Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

MULTI-FLOW PROCESS MINING AS AN ENABLER FOR COMPREHENSIVE DIGITAL TWINS OF MANUFACTURING SYSTEMS

Atieh Khodadadi[1] and Sanja Lazarova-Molnar[1,2]

[1]Institute AIFB, Karlsruhe Institute of Technology, Karlsruhe, GERMANY
[2]The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense, DENMARK

ABSTRACT

Process Mining (PM) has proven useful for extracting Digital Twin (DT) simulation models for manufacturing systems. PM is a family of approaches designed to capture temporal process flows by analyzing event logs that contain time-stamped records of relevant events. With the widespread availability of sensors in modern manufacturing systems, events can be tracked across multiple process dimensions beyond time, enabling a more comprehensive performance analysis. Some of these dimensions include energy and waste. By integrating and treating these dimensions analogously to time, we enable the use of PM to extract process flows along multiple dimensions, an approach we refer to as multi-flow PM. The resulting models that capture multiple dimensions are ultimately combined to enable comprehensive DTs that support multi-objective decision-making. In this paper, we present our approach to generating these multidimensional discrete-event models and, through an illustrative case study, demonstrate how they can be utilized for multi-objective decision support.

1 INTRODUCTION

The continuous advancement in digitalization presents new opportunities for optimization in manufacturing enterprises. The growing numbers of sensors in manufacturing systems generate substantial amounts of data. This data is a critical asset for maintaining competitiveness in the fluctuating global market (Groggert et al. 2017). Systematically analyzing data from manufacturing with computational methods enhances decision-making, thereby improving the efficiency of Smart Manufacturing Systems (SMSs) (Shao et al. 2014). SMSs represent a specialized use of big data, adapting its technologies and methods to meet manufacturing-specific requirements (O’Donovan et al. 2015). Additionally, SMSs incorporate advanced technologies such as Machine Learning (ML), simulation, the Internet of Things (IoT), and cyber-physical systems (CPSs) (O’Donovan et al. 2015). These technologies integrate the physical and digital worlds, with recent developments mainly focused on Digital Twins (DTs) (Liu et al. 2021).

A DT replicates the behavior of a physical object, process, or service. Configuring a DT of an SMS entails the integration of physical entities, their virtual counterparts, and a corresponding bidirectional interface, supporting data collection, validation, knowledge extraction, and model verification (Friederich et al. 2022). The core of DTs is the data-driven extraction of systems’ models and the simulation of the models as virtual counterparts. Due to the discrete nature of manufacturing systems, the Discrete-Event Simulation (DES) paradigm is frequently employed among various simulation paradigms (Li et al. 2021). Process Mining (PM) algorithms can extract discrete event models by analyzing event logs and extracting process behaviors (Jadrić et al. 2020). These PM algorithms facilitate the extraction and continuous updating of system models, thereby enhancing the efficiency and intuitiveness of system analysis, which is fundamental for maintaining a smooth DT lifecycle.

With the widespread use of sensors in SMSs, it is now possible to track and record the impacts of events on additional process dimensions beyond time, such as energy consumption and waste generation.

Considering these dimensions is essential for analyzing and improving system performance across multiple objectives. As a result, there is a need for methodologies that capture and model system behaviors across multiple dimensions beyond time, such as energy consumption and waste generation. The conventional DES approach tracks system changes at distinct points in time, updating the simulation clock accordingly (Zeigler et al. 2000). In comparison, our multidimensional approach treats events as instantaneous occurrences across multiple system dimensions, each with its own simulation clock that is updated similarly to the primary time-based simulation clock. To automatically extract underlying multidimensional models from event logs, we developed an innovative PM methodology, termed Multi-Flow Process Mining (MFPM) (Khodadadi and Lazarova-Molnar 2024). In this paper, we further our research by presenting a multidimensional DT framework, outlining the systematic steps required to construct a DT simulation model for an SMS utilizing MFPM. Multidimensional DTs mimic a system's behaviors across various dimensions of interest, facilitating an intuitive data-driven model extraction approach to enable comprehensive system understanding and support multi-objective decision-making. We demonstrate the extraction and validation of a multidimensional DT model through an illustrative case study.

We structured the paper as follows: In Section 2, we cover the basics of our research. In Section 3, we outline our proposed methodology for extracting multidimensional DTs. In Section 4, we present an illustrative case study to demonstrate the development process of multidimensional DTs in an SMS. Finally, in Section 5, we summarize our findings and discuss the challenges and future advancements for multidimensional DTs.

2 BACKGROUND AND RELATED WORK

In this section, we provide the foundation for understanding the various components and methodologies for DTs within SMSs, and we review related research in this area.

2.1 Digital Twins of Smart Manufacturing Systems

DTs facilitate the data-driven modeling and simulation of SMSs and provide a better insight of system behavior that supports comprehensive analysis and optimization. The application of DTs differs based on specific factors such as the stage of manufacturing, the processes involved, and the particular industry sector (Liu et al. 2024). Applications of DTs are partially classified into several key areas: product design, production management and control, manufacturing system design, system fault diagnosis, risk prevention, production data management, and manufacturing system management. These applications cover component design and manufacturing, system design and maintenance, including dynamic design execution and risk prediction, and lifecycle management with a focus on data and process management (Liu et al. 2024).

Friederich et al. (2022) developed a framework for data-driven DTs of SMSs. In this framework, the SMSs, as the modeled real-world entity, continuously collect data via its IoT devices and sensors, initiating data-driven modeling. As illustrated in Figure 1, the data-driven modeling process begins with the identification of key entities, such as production systems and control technologies, followed by data storage in structured databases. The subsequent phase involves data validation, which includes data cleaning, preprocessing, and integration. The extracted data contains information about critical events within the factory. The detection and labeling of events can be improved by using unsupervised learning techniques, such as clustering (Vaarandi 2003). Experts review the clusters to ensure accuracy, provide necessary labels, and adjust as required. ML algorithms then continuously and automatically detect events using this curated data to create comprehensive event logs. The labeled logs support the discovery of processes by employing PM algorithms (Van Der Aalst 2012), forming the basis for building a simulation model of the SMS. To ensure that the DT accurately mimics the real-world system behavior, it needs to be validated. Once the DT model is validated, the model parameters are archived for future use. The validated DT model is then employed to conduct simulation runs and various what-if analyses as part of a broader simulation study. These activities are assessed using predefined Key Performance Indicators (KPIs) to evaluate their

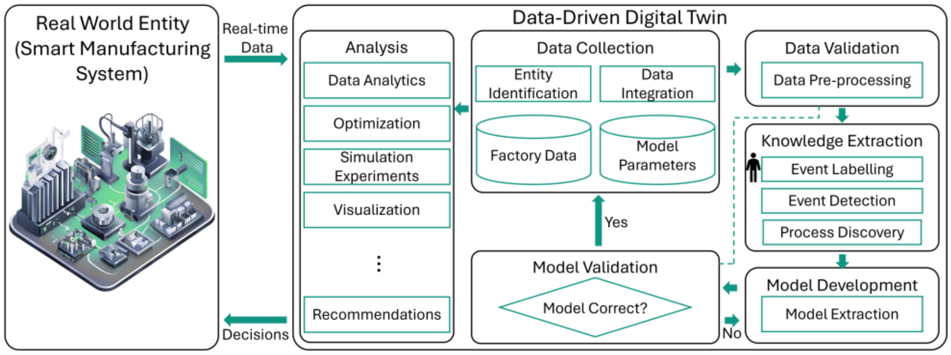

Figure 1: Framework for the data-driven Digital Twins of smart manufacturing systems (Lazarova-Molnar 2024).

effectiveness. These analyses provide stakeholders with insights for informed decisions on SMS optimization (Friederich et al. 2022).

DT can incorporate underlying simulation models, and selecting an appropriate level of abstraction for these models is a critical step in selecting simulation methodologies in DTs. The proper simulation paradigm ensures that the simulation captures essential system features while excluding non-essential details, thus developing DTs balanced between oversimplicity and complexity. DES is selected for its precision in modeling discrete event sequences, making it a preferred choice for developing DTs for SMSs (Li et al. 2021). In this paper, we focus on DES as the simulation paradigm for DTs to accurately represent system dynamics and process interactions in SMSs. In DES, systems are represented by state variables that are updated at discrete intervals (Varga 2001). Each event occurs at a distinct, predetermined point in time, and the simulation clock advances to these points to update the system's state. The clock tracks the progression of simulation time, ensuring that state variables are updated only at these discrete intervals.

2.2 Process Mining for Digital Twin Model Extraction

PM employs data-driven techniques to extract DES system models, using event logs (Van Der Aalst 2012). An event log is a structured record of events, where each event is associated with a specific case, identified by a unique case "ID". Each event includes three key elements: the case ID, the performed activity, and its corresponding timestamp. The event log consists of multiple cases, each representing a sequence of events ordered by their timestamps, reflecting the progression of activities. These logs range from complex database systems (for example, patient records in a hospital) to simpler formats such as CSV files. Entries may also include attributes such as cost, event type, and resource usage for further system analysis. The three main forms of PM include process discovery, enhancement, and conformance checking, with process discovery establishing the foundation for the other two forms (Van Der Aalst 2012). Process discovery aims to extract a process model from event log data to achieve the highest levels of comprehensiveness, clarity, and accuracy. In our research, we utilize process discovery to extract multi-flow process models from enhanced event logs containing the necessary data, as will be detailed in Section 3.

The integration of PM and DTs offers enhancements in the modeling, simulation, and monitoring of industrial cyber-physical systems (Vitale et al. 2024). PM is utilized in the extraction and validation process models from event logs of SMSs in near-real-time, which are the foundation for the development of DTs (Friederich et al. 2022). Recently, there has been a growing interest in integrating PM with DTs to enhance real-time analytics and decision-making in SMSs. For example, Vitale et al. (2024) proposed a PM framework for DT development in industrial settings and conducted a case study using a Water Distribution Testbed (WDT). Vitale et al. assessed the framework's effectiveness by extracting accurate models of the WDT and anomaly detection with machine learning algorithms. In the same context, Friederich et al. (2022) presented a case study on using PM to extract a DT model of an SMS assembling a drone part, aiming to evaluate the reliability of the assembly line. Friederich et al. employed PM to derive the system's Petri net (PN) model, facilitating the identification of bottlenecks and inefficiencies, and directly supporting enhancements in reliability and production planning. However, existing PM approaches focus only on the temporal flow of the system and, therefore, cannot extract multidimensional process flows for comprehensive multidimensional DTs. To overcome this limitation, we introduced the Multi-Flow Process Mining (MFPM) (Khodadadi and Lazarova-Molnar 2024) approach, which we will explain in detail in section 3.

2.3 Stochastic Petri Nets

Process discovery techniques extract underlying processes from event logs, which can be represented using various DES modeling formalisms, such as PNs. Several modeling methods have been developed to describe the complex behaviors of manufacturing systems, with PN being one of the most effective (Kaid et al. 2015). PNs offer an intuitive modeling style, allowing for the simultaneous handling of concurrency, a process that can be simplified for analysis. PNs are built on solid mathematical principles and provide detailed insights into the structure and behavior of the system. Additionally, PNs support both qualitative and quantitative analysis, making them highly effective for complex DES systems (Heiner et al. 2008).

In a PN diagram, two primary node types are utilized: circles, known as places, and rectangles, referred to as transitions. Places and transitions are interconnected by directed arcs, with arcs from places to transitions indicating inputs and arcs from transitions to places indicating outputs. PN operates by the distribution of markers within the net, known as tokens, which are depicted as black dots within the places. Transitions in the system destroy the required number of tokens at each of their input places and generate the defined token at each of their possible output places upon activation. (Peterson 1977).

PNs are available in various forms and extensions, each designed for distinct modeling purposes and specific applications. In our study, we utilize Stochastic Petri Nets (SPNs), as formalized and described in (Lazarova-Molnar 2005), where SPN is defined as $SPN = (P, T, A, G, m_0)$, where $P = \{P_1, P_2, \dots, P_m\}$ represents the set of places, represented as circles. $T = \{T_1, T_2, \dots, T_n\}$ forms the set of transitions, each associated with either distribution functions or weights, represented as bars. $A = \{A^I \cup A^O \cup A^H\}$ classifies the arcs into input arcs A^I, output arcs A^O, and inhibitor arcs A^H, where each arc carries a specific multiplicity. $G = \{g_1, g_2, \dots, g_r\}$ indicates the guard functions linked to various transitions. m_0 indicates the initial distribution of tokens across the places (initial marking). Each transition is represented as $T_i = (type, F)$, is classified by $type \in \{timed, immediate\}$. F is a probability distribution function for timed transitions, and for immediate transitions, F is a firing weight or probability.

2.4 Modeling Formalisms with Ability to Capture Multidimensionality

Several alternative modeling methods have been explored for representing multidimensional systems, typically limited to two dimensions of interest, such as time and cost. The Discrete Event System Specification (DEVS) (Zeigler et al. 2000) is a formalism designed for modeling and analyzing discrete event systems. DEVS supports modular and composable modeling and simulation through smaller *atomic models*, each with its own states and transitions, which can be interconnected to create more complex *coupled models*. As DEVS is increasingly applied to complex systems, such as the modeling of intricate physical continuous systems, ensuring the validity and quality of simulation data, as well as its precise management and analysis, turns into a significant challenge (Wainer and Govind 2024; Moreno et al. 2010).

Colored Petri Nets (CPNs) (Davidrajuh 2023; Gehlot and Nigro 2010) enhance traditional Petri nets by integrating data values, referred to as "colors," into tokens. This addition enables the representation of intricate system states and supports the modeling of concurrent and distributed systems. While CPNs theoretically facilitate multidimensional modeling through the assignment of structured data types to tokens, encapsulating various aspects of system behavior, their practical use poses challenges.

Another approach explored for multidimensional modeling is hybrid modeling (Fakhimi and Mustafee 2024; Brito et al. 2011), which combines various modeling techniques, such as discrete events and continuous modeling. The possible challenge in hybrid modeling is the accurate representation of continuous dynamics, which requires formulating mathematical equations that capture system behaviors over time. This task is complex, often necessitating a thorough understanding of the physical processes and intricate interactions within the system, such as temperature fluctuations or energy flows.

The modeling methods outlined in this subsection can accommodate multiple dimensions beyond temporal flow. However, they generally cannot support the representation required for automated, data-driven model extraction through PM. In this context, SPNs facilitate the modeling of a system extracted from PM (Van Dongen et al. 2009), however, they do not adequately capture the multidimensional behavior of complex systems. To address this limitation, in this paper, we extended the SPN formalism to adjust to the representation of multidimensional systems for multidimensional DTs.

3 MULTIDIMENSIONAL DIGITAL TWINS

The goal of employing DTs is to enhance decision-making in complex systems, such as SMSs, which often have multiple objectives such as increasing throughput, enhancing energy efficiency, and reducing waste and CO_2 emissions. Achieving a multi-objective optimized system requires a comprehensive understanding of the system's behavior from multiple perspectives. To address this, we propose a novel *multidimensional DT* that can extract and simulate the multidimensional behavior of a complex system based on its diverse objectives. For this, we introduce a *multidimensional DT framework*, where we employ an extension of the SPN modeling formalism, offering an intuitive and structured representation of systems' behaviors across multiple dimensions. SPNs are increasingly explored in PM research for data-driven model extraction in complex systems (Van Der Aalst 2012), and have been used as simulation backbones for DTs (Friederich et al. 2022). Our previously developed PM approach, MFPM, automatically extracts multidimensional models, enabling near-real-time systems analysis along multiple dimensions and multi-objective optimization. MFPM thus serves as a key enabler of multidimensional DTs.

In Figure 2, we illustrate our multidimensional DT framework, which builds upon the framework proposed by Friederich et al. (2022). Our multidimensional DT framework differs mainly in the data requirements, model extraction, and model development. Data is gathered from IoT devices throughout the system, capturing metrics across multiple dimensions, such as energy consumption of assets and waste production from processes. We then organize the gathered data into extended event logs, referred to as *multidimensional event logs*. To facilitate a multidimensional analysis of the system, the event log is enhanced beyond basic data points such as "Time Stamp," "ID," and "Event" by including additional metrics. These metrics offer the necessary details to understand system behavior across various relevant dimensions, including energy consumption (energy stamp), carbon footprint (CO_2 stamp), and waste

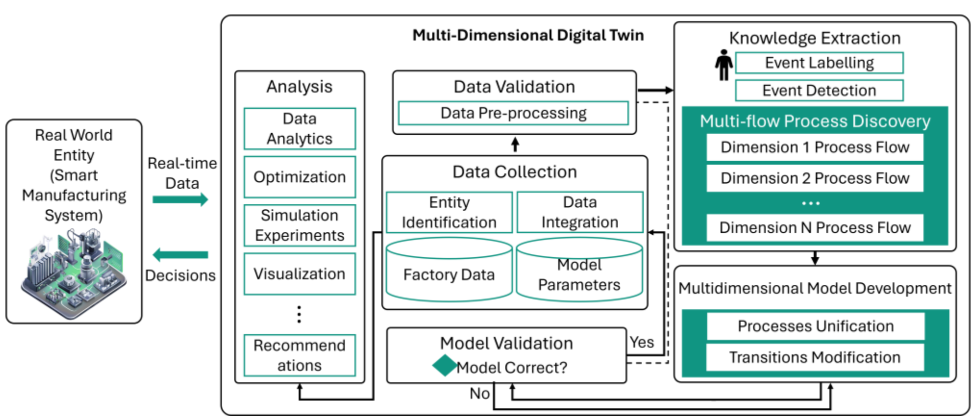

Figure 2: Multidimensional Digital Twin framework.

generation (waste stamp) at the point in time each event occurs. If needed, tracked dimensions can be further refined and categorized, such as into types of waste (water, plastic) or energy sources (battery, electricity).

Although time is typically the primary focus, PM has also been used to analyze other process dimensions, such as cost (Velásquez 2023). We apply MFPM on validated multidimensional event logs to identify a system's behavior across multiple dimensions. In MFPM, we use conventional PM and extract unidimensional models for each dimension of interest. In Figure 3, we present the workflow of MFPM, structured in two phases. The first phase involves extracting individual process flows for each dimension, designated as Process Flow 1 through N. Following the Petri nets modeling elements, the first phase involves identifying and mapping the sets of places (P) and transitions (T), the set of arcs (A) that link places to transitions, defining guard functions (G) that regulate flow based on conditions or states, and extracting the initial marking (m_0) to denote the initial token distribution across places. In the second phase, we extract the multidimensional transitions' attributes, presenting their impact on different dimensions.

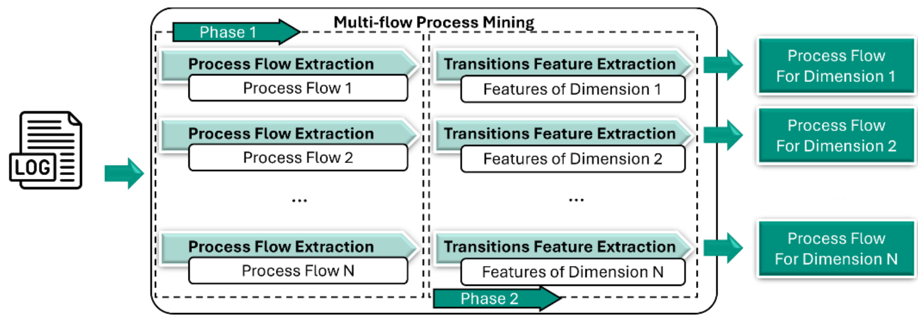

Figure 3: Multi-flow process mining framework (Khodadadi and Lazarova-Molnar 2024).

In Algorithm 1, we outline the steps of our framework for extracting unidimensional SPN models from event logs across any relevant dimension. The input, denoted as $E = \{e_1, e_2, \dots, e_n\}$, represents a multidimensional event log, capturing all system dimensions (D). The output of the MFPM is a set of unidimensional SPNs (U), each reflecting the flow and transition dynamics for a specific dimension.

Phase 1 concerns the extraction of the structures of the process flows for each dimension, denoted as $F_d = (P_d, T_d)$, where P_d is the set of places and T_d is the set of transitions in each dimension (Line 2). Phase 2 enriches these transitions with further specification and quantitative attributes relevant to the currently extracted dimension. For this, we first analyze the time dimension separately: if the currently extracted dimension is time, the duration of activities is evaluated to determine the best-fitting probability distributions ($p(T_\Delta)$) for those transitions (Line 4). In addition, for noncontributing transitions to the time dimension with multiple output selections, the probability of each output is achieved by counting the number of occurrences of each selection option (Line 5). For non-time dimensions, the impact of a contributing transition can be represented in one of two ways: as a rated value (R_d, calculated by multiplying its rate by the time duration, Line 8), or as a fixed or dynamic value which can be calculated using the best-fitting probability distributions, $p(T_\Delta)$ or ML techniques, capturing variability based on observed data ($W_d(t)$) (Line 9).

Following the extraction of the unidimensional models that consist of individual SPN models corresponding to the distinct dimensions, these individual models are integrated into a single unified comprehensive simulation model, termed the multidimensional model, as the basis of simulation, which is described next, in Algorithm 2. The resulting unified multidimensional simulation model subsequently undergoes a validation process to ensure its accuracy and reliability. The validated model is employed in systematic analyses aimed at enhancing and optimizing the system in various objectives and dimensions.

Algorithm 1: Unidimensional models extraction.

Input: $E = \{e_1, e_2, \dots, e_n\}$
Output: $U = \{F_1, F_2, \dots, F_m\}$, $F_d = (P_d, T_d)$
Foreach $d \in D$ **do:**
  //Phase 1: Process Flow Extraction
  $F_d = \{f_1^d, f_2^d, \dots, f_n^d\}$
  //Phase 2: Multidimensional Transition Modification
  Foreach transition in the time dimension:
    If contributing transition: calculate the best-fitting probability distributions: $T_\Delta = t_{i+1} - t_i$; $p(T_\Delta) = argmin_p(fit(T_\Delta, p))$
    If noncontributing transition with multiple output selection: count the number of occurrences of each selection option.
  End
  Foreach contributing transition to dimensions other than time:
    If rated value: $R_d = \frac{\Delta V_d}{\Delta t}$, $\Delta V_d = V_{i+1}^d - V_i^d$, $\Delta T = t_{i+1} - t_i$
    If fixed or dynamic value: calculate $p(T_\Delta)$ or $W_d(t) = f_{ML}(t, \theta)$
  End
End

3.1 Multidimensional Digital Twin Model

Next, we outline the methodology for integrating the extracted unidimensional SPN models into a unified Multidimensional SPN (MDSPN), a novel modeling approach introduced in this work. MDSPNs extend traditional SPNs by allowing each transition to exhibit distinct behaviors across multiple dimensions. This unified model serves as the foundation for the simulation. To simulate the MDSPN model, we assign distinct simulation clocks to each dimension. The clock in the temporal dimension tracks the progression of time, while clocks in other dimensions track updates specific to their respective attributes. These clocks advance with each transition firing, ensuring an accurate and synchronized representation of all relevant dimensions within the discrete-event system.

Algorithm 2: Unification of unidimensional models into one multidimensional model.

Input: Unidimensional SPN models $U = \{F_1, F_2, \dots, F_m\}$, $F_d = (P_d, T_d)$
Output: Unified multidimensional SPN model MDSPN, with associated "multidimensional Transitions" $\{MT_1, MT_2, \dots, MT_m\}$
Procedure:
Initialize the complete flow
Split transitions into segments corresponding to the number of dimensions (m)
Foreach transition in the model t:
  Identify the contributing and noncontributing dimensions
  Define the impact of the transition in each contributing dimension
  Adjust other SPN specifications
end

In Algorithm 2, we outline our methodology for integrating unidimensional SPN models into a unified MDSPN model. The input to this process consists of unidimensional SPN models extracted using the MFPM approach. In MDSPN, each transition is decomposed into segments corresponding to the number of dimensions of interest (m), representing their contribution type in each respective dimension. Each transition segment is color-coded, white for contributing and black for noncontributing transitions, to indicate the transition's impact on specific dimensions visually. Additionally, we specify the quantitative impact of each transition on its respective dimensions and define other model specifications such as guard functions, selection probabilities, and so forth.

In Figure 4, we illustrate an example of a unification process to extract and assess an MDSPN model. Here, we use a manufacturing line segment that incorporates the dimensions of time and energy. In this scenario, a new order arrives after a certain duration without effect on the energy dimension. Concurrently, a robot in an idle state consumes energy until a new order arrives, at which point it transitions immediately to the production process. Although the temporal flow features an immediate transition, this transition contributes added value to the energy dimension. Upon completion of production, which impacts the energy dimension, the robot reverts to its idle state, and the production advances to subsequent stages. Following the unification process, we employ multidimensional transitions (labeled MT).

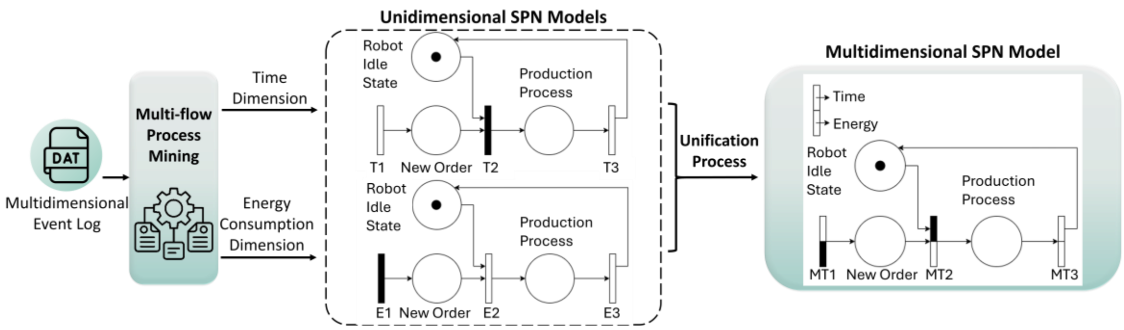

Figure 4: Unification of different dimensions models into one multidimensional model.

4 ILLUSTRATIVE CASE STUDY

To demonstrate the methodology for developing and simulating multidimensional DT models, we conducted an illustrative case study of an SMS. This study employs DES based on the MDSPN formalism to represent system dynamics. The model encompasses three key dimensions: time, energy (sourced from the grid), and product waste, each linked to specific KPIs. In our study, we simulate what we refer to as the 'ground truth model' to generate data used as a basis for (re)discovering the underlying multidimensional model. The process begins with a detailed description of the SMS, followed by identifying essential data for extracting the multidimensional models. We then apply the MFPM methodology to extract the underlying unidimensional models. Upon deriving the MDSPN model, we simulate the model using the MDPySPN simulation library (Khodadadi and Lazarova-Molnar 2025a). Finally, we validate the extracted multidimensional model against the ground truth model, comparing the defined KPIs for each dimension. The simulation code and associated resources are publicly accessible on GitHub (Khodadadi and Lazarova-Molnar 2025b).

4.1 Case Study Model Description

Our illustrative case study is a simple example of a production line focused on three dimensions of time, energy consumption, and waste generation that includes two production robots. The production process initiates with the arrival of a new order, which is randomly assigned to one of the robots, each having a 50% probability of selection. Following the production process, the product is completed and stored in the warehouse, which then alerts the user. Relevant to the energy consumption and waste generation dimensions, the robots are powered by the electrical grid and operate in two modes, active and idle, each characterized by distinct energy consumption and waste generation profiles. Robots generate plastic waste during production. In our case study, the KPIs are structured around multiple dimensions, including:

- Time Dimension (Basic): number of output products and throughput of orders.
- Energy Dimension: total energy consumption measured in kWh from the grid, tracking each asset's energy usage.
- Waste Dimension: total product waste measured in kg.

4.2 Case Study Data Requirements

To extract event logs from each asset, it is essential to catalog all activities of each asset, including both non-value-adding activities, such as idle energy states, and value-adding activities, such as machining or assembling components during the manufacturing process. Event logs are continuously generated and dynamically updated throughout operations to reflect system activities. Entries in the logs are specifically added at the start and end of activities, ensuring that any potential inefficiencies or waste occurring between events are accurately documented. Events irrelevant to a specific dimension are assigned a zero or “NA” value to signify their exclusion. Event logs, extracted from the system, encompass a 24-hour operational period of the production line, with each data point recording time details to the second. In Table 1, we presented a subset of the integrated event logs pertinent to the case study production line. For instance, the "New Order" event does not involve any assets or affect other dimensions, whereas the "Robot 2 Operation Begin" impacts all dimensions of time, energy consumption, and product waste generation.

Table 1: Multidimensional event log excerpt.

Time Stamp	ID	Asset	Energy Stamp (kWh)	Energy Type	Waste Stamp (kg)	Waste Type	Event
00:09:22	334	NA	0.0	NA	0.0	NA	Queue End
00:09:22	10043	Robot 2	340.24	Electricity	0.0	NA	Robot 2 Idle End
00:09:22	334	Robot 2	340.24	Electricity	3.14	Plastic	Robot 2 Operation Begin
00:09:26	335	NA	0.0	NA	0.0	NA	New Order
…	…	…	…	…	…	…	…

4.3 Case Study Multidimensional Digital Twin

In our case study, we employed the extracted event logs from the ground truth model as input for the MFPM. For this, we employed process discovery methods (Van Der Aalst 2012) to extract the processes. Next, we utilized SciPy (Virtanen et al. 2020) to determine the probability distributions for timed transitions. Further analysis entailed extracting probabilities associated with immediate transitions, alongside detailed energy and waste-related information, including both rate and fixed value additions for each transition that affects these dimensions. Following the extraction of the unidimensional SPN models, we integrate them into a unified MDSPN model. In Figure 5, we show the extracted MDSPN model from the case study system, which we then simulate with an extended version of MDPySPN capable of KPI extraction.

To simulate varied behaviors of a single transition represented by the MDSPNs in MDPySPN, we extended the traditional SPN simulation techniques to manage different behaviors across time and other dimensions. For example, the different behavior of one transition is in the idle state of an asset which refers to a condition where a system or component is operational but not currently engaged in any active processes, such as the production process. Here, the idle process (represented by transition MT11) is an immediate (noncontributing) transition in the time dimension but contributes to the energy dimension. To address such different behaviors by one transition, MDPySPN allows for tracking the time that tokens remain in idle states. The energy consumed during the idle state is calculated by multiplying the time by the robot's energy

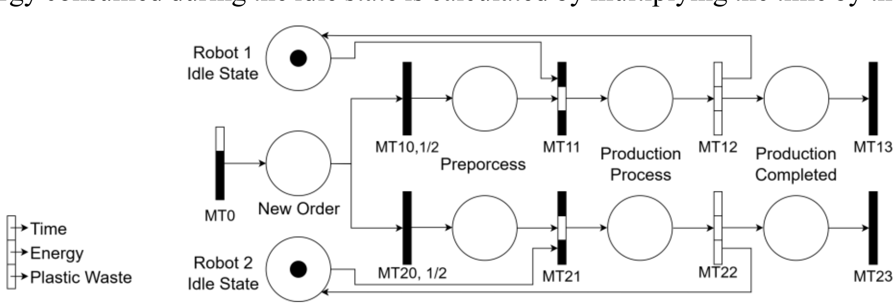

Figure 5: Unified multidimensional model of the case study.

consumption profile. Upon firing a transition, the value is consistently added to the respective dimension's clock each time the transition is activated.

4.4 Case Study Multi-Flow Structural Model Validation

We utilized a synthetic model as a ground truth model, enabling structural validation of the model extracted through the MFPM method. In real-world scenarios, such direct validation is not feasible, typically requiring an expert's assessment to confirm the fidelity of the extracted model (Alsalamah et al. 2017). To structurally validate our model (Sargent 2010), we analyze the graph representation extracted from the simulation model, which we implemented using the MDPySPN tool. This analysis involves counting places, transitions, and arcs, examining connectivity patterns (arcs). We validate correctness by comparing with the ground truth model, confirming accurate structural reflection.

4.5 Case Study Simulation Model Validation

To validate the model, we compare the predefined KPIs from the ground truth model with those from the extracted simulation model. Output validation verifies that the 95% confidence intervals of the number of output products and throughput KPIs overlap after 100 independent replications. For other KPIs, such as total energy consumption and waste generation, the validation process considers the effect of each transition's occurrence probability. Specifically, we assess the extent to which their 95% confidence intervals are close to each other. In Figure 6, we illustrate a comparison of the number of output products and throughput, demonstrating a strong alignment between our extracted DT and the ground truth model. In Table 2, we present a comparison of KPIs, including total energy consumption and total waste generation, based on their 95% confidence intervals. This comparison confirms the accuracy of the simulation model in replicating critical operational metrics of the ground truth model.

Table 2: 95% Confidence intervals for energy consumption and waste generation KPIs.

KPI	Ground Truth Model	Digital Twin's Model
Total Energy Consumption (kWh)	[2083.67, 2107.07]	[2174.31, 2206.11]
Total Waste Generation (kg)	[25.12, 25.94]	[20.85, 21.43]

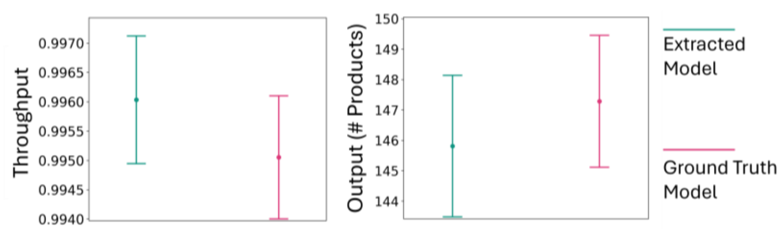

Figure 6: 95% Confidence intervals for output products and throughput KPIs.

5 SUMMARY AND OUTLOOK

Digital Twins present promising solutions for complex systems, such as in the manufacturing domain, by enabling comprehensive system analysis and enhancements. With Process Mining, Digital Twin models of real-world systems can be extracted from systems' event logs. With the widespread use of sensors in modern manufacturing systems, events can be tracked and recorded across multiple process dimensions, beyond time, such as energy and waste. Tracking systems across multiple dimensions enable a better understanding and analysis of system behavior and support stakeholders in optimized multi-objective decision-making. In this paper, we introduced a framework for extracting and simulating multidimensional Digital Twins utilizing multi-flow process mining. We demonstrate our proposed methodology through an illustrative case study of a smart manufacturing system, focusing on three key dimensions: time, energy consumption, and waste generation. Our findings identify several key challenges that must be addressed to advance the development and deployment of multidimensional Digital Twins in complex systems.

- Multi-Source Data Integration: Multidimensional Digital Twins must integrate data from multiple sources and handle a wide range of data types, from structured numerical data to unstructured textual information. The capability to process this data in (near) real-time is essential for the twin to reflect the system behavior in different dimensions correctly.
- Scalability and Dynamic Dimension: The scalability of Digital Twins to represent systems with frequent changes, along with the dynamic selection of relevant dimensions, requires effective strategies to extract accurate multidimensional Digital Twins aligned with system objectives in real-time.
- Automatic Identification of Complex Processes: In complex systems such as manufacturing, existing process mining can struggle to automatically extract certain process flows, and hinder the achievement of validated Digital Twin models. For instance, the behavior of assets in an idle state in the SPNs model.
- Clock Drift & Synchronization Delay: In real-world systems, network latency separates distributed timestamps. These errors disorder events, skew KPIs, and erode model fidelity.
- Real-World Data Validation: While our proposed technique shows promise, its application to extensive, real-world datasets has yet to be demonstrated. Thus, validating the approach of multidimensional Digital Twins with real industrial data forms part of our future work.

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

Alsalamah, A., R. Campo, V. Tanos, G. Grimbizis, Y. Van Belle, K. Hood, et al. 2017. "Face and Content Validity of the Virtual Reality Simulator 'ScanTrainer®'." *Gynecological Surgery* 14(1):18 https://doi.org/10.1186/s10397-017-1020-6.

Brito, T. B., E. F. C. Trevisan, and R. C. Botter. 2011. "A Conceptual Comparison Between Discrete and Continuous Simulation to Motivate the Hybrid Simulation Methodology." In *2011 Winter Simulation Conference (WSC)*, 3915–3927 https://doi.org/10.1109/WSC.2011.6148082.

Davidrajuh, R. 2023. *Colored Petri Nets for Modeling of Discrete Systems: A Practical Approach with GPenSIM.* Singapore: Springer Singapore https://doi.org/10.1007/978-981-99-6859-6.

Fakhimi, M., and N. Mustafee. 2024. *Hybrid Modeling and Simulation: Conceptualizations, Methods and Applications.* Cham: Springer Nature https://doi.org/10.1007/978-3-031-59999-6.

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 https://doi.org/10.1016/j.compind.2021.103586.

Gehlot, V., and C. Nigro. 2010. "An Introduction to Systems Modeling and Simulation with Colored Petri Nets." In *2010 Winter Simulation Conference (WSC)*, 104–118 https://doi.org/10.1109/WSC.2010.5679170.

Groggert, S., M. Wenking, R. H. Schmitt, and T. Friedli. 2017. "Status Quo and Future Potential of Manufacturing Data Analytics—An Empirical Study." In *Proceedings of the 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, December 10–13, Singapore, 779–783 https://doi.org/10.1109/IEEM.2017.8289997.

Heiner, M., D. Gilbert, and R. Donaldson. 2008. "Petri Nets for Systems and Synthetic Biology." In *Formal Methods for Computational Systems Biology*, edited by M. Bernardo, P. Degano, and G. Zavattaro, 215–264. Berlin, Heidelberg: Springer https://doi.org/10.1007/978-3-540-68894-5_7.

Jadrić, M., I. Ninčević Pašalić, and M. Ćukušić. 2020. "Process Mining Contributions to Discrete-Event Simulation Modelling." *Business Systems Research* 11(2):51–72 https://doi.org/10.2478/bsrj-2020-0015.

Kaid, H., A. M. El-Tamimi, E. Abouel Nasr, and A. Al-Ahmari. 2015. "Applications of Petri Nets-Based Models in Manufacturing Systems: A Review." In *Proceedings of the 2015 International Conference on Operations Excellence and Service Engineering*, September 10–11, Orlando, Florida, 516–528.

Khodadadi, A., and S. Lazarova-Molnar. 2024. "Multi-Flow Process Mining for Comprehensive Simulation Model Discovery." In *ICICM '24: Proceedings of the 2024 14th International Conference on Information Communication and Management*, November 6–8, Paris, France, 15–21 https://doi.org/10.1145/3711609.3711612.

Khodadadi, A., and S. Lazarova-Molnar. 2025a. "Multidimensional Stochastic Petri Nets: A Novel Approach to Modeling and Simulation of Stochastic Discrete-Event Systems." In *Proceedings of the 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC).*

Khodadadi, A., and S. Lazarova-Molnar. 2025b. *WSC2025.* GitHub repository. https://github.com/atikh/WSC2025 (accessed 15th August).

Lazarova-Molnar, S. 2005. *The Proxel-Based Method: Formalisation, Analysis and Applications.* Ph.D. thesis, Faculty of Computer Science, Otto-von-Guericke University Magdeburg, Magdeburg, Germany.

Lazarova-Molnar, S. 2024. "A Vision for Advancing Digital Twins Intelligence: Key Insights and Lessons from Decades of Research and Experience with Simulation." In *Proceedings of the 14th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2024)*, 5–10. Setúbal, Portugal: SciTePress. https://doi.org/10.5220/0012884800003758.

Li, W., B. H. Huynh, H. Akhtar, and K. S. Myo. 2021. "Discrete Event Simulation as a Robust Supporting Tool for Smart Manufacturing." In *Implementing Industry 4.0: The Model Factory as the Key Enabler for the Future of Manufacturing*, edited by C. Toro, W. Wang, and H. Akhtar, 287–312. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-67270-6_11.

Liu, S., P. Zheng, and J. Bao. 2024. "Digital Twin-Based Manufacturing System: A Survey Based on a Novel Reference Model." *Journal of Intelligent Manufacturing* 35(6):2517–2546 https://doi.org/10.1007/s10845-023-02172-7.

Moreno, A., J. L. Risco-Martín, and J. Aranda. 2010. "Uncovering DEVS Simulation Behaviour Throughout the Open Provenance Model." In *Proceedings of the 2010 Spring Simulation Multiconference (SpringSim '10)*, April 11–15, Orlando, Florida, Article 118, 1–8. ACM. https://doi.org/10.1145/1878537.1878661.

O'Donovan, P., K. Leahy, K. Bruton, and D. T. O'Sullivan. 2015. "An Industrial Big Data Pipeline for Data-Driven Analytics Maintenance Applications in Large-Scale Smart Manufacturing Facilities." *Journal of Big Data* 2:1–26 https://doi.org/10.1186/s40537-015-0034-z.

Peterson, J. L. 1977. "Petri Nets." *ACM Computing Surveys (CSUR)* 9(3):223–252 https://doi.org/10.1145/356698.356702.

Sargent, R. G. 2010. "Verification and Validation of Simulation Models." In *2010 Winter Simulation Conference (WSC)*, 166–183 https://doi.org/10.1109/WSC.2010.5679166.

Shao, G., S.-J. Shin, and S. Jain. 2014. "Data Analytics Using Simulation for Smart Manufacturing." In *2014 Winter Simulation Conference (WSC)*, 2192–2203 https://doi.org/10.1109/WSC.2014.7020063.

Vaarandi, R. 2003. "A Data Clustering Algorithm for Mining Patterns from Event Logs." In *Proceedings of the 3rd IEEE Workshop on IP Operations & Management (IPOM 2003)*, October 1–3, Kansas City, Missouri, USA, 119–126 https://doi.org/10.1109/IPOM.2003.1251233.

van der Aalst, W. M. P. 2012. "Process Mining." *Communications of the ACM* 55(8):76–83 https://doi.org/10.1145/2240236.2240257.

van Dongen, B. F., A. Alves de Medeiros, and L. Wen. 2009. "Process Mining: Overview and Outlook of Petri Net Discovery Algorithms." In *Transactions on Petri Nets and Other Models of Concurrency II: Special Issue on Concurrency in Process-Aware Information Systems*, edited by K. Jensen and W. M. P. van der Aalst, 225–242. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-00899-3_13.

Varga, A. 2001. "Discrete Event Simulation System." In *Proceedings of the European Simulation Multiconference (ESM'2001)*, June 6–9, Prague, Czech Republic, vol. 17, 1–7.

Velásquez, I. 2023. "Enhancing the Cost Dimension in Process Mining Through Its Application to the Mining Industry." In *BPM (Demos/Resources Forum) 2023*, September 11–15, Utrecht, The Netherlands, 28–35. CEUR-WS.org, Vol. 3469.

Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, et al. 2020. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python." *Nature Methods* 17(3):261–272 https://doi.org/10.1038/s41592-019-0686-2.

Vitale, F., S. Guarino, F. Flammini, L. Faramondi, N. Mazzocca, and R. Setola. 2024. "Process Mining for Digital Twin Development of Industrial Cyber-Physical Systems." *IEEE Transactions on Industrial Informatics* (early access). https://doi.org/10.1109/TII.2024.3465600.

Wainer, G., and S. Govind. 2024. "100 Volumes of Simulation—20 Years of DEVS Research." *SIMULATION* 100(12):1297–1318 https://doi.org/10.1177/00375497241291871.

Zeigler, B. P., H. Praehofer, and T. G. Kim. 2000. *Theory of Modeling and Simulation.* San Diego: Academic Press.

AUTHOR BIOGRAPHIES

ATIEH KHODADADI is a Ph.D. student at the Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Germany. Her research concentrates on using Digital Twins to improve energy efficiency in manufacturing systems, specifically through modeling and simulation, and data analysis. Her email address is atieh.khodadadi@kit.edu.

SANJA LAZAROVA-MOLNAR is a Professor at both the Karlsruhe Institute of Technology and the University of Southern Denmark. Her research focuses on data-driven simulation, Digital Twins, and cyber-physical systems modeling, with an emphasis on reliability and energy efficiency. She develops advanced methodologies to optimize complex systems and leads several European and national projects in these areas. Prof. Lazarova-Molnar holds leadership roles in IEEE and The Society for Modeling & Simulation International (SCS), where she currently serves as SCS Representative to the Winter Simulation Conference (WSC) Board of Directors. She was Proceedings Editor for WSC in 2019 and 2020 and serves as Associate Editor for *SIMULATION: Transactions of The Society for Modeling and Simulation International*. Her email address is lazarova-molnar@kit.edu.