Fachhochschule Nordwestschweiz

Process Mining and Simulation

University of Applied Sciences and Arts FHNW

School of Engineering

June 2019

Institute for Business Engineering
University of Applied Sciences and Arts
Northwestern Switzerland
Bahnhofstrasse 6
CH-5210 Windisch
https://www.fhnw.ch/en/

Title:

Process Mining and Simulation

Theme:

Global Implementation

Project Period:

Spring Semester 2019

Student:

Rasmus Frost Hvarregaard

Supervisor:

Raoul Waldburger

Copies: 1
Pages: 42
Date of Completion: June 28, 2019

Abstract:

The purpose of this research project is to investigate the options and possibilities of combining the two powerful tools of process mining and business process simulation. On one hand, process mining is a tool that combines data science and process modelling to create accurate data driven process models to analyse the as-is situation of any given system. On the other hand, process simulation is a tool that uses human-created process models to analyse different to-be scenarios. Hence, a combination of these tools seems to provide great potential value for companies seeking to improve their business processes. The project reveals that the current state-of-the-art in the research field is far from having such an integrated tool. The second part of this project has therefore been to contribute to the further development. The result is a solution proposal for using process mining tools to calculate process capacity, which can be used as input parameter for process simulation models. Finally, the project value in a business context is considered. It is anticipated that the solution could have both direct and in-direct value for many companies. Direct value in terms of calculating accurate capacity measures, which can improve the ability to balance supply and demand. In-direct value in terms of contributing to the further development of an integrated tool that can create accurate simulations models based on data.

Contents

Preface

This project report is created and compiled by Rasmus Frost Hvarregaard as a part of the exchange semester undertaken at Fachhochschule Nordwestschweitz, School of Engineering, in the period from 18^{th} of Febrary to 2^{nd} of July 2019. The exchange semester substitutes the 2^{nd} semester of the *Operations And Innovation Management* master programme at Aalborg University. The project is made in collaboration with Prof. Dr. Raoul Waldburger, who has been the advisor throughout the project. The theme of the project is *Process Mining and Simulation* where the combination of these two powerful tools has been investigated.

It has been a privilege to cooperate with Prof. Dr. Raoul Waldbuger whom provided support through competent guidance. He has been helpful in finding an approach to conduct the project. Finally, his knowledge and constructive criticism has been an important part of the learning experience.

Readers guidance

Throughout this project, the Harvard citing method is used. The references are compiled and arranged at the end of the report as a complete reference list. The references in the report will lead the reader to the entire reference list at the end of the report, where books are submitted by author, title, edition, and publisher. The web pages will be submitted with author, title, and URL. Figures, equations, and tables will be numbered according to the respective chapter.

Fachhochschule Nordwestschweiz FHNW, June 20, 2019

Rasmus Frost Hvarregaard
<rasmus.hvarregaard@students.fhnw.ch>

Introduction 1

This project has been conducted with the purpose of exploring the field of process mining and business process simulation. More specifically, the project is concerned with the combination of these two powerful tools, which is today still a problem that has not been resolved [van der Aalst, 2010; Martin et al., 2016; Norambuena, 2018]. A literature review of relevant articles in the field shows that only a few researchers has addressed the issue on how to integrate process mining and business process simulation in a meaningful way that can ultimately be used in practise for solving real life business cases.

The drivers behind such a combination can be identified by considering the complementary differences between both disciplines. The most significant difference seems the areas of application. On one hand, process mining is used for the discovery and analysis of existing processes i.e. the as-is situation. This is done by creating process models from raw event data extracted from the system under analysis, which in theory should enable an accurate representation of the real-world system. On the other hand, simulation is used to answer "what-if" questions by creating a simulation model based on inputs such as observations, interviews and statistics. The flaw here is that these models can be prone to human interpretation rather than being an accurate representation of the real-world system. Thus, it seems that there are flaws connected to both approaches, but that these flaws can be mitigated in a mutual way if these tools were integrated with each other. I.e. by means of a tool that would be able to create valid process and system models, based on event data, that could afterwards be used for scenario simulation for answering "what-if" questions and assist in predicting outcomes of future situations [IEEE, 2011]. Building on the potential for creating a valuable tool that can be used in a variety of applications, this project addresses the specific problem of how to integrate process mining and simulation.

1.1 Definitions and Key Constructs

Data Mining: Sifting through very large amounts of data for useful information. Data mining uses artificial intelligence techniques, neural networks, and advanced statistical tools (such as cluster analysis) to reveal trends, patterns, and relationships, which might otherwise have remained undetected. In contrast to an expert system (which draws inferences from the given data on the basis of a given set of rules) data mining attempts to discover hidden rules underlying the data. Also called data surfing [Business Dictionary, 2019b].

Process Mining: A set of techniques that allow for extracting information from raw data sets in form of event logs. For example, the audit trails of a work-flow management system or the transaction logs of an enterprise resource planning system can be used to discover models describing processes, organisations, and products [Process Mining Group, Math&CS department, Eindhoven University of Technology, 2019].

Business Process Simulation: Business process simulation is an instrument for the analysis of business processes. It is used to assess the dynamic behaviour of processes over time, i.e. the development of process and resource performance in reaction to changes or fluctuations of certain environment or system parameters. The results provide insights supporting decisions in process design or resource provision with the goal to improve factors such as process performance, process and product quality, customer satisfaction or resource utilisation [ARIS Group by Software AG, 2019].

Queuing Theory: Mathematical modelling of waiting lines, whether of people, signals, or things. It aims to estimate if the available resources will suffice in meeting the anticipated demand over a given period [Business Dictionary, 2019c].

Process Mining and Simulation 2

This chapter outlines the content of the two disciplines: process mining and business process simulation, which will be the central elements throughput this project paper. The chapter ends with a comparison on the two disciplines.

2.1 What is Process Mining?

Process mining is a branch of, and builds on principles, from the larger data science field called *data mining*. Process mining is said to be a combination of Business Process Modelling and Data Mining, which first and foremost can be used to create data driven process models. I.e. event logs extracted from information systems in companies are used to generate visual process models. These data driven process models can be used for a variety of purposes. In the Process Mining Manifesto, three main applications are identified and explained: Model discovery; conformance checking and; process enhancement. This has been illustrated in figure 2.1. The first application type, model discovery is the discipline of extracting and creating valid process models of a given system e.g. a ATM cash withdrawal, hospital patients flow or a production process, based on a historical event log generated through actions made in the systems. This is useful and distinct from traditional process modelling as it will reflect the actual system behaviour, given that the data source i.e. event log is valid, and not merely the interpretation of the process modeller. The second application type is conformance checking, which is the discipline of validating actual system behaviour against intended system behaviour. Thus, identifying unintended deviations in processes of a given system. This can for instance be used in the sense of quality conformance checking on procedures, internal or external audits in terms of accounting or ISO certification. The third and final application type is process enhancement. Obviously this concerns the improvement of existing processes, based on analysing derived models, which was the first application type. In this sense process mining gives the opportunity to identify bottlenecks, order flow, process times etc. Based on these insights, the analyser can prepare suggestions for improvements.

2.1.1 Process Mining Background

Analysing processes is the foundation for much of the work conducted in the corporate world today. This includes work with both performance enhancing purposes in mind, but also so called "process compliance checking" work. To mention a few examples of

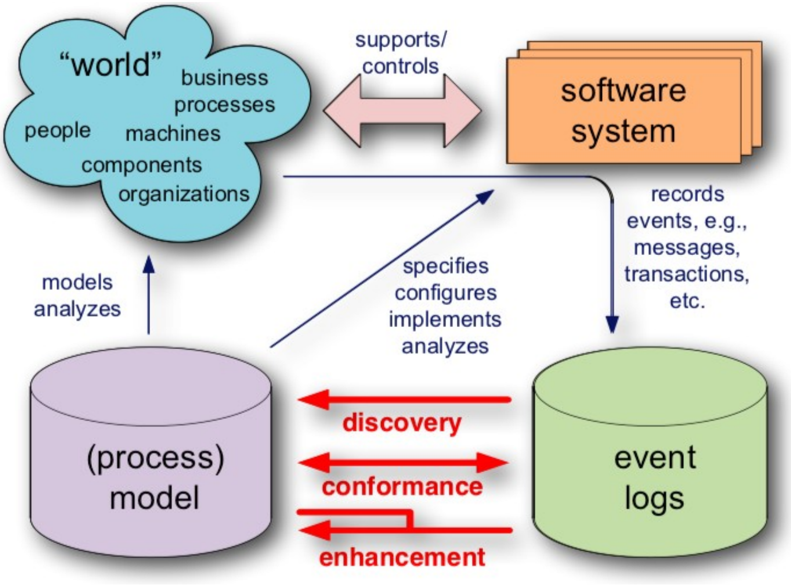

Figure 2.1: *Positioning of the three main types of process mining: (a) discovery, (b) conformance checking, and (c) enhancement. IEEE [2011]*

famous frameworks/models that are using the analysis of processes includes Total Quality Management, Six Sigma, Business Intelligence, Continuous Improvement and Business Process Management and finally compliance issues such as accounting, ISO certification or other legislative purposes. According to IEEE [2011] this is one of two key drivers behind the development and rising interest in the field of process mining. The second major driver is the continuously growing amount of available data that is being stored and provided in a variety of information systems such as enterprise resource planning systems.

As a consequence of the rising interest in the field and its applications, a range of software providers has today included process mining modules or functionality into their products. Further, in 2009 a task force were established by The Institute of Electrical and Electronics Engineers, Inc. (IEEE) with the purpose of:

- make end-users, developers, consultants, and researchers aware of the state-of-the-art in process mining,
- promote the use of process mining techniques and tools and stimulating new applications,
- play a role in standardisation efforts for logging event data,
- the organisation of tutorials, special sessions, workshops, panels,
- the organisation of Conferences/Workshop with IEEE CIS Technical Co-Sponsorship, and
- publications in the form of special issues in journals, books, articles (e.g., in the IEEE Computational Intelligence Magazine).

2.1.2 Challenges in the field of Process Mining

IEEE [2011] was also the first to publish a document that was set out to establish Process Mining as a concept and research field, which was called the Process Mining Manifesto. In this document, the basics of process mining is explained and in particular the challenges in relation to process mining is listed and summarised to encourage researchers to explore these uncovered areas. These are:

- C1: Finding, Merging, and Cleaning Event Data
- C2: Dealing with Complex Event Logs Having Diverse Characteristics
- C3: Creating Representative Benchmarks
- C4: Dealing with Concept Drift
- C5: Improving the Representational Bias Used for Process Discovery
- C6: Balancing Between Quality Criteria such as Fitness, Simplicity, Precision, and Generalization
- C7: Cross-Organizational Mining
- C8: Providing Operational Support
- C9: Combining Process Mining With Other Types of Analysis
- C10: Improving Usability for Non- Experts
- C11: Improving Understandability for Non-Experts

These challenges are all still relevant today. Even though the topic of process mining has gained more and more interest, which is clearly expressed by figure 2.2, an article by R'bigui og Cho [2017] reveal that not much progress has been done in order to solve any of the issues or challenges pointed out by IEEE [2011].

In relation to the topic of this project work, which is to combine process mining with simulation, the challenge of "Combining Process Mining With Other Types of Analysis", has also not received much attention, and is therefore still an open issue that should be investigated further [Norambuena, 2018]. In the next chapter, a state-of-the-art literature review will be presented in order to show how far the development has come within this particular problem.

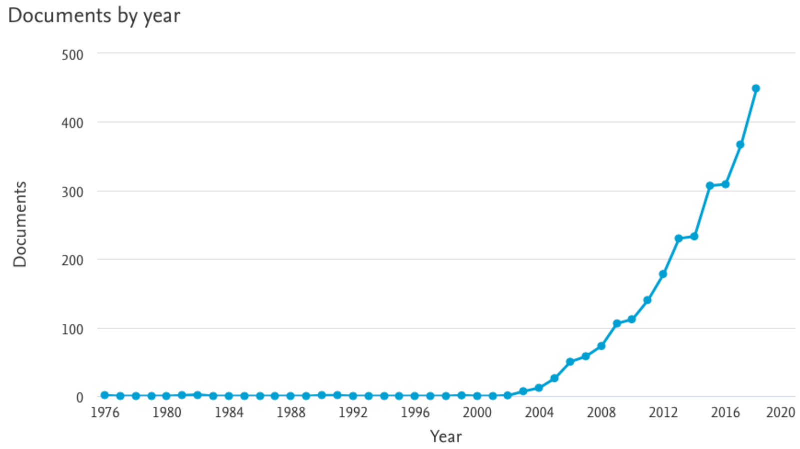

Figure 2.2: *Graph displaying the number of publications, which includes the term "process Mining", at the database Scopus, from 1976 until 2018*

2.2 What is Business Process Simulation?

Simulation provides the opportunity to create various "what-if" scenarios of existing systems based on a model that reflects the characteristics of the system under analysis. This gives the opportunity to test and compare different design alternatives to existing systems before they are carried out in real life. Thus, simulation is a tool that are valuable for decision makers in various fields as it can potentially prevent making undesirable choices. In the context of business process management, simulation can therefore be used to test different process structure design alternatives in case of re-engineering tasks. This can save both time and money for companies as it is faster to create a simulation model than it is to carry out process changes in real life. Secondly, it is cheaper to make faults in the simulation model (solutions with unintended outcomes), than it is to experience the faults only when you have changed the real world system. According to der Aalst [2015], simulation has become one of the standard analysis techniques in the field of operations management. Today, many different software tools for business process simulation exists e.g. Arena (Rockwell), Simio, Enterprise Dynamics (Incontrol) and Vensim (Ventana Systems). These tools does all have a intuitive user interface where the modeller can drag and drop modules into the model and hereby relatively little training is required before one can make use of the tools. Finally to sum up on simulation as a tool for evaluating business processes, a list of advantages and disadvantages, has been displayed in table 2.1.

Advantages
Simulation is flexible. Any situation, no matter how complex, can be investigated through simulation.
Simulation can be used to answer a wide range of questions. It is possible to assess waiting times, utilisation rates and fault percentages using one and the same model.
Simulation stimulates creativity. Simulation triggers “process thinking” without restricting the solution space upfront.
Simulation is easy to understand. In essence, it is nothing but replaying a modelled situation. In contrast to many analytical models, little specialist knowledge is necessary to understand the analysis technique used. Hence, simulation can be used to communicate ideas effectively.
Disadvantages
A simulation study can be time consuming. Sometimes, very long simulation runs are necessary to obtain reliable results.
One has to be very careful when interpreting simulation results. Determining the reliability of results can be very treacherous indeed.
Simulation does not provide any proof. Things that can happen in reality may not be witnessed during some simulation experiment.

Table 2.1: Advantages and disadvantages of simulation [der Aalst, 2015]

2.3 Comparing Process Mining and Simulation

Seen in the light of the previous two sections that has elaborated on process mining and business process simulation, this section seeks to compare the two concepts in terms of identifying the primary differences. First and foremost, there is a great difference in the sense of which questions the two tools can answer for the analyser. On one hand, process mining is useful in terms of analysing the *as-is* system, because the models are build on historical data. On the other hand, simulation is useful to evaluate different *to-be* scenarios i.e. future situations, because the models are build on statistical inputs about the system such as arrival rates, process times etc. That leads naturally to the next difference, which is the model construction it self. Process mining models are build on actual events derived from information systems i.e. they should in theory be highly accurate models that are a valid representation of the real system. Simulation models are traditionally build by "experts" who collaborate with people who are working in the system to be simulated. Thus, the simulation models are basically only the result of personal interpretations of the modeller and the responses from the people around him. The differences has been illustrated in figure 2.3.

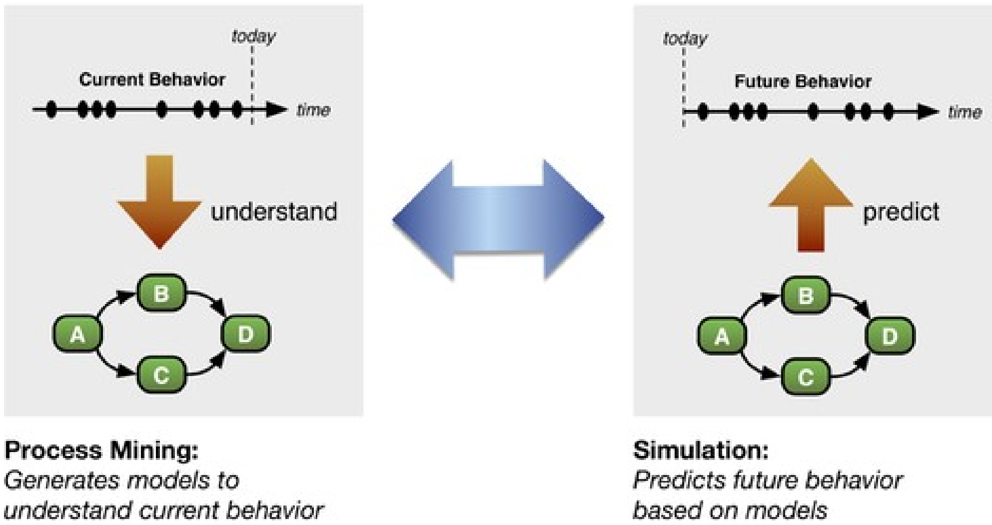

Figure 2.3: *Difference between simulation and process mining [Anne Rozinat, 2018]*

Preliminary Research Question 3

As a first step in accessing the problem of integrating process mining and business process simulation it has been decided to create an overview of the current state of research and investigate the future outlook. Thus answering the preliminary research question:

> ***What is the state-of-the-art in the research field of integrating process mining and business process simulation?***

This question will be answered through a state-of-the-art literature review in chapter 4. Findings from this review will be evaluated and synthesised into an overview of the best practice, which will be the foundation for the refined research question.

State-of-the-Art 4

This chapter contains a state-of-the-art literature study on the specific challenge of integrating/combining process mining and simulation, which is the overall topic of this project work. Thus, identifying relevant literature within this field. As mentioned in chapter 2 the challenge has initially been addressed by the Process Mining Task Force of IEEE as a central issue that should be researched further. The literature study should conclude the current state of research in the field.

The methodological approach has been to first identifying key words for searching relevant articles, which has been important in order to narrow down the investigation and focus on the issue of process mining and simulation. The selection of these key words has been an exploratory process where different opportunities has been investigated. The focus has been to narrow down the search scope to become as concise as possible, ensuring a focused effort on finding the relevant articles.

Key words	**Boolean Operator**	**Key words**	**Database**	**Articles**
Simulation	AND	"process mining"	Scopus	191
Simulation	AND	"data mining"	Scopus	7.900
Refined to:				
"Simulation modelling"	AND	"data mining"	Scopus	59
Simulation	AND	"Machine Learning"	Scopus	+10.000
Refined to:				
"Simulation modelling"	AND	"Machine Learning"	Scopus	39
Additionally, two relevant books were identified during the literature search.				

Table 4.1: Literature study search result

The search string *simulation AND "process mining"* was used as outset for the literature study. From this search, 191 articles were identified. The next step in the process has been to read through abstracts of these 191 articles to discover new key words for enriching the literature search and to sort out the relevant articles, which were addressing the topic under investigation, namely integrating/combining process mining and simulation. This selection process left 47 articles to be analysed deeper. One part of the deeper analysis has been to identify sources/sourced by works of the articles to enrich the search. The analysis revealed that some research has been done in the field of using machine learning to help the construction process of simulation models, which gave occasion to add an additional search string. Finally, because process mining builds on the field of data mining it was decided to add a search string of this topic as well in order not to oversee any relevant research. The results of the literature search has been summarised in table 4.1. The following sections has been divided into three categories where articles deals with: Integration-Approach Development; Literature Reviews and; Case Studies. The findings has also been summarised in table 4.2.

4.1 Integration-Approach Development

In this section, articles dealing with the problem of developing approaches for integrating process mining and simulation will be presented. Some of the first words written about the idea of combining process mining and simulation where done in a research study by Wynn et al. [2007] who outlines the the idea of combining process mining and simulation tools and illustrate an approach through the use of ProM 6 and CPNtools. Same year, Rozinat et al. [2008] comes with another study where the combination of process mining and simulation has focus on short-term simulation for operational decision making support, which has a proof-of-concept nature. Two years later, Rozinat et al. [2009a] writes a paper which is one of the most cited papers on the topic of combining process mining and simulation. In this research work the authors also use the approach for automatically creating simulation models on the basis of process mining and event log data by using the tools ProM 6 and CPNtools to implement the approach. Their work is highly detailed and descriptive, which has made it helpful for much further research in the field. However, the authors recognise that the approach is an early stage of development towards a full integration of the two disciplines and that more research is required. Rozinat et al. [2009b] is, together with Rozinat et al. [2009a], some of the most important works done on the topic of combining process mining and simulation. Their work has laid much of the foundation for all further research in the field. In their work, which is a continuation of Rozinat et al. [2008], they present an approach which make use of three different systems: 1) Yet Another Workflow Language (YAWL), 2) ProM 6, 3) CPNTools (CPN = Coloured Petri Net). This approach is specifically developed for operational decision support. It is argued that traditional simulation projects are mainly conducted for strategic planning of steady state systems. However, according to the authors the steady state does not exist because e.g. production systems are dynamic and continuously changing. They suggest that their approach can be used to assist managers in making day-to-day decisions by having a so called "fast-forward" function, enabling the manager to look into the near future and test different scenarios. A significant element of the approach is that it makes use of the combination of historic and current-state data, enabling a valid and true representation of the system. Finally the authors suggest that more sophisticated algorithms are needed, which can extract information about resource behaviour. Further, that the liability of the approach is that it makes use of three different systems. They envision a seamless integrated tool that can handle everything automatically. Maruster og van Beest [2009] builds on the approach inspired by Rozinat et al. [2009a] where ProM 6 and CPNtools are used. However, their approach is refined for less structured processes and validated through three case studies. Akhavian og Behzadan [2013] develop an approach for extracting knowledge from real-time sensor data in a construction site context. They use a simulation software tool called "Stroboscope" for the simulation execution. The concept of the approach is at a high level similar to the concept presented in Wynn et al. [2007], except that the approach is specifically developed for a construction site. The objective of the simulation is short-term decision support. Khodyrev1 og Popova [2014] builds on the approach developed by Rozinat et al. [2009a], but expand the approach to also include macro environmental factors that can influence the behaviour in the simulation model. The conclusion is that the extended model provides more accurate results than the original approach by Rozinat et al. [2009a]. Senderovich [2015] provides a research study that has a more narrow focus that previous studies. He investigates the opportunity for deriving queuing models from event log data, which can potentially be used in the construction of simulation models. The specific purpose is online delay prediction. Pileggi et al. [2015] apply process mining techniques to mine process models from sensor data. In the conclusion of their work they recognise the similarities between process models that can be mined from process mining and traditional simulation models. Therefore they envision an automated approach for mining simulation models directly from event log data. They identify the need for further research in regards to the event labelling and attributes to be used for developing simulation models from event log data. Martin et al. [2015] conduct a sensitivity analysis on the APRRA Framework (Arrival Rate Parameter Retrieval Algorithm). This algorithm can be used to model/mine one of the input parameters to a simulation model by using event logs. Their conclusion is that the algorithm is robust enough to determine arrival rates based on event log data. However, they suggest at further refinement in future research. In his master thesis, Liu [2015] addresses the problem of integrating process mining and discrete event simulation, which he mainly suggest is due to the lack of compatibility between process mining tools and simulation tools. For this reason, he develops a new plug-in module for the process mining software ProM 6. The plug-in is called *ProModel Export* and has the functionality of converting a so called "integrated model" developed in ProM 6 upon event data, in form of a colored petri net model, into an Excel .xls spreadsheet, which can be read by a broader range of simulation engines. His work builds on the work by Rozinat et al. [2008], who created a similar plug-in, which used an output format called .cpn that could only be read by the software *CPNtools*. Wang et al. [2015] develop an approach that is similar to Rozinat et al. [2009a]. However, a central point in their approach is the use of a method called "data to fuzzy-DEVS (D2DF) which consist of roughly three steps: (1) going from event data to event logs; (2) from event logs to transition system; (3) from transition system to Fuzzy-DEVS model. The first step is about converting the raw data to event logs, using the data structure "System Entity Structure", which allow for an expansion or enrichment of traditional event data, so that it contains attributes of additional information. This of course requires that these additional information is collected in the data logging or collection process. The second step of the approach is to create a transition model from the event logs, which is the first step of traditional process mining techniques e.g. ProM 6. Third step is to convert the transition model into a fuzzy-DEVS model, which is done in ProM 6, by using an extension module called "Convert to fuzzy-DEVS using Regions". Subsequently, the fuzzy-DEVS model is being simulated in the simulation engine SimStudio (AutoDesk). In conclusion, the article contributes with an improvement on the challenge of creating a simulation model directly from raw event data logs. However, the authors points out that further work is still required in relation to the D2DF method and in relation to validating the model. Leyer og Hüttel [2017] propose a methodology in which they combine Process Mining, Data Envelopment Analysis and Business Process Simulation. They evaluate the method by conducting a case study and concludes that better results can be yielded compared to a traditional approach. However, the method is manual and specifically developed for a context of a livestock process, which invoke some limitations on the application to other contexts. Martin [2017] is highly specific in addressing the challenge of integrating process mining and simulation methods, which he argue is important due to the liabilities of simulation model construction techniques. He first develops a conceptual overview of the use of event logs for simulation model construction by outlining the steps in the construction process. Secondly, he builds on this overview of process steps to identify modelling tasks that can be assisted by event log data. For this purpose he develops four algorithms: 1) Arrival Rate Parameter Retrieval Algorithm (ARPRA), which takes into account the queuing aspect; 2) Batch Organization of Work Identification algorithm (BOWI), which is based on a distinction between simultaneous, sequential and concurrent batching. It is the first algorithm that systematically identifies batches in an event log and calculates a set of batch processing metrics; 3) Batch Activation Rule Identification algorithm (BARI), where a batch activation rule captures the circumstances under which a resource starts processing a batch; 4) Resource Schedule Identification Method (RSIM), which generate an overview of resource availability by taking into account (i) the temporal dimension of availability, i.e. the time of day at which a resource is available, and (ii) intermediate availability interruptions. Abohamad et al. [2017] merely suggest that process mining techniques can be used to assist the construction of simulation models. However, no automatically or integrated approach is presented. The study merely uses the results from a process mining study to manually build a simulation model. Kovalchuka et al. [2018] develop a framework for the construction of simulation models in a health care system for which they use several methods of data and process mining techniques. The framework is extensive and theoretical, and does not include any automated simulation model generation. The purpose is solely to improve simulation model validity. Savickasa og Vasilecas [2018] has shown an approach where event logs are used to create Bayesian or Belief Network structures, which is a probabilistic graphical model that can show how events are conditionally related to each other. The study is highly theoretical, but the authors suggest that more research could yield an approach of how to automatically create simulation models from the Belief Networks. Pegoraro et al. [2018] comes up with a theoretical framework for semi-automatically generation of simulation models. The reason they call it semi-automatic is because the framework combines the use of event data and expert-interviews in order to create the final simulation model. Another important note about the approach is that it focuses on creating a simulation model for short term decision making, similar to Rozinat et al. [2009b]. Thus, the simulation model does not contain a sufficient high level of information to use it for strategic decision making. The research group state that their work will continue with applying the framework in practise. Elbattah og Molloy [2018] builds a research project on the fundamental idea that new methods are required in the field of modelling and simulation due to the increasing complexity of systems. For this he suggest that machine learning could be valuable. It is argued that new trends has started to emerge, where the field of hybrid simulation models are explored. Such models are per definition combining different disciplines e.g. simulation and data science, which can be complementing each other. The authors propose a framework to aspire further research which set out three key ideas: 1) Learning to Predict the System Behaviour, 2) Identify Predictable Influential Variables, 3) Incremental Learning = Adaptive Behaviour.

4.2 Literature Reviews

This section contains a presentation of the literature reviews in the field of process mining and simulation. The first follow up on the current state in the research field, van der Aalst [2010] conducted a survey on the latest approaches developed in the field of combining simulation and process mining techniques. The conclusion was that there were many challenges yet to be solved. Martin et al. [2016] made a survey on relevant process mining literature and state-of-the-art literature on integration of process mining and simulation. The conclusion is that not much research has been done and very little knowledge is available to overcome this issue. The main part of this research work is to outline the key challenges in relation to integrate process mining and simulation. They claim that their research is a starting point to structurally integrate process mining in simulation model construction. The latest literature study seems to be conducted by Norambuena [2018] in which it is revealed that the integration of process mining and simulation is still an open problem. The primary contribution is a list of key challenges in process mining, which is divided into two categories: organisational and data-centric perspective and; methodological perspective. The article concludes that one of the main challenges is the availability of suitable data for creating coherent process models, which can afterwards be used for simulation and that further research is required in order to integrate process mining and simulation in a meaningful way that can be applied in the industry.

4.3 Case Studies

In Liu et al. [2012] the approach developed by Rozinat et al. [2009a] is used to conduct a case study on work flow in a company. Aguirre et al. [2013] perform a study on business process re-design by building on the work by Maruster og van Beest [2009] and Rozinat et al. [2009a]. The expansion is set out to give more focus to the project understanding phase, in form of process scope analysis, process redesign goal setting and performance gap analysis. They suggest that further research should be done to validate the approach in other contexts and further on the necessary event log extraction necessary to create valid simulation models. Mans et al. [2013] make a case study using the approach developed by Rozinat et al. [2009a] to evaluate the impact of a new IT system on existing processes in a company. Lamine et al. [2015] seek to improve a management system in an emergency call centre by using simulation. In order to create a valid simulation model they decide to use event log data and process mining tools to generate the necessary information. Their transition approach from process model to simulation model is manual. Leyer og Moormann [2015] conduct an evaluation of different shop-floor scheduling techniques in a service company. In order to create a simulation model of this complex process, they use event log data and process mining techniques. First, they mine a process model by using ProM 6. Secondly, they use statistical methods to calculate the necessary input to create a simulation model. Their findings are preliminary, but addresses the gap between process mining and simulation. Djedovic et al. [2018] carry out an optimisation study of allocating resources. In this regard they point out the problem of traditional methods for simulation model construction. In order to overcome this challenge they suggest the use of process mining and statistical methods. Thus, no integrated approach is developed.

4.4 Literature Summary

Title	**Author**	**Year**	**Theme**	**Summary**
ML-Aided Simulation: A Conceptual Framework for Integrating Simulation Models with Machine Learning	Mahmoud Elbattah and Owen Molloy	2018	Theorethical Framework	A framework aiming for initiating a line of research in the field of integrating simulation modelling and machine learning. The main idea is to create adaptive models that can evolve and adapt automatically.
Integration of Process Mining and Simulation: A Survey of Applications and Current Research	Brian Keith Norambuena	2018	Survey	Identify important litterature in the field and synthesize this into a summary on the challenges imposed by process mining, both for its application and for the integration with simulation
Short-Term Simulation in Healthcare Management with Support of the Process Mining	Pegoraro et al.	2018	Theorethical Framework	Development of a theoretical framework that generates simulation models out of event data to support short-term decision making in healt-care
Belief network discovery from event logs for business process analysis	Titas Savickasa and Olegas Vasilecasb	2018	Theorethical Framework	The authors suggest an approach where Beleif Networks are created from event logs, which can in the future potentially be used to generate simualtion models when more research has been conducted
Simulation of patient flow in multiple healthcare units using process and T data mining techniques for model identification	Kovalchuk et al.	2018	Theorethical Framework	A theorethical framework is built for creating simulation models in a healt care system. However, the study does not involve automatically generated simualtion models

Title	**Author**	**Year**	**Theme**	**Summary**
Innovative Approach in Modeling Business Processes with a Focus on Improving the Allocation of Human Resources	Djedovic et al.	2018	Recognising the challenge	Recognises the problem in buiding simulation methods by traditional methods. Suggest to use process mining and statistical methods to aquire data that can be used to construct model manually.
A Hybrid Process-Mining Approach for Simulation Modelling	Abohama et al.	2017	Theorethical Framework	Using Process Mining to assist in the manual constrcution of a simulation model
Using Event Log Knowledge to Support Business Process Simulation Model Construction	Niels Martin	2017	Partly Approach	Creates an overview of the simulation model construction process and uses that to identify modelling tasks that can be improved by using event logs. The result is four algorithms that can generate 1) Entity arrival rate; 2) Batch processing behaviour; 3) Resource scheduling and availability.
Performance Analysis with DEA, Process Mining and Business Process Simulation on a Livestock Process	M. Leyer and S. Hüttel	2017	Conceptual Integrated Approach - Low level of details	Develop a methodology for integrating process mining, data enveloping analysis and business process simulation. Context specific.
The Use of Process Mining in Business Process Simulation Model Construction - Structuring the field	Marin et al.	2016	Survey	Making a survey of relevant process mining litterature and state-of-the-art litterature on integration of process mining and simulation. They conclude that not much research has been done and very little knowledge is available to overcome this issue. They claim that their research is a starting point to structurally integrate PM in simulation model construction.

Title	Author	Year	Theme	Summary
An integrative approach to simulation model discovery: Combining system theory, process mining and fuzzy logic	Wang et al.	2015 (Revised 2018)	Conceptual Integrated Approach - High level of detail	Developing a three step approach to discover simulation models directly from event data using ProM 6 and Simstudio. Approach needs improvements in terms validation and model enrichment. Builds on Rozinat et al. (2009).
Using Process Mining to Model Inter-arrival Times: Investigating the Sensitivity of the ARPRA Framework	Martin et al.	2015	Algorithm test	Conducting a sensitivity analysis on the ARPRA framework, which is an algorithm that can model arrival rates based on event log data.
Integrating Process Mining with Discrete-Event Simulation Modeling (Master Thesis)	Siyao Liu	2015	Conceptual Integrated Approach - High level of detail	Adressing the lacking compatibility between process mining tools and simulation tools. Develop a new plug-in for ProM 6, which should mitigate this problem by using a more generic output format in terms of Excel .xls format that can be read by simualtion engines. Early stage of development. Needs further research. Builds on Rozinat et al. (2007)
Towards Traditional Simulation Models of Context Using Process Mining	Pileggi et al.	2015	Recognising the challenge	Applying process mining techniques to mine process models from sensor data. Eventually they recognise the similarities between this model and traditional simualtion models. Therefore they envision an automated approach for mining simulation models.

Title	Author	Year	Theme	Summary
Comparing concepts for shop floor control of information-processing services in a job shop setting: a case from the financial services sector	Michael Leyer and Jürgen Moormann	2015	Theorethical Framework	Evaluating shop-floor scheduling control methods in a service company by creating a simulation model by using event log data and process mining. The approach is manual, but provides some preliminary insights to the integration of process mining and simulation.
Improving the Management of an Emergency Call Service by Combining Process Mining and Discrete Event Simulation Approaches	Lamine et al.	2015	Theorethical Framework	Seeking to improve a management system, process mining is applied to create a process model. Insights from this model is used to create a simulation model in the Witness software tool. The approach is manual.
Service Analysis and Simulation in Process Mining	Arik Senderovich	2015	Theorethical Framework - partly	Conducting a research study on the possibility to derive queuing models from event log data.
Discrete modeling and simulation of business processes using event logs	Ivan Khodyrev and Svetlana Popova	2014	Conceptual Integrated Approach - Low level of details	Expanding the integrated approach combining process mining and simulation that has previously been developed by Rozinat et al. (2009). The contribution is additon of macro environement factors.
Automated Knowledge Discovery and Data-Driven Simulation Model Generation of Construction Operations	Reza Akhavian and Amir H. Behzadan	2013	Conceptual Integrated Approach - Low level of detail	Creating an approach for generating simulation models directly from sensor data in the construction industry. Uses Stroboscope software and data mining techniques e.g. Clustering. The objective is short-term simulation.

Title	Author	Year	Theme	Summary
A process-oriented methodology for evaluating the impact of IT: A proposal and an application in healthcare	Mans et al.	2013	Conceptual Integrated Approach - High level of detail	Uses the approach developed by Rozinat et al. (2009) to analyse the impact of system changes.
Combination of Process Mining and Simulation Techniques for Business Process Redesign: A Methodological Approach	Aguirre et al.	2013	Conceptual Integrated Approach - High level of detail	Builds on the approach by Maruster and Beest (2009) and Rozinat et al. (2009) by giving more focus to the project understanding phase.
Business Process Simulation Revisited	Wil M.P. van der Aalst	2010	Survey	Gives an overview of innovative simulation approaches using process mining as support. Methods includes the use of ProM 6 and YAWL framework.
Redesigning business processes: a methodology based on simulation and process mining techniques	Laura Maruster and Nick R. T. P. van Beest	2009	Conceptual Integrated Approach - High level of detail	Building on the approach by Rozinat et al. (2009), they develop a so called "bottom-up" approach for mining and simulating process models. They illustrate the approach through three case studies.
Workflow simulation for operational decision support using event graph through process mining	Liu et al.	2012	Conceptual Integrated Approach - High level of detail	Uses the approach developed by Rozinat et al. (2009) to create a case study of work flow simulation.

Title	Author	Year	Theme	Summary
Workflow simulation for operational decision support	Rozinat et al.	2009b	Conceptual Integrated Approach - High level of detail	An approach for building simulation models on the basis of event logs is developed with focus on short-term simulation for operation decision support. The research work is the first of its kind as it focus on short-term simulation.
Discovering simulation models	Rozinat et al.	2009	Conceptual Integrated Approach - High level of detail	This work is a continuation of the prior work conducted by the same authors. The approach is refined and tested on real life data in this work.
Discovering coloured Petri nets from event logs	Rozinat et al.	2008	Conceptual Integrated Approach - High level of detail	This work focus on the integration of process mining and simulation and is the first of its kind. The paper has laid the foundation for much further work in the field.
Business Process Simulation for Operational Decision Support	Wynn et al.	2007	Proof-of-concept	The work outlines the idea of combing process mining and simulation in a proof-of-concept manner.

Table 4.2: Literature study summary

4.5 Sub-Conclusion

Upon reading through the above mentioned articles to identify the current state of research, it can be concluded that the challenge of integrating process mining and simulation is not new in research. However, one could also argue that the amount of research in the field is very limited. According to the findings in this paper, only 19 articles dealing with development of approaches for integration has been published since 2007. Moreover, none of these articles presents a finished approach that can be applied in a general context. The central issues and challenges that has been identified throughout the literature review is the following:

1. Many studies rely on a great deal of relaxing assumptions in the development of approaches for creating simulation models by the use of process mining Martin [2017].
2. A definition of the required data and data structure for mining simulation models is lacking.
3. Process mining tools are today mainly used for discovering process models. However, in order to "mine" input parameters to simulation models from event logs, additional

algorithms are required.

4. No single software tool exist, which has both mining and simulation capabilities even though the separate tools ProM 6 (with plug-ins) and CPNTools has been used together for several research projects to create a somewhat integrated approach.

As it can be seen, there are still some major challenges to be overcome in order to reach a point where process mining and process simulation can be used in symbiosis with each other. The next chapter will set the scope for the remainder of this report in order to identify a delimited research area that can be investigated within the time frame of the semester.

Refined Research Question 5

As it has been concluded in previous chapter, there are still several issues in relation to the challenge of integrating process mining and simulation. One of these issues, is the lack of methods for mining input parameters to simulation models, which is a necessity for integrating the two approaches. This issue can potentially arise from the fact that most existing process mining techniques focus on process discovery and conformance checking [Hompes et al., 2016]. Thus, little research has been done to clarify the potential for analysing business process performance, by using process mining tools and techniques. It has been stated by Milani og Maggi [2018], that most process mining techniques focus on *time* in form of process, fragment, activity and waiting duration, when talking about process performance. Additionally, few methods exist for investigating resource and quality performance of processes. Thus, there seems to be given little to no attention to the topic of measuring actual *process capacity*. One could argue that the production capacity can be calculated when the cycle time of a system is known. However, the problem with the this method is that the process time might in many cases vary significantly due to the varying conditions in a given production system or process. Thus, if the production capacity is simply assumed to be $\frac{timeperiod}{cycletime}$, there is a risk of calculating an optimistic capacity i.e. too much capacity. If optimistic capacity measures are used for planning or simulation purposes, the result might most likely be shortage of production capacity. In a production company, one of the most important aspects is the alignment of supply and demand. A mismatch between these two aspects will either lead to unsatisfied customers due to product shortage or unnecessary cost and waste in form of over-production or over-capacity. Both instances are usually associated with high costs as capacity is often in-reversible investments.

Traditional process mining techniques seems to focus on time aspects and e.g. resource utilisation rate when talking about capacity. However, it seems relevant to ask: *how can these measures be used? What is the actual time available? and how can resource utilisation be used for planning purposes? Should we increase work load if the process mining reveal 70 percent resource utilisation?* If we assume that workers "do their best" over a given period of time, we should be able to use the best case in this period as a measure of the maximum capacity? In other words, the right question seems to be if it would not be more relevant to know what we can actually do over a given period of time and identify best practise? The time period for best practise could then also be investigated to reveal why the output was higher in this period of time.

The objective for the remainder of this project will be to investigate how process mining can be used in connection with determining the *actual capacity* of a given production system, defined by output per time period, rather than statistical measures of process times, throughput times, waiting times etc. The refined research question is therefore:

> ***How can process mining and event logs be used to determine the actual production capacity in form of output per time period?***

The definition of actual capacity is in this case similar to the definition provided by Business Dictionary [2019a], which is output volume i.e. amount of products, items, weight etc. that can be produced over a given period of time. In order to answer the problem statement, relevant literature will be identified and screened for available methods that can be used either directly or modified for the purpose of this project. Finally, the proposed method will be tested on a data set.

5.1 Requirements for Solution

In this section, the requirements for the solution has been outlined in order to guide the development analysis. Based on the research question above, the solution should be an approach to determine production capacity in form of output per time period. The approach should be able to do this with *event logs* as data input since the approach should be based on process mining for integration purposes with simulation tools. It has been argued in Hompes et al. [2016] that process performance depends on a variety of factors such as difficulty of tasks and the resource performing the task. Thus, the approach developed in this project should be able to differentiate between output types and the resource(s) performing the task.

5.2 Delimitation

There might be several ways of defining and measuring capacity. However, in this project, capacity is measured as output per time period as that is assumed to be the actual capacity. Thus, the reasons (besides output type and resource) for this given capacity is ignored in this project.

Capacity Mining Approach 6

In this chapter, the purpose is primary to investigate if any available process mining techniques exist for the determination of production capacity, defined by volume per time unit. Secondary, general methods for determination of production capacity should be investigated in order to identify any relevant methods to be included in the solution development. Finally, these methods should be evaluated and potentially refined or developed for testing on a data set. Thus, this part of the study is of a rather deductive nature.

6.1 Relevant Literature

In order to seek for relevant literature, Scopus has again been used as search engine. The search strings applied and search results has been listed in table 6.1.

Key Words	Boolean Operator	Key Words	Database	Articles
"Process mining"	AND	Capacity	Scopus	22
"Process mining"	AND	"Performance analysis"	Scopus	90
"Process mining"	AND	"Capacity planning"	Scopus	1
"Process capacity"	AND	"Calculate"	Scopus	9

Table 6.1: Literature study search result

Upon screening through the search results it turned out that there were three relevant articles and one book presenting approaches to identify production capacity related measures. Further, a book were introduced by the project supervisor, Raoul Waldburger, which gave valuable insights to the determination and calculation of process capacity.

Starting with a theoretical outset for calculating capacity, the book by Damij og Damij [2014] presents a range of properties regarding process capacity. First and foremost, the book set out a definition of process capacity, originally presented by Anupindi et al. [2006], which is: *"The process capacity is the maximum sustainable flow rate of a process"*, which aligns well with the definition set out for capacity in this project. Further, great emphasis is given to the importance of resources (capital and labour) impact on process capacity. More specifically, it is argued that great attention should be given to the availability of resources required to perform the process, rather than merely analysing the actual activities performed. Another important characteristic is that a process cannot have a higher capacity than that of the bottleneck capacity. I.e. the process capacity is determined by the step in the process with the lowest capacity. Finally, the most interesting part for this project, is the formula presented to calculate *theoretical capacity* of a process *Rp*:

- Rp = 1/Tp * Load Batch * Scheduled Availability (process with one resource)
- Rp = Cp/Tp * Load Batch * Scheduled Availability (process with n resources)

Where Tp is Unit Load of a resource, which is the total amount of time the resource works to process each flow unit. Cp is the amount of resources in a resource pool (if the resources have the same Unit Load). Load Batch is the amount of flow units that can be processes simultaneously. Scheduled Availability is the time that each resource is available for the given period of time. With this calculation, it is possible to calculate the theoretical capacity for a given process, taking into account the resource performance, availability and simultaneous work/processing. This calculation is similar to the method presented by Chen et al. [2013]. This study undertakes an analysis of the capacity fluctuation on a production line. For this purpose they calculate an upper and lower boundary of capacity utilisation. The interesting part is the way that they calculate the so called *available capacity, CA*:

- CA = l * t * α * β * y

Where l is the number of similar machines/equipment in the process, t is the available capacity of a single machine/equipment (scheduled work hours). α is the non-planned downtime. β is the capacity of an operator (process time). y is the utilisation rate of the process. This method is similar to the one presented by Damij og Damij [2014]. These two methods seems to be in line with the traditional way of calculating capacities, which rely on the assumption of stable/static process times. Thus, the variation in these times might not be taken into account.

Another article by Nguyen et al. [2016], states that traditional process mining tools are good for process model discovery, which can additionally be used to identify bottlenecks, process duration, throughput times and calculate descriptive statistics of these time related measures. However, current process mining tools are not able to answer questions such as: how bottlenecks form and dissolve over time nor how the formation and dissolution of bottlenecks – and associated fluctuations in demand and capacity – affect the overall process performance. They suggest an approach called "staged process flows" where queuing theory is used to analyse and answer above mentioned questions. In their study, production capacity for each step is defined by the queuing theory concept of *departure rate*, which can be determined for each process step, part of the system or the entire system. This calculation method seems to be a useful measure for capacity as it is an expression for what the process is *actually* able to do. I.e. if it is assumed that a company is working at maximum capacity, and the sustainable departure rate is 100 items per hour, then that must also be the maximum capacity of the process. It should then be said, that there might be several reasons why this measure is not satisfying for the company. The design capacity might for instance be 200 items per hour. However, that is considered to be a topic for an additional problem analysis, beyond the scope of this project. An example where the departure rate measure could be useful is for planning purposes. In the daily operations of a company, the planning manager or operations manager might want to know *what are we able to produce of product A with resource B per hour?* in order to perform Sales and Operations Planning. Again, this might not be satisfying and an additional analysis should be undertaken to answer *why are we only able to produce this amount?*. For this project, the focus is solely, *what* is the actual capacity.

An article by Park et al. [2015], is about workload and delay analysis in manufacturing processes. The idea is to measure workload on processes in terms of volume over time or work-in-progress over time. They use these measures in a comparison analysis, where the planned situation is compared to the actual situation in order to derive deviations from the plan. Further, they investigate the delays of cases, both in terms of process time delay and start date delay, which can be used to explain the causes of delay. The relevance to this project is that one could argue that the system capacity could be derived by measuring the amount of completed tasks on time. I.e. the completed tasks on time is the actual output. However, the method presented in the article does not reflect this measure as an objective and a great deal of reconfiguration seems to be necessary. Another aspect is that the work-in-progress could also be viewed as a capacity. I.e. that the process is able to work with a given amount of orders simultaneously. Again, this does not align with the definition of actual capacity in terms of output per time.

Finally, the book by van der Aalst [2016] provides a seemingly complete and detailed overview of the applications, possibilities and limitations of process mining. The author has contributed to numerous articles and research works in the field of process mining and the book is based on his insights and his extensive amount of research. In this book, it is also recognised that the most frequent application of process mining is process discovery and conformance checking. However, the book also presents a range of additional applications such as: organisational mining, social network analysis, organisational structures, resource behaviour, time, probabilities and decision mining. Many of these applications might still to this date be an "envisioned" situation as some of these methods are still at an early stage of development and has mainly been tested in research projects. Relevant to this project, is mostly the *time* aspect, which is probably also the most developed area, additional to process model discovery. Commercial process mining tools such as Disco has integrated functions for calculating waiting times, process times, through put times, cycle times, Work-in-Progress and resource utilisation. The relevance for this project is that one could argue that capacity can be calculated on the basis of process times as mentioned above. Thus, when the average process time or throughput time is known, the calculation would be $\frac{TotalTime}{ProcessTime}$. However, by using average numbers for calculating capacity, one might risk that the calculations are wrong due to deviations in the measures. I.e. the best and worst case scenarios are "hidden".

In conclusion, based on the above presented studies and methods, it still seems like there is not a yet a complete method available for calculating actual capacity in a production system, as it has been set out in this project. However, ideas from the different studies seems to be useful in the development of such method. The approach from queuing theory presented by Nguyen et al. [2016] is on a high level the approach that seems most accurate. Thus, *departure rate* seems to be the nominator for capacity throughout the remainder of this project.

6.2 Defining the Problem

Nguyen et al. [2016] presents the term *departure rate* as a measure for capacity, which fit with the definition of capacity in this project. Departure rate is in queuing theory expressed by μ, which is the average number of items leaving the system per time unit. The inter-departure time is expressed as $\frac{1}{\mu}$, which is the average time between each item leaving the system. Thus, μ will be the foundation for the approach to be developed.

In order to fulfil the solution requirements set out in 5, the approach should take into account the type of output and the resource that produces the output. Thus, an algorithm to mine μ from event logs, that takes into account output type and resource should be defined.

According to several sources such as IEEE [2011]; Rozinat et al. [2008]; van der Aalst [2010] the foundation for process mining is the availability of data in form of event logs. Further, van der Aalst [2016] makes the distinction between "simple" event logs and "enriched" event logs. The simple event log contains cases with associated events. I.e. a sequence of cases containing events. In the simple version, the cases does not carry any identifier and the event log contains no attributes, which is simply additional information. From the simple event log, a process miner software such as ProM 6 can be used to derive a process model in form of a e.g. a Petri Net. In order to start making more advanced analysis on processes, the event log has to contain more detailed information. This additional detailed information comes in form of attributes. These attributes can contain information such as time stamps, resource, case ID, context, product type etc. Which information is needed depends on the particular situation and what is going to be the objective for analysis. Further, it should be determined in the data logging process in order to create the desired data structure from the start. Thus, according to the solution requirements set out for this project, it seems that the event logs to be used for analysis and determination of capacity in form of output/time should contain following information:

- case ID
- work station/process step
- time stamps (finished and completed)
- resource ID
- product ID

van der Aalst [2016] mentions three main types of process mining: discovery, conformance, and enhancement. The third type focus on extending or improving an existing process model. Enhancement further contains two sub-categories: repair and extension. Extension is about adding perspectives to the original process model, derived from a simple event log. Thus, this project can be seen as a contribution to the body of research within *enhancement process mining and process model extension*.

Case Study 7

The former analysis of available literature was of a deductive nature, seeking to find explanations of capacity and how to measure or determine capacity in a production system by means of theory. To evaluate these methods and to seek in a more experimental manner, this chapter will be of an inductive nature. Thus, findings from chapter 6 will be used as a basis for the case study by providing some theory for calculating capacity. However, experiments conducted on a data set [4TU.ResearchData, 2019] should provide insights on how the methods could potentially be developed or adopted in the process mining domain. In the remainder of this chapter, the experimental case study and approach development is described.

7.1 Case Description and Proposed Solution

In order to establish a context for the approach development, one can imagine the following situation. A production system in company X, where different products are produced on a range of several different machines. I.e. a production network of machines where products might follow different paths depending on the desired end-result. In such case, it can be difficult and complicated to calculate the production capacity because it depends on a broad range of different variables. As written in section 6.1, the major variables for capacity determination are: amount of resources, availability of resources and capacity of resources. However, in reality, these variables can be hard to determine and will often be based on assumptions or best guesses. When determining these variables, one should also take into account which type of product is being produced and the path that the product follows, making the calculation even more complex.

In order to seek a reduction of the complexity, it is therefore proposed to consider the capacity of such a production system in terms of *path capacities*. This is inspired by Damij og Damij [2014], who defines the term *resource pools*, which is a set of resources available to produce or serve the same product or customers. By defining a resource pool in a larger system, they are able to focus on and determine the capacity for such a resource pool. This thinking is in this project applied to the idea of *path capacity*. More specifically, capacity can be determined for different paths through the production network for different products. This should primary be enabled by having historical data available for the production system. In this way, process mining could potentially be used as a tool to first and foremost identify the different *"paths"* in the production system. Subsequently, process mining should be used to calculate the flow of products through these paths including information about the resources. This would make it possible to determine capacities for given paths under certain conditions. The proposed idea of *"paths"* has been illustrated in figure 7.1. The advantage of defining capacities for such paths in a system is that one

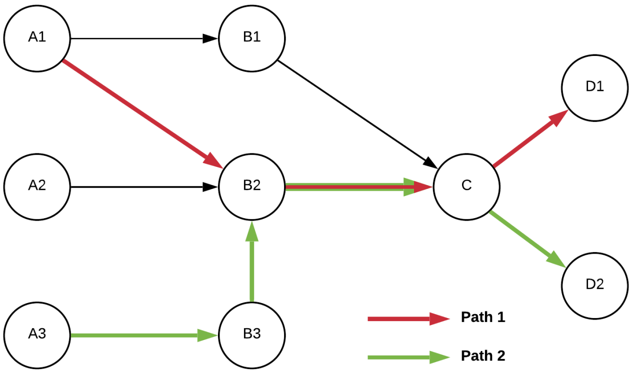

Figure 7.1: *Example of path notation*

would ensure a more aggregated level of capacity, which could easily be used for planning and scheduling of e.g. a production shop floor. I.e. the planning manager would know the consequence of using a certain path through the production system, which could quickly be used to determine KPI's like *available to promise* for a sales department.

7.2 Data Analysis Tool Selection

The first issue that occurred during the experimental stage of the project was to figure out how the capacity mining approach should be implemented. By implementation is meant the environment to be used in terms of software to carry out the data analysis. There are several tools available, and combinations hereof is possible. In order to achieve an overview of a group of selected tools with different capabilities, the following table were developed: As this project is a Business Engineering project, a full scale development of algorithms

Tool	Pros	Cons
Excel	Simple, well known software for basic data analysis. Easy and quick to use.	Limited capacity.
Python	Powerful data analysis tool, with many options.	Require programming skills and is more time consuming.
R studio	Powerful analysis tool with focus on statistics.	Require programming skills and is specific for statistics.
ProM 6	Process Mining tool with a rather open development environment.	Require programming skills in terms of coding packages that can do the desired data analysis
Disco (fluxicon)	Intuitive process mining tool. Easy to use.	Limited range of functions.

Table 7.1

in e.g. the Python or ProM 6 environment, seems to be out of scope. Therefore, Disco and Excel were selected as tools to carry out the experimental data analysis because these tools contains the necessary functions and are considered most intuitive and easy to use.

7.3 Process Mining and Data Analysis

As the tools for data analysis were selected, the actual case work could be initiated. The data set used for the case study is provided by [4TU.ResearchData, 2019] and has been used for process mining research projects by other scholars earlier. The data set is essentially an event log from a purchasing process from an unknown company. This process includes a total of 24 steps. However, not all cases goes through all of the 24 steps. I.e. there are different *paths* throughout the process, which the different purchases can follow. In total there are 98 variants of these paths. There are additionally 27 resources (people) to perform the different steps throughout the process.

First step in the data analysis was to run the event log through Disco, which is a Process Mining tool that is easy to use, and which contains the most fundamental functions of Process Mining. In this project, Disco was first and foremost used to explore the data in order to gain an overview of the content and to check if the data set was suitable for this project. The criterion, beside being an event log, were that the the data set should include information about the resources in order to distinct between these for capacity calculations. Secondly, Disco was used to identify the different *paths* in the process, which is easily done by the standard functions in Disco. Moreover, Disco can export a part of the event log by applying filters to the analysis. Thus, the most common path was identified, and the isolated event log containing the cases that had followed this particular path was exported to an Excel file for further analysis. The limitation of Disco is that one cannot perform detailed analysis on single criterion such as average time + standard deviation on single process steps for a particular resource. Therefore, the analysis had to continue in Excel, which provides more options for specific data analysis.

The data analysis steps for Excel has been described in the following list:

1. The event log was sorted according to the process steps, i.e. step 1, 2, 3 etc.
2. Data for each of the process steps were split among different sheets for individual calculations.
3. The data rows within each new sheet, for each process step, were sorted according to the resource performing the process step.
4. Sub-totals were added for average time duration and standard deviation for each resource in each process step.
5. The sub-totals were collected into a single table to give a complete overview and to summarise the process times.

The results of the Excel data analysis has been presented in table 7.2.

The purpose of the excel analysis was, as mentioned earlier, to calculate the path capacity for the most common path in the purchase process described in the data set. As it can be seen in table 7.2, there is 438 minutes difference between the fastest and slowest process

	Fastest Average			Slowest Average			Over all	
Activity	**Avg. Time**	**Std. Dev.**	**Res.**	**Avg. Time**	**Std. Dev.**	**Res.**	**Avg. Time**	**Std. Dev.**
Create Purchase Requisition	19	15	TL	38	13	MH	30	16
Analyze Purchase Requisition	6	2	FO	6	2	HG	7	12
Create Request for Quotation Requester Manager	2	0	N/A	2	0	N/A	2	0
Analyze Request for Quotation	21	9	FP	24	7	KG	22	9
Send Request for Quotation to Supplier	22	6	KG	25	6	FP	23	7
Create Quotation comparison Map	199	65	FP	229	88	MP	207	91
Analyze Quotation comparison Map	16	6	TL	26	8	CF	21	7
Choose best option	0	0	N/A	0	0	N/A	0	0
Settle conditions with supplier	535	187	FP	587	207	MP	568	203
Create Purchase Order	9	2	MP	10	3	KG	10	2
Confirm Purchase Order	18	10	KC	22	9	KK	20	9
Deliver Goods Services	1325	669	EC	1642	933	KC	1458	793
Release Purchase Order	1	0	N/A	1	0	N/A	1	0
Approve Purchase Order for payment	1	0	N/A	1	0	N/A	1	0
Send invoice	0	0	N/A	0	0	N/A	0	0
Release Supplier's Invoice	5	1	PA	5	2	KN	5	2
Authorize Supplier's Invoice payment	0	0	N/A	0	0	N/A	0	0
Pay invoice	9	3	KN	10	3	PA	9	3
Total Process Times	**2188**			**2626**			**2385**	

Table 7.2: *Results of Excel data analysis. Time in minutes.*

duration. Further, it is worth noticing that the overall (not taking into account resources) average for the process times are also very different from the slowest and fastest. This finding supports the risk that has been pointed out earlier in chapter 5. I.e. it is here proven that more accurate results can be obtained by differentiating between resources (in this case people), as different resources have different capabilities.

7.4 Capacity Calculation

After the time required to perform each process step has been calculated, the formula from earlier should be recalled from chapter 6.1. The formula presented, was for calculating *theoretical capacity* of a process *Rp*:

- Rp = 1/Tp * Load Batch * Scheduled Availability (process with one resource)
- Rp = Cp/Tp * Load Batch * Scheduled Availability (process with n resources)

- Tp = Unit Load of a resource, which is the total amount of time the resourceworks to process each flow unit.
- Cp = Cp is the amount of resources in a resource pool (if the resources have the same Unit Load)
- Load Batch = Amount of products/customers can be served simultaneously
- Scheduled availability = The time the resource(s) are available

Now, the theoretical capacity can be calculated according to the identified formula in literature from earlier. As it has been mentioned earlier in chapter 6.1, Damij og Damij [2014] points out that, in order to calculate the capacity for a given process, one can simply identify the *slowest* process step, also called the bottleneck, as the process can never be faster than the slowest link. Hence, looking at table 7.2, one can identify the slowest process step to be *Deliver Goods Services*. The capacity in this example will be calculated for both *Fastest Average, Slowest Average and Overall* in order to be able to compare the results.

Starting with Unit Load (Tp), this was the times calculated in table 7.2. This Tp should be divided with either 1 (for one resource) or Cp (the amount of identical resources e.g. machines). Thus, for this example, the Unit Load should be divided with 1 for each process step, as there is only one resource available, with the same capabilities, for each process step. Load Batch is assumed to be 1 because it is expected that the purchasing process steps will require all the resource's attention. Scheduled availability is a variable that will only influence the time period for which the capacity is calculated. I.e. in this example the capacity is calculated for one hour only (60 minutes). However, if one would like to know the capacity for a whole week, one could simply use the hours scheduled for the particular resources throughout a week. The calculations has been summarised in table 7.3.

Process Path (Bottleneck, Deliver goods services)	1/Unit Load (minutes)	*	Load Batch (items)	*	Scheduled Availability (minutes)	=	Departure Rate (μ) (orders per hour)
Fastest Average:	1/1325	*	1	*	60	=	<u>0.0453</u>
Slowest Average:	1/1642	*	1	*	60	=	<u>0.0365</u>
Overall:	1/1458	*	1	*	60	=	<u>0.0412</u>

Table 7.3: Path capacities (calculated for the bottleneck process step) for each of the scenarios for the purchasing process in terms of μ, which is the departure rate. In other words, the amount of orders that can be handled per hour in the process.

7.5 Discussion and Sub-Conclusion

In this section the proposed solution for calculating capacity will be evaluated. More specifically, the solution should be evaluated against the problem statement and requirements set out in chapter 5.1. To recall, the problem statement was:

> *How can process mining and event logs be used to determine the actual production capacity in form of output per time period?*

And requirements have briefly been listed here:

- The solution should be an approach to determine production capacity in form of output per time period.
- The approach should be able to do this with *event logs* as data input, in order to be suitable for process mining purposes or later be implemented in a process mining tool.
- The approach should be able to differentiate between output types and the resource performing the task.

Further, it was stated that the approach would be delimited to focus solely on output type and resources for determination of capacity.

Considering the first requirement, that the solution should be able to calculate capacity in terms of output per time period, the proposed solution fulfil this requirement. The results in table 7.3 is the number of served orders that the purchasing process can process per hour, i.e. μ (departure rate) as described in chapter 6.2.

The second requirement is that the capacity should be derived from an event log. This has also been accomplished. Thus, the approach should be suitable for application with any process mining tool and hereby be a contribution to the work towards integrating process mining and simulation tools. However, it has not been possible to develop the approach in a process mining tool during this project as it would be an extensive task that requires competencies from another field than business engineering.

The third and final requirement is that the capacity calculation should take into account the output types and resource performing the task or process under analysis. Regarding output type, it has not been possible to distinct between exact order types because the data set did not contain this information. However, as an alternative, the calculation has been differentiated on the basis of different paths. Thus, it has been assumed that the orders following the same path, has also been similar in terms of their type. This can of course be questioned. Considering the standard deviation of the average measures, it might indicate that the orders that has been handled has been of different types, rather that difference in performance. Regarding distinction between resources, this has been accomplished. The different capacity measures is seemingly most related to the resources that has been performing the processes.

In conclusion, the proposed solution lives up to the requirements and does overall answer the problem statement by providing an answer to how an event log can be used for "mining" capacity of a given process. Hence, the approach is hereby also a small, but important contribution towards closing the gap between process mining and simulation tools as this approach is able to derive more accurate capacity measures than those of traditional process mining, which tend to focus more on calculating highly aggregated and average measures of throughput and process times. In conclusion, it is proposed that the part of the method that has been carried out in excel, should be refined and built into a process mining tool.

7.6 Business Case Perspective

Upon having developed an approach for *capacity mining* by using event logs, it has been found worth to consider the ability of the project to create value in a business context. In order to do that, the benefits and downsides of simulation should be recalled. According to AnyLogic [2019], which is one of the largest players at the simulation market, some of the major advantages of simulation is: Low risk environment for testing solutions before they are chosen, which saves times and money; visualisation of proposed solutions; increased accuracy in decision support; ability to handle uncertainty through scenario testing. All these capabilities does seemingly have a high potential value for companies. However, there is one significant problem, also pointed out by several authors e.g. IEEE [2011]; Martin [2017]. That is, traditional simulation modelling relies upon the personal interpretation of interviews and observations by the model builder. That is in many cases a source to invalidity of the simulation models. When the simulation model is not representing the real system, the simulation is basically useless as it will not be able to provide realistic information. In worst case, the simulation model will provide wrong information, which is used for decision making, and hereby be the direct cause of wrong decision making with fatal consequences. In terms of capacity, which has been the topic for this project, wrong measures can have such fatal consequences as pointed out in chapter 5.

The over all purpose of this project has been to contribute to the body of knowledge within integration between process mining and simulation. As concluded earlier, one of the most common issues is the lack of ability to calculate input parameters for simulation models by using process mining. The solution of this project provides a small contribution in this direction in form of an approach to calculate capacity by using event logs. I.e. process mining can be used to mine the capacity measures of any given process if an event log is available. Since capacity is an important input parameter for simulation, this single issue has therefore come a little closer to be solved.

Answering how this project can have a direct value for companies is difficult to answer as it might actually not at the moment have direct value. However, by considering the aforementioned advantages of simulation, these advantages might in the future be increased by combining process mining and simulation, which should give more accurate simulation models and hereby more accurate simulation results. However, the approach for capacity calculations could also be valuable in other contexts. I.e. if the approach is integrated into a commercial process mining software, the ability to calculate capacity in a fast and accurate manner, might provide a great amount of value for planning purposes. In all companies, the balance between demand and supply is a recurring issue, which will always have economic consequences when out of balance. In case of over-capacity, resources are under-utilised and hereby costly. In case of under-capacity, the company experience market shortage, which can potentially damage the reputation of the company or merely cause lost turnover. Thus, it can be assumed that all companies are also interested in being able to accurately calculate their actual capacity in order to optimise their planning process. Thus, the proposed solution might also have a value in this context if implemented into a user-friendly tool.

Conclusion 8

Upon having finished the project, this chapter will summarise the findings through out the report. From the beginning of the project, the purpose has been to investigate how process mining techniques could be integrated with process simulation techniques as this seems to have a valuable potential for many companies. The potential was exemplified by comparing the the limitations of both techniques, which revealed that there are several complementary effects to be achieved. Most significantly is the difference in their respective areas of analysis which is: process mining techniques seeks to analyse the as-is situation whereas process simulation seeks to analyse different to-be scenarios. Further, process simulation models are prone to human interpretation, and hereby failures, in the construction process. Process mining seems to be an effective tool to overcome this uncertainty in the model building process as the process mining techniques construct models based on system data. Based on this great potential, the first step in the research work has been to identify the current state-of-the-art in the research field on process mining and process simulation integration. This study revealed some interesting findings. First and foremost, the challenge of integrating process mining and simulation is not new in research, but somehow limited in terms of the amount of articles. Further, none of the articles presents a finished integration approach that can be applied in a general context. The central issues and challenges that has been identified throughout the literature review is the following:

1. Many studies rely on a great deal of relaxing assumptions in the development of approaches for creating simulation models by the use of process mining Martin [2017].
2. A definition of the required data and data structure for mining simulation models is lacking.
3. Process mining tools are today mainly used for discovering process models. However, in order to "mine" input parameters to simulation models from event logs, additional algorithms are required.
4. No single software tool exist, which has both mining and simulation capabilities even though the separate tools ProM 6 (with plug-ins) and CPNTools has been used together for several research projects to create a somewhat integrated approach.

Upon having identified these challenges, it has been necessary to narrow down the scope of the further research work. One of the significant challenges yet to be resolved, is the development of approaches to mine input parameters to simulation models. In particular, a lack of process performance measures was identified in relation to process capacity. Therefore, the refined research question has been: *How can process mining and event logs be used to determine the actual production capacity in form of output per time period?*. In order to answer this problem statement, it was decided to undertake a two step analysis. First step was rather deductive, which meant that a search for literature in the field of process capacity calculation was necessary. This revealed some existing methods on how to calculate process capacity, which could be used for the second step of the analysis. Further, it was possible to set up some requirements to the further solution development, which was: 1) The solution should be able to identify capacity in terms of output per time period; 2) The solution should be able to extract capacity measures by using event logs; 3) The solution should be able to differentiate between resources and output type. In the second part, the aim was to develop and apply identified methods from theory on a data set in order experiment with different possibilities and finally to come up with a solution proposal that would meet the above mentioned requirements. Thus this was a more inductive approach to do the analysis work. The result was successful and ended in a final solution that could fulfil the requirements. In conclusion, this project is a rather small, but important contribution to the body of research within the field of integrating process mining and process simulation. The proposed solution should ideally be refined and built into an existing process mining tool such as the applied tool called Disco as this would enable Disco to mine capacity measures in terms of output per time period, which is used as input parameter for process simulation. Finally, the business case perspective has been considered. In conclusion, this project could have both direct and in-direct value for companies. The direct value could arise if the proposed solution were to be built into an existing user-friendly process mining tool as this would enable companies to calculate their capacity in a quick and accurate manner. Thus, improving their planning system and ability to balance supply and demand. The in-direct value is seen more in relation to the actual purpose of this project, namely to combine process mining and process simulation tools. The advantages of such a combination has been explained earlier. Hence, as this project contributes to the development of an integrated tool, the project could have a stake in the value created by using such a tool.

Bibliography

4TU.ResearchData, 2019. 4TU.ResearchData. *Production Event Log Example*. https://data.4tu.nl/repository/uuid:68726926-5ac5-4fab-b873-ee76ea412399, 2019.

Abohamad et al., 2017. Waleed Abohamad, Ahmed Ramy og Amr Arisha. *A Hybrid Process-Mining Approach for Simulation Modelling*. Winter Simulation Conference (WSC), 2017.

Aguirre et al., 2013. Santiago Aguirre, Carlos Parra og Jorge Alvarado. *Combination of Process Mining and Simulation Techniques for Business Process Redesign: A Methodological Approach*. SIMPDA 2012, LNBIP 162, pp. 24–43, 2013.

Akhavian og Behzadan, 2013. Reza Akhavian og Amir H. Behzadan. *Automated Knowledge Discovery and Data-Driven Simulation Model Generation of Construction Operations*. Proceedings of the 2013 Winter Simulation Conference, pp. 3030-3041, 2013.

Anne Rozinat, 2018. Anne Rozinat. *How Process Mining Compares To Simulation*. https://fluxicon.com/blog/2011/06/process-mining-simulation/, 2018.

Anupindi et al., 2006. Ravi Anupindi, Sunil Chopra, Sudhakar D. Deshmukh, Jan A. Van Mieghem og Eitan Zemel. *Management business process flows: Principles of operations management*. Upper Saddle River: Pearson Education., 2006.

AnyLogic, 2019. AnyLogic. *Why use simulation modelling?* https://www.anylogic.com/use-of-simulation/, 2019.

ARIS Group by Software AG, 2019. ARIS Group by Software AG. *Business Process Simulation Definition*. https://www.ariscommunity.com/business-process-simulation, 2019.

Business Dictionary, 2019a. Business Dictionary. *Production Capacity*. http://www.businessdictionary.com/definition/production-capacity.html, 2019.

Business Dictionary, 2019b. Business Dictionary. *Data Mining Definition*. http://www.businessdictionary.com/definition/data-mining.html, 2019.

Business Dictionary, 2019c. Business Dictionary. *Queuing Theory Definition*. http://www.businessdictionary.com/definition/queuing-theory.html, 2019.

Chen et al., 2013. Bing Chen, Kai Liu og Ze bing Fan. *Research on the capacity fluctuation analysis of compressor blade production line*. International Asia Conference on Industrial Engineering and Management Innovation, pp. 765–775, 2013.

Damij og Damij, 2014. Nadja Damij og Talib Damij. *Process Management - A Multi-disciplinary Guide to Theory, Modeling, and Methodology*. Springer Link, 2014.

der Aalst, 2015. Wil M.P. Van der Aalst. *Business Process Simulation Survival Guide*. Handbook on Business Process Management 1 , International Handbooks on Information Systems, Second Edition, Springer-Verlag Berlin Heidelberg, 2015.

Djedovic et al., 2018. Almir Djedovic, Almir Karabegovic, Zikrija Avdagic og Samir Omanovic. *Innovative Approach in Modeling Business Processes with a Focus on Improving the Allocation of Human Resources*. Mathematical Problems in Engineering Volume 2018, pp. 1-14, 2018.

Elbattah og Molloy, 2018. Mahmoud Elbattah og Owen Molloy. *ML-Aided Simulation: A Conceptual Framework for Integrating Simulation Models with Machine Learning*. SIGSIM-PADS'18, pp. 33-36, 2018.

Hompes et al., 2016. Bart F.A. Hompes, Joos C.A.M. Buijs og Wil M.P. van der Aalst. *A Generic Framework for Context-Aware Process Performance Analysis*. Springer International Publishing AG, LNCS 10033, pp. 300–317, 2016.

IEEE, 2011. IEEE. *Process Mining Manifesto*. Business Process Management Workshops 2011, Lecture Notes in Business Information Processing, Vol. 99, Springer-Verlag, 2011.

Khodyrev1 og Popova, 2014. Ivan Khodyrev1 og Svetlana Popova. *Discrete modeling and simulation of business processes using event logs*. Procedia Computer Science Volume 29, pp. 322–331, 2014.

Kovalchuka et al., 2018. Sergey V. Kovalchuka, Anastasia A. Funknera, Oleg G. Metskera og Aleksey N. Yakovleva. *Simulation of patient flow in multiple healthcare units using process and data mining techniques for model identification*. Journal of Biomedical Informatics, Volume 82, June, pp. 128-142, 2018.

Lamine et al., 2015. Elyes Lamine, Franck Fontanili, Maria Di Mascolo og Hervé Pingaud. *Improving the Management of an Emergency Call Service by Combining Process Mining and Discrete Event Simulation Approaches*. International Federation for Information Processing, AICT 463, pp. 535–546, 2015.

Leyer og Hüttel, 2017. M. Leyer og S. Hüttel. *Performance Analysis with DEA, Process Mining and Business Process Simulation on a Livestock Process*. Proceedings of the International MultiConference of Engineers and Computer Scientists Vol II, IMECS, March 15 - 17, Hong Kong, 2017.

Leyer og Moormann, 2015. Michael Leyer og Jürgen Moormann. *Comparing concepts for shop floor control of information-processing services in a job shop setting: a case from the financial services sector*. International Journal of Production Research, 53:4, 1168-1179, 2015.

Liu, 2015. Siyao Liu. *Integrating Process Mining with Discrete-Event Simulation Modeling*. All Theses and Dissertations. 5735. `https://scholarsarchive.byu.edu/etd/5735`, 2015.

Liu et al., 2012. Ying Liu, Hui Zhang, Chunping Li og Roger Jianxin Jiao. *Workflow simulation for operational decision support using event graph through process mining*. Decision Support Systems Volume 52, pp. 685–697, 2012.

Mans et al., **2013**. Ronny Mans, Hajo Reijers, Daniel Wismeijer og Michiel van Genuchten. *A process-oriented methodology for evaluating the impact of IT: A proposal and an application in healthcare*. Information Systems Volume 38, pp. 1097–1115, 2013.

Martin, **2017**. Niels Martin. *Using Event Log Knowledge to Support Business Process Simulation Model Construction*. BPM, 2017.

Martin et al., **2015**. Niels Martin, Benoit Depaire og An Caris. *Using Process Mining to Model Inter-arrival Times: Investigating the Sensitivity of the ARPRA Framework*. Proceedings of the 2015 Winter Simulation Conference, pp. 868-879, 2015.

Martin et al., **2016**. Niels Martin, Benoit Depaire og An Caris. *The Use of Process Mining in Business Process Simulation Model Construction*. Business Information Systems Engineering 58(1), pp. 73–87, 2016.

Maruster og van Beest, **2009**. Laura Maruster og Nick R. T. P. van Beest. *Redesigning business processes: a methodology based on simulation and process mining techniques*. Knowledge Information Systems Volume 21, pp. 267–297, 2009.

Milani og Maggi, **2018**. Fredrik Milani og Fabrizio Maria Maggi. *Comparative Evaluation of Log-Based Process Performance Analysis Techniques*. Springer International Publishing AG, LNBIP 320, pp. 371–383, 2018.

Nguyen et al., **2016**. Hoang Nguyen, Marlon Dumas, Arthur H.M. ter Hofstede, Marcello La Rosa1 og Fabrizio Maria Maggi. *Business Process Performance Mining with Staged Process Flows*. Springer International Publishing Switzerland, LNCS 9694, pp. 167–185, 2016.

Norambuena, **2018**. Brian Keith Norambuena. *Integration of Process Mining and Simulation: A Survey of Applications and Current Research*. Springer International Publishing AG, WorldCIST', AISC 745, pp. 287–294, 2018.

Park et al., **2015**. Minjeong Park, Minseok Song, Tae Hyun Baek, SookYoung Son, Seung Jin Ha og Sung Woo Cho. *Workload and Delay Analysis in Manufacturing Process Using Process Mining*. Springer International Publishing Switzerland, LNBIP 219, pp. 138–151, 2015.

Pegoraro et al., **2018**. Fábio Pegoraro, Eduardo Alves Portela Santos, Eduardo de Freitas Rocha Loures, Gabriela da Silva Dias, Lucas Matheus dos Santos og Renata Oliveira Coelho. *Short-Term Simulation in Health-care Management with Support of the Process Mining*. Springer International Publishing AG, part of Springer Nature, Á. Rocha et al. (Eds.): WorldCIST'18, AISC 746, pp. 724–735, 2018.

Pileggi et al., **2015**. Paolo Pileggi, Alejandro Rivero-Rodriquez og Ossi Nykänen. *Towards Traditional Simulation Models of Context Using Process Mining*. 7th International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN), 2015.

Process Mining Group, Math&CS department, Eindhoven University of Technology, **2019**. Process Mining Group, Math&CS department, Eindhoven

University of Technology. *Process Mining Definition*. http://www.processmining.org, 2019.

Rozinat et al., **2008**. A. Rozinat, R.S. Mans, M. Song og Wil M.P. Van der Aalst. *Discovering colored Petri nets from event logs*. International Journal of Software Tools Technology Transfer Volume 10, pp. 57–74, 2008.

Rozinat et al., **2009a**. A. Rozinat, R.S. Mans, M. Song og Wil M.P. Van der Aalst. *Discovering simulation models*. Information Systems Volume 34, pp. 305–327, 2009.

Rozinat et al., **2009b**. A. Rozinat, M.T. Wynn, Wil M.P. Van der Aalst, A.H.M. ter Hofstede og C.J. Fidge. *Workflow simulation for operational decision support*. Data and Knowledge Engineering Volume 68, pp. 834–850, 2009.

R'bigui og Cho, **2017**. Hind R'bigui og Chiwoon Cho. *The state-of-the-art of Business Process Mining Challenges*. Int. J. Business Process Integration and Management, Vol. 8, No. 4, pp.285–303, 2017.

Savickasa og Vasilecas, **2018**. Titas Savickasa og Olegas Vasilecas. *Belief network discovery from event logs for business process analysis*. Computers in Industry, Volume 100, September, pp. 258-266, 2018.

Senderovich, **2015**. Arik Senderovich. *Service Analysis and Simulation in Process Mining*. Springer International Publishing Switzerland, BPM 2014 Workshops, LNBIP 202, pp. 578–581, 2015.

van der Aalst, **2016**. Wil M.P. van der Aalst. *Process Mining - Data Science in Action*. Springer Link, 2016.

van der Aalst, **2010**. Wil M.P. van der Aalst. *Business Process Simulation Revisited*. EOMAS 2010, LNBIP 63, pp. 1–14,, 2010.

Wang et al., **2015**. Yan Wang, Grégory Zacharewicz, Mamadou Kaba Traoré og David Chen. *An integrative approach to simulation model discovery: Combining system theory, process mining and fuzzy logic*. Journal of Intelligent and Fuzzy Systems, IOS Press, 34 (1), pp. 477 - 490, 2015.

Wynn et al., **2007**. M.T. Wynn, M. Dumas, C. J. Fidge, A.H.M. ter Hofstede og Wil M.P. van der Aalst. *Business Process Simulation for Operational Decision Support*. Springer-Verlag 2008, LNCS 4928, pp. 66-77., 2007.