Data-based description of process performance in end-to-end order processing

Article *in* CIRP Annals - Manufacturing Technology · May 2020

DOI: 10.1016/j.cirp.2020.03.013

4 authors:

Günther Schuh
RWTH Aachen University
948 PUBLICATIONS **5,336** CITATIONS

Andreas Gützlaff
RWTH Aachen University
24 PUBLICATIONS **10** CITATIONS

Seth Schmitz
RWTH Aachen University
5 PUBLICATIONS **0** CITATIONS

Wil Van der Aalst
RWTH Aachen University
1,248 PUBLICATIONS **69,695** CITATIONS

Contents lists available at SciVerse ScienceDirect

CIRP Annals Manufacturing Technology

Journal homepage: www.elsevier.com/locate/cirp

Data-based description of process performance in end-to-end order processing

Günther Schuh[1] (1), Andreas Gützlaff[1], Seth Schmitz[1], Wil M.P. van der Aalst[2,3]

[1]*Laboratory for Machine Tools and Production Engineering (WZL) at RWTH Aachen University, Campus-Boulevard 30, 52074 Aachen, Germany*
[2]*Chair of Process and Data Science at RWTH Aachen University, Ahornstraße 55, 52074 Aachen, Germany*
[3]*Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*

To master ongoing market competitiveness, manufacturing companies try to increase process efficiency through process improvements. Mapping the end-to-end order processing is particularly important, as one needs to consider all order-fulfilling core processes to evaluate process performance. However, today's traditional process mapping methods such as workshops are subjective and time-consuming. Therefore, process improvements are based on gut feeling rather than facts, leading to high failure probabilities. This paper presents a process mining approach that provides data-based description of process performance in order processing and thus objectively and effortlessly maps as-is end-to-end processes. The approach is validated with an industrial case study.

Process, Performance, Machine learning, Process mining

1. Introduction

Today, manufacturing companies try to master the ongoing market competitiveness by increasing process efficiency. Process efficiency measures how economical processes are executed and is quantified by the process performance indicators (PPIs) process time, process cost and process quality [1]. To increase process performance (PP) through process improvement or process re-engineering, PP is first described. Thereby, PP is defined as how well the process, which consists of sub-processes and activities, operates to achieve its objectives and describing PP includes the mapping of as-is processes as well as the description of PPIs [2]. In general, an accurate mapping of as-is processes is required to derive lasting potentials for PP improvements and one of the biggest challenges for producing companies [1,3].

To describe the performance of the end-to-end order processing (ETEOP) process is particularly important for producing companies to ensure sustainable competitive advantages. The ETEOP process comprises all technical-operative core processes of a company, which are domain-specific business units such as sales or manufacturing, to complete customer orders (CO) in due time [4]. However, up to 96% of process mapping methods are applied in manufacturing, taking predecessor processes such as design, or successor processes such as assembly not into consideration, although they cover up to 70% of end-to-end process time [5].

Traditional methods can be used for ETEOP process mapping but only with limitations: First, mapping the process and its activities based on workshops or interviews depends highly on participants' assumptions and abstractions. Further, the as-is process mapping is time-consuming and frequently reported as the most costly stage. Thirdly, since traditional methods are paper-based, the ability to capture dynamics is limited [6]. Thus, the shortcomings are time-consuming, unsubstantiated, subjective and static process descriptions that lead to failure probabilities of up to 70% [7].

By contrast, studies since the 1998s show that the use of event logs, which are process feedback data already available in companies' information systems, improve process mapping. Studies show that process mining (PM) methods using event logs are superior to traditional methods by effortless, fact-based, objective and dynamic process mapping and thus address the aforementioned drawbacks [6,8]. However, producing companies still map ETEOP by traditional methods or applied PM just in administrative or partial processes, as no methodology for PM exists that addresses the characteristics of ETEOP. Therefore, this paper presents a methodology to describe PP in ETEOP by PM and thus supports companies by viable process improvements.

The remainder of this paper is structured as follows: Section 2 highlights the importance of PM. In Section 3, a comprehensive literature review on the state of the art is presented and the challenges of using PM in ETEOP are described. Section 4 describes the methodology, which is applied in a German SME in Section 5.

2. Importance of process mining for order processing

Today, companies understand processes usually as static, order-independent and trivial. However, real industry data shows that order numbers vary within order processing and order-related activities are too manifold to be manually mapped, wherefore real processes are not entirely known (see for instance Table 1). Further, as orders follow different sequential and parallel activities, which are executed at different times, the challenges increase [9]. Thus, an approach for the data-based mapping of ETEOP constitutes a key step to master process complexity.

PM aims to discover, monitor and control real business processes (not assumed processes) by extracting knowledge from event logs. PM must be distinguished from common data mining and machine learning disciplines by its process perspective. This paper focuses on process discovery (one type of PM) to transform data into a process model, which is a representative visualization of real processes and activities. Inputs are event log data that are collections of related events. Each event refers to at least an activity and a unique process instance, for example orders in the given context [8]. Process discovery combined with replaying event data on the process model provides a proven technology to detect bottlenecks and is especially useful in complex processes [10]. Yet, several challenges for an application in ETEOP exist, which are discussed in the following section.

Table 1
Example: Domain-specific orders and activities of a German SME

Core process domains	O	A	R
Sales	400	27	66%
Assembly	1852	391	11%
Logistics	2684	69	8%

O: No. of data-based orders; A: No. of data-based activities; R: Ratio.
$R=\frac{\text{No. of documented activities based on interviews}}{A}$.

3. State of the art

Describing PP by PM has been researched in the last 19 years. Several papers have been published, of which the most 12 relevant are presented in the following. Especially for the production domain, applying PM in real manufacturing case studies has increased significantly in the last four years [10]. Describing PPIs such as process time with PM shows that a PP improvement of up to 69.97% can be achieved [11]. However, several reasons for intensified research efforts exist for ETEOP.

In eight of the papers, PM was applied in manufacturing processes. In other papers, PM was used to describe upstream processes, such as production planning [12], or downstream processes, such as logistics [6]. Due to interactions between core process activities, however, the description of PP in partial processes is not sufficient, as the sum of the optimized partial processes does not result in an optimum for ETEOP. Rather, mapping partial processes have negative effects on ETEOP PP improvements [13].

Two papers describe PP in the administrative processes of producing companies [14,15]. However, administrative processes are characterized by consistent feedback data and consistent order-IDs [16]. As the success of PM being highly dependent on order-IDs and in ETEOP multiple order-IDs are stored in different information systems, it becomes obvious that the approaches are not suitable [8]. Instead, event correlation, which is the process to link events that belong to the same process instance but are scattered across various information systems, must be considered, that does not yet exist in industrial practice. It is indicated to be an essential step to enable end-to-end process discovery [16].

In four papers, PPIs are added as additional attributes of events in the event log [6,8,11,12]. However, it is not easy to deal with many additional attributes. Hence, current research deals with reducing mapping complexity by fewer attributes to make process models interpretable and PM applicable for industries [3]. Therefore, data should be included based on the desired PP information and a minimum viable dataset for PM in ETEOP need to be defined.

Lastly, some papers use process knowledge for model validation that consequently affects the advantage of objective process mapping [9,15,17–19]. By contrast, process knowledge for event log preparation is scarcely used, although producing industries are often challenged with semi-structured processes that require process expertise for data pre-processing [20–22]. In semi-structured processes, activities vary according to order processing, wherefore filtering of incomplete activities and grouping of rare activities is required to reduce process complexity [23]. Literature emphasizes the importance of log pre-processing for the successful application of PM in a production environment [24]. Therefore, in the following section, a new methodology is presented, which overcomes the aforementioned shortcomings of PP-related dataset pre-processing as well as dealing with typical ETEOP process characteristics, such as multiple converging and diverging order-IDs, to address the research deficit of PM applications in ETEOP.

4. Methodology for a data-based description of ETEOP

The application of PM requires an appropriate methodology that is tailored to the objectives of the application case [15]. The methodology for the data-based description of PP in ETEOP consists of three major steps that can be gone through iteratively (see Figure 1). As no standard exist, the L* life-cycle model [8] is adopted for the methodology. In the first step, the datasets for the ETEOP process are defined and event logs are extracted for each core process domain. The second step combines the multiple event logs of ETEOP, considering converging and diverging order relations. Thus, the ETEOP process is mapped by PM and the PP is described for different process levels in the third step.

The overall goal of the methodology is to apply PM to technical-operating end-to-end processes to ensure an objective and fact-based foundation for process improvements.

4.1 Extracting process-performance-related event logs

The goal of the first phase, the definition and extraction of event logs, is to initialize the datasets that are required to describe PP in ETEOP and to provide them in a pre-processed form. The underlying hypothesis is that transferring tacit process knowledge of business experts about the ETEOP is important for domain-specific event log extraction [25]. Therefore, the ETEOP process domains must initially be identified by using generic business process frameworks and integrating business experts [2].

The required data are derived for the event logs based on mutual dependencies between information and data [26]. Process time, which is the sum of the execution times of activities and the transition times in between, is the highest prioritized PPI and requires activity timestamps as data. Thereby, it depends on the core process domain and its information system whether start and/or end timestamps are logged [8]. We assume that process costs can be determined from processing times and therefore requires no additional data. Individual process times are multiplied by cost drivers after process mapping. Lastly, process quality is calculated based on the process model as the ratio of orders with process loops and total completed orders [21]. Thus, order-IDs as process instances, related activities as well as their start and/or end timestamps are consolidated for each core

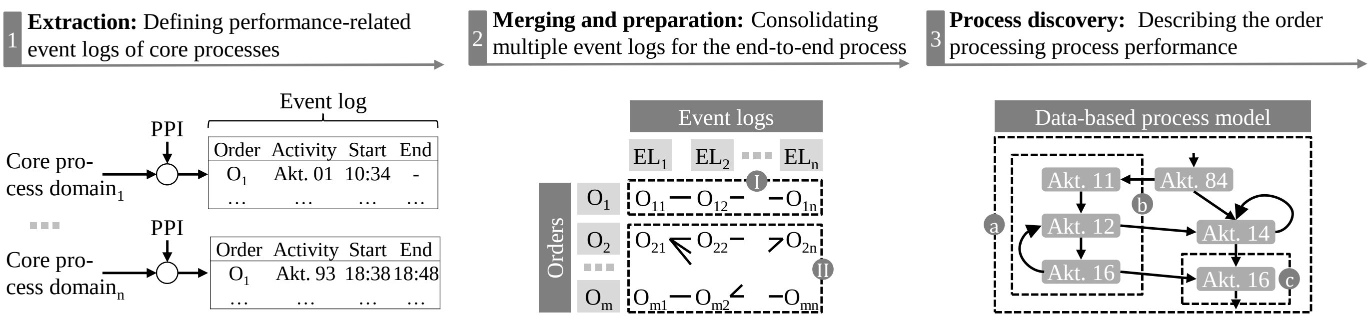

Figure 1. Methodology for data-based description of process performance in end-to-end order processing.

process into a two-dimensional, column-structured table as event log data (see step 1 of Figure 1).

The timeframe to extract the right amount of data must be contextually estimated. Thus, a representative number of high runner orders is defined that covers 80% of ETEOP orders. The timeframe is approximated by the sum of expected processing time in each core process multiplied by the number of sequential-executed orders and a factor F that covers orders occasionally need F-times longer than the expected timeframe of the domain. Having this information, event logs are extracted for each core process domain whose steps have been widely explored and are thus not detailed here [8].

Erroneous process- and order-specific data in the event log lead to unreliable process results. Therefore, data cleaning is needed. A hierarchical filtering approach is used comprising order and process filtration (see section 4.2). In the order filtration, customer-anonymous orders, for example stock replenishment orders, are removed as they distort the ETEOP process. As further data cleaning, such as additional filtration to remove duplicated orders, to increase data integrity is not ETEOP-specific and already broadly researched, it is not detailed here [27].

4.2. Merging event logs for end-to-end order processing process

Companies often log orders in different information systems that need to be merged. Customer relationship management (CRM) systems might log CO, whereas enterprise resource planning (ERP) systems and/or manufacturing execution system (MES) log manufacturing and assembly orders. Further, one-to-one (perspective I in step 2 of Figure 1) or many-to-many relationships (perspective II) can exist between the orders in ETEOP. By example, one CO generates one manufacturing order or three manufacturing orders generate two assembly orders. The underlying hypothesis is that a data relationship, for instance by reference number, exists between the multiple order-IDs. This is not considered in the previous PM approaches, as the partial processes focused so far have unique order-IDs (see section 3). To merge event logs, the method of object-centric event log correlation developed in Ref. [16] is used. Thereby, events of objects, which are domain-specific orders in the given context, are linked through object paths by existing relationships between databases. By doing so, multiple event logs are merged into one single event log, which is referred to as a Minimum Viable Dataset (MVD), containing one constituent order-ID (1), ETEOP activities (2), start (3) and/or end timestamps (4) as event log data in the above mentioned two-dimensional structure.

The timestamp format varies across ETEOP and needs to be unified for PM. In general, started and completed manufacturing orders are recorded by seconds. To avoid loss of information, all timestamps of ETEOP are transformed in the format 'DD.MM.YYY hh:mm:ss'. Therefore, activities having either start or end timestamps are enriched deterministically depending on whether activities' timeframe is a period or a moment. For periods, missing timestamps are enriched by predecessor/ successor activities. For moments, both timestamps of the activity are equated. A further process filtration removes orders which are not both started and completed within the extracted timeframe to reduce incomplete processes. By doing so, valuable event log information is retained to increase the process model reliability and performance-related aspects of the process [21].

In practice, increasing product variants lead to unstructured processes that make the mapped ETEOP process exceedingly difficult to understand [24]. Therefore, a two-step approach is applied that comprises order and process clustering. Order clustering uses incremental clustering to analyse process variants. Incremental clustering creates clusters of orders according to their similarity of processing wherefore a percentage threshold is set by the user [24]. For process perspective, processes are clustered using pattern abstractions that simplify processes by grouping directly successive activities on the desired level of granularity [28]. As a result, the ETEOP process with less causal dependencies is structured to increase the interpretability of the mapped process and to avoid an overfitting process model.

4.3. Describing end-to-end order processing performance

The PP of ETEOP can be described in two sub-steps. First, the process model is mapped using an adequate PM discovery algorithm. Afterward, PP can be calculated based on the process model for the end-to-end process (perspective a in step 3 of Figure 1), sub-processes (perspective b) and activities (perspective c).

A plethora of discovery algorithms is available and none dominates all others in every situation. The selection of a suitable discovery algorithm depends on the requirements and data [29]. For the purposes of the methodology, the inductive miner is chosen due to its robustness and formal guarantees [8,29].

In the methodology, the PP description contains two performance assessments. First, the PP is expressed by PPIs, but additional PP effects are derived from the process model, for instance loops or process interfaces. Hence, process time that can be later multiplied by costs, process model and process quality are described as PP. Further, the objective is to describe them in a valuation-neutral manner, which means without subjective identification of process weaknesses, in order to initiate well-founded process improvement measures by process experts. However, the visualization is tool-based and should be user-friendly to intuitively spot anomalies within ETEOP.

5. Application and key findings

The described methodology has been applied to a real industry case in the small series machinery sector with sales, assembly, and logistics as identified core process domains. The previously mentioned Table 1 describes the domain-specific event log characteristics of this industry case regarding the number of orders and activities for a specific timeframe. In the following, the open-source PM tool 'ProM 6.8' is used for the application. After the event logs from ERP-system, MES and CRM-system have been merged via the product serial number, a first log inspection shows 276 cases, 53502 events and 876 activities after order filtration. A further event log preparation removes 8 cases that do not both start with either 'order placement' or 'order request' and end with 'product dispatched' by using the 'filter log using simple heuristic' plug-in as process filtration. A subsequent pattern abstraction shows that several assembly activities at one assembly station can be grouped to a superior assembly activity wherefore 90 activities can be clustered to 14 activities. As a result, the final MVD contains 268 cases, 48476 events and 850 activities that is then used for PM.

Figure 2 shows the process model of ETEOP using the 'Inductive visual Miner' plug-in as the selected discovery algorithm with an activity threshold of 0.31. This implies that all events corresponding to activities that occur more than 0.31 times than the most frequent activity, remain in the MVD. Additionally, a path threshold of 0.4 was used, i.e. 60% noise filtering. The higher the thresholds are set, the more exceptions in the process behaviour of the orders are mapped, which leads to a more unstructured process model. The process model visualizes the sequential activities of sales, assembly and logistics as well as parallel activities, for instance sub-assemblies, and process inefficiencies such as process loops. Internally, the model discovered by the inductive miner is converted to a Petri net and the event data are aligned with this model to show frequencies and times, for instance execution and transition times, to describe PP. The resulting PP describes an average end-to-end process time of

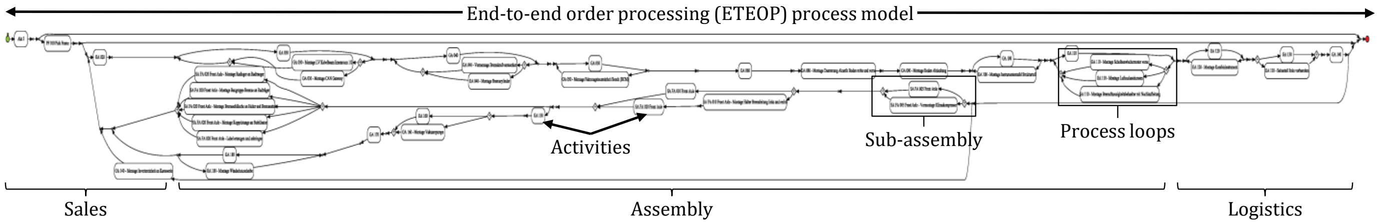

Figure 2. Data-based ETEOP process model by a process discovery inductive miner (extract).

25.67 days, from which more than three-quarter is spent in sales, and a process quality of 3%, which indicates that 260 orders have at least one unplanned process loop.

The end-to-end process model is compared to the PM process model of the assembly with an activity and path threshold of 1.0. The described ETEOP shows 12.34 longer process time by 680 more events within the core process domains sales and logistics. The results emphasize the high potential of the presented approach and promise significant contribution to increase PP when focusing ETEOP instead of partial processes (see section 1). Additionally, while it was previously not possible to describe PP due to process complexity, the approach allows full process transparency compared to traditional process mapping (see section 2).

6. Conclusion and further research

In this paper, a methodology for a data-based description of ETEOP process performance has been presented. The methodology is structured in the three steps (1) extraction of performance-related event logs, (2) merging and preparation of multiple event logs and (3) process discovery for PP description. Its innovation is the application of PM in end-to-end core processes by merging multiple domain-specific event logs. With this, PP in ETEOP can be objectively and fact-based described to derive appropriate conclusions for lasting process improvement and re-engineering projects whose starting is as-is process mapping [1]. An application of the methodology to a real industry use case is presented that shows the ETEOP process model and describes its PPIs. An investigation of PP description in partial processes compared to end-to-end core processes has been undertaken.

To further hone the methodology, several improvements can be investigated. First, the visualization can be improved for its valuation-neutral and intuitive description of PP by integrating further process analysis requirements such as swim-lanes. Further, different process discovery algorithms can be used in the third step, evaluated and analysed to describe PP. Third, the data-based process mapping can be expanded by participative methods to additionally map hidden activities that are not stored in databases. Lastly, as waste in terms of long process times, high process costs and low process quality is discovered to achieve lean processes, the integration of other lean production rules, such as identifying inventory levels, can be further researched.

Acknowledgments

The authors would like to thank the German Research Foundation DFG for funding this work within the Cluster of Excellence “Internet of Production” (Project ID: 390621612).

References

[1] Dumas M, La Rosa M, Mendling J, Reijers HA (2018) Fundamentals of Business Process Management. Springer Berlin Heidelberg, Berlin, Heidelberg.
[2] Rosenberg A (2011) Applying real-world BPM in an SAP environment. 1st ed. Galileo Press, Bonn.
[3] Malinova M, Mendling J (2013) The Effect of Process Map Design Quality on Process Management Success. ECIS 2013 Completed Research. 160.
[4] Eversheim W, St. Krumm, Heuser T, St. Müller (1993) Process-Oriented Organization of Order Processing – A New Method to Meet Customers Demands. CIRP Annals 42(1):569–71.
[5] Schönsleben P, Weber S, Koenigs S, Duchi A (2017) Different types of cooperation between the R&D and Engineering departments in companies with a design-to-order production environment. CIRP Annals 66(1):405–8.
[6] Knoll D, Reinhart G, Prüglmeier M (2019) Enabling value stream mapping for internal logistics using multidimensional process mining. Expert Systems with Applications 124:130–42.
[7] Park S, Kang YS (2016) A Study of Process Mining-based Business Process Innovation. Procedia Computer Science 91:734–43.
[8] van der Aalst W (2016) Process Mining. Springer Berlin Heidelberg
[9] Popsil M, Mates V, Hruska, Tomas, Bartik, Vladimir (2013) Process Mining in a Manufacturing Company for Predictions and Planning. International Journal on Advances in Software 6(3&4): pp. 283-297.
[10] Garcia CdS, Meincheim A, Faria Junior ER, Dallagassa MR, Sato DMV, Carvalho DR, Santos EAP, Scalabrin EE (2019) Process mining techniques and applications – A systematic mapping study. Exper Systems with Applica 133:260–95.
[11] Park J, Lee D, Zhu J (2014) An integrated approach for ship block manufacturing process performance evaluation: Case from a Korean shipbuilding company. Intern Journal of Production Economics 156:214–22.
[12] ER M, Arsad N, Astuti HM, Kusumawardani RP, Utami RA (2018) Analysis of production planning in a global manufacturing company with process mining. JEIM 31(2):317–37.
[13] Tsutsumi D, Gyulai D, Kovács A, Tipary B, Ueno Y, Nonaka Y, Monostori L (2018) Towards joint optimization of product design, process planning and production planning in multi-product assembly. CIRP Annals 67(1):441–6.
[14] R'bigui H, Cho C (2017) Customer Order Fulfillment Process Analysis with Process Mining. CSAI 2017: pp. 247–252.
[15] Suriadi S, Wynn MT, Ouyang C, ter Hofstede AHM, van Dijk NJ (2013) Understanding Process Behaviours in a Large Insurance Company in Australia: A Case Study. CAiSE 2013: pp. 449–464.
[16] Li G, Medeiros de Carvalho R, van der Aalst WMP (2018) Configurable Event Correlation for Process Discovery from Object-Centric Event Data. IEEE ICWS 2018: pp. 203–210.
[17] Viale P, Frydman C, Pinaton J (2011) New methodology for modeling large scale manufacturing process: Using process mining methods and experts' knowledge. 9th IEEE AICCSA 2011: pp. 84–89.
[18] Park M, Song M, Baek TH, Son S, Ha SJ, Cho SW (2015) Workload and Delay Analysis in Manufacturing Process Using Process Mining. AP-BPM 2015: pp. 138-151
[19] Valencia-Parra A, Ramos-Gutierrez B, Varela-Vaca AJ, Gomez-Lopez MT, Bernal AG (2019) Enabling Process Mining in Aircraft Manufactures: extracting Event Logs and Discovering Processes from Complex Data. Business Process Management Industry Forum 2019. BPM 2019, pp. 166-177
[20] Yang H, Park M, Cho M, Song M, Kim S (2014) A system architecture for manufacturing process analysis based on big data and process mining techniques. 2014 IEEE International Conference on Big Data: pp. 1024–1029.
[21] Rozinat A, Jong ISM de, Gunther CW, van der Aalst WMP (2009) Process Mining Applied to the Test Process of Wafer Scanners in ASML. IEEE Transaction on Systems, Man and Cybernetics. C 39(4):474–9.
[22] Lee S-k, Kim B, Huh M, Cho S, Park S, Lee D (2013) Mining transportation logs for understanding the after-assembly block manufacturing process in the shipbuilding industry. Expert Systems with Applications 40(1): pp. 83–95.
[23] Natschläger C, Kossak F, Lettner C, Geist V, Denkmayr A, Käferböck B (2017) A Practical Approach for Process Mining in Production Processes. ERP Future 2016. Springer International Publishing. pp. 87–95.
[24] Meincheim A, Garcia CdS, Nievola JC, Scalabrin EE (2017) Combining Process Mining with Trace Clustering: Manufacturing Shop Floor Process - An Applied Case. IEEE 29th ICTAI 2017: pp. 498–505.
[25] van Eck ML, Lu X, Leemans SJJ, van der Aalst WMP (2015) PM^2 A Process Mining Project Methodology. CAiSE 2015: pp. 297–313.
[26] Aamodt A, Nygård M (1995) Different roles and mutual dependencies of data, information, and knowledge — An AI perspective on their integration. Data & Knowledge Engineering 16(3):191–222.
[27] Bose RJC, Mans RS, van der Aalst WMP (2013) Wanna improve process mining results? IEEE CIDM 2013: pp. 127–134.
[28] Bose RPJC, Verbeek EHMW, van der Aalst WMP (2014) Discovering Hierarchical Process Models Using ProM. in Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Springer Internat Publishing: pp. 33–48.
[29] Jouck T, Bolt A, Depaire B, Leoni Md, van der Aalst WMP (2018) An Integrated Framework for Process Discovery Algorithm Evaluation. ArXiv 2018: abs/1806.07222