Data-driven Simulation in Process Mining: Introducing a Reference Model

Mahsa Pourbafrani and Wil M.P. van der Aalst
Process and Data Science Group
RWTH Aachen University
Ahornstrasse, 52070, Germany
E-mail: mahsa.bafrani@pads.rwth-aachen.de

KEYWORDS
Process mining, discrete event simulation, event log, data-driven simulation

ABSTRACT

Different approaches are proposed for simulating processes in process mining. There are open challenges while designing the simulation models of processes: (1) the quality of the designed models is mostly evaluated using simulation results, and the models themselves do not get validated, (2) the choice of process aspects to be considered in the simulation models of processes as simulation parameters is rather arbitrary, e.g., considering multitasking, and (3) the distinction between the acquiring simulation parameters step and the parameters' regeneration step is not defined. This paper aims to introduce a reference meta-model for simulation in process mining. We derive the meta-model using the provided insights from process mining and the required parameters from the simulation techniques for simulating processes, i.e., Discrete Event Simulation (DES). This model enables the creation of process simulations and the comparison of approaches in relation to the process aspects under consideration. We illustrate the use of the model in practice by developing an automatic simulation model generation approach based on the reference model.

INTRODUCTION

Data-driven simulation models of processes are able to act as digital twins (Friederich et al. 2022), i.e., by providing a digital platform that reflects the real process while the changes in the real process are reflected in the digital twin. A holistic overview of simulation in process mining is presented by van der Aalst 2018, where process mining techniques are exploited to discover describing models, and then performance analysis techniques add to the models and transform the models into prescribing models.

The focus of this paper is *on data-driven simulation approaches in process mining to provide prescribing models*. Prescription models are process models in which the flow of activities as well as the flow of process instances through the process are clear and trackable while simulating. Currently, the accuracy of the simulation results is used to select a simulation model over others. There is no reference while developing the process simulation model using event data. A model should not only present accurate outcomes but also determine how much of the available historical information, e.g., captured and presented in event logs, is used and how close the simulation model is to the real process (van der Aalst 2010). As a result, a *reference model* is required for constructing process simulation models. We define the process simulation reference meta-model as *a model that illustrates the required simulation parameters extractable from event logs, their relations, and how a simulation model of processes can be generated while considering the provided information (insights) in the event logs.*

In the area of business process simulation, some meta-models are introduced as reference models for the design phase, such as the reference models proposed by Tumay 1996 and García et al. 2014. The provided XML standard, BPSIM[1], includes the common simulation parameters such as arrival rate and duration of tasks. In addition, a high-level meta-model for a business process simulation based on event logs is proposed by Martin et al. 2014. However, some components, such as resource pooling or queuing methods, e.g., batching, are not explicitly considered. Furthermore, the execution steps are also not considered in the current meta-models.

In this paper, (1) we investigate current approaches for simulating processes in process mining and define aspects of a process from event logs. Afterward, (2) we identify the required information for accurate simulation of processes, and then, (3) we design a reference meta-model for generating simulation models of processes. Finally, (4) we present a practical implementation of the introduced model for generating data-driven simulations of processes. This model is a reference model, i.e., it is built upon the data-driven insights of processes using their event logs and also the parameters of the requirements for simulating a process. The model is also a meta-model and can be used to develop, benchmark, and categorize existing and future simulation approaches in process mining.

In the following sections of this paper, we will first present relevant concepts in the preliminary section. In the related work section, existing work for designing simulation models of processes is introduced and

[1] https://www.bpsim.org/specifications/2.0/WFMC-BPSWG-2016-01.pdf

Communications of the ECMS, Volume 37, Issue 1,
Proceedings, ©ECMS Enrico Vicario, Romeo Bandinelli,
Virginia Fani, Michele Mastroianni (Editors) 2023
ISBN: 978-3-937436-80-7/978-3-937436-79-1 (CD) ISSN 2522-2414

compared. We explain the approach for designing the reference model in the reference meta-model section and demonstrate its usage in practice, and we conclude this paper by discussing the results and presenting future work in the conclusion and discussion section.

PRELIMINARIES

Event Log. An event log represents the execution of different process instances. The process instances are referred to as traces where they contain a sequence of events, i.e., $\sigma = \langle e_1, \ldots, e_n \rangle$. Each event of σ such as e_i contains various attributes. We denote $\#_m(ei)$ which indicates the values of attribute m. The general attributes for an event in an event log are *case ID, activity, resource*, and *timestamp*. A set of process instances forms an event log (L). For instance, for an event e_i, $\#_{activity}(e_i)$ is *register request*, $\#_{resource}(e_i)$ is *Peter*, and $\#_{caseID}$ is 3. Given event logs, insights are provided using process mining techniques.

Process Mining Techniques and Insights for Simulation. According to van der Aalst 2016, four main perspectives based on the general attributes of event logs are considered the baseline of the process mining techniques, i.e., control flow, organizational, case, and time perspectives. In terms of their inputs, we consider two general categories of techniques in backward-looking process mining: Discovery (Dis) and Conformance Checking (CC). $Dis(L)$ and $CC\ (L, M)$ provide insights into four perspectives for a given event log (L) and process model (M). All the techniques under discovery use event logs, e.g., process discovery or organizational mining. Techniques under conformance checking, use both event logs and models as inputs, e.g., deviation detection. The common backward-looking process mining techniques providing insights that can be used for executing processes, i.e., simulation and what-if analysis, are:

- Process Model Discovery: process model discovery algorithms are one of the general discovery algorithms (Dis). Process model discovery algorithms such as the Inductive Miner return the flow of activities in different notations, e.g., BPMNs, Petri Nets, Directly Follows Graphs (DFG), or Process Trees. The existing information in an activity flow is the sequence of possible paths which can be taken by a process instance when entering a process (van der Aalst 2016).
- Performance Analysis: considering the time perspective in event logs, the performance of different aspects or processes such as activities, cases, or resources can be assessed (Hornix 2007). Considering the purpose of performance analysis, they can be both Dis or CC. For instance, bottleneck analysis is a CC technique since it requires the process model as well as an event log.
- Conformance Checking: the compliance of an event log with a process model is considered w.r.t. precision and fitness (Carmona et al. 2018). How the recorded event logs are in line with the discovered models is a CC technique. Conformance checking techniques such as deviation detection are able to return the unexpected paths in a process taken by the cases.
- Social Network Analysis: all techniques w.r.t. resources, shared or hand-over of tasks in processes are considered as social network analysis (Song and van der Aalst 2008). These techniques can be categorized into both categories (Dis and CC). For instance, discovering the resources performing the same roles Dis, a set of activities that should be performed by a specific set of resources, i.e., organization, or hand-over of task which requires process models to CC.
- Decision Mining: attributes of cases are taken into account to find the possible and probable path inside the control flow that cases can take. The choices in the flow of activities can be determined using decision mining techniques (de Leoni and van der Aalst 2013) as one of Dis the techniques. When alignments in decision mining are used, they are considered as CC.
- Queue Mining: we refer to the discovery of the queues inside the process w.r.t. activities, resources, or organizations and the corresponding metrics as queue mining techniques. Performance spectra project timestamps of activities visually w.r.t. the process segments they pass, i.e., taken paths by every case in the process model (Denisove et al. 2018). It reveals the existing patterns for handling the cases, e.g., handling the cases in batches.

Sendrovich et al. 2016 propose an approach to estimate the length of existing queues inside processes using event logs. Most of these techniques require both process models and event logs, therefore, we categorize them as CC.

While the majority of the techniques are supported by visualization methods, some visualization techniques, such as Dotted Charts (Song and van der Aalst 2007), provide direct insights into the processes. The introduced techniques provide the four main perspectives inside event logs for designing executable models of processes.

Discrete Event Simulation. The flowchart of DES simulations of processes is depicted in Figure 1. It is important to note that the process model discovery and initialization steps in process mining are data-driven and supported by event logs and process mining insights. In discrete event simulation, events are considered to be the arrival of a new case and the execution of an activity (Fishman 2001). Therefore, when talking about a new event, it means the system clock gets updated based on one of the two types, i.e., the simulation clock is updated based on every occurrence of events in a discrete manner. Also, it is different from the defined event notation in event logs, i.e., specific activity in a process has been performed at a specific timestamp. For instance, in the CPN tools (Ratzer et al. 2003), each of the generated tokens which flow through the models is the case (process instance) in the context of business processes.

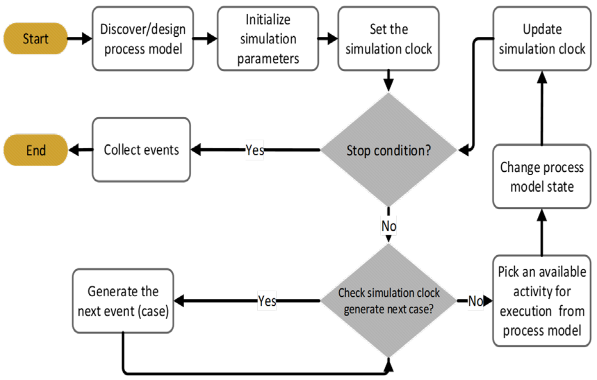

Figure 1: Process simulation flowchart using DES. It starts with discovering/designing process models and initializing the simulation parameters such as arrival rate, or execution time. The simulation models start with generating a new case (an event in DES) and after each update of the simulation clock, checking for the stop condition of the simulation.

Consider the following scenario as an example: let $s \in \mathbb{N}$ be steps of simulation and $c \in \mathbb{R}$ be the system clock. For a given set of available events to occur such as $\{A_2, E_4\}$, where A_2 is the arrival of a new case at time 2 and E_4 is the end of the activity for a case at time 4, after one step of simulation (i.e., $s = s + 1$) the earlier event (A_2) is executed and the system clock is updated (i.e., $c = c + 2$).

RELATED WORK

Different simulation techniques can be used in process mining. The most commonly used simulation techniques for simulating processes at different levels are: Discrete Event Simulation (DES) (Rozinat et al. 2009), System Dynamics (SD) (Pourbafrani and van der Aalst 2022), and Agent-Based Modeling (ABS) simulation techniques (Macal and North 2009). These simulation techniques model and simulate systems at various levels and for different objectives, see Table 1.

Table 1: Comparisons of three main simulation techniques in process mining.

Discrete Event Simulation (DES)	System Dynamics (SD)	Agent-Based simulation (ABS)
Detailed level	Aggregated level	Detailed level
Simulation steps are single events	Simulation steps are time steps	Simulation steps are single events
For decision-making and prediction	For capturing dynamic feedback behavior	Focusing on the interaction of agents in the process

Existing Data-driven Simulation Approaches

Several approaches are proposed to generate process simulation models based on process mining insights. Most business process simulations completely rely on the user and BPM techniques. They do not use event logs of processes and process mining techniques, thus it is not considered in this study. As for primary sources, we looked into Scopus and Google Scholar. Scopus queries with the main keywords *process mining*, *simulation, event logs*, and *business process simulation* were used.

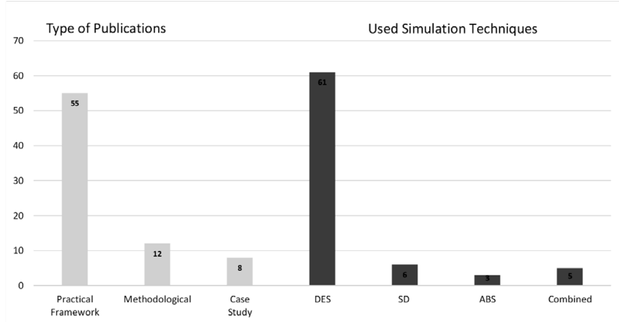

Figure 2: The number of publications w.r.t. the type of publication (black) and the used simulation techniques in the approach (gray).

Results were limited to English publications, with an emphasis on process mining keywords.[2] The search resulted in a total of 96 publications. We reviewed the results based on the focus of the publications. We identified 76 publications that directly use event logs for simulation and 62 that employ DES as the simulation technique, see Figure 2. We also categorized the publications w.r.t. their purpose, whether to design a practical framework to generate simulation models, discuss the potential of the simulation and provide methodological approaches, or use case studies such as healthcare.

[2] ALL({process mining} {simulation}{event logs})AND(LIMIT-TO(EXACTKEYWORD,"Process Mining")) AND (LIMIT-TO(LANGUAGE,"English"

Table 2: Different criteria for comparison of process simulation approaches.

Process	Activity	Resource
Case attribute Arrival rate Business hours Activity flow (process model) Decision logic (Choices in process model)	Activity duration Activity queuing pattern Activity interruption Activity resources	Schedule Shared roled (resource pooling) Social network and hand-over of work Multitasking capacity Decision logics (resource assigning rules)

For designing a reference meta-model, only practical approaches for generating DES models of processes are considered. The framework proposed by Rozinat et al. 2009 is still the most comprehensive framework in process mining for simulation. Therefore, we use this work as a baseline for comparing other approaches w.r.t. different process aspects. Table 3 represents the comparison criteria w.r.t. the process aspects. The result of comparing the approaches considering the defined criteria are shown in Figure. It is important to note that we only listed the pioneering techniques, and subsequent work that use the same approaches and do not include new aspects and parameters are not mentioned.

Data-driven Simulation Tools

To complement the review of the current approaches, we created a separate comparison for the available tools. One of the most significant components of simulation model generation is their practical execution, including the aspects that they are managed to consider. There are various tools available for business process simulation and general-purpose simulation (Jansen-Vullers and Netjes 2006).

Our focus is on the set of tools that are based on event logs or employ event logs' insights for simulation model generation and execution. Table 3 compares simulation tools in process mining. The important criteria for process simulation tools are presented in Table 2.

For instance, tools can generate a ready-to-execute simulation model such as CPN Tools as an output. Furthermore, value generators indicate various functions used for regenerating different process aspects, e.g., as random functions based on random distribution generators. Given the fact that many tools did not mention this, we excluded it from the direct comparison.

Table 3: The comparison parameters for simulation tools in process mining.

Inputs	Event logs, process models, and users.
Results	Event logs, aggregated KPIs, and executable models.
Simulation model generation	Manually designed by the user, directly simulated.
Platform	General-purpose tools, standalone tools, Python library, ProM (van Dongen, B. et al. 2005) plugin based on Java.
Visualization	Static, dynamic.
Process model notations	Petri Nets, BPMN, Process Tree, Directly Follows Graphs (DFGs).

Tool	Input	Result	Tool platform	Simulation model generation	Visualization	Process model notation
Rozinat et al.(2008a)	el, user	KPIs	ProM plugin: Java	direct simulation (results generation)	Static	PN
Rozinat et al.(2009)	el, user	exe, el	ProM plugin: Java, CPN Tools	direct transformation (model generation)	Static	PN
Pospisil et al.(2013)	el, user	KPIs	ProM plugin: Java	direct simulation (results generation)	Static	PN
Khodyrev et al.(2014)	el, user	el	ProM plugin: Java	direct simulation (results generation)	Static	PN
Gawin et al.(2015)	el, user	el	ProM plugin: Java	direct simulation (results generation)	Static	PN
Camargo et al.(2020)	el	el, KPIs	Jupyter notebook: Java (BIMP Engine), Python	direct simulation (results generation)	Static	BPMN
Pourbafrani et al. (2021)	el	exe, el	Web application: Python, CPN Tools	direct transformation (model generation)	Static	CPN
Pourbafrani et al.(2021)	el	el	Web application: Python	direct simulation (results generation)	Static	Process tree

Figure 3: Comparing simulation tools in process mining, w.r.t. execution aspects. The input can be event logs (el), users, or both. The results can be in the form of event logs (el), executable models such as CPN Tools, KPIs, or a combination of them. Models notations are Petri nets (PN), BPMN, or CPN. Simulation model generation is also considered to be directly generating a model (direct transformation), or the results (direct simulation).

Approach	Simulation Parameters												
	Case attributes	Arrival rate	Business hours	Activity flow (process model)	Activity Duration˜	Activity queuing pattern	Decision Logic (Activity Probability of happening)	role/ resource schedule	Shared Role/ resource pooling	Social networks of resources/ hand over˜	Resource Multi-tasking	Role/ Resources capacity	Decision Logic (Resource Assigning Rules)
van Beest et al.(2007)]	-	manual	-	fuzzy miner (CPN)	pm	-	pm	-	manual	-	-	manual	-
Wynn et al. (2007)	-	manual	-	manual	-	-	-	manual	manual	-	-	-	-
Rozinat et al.(2008a)	-	pm	-	general process discovery (PN)	pm	-	-		-	-	-	-	-
Rozinat et al.(2008b)	pm	pm	-	pm (WFM)	pm	-	-	-	-	-	-	-	-
Rozinat et al.(2009)	pm	pm	-	general process discovery (PN)	pm	-	pm	-	pm	-	-	-	-
Maruster et al.(2009)	pm	pm	-	heuristic miner (PN)	pm	-	pm	-	pm	-	-	-	-
Pospisil et al. (2013)	-	-	-	pm	pm	-	-	-	-	-	-	-	-
Khodyrev et al. (2014)	pm	pm	-	heuristic miner and genetic miner (PN)	pm	-	pm	-	-	-	-	-	-
Martin et al. (2014)	-	pm	-	pm, manual	pm	-	manual	manual	-	-	-	-	-
Gawin et al. (2015)	pm	pm	-	general discovery(PN)	pm	-	pm	manual	pm	-	-	-	-
Camargo et al. (2020)	pm	pm	pm	split miner (BPMN)	pm	-	pm	pm	pm	-	-	-	pm
Estrada-torta et al. (2020)	pm	pm	pm	split miner (BPMN)	pm	-	pm	pm	pm	-	pm	-	pm
Pourbafrani et al. (2021)	-	Pm	pm	inductive miner (process tree)	pm	manual	pm	-	pm	pm	-	pm	

Figure 4: The comparison of current process simulation approaches in process mining w.r.t. the presented criteria in Table 2. *pm* indicates the insights are based on event logs or process mining techniques, and WFM represents workflow nets. "-" indicates the aspect is not supported, or it is not mentioned in the paper.

DESIGNING the REFERENCE META-MODEL

Considered Process Aspects

Techniques to discover the describing models and later, values of the simulation parameters are different from the execution and regenerating values. In most of the existing approaches, the focus is only on the discovery of the simulation models and parameters while the execution of the simulation models is mostly ignored. For instance, it is considered that using event logs of processes, the average arrival rate of the cases in the process is extracted, however, it has not been systematically addressed how the extracted insights are used for reproducing the same behavior, e.g., fix values or random generator functions such as the Poisson distribution.

To create a process simulation model, two main steps should be taken: (1) the design phase for determining and extracting the required information and parameters, and (2) the execution phase for regenerating that parameter. After discovering the information such as the execution time of activities, their distribution or the aggregated value, and the way they are reproduced, using random generator functions, e.g., based on a normal distribution with the discovered average or fixed value.

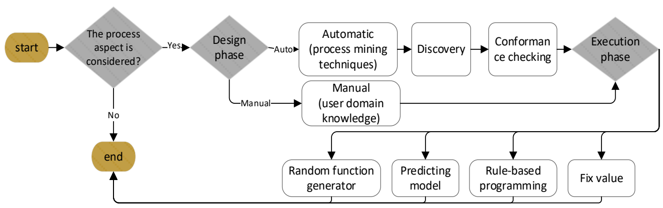

Figure 5: The designed flowchart for creating each aspect of a simulation model of a process, is based on two required phases: design and execution

The possibility of the aspects w.r.t. the two phases of designing simulation models and executing the designed model is presented in Figure 5. Throughout the design phase, the flow chart begins with one specific feature of a process and provides choices for discovering simulation parameter values automatically based on event logs and process mining techniques or setting these manually. It should be noted that the user can also use the provided information from the process mining techniques to design the simulation models and parameters manually. In such cases, we consider the design to be hybrid. For instance, performance spectrum techniques in process mining show the queue pattern of an activity visually, and the user can inject that knowledge into the simulation model. The execution phase follows, which entails reproducing the values of the aspects in the simulation. Fix values, random generator functions based on distribution, programming rules, and predictive models such as machine learning methods, e.g., Camargo et al. 2022, deep neural networks are used for generating the execution time, are all the possible methods for regenerating the simulation parameters.

Reference Meta-model

We extend the provided insights in van der Aalst 2015 to the required stimulation parameters with different possible execution models to design the reference model. In Figure 6, the high-level meta-model of a process simulation model in process mining is shown. The simulation meta-model is defined based on event logs and users' inputs.

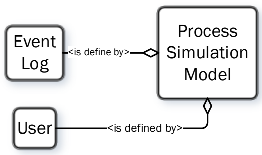

Figure 6: The high-level meta-model of processes simulation.

The designed reference meta-model is shown in Figure 7. To have an executable simulation model of processes, two main blocks are considered, the process model based on the event log, and the execution configuration. The execution configuration is the required parameter for running a simulation model, which is similar to most general simulation techniques. The start of the simulation, its duration, end time, or condition for ending the simulation. Moreover, for processes, the number of generated cases can be the determining factor for ending a simulation model.

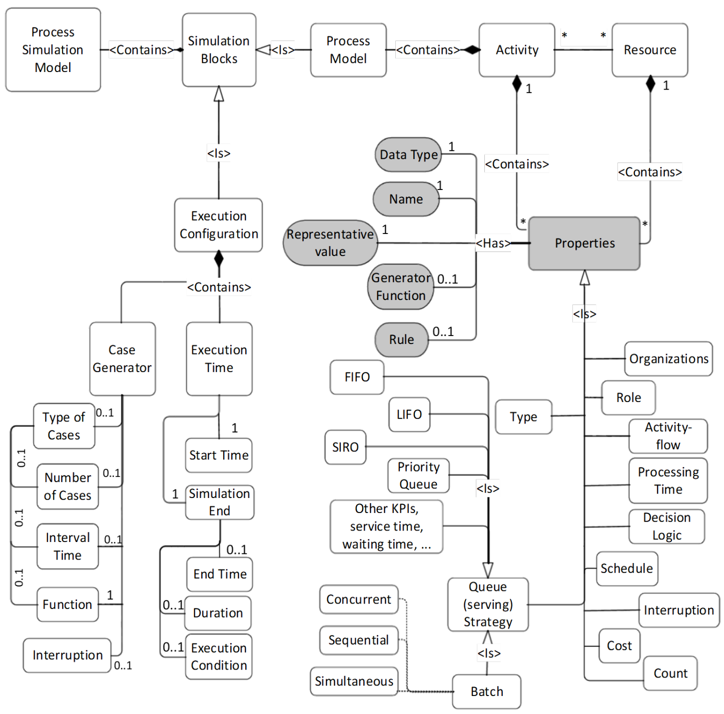

Figure 7: The Reference Meta-model of Simulation Models of Processes

In the designed reference model, the objects and their relations are defined as follows:

- Execution Configuration: (1) the Execution Time indicates the time that the simulation should start and end, (2) the simulation stop point is based on a specific timestamp, or when specific conditions are met, e.g., the number of generated cases or the duration of the simulation being 30 days.

- The Case Generator is a function (rule) generating new cases to be simulated. It can consider generating cases with attributes, the number of cases to be generated, the interval between cases, and interruption. For instance, using SML (Standard ML) with two types of customers (normal or VIP), and 1000 as the number of cases, the case generator can be defined as *if Discrete (1-10)>1, then case type =normal, else casetype= VIP*. Then a random generator function based on Poisson distribution with λ=10 (10 customers per hour) can generate 90% normal and 10% VIP customers.
- Process Model is the base of simulating processes that consist of activities. (1) Activity is the main element of a process model, and they have many-to-many relations to the Resource object. For instance, registration can be done by John and Sarah. Registration as an activity also has multiple properties, which makes its execution in the simulation possible. (2) Resource also has many-to-many relations to the Activity object. For instance, John can perform both Registration and Review tasks in the process.
- Properties define the information for Activity and Resource in the process and have the following attributes:
    - Name: the name of the properties such as activity name or organization name, e.g., the radiology department.
    - Data type: such as registration as an activity, and it has a processing time as a property where the data type of processing type is float.
    - Value: the processing time of activity registration as a property is its value, e.g., 30 minutes.
    - Function: for the processing time of activity registration as a property, with the value of 30 minutes, the function is a random generator based on a normal distribution with an average of 30.
- Properties are:
    - Organizations: every activity or resource can have a property of an organization indicating the organization they belong to, e.g., resource R belongs to the Audit department.
    - Role: in addition to organizations, every resource can be assigned to one or more roles as well, e.g., the role of Doctor or Nurse. Note that the roles might belong to different organizations.
    - Type: an activity can have a type mandatory or optional as property or resource, which can be a type of nurse.
    - Activity-flow: the flow of activities that can happen in the model, e.g., registration always happens before submitting a request.
    - Processing Time: is a property of a resource or an activity that shows how long it takes for an activity to be performed, or how long it takes for a resource to perform an activity.
    - Decision Logic: how to execute each part of the process, including the choices for specific cases.
    - Schedule: the working or active time that an activity can be executed, or a resource can execute activities, is a property of resources and activities.
    - Cost: the cost of each resource or the execution of each activity, can be presented as a property of them.
    - Count: the number of times an activity can be executed or the number of resources that exist, e.g., the number of welding machines as a resource.
    - Interruption: whether any specific state or condition, such as the end of business hours, can disrupt the resource task or activity execution.
    - Queue (serving strategy): including FIFO, LIFO, SIRO, Priority Queue, KPI-related, and Batching. Note that the mentioned condition in Martin et al. 2016 for queues, such as queue abandonment, is also covered in the introduced meta-model. For instance, a queue strategy like FIFO is a property of activity, and it can have an abandonment condition with more than 20 cases in the queue.

While designing the meta-model, the following design choices are made to make the model easier to interpret and to make it extensive enough to cover the required aspects. Under the Execution Configuration, for the Execution Time, as discussed, End Time, and Duration can be considered as the End Conditions. It is shown that they have association relations together as well. For the Case Generator, the same condition applies to the Function for generating the cases, which might include the Type of Cases or the Number of Cases. Moreover, the concept of business hours is considered in both the Case Generator by the Interruption and for Activity and Resource and their properties using Schedule.

PROCESS SIMULATION REFERENCE META-MODEL IN PRACTICE

We categorize the application areas of the designed meta-model into three main categories: (1) design and (2) verification of process simulation models, and (3) comparing different simulation models w.r.t. used insights from event logs. The introduced model serves as the foundation for creating simulation models of processes based on the insights gathered from event logs and process mining techniques. The use of the model in practice is demonstrated by (1) formally initializing the model, and (2) implementing an approach using based on the model.

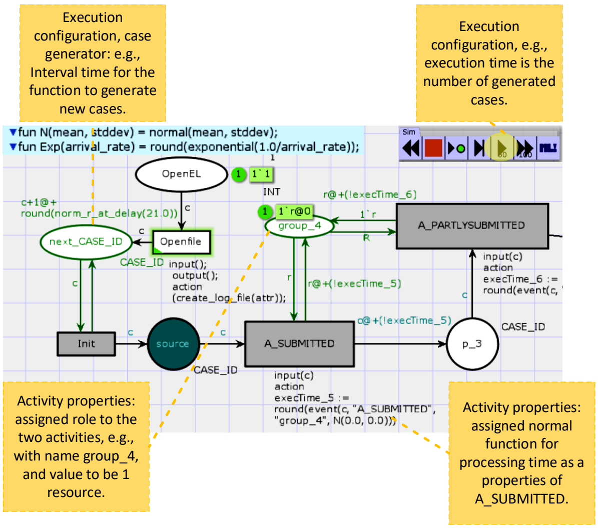

Figure 8: A Part of the generated simulation model of BPI Challenge 2012

We introduce one of the potential formal representations of the reference model based on the defined process mining concepts. Let M be the discovered process model from event log L and A be the set of activities for the process. Pr is the set of properties such that Pr *={Organizations, Role, Type, Activity-flow, Process Time, Decision Logic, Schedule, Interrup, cost, count}*. *Pin={Data Type, Name, Representative Value, Generator Function, Rule}* is the set of indicators of a property that illustrates what a property is in practice. Property function $PF: Pr \rightarrow set(Pin)$ returns the indicators of a property. An activity $a \in A$ has a set of properties, each of them having a set of indicators and a set of resources. $A \in 2^{Pr} \times 2^{R} \times PF$. For instance, $a \in A$= *({Processing Time}*, $\{r_1\}$, *PF*), where *PF(Processing Time)={float, duration, 10 minutes, Random Normal* (*μ=10 minutes, std=1 minutes)*}, and r_1 as the resource has a similar structure.
To demonstrate how the model can be used in practice, we use the model-based application to automatically generate the process simulation model in the form of Colored Petri Nets (CPN). We extract the possible automatic aspects from the event logs, including the main objects of the reference meta-model, and translate them to the CPN Tools model in the XML format, i.e., readable for the CPN Tools engine. Figure 8 is a part of the designed model for the BPI Challenge 2012 event log (van Dongen 2012) utilizing the developed tool. The code and data sets are also publicly available.[3] For instance, activity A_SUBMITTED has the properties of resource role group_4 or attribute Generator Function for that activity is based on a normal distribution.
The first main block is the process model *(M)* which the realization here is a Petri net *(M= (P, T, F))*, where P is the set of places, T is the set of transitions and $F \subseteq T \times T$. The realization of the activity block is a tuple *(R, Decision Logic, Activity-flow, Processing Time, Queuing Strategy)*, such that the resources, decision logic, flow of activities, processing time, and queuing strategy are considered. Note that except queuing strategy the rest of the properties are discovered automatically from event logs and yet the queuing strategy is considered as FIFO, and can be changed but not automatically discovered.

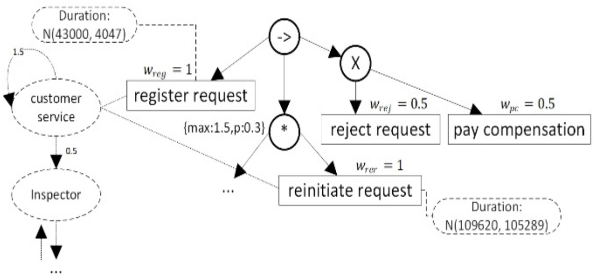

Figure 9: A sample of the process simulation model presented in Pourbafrani et al. 2021 considering the reference metamodel.

The other initialization of the meta-model for designing a simulation model of processes is presented in Pourbafrani et al. 2021. The activities have the properties of processing time and probability (decision logic). The resources also have the decision logic, as the handover matrix is also considered. A different realization of the meta-model for designing a simulation model of processes is shown in Figure 9. The code and instructions to execute the code including the implementation detail based on the reference model are publicly available[4]. In this realization of the reference meta-model for generating simulation models from event logs, the process model M is a process tree. The activity block in the design of simulation models from event logs is the tuple *(R, Decision Logic, Activity-flow, Processing Time, Count, Interruption, Queuing Strategy)*. For instance, property interruption has a type of Boolean in this implementation and has a rule that if it is set, the activity is interrupted when the working time

[3]https://github.com/mbafrani/AutomaticCPNModelGenerator

[4]https://github.com/mbafrani/SIMPT-SimulatingProcessTrees

is over, and if the activity continues outside the working time, i.e., *Pin(Interruption)= {Boolean, if 0 continue until the end, else stop until the working time starts}*.

The shown implementations, i.e., both data-driven simulation model generation processes based on event logs, illustrate the first use case of the reference meta-model. The extracted, used, and implemented blocks of the model are ready to be benchmarked and compared for the approaches required to be focused. However, the comparison of the simulation models and considered aspects is provided with the help of the referenced meta-model. This shows the requirement and support for tackling the limitation of the current techniques, i.e., the inconsistency of considered aspects for simulation, and makes the execution of simulation clear.

CONCLUSION and DISCUSSION

Process mining support can be extended to provide organizations with their digital twins, which enables faster and more confident decision-making. One of the practical realizations of digital twins, specifically in production processes, is executable simulation models. In this paper, we introduced a reference meta-model for generating simulation models based on an extensive literature review. We also exploited the potential of event logs, process mining techniques, and process simulation requirements. The comprehensive model includes all the possible extractable insights from processes as simulation parameters and all the possible directions to regenerate those insights. The reference meta-model supports designing, comparing, and evaluating simulation models of processes. This model provides the possibility of designing simulation models of production systems and checking their compliance with their historical data. However, the role of technical issues in designing and addressing all components of the process should not be overlooked. For example, there is an ongoing study into automatically detecting the type of queues inside event logs. These technical challenges, as well as the impact of human factors in generating realistic simulation models of processes, should be highlighted. Following the introduction of the comprehensive model as a reference model for process simulation, the model should be made directly executable. It should be designed in a generic framework in which, by using the generated XML format, the execution of or populating the simulation model for different tools will be possible, as illustrated by the developed tool for automatic CPN model generation.

REFERENCES

Camargo, M., Dumas, M. and Rojas, O.G., 2019. Simod: A Tool for Automated Discovery of Business Process Simulation Models. In *BPM (PhD/Demos)* (pp. 139-143).

Camargo, M., Dumas, M. and González-Rojas, O., 2020. Automated discovery of business process simulation models from event logs. *Decision Support Systems*, *134*, p.113284.

Camargo, M., Dumas, M. and González-Rojas, O., 2022, June. Learning accurate business process simulation models from event logs via automated process discovery and deep learning. In *Advanced Information Systems Engineering: 34th International Conference, CAiSE 2022, Leuven, Belgium, June 6–10, 2022, Proceedings* (pp. 55-71). Cham: Springer International Publishing.

Carmona, J., van Dongen, B., Solti, A. and Weidlich, M., 2018. Conformance checking. *Switzerland: Springer.*

de Leoni, M. and van der Aalst, W.M.P., 2013, March. Data-aware process mining: discovering decisions in processes using alignments. In *Proceedings of the 28th annual ACM symposium on applied computing* (pp. 1454-1461).

Denisov, V., Fahland, D. and van der Aalst, W.M., 2018. Unbiased, fine-grained description of processes performance from event data. In *Business Process Management: 16th International Conference, BPM 2018, Sydney, NSW, Australia, September 9–14, 2018, Proceedings 16* (pp. 139-157). Springer International Publishing.

Estrada-Torres, B., Camargo, M., Dumas, M., García-Bañuelos, L., Mahdy, I. and Yerokhin, M., 2021. Discovering business process simulation models in the presence of multitasking and availability constraints. *Data & Knowledge Engineering*, *134*, p.101897.

Fishman, G.S., 2001. *Discrete-event simulation: modeling, programming, and analysis* (Vol. 537). New York: Springer.

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

García, M.T., Barcelona, M.A., Ruiz, M., García-Borgoñón, L. and Ramos, I., 2014. A discrete-event simulation metamodel for obtaining simulation models from business process models. In *Information System Development: Improving Enterprise Communication* (pp. 307-317). Springer International Publishing.

Gawin, B. and Marcinkowski, B., 2015. How Close to Reality is the „as-is" Business Process Simulation Model?. *Organizacija*, *48*(3), pp.155-175.

Hornix, P.T., 2007. Performance analysis of business processes through process mining. *Master's Thesis, Eindhoven University of Technology*.

Jansen-Vullers, M. and Netjes, M., 2006, October. Business process simulation–a tool survey. In *Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools, Aarhus, Denmark* (Vol. 38).

Khodyrev, I. and Popova, S., 2014. Discrete modeling and simulation of business processes using event logs. *Procedia Computer Science*, *29*, pp.322-331.

Macal, C.M. and North, M.J., 2009, December. Agent-based modeling and simulation. In *Proceedings of the 2009 winter simulation conference (WSC)* (pp. 86-98). IEEE.

Martin, N., Benoît, B. and Caris, A., 2014, August. Event log knowledge as a complementary simulation model construction input. In *2014 4th International Conference On Simulation And Modeling Methodologies, Technologies And Applications (SIMULTECH)* (pp. 456-462). IEEE.

Martin, N., Depaire, B. and Caris, A., 2016. The use of process mining in business process simulation model construction: structuring the field. *Business & Information Systems Engineering*, *58*, pp.73-87.

Măruşter, L. and Van Beest, N.R., 2009. Redesigning business processes: a methodology based on simulation and process mining techniques. *Knowledge and Information Systems*, *21*, pp.267-297.

Pospíšil, M., Mates, V., Hruška, T. and Bartík, V., 2013. Process mining in a manufacturing company for predictions and planning. *International Journal on Advances in Software*, *6*(3).
Pourbafrani, M., Jiao, S. and van der Aalst, W.M., 2021, May. SIMPT: process improvement using interactive simulation of time-aware process trees. In *Research Challenges in Information Science: 15th International Conference, RCIS 2021, Limassol, Cyprus, May 11–14, 2021, Proceedings* (pp. 588-594). Cham: Springer International Publishing.
M. Pourbafrani and W. M. P. van Der Aalst, "Discovering System Dynamics Simulation Models Using Process Mining," in *IEEE Access*, vol. 10, pp. 78527-78547, 2022, doi: 10.1109/ACCESS.2022.3193507.
Ratzer, A.V., Wells, L., Lassen, H.M., Laursen, M., Qvortrup, J.F., Stissing, M.S., Westergaard, M., Christensen, S. and Jensen, K., 2003, May. CPN tools for editing, simulating, and analysing coloured Petri nets. In *Applications and Theory of Petri Nets 2003: 24th International Conference, ICATPN 2003 Eindhoven, The Netherlands, June 23–27, 2003 Proceedings* (pp. 450-462). Berlin, Heidelberg: Springer berlin Heidelberg.
Rozinat, A., Mans, R.S., Song, M. and Van der Aalst, W.M., 2008. Discovering colored Petri nets from event logs. *International Journal on Software Tools for Technology Transfer*, *10*, pp.57-74.
Rozinat, A., Mans, R.S., Song, M. and van der Aalst, W.M., 2009. Discovering simulation models. *Information systems*, *34*(3), pp.305-327.
Rozinat, A., Wynn, M., van der Aalst, W., ter Hofstede, A. and Fidge, C., 2008. Workflow simulation for operational decision support using design, historic and state information. In *Business Process Management: 6th International Conference, BPM 2008, Milan, Italy, September 2-4, 2008. Proceedings 6* (pp. 196-211). Springer Berlin Heidelberg.
Senderovich, A., Leemans, S.J., Harel, S., Gal, A., Mandelbaum, A. and van der Aalst, W.M., 2016. Discovering queues from event logs with varying levels of information. In *Business Process Management Workshops: BPM 2015, 13th International Workshops, Innsbruck, Austria, August 31–September 3, 2015, Revised Papers 14* (pp. 154-166). Springer International Publishing.
Song, M. and van der Aalst, W.M.P., 2007, December. Supporting process mining by showing events at a glance. In *Proceedings of the 17th Annual Workshop on Information Technologies and Systems (WITS)* (pp. 139-145).
Song, M. and van der Aalst, W.M.P., 2008. Towards comprehensive support for organizational mining. *Decision support systems*, *46*(1), pp.300-317.
Tumay, K., 1996, November. Business process simulation. In *Proceedings of the 28th conference on Winter simulation* (pp. 93-98).
van Beest, N.R.T.P. and Maruster, L., 2007, September. A Process Mining Approach to Redesign Business Processes-a case study in gas industry. In *Ninth international symposium on symbolic and numeric algorithms for scientific computing (SYNASC 2007)* (pp. 541-548). IEEE.
van der Aalst, W.M., 2010. Business process simulation revisited. In *Enterprise and Organizational Modeling and Simulation: 6th International Workshop, EOMAS 2010, held at CAiSE 2010, Hammamet, Tunisia, June 7-8, 2010. Selected Papers 6* (pp. 1-14). Springer Berlin Heidelberg.
van der Aalst, W.M.P., 2016. Process Mining Data Science in Action Second Edition.
van der Aalst, W.M.P., 2018, July. Process mining and simulation: A match made in heaven!. In *SummerSim* (pp. 4-1).
van Dongen, B.F., de Medeiros, A.K.A., Verbeek, H.M.W., Weijters, A.J.M.M. and van der Aalst, W.M.P., 2005. The ProM framework: A new era in process mining tool support. In *Applications and Theory of Petri Nets 2005: 26th International Conference, ICATPN 2005, Miami, USA, June 20-25, 2005. Proceedings 26* (pp. 444-454). Springer Berlin Heidelberg.
Wynn, M.T., Dumas, M., Fidge, C.J., Ter Hofstede, A.H. and van Der Aalst, W.M., 2008. Business process simulation for operational decision support. In *Business Process Management Workshops: BPM 2007 International Workshops, BPI, BPD, CBP, ProHealth, RefMod, semantics4ws, Brisbane, Australia, September 24, 2007, Revised Selected Papers 5* (pp. 66-77). Springer Berlin Heidelberg.

ACKNOWLEDGEMENT

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC-2023 Internet of Production – 390621612. We also thank the Alexander von Humboldt (AvH) Stiftung for supporting our research.

AUTHOR BIOGRAPHIES

MAHSA POURBAFRANI received her B.S. and M.S. degrees in computer science from the AmirKabir University of Technology, Tehran, Iran, in 2014. Currently, she is a research assistant in the Data and Process Science Group of RWTH Aachen University. She is a Ph.D. student under Professor van der Aalst's supervision. Her research focuses on forward-looking process mining, which employs data science methods to turn data into actionable insights. The actions are taken using simulation, what-if analysis, and predictions in process mining regarding the performance metrics of processes. She is also a scientist working on the "Internet of Production" project, which aims to combine process mining and machine learning techniques to support operations and decisions in production lines. Her e-mail address is: mahsa.bafrani@pads.rwth-aachen.de.

Prof.dr.ir. WIL VAN DER AALST is a full professor at RWTH Aachen University, leading the Process and Data Science (PADS) group. He is also the Chief Scientist at Celonis, part-time affiliated with the Fraunhofer FIT, and a member of the Board of Governors of Tilburg University. He also has unpaid professorship positions at Queensland University of Technology (since 2003) and the Technische Universiteit Eindhoven (TU/e). Currently, he is also a distinguished fellow of Fondazione Bruno Kessler (FBK) in Trento, deputy CEO of the Internet of Production (IoP) Cluster of Excellence, and co-director of the RWTH Center for Artificial Intelligence. His e-mail address is : wvdaalst@pads.rwth-aachen.de.