Process Mining in Textile Production: Insights from Penn Textile Solutions

Tobias Brockhoff, Merih Seran Uysal, Anahita Farhang Ghahfarokhi, Leon Reinsch, Thomas Kordtokrax, Andreas Meister, Franz Schütte, Tugsan Vural, Mahsa Pourbafrani, Thomas Gries, and Wil M. P. van der Aalst

Abstract

(a) *Situation faced*: In the textile industry, each individual manufacturing step is typically highly optimized. Nevertheless, inter-manufacturing step dependencies usually have great potential for further optimization. Penn Textile Solutions GmbH collects its manufacturing event data in a dedicated database on the completion of each manufacturing step. However, the data are not used to generate a holistic view of the process. In this study, we apply process mining to leverage the data, generate insights, and visualize a manufacturing production process that is focused on a single machine—the tenter frame.

(b) *Action taken*: We first focused on measuring working hour-aware time intervals (i.e., not considering off days and holidays). Then we split the analysis phase into two major parts—the manufacturing process and the quality control process. We then analyzed both parts using process models as a structuring element. For the manufacturing process analysis, we first conducted a working hour-aware analysis of lead times, process times, and machine utilization. Using a hybrid discovery approach, we discovered a process model at the

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

T. Brockhoff (✉) · M. S. Uysal · A. F. Ghahfarokhi · M. Pourbafrani · W. M. P. van der Aalst
Process and Data Science, RWTH Aachen University, Aachen, Germany
e-mail: brockhoff@pads.rwth-aachen.de; uysal@pads.rwth-aachen.de; mahsa.bafrani@pads.rwth-aachen.de; wvdaalst@pads.rwth-aachen.de

L. Reinsch · T. Gries
Institute for Textile Technology, RWTH Aachen University, Aachen, Germany
e-mail: thomas.gries@ita.rwth-aachen.de

T. Kordtokrax · A. Meister · F. Schütte · T. Vural
Penn Textile Solutions GmbH, Paderborn, Germany
e-mail: t.kordtokrax@penn-ts.com; a.meister@penn-ts.com; f.schuette@penn-ts.com; t.vural@penn-ts.com

J. vom Brocke et al. (eds.), *Business Process Management Cases Vol. 3*,
https://doi.org/10.1007/978-3-031-80793-0_7

machine level, used it to define production stages, and investigated the latter in more detail. Finally, we analyzed the quality process based on a model-induced case classification.

(c) *Results achieved*: We successfully created a process model that described the manufacturing process well. Using this model, we compared cases within and between production stages. While we were able to identify critical stages, the analysis revealed significant variance that was not straightforward to explain, even when taking the impact of COVID-19 into account. The accompanying analyses of resource load, idle times, process time, and lead times were an initial means of making machine utilization accessible. Our results sparked discussions, and by comparing results, we were able to identify the bottlenecks.

(d) *Lessons learned*: Closely integrating stakeholders helped us define realistic, realizable goals. While presenting intermediate results increased trust and understanding in the subsequent stages, it also led to interesting, unexpected discussions. Regarding the application of process mining, we found it helpful to structure the analysis using process models. However, we also recognized the limitations of existing techniques to analyze flexible processes at a detailed level. A major challenge we overcame was how to map the process in a way that took the production context into account so that we could better classify the results.

1 Introduction

The textile industry is one of the oldest industrial sectors, providing work for millions of people worldwide (Annapoorani, 2017). While most of the production has been relocated to Asia, Europe is still the leader in high-quality, special textiles, and digitalization and Industry 4.0 have been identified as key enablers for sustaining its leading role (Kemper et al., 2017). While the individual process steps are highly optimized, a holistic view of the process chain has the potential to identify further opportunities for optimization. Despite this, adoption has been slow (Kemper et al., 2017). In this case study, we investigate the application of process mining techniques to analyze the textile production process at Penn Textile Solutions GmbH (Penn). Penn is a German manufacturer producing innovative textiles to meet the demands of international customers. The production at Penn covers each step between the rolls of yarn and the finished textiles. However, the scope of this work is limited to the treatment of the knitted textile and the final quality control stage.

As described in van der Aalst (2016), traditional process mining is based on *event logs*, that is, on collections of so-called *events*. Each *event* is a recording containing information on three major concepts: A *case notion*—in our case, a textile ID—links events to a business case. An *activity notion*, in our case manufacturing operations, describes what has been done, and a *timestamp*, here the time when a manufacturing operation was performed, provides an order for the events. Following these concepts, a high-level, schematic illustration of the production is depicted in

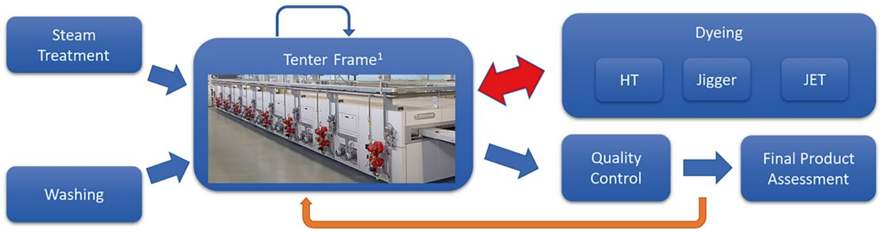


Fig. 1 Schematic illustration of the main production steps at PS. Operations performed at different machines frequently involve processing at the tenter frame (http://www.fi-tech.com/View.aspx?page=textilesproducts/tenterframesandovens). The arrows depict typical process flows. The tenter frame is used repeatedly within each case

Fig. 1. It shows the tenter frame as the central machine involved in various production steps. Two important subprocesses are dyeing, where we can distinguish between different modes, and quality control, where the textiles and colors are tested. Notably, the tenter frame is used multiple times within each case.

We applied process mining to analyze the performance of the production and quality control processes. Moreover, Penn was particularly interested in how the required rework, as determined during quality control, affects the manufacturing process. In the long term, the insight generated can be used to improve the scheduling of orders and reduce setup costs and lead times. As depicted in Fig. 1, a major challenge to answering these questions lies in the tight coupling between machines and activities. The tenter frame is required for multiple production steps for each textile, and these tasks may also differ for different textiles. While considering the low-level activities would result in an overly complex control flow, only considering the machines consolidates the control flow and, consequently, the performance analysis. However, it introduces loops centered on the tenter frame, which need to be considered carefully during the analysis. Furthermore, from a performance perspective, there are strong dependencies between the textile orders because of the shared resources, that is, machines. In our analysis, we need to consider multiple textile orders with large variations in both their manufacturing process and the order of the steps involved.

In the following, we elaborate on the situation faced including how data is currently being used at Penn as well as on the event data collected and the challenges it presented. Afterward, we present a high-level overview of the actions taken and how the project was conducted. The details of each step will be presented along with the results achieved. Finally, we highlight the lessons learned.

2 Situation Faced

At the beginning of the project with Penn, the production department had been collecting data in a dedicated database. The data was used to query fundamental statistics but it was not exploited to generate holistic insight and, ultimately, steer the process. On-site decisions (e.g., scheduling orders) were based on domain knowledge. Therefore, Penn's main interest was exploratory—What can be done with the data?—and geared toward analyzing the production planning. However, the available domain-specific information was not yet sufficient for automating the planning. Therefore, we aimed to evaluate the as-is situation.

Although the system was not designed to support process mining, most of the information required—that is, case identifier, activity concept candidates, and task completion timestamps—was directly available. Only the data on quality control steps and a few additional attributes, such as human-readable activity descriptions, are needed to be extracted from additional tables. A snippet from the composite event data, including the most important concepts used during the project, is depicted in Table 1. The Partie column represents an order and corresponds to the *case ID* in process mining terminology. KoststNr identifies the *cost center (CC)* which corresponds to a machine type (e.g., CC-20 is the tenter frame). The first digit also identifies a specific CC type (e.g., a leading three describes CCs related to dyeing). The CC ID together with the KststLfNr column can be associated with a human-readable task description given in column KstText1. The start timestamp was computed using domain knowledge (e.g., based on the complete timestamp, the textile's length, and the machine speed). Finally, the columns LaborID and SchauID reference the quality control and final product assessment subprocess, respectively.

The full dataset also contained domain-specific attributes, which are often specific to certain production tasks. However, not all machine configurations were logged. Therefore, rather than analyzing domain-specific decisions, we focused on performance. A major challenge for a detailed performance analysis is the control-flow variability in the dataset. Table 2 gives an overview of a few fundamental statistics on the data. The number of distinct activities is high, and there are many variations in how a textile is processed. The most frequent variant covers only 6% of all cases while, together, the five most frequent variants cover 23% of the event log. As a detailed performance analysis requires that all cases are considered, this variability is particularly challenging. Moreover, the data contain many loops involving the tenter frame, which is used multiple times within each individual case. Furthermore, the tenter frame is the most expensive machine; therefore, its load should be high, and the tenter frame was thereby an anticipated bottleneck.

3 Action Taken

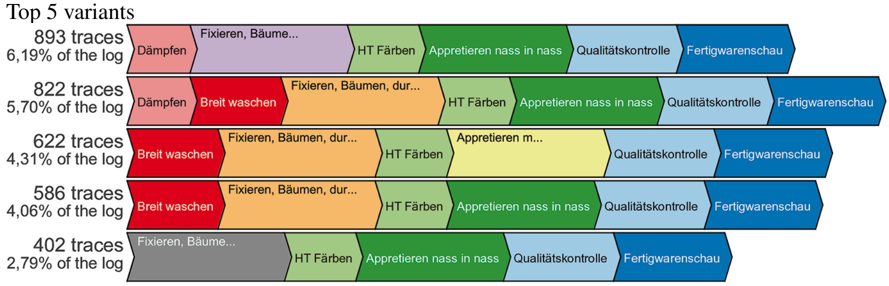
In this use case, we applied process mining techniques to generate insight into the manufacturing data. As indicated in Sect. 2, we focused on the performance dimension where sub-aspects were identified together. Given Penn's interest in an exploratory analysis, details on intermediate steps were also presented and discussed. Figure 2 depicts a detailed illustration of the steps taken. At a high level, we identified three steps: data extraction and preprocessing, analysis of the manufacturing process, and analysis of the quality control process. Each step entailed discussing

Table 1 Reduced snippet of the manufacturing data

Partie	Zeile	KststNr	KststLfNr	LaborID	SchauID	KstText1	Start	Complete
7004	0	10	1			Steaming	DD.MM.YYYY hh:mm	DD.MM.YYYY hh:mm
7165	0	10	1			Steaming	…	…
7165	1	12	1			Washing	…	…
22,426	3	30	1			HT dyeing	…	…
22,426	4	20	17			Finishing	…	…
22,426	5	90	2	394,759		Quality control	…	…
22,426	6	60	1		817,491	Examination	…	…

Table 2 Fundamental event log statistics using task descriptions as activity notion

Statistic	Value
#Cases	14,417
#Events	101,482
#Activities	40
#Variants	2851
Top 5 variants	



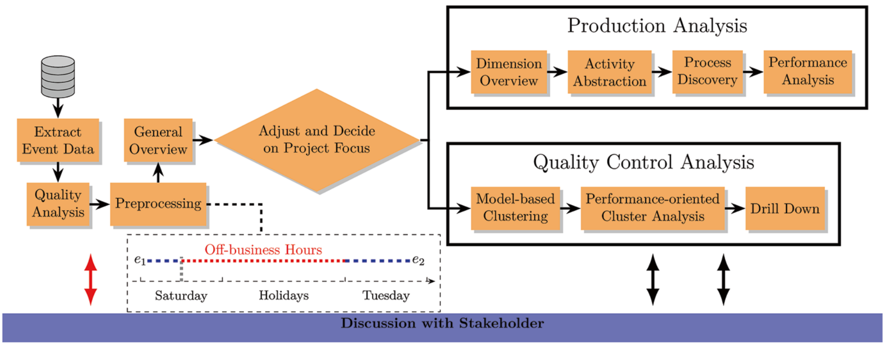
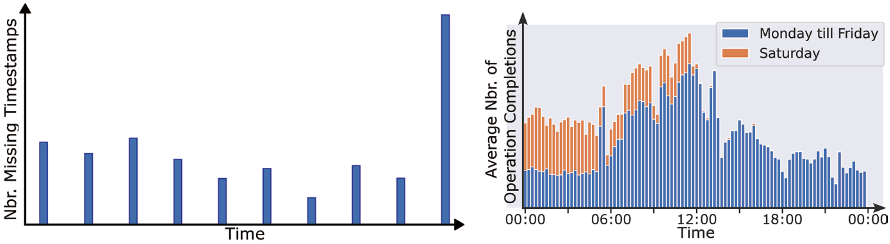


Fig. 2 Overview of the steps taken during the analysis. We distinguish three parts: preprocessing—particularly measuring business hour-aware time intervals—and initial overview, performance-oriented analysis of the production and performance-oriented analysis of the quality control process. The analysis was closely accompanied by discussions with Penn

the results with Penn, and considering the importance of these discussions was one of the lessons learned. In particular, we highlight the discussion of intermediate results during the preprocessing phase.

As depicted in Fig. 2, we first identified the relevant tables based on domain knowledge and extracted an event log as shown in tabular form in Table 1. Then we analyzed the data quality while focusing on the potential for performance analysis and discussed the initial results with the company. The subjects discussed ranged from relatively technical insight (e.g., the distribution of missing timestamps depicted in Fig. 3a) to more relatable statistics, such as the distribution of manufacturing operation completions on a particular day (see Fig. 3b).



(a) Missing Timestamp (b) Daily Work Distributions

Fig. 3 Example of timestamp-related results discussed in the initial project phase: (**a**) The distribution of missing timestamps based on the latest observed timestamp of the associated case. (**b**) The daily average number of completed operations. The capacity of the manufacturing line changes over the day

The steps showed that the data quality was high. Many missing timestamps were associated with cases for which the latest event was recorded close to the end of the extraction window. This can be explained by information not yet being filled in completely. We later decided to drop these events. The daily distribution of timestamps in Fig. 3b reflects that Penn is, on working days, working in three shifts (i.e., 8 h per shift, 24 h per day). On Saturdays, work is only conducted for half of the day. Moreover, it shows that the capacity of the production varies over the day. The activity level at night is lower than during daytime. This makes measuring business hour-aware time intervals challenging. As illustrated in Fig. 2, we accounted for holidays when measuring case durations. However, to effectively evaluate waiting times, the current capacity would need to be considered.

In general, Fig. 3b sparked discussions on when and how work is conducted. However, missing timestamps and high intraday activity-level variance rendered the analysis of the scheduling difficult. In addition, the start timestamps were derived based on the complete timestamps and were, therefore, not precise. For an accurate analysis and automated planning, additional information would be required. Accordingly, the focus was shifted to an investigation and visual assessment of the overall performance of the manufacturing and quality control process. In doing so, we again robustly measured and compared time intervals. As there was no precise information on capacity changes during the day, this was not considered.

3.1 Production Model Discovery

To investigate and measure the performance of specific production stages, we first structured the production using a process model. Based on this process model, we identified specific parts of the process of interest and investigated them using a *model-based performance spectrum (PS)* (van der Aalst et al., 2020). A PS is a visualization of per-case throughput times for user-selected sections of a process model. Each case is represented by a line whose endpoints correspond to the timestamp indicating when it entered or exited that part of the process. Therefore, PS is suitable for analyzing performance on a low level.

We only used the process model to structure the process and analyze its performance. The control flow and potential deviations were mostly uninteresting from a domain expert's perspective. Considering the level of granularity in the event data at hand, the upfront assumption is that control flow decisions made on the shop floor are easily explainable. Valuable insight into control flow would require more detailed consideration of the activities and the incorporation of, for example, machine configurations. The model for the performance analysis was designed as follows:

(a) Specify five activity label candidates on different abstraction levels, specifically task descriptions, aggregated task descriptions (domain knowledge), CCs, CCs grouped by function, and CCs grouped by function and loop unrolling.
(b) Process discovery using the inductive miner algorithm (IM) (Leemans et al., 2013) and a grid search over activity candidates, sub-logs containing the frequent variants covering [25, 50, 75, 85, 100] percent of the log, and IM noise thresholds [0.1, 0.2, 0.4].
(c) Evaluate by investigating fitness, precision, and F1-score.
(d) Manual extension based on the inspection of alignments (Adriansyah et al., 2011).

We tested loop unrolling to better unravel the process around the tenter frame. If the tenter frame is repeatedly used at almost the same position within the process, loop unrolling might better capture the different phases.

3.2 Quality Control Analysis

When analyzing the quality control process, we applied a clustering-based approach to assess the effects of control flow variance on quality control lead times. To this end, two to-be models (according to domain experts) were created that described the expected behavior when quality control and final product assessment were conducted without objection, as well as when rework was required and conducted as expected. We then analyzed lead times after splitting the log into three classes, one for each of these models and one log for the remaining, unfitting cases. The aim was to analyze whether cases that deviate from the well-structured control flow defined by the models exhibit worse lead times. If so, better structuring the quality control and rework workflows would be valuable.

4 Results Achieved

In this section, we discuss and report the results achieved based on the actions taken. Our focus is on the production and quality control analysis. Due to confidentiality, we frequently leave out precise axis tick labels in the examples provided.

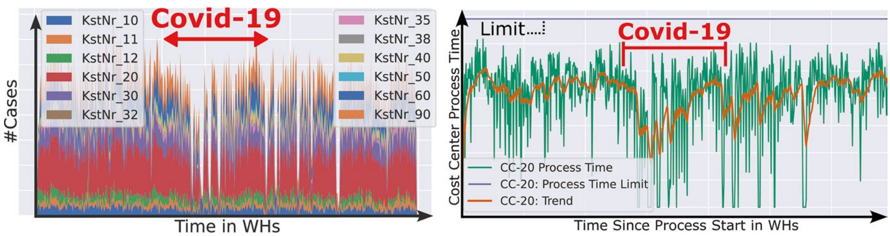
4.1 Production Analysis

The goal of the production analysis was to identify and evaluate the subprocesses and behaviors responsible for the performance variations. We started from the hypothesis that scheduling orders and adapting the schedule of tasks—for example, due to rework—was a main driver for lead time variance. To underpin and refine the hypothesis, we started with a brief analysis of two additional, potentially important factors—namely, machine/resource utilization and process times. By analyzing the machine utilization, we verified that there were no obvious inefficiencies in the scheduling (e.g., machine idle time), and by analyzing the process times, we verified that the impact of product-specific properties on lead times was low.

Two high-level results of this initial analysis are depicted in Figs. 4 and 6. Figure 4a shows the CC load in terms of the number of completed tasks within a sliding window of 48 hours. Note that we applied a working hour-aware sliding window to compensate for Sundays, part-time work on Saturdays, and holidays. The figure confirms that the utilization of critical resources, such as the tenter frame (CC-20, red), is high. The period of high variance in the middle of the figure can be attributed to the effects of COVID-19, while shorter outlier periods were caused by misclassified off days. Figure 4b focuses on the tenter frame and depicts the utilization achieved and the utilization's upper limit. As the data were not available, setup times were not considered.

Other than the period affected by COVID-19, the utilization achieved is relatively stable but still not yet perfect. Finally, the comparison of order lead times to total process times in Fig. 6 shows that processing time contributes little to the overall lead time. Again, we measured working hour-aware time periods with noise sometimes causing short or even negative process times.

As illustrated in Sect. 3, we followed a hybrid approach to discover a process model based on which we could structure the subsequent analysis. We determined that the model based on CCs would be most suitable. Considering the task



(a) Station Load (b) Tenter Frame Utilization

Fig. 4 Working hour-aware overview of CC load and capacity utilization of the tenter frame: (**a**) a stacked plot showing a two-day sliding window of the number of completed tasks. The resource utilization of critical machines (e.g., tenter frame (red) and HT dyeing (purple)) is high. (**b**) The achieved daily tenter frame utilization (turquoise), the trend line (orange), and the theoretical maximum (purple) do not consider setup time descriptions led to large models, even though fitness and precision were decent. Aggregated task descriptions faced the problem that previously distinct activities, which usually occurred at different positions in the process, could not be properly handled by IM. Grouping CCs did not create significant improvement over the CC baseline and straightforward loop unrolling had a negative effect because a few CCs repeatedly occurred at different positions for different tasks. The resulting process model, as well as the tools used in the hybrid model discovery, are depicted in

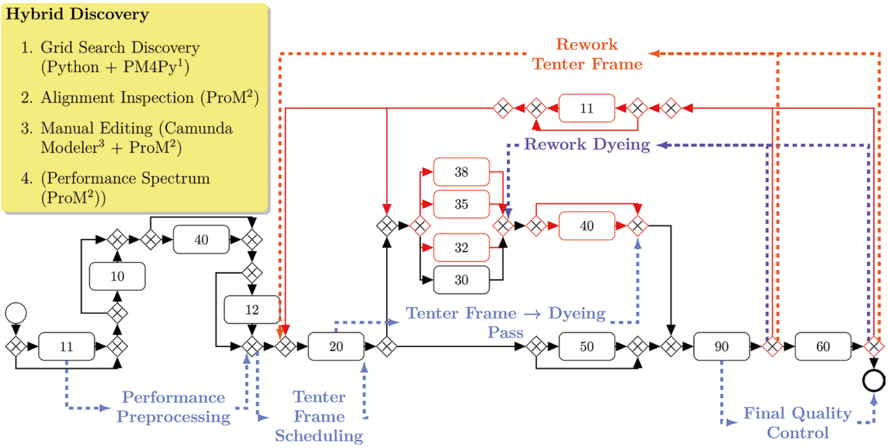


Fig. 5 Tools (PM4Py: https://pm4py.fit.fraunhofer.de/; ProM: https://promtools.org/; Camunda Modeler: https://camunda.com/download/modeler/) and steps of the hybrid discovery approach and the resulting BPMN model. Each task corresponds to a processing step at a CC. The manual extensions of the model are highlighted in red. The dashed lines illustrate the sections of the model that were analyzed in relation to their performance using PS

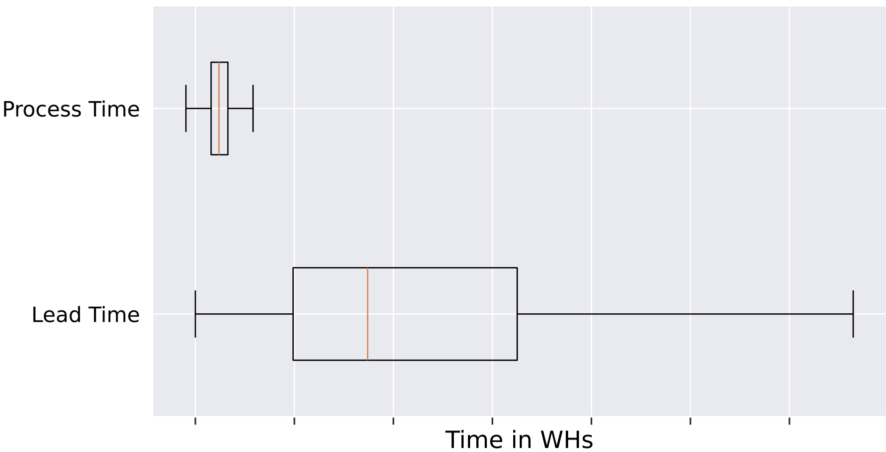


Fig. 6 Comparison of lead times to total per-case process times

Fig. 5. After inspecting deviations in ProM, we added three less frequent dyeing procedures, explicitly modeled the (infrequent) rework, and accounted for CC-11 and CC-40 being used in multiple positions, which IM cannot handle natively. Based on this model, we defined six stages of interest:

(a) Initial processing of the textile before the first run on the tenter frame.
(b) Rework that is directly scheduled for dyeing.
(c) Time until a preprocessed textile is scheduled on the tenter frame.
(d) Rework that directly requires processing on the tenter frame.
(e) A single tenter frame run with subsequent dyeing.
(f) Completion of the final quality control and product assessment stage.

For each stage, we computed a model-based performance spectrum based on the model depicted in Fig. 5. The resulting spectra are depicted in Fig. 7. It shows that any of the defined stages can be associated with high lead times. While the average load of the tenter frame is high (comp. Fig. 4a), Fig. 7a, c suggest that this is partially the result of scheduling many orders. In a significant number of cases, preprocessing is started and completed much later. Moreover, preprocessed cases may need to wait to be processed on the tenter frame. Once an order is scheduled on the tenter frame and for dyeing, this process stage is usually completed without major delays (comp. Fig. 7e). However, there are some exceptions (e.g., sample products). As expected, rework can require a significant amount of time. In particular, rescheduling products on the tenter frame can cause a bottleneck (comp. Fig. 7d). Finally, Fig. 7f shows that finalizing the order and conducting the final product assessment step can have a major impact on lead times.

Our analysis confirmed that there was a high level of resource utilization, resulting in the tenter frame being a main bottleneck. Moreover, by structuring the analysis through a process model, we visualized the impact of scheduling rework on the tenter frame and compared it to the impact of other production stages. The analysis

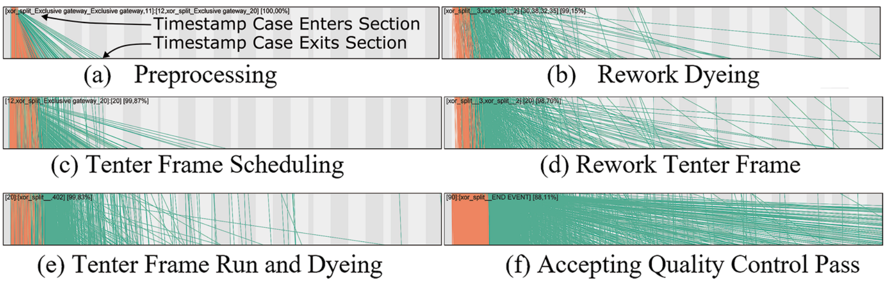
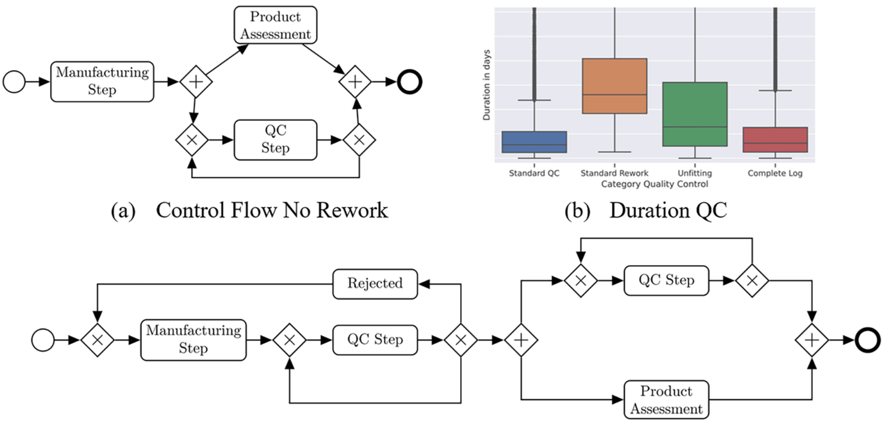


Fig. 7 Performance spectra for the sections defined in Fig. 5. Each spectrum shows the working hours relative to the case's start on the x-axis. The left column shows spectra related to manufacturing sections, while the right column focuses on quality control and rework. The orange lines reflect the 80% of cases with the lowest lead times, while the turquoise lines reflect the 20% with the highest lead times. Important causes for delay were identifed as (**f**) time till customer acceptance and (**d**) rework scheduled on the tenter frame

provided evidence that high resource utilization is partly achieved by having many pending products at the same time. However, a primary driver for long lead times seems to be the finishing stage.

4.2 *Quality Control Analysis*

As quality control and final product assessment can have a big impact on lead times, the final step of the analysis focused on this stage. First, we filtered the event log so that it only contained the last production step preceding quality control and the production steps that occurred during the quality control stage. The resulting log was classified by replaying it on the models depicted in Fig. 8a, c. We then analyzed the lead times for each cluster, focusing on the unfitting cases in particular. An overview of the lead times is depicted in Fig. 8b. As expected, lead times for the final stage were higher for orders that required rework. However, the overall lead time did not increase for the unfitting cases. In fact, allowing for additional flexibility did result in lower lead times than the rework process defined by the model presented in Fig. 8c. At this point in time, we have not completed analyzing the process to identify generalizable recommendations based on this observation. However, drilling down into the final stage lead times revealed that customer response times play an important role and vary among customers. Further details on this are omitted here to maintain confidentiality.



(a) Control Flow No Rework

(b) Duration QC

(c) Control Flow Rework

Fig. 8 Classification of cases according to whether rework is required, the order is finalized without objection, or neither of the former applies. The models used for classification are depicted in (**a**) and (**c**). Note that (**a**) contains a case-dependent optimization of simultaneous, early quality control. The lead times of the final production stages are compared in (**b**)

5 Lessons Learned

In this section, we elaborate on the lessons learned over the life cycle of the project—that is, project setup, event data preprocessing and overview, and in-depth process analysis. As discussed in Sect. 2, with Penn being interested in advancing their digitalization, the initial scope was an exploratory analysis. Even though there were specific questions, such as how to improve the scheduling, the extent to which they could be answered was unclear given the available data. For example, an analysis of the scheduling process might require additional information on when orders would have been available (rather than only seeing them once they were scheduled), precise process times, and setup times and costs. Thus, the initial project was scoped as a larger exploratory analysis showing, relating, and quantifying insight into the performance of the process. Rather than defining potentially unrealistic objectives upfront, we, together with Penn, iteratively defined sub-aspects of interest for the analysis. Given the current results, realistic objectives for the next stage will be better defined. We found that it can also be helpful to sketch intermediate steps and results. For example, the distribution of task completions during the day, shown in Fig. 3b, has been used as a sanity check for timestamps and for defining working hours. While these results can sometimes lead to somewhat unexpected discussions, for example, Fig. 3b sparked a discussion on when certain tasks were performed, intermediate results such as these also helped increase trust (e.g., when the behavior experienced aligns with the analysis) and understanding of subsequent stages. For example, Fig. 3b shows outliers relating to the usual working hours, which can lead to negative working hour-aware timespans. Having previously seen Fig. 3b makes this problem more accessible for stakeholders.

Thereby, the project increased trust and interest in applying process mining to gain new insights. Yet the cautious approach did not result in significant on-sight changes. Rather, it sparked discussions and increased stakeholder awareness of the process. We presume that real change would require the incorporation of additional domain-specific information and constraints. For example, specific production decisions are currently made based on operator experience. Taking the opportunity provided by a follow-up project, which would potentially require less domain knowledge, we discussed broadening the scope. Creating transparency for the integration of procurement, production, and sale was considered an interesting topic for future work. A data-driven approach based on event data could help connect less strongly coupled departments.

During our analysis, we focused, in particular, on model discovery and detailed performance analysis using performance spectra. We used process models to structure the process for the performance analysis. The control flow modeled was, itself, from a domain perspective point of view of relatively little value. In fact, this might generalize to other production processes. In contrast to virtual business cases where, depending on the configuration of the information system, many actions might be enabled at any point in time, in this case study, each production case has a physical state that restricts the next step. Furthermore, it is often impossible to undo an operation; therefore, performing tasks in the incorrect order can irrecoverably destroy the product. Moreover, production operations can have complicated physical effects that are dependent on low-level machine configurations. Medium-level task descriptions are, for instance, insufficient to predict rework. For example, to predict rework based on the model shown in Fig. 6, it is highly likely that factors such as tenter frame configurations, product colors, and textile properties would need to be incorporated. Consequently, the modeled control flow gives little insight into individual cases. In fact, it might be assumed that the order of task executions is valid as it is. Nevertheless, we found the model to be useful for structuring the analysis and for concretely discussing production stages. Finally, we applied automatic model discovery in a grid search to effectively test configurations and activity concepts to come up with an initial model which we then manually edited based on conformance diagnostics.

Regarding the detailed performance analysis, we identified limitations in existing process mining techniques. Orders were scheduled by a domain expert and no obvious inefficiencies, for which no reasonable explanation exists, were found. Assessing and evaluating the performance is difficult as it is context-dependent. At a detailed level, the state of the production depends on the number of available orders, the size of the orders, setup costs related to colors, the type of orders which influence the conducted tasks, and the due dates. While determining the next stages of the analysis based on discussions of intermediate results helped us adapt to the limitations of the data and led to "*interesting*" insights, we encountered difficulties converting specific retrospectively obtained insights into action. Identifying opportunities for optimization in an already optimized process requires consideration of a broader context and many additional implicit constraints.

References

Adriansyah, A., van Dongen, B.F., van der Aalst, W.M.P., 2011. *Conformance checking using cost-based fitness analysis*. 15th international enterprise distributed object computing conference.

Annapoorani, G. S. (2017). Social sustainability in textile industry. In *Sustainability in the textile industry* (Textile science and clothing technology). Springer.

Kemper, M., Gloy, Y. S., & Gries, T. (2017). The future of textile production in high wage countries. *IOP Conference Series: Materials Science and Engineering, 254*(20), 202002.

Leemans, S. J. J., Fahland, D., & van der Aalst, W. M. P. (2013). *Discovering block-structured process models from event logs – A constructive approach*. PETRI NETS.

van der Aalst, W. M. P. (2016). *Process mining – Data science in action* (2nd ed.). Springer.

van der Aalst, W. M. P., Tacke Genannt Unterberg, D., Denisov, V., & Fahland, D. (2020). *Visualizing token flows using interactive performance spectra*. PETRI NETS.

Tobias Brockhoff is a research assistant in the Process and Data Science Group at RWTH Aachen University. He obtained his bachelor's and master's degrees in computer science from RWTH Aachen. His research interests include process mining and focusing on probabilistic approaches for comparative process mining. As a member of the Cluster of Excellence "Internet of Production," he is also interested in applying and developing process mining techniques in production contexts.

Merih Seran Uysal worked as an IT consultant in the industry on numerous projects after completing her computer science diploma at RWTH Aachen University. After a couple of years, she began working as a research associate and doing her PhD at RWTH Aachen University. After completing her PhD, she started working as a group leader at the newly established Chair of Process and Data Science and supervising PhD students. She took full responsibility for managing activities of the Cluster of Excellence project "Internet of Production," acting as the coordinator for industrial projects and being responsible for the acquisition of industrial partners. She is also a lecturer at the Business School of RWTH Aachen. Her research focuses on process mining applied in industry, comparative process mining, process mining of uncertain event data, and process matching and similarity for business process models. She has published 47 peer-reviewed publications and given 26 talks at companies, conferences, and universities.

Anahita Farhang Ghahfarokhi received her bachelors in automation engineering from the Isfahan University of Technology, Isfahan, Iran, in 2015, and a master's degree in industrial management from Tehran University, Iran, in 2019. In 2019, she began working as a research assistant in the project Internet of Production (IoP) at RWTH Aachen University. Her research interests include data science, process mining, and Industry 4.0.

Leon Reinsch worked as a research assistant at the Institut für Textiltechnik of RWTH Aachen University. He obtained his bachelor's and master's degree from RWTH Aachen in Mechanical Engineering and specialized in textile machinery. As a member of the Cluster of Excellence project, "Internet of Production," he participated in interdisciplinary research on applied digitalization. Since 2023, he has been working at Reifenäuser Reicofil as a digital engineer in the R&D department.

Mahsa Pourbafrani is a research assistant in the Data and Process Science Group of RWTH Aachen University. She is a last-year PhD student under Professor van der Aalst's supervision. Her research focuses on forward-looking process mining, which employs data science methods to turn data into actionable insights. The actions are taken using simulation, what-if analysis, and predictions in process mining regarding the performance metrics of processes. She is also a scientist working on the "Internet of Production" project under the Germanys Cluster of Excellence, which aims to combine process mining and machine learning techniques to support operations and decisions in production lines.

Thomas Gries was born in Cologne, Germany, in 1964. He received his Dipl. Ing. degree in Mechanical Engineering and a Dipl. Wirt. Ing. degree from RWTH Aachen University, Germany, in 1989 and 1992, respectively, and a Ph.D. degree from RWTH Aachen University in 1995. From 1995 to 2000, he worked at the Lurgi Zimmer AG, Frankfurt/Main, Germany, as the Head of the Department and Central Division of Technologies for Fibres. In 2001, he became a professor of textile machinery at RWTH Aachen University and the Director of the Institut für Textiltechnik (ITA).Since 2013, he has been an honorary professor with Lomonosov Moscow State University, Moscow, Russian Federation. He has authored and coauthored more than 3000 publications (thereof 1500 scientific papers) and held more than 625 presentations (thereof 140 presentations as keynote or invited speaker), mainly at international congresses.

Wil M.P. van der Aalst is a full professor at RWTH Aachen University, leading the Process and Data Science (PADS) group. He is also the chief scientist at Celonis, a part-time affiliate with the Fraunhofer FIT, and a member of the Board of Governors of Tilburg University. His research interests include process mining, Petri nets, business process management, workflow management, process modeling, and process analysis. van der Aalst has published over 900 articles and books and is typically considered to be one of the top 15 most cited computer scientists with an H-index of 173 and more than 130,000 citations. van der Aalst is an IFIP Fellow, IEEE Fellow, and ACM Fellow and has received honorary degrees from the Moscow Higher School of Economics and Hasselt University.