Data-driven decision support tool for production planning: a framework combining association rules and simulation

Data-driven decision support tool for production planning: a framework combining association rules and simulation

Virginia Fani[a1], Sara Antomarioni[b], Romeo Bandinelli[a], Maurizio Bevilacqua[b]

[a] Department of Industrial Engineering, University of Florence, Florence, Italy.

[b] Department of Industrial Engineering and Mathematical Science, Università Politecnica delle Marche, Ancona, Italy.

[1] Corresponding author: virginia.fani@unifi.it, Florence, IT

Abstract

Nowadays, guaranteeing the highest product variety in the shortest delivery time represents one of the main challenges for most of industries. The dynamic contexts where they have to compete push them to quickly readapt their processes, increasing the need for reactive decision-support tools to identify targeted actions to improve performance. Starting from the analysis of existing decision-support tools separately adopting simulation or data mining techniques, a framework that combines Association Rule Mining (ARM) and simulation has been developed to capitalize on the benefits brought by both techniques. On the one hand, ARM supports companies in identifying the main criticalities that slow down production processes, such as different causes of stoppage, giving a priority ranking of interventions. On the other hand, data-driven simulation is used to validate the ARM results and to conduct scenario analyses to compare the KPIs values resulting from different configurations of the production processes. Once the best-impacting mitigating actions have been implemented, the proposed framework can be iteratively used to define an updated set of intervention areas to enhance, promoting continuous improvement. This data-driven approach represents the key value of the framework, guaranteeing its easy-to-readapt and iteratively application.

Theoretical contributions refer to the use of simulation with ARM not only to validate relations but to perform scenario analyses in an iterative way, as well as to the novelty application in a low-tech sector. From a

practical point of view, a case study in the fashion industry demonstrates the usability and reliability of the proposed framework.

Keywords

Decision support tool; Association Rules; Simulation; Data-driven; Production.

1. Introduction

Nowadays, most industries are characterized by high product variety and short delivery time, making production planning and control critical. More in detail, the fast-moving context where companies have to compete pushes for quickly readapting their processes to guarantee the best performance, especially in terms of production lead times. In addition, dynamic sectors require frequent identification of criticalities to be solved in order to define targeted mitigation actions to implement to reach the best performance, such as lead time reduction or deviations from expected lead times identification. To increase their competitiveness, companies have to iteratively manage this process, following a continuous improvement approach. Due to the fast-changing scenario, decision tools to support companies in both setting out intervention priority and comparing their impacts in terms of KPIs improvement, such as lead time reduction, are highly recommended. In comparison with other strategical approaches (e.g., Business Process Reengineering), companies can benefit from the adoption of decision-support tools at the operative level, as confirmed by Khan et al. (2018).

To ensure a quick data analysis to identify the main criticalities that slow down production processes, such as different causes of stoppage, statistic tools should be substituted by data-driven approaches since they are more effective in providing practical insights for process improvement (Patwardhan et al., 2016); among them, Association Rule Mining (ARM) can be successfully applied since it presents several advantages over the statistical techniques, like being able to consider several variables at once and not needing basic hypothesis or assumptions prior to the analysis (Antomarioni et al., 2022; Ciarapica et al., 2019). On the other hand, simulation has been widely used to compare different scenarios in terms of how changes in input parameters impact performance values, even in low-tech and dynamic industries such as the fashion one (e.g., Cagliano et al., 2011; Fani et al., 2017a, 2017b; Kim et al., 2018). Moreover, like the ARM, even discrete simulation models can be developed following a data-driven approach (Fani et al., 2021) in order to guarantee the least effort possible for companies after each iteration once the model has been set.

According to the described context, the aim of the paper is to define a framework to be used as a decision support tool for companies who need to quickly evaluate the best parameter configuration to enhance production performance and planning, combining data-driven ARM and simulation approaches in a decision support tool aiming to: (a) estimate lead time deviations from standard times; (b) anticipate process stoppages; (c) understand process modification to overcome the identified criticalities.

In other words, the tool will support production planners in identifying the most critical issues that impact the production KPIs (i.e., lead time, in the proposed case study) in order to define the mitigation actions needed and set out intervention priorities. ARM and simulation are often used separately as decision support systems (DSS). However, their joint implementation can be found in the fault detection application area, especially in IT-oriented sectors. The existing contributions are mainly focused on using the simulation to validate the relationships identified through the ARM (e.g., Glowacka et al., 2017a; Xin Zhao et al., 2018a), leaving unexplored the possibility of using the simulation to conduct scenario analyses and iteratively define corrective actions for the production planning process, taking into account the most likely deviation from the expected lead time during activity scheduling. Once the selected actions have been implemented, the proposed framework can be iteratively used to define an updated set of intervention areas to enhance, according to a continuous improvement approach. Applying the proposed framework to dynamic but low-tech contexts, like the fashion industry, represents a further novelty of the study. Indeed, ARM and simulation jointly applied to fashion industry case studies could not be found.

The paper is therefore organized as follows: the main results of the literature review are proposed in Section 2; in Section 3, the proposed model is detailed step-by-step; in Section 4, the application of the framework to a real company is described; the main results and implications from the case-study are discussed in Section 5; finally, conclusions and further developments are resumed in Section 6.

2. Literature review

2.1 Literature review on data-driven DSS

Data-driven framework to support the decision-making process are vastly applied in existing literature on production processes in dynamic contexts. Such contributions agree on stating that the quality of the model relies on the accuracy of the information extracted by data (Ren et al., 2023). In these works, several framework are proposed basing on different techniques: for example, rail traffic can be controlled through the implementation of deep forest ensemble learning to manage dwelling and running times (Luo et al., 2022). Kabadurmus et al. (2022), instead, propose a data-driven DSS to limit food waste and maximize profits: they use a simulator that dynamically updates prices of grocery products on the bases of their freshness levels. Similarly, Xu et al. (2020) develop a data-driven quality degradation model to solve the same problem, dynamically optimizing the supply network. Different kind of flows can be balanced through data-driven approaches: for instance, ARM can be applied to solve the bike sharing problem, by balancing agents' flows considering recurrent patterns (Cipriano et al., 2021). Focusing on the production area, Parhizkar et al. (2020) propose a data-driven framework for risk-informed decision making using a Bayesian network model aiming to support the operators by providing information on the status of the system and possible variations from the normal behavior; a case study on the dynamic positioning of drilling operators is used to present an validate the case study. In Kaniappan Chinnathai et al. (2021), a data-driven framework based on discrete event simulation and genetic algorithm is proposed to identify the best production process configurations taking into account customer demand. The same approach is of particular interest if applied to industries characterized by high customer variability. For instance, semiconductor production is the execution of what-if scenarios critical in terms of results accuracy (Sakr et al., 2019). Psarommatis and Kiritsis (2022) show how the implementation of a DSS can enhance the performance of the whole process: they propose a data-drive approach integrated with an ontology-based one focusing on supporting defect detection in a zero-defect manufacturing perspective. The validation of the proposed framework is provided through a case study. Similarly, at an operational stage, Gopalakrishnan et al. (2020) elaborate a data-driven DSS assessing the criticality state of the machine in the production process, aiming to improve the throughput time. They adopt simulation to test and validate the proposed case study. Considering the insights obtained by a first excursus on the literature, recurrent patterns and implication as well as the simulation of different case scenarios can provide useful support for decision makers. Hence, the existing literature conjugating these data-driven methods is detailed in the following.

2.2 Literature review on ARM and Simulation

A systematic literature review has been conducted to identify in which works simulation and ARM are jointly applied to the production environment. Scopus has been chosen as database and the only papers written in English are included in the analysis. In Table 1, the keywords and a summary of the total number of papers extracted from Scopus and the detail of the selected ones are provided.

Table 1 Summary of the keywords used in the literature review

Keywords	Number of papers	Relevant
“Data-driven” and “Association rule mining” and “Simulation”	8	2
"Simulation" and "Association rule mining" and "Framework"	29	5
"Discrete event simulation" and "Association rule"	23	6
"Simulation" and "Association rule mining" and "Fault"	29	16
"Simulation" and "Association rule mining" and "Scheduling"	13	2
"Simulation" and "Association rule mining" and "Production"	10	4

Several applications of ARM and simulation can be found in the literature focusing on different areas. For example, process traceability can be deployed by combining ARM and simulation, as exemplified by Lu et al. (2020). Indeed, the Association Rules (ARs) are used to track and identify the meat pigeon quality; the simulation ensures that the method performs well in real-time. In Glowacka et al. (2017b), instead, ARs are used to predict patients' no-shows in medical clinics. At the same time, through the simulation, their arrival is sequenced to optimize the performance of the clinic. In Xin Zhao et al. (2018b), ARM and simulation are used to define the optimal material handling trajectory on a manufacturing shop floor. The former technique mines the best trajectory from RFID data; the latter validates the approach. ARM can also be applied to understand simulation results in complex applications, such as geological applications, to investigate the impact of geological features on oil recovery and behavior in marine reservoirs (Suzuki et al., 2016).

Considering the energetic field, Abediniangerabi et al. (2020) propose a data-driven framework to design energy-efficient building façades, using ARM to understand energy-saving characteristics of the covering materials and simulating the thermal behavior of the building. Similarly, in Zhou et al. (2019), the ARs are mined to optimize the operational parameters of a chiller plant; through the simulation, the performance of the proposed approach is assessed. In Viet et al. (2020), instead, the two methodologies are applied to redesign the distribution process of an agri-food supply chain: the ARM is applied for the supplier selection, while the simulation is used for assessing the benefits of anticipatory shipping. ARM can be used for the automated diagnosis of faults. In Li et al. (2019), ARs between features and faults of rolling bearings are mined, and simulated scenarios are used to test the goodness of the approach.

ARM can also help identify the relationships between faults and alarms in communication networks (e.g., P. Liu et al., 2013; Zheng et al., 2020b) or distribution networks (Xiangwen Zhao et al., 2019b); moreover, simulation results in being a valuable tool to assess the effectiveness of the approach (Xiangwen Zhao et al., 2019a; Zheng et al., 2020a). Similarly, simulation tests the fault detection and classification capabilities of an adaptive neuro-inference system that uses ARM to identify the input dataset in a power network (Katooli & Koochaki, 2020). In contrast, in Chen et al. (2020), it is used to verify the data processing efficiency of an enhanced ARM method for smart substation fault diagnosis. A similar approach is proposed in Cai et al. (2017), where a real-time monitoring system for marine diesel engines based on support vector machine, ARM, and simulation is used for fault classification and relationship identification. In Dautov and Mosin (2018), ARM and k-nearest neighbors are used to create a classifier discriminating faulty behavior from fault-free ones, and simulation validates its accuracy.

The fault location problem is addressed in Leng & Li (2012), where parallel fuzzy ARM is proposed to identify the relationships within inner-and inter-domain alarms to define the fault's location in a large distributed communication network; simulation is used to validate the approach. In J. Liu et al., (2010), ARM, artificial neural networks, and simulation are combined to define the optimal maintenance interval by considering components' failure probability.

The existing contributions confirm that data-driven approaches support decision-making processes in several sectors. Extant literature jointly applying ARM and simulation is mainly focused on using the latter technique to validate the relationships identified through the ARM. Research opportunities concerning the use of simulation to conduct scenario analyses and iteratively define improvement actions are missed. Indeed, the proposed joint implementation of such techniques will allow not only to identify inefficiency through ARM validated through simulation, as most of the contributions in the literature, but also to quantify the improvement that can be obtain implementing different actions through scenario analysis. Finally, the joint use of ARM and simulation in dynamic but low-tech sector, like the fashion industry, is still missing.

3. Research approach

To overcome the literature limitations and reach the highlighted research opportunities, a 3-step framework that include the joint application of ARM and simulation techniques has been proposed. The general procedure, reported in Figure 1, will be described in the following sections. Each step of the framework is dedicated to a specific activity, i.e., data management, data analytics, and execution and control. The general procedure, reported in Figure 1, will be described in the following sections.

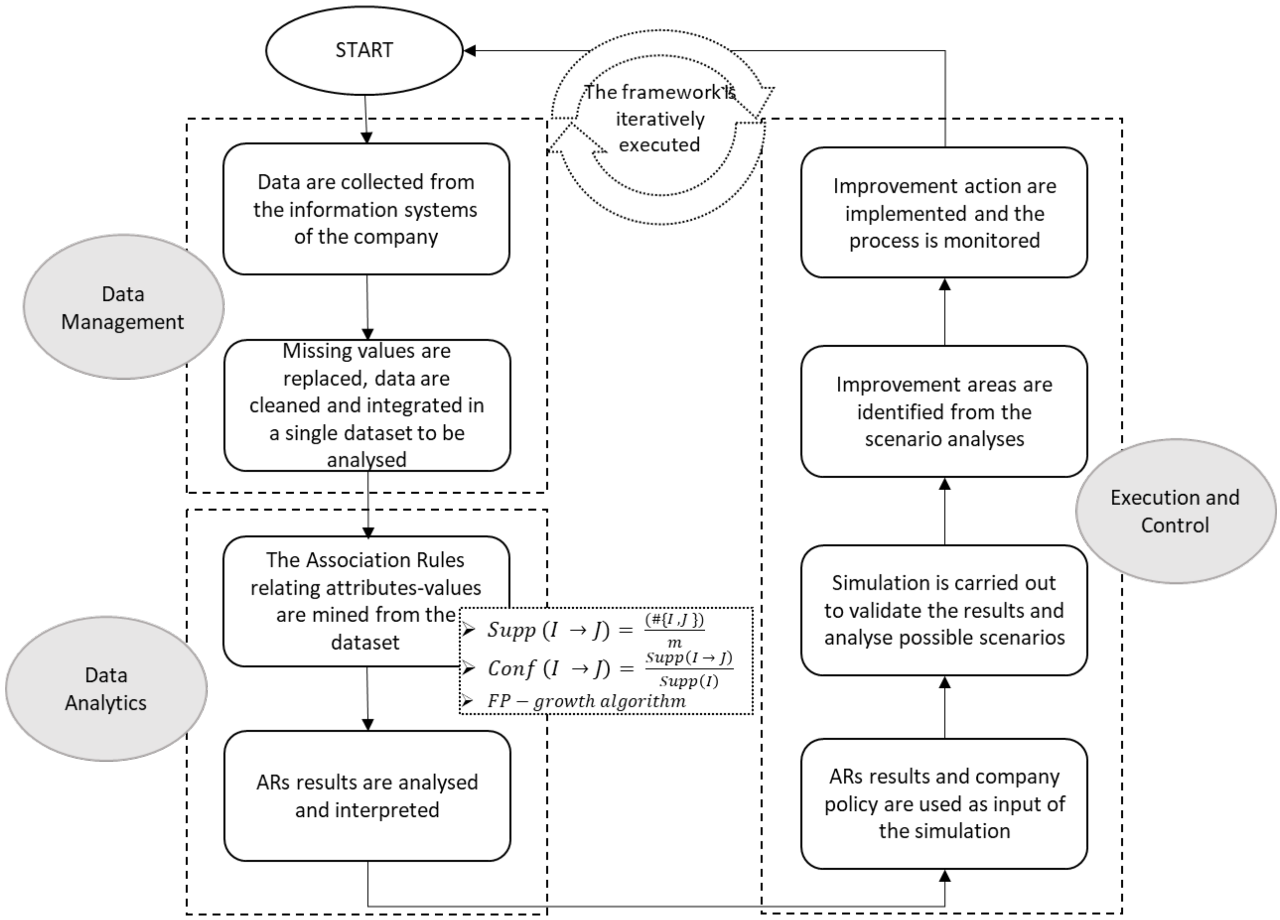

Figure 1 Proposed framework

3.1 Data management

The first step of the framework involves collecting and cleaning the data available to the company under investigation. The quality of the entire framework relies on this step. Indeed, any unnoticed inconsistency at this stage may reflect on the downstream process.

Different data sources should be considered to implement the proposed approach, for example, order lists, raw materials, production cycles, and plant stoppages. The correctness of the collections has to be ensured, different sources have to be integrated, and missing values should be replaced. The output of this step is a unified dataset integrating all the relevant sources.

3.2 Data analytics

The analytics of the integrated dataset is carried out in the second step of the framework. Specifically, the ARM involves identifying pairs of events frequently occurring together. Considering this assumption as a starting point and depending on the data sources available to develop the case study, specific goals can be set for the analysis, and the ARs are mined to address the specific issue. For example, the focus can be centered on identifying the relationships between fault causes in the production process and delays in the lead times or the impact of specific workflows on fault occurrences. The results obtained at this step, namely the ARs, can be used to arrange specific corrective actions to improve the process flow, its reliability, and, thus, the on-time product delivery.

More in detail, ARM is a data analytics approach whose objective is to find useful attribute-value relationships in large datasets (Buddhakulsomsiri et al., 2006) in order to extract hidden and previously unknown insights supporting the decision-making process.

Considering the following sets:

I. BD= {bd_1, bd_2,..., bd_n} be a set of boolean data called items;

II. T = {t_1, t_2, …, t_m} be a set of transactions, each of whom contains a subset of the items (hereafter, an itemset) taken from the set BD;

an Association Rule can be defined as an implication between a couple of itemsets taken from the set BD and cannot have any items in common:

$$\alpha \longrightarrow \beta, (\alpha, \beta \subseteq \mathrm{BD}, \alpha \cap \beta = \emptyset)$$

Several metrics can be used to measure the quality of an AR. In this work, the support (Supp) and confidence (Conf) are recalled:

- The support of the rule α —> β, $Supp\ (\alpha \rightarrow \beta) = \frac{(\#\{\alpha,\beta\})}{m}$, represents the probability of simultaneously finding α and β in a transaction, i.e., the joint probability;
- The confidence of the rule α —> β, $Conf\ (\alpha \rightarrow \beta) = \frac{Supp(\alpha \rightarrow \beta)}{Supp(\alpha)}$, represents the conditional probability of the rule, i.e., the probability of finding item β in a transaction containing α.

To mine the ARs, the procedure starts by defining the cut-off values to limit the solution space, i.e., minimum support and minimum confidence. Frequent Itemsets (FIs) meeting the support threshold have to be generated following the FP-growth algorithm proposed by Han et al. (2007). The identified frequent itemsets are combined to create the ARs, ensuring that only the ones having a confidence higher than the minimum confidence threshold are considered.

3.3 Execution and control

Once the dataset has been consolidated, and the ARs have been extracted, the final step of the framework includes the use of simulation for a twofold aim. On the one hand, it will support the validation of the results from applying the ARM. On the other hand, simulation will support decision-makers in conducting scenario analysis to define the optimal resource allocation. In terms of validation, a double-check of the correctness of the application of the ARM to the cleaned dataset will be done by simulating how the highlighted relations impact real processes. An in-depth analysis will be conducted to define the simulation approach that better fits with the company's purpose (i.e., Discrete Event Simulation - DES -, Agent-Based Modelling - ABM -, System Dynamics - SD -, Hybrid Simulation), even DES widely represents the most adaptable solution for on-field application to production processes. Therefore, the conduction of direct interviews with the company's operations managers represents the step 0 to identify the most adequate simulation approach to follow. As presented in Fani et al. (2022), these approaches can be implemented separately or even jointly following a hybrid perspective. The developed simulation model, indeed, could differ from one to another company according to the analyzed processes. For instance, single- and multi-assortment could be both managed by organizing the input datasheet in order to include "SKU", "phase", "sequence", and "time" to describe the production cycle and "SKU", "quantity", and "date" to list when they have to be processed: according to the company's production mix, a single SKU or several could be included in the production plan datasheet to generate agent entering the simulation model that will follow each own production cycle according to the "SKU" parameter they have. Similarly, the involved resources could be set to model different system configurations, such as job or flow shops, according to the "sequence" expressed in the SKU cycle. Beside the parameters set by the company (e.g., product mix,), the other inputs of simulation modeling are represented by the ARs results. Once the simulated model confirms the effectiveness of the ARM results as the first aim of simulation in the proposed framework, it could be used by the decision-maker to understand how changes in process configuration could impact the overall KPIs through scenario analyses.

Improvement areas could be therefore identified, considering the changes in input parameters that result in increased performance. For instance, comparing scenarios could support companies in choosing targeted actions to implement among several that look similar but impact differently. Once the selected action has been implemented, processes have to be monitored to track production performance and collect updated data.

For the simulation modelling, a data-driven approach has to be followed, defining a datasheet to be automatically read at the model start-up that includes all the data needed as input. Several simulators, like AnyLogic®, have in fact embedded database that will be filled with information about resources and flows according to a structured datasheet defined at the begin. It includes, on the one hand, the fixed input parameters defined by the company and, on the other hand, the results coming from the ARM that will change in terms of contents but not as structure after each iteration of the framework.

The data-driven approach to follow represents the key value of the proposed framework, guaranteeing its easy-to-readapt and iteratively application. While the type of system and process model to follow (e.g., DES or SD; multi- or single-assortment) is left as free choice, using a data-driven approach for simulation is mandatory for an effective application of the proposed framework. Besides the described one-shot implementation towards the definition of improvement areas and monitoring of the results, in fact, an iterative application of the framework has been proposed and can be easily implemented only using data-driven simulation. More in details, the evidences from the process monitoring will be used as input to integrate, or even redefine, the ARs that will be used for the next framework application, during which the data-driven simulation model will automatically read from the datasheet the updated results of the ARM. According to this, a data-driven approach will allow to define the simulation model once at the beginning and iteratively apply the framework as a decision-support tool.

4. Application

The described framework has been applied within a real scenario to demonstrate its usability and the main benefits that could be gained as a decision support tool. After a brief introduction to the company involved as case study, the detailed implementation of the proposed framework is described in the following paragraphs.

4.1 Case study

The developed framework has been applied to a fashion-luxury brand owner that operates in the leather goods market segment. The company's production processes can be resumed into 4 macro-steps: cutting, preparing components, assembling, and finally, quality check and packing. Cutting refers to the first step needed to realize a leather good and represents a critical process for its impact on the final product quality and cost. The second step refers to preparing components needed for the third phase, including cut leather, metal accessories, and other technical components. In fact, the assembly step to be smartly performed requires a previous kitting process to collect all the materials needed to realize each assembly unit included in the daily scheduling program. During the assembling step, activities such as stitching, lining, waxing, and polishing are performed according to the specific production cycle of the final product. For instance, the leather pieces are stitched together, and accessories are applied. The final step includes the quality check and the packing activities to avoid that defects could be found in the final products and to guarantee that all packaging materials, such as paper to keep the product shape, are included.

The described production steps could be performed following a make or buy strategy. More in detail, the analyzed company internally manages the cutting, quality check, and packing phases, while chooses between internal production and outsourcing for the preparing components and assembling steps. Moreover, the cutting phase could affect units to be internally processed or given to external companies, while the outsourced steps previously mentioned could be assigned to the same or different suppliers. The resulting six types of work orders managed by the analyzed company are shown in Figure 2. Beside the "Central Cutting" work order type that impacts only on the capacity of the cutting step, from the company point of view, internal or external production refers to work orders that are internally or externally assembled, respectively. According to this, the company splits the internal production into "Internal Production Int + Int" (i.e., internal preparing + internal assembling) and "Internal Production Ext + Int" (i.e., external preparing + internal assembling) work order types and the external production in "Different Supplier Int + Ext" (i.e., internal preparing + external assembling), “Different Supplier Ext + Ext” (i.e., external preparing + external assembling made by different suppliers) and “Same Supplier” (i.e., external preparing + external assembling made by the same supplier) work order types.

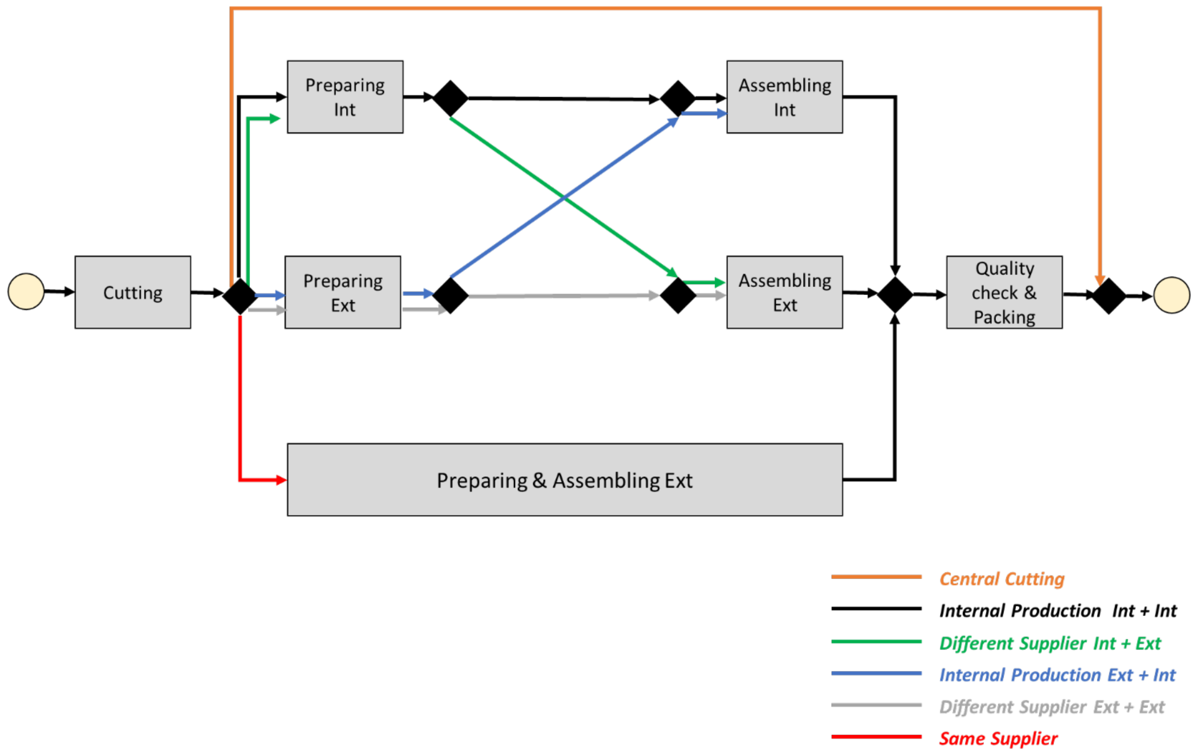

Figure 2: Production flows

Even if the quality check on the final product represents the last phase of the described production process, defects and other causes of stoppage could occur over all the steps. More in details, the cause of stoppage occurring in the preparing and assembling phases are concurrently collected and recorded at the end of assembly. The company manages 12 causes of stoppage that could be clustered into six macro-classes, as shown in Table 2. Due to privacy issues, causes of stoppage are anonymously reported.

Table 2 – Causes of stoppage (CoS)

Cause of stoppage	Cause of stoppage group
CoS_1	Feasibility
CoS_2	Feasibility
CoS_3	Engineering
CoS_6	Final product quality
CoS_7	Feasibility
CoS_9	Raw material quality
CoS_10	Raw material quality
CoS_12	Feasibility
CoS_14	Accessories quality
CoS_15	Final product quality
CoS_16	Feasibility
CoS_19	Workflow

The "Feasibility" stoppage group refers to elements which compromise the technical feasibility across the production line, for instance, due to the late arrival of materials or components needed to start a specific process. "Engineering" covers changes to the product industrialization that could involve even the bill of materials or the bill of labor. Quality issues could concern products at different stages, from raw materials to final products as well as components or accessories to be applied on (i.e., "Raw material quality", "Final product quality", "Accessories quality"). Even if the described stoppages refer to quality and result in delays due to reworks or replacements, they could be internally or externally generated. Finally, the "Workflow" cause of stoppage refers to missing or delayed approvals that could occur that often require the involvement of supervisors.

As widely known, one of the main critical success factors for companies belonging to the fashion industry refers to dealing with the small window for the products to be on the market, making the Lead Time (LT) a KPI to be constantly monitored and reduced as much as possible (Wang et al., 2020). For the analyzed company, minimizing LT means focusing on reducing production LT as the sum of the LT of the described steps. Moreover, stoppages have been identified as one of the main issues that have a negative impact on the LT, representing the aspect to be firstly investigated. According to this, the application of the proposed framework for the analyzed company has a twofold aim: on the one hand, to identify the most occurrent causes of stoppage for each production step; on the other hand, to evaluate the most impactful causes of stoppage in terms of delay on the production phase where they occur. The evidences will represent guidelines for the company to better understand where risk mitigation activities should be focused to minimize the production LT, reducing the occurrence of the most impactful causes of stoppage.

4.2 Data management

Considering the production process characteristics highlighted in the previous section, several data categories have to be merged to prepare the dataset for the analysis. For each product, the following data have to be extracted from the data management systems and integrated: customer orders have to be retrieved and integrated with the bill of materials and the bill of labor. In addition, information about the raw materials used to process each product (or batch) must be found so that any material quality defects can be linked to the specific supplier. It is also necessary to have a complete overview of the production process: the work order related to each product has to be noted, as well as the processing time for each phase, possible stoppage, and the related cause.

A cleaning process has been necessary to make the data suitable for the analysis. More specifically, the dataset initially constructed consisted of 2096 rows. 146 of these were removed due to lack of information on the LTs (null or negative values), while another 642 presented a null value under "work order".

An excerpt of the dataset used for the analysis is reported in Table 3: it collects the work order category for each production batch, the stoppage phase - also indicating whether it is internal or external - and its cause. Since the objective of the study is to identify the deviation from the expected LT due to stoppages of the process, Table 3 does not report the actual processing times per phase but only the deviations. For instance, in the first row, the work order type is “Same Supplier”, and a stoppage occurs during the Preparing & Assembling phase of the production process. The stoppage is charged to the cause *CoS_2* and implies a delay in the expected LT of 4.72 days on the same phase.

Table 3 Excerpt of the dataset created for the analytics phase

Work OrderType	StoppagePhase	CoS	DeltaLT_Cutting	DeltaLT_Preparing & Assembling	Deltalt_Quality check & Packing
Same Supplier	Preparing & Assembling	*CoS_2*	0.00	4.72	0.00
Different Supplier Ext + Ext	Preparing & Assembling	*CoS*_12	0.00	9.62	0.00
Same Supplier	Preparing & Assembling	*CoS_3*	0.00	2.72	5.14
Different Supplier Int + Ext	Preparing & Assembling	*CoS_2*	0.00	4.67	2.39
...	...	...	...	...	...

The next session shows the results obtained by applying the ARM to the dataset.

4.3 Data analytics

The analytics phase has been carried out, taking into account a twofold objective. On the one hand, the aim of the study regards the definition of a risk management plan related to identifying the causes of the stoppages. To this end, the ARs relating the work order category and the stoppage phase to the stoppage cause have been mined. In this way, it is possible to identify, given a work order and the phase, which is the most likely cause of stoppages.

The best minimum support and confidence thresholds are set through a set of experiments since they should be defined considering a trade-off not to lose relevant rules and not considering trivial ones (Koh, 2008). The relevant attributes are selected on the basis of the objective mentioned before: in Table 4, the tests carried out to define such thresholds are summarized. For each test, the hypothesized minimum support and confidence thresholds are reported, as well as the minimum values actually obtained from the ARM, the total number of rules, and the ones relevant for the study (i.e., the ones composed of all the considered attributes). In addition, the time required for the ARM process is added. From the table, it can be seen that for min_supp values lower than 0.001, the number of relevant rules and the associated parameters do not change. While increasing it above 0.001 causes a loss of relevant rules. Hence, the min_supp is set to 0.001, and the min_conf is set to 0.01.

Table 4 Experiments carried out to define acceptable min_supp and min_conf thresholds using an Intel Core i5-7200U CPU – [2.50GHz; 2.70 GHz] and 8.00 GB RAM.

Min_supp threshold	Min_supp value	Min_conf threshold	Min_conf value	Number of rules	Number of relevant rules	Min_conf of relevant rules	Duration
0.000001	0.000765	0.010	0.013	297	32	0.019	< 0.01 s
0.000765	0.000765	0.010	0.013	297	32	0.019	< 0.01 s
0.001	0.001529	0.010	0.013	287	32	0.019	< 0.01 s
0.005	0.005352	0.010	0.021	228	24	0.084	< 0.01 s

Table 5 shows some of the rules extracted to this end, while the extended is presented in Appendix 1: considering rows 1 and 2, it can be noticed that when the work order type is “Central Cutting” and the stoppage occurs on the cutting phase, then in the 66.67% of cases the cause of stoppage will be *CoS_7*, while in the remaining 33.33% will be *CoS_9*. This information can be used in order to identify the type of interruption of the productive flow, i.e., verifying which cause it deals with in order of confidence that represents occurrence probability. Considering the third row, instead, when a work order of category “Internal Production Int + Int” is processed on the cutting phase and a stoppage occurs, the cause of stoppage will be *CoS_7* in 100% (Confidence = 1) of cases. Despite the previously mentioned examples reporting high confidence values, they are characterized by low support values. This means that the joint probability of occurrence of the above conditions is quite rare, although the conditional probability is high. In Table 5, the rule associating Work Order Type = “Different Supplier Ext + Ext”, Stoppage phase = “Preparing & Assembling”, and Cause of stoppage = *CoS_2* has more relevant support (0.110), which stands for a higher joint probability to obtain the occurrence of the same condition. Given the work order category and the interruption flow phases, the confidence of the rule is 50.2%. Alternatively, the other expected causes of stoppage are *CoS_3* (33.1%) and *CoS_12* (16.0%). The remaining percentage is not classifiable since, in the current dataset, there is not a rule strong enough to be considered. For the aim of the current study, the confidence value is used to determine the probability of occurrence of a cause of stoppage, given the work order and the stoppage phase.

Table 5 Excerpt of the ARs relating work order type, stoppage phase, and cause of the stoppage

X	Y	Supp	Conf
Work Order Type = *Central Cutting*, Stoppage phase = *Cutting*	Cause of stoppage = *CoS_7*	0.034	0.667
Work Order Type = *Central Cutting*, Stoppage phase = *Cutting*	Cause of stoppage = *CoS_9*	0.017	0.333
Work Order Type = *Internal Production Int + Int*, Stoppage phase = *Cutting*	Cause of stoppage = *CoS_7*	0.005	1
Work Order Type = *Different Supplier Ext + Ext*, Stoppage phase = *Preparing & Assembling*	Cause of stoppage = *CoS_2*	0.110	0.502
Work Order Type = *Different Supplier Ext + Ext*, Stoppage phase = *Preparing & Assembling*	Cause of stoppage = *CoS_3*	0.073	0.331
Work Order Type = *Different Supplier Ext + Ext*, Stoppage phase = *Preparing & Assembling*	Cause of stoppage = *CoS_12*	0.035	0.160

On the other hand, the second objective pursued in this study is identifying the expected deviation from the lead time due to the occurrence of stoppage. Hence, the second group of ARs has been mined in order to relate the work order, cause, and phase of stoppage to the expected LT deviation (Table 7 - see Appendix 2 for the complete list) due to the check and reworking of the products. The rationale for setting the min_supp and min_conf thresholds is the same as in the previous case (Table 6).

Table 6 Experiments carried out to define acceptable min_supp and min_conf thresholds using an Intel Core i5-7200U CPU – [2.50GHz; 2.70 GHz] and 8.00 GB RAM.

Min_supp threshold	Min_supp value	Min_conf threshold	Min_conf value	Number of rules	Number of relevant rules	Min_conf of relevant rules	Duration
0.000001	0.0001	0.01	0.01	10057877	63	0.118	49 s
0.0001	0.0001	0.01	0.01	10057877	63	0.118	26 s
0.001	0.001209	0.01	0.01	4393625	63	0.118	14 s
0.005	0.005352	0.01	0.01	384551	52	0.118	1 s

LT deviations have been classified following the company object of the case study: the most suitable range considered was found to be five days (i.e., a workweek). It is interesting to note that the rule that links work orders "Central Cutting" and "Internal Production Int + Int" with a stoppage in the cutting phase of the production process implies a delay in the LT always limited to 5 days. Instead, when dealing with work order type "Same Supplier", a wider variability can be noticed: if the work order type is "Same Supplier", the phase is "Cutting" and the cause of stoppage is *CoS_7*, then the deltaLT will be within 5 days in 75% of cases, while in the remaining 25% it will be between 80 and 85 days. This second condition, while very severe, is actually less likely, according to the support (0.006 versus 0.019). It is, however, an eventuality to be taken into consideration when defining the delivery times of the produced lot.

Table 7 Excerpt of the ARs relating work order type, stoppage phase, cause of stoppage, and deltaLT

X	Y	Supp	Conf
Work Order Type = *Central Cutting*, Stoppage phase = *Cutting*, Cause of stoppage = *CoS_7*	DeltaLT_Cutting = [0 - 5]	0.264	0.955
Work Order Type = *Central Cutting*, Stoppage phase = *Cutting*, Cause of stoppage = *CoS_9*	DeltaLT_Cutting = [0 - 5]	0.138	1.000
Work Order Type = *Internal Production Int + Int*, Stoppage phase = *Cutting*, Cause of stoppage = *CoS_7*	DeltaLT_Cutting = [0 - 5]	0.044	1.000
Work Order Type = *Same Supplier*, Stoppage phase = *Cutting*, Cause of stoppage = *CoS_7*	DeltaLT_Cutting = [0 - 5]	0.019	0.750
Work Order Type = *Same Supplier*, Stoppage phase = *Cutting*, Cause of stoppage = *CoS_7*	DeltaLT_Cutting = [80 – 85]	0.006	0.250
Work Order Type = *Same Supplier*, Stoppage phase = *Preparing & Assembling*, Cause of stoppage = *CoS_14*	DeltaLT_Preparing & Assembling = [60 - 65]	0.001	1.000
Work Order Type = *Different Supplier Ext + Ext*, Stoppage phase = *Preparing & Assembling*, Cause of stoppage = *CoS_14*	DeltaLT_Preparing & Assembling = [45 - 50]	0.001	0.500
Work Order Type = *Different Supplier Ext + Ext*, Stoppage phase = *Preparing & Assembling*, Cause of stoppage = *CoS_14*	DeltaLT_Preparing & Assembling = [35 - 40]	0.001	0.500

The defined ARs are used as inputs to the execution and control phase of the research approach. In fact, they are useful in defining the probability of occurrence of each event when combined with discrete event simulation. For what concerns the rules presented in Table 9, a practical method for taking into account both the conditional probability of occurrence and the magnitude of the delay in lead time is to weigh the latter by confidence. In this way, we can correctly average each of the occurrences presented in column X (i.e., work order, stoppage phase, and cause of stoppage) for the delay they may entail. As an example, let us consider the rule relating Work Order Type = “Same Supplier”, Stoppage phase = “Preparing & Assembling”, Cause of stoppage = *CoS_14,* and the corresponding LT delay: the confidence indicates that in 100% of cases, the deltaLT associated with the occurrence of such conditions ranges between 60 and 65 days. In the DES step for the execution and control of the approach, the value associated with the rule will be 100*62.5 (i.e., the mean value of the deltaLT range). Instead, LT changes associated with Work Order Type = “Different Supplier Ext + Ext”, Stoppage phase = “Preparing & Assembling”, and Cause of stoppage = *CoS_14* belong to two different intervals: between 40 and 45 days, with 50.0% confidence, between 35 and 40 with the remaining 50.0% confidence. In this case, the expected deltaLT is calculated as follows:

$$\frac{40\ +\ 45}{2} * 50.0\%\ + \frac{35\ +\ 40}{2} * 50.0\%$$

The following section explains in detail how these results are capitalized to simulate the final steps of the proposed approach.

4.4 Execution and control

4.4.1 First iteration

Starting from the results of the ARs and the described production flows, a simulation model has been developed according to the peculiarities of the analyzed company. According to the case study, discrete event simulation has been used to model the described multi-assortment manufacturing process. First, the proposed framework has been applied in the presented real-case scenario in order to validate the results coming from the ARs in terms of stoppages occurrence and delays for reworking. The KPI used to confirm the model's reliability has been the production LT, considered by the company from the beginning of the cutting to the end of the quality check and packing processes. More in details, the developed DES model represents the six blocks shown in Figure 2 as black boxes. A data-driven approach has been followed, defining a datasheet to be automatically read on the model start-up that includes all the data needed as input. On the one hand, the datasheet has been filled according to the company characteristics, such as the managed work order types and their proportion within the order portfolio and the production flow they follow moving across internal or external stages. For instance, the flows to be simulated are listed on the datasheet detailing "work order type", "phase", "sequence", "LT". All the information are modeled in the simulator as written in the datasheet, allowing to run several scenarios updating the input datasheet without making any changes to the model. As the modeled stages will be the ones listed as "phase" and the production flow per work order type is expressed by the "sequence", any scenario the company wants to run, changing the stages each work order type follows, will only require to update "phase" and "sequence" on the datasheet. On the other hand, updates about stoppages occurrence and delays for reworking will be automatically acquired from the ARs results and organized in a datasheet with "work order type", "stoppage phase", "cause of stoppage", "cause of stoppage occurrence_alpha", "cause of stoppage occurrence_beta", "delta LT". According to this, each production step is represented as a black box, as shown in Figure 3, and the number of steps to be included in the simulation model is read from the datasheet. Work orders are generated in batches of 20 SKUs as written in the datasheet, which includes the information about the daily mix expressed as "work order type", "batch dimension", "quantity", "day". Once generated, work orders enter one production step from the "enter_From_Previous_Step" block of the simulation model, according to the work order cycle specified in the input datasheet. Once it enters a phase, it could wait to be processed in the "queue_Pre_Processed" block according to the processing capacity of the step (i.e., "Processing Capacity" parameter in Figure 3): if the number of in-step work orders has reached this value, the work order has to wait; otherwise it can be processed. The graphs "count_Processing_Capacity_Start" and "count_Processing_Capacity_End" in Figure 3 evaluate it and regulate the flow. According to the fact that production steps have been modeled as black boxes, work orders have then processed considering a processing LT (see the "in_To_Be_Processed", "processing_LT_Step", "out_Processed" and "queue_Processed" blocks in Figure 3). The processing LT for work order type related to each step is read from the input datasheet. The processed work orders could be stopped for several causes that are the object of applying the proposed framework for the analyzed company. On the one hand, the percentage of stoppages is communicated by the company and read in the “Stoppage_Perc” block from the datasheet. On the other hand, detailed information about stoppages, such as the occurrence of a single cause of stoppage and the LT required to solve it, is not directly managed by the company even if raw data are available. For instance, even if the company is confident that around 10% of work orders have to be stopped and checked within a step, the occurrence of a specific cause of stoppage is not perceived. To use that information as input of the simulation model and parameter to change in order to conduct scenario analyses, ARs have been used to rationalize raw data and include the occurrence of a single-cause of stoppage and the LT required to solve it in the input datasheet. As for the processing capacity, the number of work orders that presents a cause of stoppage that could be processed is limited by the “CoS_Mgmt_Capacity” communicated by the company and read from the datasheet. Once the work orders are ready for the next step (i.e., “queue_Ready” block in Figure 3), they move to it through the “exit_To_Next_Step” block. Finally, the time measurement elements “t_Start_Step” and “t_End_Step” are used to trace the production phase LT, including both the processing and the time needed to solve the cause of stoppage if occurred.

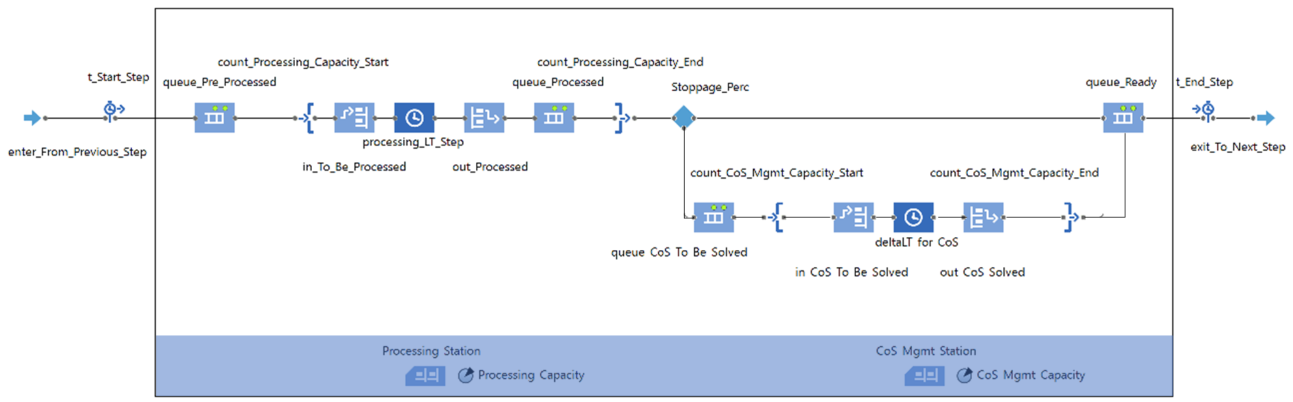

Figure 3 - Single-step detail

The production LT value obtained running the DES model using the results of ARs application as input is aligned to the one previously communicated by the company, confirming the reliability of the proposed model. More in details, the average production LT simulated and communicated by the company shows a less than 5% gap, moving from around 43 to 45 days. This difference has been considered acceptable by the company due to the simplifying hypotheses done in terms of percentage of work order types entering the simulation model, as well as related single-step LT and stoppage occurrence.

Once the model has been validated, the scenario analyses are carried on. Improvements to be evaluated have been defined starting from the results of ARs, which support the company in identifying the main criticalities. For instance, the causes of stoppages per work order type per production step have been ordered from the highest to the lowest in terms of priority, defined by multiplying the related conditioned probability of occurrence (i.e., the confidence of the rule) by the LT per reworking. These are, in fact, the parameters considered by the company to define priorities for action: on the one hand, the occurrence measure the probability a specific cause of stoppage could happen; on the other hand, how it impacts on the overall average production LT strictly depends on the related LT for reworking. The cause of stoppages to be firstly reduced should be the most occurring and impacting. According to the coefficient described above, the most critical causes of stoppages have been listed in Table 8.

Table 8 - Critical causes of stoppage in terms of occurrence and delays for reworking

Work order type	Stoppage phase	Cause of stoppage
Different Supplier Ext + Ext	Preparing & Assembling	CoS_12
Same Supplier	Preparing & Assembling	CoS_3
Different Supplier Ext + Ext	Preparing & Assembling	CoS_3
Different Supplier Ext + Ext	Cutting	CoS_7
Different Supplier Ext + Ext	Quality check & Packing	Cos_15
Different Supplier Ext + Ext	Quality check & Packing	CoS_3
Quality check & Packing	Preparing & Assembling	CoS_2
Quality check & Packing	Quality check & Packing	CoS_16
Different Supplier Ext + Ext	Preparing & Assembling	CoS_2
Same Supplier	Preparing & Assembling	CoS_2

Besides the previous assumption, the company has asked to evaluate how possible improvements will change if the causes of stoppage that mostly occur in production are considered the critical ones, excluding their impact in terms of delays for reworking from the priority identification. The most critical causes of stoppages in terms of occurrence have been listed in Table 9.

Table 9 - Critical causes of stoppage in terms of occurrence

Work order type	Stoppage phase	Cause of stoppage
Same Supplier	Preparing & Assembling	CoS_3
Internal Production Int + Int	Quality check & Packing	CoS_16
Different Supplier Ext + Ext	Preparing & Assembling	CoS_2
Internal Production Int + Int	Preparing & Assembling	CoS_2
Different Supplier Ext + Ext	Cutting	CoS_7
Different Supplier Ext + Ext	Preparing & Assembling	CoS_3
Central Cutting	Cutting	CoS_7
Different Supplier Ext + Ext	Quality check & Packing	CoS_3
Different Supplier Int + Ext	Preparing & Assembling	CoS_2

Scenario analyses have therefore been defined to support the company in terms of actions to implement. More in details, company has asked to compare two action types: to drastically reduce the most critical causes of stoppage in terms of only occurrence, as it represents the most criticalities perceived by the company itself, or considering the combined effect of occurrence and impact in terms of delays for reworking. In both scenarios, the actions to evaluate as future implementation for the company refer to reducing the occurrence of the most critical cause of stoppage. According to the evidences from Table 8 and Table 9, first scenario, therefore, includes the reduction of the average occurrence of *CoS_3* cause of stoppage by 50%, while the second one has the same improvement for the *CoS_2* cause of stoppage. Indeed, the company advised ignoring *CoS_12* and *CoS_16* since they are direct consequences of *CoS_*2 and will surely be interested in the improvement actions planned to address it. The results show that improvements on *CoS_3* cause of stoppage overcome the ones on *CoS_2*, reducing the average production LT from 45 days to 40 and 42 days, respectively. Even if significant benefits could be expected from the first improvement, quantifying the gap between production LT from the two scenarios could support the company in evaluating the trade-off between costs needed to reduce a specific cause of stoppage and the overall LT reduction. Comparing the results, the company has decided that the best trade-off could be reached by working on reducing the occurrence of the *CoS_2* cause of stoppage.

4.4.2 Second iteration

Once the planned actions have been implemented, a second iteration of the proposed framework has been conducted after six months in order to evaluate if expected benefits have been obtained and identify the new set of critical causes of stoppage according to a continuous improvement approach. Once data have been managed and ARs applied according to the first steps of the framework, the most critical causes of stoppages in terms of occurrence collected in the second iteration have been found as listed in Table 10.

Table 10 - Critical causes of stoppage in terms of occurrence: 2nd iteration

Work order type	Stoppage phase	Cause of stoppage
Different Supplier Ext + Ext	Preparing & Assembling	CoS_12
Internal Production Int + Int	Preparing & Assembling	CoS_3
Different Supplier Int + Ext	Preparing & Assembling	CoS_12
Different Supplier Ext + Ext	Quality check & Packing	CoS_16
Internal Production Ext + Int	Preparing & Assembling	CoS_3
Internal Production Int + Int	Quality check & Packing	CoS_3
Central Cutting	Cutting	CoS_10

The comparison between the results of the first and second iteration (see Table 9 and Table 10) shows that the actions implemented to mitigate the effects related to the *CoS_2* cause of stoppage have effectively been gained. Moreover, using as input occurrences and delays for reworking calculated by the application of ARs, the simulated average production LT shows a significant reduction, moving from 45 to 38,3 days, due to the combined effect of several improvements the company has implemented. Even in the second iteration, the LT resulting from the combined framework confirms its reliability, since it is aligned to the one communicated by the company after the implementation of the mitigation actions addressing *CoS_2*.

For the second framework iteration, scenario analyses have been differently defined. More in details, the company has asked to compare a new different set of actions, moving from improvements focused on the significant reduction of occurrences of causes of stoppage related to a single-cluster to slighter improvements that cover more than one cluster. According to this, the scenario analyses aimed to support the company in choosing the most performant set of initiatives to implement. More in details, the scenarios include different percentages of occurrence reduction (i.e., 20%, 50%, 80%) and the application of a single or more than one cluster. The results of the scenario analyses are shown in Table 11, where the average production LT per scenario (i.e., column “AVG LT Scenario_X [days]”) and the percentage gap compared to the scenario AS IS for the second iteration (i.e., column “AVG LT Scenario_X vs. Scenario 0 [%]”) are listed.

Table 11 - Results of scenario analyses: 2nd iteration

#	Scenario	AVG LT Scenario_X [days]	AVG LT Scenario_X vs Scenario 0 [%]
0	Scenario AS IS_2nd iteration	38,33	
1	Reduce_20%_Feasibility	37,91	1,1%
2	Reduce_50%_Feasibility	37,04	3,4%
3	Reduce_80%_Feasibility	36,90	3,7%
4	Reduce_50%_Engineering	37,35	2,6%
5	Reduce_50%_Raw material quality	38,05	0,7%
6	Reduce_50%_Final product quality	38,14	0,5%
7	Reduce_50%_Accessories quality	38,28	0,1%
8	Reduce_50%_Workflow	38,31	0,1%
9	Reduce_50%_All	35,58	7,2%

Starting from Table 11, the trade-off between the efforts needed and the obtainable results in terms of average production LT reduction has been evaluated. According to this, the company has defined a short-list to improvements for a second round of evaluation, including scenarios number 2 and 4. In fact, even though the last scenario shows the best improvements, the effort needed to reduce the occurrence of all the causes of stoppage by 50% has been considered not affordable for the company. Similarly, reducing the occurrence of feasibility clusters by 80% has been evaluated as unrealistic. Scenarios 2 and 4 therefore represent the best performant considering a medium effort on a single cluster. Comparing the efforts needed to reduce causes of stoppage related to feasibility and engineering clusters, the company evaluates to identify improvements on engineering issues more affordable because they mainly refer to internal criticalities, while feasibility problems often involve other supply chain actors (e.g., delay in material supply).

5. Discussion

5.1 Rationale behind the joint implementation of ARM and data-driven simulation

Considering the situation immediately before the implementation of the approach proposed in this article, the company did not implement any improvement actions in the presence of high delays. Basically, it was simply switching suppliers, and going through all the hassles involved (e.g., loss of time in finding a replacement and training, lack of a long-term relationship). In this sense, it could not be effectively solved by not identifying the actual cause of the problem.

Implementing only the ARs would result in an effective analysis from the point of view of identifying the conditions that lead to stop and delay causes and efficient (see, in this regard, the time values given in Table 4 and Table 6). On the downside, one would lose the dimension associated with improving the average times of the overall process. On the other hand, implementing only a simulation-based approach implies compromising the efficiency of the analyses, as one has to make assumptions to decide which statistics to calculate, as well as recognize outliers independently. The amount of time required to perform this step strictly depends on the number of statistics and performance indicators that have to be calculated. Of course, an indication of improvement can be obtained. Still, it is not a decision support tool that can be updated in a reasonable time interval. Therefore, such analysis is potentially effective but not efficient.

5.2 Theoretical and practical implications

Both theoretical and practical contributions can be extracted from the implementation of the framework. The amount of data characterizing current production processes surely requires introducing a tool that guides the analysis and ensures the extraction of meaningful knowledge. Hence, the sole adoption of traditional statistical techniques is no longer sufficient (Antomarioni et al., 2021), but new techniques and analyses are needed to improve the business decision-making processes (Hwang et al., 2004). To this end, adopting a data-driven approach combining ARM and simulation has shown promising results in supporting company managers during the decision-making process (Fani et al., 2021). The rationale behind the adoption of the ARM is due to the potentialities inherent in the technique. On the one end, it ensures the definition of patterns characterized by potentially unknown relationships, whose co-occurrence should be addressed to limit their impact on the production process. Conversely, it provides intuitive results that can be easily interpreted even by non-experts in the domain. This technique is widely applied in current literature and, as previously presented, the implementation areas are limitless. A further advantage from the adoption of the ARM is related to the possibility of analyzing the dataset before the formulation of research hypotheses. In this way, the search space for useful patterns remains unrestricted and leads to unbiased decision-making. Nonetheless, data quality represents a crucial aspect in the implementation of such a process: if data are not correctly managed, the analytics phase will be based on inconsistent datasets and, consequently, decisions are made based on misleading information that does not add any value to the current framework.

Appropriate actions are needed in this direction if the company has not developed the so-called "data culture" (Dubey et al., 2019) or, at least, to ensure that the value from a consistent data management is shared across all the organizational levels.

For what concerns the results obtained in the current application, it is evident that using the ARM, a prioritization of the improvement actions to undertake is provided. In this way, the scenarios tested and simulated focus on the most critical stoppage causes or phases, avoiding wasting time on activities that would bring more marginal improvements. Besides this, using data-driven simulation to validate the ARs results strengthen the model. Vice versa, ARs results may represent input for a simulation model that companies are not able to define, such as relations between variables that are not immediately identifiable looking at the data or basic statistics. In addition, being able to simulate different scenarios provides a clearer vision for the company of the expected results. In fact, even companies may identify the critical parameters that impact the KPIs; it is more difficult to choose the best performing among different similar actions. For instance, it is clear that reducing a cause of stoppage will improve company's performance, but it is not so easy to quantify the gainable benefit and to compare the reduction of two similar causes of stoppage in order to choose which of them to implement first. Also, the iterative nature of the framework allows a continuous improvement in terms of stoppage cause reduction and ensures that, gradually, improvement actions in all areas can be undertaken. In this regard, updating the dataset from which the analysis is carried out represents a critical aspect or, at least, one to be considered. In fact, the production processes continue to produce data, and one might decide to continue analyzing the entire database (i.e., going on to add new records to the existing set) or to eliminate old occurrences, creating a new set. This decision is closely linked to the corrective action taken. If the change does not concern the structural conditions of the production process, then it is possible to continue to integrate the dataset. If this action has been implemented effectively, we will have rules characterized by a lower, but not null, support. If, on the other hand, the change is deemed to result in a substantial change to the production process, then a new dataset must be created for analysis. This implies that some ARs may no longer be extracted, considering that the issue is resolved.

It should also be pointed out that other different issues can be addressed following the proposed framework: indeed, modifying the attributes composing the ARs, the focus of the execution and control measures can be driven to other aspects such as, for example, material quality or human error. Similarly, once the model has been validated, simulation can be used not only to check the mitigation actions to implement considering the current supply chain structure of the company but even to understand if other configurations could better respond to the most critical causes of stoppages. For instance, the case study has been focused on identifying the main causes of stoppage to reduce, but a different proportion of worker order types to internally or externally produce could also be evaluated.

6. Conclusions

The present study proposes a framework devoted to data-driven production planning and control, capitalizing on the combination of two widely known techniques, Association Rule Mining (ARM) and simulation. Specifically, the former is used to extract the co-occurrences of production stoppages per work order type, focusing on the specific stoppage causes and phases. The impact of such stoppages on the reworking lead time is also inquired. The Association Rules (ARs) mined are considered on the basis of these aspects so that improvement actions can be defined and, if appropriate, implemented. In order to quantify the benefits deriving from the improvement actions, a simulation of the modified process can be carried out. The scenario analysis is used to compare different possibilities and to assess the most convenient one, taking into account both the ARs mined and company policy. The results highlight which are the actions that should be undertaken at first in order to bring a major benefit to the process: having a clearer representation of its limitations or criticalities ensures a better consciousness during the planning and control phases.

The data-driven approach to follow is the key value for the effective application of the framework, making it easier to be iteratively run according to the updated ARs. If the first round requires the identification of the product parameters to investigate with ARs and the input needed for modelling the identified company's processes through simulation, the next rounds only require to automatically uploading the updated results from the ARM. In case of structural changes to the current production system, both in terms of ARs related attributes or process configuration, integrations are needed. For instance, product attributes not included before (e.g., the product area where the defect occurs) can be added to investigate new ARs, requiring minor changes in the ARM input data and in the model set. On the other hand, new process configuration (e.g., new resources or changes in the production cycle) could be managed by updating the datasheet read as input for the simulation modeling. However, the listed reasons to manually readapt the ARM and simulation settings reflect one-off changes, guaranteeing an easy daily application of the framework as a decision-support tool.

The beneficiaries of the implementation of the proposed framework are both the company itself and customers and suppliers: having better clarity, in fact, it will be possible to schedule production, orders, and deliveries in a more timely manner, without incurring material shortages or delays.

From a theoretical point of view, the main contribution can be represented by an iterative decision support tool based on ARM and simulation; specifically, the simulation is not only used to validate the results obtained through the ARM, as already identified in the literature, but also for the definition of improvement actions on critical areas, following an iterative process.

The major limitation of this work lies in the fact that the framework was tested on a single company and targeted a specific problem. Future developments of the approach, in fact, will address other companies and take into account different types of recurring co-occurrences.

References

Abediniangerabi, B., Shahandashti, S. M., & Makhmalbaf, A. (2020). A data-driven framework for energy-conscious design of building facade systems. *Journal of Building Engineering*, *29*, 101172. https://doi.org/10.1016/J.JOBE.2020.101172

Antomarioni, S., Ciarapica, F. E., & Bevilacqua, M. (2022). Data-driven approach to predict the sequence of component failures: a framework and a case study on a process industry. *International Journal of Quality and Reliability Management*. https://doi.org/10.1108/IJQRM-12-2020-0413/FULL/XML

Antomarioni, S., Lucantoni, L., Ciarapica, F. E., & Bevilacqua, M. (2021). Data-driven decision support system for managing item allocation in an ASRS: A framework development and a case study. *Expert Systems with Applications*, *185*, 115622. https://doi.org/10.1016/J.ESWA.2021.115622

Buddhakulsomsiri, J., Siradeghyan, Y., Zakarian, A., & Li, X. (2006). Association rule-generation algorithm for mining automotive warranty data. *International Journal of Production Research*. https://doi.org/10.1080/00207540600564633

Cagliano, A. C., Demarco, A., Rafele, C., & Volpe, S. (2011). Using system dynamics in warehouse management: A fast-fashion case study. *Journal of Manufacturing Technology Management*, *22*(2), 171–188. https://doi.org/10.1108/17410381111102207/FULL/XML

Cai, C., Weng, X., & Zhang, C. (2017). A novel approach for marine diesel engine fault diagnosis. *Cluster Computing 2017 20:2*, *20*(2), 1691–1702. https://doi.org/10.1007/S10586-017-0748-0

Chen, L., Wang, L., He, Q., & Liu, H. (2020). Fault diagnosis method of intelligent substation based on improved association rule mining algorithms. *Lecture Notes in Electrical Engineering*, *585*, 403–413. https://doi.org/10.1007/978-981-13-9783-7_32

Ciarapica, F., Bevilacqua, M., & Antomarioni, S. (2019). An approach based on association rules and social network analysis for managing environmental risk: A case study from a process industry. *Process Safety and Environmental Protection*, *128*, 50–64. https://doi.org/10.1016/j.psep.2019.05.037

Cipriano, M., Colomba, L., & Garza, P. (2021). A Data-Driven Based Dynamic Rebalancing Methodology for Bike Sharing Systems. *Applied Sciences 2021, Vol. 11, Page 6967*, *11*(15), 6967. https://doi.org/10.3390/APP11156967

Dautov, R., & Mosin, S. (2018). A technique to aggregate classes of analog fault diagnostic data based on association rule mining. *Proceedings - International Symposium on Quality Electronic Design, ISQED*, *2018-March*, 238–243. https://doi.org/10.1109/ISQED.2018.8357294

Dubey, R., Gunasekaran, A., Childe, S. J., Blome, C., & Papadopoulos, T. (2019). Big Data and Predictive Analytics and Manufacturing Performance: Integrating Institutional Theory, Resource-Based View and Big Data Culture. *British Journal of Management*, *30*(2), 341–361. https://doi.org/10.1111/1467-8551.12355

Fani, V., Bandinelli, R., & Rinaldi, R. (2017a). Optimizing production allocation with simulation in the fashion industry: A multi-company case study. *Proceedings - Winter Simulation Conference*, 3917–3927. https://doi.org/10.1109/WSC.2017.8248102

Fani, V., Bandinelli, R., & Rinaldi, R. (2017b). A simulation optimization tool for the metal accessory suppliers in the fashion industry: A case study. *Proceedings - 31st European Conference on Modelling and Simulation, ECMS 2017*, 240–246. https://doi.org/10.7148/2017-0240

Fani, V., Bindi, B., & Bandinelli, R. (2021). Designing And Optimizing Production In A High Variety/Low Volume Environment Through Data-Driven Simulation. *Proceedings - European Council for Modelling and Simulation, ECMS*, 10–15. https://www.scs-europe.net/dlib/2021/ecms2021acceptedpapers/0010_bpmi_ecms2021_0051.pdf

Fani, V., Pirola, F., Bindi, B., & Bandinelli, R. (2022). *Design Product-Service Systems by Using a Hybrid Approach : The Fashion Renting Business Model*. 1–16.

Glowacka, K. J., Henry, R. M., & May, J. H. (2017a). A hybrid data mining/simulation approach for modelling outpatient no-shows in clinic scheduling. *Https://Doi.Org/10.1057/Jors.2008.177*, *60*(8), 1056–1068. https://doi.org/10.1057/JORS.2008.177

Glowacka, K. J., Henry, R. M., & May, J. H. (2017b). A hybrid data mining/simulation approach for modelling outpatient no-shows in clinic scheduling. *Https://Doi.Org/10.1057/Jors.2008.177*, *60*(8), 1056–1068. https://doi.org/10.1057/JORS.2008.177

Gopalakrishnan, M., Subramaniyan, M., & Skoogh, A. (2020). Data-driven machine criticality assessment – maintenance decision support for increased productivity. *Https://Doi.Org/10.1080/09537287.2020.1817601*, *33*(1), 1–19. https://doi.org/10.1080/09537287.2020.1817601

Han, J., Cheng, H., Xin, D., & Yan, X. (2007). Frequent pattern mining: Current status and future directions. *Data Mining and Knowledge Discovery*. https://doi.org/10.1007/s10618-006-0059-1

Hwang, S. Y., Wei, C. P., & Yang, W. S. (2004). Discovery of temporal patterns from process instances. *Computers in Industry*, *53*(3), 345–364. https://doi.org/10.1016/J.COMPIND.2003.10.006

Kabadurmus, O., Kayikci, Y., Demir, S., & Koc, B. (2022). A data-driven decision support system with smart packaging in grocery store supply chains during outbreaks. *Socio-Economic Planning Sciences*, 101417. https://doi.org/10.1016/J.SEPS.2022.101417

Kaniappan Chinnathai, M., Alkan, B., & Harrison, R. (2021). A novel data-driven approach to support decision-making during production scale-up of assembly systems. *Journal of Manufacturing Systems*, *59*, 577–595. https://doi.org/10.1016/J.JMSY.2021.03.018

Katooli, M. S., & Koochaki, A. (2020). Detection and Classification of Incipient Faults in Three-Phase Power Transformer Using DGA Information and Rule-based Machine Learning Method. *Journal of Control, Automation and Electrical Systems 2020 31:5*, *31*(5), 1251–1266. https://doi.org/10.1007/S40313-020-00625-5

Khan, M. A. A., Butt, J., Mebrahtu, H., Shirvani, H., & Alam, M. N. (2018). Data-driven process reengineering and optimization using a simulation and verification technique. *Designs*, *2*(4), 1–22. https://doi.org/10.3390/designs2040042

Kim, J., Chung, B. Do, Kang, Y., & Jeong, B. (2018). Robust optimization model for closed-loop supply chain planning under reverse logistics flow and demand uncertainty. *Journal of Cleaner Production*, *196*, 1314–1328. https://doi.org/10.1016/J.JCLEPRO.2018.06.157

Koh, Y. S. (2008). Mining Non-coincidental Rules without a User Defined Support Threshold. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5012 LNAI*, 910–915. https://doi.org/10.1007/978-3-540-68125-0_92

Leng, X., & Li, X. (2012). Alarm fuzzy association rules parallel mining in multi-domain distributed communication network. *International Conference on Communication Technology Proceedings, ICCT*, 501–506. https://doi.org/10.1109/ICCT.2012.6511270

Li, Y., Wang, J., Duan, L., Bai, T., Wang, X., Zhang, Y., & Qin, G. (2019). Association rule-based feature mining for automated fault diagnosis of rolling bearing. *Shock and Vibration*, *2019*. https://doi.org/10.1155/2019/1518246

Liu, J., Ji, H. P., & Zhu, Q. X. (2010). One Method to Determine the Optimal Maintenance Time Based on Selective Attrition. *Advanced Materials Research*, *139–141*, 2578–2581. https://doi.org/10.4028/WWW.SCIENTIFIC.NET/AMR.139-141.2578

Liu, P., Li, X. M., & Feng, Y. G. (2013). An algorithm of multi-level fuzzy association rules mining with multiple minimum supports in network faults diagnosis. *Proceedings - International Conference on Natural Computation*, 884–888. https://doi.org/10.1109/ICNC.2013.6818101

Lu, X., Liu, P., Nie, W., & Zhang, H. (2020). Whole process tracing model of pigeon quality in block chain environment. *Proceedings - 2020 International Conference on Computer Information and Big Data Applications, CIBDA 2020*, 435–439. https://doi.org/10.1109/CIBDA50819.2020.00104

Luo, J., Peng, Q., Wen, C., Wen, W., & Huang, P. (2022). Data-driven decision support for rail traffic control: A predictive approach. *Expert Systems with Applications*, *207*, 118050. https://doi.org/10.1016/J.ESWA.2022.118050

Parhizkar, T., Hogenboom, S., Vinnem, J. E., & Utne, I. B. (2020). Data driven approach to risk management and decision support for dynamic positioning systems. *Reliability Engineering & System Safety*, *201*, 106964. https://doi.org/10.1016/J.RESS.2020.106964

Patwardhan, A., Verma, A. K., & Kumar, U. (2016). A Survey on Predictive Maintenance Through Big Data. *Lecture Notes in Mechanical Engineering*, 437–445. https://doi.org/10.1007/978-3-319-23597-4_31

Psarommatis, F., & Kiritsis, D. (2022). A hybrid Decision Support System for automating decision making in the event of defects in the era of Zero Defect Manufacturing. *Journal of Industrial Information Integration*, *26*, 100263. https://doi.org/10.1016/J.JII.2021.100263

Ren, Z., Han, X., Yu, X., Skjetne, R., Leira, B. J., Sævik, S., & Zhu, M. (2023). Data-driven simultaneous identification of the 6DOF dynamic model and wave load for a ship in waves. *Mechanical Systems and Signal Processing*, *184*, 109422. https://doi.org/10.1016/J.YMSSP.2022.109422

Sakr, A. H., Yacout, S., & Bassetto, S. (2019). A Discrete Event Simulation logic for Semiconductor Production Planning and Control within Industry 4.0 Paradigm. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 172–182.

Suzuki, S., Stern, D., & Manzocchi, T. (2016). The Use of Association-Rule Mining and High-Dimensional Visualization To Explore the Impact of Geological Features on Dynamic-Flow Behavior. *SPE Journal*, *21*(06), 1996–2009. https://doi.org/10.2118/174774-PA

Viet, N. Q., Behdani, B., & Bloemhof, J. (2020). Data-driven process redesign: anticipatory shipping in agro-food supply chains. *International Journal of Production Research*, *58*(5), 1302–1318. https://doi.org/10.1080/00207543.2019.1629673

Wang, B., Luo, W., Zhang, A., Tian, Z., & Li, Z. (2020). Blockchain-enabled circular supply chain management: A system architecture for fast fashion. *Computers in Industry*, *123*, 103324. https://doi.org/10.1016/J.COMPIND.2020.103324

Xu, X., Guo, W. G., & Rodgers, M. D. (2020). A real-time decision support framework to mitigate degradation in perishable supply chains. *Computers & Industrial Engineering*, *150*, 106905. https://doi.org/10.1016/J.CIE.2020.106905

Zhao, Xiangwen, Luo, L., Ma, G., Cai, Z., Gu, Z., & Wang, Q. (2019a). Operation and Maintenance Management and Decision Analysis in Distribution Network Based on Big Data Mining. *2018 International Conference on Power System Technology, POWERCON 2018 - Proceedings*, 4855–4861. https://doi.org/10.1109/POWERCON.2018.8601825

Zhao, Xiangwen, Luo, L., Ma, G., Cai, Z., Gu, Z., & Wang, Q. (2019b). Operation and Maintenance Management and Decision Analysis in Distribution Network Based on Big Data Mining. *2018 International Conference on Power System Technology, POWERCON 2018 - Proceedings*, 4855–4861. https://doi.org/10.1109/POWERCON.2018.8601825

Zhao, Xin, Meng, H., Zhang, W., Li, X., & Ren, S. (2018a). An Association Rule Mining Approach for Shop Floor Material Handling Based on Real-Time Manufacturing Big Data. *Advances in Intelligent Systems and Computing*, *885*, 706–713. https://doi.org/10.1007/978-3-030-02804-6_92

Zhao, Xin, Meng, H., Zhang, W., Li, X., & Ren, S. (2018b). An Association Rule Mining Approach for Shop Floor Material Handling Based on Real-Time Manufacturing Big Data. *Advances in Intelligent Systems and Computing*, *885*, 706–713. https://doi.org/10.1007/978-3-030-02804-6_92

Zheng, Q., Li, Y., & Cao, J. (2020a). Application of data mining technology in alarm analysis of communication network. *Computer Communications*, *163*, 84–90. https://doi.org/10.1016/J.COMCOM.2020.08.012

Zheng, Q., Li, Y., & Cao, J. (2020b). Application of data mining technology in alarm analysis of communication network. *Computer Communications*, *163*, 84–90. https://doi.org/10.1016/J.COMCOM.2020.08.012

Zhou, X., Wang, B., Liang, L., Yan, J., & Pan, D. (2019). An operational parameter optimization method based on association rules mining for chiller plant. *Journal of Building Engineering*, *26*, 100870. https://doi.org/10.1016/J.JOBE.2019.100870