Multi-flow Process Mining for Comprehensive Simulation Model Discovery

Atieh Khodadadi*
AIFB
Karlsruhe Institute of Technology
Karlsruhe, Germany
atieh.khodadadi@kit.edu

Sanja Lazarova-Molnar
AIFB
Karlsruhe Institute of Technology
Karlsruhe, Germany
University of Southern Denmark
Odense, Denmark
sanja.lazarova-molnar@kit.edu

Abstract

Process mining has proven effective in explaining the underlying processes of systems, thereby improving systems' understanding, analysis, and operational efficiency. Process mining, however, often falls short in addressing multiple dimensions of systems' behaviors, limiting its ability to provide comprehensive insights for systems' performance and optimization opportunities. In this paper, we introduce an enhancement to conventional process mining that we term Multi-flow Process Mining (MFPM), which effectively extracts process flows across different system dimensions, such as time, energy, waste, and carbon footprint. MFPM enables a more comprehensive view of a system's dynamics, enabling holistic decision-making for enhanced system efficiency. We detail the framework of MFPM, outline corresponding data requirements, and introduce an expanded version of Petri nets—used here as a modeling formalism to describe and analyze multi-flow system processes. Through a detailed case study, we demonstrate the practical application of MFPM in capturing and analyzing multifaceted aspects of systems.

CCS Concepts

• Computing methodologies; **• Modeling and simulation**; **• Model development and analysis**; **• Modeling methodologies**;

Keywords

Process Mining, Multi-flow Processes, Petri nets, Multi-objective Decision Support

ACM Reference Format:
Atieh Khodadadi and Sanja Lazarova-Molnar. 2024. Multi-flow Process Mining for Comprehensive Simulation Model Discovery. In *2024 The 14th International Conference on Information Communication and Management (ICICM 2024), November 06–08, 2024, Paris, France*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3711609.3711612

*Corresponding Author

ICICM 2024, Paris, France

ACM ISBN 979-8-4007-1747-5/2024/11
https://doi.org/10.1145/3711609.3711612

1 Introduction

Process Mining (PM) is a developing field that facilitates intelligent integration between data science, business process management, and information systems. PM techniques facilitate the discovery of simulation models from event logs, which supports the structured creation and ongoing updates of simulation models especially for data-driven simulation models such as Digital Twins[1].

A major challenge in traditional PM is its limited scope, focusing mainly on the temporal aspects of process flows [2]. This narrow focus can lead to suboptimal decision-making, often overlooking other critical dimensions, such as cost, energy usage, or waste generation. To effectively address multiple objectives of complex systems, such as energy efficiency or cost reduction, monitoring changes across various dimensions rather than just time is essential—a capability that current PM frameworks do not adequately support. Such omissions can significantly impact the accuracy and usefulness of the generated insights, potentially leading to decisions that might not align with the organization's operational or sustainability goals. For instance, a process that is optimized for time efficiency might incur excessive costs or environmental burdens, thereby negating the perceived benefits.

In this paper, we introduce a novel approach and extension of PM that we term "Multi-flow Process Mining" (MFPM). With MFPM, we extend traditional PM by tracing process flows across multiple dimensions—beyond time—to include other system metrics, such as energy consumption, cost, carbon footprint, waste generation, etc., based on the specific goals of the systems of interest. With this, MFPM enables the discovery of comprehensive models, facilitating multi-objective optimization. MFPM enhances the understanding of processes by incorporating elements that directly affect both operational efficiency and strategic decision-making. Our approach is motivated by our work on Energy-Oriented Digital Twins for manufacturing systems [3], where we extract simulation models that capture both temporal process flows and the relevant energy-oriented process flows. For this, we needed to extend the original PM approach and adjust the related data requirements accordingly, as we generalize and detail in the following.

The structure of this paper is outlined as follows: In Section II, we provide a background of PM covering paradigms, PM model formalisms, and the existing challenges. In Section III, we introduce the concept of MFPM, detailing the data requirements and the methodology of MFPM. In Section IV, we present a case study to showcase the MFPM application in a manufacturing system. Section V addresses the challenges and limitations encountered in the application of MFPM. Finally, in Section VI, we summarize our findings and discuss potential future directions in the research and application of MFPM.

2 Background

In the following, we provide a comprehensive overview of the foundational paradigms and formalisms related to PM, particularly focusing on how we integrate Petri Nets (PNs) in MFPM. Furthermore, we address challenges and limitations inherent to existing PM formalisms, highlighting areas that require further research and development.

2.1 Process Mining

PM forms a fundamental framework in business process management. PM's goal is to extract and analyze process data from event logs to enhance organizational efficiency and compliance [4]. Each entry in an event log represents a step in a process, such as the activity start or end, cost, resource used, and the sequence of activities (trace) for each process at a particular point in time. An event log entry may also include supplementary details or attributes, such as cost, event type, resource, etc. [5]. Event log entries associated with a case are arranged in order of time, each representing a single execution of the system's process, often termed an "ID". The series of activities performed for a case is referred to as a trace. Consequently, an event log can be understood as a collection of traces, where multiple cases may belong to the same sequence of activities [6]. For instance, in manufacturing, assembling a car is a case, with events such as "body welding" and " body painting" recorded as a trace under the car's serial number as the case ID.

The three core paradigms of PM are process discovery, conformance checking, and process enhancement. Process discovery entails the automated generation of process models or diagrams from event logs. Discovery tools and algorithms, such as the alpha algorithm and heuristic miners, enable organizations to visualize and understand their operational processes in depth [7]. Conformance checking compares the actual behavior recorded in logs with the behavior defined by a process model to ensure that the observed process actions align with the expected or modeled behaviors [8]. Conformance checking aids organizations in maintaining governance, managing risks, and ensuring operational integrity. Process enhancement focuses on developing models that achieve specific properties, which is further divided into process extension and process improvement. Process improvement refines models to eliminate behaviors that lead to unsatisfactory outcomes or violate regulations, ensuring these actions are not permitted by the model [9].

In our work, we employ the process discovery paradigm of PM to effectively extract multi-flow process models from enhanced event logs that feature required data, as detailed in Section III. In the context of process discovery applications, Friederich et al. [10] demonstrated the application of PM in deriving reliability models from event data in Smart Manufacturing Systems (SMS), using a case study of a flow shop with parallel operations. Similarly, Bemthuis et al. [11] developed a proof-of-concept for extracting agent models from event logs, utilizing Schelling's segregation model to validate the effectiveness of PM techniques in agent model extraction.

2.2 Petri Nets

PNs can very well complement PM by providing a formal modeling formalism that can be used to represent, analyze, and improve business processes based on real data. One of the main applications of PM is the automatic discovery of PNs from event logs, which involves analyzing sequences of events recorded in information systems and constructing PNs that model underlying processes. Discovered PN models provide insights into process flows, deviations, and bottlenecks, facilitating optimization of the corresponding processes [12].

In a PN graph, the system's static properties, such as the topology of the network, the initial marking of tokens across the places, and the capacities of places, are represented [13]. PN graph consists of two types of nodes: circles, referred to as places, and bars, referred to as transitions. Places and transitions are linked by directed arcs, with transitions serving as input for places and vice versa. The execution of a PN is governed by the removal and creation of markers, known as tokens, represented by black dots, within the net's circles. Tokens are shifted through firing of transitions within the net, requiring transitions to be enabled by presence of tokens in their input places. Firing a transition implies removing enabling tokens from input places and generating new tokens, which are then deposited into output places [14]. There are several types of PNs, each designed for specific applications and modeling requirements. In our research, we use Stochastic Petri Nets (SPNs) as formalized and described in [15], where SPN is defined as $SPN = (P, T, A,\ G,\ m_0)$, where:

- $P = \{P_1, P_2,\ \ldots, P_m\}$ represents the set of places, depicted as circles;
- $T = \{T_1, T_2,\ \ldots, T_n\}$ comprises the set of transitions, each paired with their respective distribution functions or weights, illustrated as bars;
- $A = \{A^I \cup A^o \cup A^H\}$ categorizes the arcs, with A^O indicating output arcs, A^I denoting input arcs, and A^H representing inhibitor arcs, each arc assigned a specific multiplicity;
- $G = \{g_1, g_2,\ \ldots, g_r\}$ denotes the set of guard functions linked with various transitions;
- and m_0 signifies the initial marking that outlines the initial token distribution across the places.

Each transition is denoted by $T_i = (type,\ F)$, where $type \in \{timed,\ immediate\}$ denotes transition's type, and F is a probability distribution function for timed transitions or a firing weight or probability for immediate transitions.

2.3 Challenges in Existing Process Mining Approaches

One of the significant challenges in existing PM approaches is their limited ability to encompass multiple dimensions of processes beyond the basic temporal process flow. Current PM techniques generally neglect these other critical dimensions, such as time, cost, energy, manpower, and their interactions and interdependencies. This oversight can lead to an incomplete understanding of process dynamics, particularly in complex environments such as manufacturing systems. Existing PM frameworks, including those based on well-established models, such as PNs, struggle to effectively integrate and represent these multiple flows within their existing structures and features presenting an obstacle to achieving comprehensive process optimization. Challenges with multi-flow processes necessitate enhancements in PM approaches to reflect and improve real-world systems, ensuring that all relevant dimensions are considered for a thorough analysis and optimization more accurately.

There has been research in what has been termed as multi-dimensional [16] or multi-perspective PM [17], which differs from our approach, leading us to define the term Multi-flow Process Mining to distinguish from these approaches. Existing multidimensional PM approaches are mainly applied to healthcare and educational systems and utilize multi-dimensional data from various sources to extract entities' or process' features to refine time-oriented process flows without fully exploring different dimensions of process flows in systems such as energy consumption. For instance, the works by Vogelgesang et al. [16] focused on healthcare systems, using patient and process characteristics to optimize healthcare delivery but do not extend to other dimensions such as energy or waste. Similarly, while multi-perspective process mining, as discussed by Mannhardt et al. [17], aims to integrate diverse types of information such as control flow, data, resources, and time to create a more detailed view of processes, still considering only their temporal flow.

Noted limitations accentuate a significant gap in traditional PM approaches, where the singular focus on the temporal flow of the system omits other critical dimensions of process flows that cannot be captured by time-oriented analysis alone. For instance, heat waste generated by an instantaneous activity in a time-oriented process flow has significant implications for waste generation dynamics. However, it remains unaccounted for in time-oriented process flow in conventional PM methodologies.

3 Multi-flow Process Mining

To enhance capabilities of conventional PM in capturing different process dimensions, we introduce MFPM. MFPM effectively captures dynamics in systems in different dimensions, such as time, cost, energy, and waste generation, enabling a comprehensive analysis of system behaviors and interdependencies. To better illustrate our understanding of process flow dimensions, consider the following illustration. Process flow encompasses the sequential progression of activities within a process, with, e.g., possible focus on dimensions, such as **time** which involves the sequence and duration of activities from start to finish; **cost**, representing financial resources consumed at each step of the process; **energy**, detailing amount of energy required or consumed during process activities; **waste generation**, concerning the creation of waste materials as a result of process operations; and **carbon footprint** measuring the total amount of greenhouse gases emitted directly or indirectly by the process.

MFPM reveals that processes often vary significantly across different dimensions. For instance, an activity might consume time in a time-oriented process flow but not consume energy in an energy-oriented process flow. This highlights MFPM's role in providing comprehensive insights in system's dynamics.

Furthermore, MFPM allows us to analyze processes from multiple perspectives, enhancing our ability to identify and address previously unnoticed inefficiencies. E.g., by analyzing the energy consumption alongside the time taken for each activity, MFPM can uncover instances where a machine remains idle but still consumes power. By focusing on specific dimensions, MFPM facilitates the achievement of targeted goals, supporting multi-objective optimization of systems. Based on the objectives of the system, MFPM can extract distinct process flows across different dimensions, which can subsequently be integrated into comprehensive multi-dimensional models, enabling more holistic decision support.

3.1 Data Requirements for Multi-Flow Process Mining

Events form the backbone of event logs, enabling detailed analysis of processes and system behaviors. To enable MFPM, we first define the structure of an event log to align with the specific objectives of the system. For this, we need to capture, besides basic data points such as "Time Stamp", "ID", and "Event", additional data points that reflect the broader relevant context of the system's processes and objectives, such as energy consumption, carbon footprint, waste generation, etc., at the point of the event occurrence.

Moreover, based on the characteristics of the system, it may be necessary to categorize resources, such as types of waste (e.g., water, plastic) and types of energy sources (e.g., battery, electricity, oil). This resource categorization facilitates more effective resource management and enhances the decision-making of the system. Thus, in MFPM, we define the event log as $\{E_1, E_2, \ldots, E_m\}$, where $E_i = (ID, Timestamp, Event, \{D\_\text{Attributes}\})$. Each E_i captures essential data points utilized by the MFPM, ensuring that every relevant dimension of the process is documented for comprehensive analysis. MFPM facilitates a detailed and insightful analysis across various system dimensions by systematically integrating multi-dimensional data points into an event log. Models extracted through MFPM enable comprehensive insights and enhance understanding of the interactions between different dimensions and their KPIs, ultimately affecting overall system efficiency. MFPM leads to more effective optimizations, allowing for targeted improvements in system performance.

Based on the system configuration, data can be extracted through various tools and interfaces. While each system tool and interface may generate separate event logs, it is essential to integrate all of the separate event logs into a unified event log that encompasses all relevant dimensions. The integration process ensures that multiple logs are correlated based on common data points such as "Time Stamp" and "Event". This method of integrating data from diverse sources into a single or unified event log enables comprehensive analysis, leading to more precise conclusions about different aspects of the process behaviors, ensuring that all relevant data dimensions are considered collectively.

3.2 Multi-Flow Process Mining for Process Discovery

A key advancement in MFPM is the refined definition of process flows in extracted models, which goes beyond the traditional focus on temporal process flow. MFPM also incorporates changes in system variables across other dimensions. In MFPM-extracted models, events are related not only to the temporal progression of

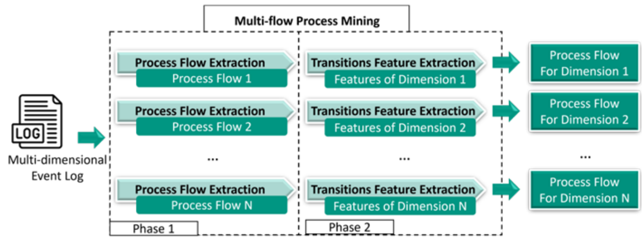

Figure 1: Multi-flow process mining framework.

processes but also to advances in other dimensions, such as energy consumption. For instance, the quantity of waste generated within a production activity might be a constant value, independent of the activity's duration, and instead influenced by the product's material composition (e.g., plastic, iron).

As we illustrated in Fig 1, the MFPM is implemented in two phases. In the first phase, we begin with extracting process flows for each dimension of interest, labeled as Process Flow 1 to N. Each flow corresponds to a specific operational aspect, such as energy consumption, carbon footprint, or waste generation. According to the SPN formalism, in this phase, the set of places (P) and transitions (T) are identified and mapped. In the second phase, the focus shifts to extracting transition features for each dimension. The second phase ensures that the specifics of each process flow are accurately represented. As a result of utilizing MFPM, we extract multi-flow models that serve as the foundation for data-driven simulation models supporting multi-objective decision-making.

Considering the SPN formalism, the second phase defines the attributes of transitions and their effects on the various dimensions, involving determining the arcs (A) that connect places to transitions and vice versa, defining guard functions (G) that control the flow based on conditions or variable states, and establishing the initial marking (m_0) that specifies the initial distribution of tokens in places. The second phase ensures that all relevant factors influencing transitions, such as energy consumption or waste generation, are integrated in the SPN model.

In Algorithm 1, we detail our framework approach to extract multi-flow models from event logs in two phases that capture different dimensions, such as energy, carbon footprint, and waste generation, in addition to time in production flow. We must ensure that the event log contains the entire process of interest, including activities that do not directly involve the dimension being analyzed, such as when an activity does not generate waste in the waste-oriented (waste generation) dimension. This comprehensive logging is essential for extracting the complete process flow and obtaining a holistic system view. With PM techniques, we extract individual process flows from the event logs. Our preferred underlying extracted model is an SPN, characterized by incorporating places and transitions, with connections extending from a place to a transition or vice versa. Effective implementation of the SPN model requires detailing the attributes of each transition, including the distribution of timed transitions and the weights of immediate transitions. We apply statistical analysis to both types for timed transitions in the process flow to identify the best-fitting probability distributions for their times.

We extract metrics beyond time duration to determine transition values for dimensions that are not time-oriented. E.g., for assessing waste generation, we extract the rate or the probability distribution that quantifies the total waste output associated with a particular activity. When determining the value of a transition where a fixed rate is not suitable, we utilize statistical methods, such as regression models or other machine learning tools, which help predict values based on the duration of the activity along with other factors, enabling a more precise extraction of activities impacts on the dimension of interest.

Algorithm 1 Extraction of the Multi-flow Models

Input: multi_dimentional_event_log
Output: multi_dimentional__model
Procedure:
For each dimension:
<u>Phase 1: Process flow extraction</u>
• extract process flows.
<u>Phase 2: Transition modification</u>
For Immediate Transitions
• Determine the probability of occurrence for each related event by dividing the count of specific event occurrences by the total count of total related events.
For Timed Transitions:
• Calculate the temporal occurrence or the duration by the difference between consecutive events in the process flow.
• Apply statistical analysis to find the best-fitting probability distribution for timed transition durations.
For Multi-Dimensional Transitions (Non-time):
If Fixed Rate:
• Track changes in dimension value from the activity start to the next event in the process flow.
• Calculate the average rate of change per unit of time for each related activity.
If Dynamic Value:
Utilize simple regression models and machine learning tools to estimate values based on the activity duration.

4 Case Study

To demonstrate the MFPM methodology, we conducted a case study of a small-scale manufacturing system. In this study, we use a Discrete Event simulation (DES) model to generate event log data, tracing energy consumption and waste generation along the production line. DES, a commonly used modeling technique for studying and analyzing manufacturing processes [18], captures each event in DES at a specific instant and marks a state change in the system. We utilized this "original model" (or ground-truth model) to generate data, which we then subjected to MFPM to rediscover the underlying multi-flow PN. In the following, we describe the case study system, and the data required to extract the multi-flow model. We, then, detail the MFPM methodology applied to our case study and provide a comprehensive overview of the process used to extract the multi-flow PN model.

Table 1: Multi-dimensional event log excerpt.

Time Stamp	ID	Asset	Energy Stamp (kWh)	Power Type	Waste Stamp (kg)	Waste Type	Event
02:59:57	280	Cell1	84.67	Electricity	1.004	Plastic	Cell1 Operation Begin
03:04:33	281	Cell2	184.20	Electricity	0.525	Plastic	Cell2 Operation End
03:04:33	281	NA	0.0	NA	0.0	NA	Order Completed
03:05:54	280	Cell1	93.82	Electricity	1.115	Plastic	Cell1 Operation End
...	...	...	...	...	...	...	...

4.1 Case Study Description

Our case study features a production line equipped with three key assets: a warehouse with automated ordering, and two assembly collaborative robots (cobots). The process begins when a customer places an order. The warehouse prepares the necessary parts, with a 50% probability of being processed on one of two lines. During assembly, plastic waste is generated. Once assembled, the product is stored in the warehouse, and the customer is notified. Both cobots operate on electricity, with the cobot on line 1 having a lower energy consumption profile than the cobot on line 2. Each cobot operates in two states: idle and active. In terms of waste, line 1 produces less waste than line 2, and the volume of waste varies within a specified range of waste generation for each line.

4.2 Data Requirements

To extract a multi-flow model using MFPM for our case study, we require a multi-dimensional event log capturing the entire production process and detailed information for each relevant dimension. Accurate extraction depends on identifying all activities linked to each asset. In our case study, we maintain event logs of the entire production system—from order placement to production completion. In [3], we defined the data requirements for extracting an energy-oriented Digital Twin model using PM techniques that we here supplement with waste generation data.

In TABLE I, we show excerpts of the event logs that encompass a 24-hour operational duration of the production line. Each entry in the table is timestamped to the second, ensuring precision in tracking and analysis. The columns capture diverse dimensions, such as energy usage measured in kilowatt-hours, the type of power utilized, and the amount of waste generated in kilograms alongside the type of waste. Events listed include the start and end of operations for specific production cells, reflecting the dynamic nature of the manufacturing environment and also serve as critical markers for identifying inefficiencies in further system analyses.

4.3 Multi-flow Process Mining of the Case Study

Next, we outline how we used MFPM to extract process flows that capture different dimensions of our case study system. We employ the PM4Py Python library [19] to extract the three distinct process flows of our case study system: time-oriented, energy-oriented, and waste-oriented process flows. The extracted processes can be then integrated into a single, unified process flow that represents all operations. Following the extraction of process flows, we extract time duration probability distributions and other dimensional values. For probability distribution fitting for timed transitions, we use libraries, such as SciPy [20]. For immediate transitions with weights, we calculate the rate of each transition by dividing the number of events for each line by the total number of events outputted from the transition. For the energy dimension, we analyze each cobot's event log to aggregate activities, calculate the energy consumed, and the duration until the next event, allowing us to compute the average energy consumption rate for each activity. For waste generation, we compile activity data from the cobots and use regression modeling to estimate waste from each activity, independent of the time of day. E.g., during assembly, the waste calculation is based on the activity data from the cobot operating on that line.

With MFPM, we aim to extract PNs that model process flows across the different dimensions of our case study system. Extracted PN models serve as simulation models to assess various what-if scenarios for system analysis and enhancement. As illustrated in Fig 2, we extracted three distinct PNs: time-oriented, waste-oriented, and energy-oriented, where each PN follows the same process flow but differs in the transitions. For instance, Transition T1, representing the new order state, is time-consuming and depicted as a white rectangle in the time-oriented PN. However, in the energy-oriented and waste-oriented PNs, the corresponding transitions E1 and W1 are shown as black rectangles, indicating immediate transitions in which no energy or waste is consumed and generated during this transition. The extracted models can be used to enhance system analysis and provide decision-making support. The conventional PM approach handles only the temporal aspect and, subsequently, extracts more limited models. With MFPM, we gain insights into various aspects of the system. E.g., through the extracted energy-oriented model, we can understand the system's energy-related behaviors and assess energy usage. Similarly, the waste-oriented model allows us to pinpoint where and how much waste is generated within the system. Utilizing these multi-flow models supports multi-objective decision-making, ensuring that enhancements address all dimensions of the system's efficiency. The models also facilitate the definition of specific Key Performance Indicators (KPIs) for our case study. We measure KPIs such as throughput and output for the production line. In the energy-oriented model, we prioritize energy efficiency, specifically electricity usage, while the waste-oriented model focuses on the volume of waste as a KPI.

5 CHALLENGES AND LIMITATIONS

In our study, we identified the following key challenges and limitations associated with MFPM:

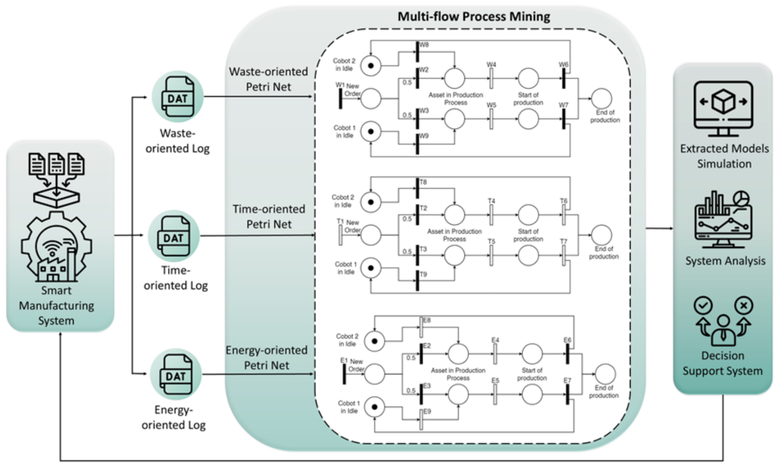

Figure 2: Multi-flow process mining for comprehensive simulation model of the case study.

- **Challenges in Extracting Multi-Dimensional Event Logs**: The extraction and integration of multi-dimensional event logs are challenging due to the need to capture data from various system dimensions simultaneously. This process often demands advanced data collection and integration technologies and techniques to ensure accurate representation and synchronization of all relevant dimensions.
- **Integration of Multi-Dimensional Models:** The challenge of integrating and unifying models that encapsulate various dimensions proves complex, particularly due to the need to synchronize and align models from each dimension. This integration demands development of novel methodologies that ensure seamless interaction among the models, accurately reflecting the system's multi-dimensional behavior.
- **Specific Behaviors in Petri Nets**: Specific behaviors in the extracted Petri net models, such as battery usage and charging dynamics in the energy dimension, need to be accurately represented to ensure that the models faithfully mirror real-world operations. To achieve this level of detail, expert knowledge may be required for precise model extraction and validation of these behaviors.

6 SUMMARY AND OUTLOOK

Process Mining (PM) enables data-driven simulation model extraction. In this paper, we introduced Multi-flow Process Mining (MFPM) as extension of traditional PM to include alternative system dimensions, beyond only time, such as energy consumption, cost, carbon footprint, waste generation, etc., based on the system and system efficiency's objectives. MFPM facilitates multi-objective optimization, enabling concurrent improvements across different process outcomes. MFPM provides a thorough understanding of processes by integrating factors that influence both operational efficiency and strategic decision-making. Furthermore, the MFPM significantly enhances the scalability of PM when applied to very large or highly complex systems by allowing each dimension to be isolated and addressed individually. This segmentation simplifies the analysis and significantly scales the MFPM approach, making handling systems with multiple complex dimensions feasible. Through a case study of a smart manufacturing system, we showcased the methodology of MFPM for three dimensions: time, energy, and waste flow. Our case study demonstrated how with MFPM we can extract system behaviors for different aspects and enable data-driven simulation models supporting multi-objective decision support.

We aim to apply MFPM to more complex systems in the future, incorporating a broader range of dimensions. Our goal is to develop a methodology that combines process flows from each dimension into a unique, multi-flow process to integrate various aspects such as energy usage, waste generation, and other relevant dimensions into a cohesive framework. Furthermore, we plan to develop a multi-flow process simulation tool to streamline the implementation of MFPM across various systems.

Acknowledgments

The authors extend their thanks for the funding received from the ONE4ALL project funded by the European Commission, Horizon Europe Programme under Grant Agreement No. 101091877.

References

[1] Jonas Friederich, Deena P Francis, Sanja Lazarova-Molnar, and Nader Mohamed. 2022. A framework for data-driven digital twins of smart manufacturing systems. *Computers in Industry* 136 (2022), 103586.
[2] Karim Nadim, Ahmed Ragab, and Mohamed-Salah Ouali. 2023. Data-driven dynamic causality analysis of industrial systems using interpretable machine learning and process mining. *Journal of Intelligent Manufacturing* 34, 1 (2023), 57–83.
[3] Atieh Khodadadi and Sanja Lazarova-Molnar. 2024. Essential Data Requirements for Industrial Energy Efficiency with Digital Twins: A Case Study Analysis. *Procedia Computer Science* 238 (2024), 631–638.
[4] Dominik Andreas Fischer. 2023. Advancing Process Mining from the Core: Managing Process Mining Project Portfolios from Data Processing to Process Improvement. Ph. D. Dissertation.
[5] Mohammed Oussama Kherbouche, Nassim Laga, and Pierre-Aymeric Masse. 2016. Towards a better assessment of event logs quality. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 1–8.

[6] Wil MP Van der Aalst. 2014. Process mining in the large: a tutorial. Business Intelligence: Third European Summer School, eBISS 2013, Dagstuhl Castle, Germany, July 7-12, 2013, Tutorial Lectures 3 (2014), 33–76.

[7] Mohammed Abdulhakim Al-Absi and Hind R'bigui. 2023. Process Discovery Techniques Recommendation Framework. *Electronics* 12, 14 (2023), 3108.

[8] Paolo Felli, Alessandro Gianola, Marco Montali, Andrey Rivkin, and Sarah Winkler. 2023. Data-aware conformance checking with SMT. *Information Systems* 117 (2023), 102230.

[9] Massimiliano de Leoni. 2022. Foundations of process enhancement. In *Process Mining Handbook.* Springer, 243–273.

[10] Jonas Friederich and Sanja Lazarova-Molnar. 2022. Data-driven reliability modeling of smart manufacturing systems using process mining. In *2022 Winter Simulation Conference (WSC)*. IEEE, 2534–2545.

[11] Rob H Bemthuis and Sanja Lazarova-Molnar. 2022. Discovering agent models using process mining: Initial approach and a case study. *In 2022 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*. IEEE, 163–172.

[12] Wil MP Van der Aalst and Anton JMM Weijters. 2004. Process mining: a research agenda. C*omputers in industry* 53, 3 (2004), 231–244.

[13] Ina Koch. 2010. Petri nets–a mathematical formalism to analyze chemical reaction networks. *Molecular Informatics* 29, 12 (2010), 838–843.

[14] James L Peterson. 1977. Petri nets. *ACM Computing Surveys (CSUR)* 9, 3 (1977), 223–252.

[15] Sanja Lazarova-Molnar. 2005. *The proxel-based method: Formalisation, analysis and applications*. Ph.D. Dissertation. Otto-von-Guericke-Universität Magdeburg, Universitätsbibliothek.

[16] Thomas Vogelgesang, Georg Kaes, Stefanie Rinderle-Ma, and Hans-Jürgen Appelrath. 2016. Multidimensional process mining: Questions, requirements, and limitations. (2016).

[17] Felix Mannhardt. 2018. Multi-perspective process mining. In *16th International Conference on Business Process Management (BPM 2018)*. CEUR-WS. org, 41–45.

[18] Oleghe Omogbai and Konstantinos Salonitis. 2016. Manufacturing system lean improvement design using discrete event simulation. *Procedia CIRP* 57 (2016), 195–200.

[19] Alessandro Berti, Sebastiaan van Zelst, and Daniel Schuster. 2023. PM4Py: A process mining library for Python. *Software Impacts* 17 (2023), 100556.

[20] Ralf Gommers, Pauli Virtanen, Evgeni Burovski, Warren Weckesser, Travis E Oliphant, Matt Haberland, David Cournapeau, Tyler Reddy, Pearu Peterson, Andrew Nelson, *et al*. 2022. scipy/scipy: SciPy 1.9. 0rc2. *Zenodo* (2022).