Task allocation with tardiness minimization for maintenance delivery of Smart Product-Service Systems

Sala R.*, Pinto R.*, Pirola F.*, Pezzotta G.*

* *University of Bergamo, Department of Management, Information and Production Engineering, Viale Marconi, 5, Dalmine (BG), 24044, Italy (roberto.sala@unibg.it, roberto.pinto@unibg.it, fabiana.pirola@unibg.it, giuditta.pezzotta@unibg.it)*

Abstract: Data collection, analysis, and exploitation are among the most discussed topics in current researches related to Industry 4.0, especially as far as production themes are concerned. Some authors have discussed the benefits that could originate from the exploitation of product functioning data for maintenance delivery in the Product-Service System (PSS) field. Despite this, a comprehensive approach considering both the asset and the service perspectives for the improvement of the service delivery process is still missing. The authors try to deal with the service delivery decision-making problem presenting a task allocation model aimed at minimizing the total tardiness in maintenance delivery. The model considers actual information from both the asset (e.g. the Residual Useful Life (RUL) of a component) and the service (e.g. the operator's calendar, the various resolution approaches available and the mean time to repair with a specific approach, customer features) to match the tasks and the operators. The paper describes the model development, discussing the benefits of data exploitation in the decision-making process. Moreover, the role and benefits of such a model in a data-driven decision-making process for PSS delivery are discussed. Finally, the paper presents the model limitations and its possible extensions, considering additional constraints and different objective functions.

Keywords: Maintenance, Product-Service Systems, decision-making, task allocation.

1. Introduction

In the current servitization scenario, the efficient delivery of services represents an important instrument for manufacturing companies who want to create long and strong relationships with their customers (Kindström & Kowalkowski, 2014). One of the effects of the spread of the Industry 4.0 programs in the manufacturing sector is the possibility to generate, share and analyse unprecedented quantities of data related to the production process and the behaviours of the assets (Cattaneo et al., 2018; Moghaddass & Zuo, 2014). This data could be used to monitor the productivity but also to detect and predict failures in the production process, which could be related to the process structure or problems in the machines (Tamilselvan & Wang, 2013), and tackle them in a promptly (Hu et al., 2012). In particular, one of the most discussed approaches to data analysis concerns the usage of Machine Learning in the scope of predicting or classifying machines status (Kotsiantis et al., 2006). Therefore, data exploitation could play a pivotal role in the servitization journey of manufacturing companies (Bagozi et al., 2017; Coreynen et al., 2015).

Asset monitoring is not the only factor concurring in guaranteeing efficient productivity of the machines, but also the structure of the service offering, such as maintenance delivery, plays a fundamental role (Rahman et al., 2017). The effectiveness of maintenance delivery depends upon the definition of proper maintenance policies, which usually are established after the company knowledge on the behaviours of the assets and, thus, on the company experience (Potes Ruiz et al., 2013). Similarly, decisions at the operational level, such as the allocation of the maintenance intervention to the operator, are performed with limited support of specific tools (Gopalakrishnan et al., 2015). Data gathering and analysis results to be not fully exploited if decision-makers rely mainly on their experience instead of the data from the field.

The authors have proposed a framework for data-driven maintenance delivery that, through the exploitation of data related to the service and the asset, intend to improve the maintenance delivery decision-making phase (Sala, Pirola, et al., 2019). The framework proposes a structured approach to data collection and analysis and also specific tools to support these activities. One of the tools under development in the framework is aimed at supporting the planner in the task allocation phase. This paper proposes an initial study on the task allocation phase through the definition of an offline optimization model that, considering asset and service-related data, aims at minimizing the total tardiness of the interventions.

The paper is structured as follows: Section 2 deals with a literature review on the topics of service delivery and task allocation. Section 3 describes the general problem, its formulation and the following analysis on the model computational time. Section 4 discusses the possible extensions of the current model. Eventually, Section 5 concludes the paper delineating the future steps in the research.

2. Literature Review

2.1 Maintenance Delivery Process

When it comes to creating economically sustainable PSS, the service delivery aspect, which is constituted by all those actions performed in back-office or front-office by the customer and the employees to allow the correct delivery of a specific service (Mathieu, 2001), cannot be neglected. According to the previous definition, service delivery encompasses a series of activities and relations between actors that should guarantee a satisfactory result for the customer and the supplier. This is not always true, many times companies are not able to guarantee a satisfactory delivery of their services because of problems distributed along the process (Sala, Pezzotta, et al., 2019). These problems could originate from different sources like the activities performed by the actors, the interactions between them, the decisions that the actors make, the information flow inside the process or others (Rondini et al., 2018). Each of these factors, combined with others like resource availability, which usually are rather limited, contributes to influencing the way service delivery happens. (Sala, Pirola, et al., 2019) describes a framework aimed at improving maintenance delivery proposing a structured data-driven decision making-process that, considering data collected from the asset and service can support the planner in organizing the delivery process by allocating the intervention requests to the operators. In particular, the framework proposes to establish a strategy for data collection and analysis for the asset and service so that it would be possible to create a continuous flow of information that if properly managed, can lead to the optimal allocation of the maintenance intervention requests.

In the case of maintenance, the exploitation of the knowledge created by the company over the years constitutes the base on which constructing a reliable maintenance engineering, able to define suitable approaches for the prevention of unexpected failures (Ruiz et al., 2014). Depending on the asset and on the specific component, different maintenance policies and approaches could be adopted, resulting in a different way to manage and deal with each failure (Andreacchio et al., 2016). For example, with failures that to be fixed require expensive components and/or prolonged downtimes, it would be advisable to apply preventive and/or proactive maintenance policies. On the other side, for failures repairable in a short time, without the need for expensive components, it would be advisable to adopt corrective policies. It is important to evaluate the way assets and components are monitored and maintained. The frequent substitution of the components may result in an inefficient way to manage the asset maintenance resulting thus in the substitution of components before the end of their Residual Useful Life (RUL). This is not optimal from the economical point of view since it implies higher expenditures for maintaining the components compared with the necessary ones (Andreacchio et al., 2016). Thus, the usage of process data, as proposed in (Sala, Pirola, et al., 2019), could be useful in supporting analysis to determine the health status of certain components and intervene promptly.

In light of the spread of Industry 4.0, software, data collection and analysis become a major topic of discussion for practitioners and researchers working on maintenance delivery. The problem has been approached mainly from the asset monitoring aspect, with many papers dealing with the use of Machine Learning techniques to predict the components failure (Carvalho et al., 2019; Chopra & Priyadarshi, 2019; Ruiz-Sarmiento et al., 2020). The introduction of an approach able to monitor the status of the components and predict their failure in advance can have positive consequences on maintenance intervention scheduling, which strongly influences customer satisfaction. Effective results on customer satisfaction could be achieved only if the scheduling approach is coherent with the content of the service contracts established with customers (e.g. minimization of the tardy tasks if there is a penalty for the late completion of the maintenance interventions). The development of such a schedule requires to consider information from both the service and the asset since the joint use of information on the asset status and on the service operators' competences and performance could help in the intervention allocation. Thus, the need for an approach considering these aspects emerges.

2.2 Task allocation

The definition of an effective schedule for operators able to satisfy all the customers and maximising the service supplier is a tricky task. Many variables, going from operators' calendar, to travel costs, to travel time and others, has to be taken into account when making decisions related to this topic. Frequently maintenance planners are requested to work with limited resources, evaluating and prioritizing the customers' requests. To do so, planners need to take into account several aspects and consider multiple variables, relying, in the majority of cases, on their experience and knowledge to make decisions related to the allocation of the interventions to the operators (Gopalakrishnan et al., 2015).

Many authors have discussed the problem of the task allocation over the years (Agnihothri & Mishra, 2004; Anim-Ansah et al., 2006; Hurkens, 2009; Lagemann et al., 2014; Xu et al., 2014), proposing different approaches aimed at maximising or minimising different objective functions depending on the problem under investigation. An example of this can be found in (Pal et al., 2017), which propose an optimization model for maintenance scheduling in process industries aimed at optimizing the production of the company while considering maintenance and power-balance constraints in the production units. Another example can be found in (Huang et al., 2018), which describes an optimization model that, in the light of Industry 4.0 data collection possibilities, is focused on the minimization of maintenance costs thanks to the exploitation of process data and health status data. (Agnihothri & Mishra, 2004) proposes an optimization model discussing the effects of cross-training on efficient and cost-effective service delivery. Similarly, (Xu et al., 2014), in their model, show the importance of cross-training for operators when it comes to optimize the service delivery reducing the provision cost and maximise the customer satisfaction. Besides operators' skills, other inputs should be considered during the operators scheduling. For example, (Anim-Ansah et al., 2006) discusses, in addition to the importance of the operators' skills, the influence of travel time and operators' availability on the schedule definition while trying to maximise the service quality.

At the operative level, the presence of an effective decision-making tool to support the planner may be decisive for the effective allocation of the interventions, especially when it comes to considering the time criticality aspect. This aspect affects not only customer satisfaction but also its productivity. In the case of corrective maintenance, the prompt allocation of an operator to the customer requests may be fundamental to reduce the machine downtime and allow the restart of the production. The integration of service data with information on the assets health status mixed with the usage of proper decision-making tools could improve the intervention allocation phase (Sala, Pirola, et al., 2019).

Thus, this paper tries to continue the research on maintenance delivery proposing a preliminary version of an optimization model that, integrating the service and asset information is aimed at improving the allocation of the intervention requests to the operators.

3. Problem Definition and Formulation

This section is devoted to the proposal and discussion of an optimization model that, considering information collectable from the service and the asset, proposes the allocation of intervention requests to the operators. Under the hypothesis of the existence of service contracts between the supplier and the customers, the model aims to allocate the intervention to the operators in a way that minimizes the interventions that end later than established in the contract and, thus, minimize the total tardiness.

3.1 Definition

Let us consider a set K of intervention requests required to Company A to be executed in the following week. At the end of the current week, the planner collects all the requests K to fulfil the allocation task. Since the complete pool of requests to be allocated is known before the beginning of the activity, this can be defined as an offline allocation problem.

Each intervention request in K can be fulfilled in more than one way, depending on the request itself. Resolution methods can span from remote support to on-field intervention performed sending an operator to the calling customer's facility. The resolution method depends upon the type of intervention request and the experience of the customer in handling the failures.

To be fulfilled, each task requires different skills, time and costs. Therefore, the service provider has to select the task for each intervention that better satisfies the customer's request minimizing the costs for the provider and maximising the customer satisfaction by respecting the constraints established in the contract. In the scenario analysed by the current model, the costs are mainly related to the possibility of unmet the SLA, as the other costs are covered by the service contract established between the stakeholders.

3.2 Formulation

We formulate the problem as a parallel machine scheduling problem, considering the operators as machines running in parallel to perform tasks. The allocation of interventions to operators is similar to the problem of allocating jobs to (unrelated) parallel machines. We address the parallel machine problem described above using the assignment and positional variables (Unlu & Mason, 2010). Thus, in the proposed formulation, the typical schedule structure at the beginning of the week is represented by a series of empty "positions" p available in the schedule of each operator. We refer the reader to the paper by Unlu & Mason (2010) for the details. The model in Unlu & Mason (2010) is used as a starting point to formulate the management of the operators' calendar while other variables have been considered to structure the analysis. As introduced in Section 2, the model is a part of a service- and asset- related framework supporting the improvement of maintenance delivery and, thus, uses as input historical and real-time information from both sides, finding in this its originality.

To formulate the problem described in Subsection 3.1 we make use of the following notation:

- K: set of intervention requests;
- M_k: set of tasks (i.e. resolution interventions) that satisfy the request $k \in K$;
- O: set of available operators;
- P: set of positions for the intervention requests allocation.

Each intervention request $k \in K$ defines a set of tasks M_k that fulfil the request. In our model, we make the following assumptions:

- Each operator $j \in O$ can execute every task (i.e. has all the required skills). Operators differ in the execution time of the tasks, which depends on the operators' experience (e.g. one operator performs better with the on-site interventions while the other performs better with the remote support. Both can carry out the two tasks but their execution time is different);
- At the time of requests allocation, the schedule of the operators are blank, no tasks are assigned;
- To execute the tasks, all the operators depart and return to the headquarters before going to the next customer.

Let us define the following parameters:

- DD_k: due date of intervention request k, defined as $\min\{SLA_k; RUL_k\}$, where SLA_k is the date before which the request k must be fulfilled according to the service level agreement agreed with the customer, and RUL_k is the residual life before the breakdown of the component associated to request $k \in K$;
- t_{km}^{TOP}: travelling time for the operator to reach (and come back from) the location of the intervention request k addressed with the task $m \in M_k$. Since all the operators depart and return to the headquarters, this value is not dependent on the single operator $j \in O$;
- t_k^{SS}: time to get the spare parts in place for the execution of task $m \in M_k$ fulfilling the intervention request $k \in K$. This time is dependent only upon the intervention request k, because it is the intervention required that determines the necessity of spares;
- t_{kmj}^{INT}: time required to perform task $m \in M_k$ by the operator $j \in O$;
- M: a constant, large number for modelling purpose.

Finally, we make use of the following variables:

- γ_{pj}= end date of the task scheduled in the *p*-th position for operator $j \in O$;
- C_k= completion time of intervention $k \in K$;
- $T_k = \begin{cases} C_k - DD_k \ if \ C_k > DD_k \\ 0 \ otherwise \end{cases}$ = tardiness

Each intervention request has to be allocated to a single operator, and all the tasks have to be allocated. The allocation of the intervention requests is done using a binary variable x_{kmjp} equal to 1 if intervention requests $k \in K$ is assigned to the position p of the schedule of the operator $j \in O$ who is tasked with task $m \in M_k$, and 0 otherwise.

Thus, the problem is formulated as follows:

$$\min Z = \sum_{k \in K} T_k \tag{1}$$

$$\sum_{\substack{k \in M_k \\ j \in O \\ p \in P}} x_{kmjp} = 1 \quad \forall k \in K \tag{2}$$

$$\sum_{\substack{k \in K \\ m \in M_k}} x_{kmjp} \leq 1 \quad \forall j \in O, p \in P \tag{3}$$

$$\gamma_{pj} \geq \sum_{\substack{k \in K \\ m \in M_k}} \left(t_{kmj}^{INT} + t_{km}^{TOP}\right) \cdot x_{kmjp} \quad \forall j \in O, p = 1 \tag{4}$$

$$\gamma_{pj} \geq \gamma_{(p-1)j} + \sum_{\substack{k \in K \\ m \in M_k}} \left(t_{kmj}^{INT} + t_{km}^{TOP}\right) \cdot x_{kmjp} \quad \forall j \in O, p \in P \backslash \{1\} \tag{5}$$

$$C_k \geq \gamma_{pj} - M\left(1 - x_{kmjp}\right) \ \forall k \in K, m \in M_k, j \in O, p \ in \ P \tag{6}$$

$$T_k \geq C_k - DD_k \ \forall k \in K \tag{7}$$

$$C_k - \sum_{\substack{m \in M_k \\ j \in O \\ p \in P}} t_{kmj}^{INT} \cdot x_{kmjp} \geq \sum_{\substack{m \in M_k \\ j \in O \\ p \in P}} \frac{t_{kmj}^{TOP}}{2} \cdot x_{kmjp} \ \forall k \in K \tag{8}$$

$$C_k - \sum_{\substack{m \in M_k \\ j \in O \\ p \in P}} t_{kmj}^{INT} \cdot x_{kmjp} \geq t_k^{SS} \ \forall k \in K \tag{9}$$

$$T_k \geq 0 \quad \forall k \in K \tag{10}$$

$$x_{kmjp} \in \{0,1\} \ \forall k \in K, \forall m \in M_k, j \in O, p \in P \tag{11}$$

$$\gamma_{pj} \geq 0 \ \forall p \in P, j \in O \tag{12}$$

$$C_k \geq 0 \quad \forall k \in K \tag{13}$$

The objective function (1) minimizes the total tardiness. In the hypothesis of the definition of a service contract based on the SLA with penalties due to late interventions, the service provider needs to minimize the number of tardy interventions delivered to customers. Such an objective is relevant in the considered case because it minimizes the number days late and, thus, reduces the penalties that the company has to face. Moreover, tardiness results to be an important factor in consideration of the machine productivity, which results to be furtherly reduced in case of tardy maintenance interventions. In this sense, it can be noted that tardiness is also related to costs, expenses, and profitability of a company. Constraint set (2) stipulates that each intervention request is allocated exactly once, whereas constraint set (3) guarantees that each position on every slot contains at most one task. Constraint set (4) and (5) defines the completion time of the task in position p for the operator j. Constraint set (6) defines the completion time of intervention k, whereas constraint set (7) sets the tardiness variable T_k. Constraint sets (8) and (9) stipulate that the start of the intervention must occur after the spares are delivered to the location and after the operator has arrived. Finally, constraint sets from (10) to (13) define the domains of the variables.

3.3 Numerical analysis of computational time

The performance of the model should be tested and assessed against the current process and methods commonly adopted in practice. However, at this point of the research, we are more interested in understanding the general characteristics of the problem concerning its practical implementation and adoption in a real service planning process. Due to space limitation, we focus now on the analysis of the average computational time that the model requires to achieve a solution. The solution time of a model may hinder of favour its practical adoption in a real context. If the running time is too high, it may be difficult to use such a model in a day-by-day process, where it may be necessary to change the considered data or change the planning because of disruptions and issues occurring in the network (i.e. delay in task execution, problems with the spare parts, last-minute operator related issues).

With this scope, the authors formulated the problem and solved multiple instances considering a growing number of intervention requests and the number of operators. Random instances have been used to solve the problem. In particular:

- Number of intervention requests: from 50 to 300, with increments of 50;
- Number of operators: from 10 to 100, with increments of 10;
- Number of tasks: 3.

For each pair (intervention requests and operators), the authors solved 5 instances, for a total of 120 overall.

Figure 1 reports the average time required to solve the model using an Intel® Core™ i5-7200 CPU @ 2.50 GHz, 2 core, using Cplex12. Analysing Figure 1 it is possible to notice how the resolution time starts growing consistently when the number of interventions is higher than 150 and the number of operators is higher than 40. Besides, the resolution time grows exponentially when the number of intervention requests is higher than 250 and the number of operators is higher than 70. Thus, the current allocation model may be suitable for problems that have a limited number of variable to handle.

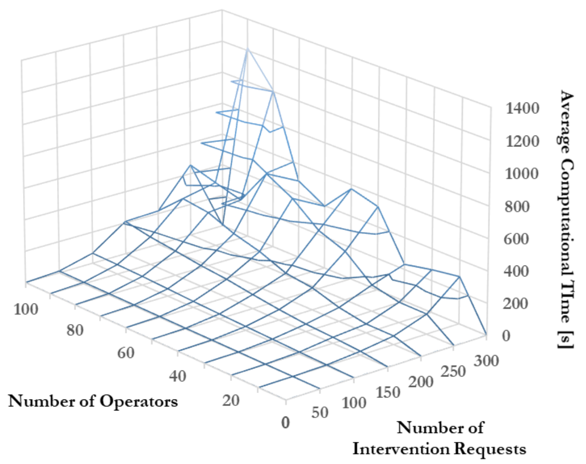

Figure 1: Computational Time

4. Model extensions

The task allocation problem addressed in this work represents a first step in the definition of a comprehensive optimization model able to consider all the services and asset-related information part of the framework described in (Sala, Pirola, et al., 2019). The current model is thus expected to be extended in the next future through the implementation of additional constraints or by changing the objective function considering new resolution interests.

The first change that could be applied to the model is related to the criticality of the interventions, which could be introduced assigning weights that changes according to the request type. Thus, equation (1) becomes:

$$\min Z = \sum_{k \in K} w_k * T_k \tag{14}$$

Considering the current objective function, equation (1), this could be changed if the interest is to minimize the total number of tardy interventions instead of the total tardiness in the schedule. Thus, the objective function changes as follows:

$$\min Z = \sum_{k \in K} U_k \tag{15}$$

In equation (15), U_k is the sum of the tardy interventions. In particular, $U_k = 1$ if the intervention $k \in K$ is satisfied after the due date DD_k, and equal to 0 otherwise. The definition of such an objective function changes the nature of the problem. If in the first case the aim was to minimize the total tardiness considering all the interventions, in the second case the aim is to minimize the number of tardy interventions. The perspective of the problem changes. In fact, in the first case a T_k may be the result of one intervention with the tardiness of 10 or 10 interventions with the tardiness of 1 each. Instead, in the second case, $U_k = 10$ means that the company has 10 interventions that are tardy. Thus, the decision related to the objective functions is strongly influenced by the company strategy and the service contracts established with the customers. In some cases, it would be better to have only one intervention in considerable tardiness, in other cases it would be better to have several interventions with small tardiness.

Considering the updated objective function (15), also the constraint set (7) has to be updated

$$C_k \leq DD_k - M \cdot U_k \tag{16}$$

Constraint set (15) is true in two cases. In one case, the completion time C_k is actually lower than the due date DD_k. In the other case, C_k is higher than DD_k, and the constrain is valid only if $U_k = 1$, which means that the intervention is tardy. The definition of such a constraint allows counting the number of tardy interventions, supporting the use of the objective function written in the equation (15).

Another aspect that could be taken into account when extending the model is the possibility that the operators are not skilled for every intervention. The current model assumes that all the operators are skilled to execute every task associated with the different intervention requests. The difference between the operators can be found in the time requested to execute the interventions. In a real context, not all operators can execute all the interventions. To implement this into the model, it is necessary to introduce a constraint set that considers the skills required by the task and the skills that the operator has. It would be necessary thus, to know for every intervention, and related task, the skills required to fulfil it optimally. To implement such a constraint in the model, it would be necessary for the service provider to map and store all the competencies of the resources working in the service department. Once mapped, the competence database could be used to filter the list of operators able to execute the required intervention reducing the risk of allocating the requests to operators who are not sufficiently skilled to execute it efficiently.

Another possibility for the extension of the model is related to the introduction of availability windows or calendars for operators. As of now, the model considers an empty schedule for the operators at the beginning of the planned week. For this reason, the model can assign each intervention without the necessity to consider blocked and unavailable periods. The introduction of such a hypothesis would mean to reduce the availability of the operators, increasing the difficulty for the optimization model to identify a configuration that minimizes the total tardiness (if the objective function is equation (1)) or that minimizes the number of tardy interventions (15). At the same time, it would also be possible to assign more than one intervention to the same time window, allowing an operator to perform two or more consecutive interventions as long as the time window is not closed.

In future extensions, the model should better clarify the impact of costs in the definition of the solution. As of now, the model works under the assumption that costs are only related to tardy interventions, and are modelled only indirectly through the minimization of the tardiness. Future development should consider the quantification of these costs in the model considering also the weight of the tardy intervention, including both the temporal and economic aspects in the resolution.

Finally, the model as of now adopts a deterministic view of the problem (i.e. concerning the time required for the operator's travel or the execution of the intervention), while stochasticity is neglected. An extension should consider this aspect to provide sound results. Similarly, a sensitivity analysis could be used in the current model to test the stability of the solution.

5. Conclusions

The allocation of intervention requests to operators is a critical task for companies who intend to offer a satisfactory maintenance service to their customer. The new possibilities disclosed by Industry 4.0 in terms of data extraction and elaboration allow improving the way components health status is monitored. Data-driven decision-making could contribute to the way maintenance interventions are scheduled since, using suitable approaches, it could be possible to anticipate the failure, intervening only in the proximity of it, avoiding a too early repair activity that would prevent the company from exploiting at maximum the RUL of the components as much as possible. Moreover, it would be possible to support the daily tasks of the planners and the service department through the elaboration of aggregated data and the exploitation of specific useful information.

This paper contributes to this research stream through the proposal of an optimization model that, considering information incoming from different sources, supports the planner in allocating the intervention requests to the operators to minimize the total tardiness of the interventions. The information sources can be divided into two main categories: service-related and product-related. Service information is mainly related to the available service resources and to the time constraints (e.g. set of tasks, set of operators, with the related schedule, SLA, time required to perform an intervention and to travel to the customer for the operator and the spare parts). Product information is mainly related to the RUL associated with a specific component or to an asset (depending on the granularity of the analysis).

The introduction in service departments of such a model could contribute to the improvement of the service performance by reducing the time required to allocate the task to operators and, at the framework level, as discussed in Section 2, introducing a continuous improvement logic that supports the identification of the problems that slow down maintenance delivery at the service and asset levels. In turn, this analysis would impact also on the intervention allocation in the medium term, since more specific information on the operators' performance and time required to execute the interventions will be available and, this, will be elaborated.

The application of the model discussed in this paper can have effects also on the definition of the service contracts. The knowledge created can be used to propose tailored contracts based on customer necessities and supplier capabilities. For example, only specific service levels could be proposed to customers, depending on supplier capabilities. Besides, more expensive contracts could be signed if some customers are willing to have a higher service level guaranteed.

The model proposed in the paper is an initial version of a more comprehensive model able to consider more aspects and variables, as discussed in Section 4. Future works will encompass the update of the model considering what has been discussed in Section 4 and its implementation in a real context.

References

Agnihothri, S. R., & Mishra, A. K. (2004). Cross-training Decisions in Field Services with Three Job Types and Server-Job Mismatch*. *Decision Sciences*, *35*(2), 239–257.

Andreacchio, M., Bekrar, A., Benmansour, R., & Trentesaux, D. (2016). Balancing preventive and corrective maintenance of aircraft assets: A cyber-physical systems approach. *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)*, 500–503.

Anim-Ansah, G., Kern, M., & Voudouris, C. (2006). Field service planning as an enabler for field service optimisation. *3rd International IEEE Conference Intelligent Sysems*, 156–162.

Bagozi, A., Bianchini, D., Antonellis, V. De, Marini, A., & Ragazzi, D. (2017). Interactive Data Exploration as a Service for the Smart Factory. *2017 IEEE International Conference on Web Services (ICWS)*, 293–300.

Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers and Industrial Engineering*, *137*, 106024.

Cattaneo, L., Fumagalli, L., Macchi, M., & Negri, E. (2018). Clarifying Data Analytics Concepts for Industrial Engineering. *IFAC-PapersOnLine*, *51*(11), 820–825.

Chopra, V., & Priyadarshi, D. (2019). Role of machine learning in manufacturing sector. *International Journal of Recent Technology and Engineering*, *8*(4), 2320–2328.

Coreynen, W., Matthyssens, P., & Van Bockhaven, W. (2015). Boosting servitization through digitization: Pathways and dynamic resource configurations for manufacturers. *Industrial Marketing Management*, *60*, 42–53.

Gopalakrishnan, M., Bokrantz, J., Ylipää, T., & Skoogh, A. (2015). Planning of maintenance activities - A current state mapping in industry. *Procedia CIRP*, *30*, 480–485.

Hu, C., Youn, B. D., Wang, P., & Taek Yoon, J. (2012). Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliability Engineering and System Safety*, *103*, 120–135.

Huang, K.-L., Blue, J., Weng, H.-C., & Liu, S.-H. (2018). Single machine scheduling with consideration of preventive maintenance and machine health. *MATEC Web of Conferences*, *192*, 1009.

Hurkens, C. A. J. (2009). Incorporating the strength of MIP modeling in schedule construction. *Res*, *43*, 409–420.

Kindström, D., & Kowalkowski, C. (2014). Service innovation in product-centric firms: a multidimensional business model perspective. *Journal of Business & Industrial Marketing*, *29*(2), 96–111.

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, *26*(3), 159–190.

Lagemann, H., Dorka, T., & Meier, H. (2014). Evaluation of an IPS2 delivery planning approach in industry - Limitations and necessary adaptations. *Procedia CIRP*, *16*, 187–192.

Mathieu, V. (2001). Service strategies within the manufacturing sector: Benefits, costs and partnership. *International Journal of Service Industry Management*, *12*(5), 451–475.

Moghaddass, R., & Zuo, M. J. (2014). An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process. *Reliability Engineering and System Safety*, *124*, 92–104.

Pal, D., Vain, J., Srinivasan, S., & Ramaswamy, S. (2017). Model-based maintenance scheduling in flexible modular automation systems. *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, 1–6.

Potes Ruiz, P. A., Kamsu-Foguem, B., & Noyes, D. (2013). Knowledge reuse integrating the collaboration from experts in industrial maintenance management. *Knowledge-Based Systems*, *50*, 171–186.

Rahman, A. R. A., Husen, C. Van, Pallot, M., & Richir, S. (2017). Innovation by Service Prototyping Design Dimensions & Attributes, Key Design Aspects, & Toolbox Abdul. *23rd ICE/IEEE International Technology Management Conference*, 587–592.

Rondini, A., Pezzotta, G., Cavalieri, S., Ouertani, M. Z., & Pirola, F. (2018). Standardizing delivery processes to support service transformation: A case of a multinational manufacturing firm. *Computers in Industry*, *100*(February), 115–128.

Ruiz-Sarmiento, J. R., Monroy, J., Moreno, F. A., Galindo, C., Bonelo, J. M., & Gonzalez-Jimenez, J. (2020). A predictive model for the maintenance of industrial machinery in the context of industry 4.0. *Engineering Applications of Artificial Intelligence*, *87*, 103289.

Ruiz, P. P., Foguem, B. K., & Grabot, B. (2014). Generating knowledge in maintenance from Experience Feedback. *Knowledge-Based Systems*, *68*, 4–20.

Sala, R., Pezzotta, G., Pirola, F., & Huang, G. Q. (2019). Service Delivery Process improvement using Decision Support Systems in A new methodology to analyze the functional and physical architecture of two manufacturing companies existing products for an assembly oriented product family identification. *Procedia CIRP*, *83*, 248–253.

Sala, R., Pirola, F., Dovere, E., & Cavalieri, S. (2019). A Dual Perspective Workflow to Improve Data Collection for Maintenance Delivery: An Industrial Case Study. In F. Ameri, K. Stecke, G. von Cieminski, & D. Kiritsis (Eds.), *Advances in Production Management Systems. Production Management for the Factory of the Future. APMS 2019.* (pp. 485–492). Springer, Cham.

Tamilselvan, P., & Wang, P. (2013). Failure diagnosis using deep belief learning based health state classification. *Reliability Engineering and System Safety*, *115*, 124–135.

Unlu, Y., & Mason, S. J. (2010). Evaluation of mixed integer programming formulations for non-preemptive parallel machine scheduling problems. *Computers and Industrial Engineering*, *58*(4), 785–800.

Xu, Z., Ming, X. G., Zheng, M., Li, M., He, L., & Song, W. (2014). Cross-trained workers scheduling for field service using improved NSGA-II Cross-trained workers scheduling for field service using improved NSGA-II. *International Journal of Production Research*, *53*(4), 1255–1272.