Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 00 (2021) 000–000

www.elsevier.com/locate/procedia

54th CIRP Conference on Manufacturing Systems

A Real-World Application of Process Mining for Data-Driven Analysis of Multi-Level Interlinked Manufacturing Processes

Alexander Birk[a], Yannick Wilhelm[a,b,*], Simon Dreher[a], Christian Flack[a], Peter Reimann[a,c], Christoph Gröger[c]

[a]*Graduate School of Excellence Advanced Manufacturing Engineering - University of Stuttgart, Nobelstr. 12, 70569 Stuttgart, Germany*
[b]*Festo SE & Co. KG, Ruiter Straße 82, 73734 Esslingen, Germany*
[c]*Institute for Parallel and Distributed Systems - University of Stuttgart, Universitätsstr. 38, 70569 Stuttgart, Germany*

* Corresponding author. Tel.: +49-711-347-590-28. *E-mail address:* yannick.wilhelm@gsame.uni-stuttgart.de

Abstract

Process Mining (PM) has huge potential for manufacturing process analysis. However, there is little research on practical applications. We investigate a real-world manufacturing process of pneumatic valves. The manufacturing process comprises interlinked events at the superordinate business process level and at the subordinate machine level, making its analysis based on PM challenging. We show how to integrate heterogeneous data sources and give examples how PM enables a deeper understanding of the manufacturing process, thereby helping to uncover optimization potentials. Furthermore, we discuss challenges in data integration and point out limitations of current PM techniques in manufacturing.

Keywords: Process Mining; Multi-level Interlinked Manufacturing Process; Heterogeneous Data Sources; Data Integration

1. Introduction

In recent years, Process Mining (PM) has become an important data-driven approach to analyze and optimize processes of any kind [1]. PM is usually based on applying data mining and business process management techniques on event data. Continuous process optimizations are crucial especially for industrial companies with cost-intensive manufacturing processes. Applying PM to such processes has huge potential for achieving higher efficiency and effectiveness. Thereby, industrial companies can analyze their manufacturing processes in depth to gain the much needed process transparency based on data [6, 12].

Literature reviews show that the application of PM in the manufacturing industry has found less attention in research compared to other domains, e. g., healthcare [3, 8]. Yet, academia does not entirely ignore the analysis of manufacturing processes using PM. In fact, Dreher et al. [5] give an overview on various applications and case studies for the analysis of manufacturing-related processes. However, almost all studies apply PM exclusively on one data source, e. g., Enterprise Resource Planning (ERP). So, existing studies leave the integration of heterogeneous data sources such as ERP and machinery data open for future research [5, 9]. An integration of various data sources seems promising as this enables both an overall view on and a deep dive into complex processes [7, 9].

Especially real-world complex processes pose several challenges for a successful application of PM [2, 5, 7, 12]. In manufacturing, process complexity is often associated with multi-level interlinked process structures. Such structures are characterized by coupled processes with complex interdependencies on different organizational levels, with parallel and asynchronous sub-process executions, and with different levels of automation. Although such process structures are typical for the manufacturing area, their effects on the whole manufacturing system and the application of PM on them have not yet been sufficiently researched [2, 5, 6, 12].

We address this research gap by applying PM on a real-world manufacturing process of pneumatic solenoid valves. It comprises events at the superordinate business process level and at the subordinate machine level, making it a typical example for multi-level interlinked manufacturing processes.

In this paper, we (1) present an approach to integrate event data from heterogeneous data sources into one event log. Moreover, we (2) propose a way to represent interdependencies of multi-level interlinked process structures directly within the event log. This enables a holistic analysis of real-world interlinked manufacturing processes and of all involved subprocesses at different organizational levels. Based on this, we (3) discuss how to identify optimization potentials for the considered manufacturing process by combining PM results and available domain and expert knowledge. Lastly, we (4) point out limitations of current PM techniques in manufacturing.

The structure of the paper is as follows: Section 2 introduces the investigated use case. The approach to data integration is described in Section 3, while Section 4 discusses the conducted application of PM on the use case. We evaluate and discuss the PM results in Section 5. Section 6 concludes the paper.

2. Use Case – Manufacturing of Pneumatic Solenoid Valves

The prerequisite for PM is event data, i. e., process-oriented events logged with related timestamps and identifiers. In this paper, we investigate the application of PM to multi-level interlinked manufacturing processes of pneumatic solenoid valves of an industry partner. Figure 1 depicts the target process execution sequence. At its superordinate enterprise control level, the multi-level process structure comprises event data of business processes of the production program planning and the production order processing. We have exported this event data from the company's ERP system in text file format. It includes the events *Production Order Created*, *Order Released*, *Order Printed*, *Order Confirmed*, *Order Delivered* and *Order Last Basic Fn*.

Production program planning organizes the production in an order-oriented way. All events at business process level are characterized by a start and end date as well as by an unique production order identifier. Since the original purpose of the ERP data is not process analysis, some events such as *Order Released* and *Order Printed* have the same date as start and end date. Other events, e. g., *Order Last Basic Fn.* have no time specifications in their start and end date at all.

The events of the manufacturing process at machine level are situated organizationally and chronologically between the events *Order Printed* and *Order Confirmed*. The highly automated and multi-stage assembly line produces the valves in multiple variants. In our analysis, we consider four orders of three different valve variants and a total quantity of 3,678 valves. The production flow of the assembly line comprises 20 steps, six of which are responsible for the quality-relevant properties of the valve and acquire manufacturing process data (hereinafter referred to as P1-P6). Seven fully automated test steps follow to check whether the assembled valves meet quality requirements. The first four test steps (T1-T4) are passed sequentially. The last three steps (T5/1-T5/3) are functionally identical and arranged in parallel to achieve the required cycle time. Thereby, each valve passes through exactly one of them.

The Supervisory Control and Data Acquisition (SCADA) system of the assembly line provides all the machinery data. The data is stored in a relational database system that enables queries in Structured Query Language (SQL) for data access. Unique identifiers assign sensor measurements and timestamps

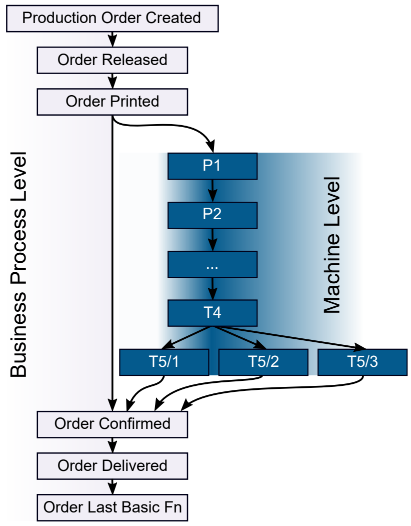

Fig. 1. Target process model of the considered multi-level interlinked manufacturing process.

to the manufactured valves. In addition, every valve identifier is associated with an order identifier from the ERP system.

3. Data and Information Integration

Figure 2 shows the general PM workflow we propose and apply to extract knowledge from process event data. Preparing the data of the process structure in Figure 1, i. e., the blue steps in Figure 2, is one of the main challenges of our work. In fact, the data and information integration steps took up about 80 % of total time on this project. In the following, we present how we conducted the data integration in more detail, paying special attention to the complications that arise from our multi-level interlinked process structure (cf. Section 3.2).

3.1. *Format for Event Logs*

In an event log, each row of data describes exactly one event that represents one execution of a particular process step. Each event involves exactly one instance, i. e., one distinct element passing through the process. For example, each time one piece of product passes a manufacturing process step, an event is created. We integrate the data from the ERP and SCADA systems into the eXtensible Event Stream (XES) [11] data format. The rationale behind this choice is that XES is an open standard XML-based format that offers flexibility through a multitude of extensions. Hence, it is supported by most of the available PM software tools. XES data consists of *events* that are collected in *traces*, where each trace represents an individual instance's path through the process. To get an useful event log for PM, every event needs to contain a timestamp and identifiers for both the instance and the process step that triggered the event.

3.2. *Subordinate Processes and Multi-Instantiation*

The subordinate process steps at machine level (P1 to T5/x in Figure 1) are instantiated multiple times for each execution

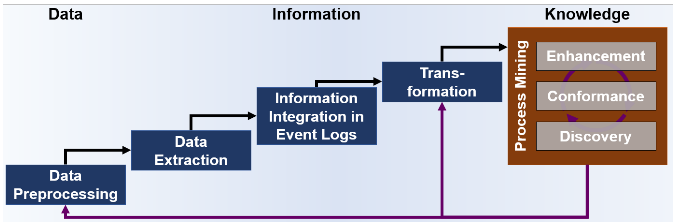

Fig. 2. PM workflow summarizing our methodical approach. Note that this workflow incorporates several iteration loops, indicated by purple arrows.

of the superordinate business process. This is commonly referred to as *multi-instantiation of sub-processes* [13]. Such process structures present a fundamental problem for the event log description: In order to analyze a given process, event logs require each instance to traverse the whole process structure. In a multi-level interlinked process, however, there are fundamentally different, yet interdependent instances on the different levels that do not flow through the same process. We considered three ways to cope with this and to clearly distinguish between instances of process steps at both levels.

The first way is to use the XES *micro* extension, which introduces the concept of sub-processes. However, it does not support multi-instantiation and therefore does not provide multi-level process understanding. The second way is to analyze superordinate and subordinate processes separately. This would go against the original purpose of this investigation as we want to provide a consolidated view on both levels for the whole manufacturing process. As a third alternative, we may analyze both levels side-by-side, linking them through common attributes. This means that the logical connection between both levels is not conveyed to the PM algorithm, but it can at least be presented to the user visually and through appropriate filtering.

We concluded that there is no perfect solution. We chose the third option because it requires the least change in data while still enabling structured analysis. To support this further, we introduce an artificial event to the business process called *Assembly and Test*. This event begins when the first valve of the order enters assembly and ends when the last one leaves testing, thus embedding the production sub-process in the business process.

3.3. Data Integration with Talend Open Studio

We implemented our solution for integrating the process data from the ERP and SCADA systems into XES logs with the software Open Studio for Data Integration by Talend[1]. It is open-source, allows building information integration processes graphically, and offers support for all necessary input and output formats. The individual steps we implemented are as follows (cf. Figure 2):

Data extraction. Accessing the existing company data pools through software interfaces was a challenge mainly due to lacking documentation of data syntax and semantics. This required significant expert knowledge.

Data preprocessing. Removing data points with irrelevant or duplicate information, inserting meta-information or additional artificial events are major steps in data preprocessing. The aim is to unify and clean up data from differently structured sources such that it all matches the event-oriented description.

Information integration into event logs. After uniforming the data, we are able to build the event log itself. Talend Open Studio does not support XES as output natively. However, it can output XML according to a user-defined XML schema, which we hence implemented for the XES event log format.

Data transformation. The transformation step can be carried out via Talend Open Studio or within the PM tools themselves. In this step, we mainly filter the event log for specific classes of traces or outliers to enhance visibility of specific characteristics within the data, but also to increase its quality.

We found that systematic problems in the data set or integration only became visible once we created the first process graph with a PM tool (see Section 4); the source data lacked the structure to notice them beforehand. Therefore, the data integration step should be thought of as iterative, cf. Figure 2.

4. Application of Process Mining

We analyzed the XES logs with the open source PM tool *ProM* [4] since it offers numerous functionalities in a flexible way. We also used the commercial PM tool *Disco*[2] from Fluxicon because it is efficient and user-friendly. The following sections describe the application of these PM tools to the use case. Furthermore, we discuss the achieved results of our PM analysis. The structure of this section follows the typical three phases of PM (cf. right part of Figure 2): (1) Discovery, (2) Conformance Checking and (3) Enhancement of processes [1].

4.1. Process Discovery

A common process model in PM is the process map, which we hence use as a basis to present the major results for our use case [1]. The process map with its associated statistics enables the discovery phase of PM. This phase aims to create an understanding of the actual as-is process structure and how it is executed. Figure 3a shows the discovered process map of the manufacturing process introduced in Section 2. We created this process map with the PM tool Disco. According to van der Aalst [1], Disco uses an advanced further development of the Fuzzy Miner from Günther and van der Aalst [10]. The process map shown in Figure 3a includes all events and traces contained in the XES log of the use case. Events of the superordinate business process level are on the leftmost path after the origin at the top. The events at machine level follow to the right. Directed edges between the events, i. e., between process steps, indicate the chronological process execution sequence. In our work, the number of cases corresponds to the number of production orders or the number of manufactured valves, respectively, which flow through the individual paths of the multi-level interlinked

[1] Talend Open Studio for Data Integration: https://www.talend.com/products/talend-open-studio/

[2] Disco Process Mining Software: https://fluxicon.com/disco/

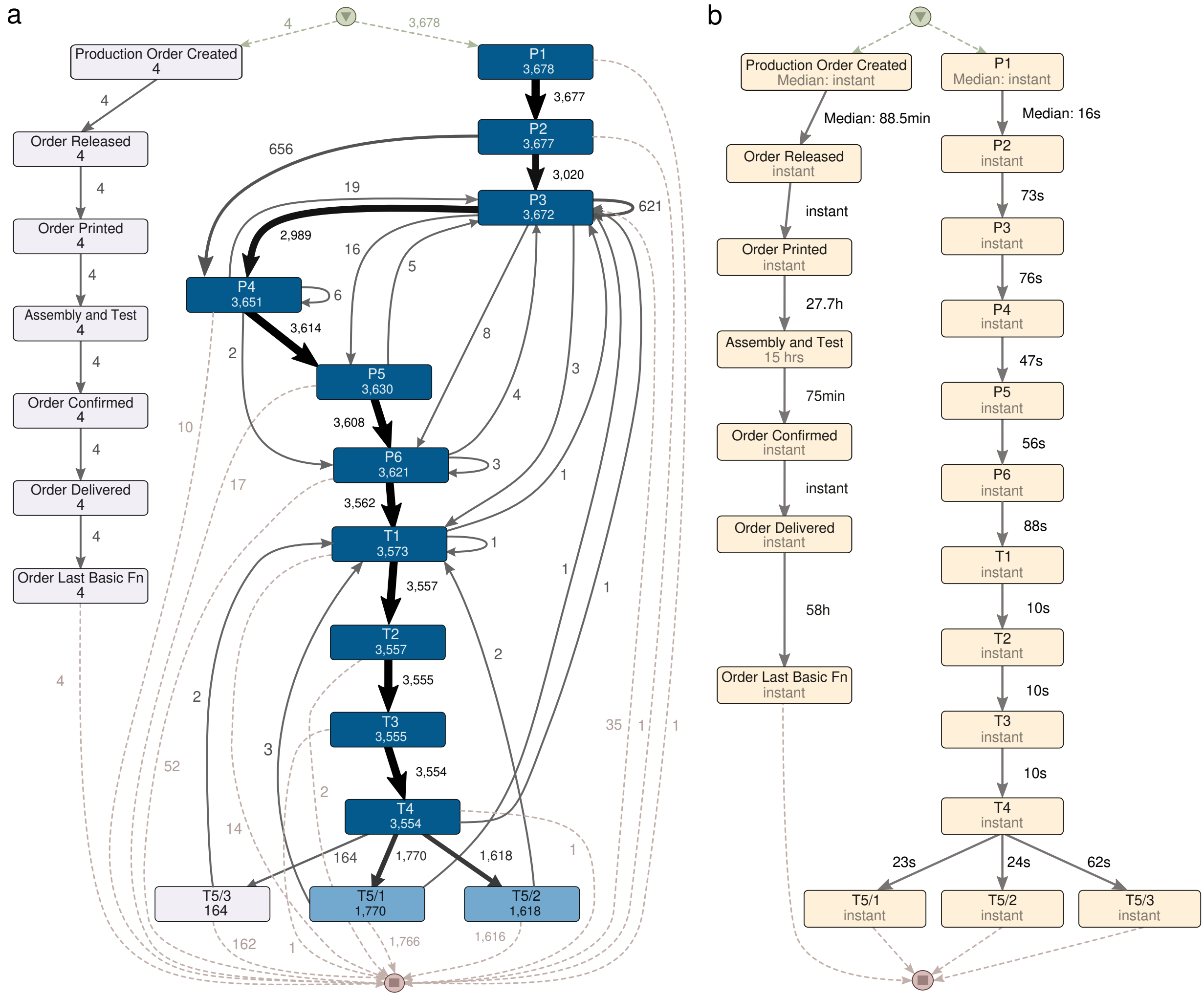

Fig. 3. (a) Process map of the multi-level interlinked manufacturing process as it is actually executed. The map was created with the PM tool Disco. The arcs are labeled with the number of orders or respectively valves passing the process steps. (b) Filtered target process. The process map includes statistics of the process execution. (An artificial bias was added to all statistics for confidentiality reasons.)

process structure. In the following, we discuss the major conspicuous features in the process structure that may be derived from this process map.

Concurrency of interlinked process levels. Due to the limitations of the event log in representing multi-level interlinked process structures, we integrated both process levels into concurrent traces, linking them via attributes representing common order identifiers (cf. Section 3.2). This results in the process map in Figure 3a with the different process levels depicted in parallel. The explicit interlinked structure is thereby visually lost. Nevertheless, it allows for a holistic analysis of the multi-level interlinked process structure. As an alternative, we may investigate the chronological execution sequence on both concurrent process levels. For this purpose, both Disco as well as ProM provide functions to animate the process map with respect to its chronological execution. This way, we can make sure that the processes at machine level are executed as expected, i. e., between the *Order Printed* and *Order Confirmed* events.

Process termination at machine level. The process map helps to discover terminated process execution sequences. These are indicated by paths that go from an early process step directly to the end of the process structure. Such terminated sequences can be identified using the PM tools' filters. There are numerous possible reasons for process terminations that must be verified individually by process experts. One of the main reasons are cases where manufactured valves do not comply with quality requirements of a process step and thus are rejected.

Non-sequential execution of P2-P3-P4. The process map shows a significant amount of 656 valves which skip process step P3. However, we expect that the normal execution of the sub-sequence considered here is P2-P3-P4 (see also Figure 1). The reasons of this irregular process step sequence must be further investigated by process experts.

Loops. Some sub-sequences of process steps such as T1-T2-T3-T4-T5/1-T1-T2-... comprise loops. These indicate repeated process steps or failed process executions. Process experts confirmed that this mainly concerns valves that are rejected as they do not fulfill quality requirements. These valves are then manually repaired and reinserted into the same or previous steps of the assembly line.

4.2. Process Conformance Checking

We conducted the conformance checking with Disco and a rule-based approach [1]. We defined rules for filtering the XES log of the as-is process as shown in Figure 3a, so that the filtered cases are only those that match the target process execution sequence. The rules are derived from a priori process knowledge given by the target process structure according to Figure 1 and by the process experts of our industry partner. Our solution for representing multi-level interlinked processes in XES logs enables us to filter all traces at both process levels by using the common attribute of the order identifier. This way, we obtain the filtered target process (cf. Figure 3b) from the as-is process map, which we then may check holistically for conformance. As we see, the filtered process structure shown in Figure 3b complies with the given target process structure of the process diagram depicted in Figure 1. Furthermore, the execution sequence of Figure 3b represents the majority of all cases, i. e., of the amount of manufactured valves. So, this majority of valves passes the process steps in a chronological sequence that complies with the intended target process structure.

However, the concurrency between the production order workflow at business process level (left path in Figure 3b) and the multiple instances of the sub-processes at machine level (right path in Figure 3b) make an intuitive conformance checking difficult, since the concurrent processes are not shown in an interlinked way. In order to check the conformance of the chronological execution sequence between the superordinate and subordinate process levels, a detailed analysis of single timestamps at both levels is necessary. A more convenient way to do this is a visual analysis. Here, we animate the multi-level process sequence execution of the filtered cases at business process level (i. e., orders) and at machine level (i. e., valves). Both Disco and ProM support this functionality, as mentioned in Section 4.1. This way, we may confirm that the chronological process execution complies with the filtered target process.

4.3. Process Enhancement

A statistical analysis of the filtered as-is process may further uncover optimization potentials to improve the manufacturing processes. Figure 3b shows the median duration of the process steps and their transitions on the superordinate business process level. Note that we modeled *Assembly and Test* as a dedicated step at business process level. Thereby, it has a defined start and end event given by the timestamps of the first and last assembled valve per order (cf. Section 3.2). By analyzing these timestamps, we can calculate the total duration of the actual assembly and test at machine level with respect to the duration of the whole production order processing steps at business process level. Specifically, the process step *Assembly and Test* has a median duration of 15 h. It follows that the actual assembly and test of the valves amount to only 14.5 % (median) of the total time of the production order workflow. This calls for optimizations of the whole business process in order to increase the proportion of time of the value-adding *Assembly and Test* step in relation to total processing time.

Figure 3b also gives an overview of the statistics for the process step transitions at machine level. These statistics refer to the manufacturing time of one valve and thus to one instance of the subordinate process at machine level. We solely focus on median values since the mean values are not robust enough to outliers in the data. Removing and reinserting manually reworked valves into the assembly line leads to large deviations in process execution time. This leads to some outliers that may significantly increase the mean values. Here, the median calculation is more robust due to the large number of valves passing through the process normally. The median values can help to identify the bottleneck of the assembly line. According to the statistical analysis, the longest transition time is located between P6 and T1. This corresponds to the transition between the last assembly and the first test step. The assembly line may be optimized by removing this bottleneck. Furthermore, the transition T4-T5/3 shows a significantly longer duration compared to the functionally equal and parallelized transitions T4-T5/1 and T4-T5/2. This is due to the serial arrangement of the equal test steps in the assembly line, with T5/3 being executed at the end only when the line is at maximum capacity.

5. Evaluation and Discussion

Our results presented in Section 4 show the applicability of PM on multi-level interlinked manufacturing processes: We successfully conducted all three phases of PM on the data of the real-world use case. Conspicuous features of the actual as-is process execution could be identified. We could locate bottlenecks at machine level by analyzing the statistics of the execution times of the manufacturing process. The optimization potentials we have uncovered show that PM represents a viable data-driven approach for process analysis and process enhancement in manufacturing. Manufacturing environments are even predestined for PM analysis due to their event-driven nature leading to a high availability of event data. However, the use of PM techniques in manufacturing is non-trivial and challenging as we found in our work. In the following, we shed light on these challenges and current limitations, but also potential solutions for them. Our main focus there lies on the data integration and representation in XES logs.

The data of the production order workflow at business process level of our use case is originally not intended for process analysis. It contains imprecise or even missing timestamps and has a low resolution of process events representing a process execution of interest. In future, process analysis based on PM should be therefore considered as an use case in designing data acquisition methods.

Considerable effort and lots of process knowledge are required for integrating data from several data sources at different organizational levels into the event log.To cope with this challenge, we propose a universal information integration approach. We have implemented this approach in our work using the Talend Open Studio for Data Integration software. Once the event log is created, the generation of process models based on PM does not require process knowledge or great effort. However, the analysis of these process models requires a lot of a priori expert knowledge. This is also confirmed by the work of Intayoad et al. [12]. For example, in order to filter the process map into the target process execution, we had to leverage knowledge about the parallel execution of T5/1-T5/3 and the cases related to normal execution sequences. This shows that a group of PM and process experts must closely cooperate during PM analysis in order to elicit optimization potential from PM results.

A further limitation of PM techniques is the standardized XES log not allowing adequate representation of multi-level interlinked process structures. PM in its current form requires cases to be tracked and uniquely identified from beginning to end within the process. This requirement cannot be satisfied in a multi-level process containing sub-processes that are instantiated multiple times. Therefore, we had to resort to analyzing the business and machine process levels side-by-side in our use case analysis. We achieved this by integrating both interlinked process levels into concurrent traces that are linked through a common attribute (cf. Section 3.2). In this way, the hierarchy of the multi-level processes is lost, so that involved processes are interpreted by PM as concurrent processes as visualized in Figure 3. It does not explicitly indicate that the process at machine level is a sub-process that is instantiated multiple times at the *Assembly and Test* step of the process at business level.

PM algorithms are also not able to analyze all chronological interdependencies of the interlinked process structure at its execution. Nevertheless, the shared attributes of both concurrent processes in the XES log still allow for analyzing the multi-level interlinked process structure to a certain extent. This is enabled by cross references of cases between both process levels. These cross references help process experts to investigate them with regard to their interlinked structure via common filtering options for both levels. However, these cross references do not represent a concept that PM tools can automatically interpret themselves. Since we have not been able to find a completely satisfactory solution to this problem in literature, we conclude that a fundamental extension to the event log concept is necessary. This extension has to make it possible to represent multi-level interlinked process structures as such within event logs. This ultimately facilitates more holistic PM analyses.

6. Conclusion

Manufacturing processes are often characterized by complex interdependencies between sub-processes at different organizational levels. Such multi-level interlinked process structures must be analyzed holistically to uncover optimization potentials. In this paper, we discuss how to successfully apply PM to real-world multi-level interlinked manufacturing processes of pneumatic solenoid valves. We thereby offer a specialized information integration approach that is able to integrate event data from diverse data sources into the XES log. PM approaches lack appropriate means to represent interdependencies between (sub-)processes within event logs. We hence propose an approach that provides cross references between individual interlinked processes on concurrent process levels. These cross references help experts to interpret the interdependencies between superordinate and subordinate processes. Altogether, we have been able to extract process knowledge and get a deeper understanding about the manufacturing processes of our use case. By leveraging available expert knowledge, we were able to uncover optimization potentials with PM and to improve the manufacturing processes. Future work may aim at developing further solution approaches, e. g., to natively represent multi-level interlinked process structures within event logs used for PM.

Acknowledgements

We thank the German Research Foundation (DFG), the Ministry of Science, Research and Arts of the State of Baden-Württemberg, and the Festo SE & Co. KG for financial support within the Graduate School of advanced Manufacturing Engineering. We also thank the Festo SE & Co. KG for providing the data and supporting this work with their expert knowledge.

References

[1] van der Aalst, W., 2016. Process Mining. Springer Berlin Heidelberg.

[2] Becker, T., Lütjen, M., Porzel, R., 2017. Process Maintenance of Heterogeneous Logistic Systems – A Process Mining Approach. Freitag, Kotzab et al. (Ed.) – Dynamics in Logistics 9, 77–86.

[3] Dakic, D., et al., 2018. Business Process Mining Application: A Literature Review, in: Katalinic, B. (Ed.), Proc. of the 29th DAAAM International Symposium, Vienna, Austria. pp. 866–875.

[4] van Dongen, B.F., et al., 2005. The ProM Framework: A New Era in Process Mining Tool Support, in: Applications and Theory of Petri Nets, Springer Berlin Heidelberg. pp. 444–454.

[5] Dreher, S., Reimann, P., Gröger, C., 2020. Application Fields and Research Gaps of Process Mining in Manufacturing Companies, in: Reussner, R.H., Koziolek, A., Heinrich, R. (Eds.), Proc. of the 50th Jahrestagung der Gesellschaft für Informatik, Karlsruhe, Germany.

[6] Er, M., Astuti, H. M., Wardhani, I. R. K., 2015. Material Movement Analysis for Warehouse Business Process Improvement with Process Mining: A Case Study, in: Bae, J., Suriadi, S., Wen, L. (Ed.), Asia Pacific Business Process Management. Springer International Publishing. volume 219 of *Lecture Notes in Business Information Processing*, pp. 115–127.

[7] Flack, C., et al., 2020. Process Mining in der Produktion: Spezifische Herausforderungen bei der Anwendung. Zeitschrift für wirtschaftlichen Fabrikbetrieb 115, 829–833.

[8] Garcia, C.d.S., et al., 2019. Process Mining Techniques and Applications – A Systematic Mapping Study. Expert Systems with Applications 133, 260–295.

[9] Gröger, C., et al., 2016. A Mobile Dashboard for Analytics-based Information Provisioning on the Shop Floor. International Journal of Computer Integrated Manufacturing 29, 1335–1354.

[10] Günther, C.W., van der Aalst, W.M.P., 2007. Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics, in: Alonso, G., Dadam, P., Rosemann, M. (Eds.), Proc. of the 5th International Conference on Business Process Management, Brisbane, Australia. pp. 328–343.

[11] IEEE Standard, 2016. IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams. IEEE Std 1849-2016 .

[12] Intayoad, W., Becker, T., 2018. Applying Process Mining in Manufacturing and Logistic for Large Transaction Data. Freitag, Kotzab et al. (Ed.) – Dynamics in Logistics 37, 378–388.

[13] Weber, I., et al., 2015. Mining Processes with Multi-Instantiation, in: Proceedings of the 30th Annual ACM Symposium on Applied Computing (SAC), Salamanca, Spain. pp. 1231–1237.