Process Mining for Production Processes in the Automotive Industry

Merih Seran Uysal[1⋆], Sebastiaan J. van Zelst[1,2], Tobias Brockhoff[1], Anahita Farhang Ghahfarokhi[1], Mahsa Pourbafrani[1], Ruben Schumacher[4], Sebastian Junglas[4], Günther Schuh[3,4], and Wil M. P. van der Aalst[1,2]

[1] Process and Data Science Chair, RWTH Aachen University, Aachen, Germany
[2] Fraunhofer Institute for Applied Information Technology, Sankt Augustin, Germany
[3] Lab. for Machine Tools and Production Engineering, RWTH Aachen University, Aachen, Germany
[4] e.GO Mobile AG, Aachen, Germany
uysal@cs.rwth-aachen.de

Abstract. The increasing digitization of organizations leads to unprecedented amounts of data capturing the behavior of operational processes. On the basis of such data, process mining techniques allow us to obtain a holistic picture of the execution of a company's processes, and their related events. In particular, production companies aiming at reducing the production cycle time and ensuring a high product quality show an increased interest in utilizing process mining in order to identify deviations and bottlenecks in their production processes. In this paper, we present a use case study in which we rigorously investigate how process mining techniques can successfully be applied to real-world data of the car production company e.GO Mobile AG. Furthermore, we present our results facilitating more transparency and valuable insights into the real processes of the company.

Keywords: Process Mining · Internet of Production · Operations Management · Operational Processes · Automotive Industry.

1 Introduction

Process mining [1] is an emerging scientific discipline that allows for extracting knowledge from *event logs*, i.e., collections of historical execution data, available in modern business information systems. Process mining is primarily used to discover, monitor, and improve processes by applying various techniques to event logs generated by the execution of processes [10]. Process discovery, conformance checking, and process enhancement form the three main tasks in process mining, which have been extensively applied to business processes in numerous application fields, such as finance, logistics, and health care. In the literature, very few authors report on the application of process mining in production processes [3,5], e.g., when compared to other application domains such as sales, procurement, banking, and insurance. In addition, existing works mostly focus on use cases at an abstract level [4].

⋆ Corresponding author

Manufacturing companies aim at establishing automated production processes in order to sustain high production quality and decrease the overall costs of their manufacturing processes. Founded in 2015, e.GO Mobile AG is a young manufacturer of cost-effective and customer-oriented electric vehicles for short-distance traffic. The company was primarily established as a semi-automated manufacturer of electric vehicles for which the involvement of the human resources in the production processes comprises the levels of supervisors and operators at the stations of the factory. As such, e.GO adopted an *agile manufacturing model* with Autonomous Guided Vehicles (AGVs) navigating each car through its production process autonomously. Furthermore, every step within the production process is recorded, i.e., creating a basis for a *digital shadow* of the production process [2]. Within the Internet of Production (IoP) researchers from RWTH Aachen University are developing a reference architecture supporting digital shadows in production. In this context, e.GO is a highly relevant use case for IoP.

On the one hand, e.GO typifies the application of the digital shadow of its production processes by elaborately taking technical and organizational changes into account. On the other hand, the fact that the human operators are an essential part of the production line results in various challenges to deal with. Both the notion of a digital shadow and the inherent variability and flexibility caused by manual operations justify the use of process mining. Hence, it is interesting to examine this semi-automated car manufacturer as a use case and derive insights into its real production processes using process mining.

This paper presents a thorough application and examination of process mining techniques on event logs of semi-automated production processes of e.GO where human resources serve as an active, indispensable part of the production processes. Furthermore, we shed light on the arising challenges in operational processes for which process mining can provide tangible results and support by discovering performance and compliance problems. Consequently, the company is able to leverage its data to gather full transparency about how processes are executed. Our paper is structured as follows: Section 2 introduces the business challenges faced and Section 3 shows the solution approach to deal with these challenges. Section 4 discusses the results and benefits from the applied process mining techniques and Section 5 elaborates on the lessons learned.

2 Business Challenge Faced

In this section, we explain the basic manufacturing process at e.GO, the structure of the obtained process data, as well as the main problems and challenges observed.

2.1 As-Is Manufacturing Process

As indicated before, an autonomous manufacturing model is adopted at e.GO. After basic mounting of a to-be produced car's chassis, i.e., including its electronic engine, etc., AGV navigates the car through the production process. Consider Figure 1 for a schematic overview of the general manufacturing process. AGV navigates the car to a number of *General Assembly Stations* (GAS's), i.e., the horizontal blue chevrons in Figure 1 represent these stations (GA0, ..., GA28). In each GAS a specific (sub)-part of

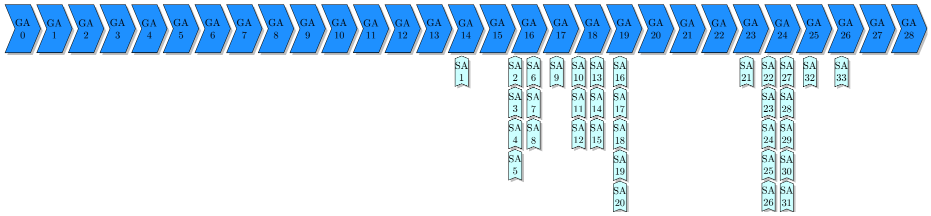

Fig. 1. An overview of the process model generated by the company. The general assembly line consists of *general assembly stations* (e.g., GA0, GA1) and sub-assembly lines include *sub-assembly stations* (e.g., SA1, SA2), which are numbered consecutively.

Table 1. Basic information on the event data used in the case study.

Time range	Two months in 2019
Cars (i.e., cases)	116
Events	12.062
Case attributes	6 (e.g., car ID, color, release version)
Event attributes	10 (e.g., event name, timestamps)

the car assembly is performed, e.g., mounting the doors of the cars. The control-flow of the manufacturing process is strictly sequential, i.e., first GA0 is completed, then GA1, etc. In some cases, preliminary work, i.e., not part of the "general assembly", needs to be performed prior to performing work at a GAS. Such work is executed in a *Sub-Assembly Line* (SAL) comprising *Sub-Assembly Stations* (SAS's), i.e., represented by vertical blue chevrons in Figure 1. Hence, for some GAS, the completion of activities depends on the completion of activities at one or more SAS's.

In its current form, e.GO does not apply any form of *buffering*, i.e., allowing cars to temporarily park somewhere after completion at a GAS. As such, a car awaits completion of the car at the subsequent station in order to proceed.

2.2 Dataset Description

In Table 1, we provide some information on the data which are used in this case study[1]. The event data capture the production of a specific release of the basic low-cost electric car produced by e.GO, i.e., the *e.GO Life*[2]. Within the provided event log, the production of a total of 116 cars is captured. The selected dataset spans a total of 12.062 different events. These events are captured on the GAS/SAL level, i.e., the duration at which the car is present at a GAS, as well as the duration of SAL-based activities is captured. Note that some properties of the car are logged, e.g., the color of a car, as well as some properties specific to the captured events, e.g., which operator was responsible

[1] In agreement with the confidentiality policy of e.GO, we apply anonymization and do not reveal station names, process descriptions, release numbers, and properties of cars and events.

[2] A release of a car is a specific configuration of the car, i.e., in terms of sub-assemblies, software, etc. The data reported here are the data considered after preprocessing, i.e., the data obtained after the removal of earlier car releases.

for a specific production step. During the provided period, the average production rate of cars of the specific release considered is ~ 2.7 cars per working day.[3]

2.3 Problems Faced

While preprocessing the data several challenges were encountered. Here we mention two: *(i) Improper Logging*: Timestamps of entering and leaving stations were automatically recorded by AGVs. However, start and end time of activities at the station level were not always recorded properly by the operators. *(ii) Data Quality Issues*: In the information system, the business analyst rarely modified some data related to the rework operations, which also affected the timestamps of some operations in the production line. Thus, the actual time an event takes place and its timestamp in the log differ from each other. As a result, the order of such events seems to be unreliable in the information system, which results in difficulties in attaining reliable process mining results [1,8].

From the productivity management perspective, the company had further challenges to tackle. First, it was of high interest to reveal which stations and activities often cause a delay in production. Second, they were interested in perceiving whether the observed patterns change over time. Third, the question arose about whether the execution of the sub-assembly lines conform with their normative models. The following section elaborates on the methodology considered to overcome these challenges.

3 Solution Approach

The case study performed together with e.GO followed the general steps of the *process mining project methodology* (PM^2) [7]. The preprocessing concerned the estimation of the true service times, since the automatically recorded end timestamps by AGV include both the service times and waiting times. Thus, assuming that the operator confirms the completion of all tasks at a GAS before starting to wait for the next car, we considered the timestamp of his last operation to estimate the service time. Furthermore, as mentioned before, we filtered the dataset for a specific release `RX`.

For an initial overview of the data, we first visualize production activities over time by using the dotted chart implemented in ProM. Second, for the identification of the bottlenecks, we use PM4Py (open source process mining platform in Python) and Pandas (open source data analysis tool built in Python) to generate the results. Since commercial tools, such as Celonis or Disco, based on directly-follows graph cannot deal with the concurrency, they cannot visualize the production line properly. Therefore, we use a custom visualization to visualize the obtained statistics. Third, we analyze the control-flow behavior of particular GA stations indicating bottlenecks and their linked SA lines by using ProM. Since, as expected, the high-level control-flow of the car production is in line with the reference model (see Figure 1)[4], we do not focus on presenting results of conformance checking applied to GA line. Last, to gain deeper insights in process evolution, we examine the process performance on a weekly basis.

[3] The true production rate is higher, i.e., only cars of a specific release are included, however, at the beginning of the dataset, cars of an earlier release type were still in production, cf. Figure 2.

[4] These results were obtained by applying conformance checking using car production data and the reference model as presented in Figure 1, i.e., solely using the dark blue chevrons.

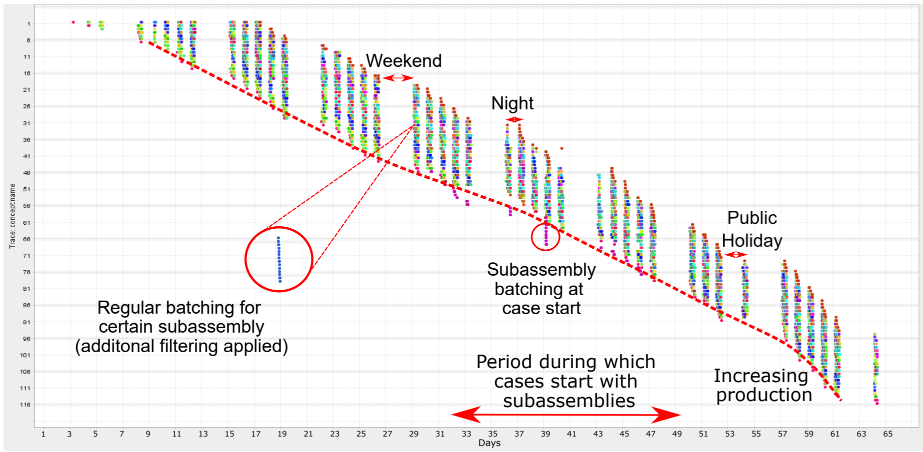

Fig. 2. Dotted Chart Analysis: Visualization of the production of e.GO Life RX over time. Each line on the vertical axis represents a car, each dot represents a production activity performed for a car. Production activities either represent leaving a GA station, or the completion of an SA line. The data is sorted on the first production event logged for a car.

4 Benefits

In this section, we present the main results and benefits of the conducted process mining case study, based on the provided e.GO production data.

4.1 Basic Bottleneck Identification

To gather a basic overview of the production of e.GO Life RX, consider Figure 2, in which we visualize a *Dotted Chart* [9] analysis of the data for which the temporal range spans two months. Despite its rather simple nature, the dotted chart provides various interesting insights w.r.t. the general behavior of the process, as highlighted in the figure. When considering the production activity captured on the first few days of the production (the number of dots per vertical line in the graph), we observe a rather low activity. However, this is explained by the fact that the data only describe the production of e.GO Life RX, i.e., on these days the production of earlier versions is still performed.

Furthermore, we can see typical patterns in the event data. First, we clearly observe weekends and nights in the data, i.e., there is no weekend shift or night shift which continues production. Second, for some sub-assembly lines, we observe *batching behavior* due to improper logging, i.e., a large amount of sub-assemblies are completed and/or executed at the same time. This is illustrated by several similarly colored dots appearing at the same vertical line in the dotted chart. Third, no production activity is recorded on one particular day which coincides with a public holiday in Germany.

To gain better insights in potential bottlenecks, consider Figure 3, in which we present relative performance information on top of the production line. In the figure, the waiting times are visualized in-between stations, using a grey-purple-blue color

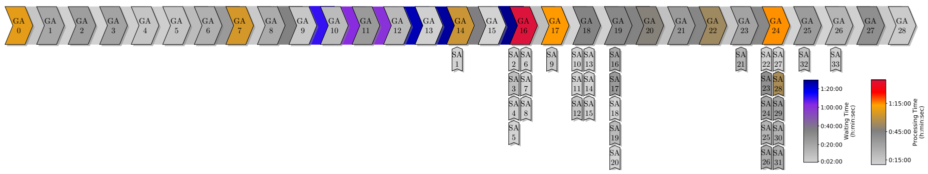

Fig. 3. Visualization of median waiting time (blue color scale) and service time (red color scale) on the GA line and SA stations. The major bottleneck in the process is formed by the general assembly station GA16. Due to the sequential nature of the production process, waiting times, and to some degree service times, of prior stations (GA9,...,GA15) are significantly high.

scale. A light gray color reflects a low waiting time, a purple color reflects an average waiting time, a (dark) blue color reflects a long waiting time. Analogously, the service times of the stations are colored in a gray-orange-red color scale, and are visualized on top of the stations in the production process. Clearly, station *GA 16*, i.e., roughly in the middle of the production process takes the longest. However, note that, this is the first step in the production line which depends on the sub-assembly lines including many SA stations. The idle time, i.e., cars' waiting to proceed to the next station, is highest in-between GA15 and GA16, GA13 and GA14, GA12 and GA13, and GA9 and GA10. Interestingly, the idle time in-between GA14 and GA15 is relatively low, which can be elucidated by the slightly increased service time at station GA14. Finally, note that, out of the 9 general assembly stations that require input from sub-assembly lines, only 4 show a significantly reduced performance. However, all stations depending on sub-assembly lines, perform worse than those without such a dependency.

4.2 Performance Evolution

To better explain the observed performance, we examine the *influx* (aka *arrival rate*) of cars in the production line. Within the dotted chart analysis (Figure 2), the arrival rate is schematically depicted by means of a dashed red line. We observe that an increase in the steepness of said red line corresponds to an increase in influx. Correspondingly, we expect such an increase to be in line with an increased production rate due to no buffering possibilities in the production line. Given the sequential nature of the production line, and, the lack of buffering possibilities, it is likely that the overall production rate, i.e., the average number of produced cars per day, has increased.

Consider Figure 4, in which we depict the process performance for six subsequent weeks in the production ramp-up phase (weeks are presented from top to bottom, e.g., the first week is presented at the top). This further inspection indicates that the production rate almost doubled from $\sim$2.5 cars per day in the first week, to $\sim$4.5 cars per day in the last week. Interestingly, the first week shows a large amount of waiting time and several assembly stations with slow service. In the second week, we observe that certain stations performed better compared to the first week, however, many stations performed worse, and waiting time in-between stations was reduced. In the subsequent weeks, we observe that service times and waiting times were gradually reduced, finally leading to the peak performance observed in the last week captured within the data.

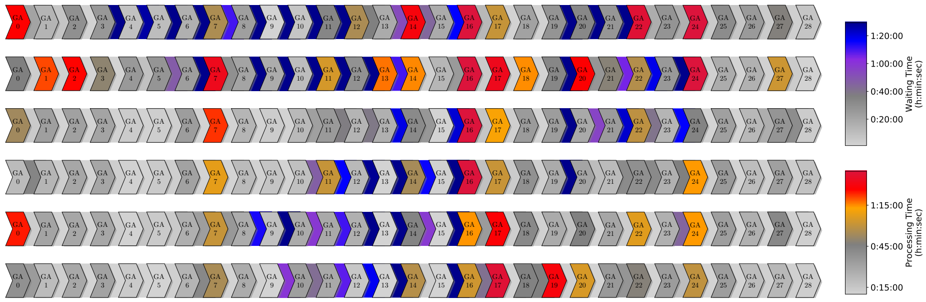

Fig. 4. Visualization of the process performance on a weekly basis, including six consecutive weeks. The most significant waiting times within the production process change over time.

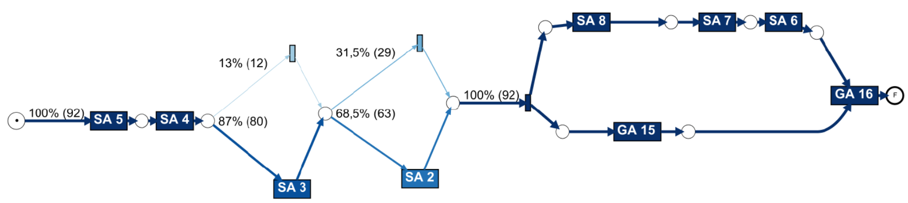

Fig. 5. Discovered control-flow result [6] of the interaction between sub-assembly stations SA2, ..., SA8 and general assembly stations GA15 and GA16. Colors indicate frequency of execution.

4.3 Control-Flow Analysis

Due to the sequential nature of the production process, the application of automated process discovery techniques and/or conformance checking techniques adds little to no value. However, it is worth examining the scheduling of the sub-assembly stations. For example, in Figure 1, we consider the sub-assembly stations which need to be executed, prior to GA16 (i.e., the main bottleneck in the process). There is a clear order between stations SA5, SA4, SA3, SA2, and, also at the same time, between stations SA8, SA7, and SA6. However, these two SA lines are allowed to be executed independently and/or concurrently. We discover a process model for stations SA2, ..., SA8 and GA15 and GA16, as given in Figure 5. Interestingly, the model describes that SA5, SA4, SA3 and SA2 are executed, prior to SA8, SA7 and SA6. Furthermore, we observe that SA stations are indeed executed in the correct order. Note that, in the data, SA3 is skipped in 13% of the cases, and, SA2 is skipped in 31.5% of the cases, which can be attributed to improper logging and data quality issues. Finally, we observe that the full concurrency potential of the sub-assemblies is not achieved within this batch of production.

5 Lessons Learned

We presented a case study using a range of process mining techniques on event logs of a semi-automated automotive production company where human resources conduct assembly sets. In order to attain tangible results and tackle performance and compliance problems, we applied the PM^2 process mining project methodology and analyzed the process execution in the production line comprising general assembly and sub-assembly lines. Furthermore, we observed that automotive production processes are often sequential by nature, allowing for reliable, and, equally important, understandable performance statistics and visualizations.

From both empirical and methodological points of view, an important lesson to learn is that the improvement of the data quality and enabling correct logging will result in more reliable process mining analyses. Consequently, the results will facilitate more transparency and valuable insights into the real processes, which overall serves as a basis for the opportunity to increase the real production rate. Moreover, we will be able to gather more insights into the execution of sub-assembly stations in terms of concurrency, which shows the potential for reducing the production cycle time.

As expected, another important finding emerging during the study is that a bottleneck general assembly station influences the throughput at preceding stations, since no buffering is applied in the production line. In order to circumvent this situation and increase the productivity rate, it is worth considering and simulating bypassing such stations causing substantial delays, which constitutes an interesting direction for future work. Unlike traditional backward-looking approaches, forward-looking techniques, e.g. simulation approaches can be used to answer what-if questions regarding the whole production line and explore various process design alternatives.

Acknowledgments This work is 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.

References

1. W. van der Aalst. *Process Mining: Data Science in Action*. Springer-Verlag, Berlin, 2016.
2. T. Bauernhansl et al. The digital shadow of production - a concept for the effective and efficient information supply in dynamic industrial environments. *Pr. CIRP*, 72:69 – 74, 2018.
3. K. Demtroeder and P. Lechner. Agilitaet durch Process Mining, Manager Magazin, Manager Wissen, June 2019.
4. L. Reinkemeyer. *Process Mining in Action*. Springer-Verlag, Berlin, 2020.
5. P. Lechner. *BMW: Process Mining @ Production*, pages 65–73. Springer, Cham, 2020.
6. F. Mannhardt, M. de Leoni, and H. A. Reijers. The multi-perspective process explorer. In *Proc. of BPM Demo Session co-located with BPM 2015*, pages 130–134, 2015.
7. M.L. van Eck et al. PM^2: A Process Mining Project Methodology. In *CAISE 2015*, volume 9097 of *Lecture Notes in Computer Science*, pages 297–313. Springer-Verlag, Berlin, 2015.
8. R. Andrews et al. Leveraging Data Quality to Better Prepare for Process Mining: An Approach Illustrated Through Analysing Road Trauma Pre-Hospital Retrieval and Transport Processes in Queensland. *Int J Environ Res Public Health*, 16(7):1138, 2019.
9. M. Song and W.M.P. van der Aalst. Supporting Process Mining by Showing Events at a Glance. In *WITS 2007*, pages 139–145, 2007.
10. W. van der Aalst et al. Process mining manifesto. In *Business Process Management Workshops*, pages 169–194, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.