood-based.

Overbeck et al. (2021) introduced a method aimed at bridging the disparity between model representations and real-world scenarios through ongoing, iterative updates facilitated by linking the simulation model with IT systems and employing intelligent data analysis techniques.

Park et al. (2022) proposed methods for information fusion and systematic logic library (SLL) generation to assist in the self-configuration of autonomous digital twins. Information fusion extracts essential elements from various sources, while the SLL generation method creates representations of functional units within the physical asset. The proposed methods are based on standards such as ISA-95 and OPC UA and support the development of an autonomous digital twin framework by enabling the extraction of asset description objects and SLL through artificial intelligence.

2.5 Services to Production Systems

Several contributions propose frameworks or approaches that aim to directly provide specific services to the production systems; these works are classified in the following sections.

2.5.1 Anomaly Detection

Chiò et al. (2021) proposed a method to identify alterations within a production system through the analysis of its sensor data. Raw sensor data are processed to calculate and visualize pertinent system metrics conducive to change detection. To evaluate its efficacy, simulation is employed to test the proposed procedure within a medium-sized production line.

2.5.2 Bottleneck Identification

Kumbhar et al. (2022) introduced a data-driven approach leveraging data integration, process mining, and analytics rooted in factory physics to pinpoint bottlenecks. The general idea is to use process mining to first generate a factory physics simulation model of the system, which is then fed with system data and used to identify the bottlenecks. The authors conducted a case study within an industrial setting, constructing and validating a discrete event simulation model. Similarly, Kumbhar et al. (2023) presented the development of a digital twin framework designed to identify, diagnose, and enhance bottleneck resources through utilization-based bottleneck analysis, process mining, and diagnostic analytics. The approach leverages enterprise data from various sources such as production planning, process execution, and asset monitoring, to create an event log directly feedable into the digital twin.

2.5.3 Performance Evaluation

Yang et al. (2022) presents a digital twin discovery framework including process mining. The resultant digital twin gathers its attributes within the event log. The proposed digital twin framework is based on cloud computing and it is used to forecast the remaining manufacturing cycle time employing machine learning, specifically through a transition-based embedding approach.

3 CHALLENGES OF CURRENT APPROACHES

Despite the advancements of process mining-based approaches, several limitations remain for applications to manufacturing environments. This section elaborates on such limitations, classified accordingly.

3.1 Sensitivity to System Type and Complexity

Despite the significant advances of available approaches, each contribution has been developed to tackle a particular type of production system. Meanwhile, there are notable variations among different types of systems that can be identified. For instance, job shops exhibit multiple independent part flows, potentially complicating the identification of the system structure. Some discovery methods rely on parameters that must be configured based on prior knowledge (Lugaresi and Matta 2021b). Estimating these parameters for adapting to a different system type can be a laborious task and may require manual interventions, thus contradicting the initial proposition of automated model discovery.

3.2 Limited Model Tuning Capabilities

Relying solely on an up-to-date process model may prove insufficient, as an automated model generation process could excessively detail certain aspects of the system (i.e., the *"spaghetti model"* effect (Van Der Aalst 2016)). Such models not only tend to be overly large and challenging to comprehend, but also inaccurate in estimating performance metrics. Consequently, it is essential to assess whether the level of detail in a generated model aligns with its intended use. Model tuning or adjustment approaches involve changing a model to achieve a proper size. Despite several proposed approaches, there is no general consensus on a specific method to handle over-complex models in manufacturing systems. Table 2 lists selected approaches that can be effectively used, with a comparison of the main advantages and disadvantages in their application.

Traditional process mining approaches (Van Der Aalst 2016) can be used to generate a complete graph representation of a system. In this case, the selection of relevant parts to be kept in the model is left to the user. Hence, the risk of relevant information loss is low, but at the cost of a higher manual intervention.

Fuzzy process mining approaches (Günther and Van Der Aalst 2007) assign metrics to nodes and arcs using significance and correlation scores. Then, less significant nodes are either eliminated or grouped into clusters, which inherit the precedence relationships and connections between nodes. Thanks to the correlation measures between different activities, a graph can be reduced in size within a reasonable time. However, all the steps are guided by user-defined thresholds for node and arc cut-off parameters. The definition of these thresholds is not trivial and it is strictly case dependent. Also, in several fuzzy mining approaches the activity names are used to generate correlation measures based on string similarities to facilitate clustering. This approach might reveal of limited utility in production systems, where activity names often consist of sensor or machine identifiers.

In trace clustering and similar approaches (Song et al. 2008; Bose and der Aalst 2009), the event log is divided into uniform subsets, each corresponding to an observed part in the system. Distinct characteristics are computed for each case and a resultant feature vector is attributed to each case. This way, clustering methods can be employed to categorize similar traces, and process mining is applied individually on each cluster. Trace clustering methods effectively lead to models with reduced complexity. Nevertheless, this holds true solely for a subset of identified clusters, which emphasize entities adhering to straightforward

Table 1: Selected contributions from the literature review of section 2 (OCEL: Object-Centric Event Log, DGM: Directed Graph Mining, DES: Discrete Event Simulation, DG: Directed Graph, GM: Graph Model, FMS: Flexible Manufacturing System).

Reference	Main Scope	Framework	Input Data	Discovery Method	Output Model	System Type
Lugaresi and Matta (2021a)	Model Generation	-	Material Flow	Custom	DES	Flow Line
Takakura et al. (2023)	Event Log Preparation	-	Material Flow	-	-	General
Overbeck et al. (2021)	Model Update	-	Material Flow	Alpha Algorithm	DES	Flow Shop
Park et al. (2022)	Event Log Preparation	•	Event Logs	-	DES	General
Bayomie et al. (2022)	Event Log Preparation	-	Sensor Data	DGM	DG	Single CNC
Kumbhar et al. (2023)	Bottleneck Identification	•	Custom Event Log (Generator)	DGM	FACTS	Flow Line
Francis et al. (2021)	General Framework	•	Material Flow	-	-	General
de Oliveira et al. (2023)	Model Generation	•	Material Flow	DGM	DES	Job Shop
Lugaresi and Matta (2023)	Model Generation	-	Material Flow	Custom	GM	Assembly
Friederich et al. (2022)	General Framework	•	Material Flow	Alpha Algorithm	Petri Net	FMS
Overbeck et al. (2023)	Model Update and Validation	-	Material Flow		DES	Flow Shop
Lugaresi and Matta (2021c)	Model Generation	-	Material Flow	Custom	GM	Assembly
Tan et al. (2023)	Event Log Preparation; Model Generation	-	OCEL	Machine Learning	Petri Net	*Semiconductor
Yadav et al. (2023)	Model Generation	•	Condition Monitoring, Event Data, State Data	Alpha Algorithm	Petri Net	Flow Line
Chiò et al. (2021)	Anomaly Detection	-	Material Flow	-	-	Flow Line
Telli et al. (2023)	Event Log Preparation	-	Fault Data	-	-	-
Kumbhar et al. (2022)	Bottleneck Identification	-	Material Flow	DGM	DG	FMS
Yang et al. (2022)	Performance Evaluation	•	Material Flow	DGM	DG	Job Shop
Lugaresi and Matta (2021b)	Model Generation, Model Tuning	-	Material Flow	Custom	GM	Flow Line
Novák and Vyskočil (2022)	Model Generation	•	Material Flow	DGM	DG	FMS
Mayr et al. (2022)	Event Log Preparation	-	Material Flow, State Data	DGM	DG	Flow Line

Table 2: Comparison of selected model adjustment approaches.

Method	Advantages	Disadvantages
Process Mining	Complete graph is easier to understand for users.	Almost all adjustments are manual; it forces a selection of nodes.
Fuzzy Mining	Correlation measures among activities; simplified graph can be obtained in a short time.	Forces to choose a metric for clustering; ineffective on some types of graphs.
Trace Clustering	Selection of graphs with a lower complexity.	Clusters with a high importance might be excluded; separate views for each model.
Model Tuning	System properties are kept in the representation; reduction of size in a short time.	Search parameters have to be set in advance; no guarantee of optimal model.

paths. Meanwhile, other clusters that could potentially be associated with intricate production dynamics persist in the log.

Model tuning approaches (Lugaresi and Matta 2021b) aim to generate digital twins with a proper level of detail according to the user's requests. The pertinent traits of a production system are extracted from data logs automatically, and used to drive the adjustments toward a of a proper size. The advantageous aspect of this approach is that the system properties are kept in the final representation. However, the tuning criteria are static and lack dependence on the distinctive features of the manufacturing system. Also, the current tuning search algorithm prioritizes speed of execution, potentially producing local optima.

3.3 Discovery of Non-Linear Material Flows

The application of model generation techniques in systems presenting converging or diverging material flows has shown some limitations of current approaches. Indeed, several process mining methodologies operate under the premise of a single part identifier within the dataset used to generate the model, while in realistic settings, multiple object types often participate in production steps. For instance, in assembly systems, part identifiers often transition between production stages. When gathering event data from these systems, several data flattening options exist, leading to disparate perspectives. The system's structural overview may be rapidly lost as event data must be extracted multiple times for various viewpoints. Meanwhile, the digital model must account for the assembly phases as the availability of all necessary materials upstream serves as a blocking condition at assembly points. Neglecting this condition during model generation will therefore result in a model that overestimates the real system's performance.

Current approaches to handle merging material flows rely on strict assumptions which limit their applicability to industrial settings. For instance, the approach proposed by Lugaresi and Matta (2021c) relies on temporal proximity and is therefore not able to represent batched operations, causing the inability to reproduce certain types of manufacturing systems (e.g., semi-conductor production systems). Also, it assumes perfect traces and the complete availability of bill of materials, hence it is not possible to discern among a system change or wrong inputs.

3.4 Discovery of Production Policies

Process mining is also adept at unveiling the production policies implemented on the shop floor. Basic policies often revolve around the production plan and understanding splitting ratios in cases of multiple alternative paths. However, realistic environments feature an array of rules influencing their performance, such as maintenance policies, priority rules, as well as adherence to lean production principles. Current literature lacks methodologies capable of comprehensively identifying the full spectrum of production policies, indicating the need for further research in this area. Precisely distinguishing among policies can also enhance model generation, enabling easy differentiation of various elements within a manufacturing system. For instance, a conveyor serving as a buffer can be characterized by a first-in-first-out policy (Milde and Reinhart 2019).

3.5 Management of Rare Events

A common challenge faced by data-driven approaches and often cited in the process mining literature is the effective handling of rare events. Within manufacturing environments, these events might manifest as machine breakdowns, unplanned intensive maintenance, or extreme weather conditions. Such rare occurrences tend to be more disruptive than more frequent ones, often resulting in significant detriment to the system performance and profitability. Unfortunately, identifying these rare conditions is not straightforward and is frequently overlooked in model generation techniques. For example, in model tuning methodologies, rare events are often disregarded based on frequency- or causality-based scores. Additionally, automatically generated models may inadvertently introduce rare conditions themselves, such as unsound Petri Nets susceptible to deadlocks. Herein lies the challenge of distinguishing between accurately modeling rare conditions and modeling errors.

3.6 Integration of Expert Knowledge

Competitive edge in manufacturing depends greatly on the proficiency of the workforce (Barney and Clark 2007). In this context, expert knowledge plays a pivotal role in driving successful production operations. Conventional process mining approaches depend solely on event logs as their primary information source. Hence, the integration of a company know-how into model generation methodologies can enhance their effectiveness (Dixit et al. 2015). Innovative model generation approaches may be combined with machine learning techniques to achieve improved functionalities, such as predicting the results of certain actions or classifying different production conditions. In this context, expert knowledge can reveal essential to validate such approaches before applying them online.

4 RESEARCH OPPORTUNITIES

Based on the aforementioned limitations and on the literature review of section 2, this section provides a preliminary list of promising research directions for process mining-based digital model generation and exploitation for manufacturing applications.

4.1 Joint Process Mining Approaches

In several applications, a single process mining algorithm is relied upon by the model generation method, necessitating a compromise between specific algorithms performance. For example, inductive algorithms demonstrate scalability, fitness to the log, and precision, while heuristic algorithms excel in discovering diverse process patterns. To address this limitation, more extensive methodologies need to be devised. One approach could involve the integration of multiple process mining algorithms within the same analysis. For instance, by employing a joint methodology that exploits both inductive and heuristic mining algorithms, it would become possible to leverage the strengths of each algorithm while keeping the model generation computation times reasonable.

Also, most approaches consider only isolated aspects of systems (e.g., reliability models) and focus on a specific modeling purpose (e.g., material flow identification). Novel approaches can enable the combination of models extracted through different process mining algorithms (Friederich et al. 2022). Further research will benefit from an assessment on the relation between available data in common manufacturing information systems (e.g., ERP, MES) and the type of models that can be developed with mining approaches. Moreover, the compatibility among modeled features and each modeling formalism has to be addressed. For instance, Petri Nets are a good choice for representing blocking conditions, which are typical of machine failures or assembly points. Hence, the final modeling formalism is driven by the choice to model such features. The development of techniques for the exploitation of the automatically built models will enhance their exploitation in forward looking scenarios, and may imply the introduction of machine learning approaches. For example, the discovery of a maintenance policy can be done with process mining approaches, while its optimization may benefit from the use of reinforcement learning algorithm applied to the digital instances. Last but not least, the integration of continuous data, such as condition monitoring data, into the overall modeling framework is yet another challenge. For instance, degradation states of machines may first need to be detected and diagnosed using signal processing approaches.

4.2 State-based System Discovery

A manufacturing system behavior undergoes significant changes based on its state. For example, certain scheduling policies might only apply when a station is in a degraded condition, and dispatching rules dictating which parts to load from a buffer into a machine may be changed if the machine operates under non-nominal conditions or fails (Zhu et al. 2023). Process mining methodologies follow a sequential discovery approach that do not monitor the evolution of the system state. As a consequence, the system attributes are estimated over aggregated time horizons and disregarding their state-dependent characteristics. New model generation approaches should therefore recognize and maintain visibility over the system states during the discovery phase (Zhu et al. 2023).

4.3 Event Graph-based Process Monitoring

Graph models discovered with process mining can be used to compare the as-is situation in the factory and the to-be, desired behaviour. The discovered graphs may also be used to relate quality measures. For example, an enterprise may know that certain paths will lead to faulty products or a higher failure probability. Process mining can be used to discover the graphs of the current system behaviour. Each product family could be associated with a cluster of graphs and the probability of leading to a faulty product can be associate to each graph variant. Event knowledge graphs are data structure that allow to naturally model behavior over multiple entities as a network of events (Fahland 2022). It has been shown that they can be used to construct, query, and aggregate event knowledge graphs to get insights into complex behaviors. Event knowledge graphs are proposed as a very versatile tool that opens the door to process mining analyses in multiple behavioral dimensions at once. Adequate query languages that also can handle process-relevant phenomena such as frequency, noise, performance in relation to multiple entities need to be considered. Also, more complex behavioral dynamics can be discovered. For example, enriching the event knowledge graph with the activity dimension to derive the performance spectrum allows detecting sub-graphs that indicate high workload or a dynamic bottleneck (Toosinezhad et al. 2020).

4.4 Video-based Process Mining

The exploitation of video-sources to track material flow behaviour in the factory has been recently recognized as promising research direction (Chen et al. 2020). Video-based part recognition is a standardized practice in digitized companies. Among other scopes, it is used for part identification, quality control, and assembly monitoring. Despite the very promising relationship, noteworthy approaches linking video monitoring to model generation are scarce. The existing contributions remain conceptual and prototypical (Lepsien et al. 2022). Recent demonstrators shown promising connections to link video-based data sources on the shop-floor to event logs and model generation approaches (Lugaresi et al. 2024).

4.5 Digital Model Enrichment with Artificial Intelligence

The application of artificial intelligence to manufacturing has been proved to be a promising and rich research field (Nti et al. 2022). Yet, it is not clear how it can be used as integrated with process mining for model generation. In a realistic simulation-based digital twin it is essential to be able to change the statistical distribution functions once the state of the system changes. However, in most realistic systems the distributions and their co-variate relationship with input parameters are unknown. Recent research showed the capability of state-of-the-art neural networks to predict not only output values, but also statistical

Table 3: Suggested connections between the presented challenges and research opportunities in this paper, together with proposed use cases (SS: Single Station, JS: Job Shop, FS: Flow Shop, FMS: Flexible Manufacturing System, AL: Assembly Line, FL: Flow Line, CLS: Closed-loop Systems).

	Research Opportunities					*Suggested Use Cases*						
Challenges	4.1	4.2	4.3	4.4	4.5	SS	JS	FS	FMS	AL	FL	CLS
Sensitivity to System Type	•		•		•		•	•	•	•		•
Limited Model Tuning Capabilities		•	•		•		•	•		•		
Discovery of Non-Linear Material Flows		•			•		•			•		
Discovery of Production Policies	•	•				•		•			•	•
Management of Rare Events	•			•	•						•	
Integration of Expert Knowledge	•		•		•	•	•					•

distributions (Reed et al. 2021). This approach is promising because situations of partial knowledge are very common in manufacturing systems. For instance, the distribution of service times of a particular operator on an assembly line, a conditional distribution of processing times given the degradation states of a machine, or the update of an established simulation model that did not previously include a specific distribution. Other challenges must be taken into account for this research direction, such as how to cope with the large number of input parameters (e.g., number of nodes per layer, hidden layers, loss functions).

5 FINAL REMARKS

The presented research challenges and opportunities are inevitably linked by the characteristics of the underlying systems and related issues. Table 3 attempts to propose a link between the topics as a proposal for further research. The sensitivity of model generation approaches to the system type can be addressed by joint mining approaches, which could be used to wisely select between which discovery method to use for the representation of a specific part of the system (e.g., a directed graph for the main material flows, a Petri Net for the resource reliability model). Experiments on systems rich in interactions such as job shops could reveal beneficial. The development of approaches that aim to overcome the limited model tuning capabilities may benefit from use cases on complex systems such as job and flow shops, in which the dimension of the initial graph is typically large. It is also worth to notice that system discovery has not been tested extensively with closed-loop systems. Several issues are foreseen: (1) incorrect discovery of the graph: if pallets do not possess single identifiers, the discovery algorithm finds open models. (2) if single pallet identifiers are available, the buffer size discovery has to be reviewed. Indeed, the traces can become infinite and a new method should be designed to cope with this case.

This work presented a preliminary list of limitations and research directions for the application of process mining in digital twins for manufacturing systems. This work is by no means complete. The literature review has been done systematically only on one research database. A more formal review should be done including multiple sources as well as backward and forward citation analyses. Also, the review did not include significant contributions from the business process modeling literature, which could be used as inspiration for innovative approaches. This work could also benefit from extensive literature reviews that include methodologies developed in other fields than manufacturing.

REFERENCES

Barney, J. B. and D. N. Clark. 2007. *Resource-based theory: Creating and sustaining competitive advantage*. Oxford University Press on Demand.

Bayomie, D., K. Revoredo, S. Bachhofner, K. Kurniawan, E. Kiesling and J. Mendling. 2022. "Analyzing Manufacturing Process By Enabling Process Mining on Sensor Data.". In *PoEM Workshops*.

Biesinger, F., D. Meike, B. Kraß, and M. Weyrich. 2019. "A digital twin for production planning based on cyber-physical systems: A Case Study for a Cyber-Physical System-Based Creation of a Digital Twin". *Procedia CIRP* 79:355–360.

Bose, R. P. J. C. and W. M. P. der Aalst. 2009. "Abstractions in process mining: A taxonomy of patterns". In *International Conference on Business Process Management*, 159–175. Springer.

Cardin, O. and P. Castagna. 2012. "Myopia of service oriented manufacturing systems: benefits of data centralization with a discrete-event observer". In *Service Orientation in Holonic and Multi-Agent Manufacturing Control*, 197–210. Springer.

Chen, C., C. Zhang, T. Wang, D. Li, Y. Guo, Z. Zhao *et al*. 2020. "Monitoring of assembly process using deep learning technology". *Sensors* 20(15):4208.

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

de Oliveira, A. B., A. L. Micosky, C. F. dos Santos, E. de Freitas Rocha Loures and E. A. P. Santos. 2023. "A Hybrid Model to Support Decision Making in Manufacturing". In *International Conference on Flexible Automation and Intelligent Manufacturing*, 651–658. Springer.

Dixit, P. M., J. C. Buijs, W. M. van der Aalst, B. Hompes and J. Buurman. 2015. "Using domain knowledge to enhance process mining results". In *International Symposium on Data-Driven Process Discovery and Analysis*, 76–104. Springer.

Fahland, D. 2022. "Process mining over multiple behavioral dimensions with event knowledge graphs". In *Process Mining Handbook*, 274–319. Springer.

Francis, D. P., S. Lazarova-Molnar, and N. Mohamed. 2021. "Towards data-driven digital twins for smart manufacturing". In *Proceedings of the 27th International Conference on Systems Engineering, ICSEng 2020*, 445–454. Springer.

Friederich, J., D. P. Francis, S. Lazarova-Molnar, and N. Mohamed. 2022. "A framework for data-driven digital twins of smart manufacturing systems". *Computers in Industry* 136:103586.

Friederich, J., G. Lugaresi, S. Lazarova-Molnar, and A. Matta. 2022. "Process Mining for Dynamic Modeling of Smart Manufacturing Systems: Data Requirements". *Procedia CIRP* 107:546–551.

Günther, C. W. and W. M. P. Van Der Aalst. 2007. "Fuzzy mining–adaptive process simplification based on multi-perspective metrics". In *International conference on business process management*, 328–343. Springer.

Kumbhar, M., A. H. Ng, and S. Bandaru. 2022. "Bottleneck detection through data integration, process mining and factory physics-based analytics". In *10th Swedish Production Symposium (SPS2022), Skövde, April 26–29 2022*, 737–748. IOS Press.

Kumbhar, M., A. H. Ng, and S. Bandaru. 2023. "A digital twin based framework for detection, diagnosis, and improvement of throughput bottlenecks". *Journal of manufacturing systems* 66:92–106.

Lepsien, A., J. Bosselmann, A. Melfsen, and A. Koschmider. 2022. "Process Mining on Video Data.". In *ZEUS*, 56–62.

Lugaresi, G., L. Kovacs, and K. Tamas. 2024. "Digital Twin Driven Assembly Line Re-Balancing and Decision Support". In *2024 38th ECMS International Conference on Modeling and Simulation*.

Lugaresi, G. and A. Matta. 2021a. "Automated digital twins generation for manufacturing systems: a case study". *IFAC-PapersOnLine* 54(1):749–754.

Lugaresi, G. and A. Matta. 2021b. "Automated manufacturing system discovery and digital twin generation". *Journal of Manufacturing Systems* 59:51–66.

Lugaresi, G. and A. Matta. 2021c. "Discovery and digital model generation for manufacturing systems with assembly operations". In *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, 752–757. IEEE.

Lugaresi, G. and A. Matta. 2023. "Automated digital twin generation of manufacturing systems with complex material flows: graph model completion". *Computers in Industry* 151:103977.

Mayr, M., S. Luftensteiner, and G. C. Chasparis. 2022. "Abstracting process mining event logs from process-state data to monitor control-flow of industrial manufacturing processes". *Procedia Computer Science* 200:1442–1450.

Milde, M. and G. Reinhart. 2019. "Automated Model Development and Parametrization of Material Flow Simulations". In *2019 Winter Simulation Conference (WSC)*, 2166–2177. IEEE.

Ng, A. H., J. Bernedixen, M. U. Moris, and M. Jägstam. 2011. "Factory flow design and analysis using internet-enabled simulation-based optimization and automatic model generation". In *Proceedings of the 2011 winter simulation conference (WSC)*, 2176–2188. IEEE.

Novák, P. and J. Vyskočil. 2022. "Digitalized automation engineering of Industry 4.0 production systems and their tight cooperation with digital twins". *Processes* 10(2):404.

Nti, I. K., A. F. Adekoya, B. A. Weyori, and O. Nyarko-Boateng. 2022. "Applications of artificial intelligence in engineering and manufacturing: a systematic review". *Journal of Intelligent Manufacturing* 33(6):1581–1601.

Overbeck, L., O. Brützel, M. Teufel, N. Stricker, A. Kuhnle and G. Lanza. 2021. "Continuous adaption through real data analysis turn simulation models into digital twins". *Procedia CIRP* 104:98–103.

Overbeck, L., S. C. Graves, and G. Lanza. 2023. "Development and analysis of digital twins of production systems". *International Journal of Production Research*:1–15.

Park, K. T., S. H. Lee, and S. D. Noh. 2022. "Information fusion and systematic logic library-generation methods for self-configuration of autonomous digital twin". *Journal of Intelligent Manufacturing* 33(8):2409–2439.

Qi, Q., F. Tao, Y. Zuo, and D. Zhao. 2018. "Digital twin service towards smart manufacturing". *Procedia Cirp* 72:237–242.

Rao, Y., F. He, X. Shao, and C. Zhang. 2008. "On-Line simulation for shop floor control in manufacturing execution system". *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5315 LNAI(PART 2):141–150 https://doi.org/10.1007/978-3-540-88518-4_16.

Reed, S., M. Löfstrand, and J. Andrews. 2021. "Modelling stochastic behaviour in simulation digital twins through neural nets". *Journal of Simulation*:1–14.

Santillan Martinez, G., S. Sierla, T. Karhela, and V. Vyatkin. 2018. "Automatic Generation of a Simulation-based Digital Twin of an Industrial Process Plant". In *Proceedings of the 44th Annual Conference of the IEEE Industrial Electronics Society, IECON 2018*, Proceedings of the Annual Conference of the IEEE Industrial Electronics Society, 3084–3089. United States: Institute of Electrical and Electronics Engineers https://doi.org/10.1109/IECON.2018.8591464.

Scopus 2024. "List of papers". https://tinyurl.com/pm-for-dt.

Song, M., C. W. Günther, and W. M. P. der Aalst. 2008. "Trace clustering in process mining". In *International conference on business process management*, 109–120. Springer.

Takakura, Y., J. Mori, and H. Kobayashi. 2023. "A Log-Likelihood-Based Evaluation Metric for the Reproducibility and Simplicity of Logistics Graphs". *IFAC-PapersOnLine* 56(2):6174–6180.

Tan, W. J., M. G. Seok, and W. Cai. 2023. "Automatic Model Generation and Data Assimilation Framework for Cyber-Physical Production Systems". In *Proceedings of the 2023 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*, 73–83.

Tavakoli, S., A. Mousavi, and A. Komashie. 2008. "A generic framework for real-time discrete event simulation (DES) modelling". In *Proceedings - Winter Simulation Conference*, 1931–1938. Miami, FL https://doi.org/10.1109/WSC.2008.4736285.

Telli, A., T. G. Erdogan, and A. Kolukısa. 2023. "Detecting Novel Behavior and Process Enhancement with Multimodal Process Mining". In *2023 4th International Informatics and Software Engineering Conference (IISEC)*, 1–6. IEEE.

Toosinezhad, Z., D. Fahland, Ö. Köroğlu, and W. M. Van Der Aalst. 2020. "Detecting system-level behavior leading to dynamic bottlenecks". In *2020 2nd International Conference on Process Mining (ICPM)*, 17–24. IEEE.

Van Der Aalst, W. M. P. 2016. *Process mining: Data science in action* https://doi.org/10.1007/978-3-662-49851-4.

Wozniak, L. and P. Clements. 2015. "How Automotive Engineering is Taking Product Line Engineering to the Extreme". In *Proceedings of the 19th International Conference on Software Product Line*, 327–336. ACM.

Yadav, R., Y. M. Roopa, M. Lavanya, J. Ramesh, N. T. Chitra and G. R. Babu. 2023. "Smart Production and Manufacturing System Using Digital Twin Technology and Machine Learning". *SN Computer Science* 4(5):561.

Yang, M., J. Moon, J. Jeong, S. Sin and J. Kim. 2022. "A novel embedding model based on a transition system for building industry-collaborative digital twin". *Applied Sciences* 12(2):553.

Zhu, L., G. Lugaresi, and A. Matta. 2023. "Automated Generation of Digital Models for Production Lines Through State Reconstruction". In *2023 IEEE 19th International Conference on Automation Science and Engineering (CASE)*, 1–8. IEEE.

AUTHOR BIOGRAPHIES

GIOVANNI LUGARESI is Assistant Professor at the Department of Mechanical Engineering at KU Leuven. His research focuses on production planning and control, smart manufacturing systems, digital twins, robust optimization for the design and operation of production systems. His email address is giovanni.lugaresi@kuleuven.be and his website is https://www.kuleuven.be/wieiswie/en/person/00163811.