IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 19, 2022, accepted January 26, 2022, date of publication February 16, 2022, date of current version March 8, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3152211

Screening Process Mining and Value Stream Techniques on Industrial Manufacturing Processes: Process Modelling and Bottleneck Analysis

JULIA RUDNITCKAIA[1], HARI SANTHOSH VENKATACHALAM[2], ROLAND ESSMANN[3], TOMÁŠ HRUŠKA[1], AND ARMANDO WALTER COLOMBO[2], (Fellow, IEEE)

[1]Faculty of Information Technology, Brno University of Technology, 612 00 Brno, Czech Republic
[2]Hochschule Emden/Leer, University of Applied Sciences Emden/Leer, 26723 Emden, Germany
[3]Head of Production Intelligence, Honeywell, 49504 Lotte, Germany

Corresponding author: Julia Rudnitckaia (irudnickaia@fit.vutbr.cz)

This work was supported in part by the Brno University of Technology Project "Application of AI Methods to Cyber Security and Control Systems" (2020–2022) under Grant FIT-S-20-6293, and in part by the German Academic Exchange Service (DAAD) through the Deutsch-Argentinisches Hochschulzentrum (CUAA-DAHZ) by the Binational Studies with Double-Degree: Master in Industrial Informatics with Specialization in Industrial Cyber-Physical Systems under Project 57530192.

ABSTRACT One major result of the Industrial Digitalization is the access to a large set of digitalized data and information, i.e. Big Data. The market of analytic tools offers a huge variety of algorithms and software to exploit big datasets. Implementing their advantages into one approach brings better results and empower possibilities for process analysis. Its application in the manufacturing industry requires a high level of effort and remains to be challenging due to product complexity, human-centric processes, and data quality. In this manuscript, the authors combine process mining and value streams methods for analyzing the data from the information management system, applying the approach to the data delivered by one specific manufacturing system. The manufacturing process to be examined is the process of assembling gas meters in the manufacture. This specific and important part of the whole supply-chain process was taken as suitable for the study due to almost full-automated line with data about each process activity of the value-stream in the information system. The paper applies process mining algorithms in discovering a descriptive process model that plays the main role as a basis for further analysis. At the same time, modern techniques of the bottleneck analysis are described, and two new comprehensible methods of bottlenecks detection (TimeLag and Confidence intervals methods), as well as their advantages, will be discussed. Achieved results can be subsequently used for other sources of big data and industrial-compliant Information Management Systems.

INDEX TERMS Bottleneck analysis, manufacturing process, process mining, process modelling, information management system, value stream.

I. INTRODUCTION

In a digitalized eco-system manufacturing companies find themselves possessing a considerable pool of production-system and -process data having a tremendous value which is yet to be unlocked. Leveraging this data by means of new analytics tools offers opportunities to foster data-driven decision-making at all levels of an enterprise to increase both efficiency and effectiveness of existing business processes. The pressure of the constant competition, globalization, and rapid technological change, particularly the penetration of digitalization in the industrial ecosystem [1] have pushed manufacturing companies moving towards primarily performance factors: time, costs, risks, and quality service. All of them are highly interdependent and changing one parameter influences others. Among them, time is a more crucial factor that defines optimization, service level and equipment reliability. Moreover, time plays a central role in selecting

The associate editor coordinating the review of this manuscript and approving it for publication was Vlad Diaconita.

and implementing adequate risk management strategies at different levels (Operational Technology (OT), Information Technology (IT) Levels) of an industrial eco-system, conducting to guaranteeing a minimum set of Key Performance Indicators (KPI) associated to the quality services.

Therefore, all methods proposed in the paper consider time parameter and imply its accuracy and availability in the information system.

Prior knowledge of delays in the production time, cycle time and process behavior are extremely important for manufacturing companies. Since increasing amounts of data are being generated, turning them into useful information becomes a major challenge [2]. Most of the modern tools are able to process big data and to provide appropriate results. However, their integration brings higher efficiency and fill the gaps for each other.

The paper contributes with a novel innovation approach combining Process Mining and Value Stream Techniques for Process Modelling and Bottleneck Analysis in digital manufacturing processes. It provides the foundation of essential steps for process modelling and using it for bottleneck analysis. The goal of the study lies in not only finding bottlenecks on the modelled process, but also in the detection of the bottleneck's roots. The authors go beyond what is currently known and the practices, developing, implementing, and integrating existing methods and tools by using historical data from production-system and -processes, provided by the real industrial information management system, can support both, the value stream analysis as well as the process mining. Note: For the exemplary application, the authors are using data and information provided by the industrial information management system SAP ME.[1] The generic framework can be extended for different production processes and data sources (different information management systems).

This paper is structured as follows: Section 2 (theoretical background) provides the state of the art of manufacturing processes and their analysis in terms of process mining and value stream techniques, followed by Section 3 with tools, existing common methods of bottleneck detection and some definitions for further understanding. Section 4 (proposed methods) presents necessary steps for process modelling based on data provided by an Information Management System, i.e. the SAP ME system, as well as the proposal of two methods for bottleneck analysis. Section 5 shows a use case example of proposed methods implementation followed by a comparison to discussed methods and their evaluation. Finally, Section VI summarizes the contributions of the paper and outlooks possible future prospects and improvements.

II. RELATED WORK

Process Mining offers a promising way to acquire insights into the real dynamics in manufacturing. Despite this method being relatively newly developed, many analysts use it in healthcare systems [3], banking, logistics companies [4], [5]. Manufacturing processes are most suitable for process mining techniques because they usually have a substantial amount of detailed data from sensors recorded in the information system. Also, many operations are defined, structured and not so much human-centric like in other fields. This paper and most other studies investigated data perspectives, while there are studies [6], [7] with different key performance perspective – cost. Other studies like [8]–[10] used some process mining discovering techniques to understand the production line. There was also a study on bottlenecks analysis with process mining techniques [11]. In order to give the readers an actualized view of the current application of process mining techniques in different industrial sectors, similar to the application addressed in this paper, it is recommendable to access [12]. As one can notice, the application of process mining techniques is quite extensive, has many perspectives and options, and is flexible and adaptive. It provides a meaningful basis for supporting further analysis.

Modern technologies can already store large data quantities. Many companies need to require stream data processing to accelerate insights obtaining. Automated pre-processing and stream data mining should be adopted to accuracy, time, memory complexity parameters as well as to concept drifts. The analysis of the value stream, as it is performed in this work, is objective for research and innovation in many different industrial sectors, see for example [13]–[17].

Value streams can be presented as an ongoing data handling, data pre-processing for further analysis that is one of the assumptions of the Industrial Digitalization following the major specifications of the Reference Architecture Model for Industry 4.0 (RAMI 4.0) [18], [19]. It helps to avoid the loss of information throughout the process [20]. Therefore, integration of all steps and their automatization brings insights for companies. Changing input parameters leads to obtaining more dynamic results and information decision markers, which is responding to the real volatile nature of processes.

III. MATERIALS AND METHODS

A. SAP MANUFACTURING EXECUTION (SAP ME)

After analyzing the set of potential production management and control components located at the IT level of the company infrastructure, e.g., SCADA (Supervisory Control and Data Acquisition), MES (Manufacturing Execution Systems), IME (Information Management Systems), CRM (Customer Relationship Management), ERP (Enterprise Resource Planning, SCM (Supply Chain Management), the authors decided to use the Information Management System, specifically the SAP ME, for being used as a source of data associated to the manufacturing process and value stream.

SAP Manufacturing Execution is a plant-centric shop floor system that provides manufacturing execution planning, production monitoring and control, quality control, tracking or traceability. It digitizes manufacturing processes and integrates business systems using a cost-effective, high-quality,

[1] https://me.sap.com/

and resource-efficient methodology based on the Industry 4.0 technology [21], [22].

The main parameters of this system that are used for further algorithms are described as follows:

•The Shop Floor Control (SFC) number is the primary metric of tracking the individual production unit in SAP ME. It is a unique identifier representing a specific instance of a particular material being built during the manufacturing process. In this study, items and SFC number are equal to represent methods, but it is important to notice that they are two different insights in the real manufacturing process.

- Operation – a basic step in the manufacturing process.
- Times:
  - ELAPSED_TIME (ET) - The time of execution of an operation
  - ELAPSED_QUEUE_TIME (EQT) – The waiting time for an operation to commence
  - DATE_TIME – Date and Time entry logged in UTC
- Resource is a machine or piece of equipment that performs work at an operation.
- Work center (WC) is a sequence of resources required to produce or assemble material; a specific location in the site, where inventory is manufactured, processed, or stored.
- Nonconformance Code (NC) – a distinct set of characters that represents a failure, defect, or repair code indicating that SFC number does not meet the approved product quality definition.

Once, analogues of these parameters are found in other information systems, then all methods described in the paper can be applicable and extended.

B. PROCESS MINING FRAMEWORK

Process Mining provides a generic set of techniques to turn event data into valuable insights, a basis for Machine Learning and data mining algorithms, predictions, and recommendations. The three minimum requirements for process mining should be defined: A Case ID (a specific instance of the process), an activity name (one step in the process) and at least one Timestamp column.

Figure 1 presents the workflow of the framework of process mining applied to the Information Management System "SAP ME" in terms of discovery and bottleneck analysis.

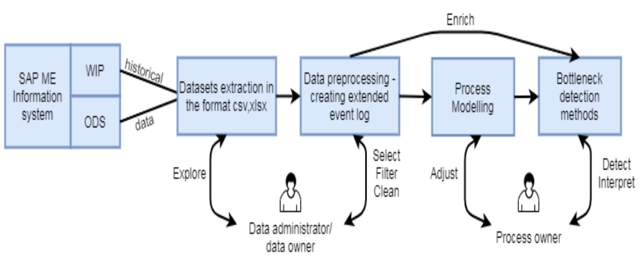


FIGURE 1. Workflow of the Analysis framework with roles and activities.

If instead of historical data, current data are getting from the information management system in Figure 1, then all the following steps can be automatized.

Extracting data from SAP ME will combine datasets from Work-in-Progress (WIP) and Operational Data Store (ODS) parts depending on the structure of the Information System in the company, analysis and required results. The ODS tables provide storage for SAP ME data needed for internal and external reporting. Work in progress refers to the raw materials, labour, and overhead costs incurred for products that are at various stages of the production process.

Dataset's extraction implies exploring and understanding the data structure and includes the next steps [5]:

- Selecting a process instance – Case ID (for the SAP ME, it can be SFC/item, work-center, resource, shop order, etc.)
- Identifying a process activity (operations, steps, resources, etc.)
- Defining or creating additional process attributes (productivity, speed, costs, risks, number of NC, etc.)

According to the recent study [23], industrial data are not flattened. Therefore, an event log should represent different objects and perspectives. For instance, in order to follow the end-to-end process, one needs operations for the whole cycle of production (SFC->item->shop order).

Data processing, in turn, comprises:

- *Preprocessing* data based on the time parameter – choosing an appropriate time frame, calculating waiting times, end timestamps (if they are absent)
- *Filtering* of completed cases (define first and last operations for each process and subprocess)
- *Cleaning data* from noises and incompleteness (delete unnecessary missing values and replace them with a mean /median value or according to production rules)
- *Regrouping/clustering data* to reduce diversity (for work with big data)
- *Converting data* into the MXML or XES formats, which are common formats for event logs

These five steps are crucial for process mining algorithms because each mistake in data leads to incorrect results. Also, the algorithms are not able to work with missing values and just skip them in calculations.

Process Modelling:

- *Discover* the real process model with various algorithms (alpha miner, fuzzy miner, inductive miner, etc.)
- *Use different resulted notations* to avoid their limitations (workflow, Business Process Model and Notation (BPMN), Directly-Follows Graph (DFG)) [24].
- *Explore* main process flow and subprocesses in detail using dotted chats and process variants
- *Considering* performance, resource and time perspectives.

Eventually, it is highly recommended to adjust all obtained results and models with data and process owners to avoid wrong interpretation.

Discovering the real descriptive model is essential for further analysis. There are some reasons describing why existing models don't reflect the real situation in the manufacture:

- *Subjectivity*. Depending on the roles and perspectives, everyone has a subjective picture of the process. Therefore, interviews, workshops and expert knowledge are not enough to discover the real process.

- *Partial view*. Processes are usually complex and include the participation of multiple people, teams, departments and so on. The challenge is that there is no single person, who performs the complete process and know it from the beginning till the end.

- *Change*. Processes are changeable under the influence of different causes (seasons, product changes, weather conditions, covid lockdown, etc.). The real process model should present relevant data, be flexible and actual at the time of analysis.

- *Invisibility*. Some cases can be unfollowed because of system breakdown or data issues. Also, some "black-box" subprocesses might be omitted due to their trivial nature, but they can impact the results of the analysis later.

C. APPROACH FOR COMBINING TWO METHODS

Having provided the foundation for the Process Mining and the Value Stream, the novelty of the contribution of this paper lies in a novel innovation approach combining Process Mining and Value Stream Techniques for Process Modelling and Bottleneck Analysis in digital manufacturing processes. Moreover, since the approach and associated framework are applied to real industrial manufacturing scenarios, mitigation of risks, process optimization and quality services, among other KPIs, can be a better guarantee if not only bottlenecks on the modelled process are found, but the bottleneck's roots also can be detected and identified.

As already highlighted in the introduction, the authors go beyond what is currently known and the practices applied in the industrial manufacturing, developing, implementing, and integrating existing methods and tools that use historical data from production-system and -processes, provided by the real industrial information management system, which able to support both, the value stream analysis as well as the process mining.

Figure 2 shows the application of two branches, each with its methods, a value stream analysis and a process mining, including recommendable tools. The lower branch is dedicated to process modelling and bottleneck analysis including process visualization. The upper branch is dedicated to the analysis and validation of the value stream [25].

Both techniques, Process Mining and Data Stream analysis, start with data extracting from the information management system and data preparation, which includes data filtering, cleaning, processing noises and incomplete cases, and feature selection. This step is time-consuming and needs to be carefully performed because all algorithms are sensitive to input data and the results of the analysis are the most impacted at this point. It follows the implementation of the

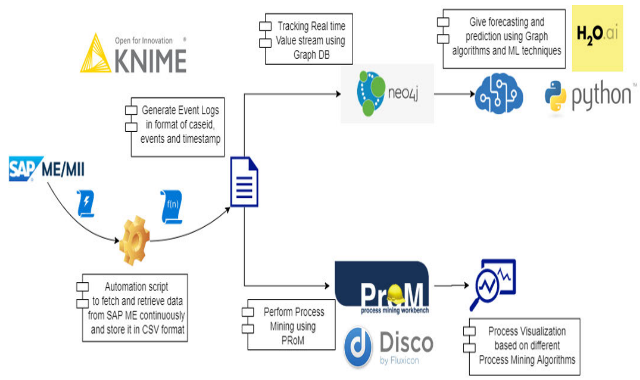


FIGURE 2. **A novel approach for combining Data value stream analysis and Process Mining (including tools).**

two branches: (i) value stream analysis and (ii) process mining to discover the process model, non-confirmed process flows and bottlenecks through the model. It is important to notice, that even the two directions are separated, they influence and improve the results mutually. For example, a process model gives a better understanding of problematic parts in the process as well as process variants for different types of items, and bottleneck analysis, in turn, shows essential features, which can lead to increased cycle time for the whole production process.

The value stream analysis will also involve building a bottleneck identification mechanism that leverages the data prepared from the generic framework and uses bottleneck detection algorithms to identify multiple bottlenecks alongside their root cause. This can be integrated easily into a central data warehouse or data lake for further analysis. As value stream analysis is a continuous approach it should be carried out and worked on continuously to realize the impact in the long run [26].

IV. EXISTING BOTTLENECK DETECTION METHODS – PROBLEM DESCRIPTION

The bottleneck detection in manufacturing is the key to improving production efficiency and stability in order to improve capacity. The existence of bottlenecks are situations that influence the throughput of the entire system. The larger the influence, the more significant the bottleneck [27]. In this part, the most common methods for bottleneck detection and limitations are described. The authors focused on statistical methods, bottleneck walk and process mining performance analysis.

A. PERFORMANCE ANALYSIS WITH STATISTICS

Statistical metrics are widely used and do not require sophisticated calculations. To detect the most common roots of bottlenecks, statistical characteristics for EQT (delays between operations in queues) and ET (breakdowns or resource absence) values might be chosen (Figure 3) [25]. According to the historical data, results can highlight three of the longest operations in the column for each position.

Operation	Statistics Values for Active Time (Elapsed_Time),s						Statistics Values for NonActive Time (Elapsed_Queue Time),s					
	Min	1st (25%)	2nd (50%)	3rd (75%)	Max	Mean	Min	1st (25%)	2nd (50%)	3rd (75%)	Max	Mean
A011600	4.93	5.77	6.03	6.29	105.09	6.04	10.00	11.00	12.00	25.00	8313.00	27.26
A011800	4.20	5.05	5.33	5.60	1788.52	6.05	6.00	3.00	13.00	32.00	8318.00	33.45
A011900	0.49	2.21	3.15	4.82	4194.62	7.21	6.00	7.00	11.00	18.00	8027.00	19.18
A012000	7.85	8.98	9.28	9.57	239.47	9.31	10.00	12.00	17.00	24.00	7894.00	21.63
A012200	2.86	3.71	4.03	4.34	936.77	5.00	3.00	4.00	4.00	5.00	1721.00	7.65
A012300	3.00	6.81	8.24	9.94	600.05	8.56	9.00	12.00	14.00	20.00	8175.00	26.10
A012500	8.00	9.22	9.51	9.80	56.85	9.57	7.00	8.00	10.00	17.00	7885.00	16.85
A012600	7.31	8.93	9.53	10.09	26.19	9.53	3.00	5.00	5.00	6.00	7867.00	8.28
A012800	17.14	18.99	19.46	20.26	62.10	19.84	30.00	47.00	86.00	153.00	17627.00	132.24
A012850	0.00	0.00	0.00	0.00	0.03	0.00	33.62	35.91	37.01	38.65	488.25	39.57
A012900	11.06	11.92	12.22	12.51	860.74	12.75	41.00	43.00	47.00	49.00	2540.00	53.05
A013500	0.00	0.00	0.00	0.00	4.30	0.00	10.75	56.14	60.32	74.91	8802.76	69.94

FIGURE 3. Statistical characteristics for each operation.

	OPERATIONS																			
	A011600		A011800		A011900		A012000		A012200		A012300		A012500		A012600		A012800		A013500	
Items in queue	4		3		1		2		0		1		1		1		11			
	Real	ProdLog	Real	ProdLog	Real	ProdLog	Real	ProdLog	Real	ProdLog	Real	ProdLog	Real	ProdLog	Real	ProdLog	Real	ProdLog	Real	ProdLog
Elapsed Time,s	8	6	5	6	11	7.2	10	9.3	5	5	7	8.56	9	9.57	8	9.5	25	20	0	0
Elapsed Queue Time,s	50	27.3	38	33.5	15	19.18	24	21.6	1	7.65	10	26.1	8	16.85	5	8.28	52	132.24	37	69.9

FIGURE 4. Comparing statistical and real (measured) times.

Usually, median (2nd quartile represents 50% of the data lies below the value) time reflects better values. Figure 4 shows the times estimation for each operation with mean times. Manual observations of reality could have had some inaccuracies due to human factors. It contains the next parameters:

- Items in the queue (buffer) – items/SFC capacity that can be placed before the operation on the conveyor belt.
- ET – time is needed for the operation to be completed.
- EQT – time, during which the item stayed in the queue before starting the operation.

For example, for Operation A011800 – 3 items can be in the queue before this operation, the real elapsed time is 5s, the mean elapsed time from the production log is 6s, the real-time for an SFC in the queue before this operation (on the way from A011600 to A011800) is 38s and the mean elapsed time obtained from production log is 33.5s.

The values from Figures 3 and 4 were taken for further bottleneck detection methods: the bottleneck time-lag method and the confidence interval method.

The statistics represent real data for a month and do not include process flow perspective, concept drifts (process changes), process variants, and NC cases. This method is not able to follow the flexibility of production lines. Mathematical models are made for common cases and are useless for the specific minority of processes. Also, the method strongly depends on historical information, more than on expert knowledge. Moreover, outliers in data impact results to a great extent and should be processed separately.

B. BOTTLENECK WALK METHOD

The bottleneck walk is based on observations of different process and resources states. These data are collected during a walk along the flow line. The gathered data are evaluated in a systematic process. The result of these two steps is a ranking of bottleneck sets that limit the output of the flow line during the period observed (Figure 5). There are three possible system states and the conclusion about the bottleneck:

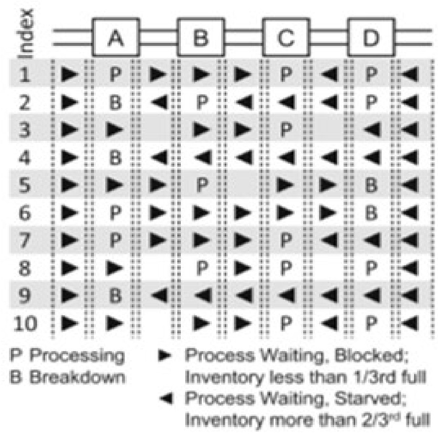


FIGURE 5. Bottleneck detection walk of a fast-changing valve assembly.

- May be the bottleneck working; breakdown; setup; maintenance; scheduled break, etc.
- Starved bottleneck is upstream - the inventory is empty or rather empty.
- Blocked bottleneck is downstream - the inventory is full or rather full.

The method is suited for practical use by shop floor supervisors and clerks. The direct observation of the bottleneck also gives additional information about the underlying causes of the bottlenecks, simplifying the improvement of the system capacity [27].

The method is functional and is applied to the manufacturing process and more. However, it finds queues and weak points manually by counting items before and after the operation. The detection can be automated using sensors, production logs and information about buffer of items before and after the operation (Figure 4).

In this case, another issue occurs, when a manual resource takes an item (NC) out of the production line/conveyer or adds new items in a certain part of the production line to rework. Therefore, items just emerge in the production log through the process and it complicates to follow them. Also, the method is supposed to work with a direct flow process, not processing conjunctions and parallel flows.

C. BOTTLENECK ANALYSIS WITH PROCESS MINING AND ANIMATION

As soon as the adjusted descriptive process model has been created, statistical metrics can be added to it and give an integrated perspective. In order to get execution times for all steps of the production process, both start and completion timestamps in the data set should be identified.

Performance analysis allows seeing mean, median, min/max time, frequency of cases through the process. Filtering by process variants, start/end timestamps, case duration, etc. helps to focus on specific cases with a detailed view.

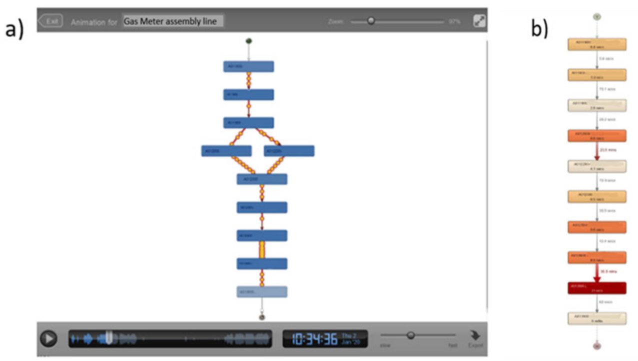


FIGURE 6. The sample of the process visualization from flow and performance perspectives: a) Process animation to detect bottlenecks in real-time (Disco), b) Process map with Performance perspective1 (it represents only the visualization and process model will be shown in detail in the next part).

When one wants to intuitively spot and highlight bottlenecks in the process, process animation can be applied. In such bottleneck situations, you have a lot of cases that accumulate in certain paths. Real-time monitoring shows when many cases pile up on a certain arc and are causing congestion. Process animation[2] groups these cases into larger "bubble clusters" (Figure 6a)

Another process map view, which is shown in Figure 6b, presents performance characteristics for each operation and paths between them. Red colours detect the longest times in comparison with other parts. The process model enriched with information about times makes the detection more transparent and comprehensible. Figure 6 presents an abstract view of the process bottlenecks with "bubble clusters"; part b will be shown in detail in Figure 9. All operation names were hidden according to the confidential policy of the manufacturing company.

Although the process model combined with statistic metrics gives a better understanding of the process and possible bottlenecks, as well as their roots, the detection in this method is relative. All activities and paths are compared with each other based on historical data. Therefore, if one operation takes 25min (and this time is regular for the operation), but another one takes 5s, then the first operation will be detected as a weak place in the process. This situation can be omitted by time processing and considering each operation as instant. For example, start_time and end_time for each activity should be equated to DATE_TIME. In this case, the waiting time between two operations will combine EQ (operation duration time) and EQT (time in the queue). All operations will be instant. This makes it impossible to differ bottleneck roots – breakdown of a resource or delays on the path between activities.

D. PROPOSED BOTTLENECK DETECTION METHODS

Each industrial process is having its time period (frequency) and this period set up the reference for the application of detection methods. The real process set up the reference period/frequency (time complexity) and our algorithm has to be n2 or n3 (Big-O notation) for guaranteeing effectiveness.

1) TimeLag METHOD

The prior method used will give the previous and the next SFC from the operation where the bottleneck happened, and search where the delay happened in the first place and highlight that operation->work center->resource which played a role in creating that ripple effect (Figure 7a).

It helps in investigating the operation which caused the delay after segregating the initial bottleneck and finding the previous and next SFC for that incident operation. The inputs can be SFC given or the operation of interest which results in the bottleneck analysis flow.

1. Identifying an operation of interest to calculate the root cause of bottlenecks ensure they are clustered and sorted by DateTime properly, find the previous and successive SFC's, identify the big transfer times and isolate them for investigation.

2. The isolated time span is taken, the previous 50 items processed, and the next 50 items processed during the DateTime period is taken.

3. During the isolated time range, the operations which are running indicates they were functioning well during that time, and the missing operations indicate that it was the first operation to break and result in a bottleneck which resulted in the ripple effect.

It is able to find the biggest bottleneck along the production process, what is useful in the theory of constraints [28]. However, if the process has a sequence of bottlenecks, the method should be readjusted [25]. Therefore, this method was not taken as a more suitable and effective one, conducting the authors to use Confidence interval method.

2) CONFIDENCE INTERVAL METHOD

Confidence interval method (Figure 7b) shows an adequate level of robustness for the production line, supporting identifying many different kinds of bottleneck situations in the real production process under analysis.

To implement this method, one should follow the next steps:

- Calculate median queue times EQT for all operations. ET is not considered because it will show resource breakdowns of resources and this information are known.
- For each case – the SFC compare real times with previous ones and give status Slower or Faster based on the result. It's preferable to add additional 6 - 7s to the confidence interval to avoid sensitive calculations. For example, if median EQT for operation A12800 is 19.46s accordingly, then the confidence interval might be less than 26s (a low threshold is not important, since just a longer time is considered as a bottleneck).
- For all operations with status Slower, difference Queue time – Median queue time should be compared between

[2]www.fluxicon.com

previous and next operation. It will show the direction of the bottleneck root – upstream or downstream.

It is possible to consider three possible reasons for a bottleneck: resource unavailability (for manual resource – absence at the workstation, for machine resource – breakdown), the downstream impossibility of shipment because of issues, upstream lack of materials.

Algorithm 1 Bottleneck Detection Through the Process

```
  // the list of EQT median values for
each operation in the process based
  // on historical data, can be chosen
upper quartile Q3, mean value or the
value // defined be expert group
  Median = [value for 1st
  operation,...value for nth operation]
  diff = []
  result = []
  for i = 1 to N do
  // to make calculations not too
  sensitive, it is recommended to add
  8-9 seconds,
  // depending on the operation type
    diff = Median [i] - EQT[i]
    if diff < 0 then
        result = "Slower"
    else
        result = "No bottleneck"
```

Therefore, the minimum value of the difference between median time and EQT – min(diff) – shows the longest bottleneck, diff.index(min(diff)) – shows the index of the operation with the longest waiting time and index.result = "Slower" presents all operations with bottlenecks through the process.

Algorithm 2 Bottleneck Direction

```
if result = "Slower" then
    if diff[i] < diff[i-1] then
      direct = "Downstream"
    else
      direct = "Upstream"
else
    result = "No bottleneck"
```

The algorithms work with columns in table frames. Also, for the last operation in the process if result = "Slower", then status automatically will be *Bottleneck-shipping reason*. Usually, the process in the workcenter ends with transportation to another workcenter.

In order to get a third possible reason – the breakdown of resource, one should compare ET with the inspect time interval. As a rule, all machine resources are connected to the information management system and detection of bottleneck happens immediately.

The algorithm allows us to detect a bottleneck root in the straightforward process as well as multilevel production lines, where we work with some inputs to one resource. In terms of the manufacturing line, the flow of materials can come from different floor's lines as well as from the same level.

Remark: A qualitative comparison of both named methods allow us to reinforce the following facts: the application of the Time Lag method facilitate detecting only major bottlenecks but the Confidence Interval Method support detecting many of them including the positions along the value stream.

V. TEST CASE STUDY

A. PROCESS FLOW MODELLING

Process mining algorithms are able to discover, monitor, enrich and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today's information systems. The real descriptive model is the basis for valuable types of analysis like conformance checking (comparing two models – descriptive and normative, or the real model with event log for a new month to find deviations), streaming (monitoring the production process in real-time), performance analysis (bottlenecks, resource analysis), recommendation and decision support system (decision rules for KPI – key performance indicators). Each node in the process can be enriched with required data like resources, risks, probabilities, productivity characteristics.

Process mining techniques are applied to manufacturing processes event logs to obtain the process model and activity related statistic information. The manufacturing process of assembling a line of gas meter is considered, including all resources (machines, employees, sensors) along with data from the information system SAP.

For the production line of the working center, the goal is to follow item production. Therefore, SFC is taken as the case (process instance), Operation is activity for the process mining algorithms, DATE_TIME and DATE_TIME+ ET are start and end timestamps accordingly. To discover more dependencies in the data, the whole process for different work centres or to enrich the process model following additional attributes can be used: shop order number, resource, item, router, work center. Also, based on the main existing parameters, new more complicated ones can be calculated: flow time, waiting time, rework, deviations, process variants or service-level agreement (SLA) measurements (window time for each level of service, cost trade-offs, frequencies).

Figure 8 shows three of the most popular algorithms of Process Mining: Alpha Miner with Petri net notation, Inductive Miner – Petri net, Fuzzy Miner – directly follower graph.

It's necessary to build and check the process model in various notations to avoid the limitlessness of each algorithm [24]. Rectangles/nodes represent operations and arches are relations between them. Most process miners usually apply fuzzy miner and work with direct follower graph. As one can notice, Figure 8c shows three possible scenarios for the production process. The process variant (sequence of

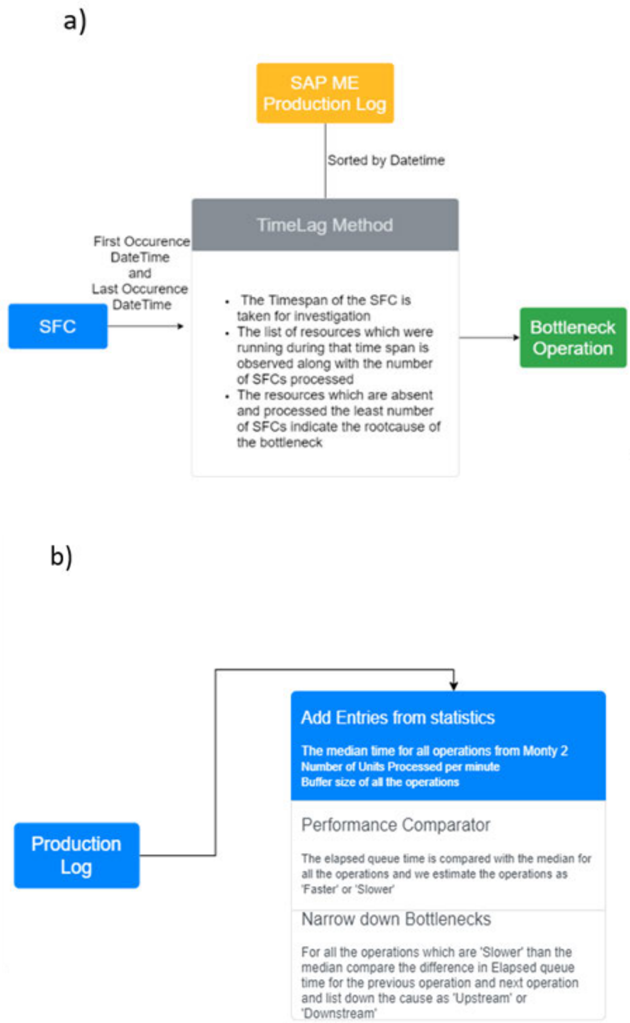


FIGURE 7. **Methods for the detection of bottleneck roots: a) Time Lag method; b) Confidence interval method.**

operations) with the highest frequency will be taken, it was conducted for most items, instead of specific cases.

So, the loop in the process, which Figure 8 presents, is avoided and statistics values for operations A012800 and A013500 are improved respectively.

All three notations were considered, and Directly-Follows Graph was chosen as a basis for presenting the bottleneck method.

B. APPLYING CONFIDENCE INTERVAL METHODS

In order to estimate the confidence interval methods, one production line was taken for experiments. Process variants (sequence of operation) varies for different types of items, so the process flow that is common for many cases was chosen as priority one.

Figure 9 shows an example of applying the method for a specific SFC.

To present calculations for each operation and to avoid an overloaded set of data, let's the orderly number the sequence of operations.

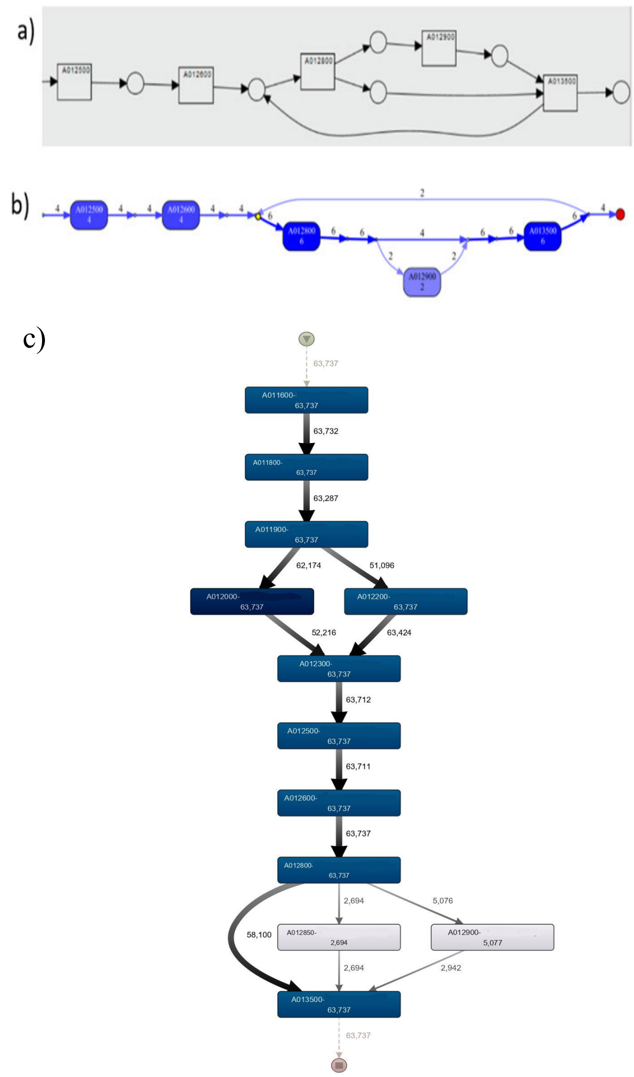


FIGURE 8. **Three PM algorithms to discover the process model: Alpha Miner - Petri net (PROM), b) Inductive Miner - Petri net (RapidMiner), c) Fuzzy Miner - direct follower graph (Disco). Note: See Appendix A with High-Quality Picture.**

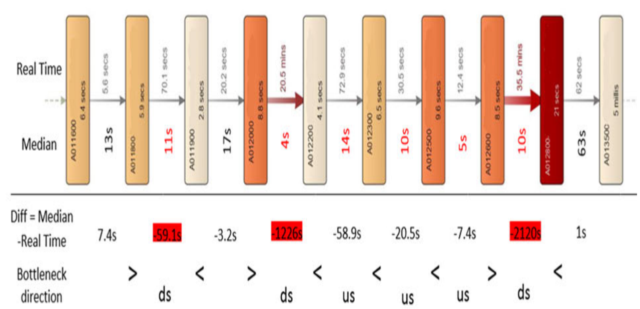


FIGURE 9. **Example of the case with applied Bottlenecks Confidence method.**

Applying algorithm 1 to this example, the following sets can be defined:

- EQT =<5.61, 70.12, 20.23, 12304, 72.95, 30.56, 12.47, 2130.8, 62.9> - elapsed queue time for each operation;

- Median =<13.1, 11.2, 17.3, 4.4, 14.5, 10.6, 5.7, 10.8, 63.9> - median time for EQT was taken from Figure 3 with statistics of historical data and shows the common time for general amounts of cases in the certain position of the process. This time has to be discussed with experts and process owners. It's important to notice that for more realistic median values the main process variant (sequences of operations) was taken and median values for operations A012800 and A013500 are different than in Figure 3 because the process variant with loop for scrap items was not considered as the main one (see Figure 8). Instead of the median value, other statistics like quartiles, modes, mean values should be taken into account. Also, the term Confidential interval was used to define correct boundaries for each part of the process. Low thresholds are not so important and will be omitted in calculations because they show faster cases, which are not necessary for the analysis. A high threshold, in a turn, determines the sensitivity of algorithms and the accuracy of results. For example, the delay for index = 3 is 3.2s and was found as not crucial.
- Diff =<7.41, −59.12, −3.23, −1226, −58.95, −20.56, −7.47, −2120, 1> - the difference between median and real time.
- Status "Bottleneck" is defined for all values in the set Diff that are lower than 0. Checking three longest bottlenecks presents next values (index:diff) =<8:-2120; 4:-1226; 2:-59.1>. These vulnerabilities were discussed with the process owner later and their influence on the whole process was proved.

Further, using the Algorithm 2, directions for all bottlenecks are identified (ds-downstream, us-upstream, nb-no bottleneck): direct $=<nb^1, ds^2, nb^3, ds^4, us^5, us^6, us^7, ds^8, nb^9>$. Indexes link to operations' numbers.

According to the theory of constraints [28], it's important to identify the most limiting factor (i.e., constraint) that stands in the way of achieving a goal and then systematically improving that constraint until it is no longer the limiting factor. The longest bottlenecks are examined and can be detected after a comparison of differences between Median and Real-time for each operation. The last operation of the process A013500 can not be compared with the next operation, so if elapsed queue time is higher than median, the bottleneck can be marked at that point. The confidence intervals and reasons for bottlenecks must be discussed with process experts. Elapsed times can be also included in calculations for further analysis. It's important to notice that one delay has a riffle effect and other items ate stuck in the line. So, the bottleneck detection for one specific SFC can prevent big lines and delays.

After estimation root causes directions for bottlenecks, it is possible to make suggestions about real reasons for delays. Table 1 below shows the list of possible reasons for each operation in the production line.

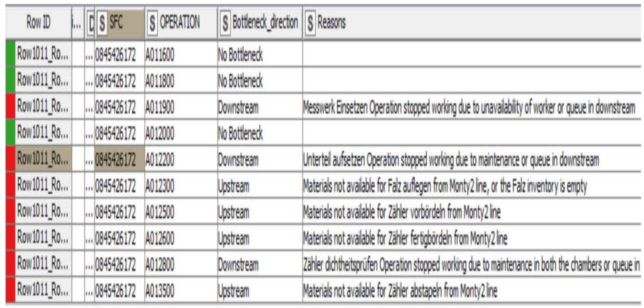
Based on the algorithm and the list, one can make production process rules and offer some probable reasons for delays. Also, this recommendation table can be enriched with information about breakdowns in the production line, that leads to making bottleneck reasons more accurate.

Eventually, Figure 10 is automatically created, based on two algorithms of detection bottleneck position and the table of reasons for different statuses [25].

It is necessary to add, that finding bottleneck for one SFC represents the same bottlenecks for a batch of SFCs depending on the buffer size of the production line. If the production line processes 14 SFC, then all of them will have a similar time to be completed and delays in the different parts of the process according to capacities (items in the queue) between parts. However, the roots of bottleneck for all these SFCs are the same.

TABLE 1. Possible reasons for each operation in the process with directions.

Operation	Direction	Reasons
A011600	Downstream	Operation A011600 stopped working due to maintains or queue downstream
A011600	Upstream	Materials are not available for this operation
A011800	Downstream	Operation A011800 stopped working due to unavailable shipment
A011800	Upstream	Materials are not available for this operation/ waiting for materials from another work center
…	…	…



Row ID	i...	D	S SFC	S OPERATION	S Bottleneck_direction	S Reasons
Row1011_Ro...		...	0845426172	A011600	No Bottleneck	
Row1011_Ro...		...	0845426172	A011800	No Bottleneck	
Row1011_Ro...		...	0845426172	A011900	Downstream	Messwerk Einsetzen Operation stopped working due to unavailability of worker or queue in downstream
Row1011_Ro...		...	0845426172	A012000	No Bottleneck	
Row1011_Ro...		...	0845426172	A012200	Downstream	Unterteil aufsetzen Operation stopped working due to maintenance or queue in downstream
Row1011_Ro...		...	0845426172	A012300	Upstream	Materials not available for Falz auflegen from Monty2 line, or the Falz inventory is empty
Row1011_Ro...		...	0845426172	A012500	Upstream	Materials not available for Zähler vorbördeln from Monty2 line
Row1011_Ro...		...	0845426172	A012600	Upstream	Materials not available for Zähler fertigbördeln from Monty2 line
Row1011_Ro...		...	0845426172	A012800	Downstream	Zähler dichtheitsprüfen Operation stopped working due to maintenance in both the chambers or queue in
Row1011_Ro...		...	0845426172	A013500	Upstream	Materials not available for Zähler abstapeln from Monty2 line

FIGURE 10. Bottleneck analysis for one specific SFC. Note: See Annex A with High-Quality Pictures.

VI. CONCLUSION AND OUTLOOKS

As a consequence of the complexity and rapid changes of the manufacturing processes, current process models become obsolete and should be discovered with advanced methods. The work was centred around building a process model and using it as a basis for the following bottleneck detection methods. This extraordinary amount of data provides unprecedented opportunities for data-driven decision making and knowledge discovery. The analysis started with understanding the production by process modelling to visualize the process flow. Process Mining tools enabled the ability to take a deeper look into production by seeing statistics and measuring the run rate of production to help identify the different variants in the process. Also, Process Mining offers

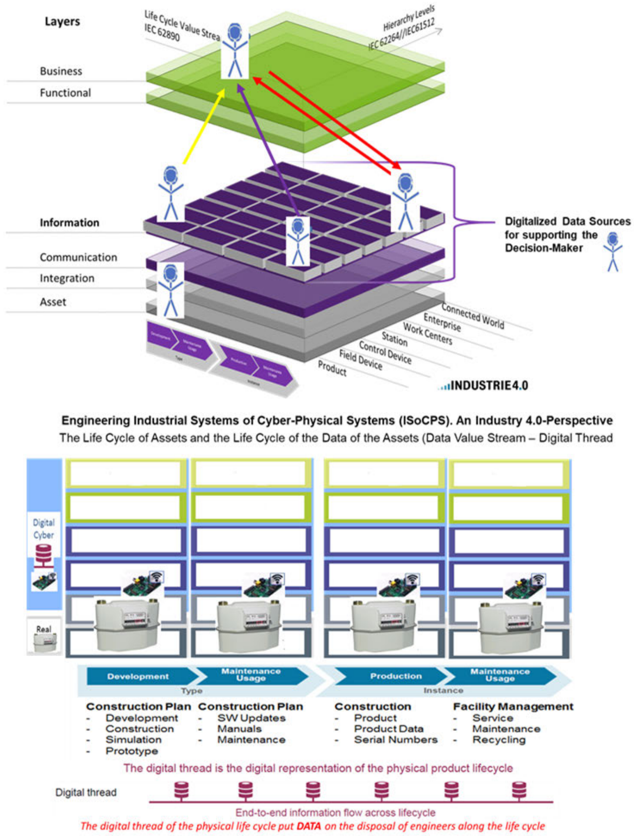


FIGURE 11. DIN SPEC 91345 compliant Digitalization for providing Digitalized Data Sources in a Digital Thread.

the necessary algorithms to obtain the real descriptive process model as a basis for future analysis.

The bottleneck detection is a crucial part of the continuous improvement manufacturing processes. Commonly used methods were discussed throughout the work. However, the flexible and appropriate method of bottleneck detection stays relevant. This paper was focused on finding the most time-consuming places in the process and the reasons that caused them. Using the time lag method and the confidence interval method, the workflow will identify the bottlenecks from the dataset and add the root cause for the bottlenecks based on the operation and delay incurred by the downstream or upstream process. The underlying process model makes the detection more transparent, accurate and valuable. Using modelled process and knowledge about time parameters from historical data allows following the value stream in real-time and give immediate scenarios for arising issues. Thus, the proposed approach helps not just detect a bottleneck but, what's more efficient for manufactures, give a possible root cause. Also, the process model enriches the method giving more options to monitor reasons for deviations.

In contrast with widely used approaches, which are described in Related works, the offered method doesn't rely on manual resource/worker who will follow the process and make calculations on the shopfloor. It can be applied to multilevel shopfloor including intersections in the process.

The methods have been developed based on research done at Honeywell, Production Intelligence Department, Germany (see https://www.elster-instromet.com/de/index).

Given that the industrial eco-system is digitalized following e.g. the DIN Specification 91345 RAMI 4.0, as addressed in Section II C, for each kind of Asset positioned within the Hierarchy standardized according to the IEC 62264 / IEC 61512, and considering the different phases of its life cycle according to the IEC 62890, a correctly specified and implemented layer 4 of the digitalization dimension implies the specification and implementation of an unprecedented set of digitalized data sources, as depicted on the upper part of Figure 11 (see [29]).

Customizing this Industry 4.0-compliant digitalization approach for a gas metering component located in the level "product" of the hierarchy, as shown on the lower part of Figure 11, allows the definition, specification and implementation of its Digital Thread. A considerable set of outlooks from the approach presented here can then be highlighted, among others, (i) the aspects of interoperability of the data along with the Digital Thread, (ii) the collaborative Decision-Making process that actors along the value stream can provide and (iii) the application of the methodologies and tools addressed in this manuscript to examining and analysing the big data provided by such a big amount of heterogeneous data sources.

An essential future work should be the validation of the proposed innovation solution defining metrics with measurements addressing among other parameters of the solution like

Row ID	i...	D	S SFC	S OPERATION	S Bottleneck_direction	S Reasons
Row1011_Ro...	...		0845426172	A011600	No Bottleneck	
Row1011_Ro...	...		0845426172	A011800	No Bottleneck	
Row1011_Ro...	...		0845426172	A011900	Downstream	Messwerk Einsetzen Operation stopped working due to unavailability of worker or queue in downstream
Row1011_Ro...	...		0845426172	A012000	No Bottleneck	
Row1011_Ro...	...		0845426172	A012200	Downstream	Unterteil aufsetzen Operation stopped working due to maintenance or queue in downstream
Row1011_Ro...	...		0845426172	A012300	Upstream	Materials not available for Falz auflegen from Monty2 line, or the Falz inventory is empty
Row1011_Ro...	...		0845426172	A012500	Upstream	Materials not available for Zähler vorbördeln from Monty2 line
Row1011_Ro...	...		0845426172	A012600	Upstream	Materials not available for Zähler fertigbördeln from Monty2 line
Row1011_Ro...	...		0845426172	A012800	Downstream	Zähler dichtheitsprüfen Operation stopped working due to maintenance in both the chambers or queue i
Row1011_Ro...	...		0845426172	A013500	Upstream	Materials not available for Zähler abstapeln from Monty2 line

FIGURE 12. Bottleneck analysis for one specific SFC.

robustness, scalability, etc. and comparing the results with other tools.

APPENDIX

APPENDIX A

See Figure 12.

APPENDIX B

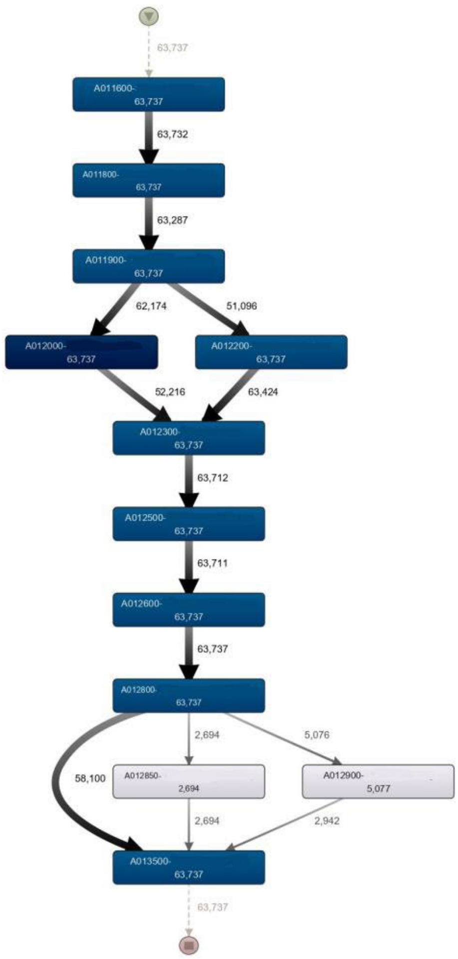


FIGURE 13. Fuzzy Miner - Directly-Follows Graph (Disco). Rectangles represent each Activity/Operation and arches are the connections/relations between them.

REFERENCES

[1] A. W. Colombo, M. Gepp, J. Barata, P. Leitao, J. Barbosa, and J. Wermann, *Digitalized and Harmonized Industrial Production Systems: The PERFoRM Approach*. Boca Raton, FL, USA: CRC Press, 2019.

[2] E. Ruschel, E. A. P. Santos, and E. D. F. R. Loures, "Establishment of maintenance inspection intervals: An application of process mining techniques in manufacturing," *J. Intell. Manuf.*, vol. 31, no. 1, pp. 53–72, Jan. 2020, doi: 10.1007/s10845-018-1434-7.

[3] Á. Rebuge and D. R. Ferreira, "Business process analysis in healthcare environments: A methodology based on process mining," *Inf. Syst.*, vol. 37, pp. 99–116, Apr. 2012.

[4] J. Rudnitckaia, W. Intayoad, T. Becker, and T. Hruska, "Applying process mining to the ship handling process at oil terminal," in *Proc. IEEE Int. Conf. Ind. Cyber Phys. Syst. (ICPS)*, May 2019, pp. 552–557, doi: 10.1109/ICPHYS.2019.8780305.

[5] Y. Wang, F. Caron, J. Vanthienen, L. Huang, and Y. Guo, "Acquiring logistics process intelligence: Methodology and an application for a Chinese bulk port," *Expert Syst. Appl.*, vol. 41, no. 1, pp. 195–209, Jan. 2014.

[6] T. B. Hong Tu and M. Song, "Analysis and prediction cost of manufacturing process based on process mining," in *Proc. Int. Conf. Ind. Eng., Manage. Sci. Appl. (ICIMSA)*, May 2016, pp. 1–5, doi: 10.1109/ICIMSA.2016.7503993.

[7] M. T. Wynn, W. Z. Low, and W. Nauta, "A framework for cost-aware process management: Generation of accurate and timely management accounting cost reports," in *Proc. Conf. Res. Pract. Inf. Technol. Ser.*, vol. 143, 2013, pp. 79–88.

[8] A. Bettacchi, A. Polzonetti, and B. Re, "Understanding production chain business process using process mining: A case study in the manufacturing scenario," in *Advanced Information Systems Engineering Workshops* (Lecture Notes in Business Information Processing), vol. 249. Cham, Switzerland: Springer, 2016.

[9] S. Rinderle-Ma and J. Mangler, "Process automation and process mining in manufacturing," in *Proc. Int. Conf. Bus. Process Manage.* Cham, Switzerland: Springer, 2021, pp. 3–14.

[10] F. Stertz, J. Mangler, B. Scheibel, and S. Rinderle-Ma, "Expectations vs. experiences-process mining in small and medium sized manufacturing companies," in *Proc. Int. Conf. Bus. Process Manage.* Cham, Switzerland: Springer, 2021, pp. 195–211.

[11] R. Bemthuis, N. van Slooten, J. Arachchige, J. Piest, and F. Bukhsh, "A classification of process mining bottleneck analysis techniques for operational support," in *Proc. 18th Int. Conf. e-Bus.*, 2021, pp. 127–135.

[12] *Process Mining Camp 2020 Webinar Series*. Accessed: Dec. 20, 2021. [Online]. Available: https://fluxicon.com/camp/2020/

[13] P. Vrablecová, "Predictive analytics on data streams," in *Proc. Inf. Sci. Technol., Bull. ACM Slovakia*, vol. 2, 2019, pp. 10–18.

[14] T. Meudt, J. Metternich, and E. Abele, "Value stream mapping 4.0: Holistic examination of value stream and information logistics in production," *CIRP Ann.*, vol. 66, no. 1, pp. 413–416, 2017.

[15] D. Knoll, G. Reinhart, and M. Prüglmeier, "Enabling value stream mapping for internal logistics using multidimensional process mining," *Expert Syst. Appl.*, vol. 124, pp. 130–142, Jun. 2019.

[16] M. Pegoraro, M. S. Uysal, and W. M. van der Aalst, "Discovering process models from uncertain event data," in *Business Process Management Workshops* (Lecture Notes in Business Information Processing), vol. 362. Cham, Switzerland: Springer, 2019, pp. 238–249.

[17] D. Schuster, S. J. van Zelst, and W. M. van der Aalst, "Incremental discovery of hierarchical process models," in *Research Challenges in Information Science* (Lecture Notes in Business Information Processing), vol. 385. Cham, Switzerland: Springer, 2020.

[18] K. Nagorny, S. Scholze, A. W. Colombo, and J. B. Oliveira, "A DIN spec 91345 RAMI 4.0 compliant data pipelining model: An approach to support data understanding and data acquisition in smart manufacturing environments," *IEEE Access*, vol. 8, pp. 223114–223129, 2020, doi: 10.1109/ACCESS.2020.3045111.

[19] P. Leitao, F. Pires, S. Karnouskos, and A. W. Colombo, "Quo vadis industry 4.0? Position, trends, and challenges," *IEEE Open J. Ind. Electron. Soc.*, vol. 1, pp. 298–310, 2020, doi: 10.1109/OJIES.2020.3031660.

[20] N. Ayvarnam and P. S. Mayurappriyan, "Dynamic scheduling of machines towards the vision of industry 4.0 studio—A case study," in *Proc. 2nd Int. Conf. Intell. Comput. Appl.* Singapore: Springer, 2017, pp. 103–111, doi: 10.1007/978-981-10-1645-5_9.

[21] J. Chandan and S. Dipankar, *Implementing SAP Manufacturing Execution*, 1st. ed. Quincy, MA, USA: Rheinwerk Publishing, 2016, p. 480.

[22] A. W. Colombo, T. Bangemann, S. Karnouskos, J. Delsing, P. Stluka, R. Harrison, F. Jammes, and J. Lastra, "Industrial cloud-based cyber-physical systems," in *The IMC-AESOP Approach*. Cham, Switzerland: Springer, 2014.

[23] W. M. P. van der Aalst and A. Berti, "Discovering object-centric Petri nets," *Fundamenta Informaticae*, vol. 175, nos. 1–4, pp. 1–40, Sep. 2020.

[24] W. M. P. Van Der Aalst, "A practitioner's guide to process mining: Limitations of the directly-follows graph," *Proc. Comput. Sci.*, vol. 164, pp. 321–328, Jan. 2019.

[25] H. S. Venkatachalam, "Value stream analysis of gas meter production by simulation and forecasting," M.S. thesis, Inst. Ind. Inform., Automat. Robot., IP-Protected, Disposal Request, Univ. Appl. Sci. Emden, Emden, Germany, 2020.

[26] J. Lee, E. Lapira, B. Bagheri, and H.-A. Kao, "Recent advances and trends in predictive manufacturing systems in big data environment," *Manuf. Lett.*, vol. 1, no. 1, pp. 38–41, Oct. 2013.

[27] C. Roser, K. Lorentzen, and J. Deuse, "Reliable shop floor bottleneck detection for flow lines through process and inventory observations: The bottleneck walk," *Logistics Res.*, vol. 8, no. 1, pp. 63–68, Dec. 2015.
[28] J. Cox III and J. Schleier, *Theory of Constraints Handbook*. New York, NY, USA: McGraw-Hill, 2010.
[29] A. W. Colombo, G. J. Veltink, J. Roa, and M. L. Caliusco, "Learning industrial cyber-physical systems and industry 4.0-compliant solutions," in *Proc. IEEE Conf. Ind. Cyberphysical Syst. (ICPS)*, Jun. 2020, pp. 384–390, doi: 10.1109/ICPS48405.2020.9274738.

JULIA RUDNITCKAIA received the M.Eng. degree in information systems from Admiral Ushakov Maritime State University, Russia. She is currently pursuing the Ph.D. degree in the field of computer science and engineering with the Faculty of Information Technologies, Brno University of Technology, Czech Republic. Since 2014, she has been working on four international projects, including traineeships in different countries (Norway, Germany, and Russia). As a Ph.D. candidate, she is researching the topic of process mining, process analysis and diagnosis in the logistics domain, specifically the seaport segment.

HARI SANTHOSH VENKATACHALAM, photograph and biography not available at the time of publication.

ROLAND ESSMANN, photograph and biography not available at the time of publication.

TOMÁŠ HRUŠKA received the Ing. (M.Sc.) and C.Sc. (Ph.D.) degrees from the Brno University of Technology, Czech Republic. Since 1978, he has been working with the Brno University of Technology, and since 1998, as a Full Professor. From 2002 to 2007, he was the first Dean of the New Faculty of Information Technology. From 1978 to 1983, he dealt with research in the area of compiler implementation for simulation languages. Since 1990, he has been working on the implementation of information systems. Since 2005, he has been working on the implementation of the Codasip project of the processor description language and methods for transformations of processor models. He deals with the analysis of industrial processes as well.

ARMANDO WALTER COLOMBO (Fellow, IEEE) received the B.Sc. degree in electronics engineering from the National Technological University of Mendoza, Argentina, in 1990, the M.Sc. degree in control system engineering from the National University of San Juan, Argentina, in 1994, and the Ph.D. degree in engineering (manufacturing automation and systematization) from the University of Erlangen-Nuremberg, Germany, in 1998. He joined the Department of Electrotechnical and Computer Sciences, University of Applied Sciences Emden/Leer, Germany, became a Full Professor, in August 2010, and the Director of the Institute for Industrial Informatics, Automation and Robotics (I2AR), in 2012. From 2001 to 2018, he worked as the Manager for Collaborative Innovation Programs and also as an Edison Level 2 Group Senior Expert at Schneider Electric Corporation. He has over 30 industrial patents and more than 300 peer-reviewed publications. With his contributions, he has performed scientific and technical seminal contributions that are nowadays being used as one of the basis of what is recognized as "The 4th Industrial Revolution." His research interests include the fields of industrial cyber-physical systems, industrial digitalization, engineering of industry 4.0-compliant solutions, system-of-systems engineering, industrial Internet of Things, and intelligent production automation systems. He is a Distinguished Lecturer of the IEEE Systems Council and the Co-Editor-in-Chief of the IEEE Open Access Journal of the Industrial Electronics Society (OJIES). He is listed in Who's Who in the World/Engineering 99-00/01 and in Outstanding People of the XX Century (Bibliographic Centre Cambridge, U.K.).