International Journal for Quality Research 18(2) 587–602
ISSN 1800-6450

Goran D. Putnik[1]
Leonilde Varela
Pedro Pinheiro
Sara Maia
Cátia Alves

Article info:
Received 20.10.2023.
Accepted 03.02.2024.

UDC – 004.4`414
DOI – 10.24874/IJQR18.02-17

CC BY NC

IDENTIFICATION OF THE MAJOR TRENDS IN CURRENT MAINTENANCE POLICIES USING LDA METHODOLOGY FOR SEMANTIC ANALYSIS OF THE PUBLISHED RESEARCH RESULTS

Abstract: *This paper presents a semantic analysis of a collection of articles on maintenance policies through the Latent Dirichlet Allocation (LDA) methodology. The research comprises two periods. The first, from 2000 and 2018, aims to reveal the habits of organizational maintenance policies before the second period, from 2019 to 2023, and verify the differences in current trends in maintenance policies compared to the past.*

With technologies associated with cyber-physical systems and Industry 4.0, a change in organizational models and maintenance policies is expected. The results obtained through the LDA methodology show changes in the maintenance design, namely the intensification of the maintainability concept, schedule optimization, and the integration of technologies enablers of smart production.

Keywords: *Maintenance Policies, Semantic analysis, Latent Dirichlet Allocation (LDA)*

1. Introduction

Maintenance activities are essential for ensuring systems' reliability and smooth operation (Kamalraj, Verma, & Joshi, 2023) in industrial context, where numerous sources of uncertainty exist and is critically important to treat these sources of uncertainty properly (Kim, An, & Choi, 2017).

This article, aims to apply a Machine Learning (ML) technique the Latent Dirichlet Allocation (LDA) methodology, for extraction of "hidden" information from scientific publications addressing maintenance policies. The hypothesis of the article is that there are differences of the research topics regarding maintenance policies in last years in comparison with the earlier period.

The collection of articles for analyses by the LDA methodology is extracted from Scopus data base (scopus.com), by the keywords "maintenance policies" AND (manufacturing OR industry). The collection extracted is divided in 2 collections of articles for 2 periods, from year 2000 to 2018 and from 2019 to 2023. The analyses of 2 different periods allows to observe whether current trends in maintenance policies diverges from previous ones and whether new concepts emerge in maintenance policies scientific field.

The LDA methodology is applied for a semantic analysis for identification of relevant topics associated to the maintenance policies. The model considers the words in the text that are most plausible for a topic.

Corresponding author: Goran D. Putnik
Email: putnikgd@dps.uminho.pt

The words considered are from the articles abstracts.

Following, the article presents a traditional literature review related to maintenance policies, in chapter 2. Next, the LDA methodology is presented in chapter 3. After, the results obtained by application of LDA methodology are shown in chapter 4 with commentaries. The article finish with conclusions in chapter 5.

This article, is an extended version of the conference paper presented to “The Quality Festival 2023”, which took place 24-27 May 2023, at Faculty of Engineering, University of Kragujevac, Serbia.

2. Literature review

"Maintenance strategies" is often considered a synonym of "maintenance policies". In the literature, "maintenance strategies" and "maintenance policies" are used interchangeably (Santos, Cavalcante, & Wu, 2022).

Many may suspect that production and scheduling problems deal with systems remaining available forever. This fallacy is far from reality, as there are times when machines break down and/or require repair to be maintained (Mobley, 2002), implying that production planning and scheduling cover an infinite number of models.

In general, the production process planning/scheduling focuses on a single machine, modelling its deterioration and, simultaneously, allowing the analysis and configuration of more complex group technology (Wang, 2002), such as achieving integrated production and maintenance scheduling, optimizing maintenance time and avoiding costs associated with machine failures (Ladj, Varnier, & Tayeb, 2016). However, the joint optimization of maintenance and production plans deserves greater attention and more in-depth studies (Bajestani, Banjevic, & Beck, 2014; Rastgar, Rezaeian, Mahdavi & Fattahi, 2023; Liao, Pan, & Xi, 2010; Pan, Liao, & Xi, 2010), in addition to taking on integrated maintenance scheduling and machine degradation modelling, where they are taken into consideration separately (Mourtzis & Vlachou, 2018; Zhai, Riess, & Reinhart), aiming for perfect reliability, and availability for complex systems can be simplified by scheduling the maintenance staff schedules.

During the last decades, maintenance has changed significantly taking into account the new manufacturing paradigms (Vrignat, Kratz, & Avila, 2022). Nowadays organizations try to implement smart and efficient maintenance processes to raise productivity, increase customer satisfaction, reduce machine downtime, and decrease delays (Dakkak, Irhirane, & Bounit, 2019; Hajej, Nidhal, Anis, & Bouzoubaa, 2020). Maintenance policies aim, for example, to reduce costs, increase reliability or improve the quality of services and products.

Increasingly complex manufacturing systems have made maintenance more relevant to the industry with the increase in dynamics of production activities (Sezer, Romero, Guedea, Macchi, & Emmanouilidis, 2018).

With the challenges that occur due to stochastic failures and multistate deterioration in production systems, and with the prospect of increased flexibility in production scheduling management (Sun & Geng, 2019), corrective maintenance (CM) and preventive maintenance (PM) policies attempt to keep different production systems functional (Cui, Lu, Li, & Han, 2018). CM was the first to be applied. CM is done only after the failure occurs and is detected. Over time, PM began to be implemented. PM has specific schedules for performing maintenance (Matyas, Nemeth, Kovacs, & Glawar, 2017; Nemeth, Ansari, Sihn, Haslhofer, & Schindler, 2018).

Although several researches have focused on flexible workshop scheduling and often assume a static environment in production, with machines unoccupied and available for production, it does not reflect realistic production environments, with dynamic environments, with constant changes in the entry of new orders (Gao et al., 2015), variations in processing times and changes in expiration dates (Baykasoğlu, Madenoğlu, & Hamzadayı, 2020). As machines age, the likelihood of the need for downtime caused by scheduled PM tasks increases (Zandieh, Khatami, & Rahmati, 2017; Zhang, Liao, Zeng, Shi, & Zhao, 2021), as well as occasional machine failures, which trigger CM.

In 1970, a Japanese concept of Total Productive Maintance (TPM) emerged, with the main objective of increasing the productivity of existing equipment, expanding PM to become more productive maintenance and complementary to other global strategies (Zlatić, 2019). The implementation of TPM allows you to maximize the use of all equipment, eliminating losses resulting from failures, maximizing the efficiency of equipment operation and building a comprehensive PM system (Rasztorf, Urbaniak, & Zimon, 2023).

PM encompasses time-based preventive maintenance (TBPM) and condition-based maintenance (CBM) (Ben-Daya & Rahim, 2000). TBPM is a traditional version of PM as maintenance is also predetermined for specific schedules, CBM, on the other hand, selects maintenance activities considering the true conditions of the system, i.e., it monitors degradation and predicts failures. However, this subcategory encompasses predictive maintenance (PdM), which uses data and time understanding to signal the possibility of a failure and to prevent downtime as well. Predictive maintenance (PdM) and predictive analytics represent one of the recent innovations that provokes curiosity in researchers and industry. PdM skills are essential with the rise of complex industrial processes (Shrivastava, Singhal, & Bhuvana, 2023). There are relevant developments concerning complex ML models (Putnik, Manupati, Pabba, Varela, & Ferreira, 2021), and considering Remaining Useful Lifetime (RUL) (Manupati et al., 2019).

Real-time optimization models are the most required to be applied in practice, the expression "real time" is related to the system response to stops and unforeseen events that do not interrupt the production process (Harmonosky & Robohn, 1991). Manufacturing systems are exceedingly dynamic and prone to interruptions and unforeseen events. Optimization models allow to assist the manufacturing system to withstand and adjust to these unforeseen events in real time (Cheng et al., 2018; Zheng et al., 2018). These systems include smart objects that represent a sub-system of broader concepts of advanced manufacturing systems, such as Industrial Internet of Things (IIoT), ubiquitous and cloud manufacturing systems, cyber-physical systems, digital factories, factories of the future, and Industry 4.0, which is related to the phenomena of Big Data and associated technologies and techniques (Putnik et al., 2015). It emphasizes that without an information system based on the application of modern information technology, it cannot achieve an efficient and effective management of maintenance, which must be open and adaptable to grow with the information system and included processes, facilitating completion of this information system after a few years (Arsovski, Pavlovic, & Arsovski, 2008). The data produced by machines can improve maintenance in different areas (Kim et al., 2017) based on the IIoT and Artificial Intelligence (AI) (Zonta et al., 2020), allowing to improve fault detection, quality control and decision support (Shivajee, Singh, & Rastogi, 2019). With the accelerated development of the IIoT, Industry 4.0, Big Data and Cloud Computing (Chen, Mao, & Liu, 2014; Xu, He, & Li, 2014), companies are looking for strategies to differentiate themselves from competitors and improve quality of its processes, products and services, which requires the collection of data on machines, and through its analysis, quality flaws can be corrected (Köksal, Batmaz, & Testik, 2011). Some studies (Vaidya et al., 2018; Pierdicca et al., 2017; Jones et al., 2019) agree on the evolution of technology in 9 pillars, and these pillars are necessary to provide the improvement in the areas pointed out. The nine pillars encompass the creation of information from horizontal and vertical systems, big data and analytics, cloud computing, the IIoT, simulation, Autonomous Robots, Cyber Security, Augmented Reality, and Additive Manufacturing. These pillars transform the factory, meaning that it becomes fully autonomous, integrated and optimizes the production process.

Using previous technologies, Prognosis and Health Management (PHM) has emerged as a higher-level adaptation of CBM, intending to achieve more efficient maintenance that applies RUL for decision support (Vrignat et al., 2022). However, the PHM does not end with the RUL forecast. System health management goes beyond failure time predictions and supports optimal maintenance and logistics decisions, considering available resources, the operational context and the economic consequences of different failures. Health management is the process that enables timely and optimal maintenance actions to be taken based on diagnostic and prognostic results, available resources and operational demand (Fink et al., 2020).

Uncertainty in forecasts is an inevitable part (Kim et al., 2017), requiring the use of advanced predictive tools that reduce the impacts of these uncertainties (Lee, Lapira, Bagheri, & Kao, 2013). PHM systems occupy a central place in the opportunities created by cyber-physical systems and Industry 4.0, aiming to reduce the probability of extreme failure events (Biggio & Kastanis, 2020).

However, most maintenance optimization works do not consider PHM. Eventually, this gap arises because most studies apply detection, diagnostic and prognostic models to individual components of different types, while optimization systems require the analysis of integrated systems (Pinciroli, Baraldi, & Zio, 2023). PHM can be considered a holistic approach to a management system (Fink et al., 2020), with integration into complex industrial manufacturing systems being a differentiating element of the algorithms applied in PHM compared to other methodologies (Lee, Bagheri, & Kao, 2014).

With Industry 4.0, production planning is a priority, as achieving efficiency and speed in maintenance requires organizations to have a profitable production system (Rødseth, Schjølberg, & Marhaug, 2017). Therefore, the maintenance activities should be oriented and adapted to maximize the availability of installed systems and equipment in line with customer needs, with efficiency, effectiveness and sustainability, considering costs, production quality, environmental protection, safety and legality (Moutinho & Oliveira, 2015).

3. Methodology

In this semantic analysis, two sets of article abstracts extracted from Scopus (scopus.com) were used. The first collection contains 335 articles, published from 2000 to 2018 and the second one contains 214 articles, published from 2019 to 2023. The keywords used in the search were: “maintenance policies” AND (manufacturing OR industry).

In this article, the data analysis method chosen was topic modelling. Topic modelling automatically maps hidden topics in a set of text through ML algorithms. Among the different ML techniques, LDA methodology is chosen. LDA is a set of topic modelling techniques that automatically finds latent structures in a corpus of unstructured documents using word frequency statistics (Jelodar et al., 2019). A word is seen as a basic unit of data measurement, the corpus as a set of documents that encompass the data set, and each document as a sequence of words. The different words in a corpus represent the vocabulary and the topics represent the probability distributions of the vocabulary words (Vayansky & Kumar, 2020). Therefore, it is expected that certain words stand out more frequently in documents, as each document is reflexively related to a certain topic. Topics are semantic sets formulated by words that stand out in documents and in the corpus (Blei, 2012).

Fitting the LDA model requires an assessment of the results. Among the best-known measures are relevance (Sievert & Shirley, 2014) and coherence (Stevens, Kegelmeyer, Andrzejewski, & Buttler, 2012). Given the reasons presented by Sievert and Shirley (2014), we adopted the value of λ=0,6 for the measure of relevance for defining the topics.

Considering the size of the corpus, and because it is a coherence measure that considers the total size of the samples, the choice of the coherence CUmass measure seems to be the most appropriate (Stevens et al., 2012). The CUmass coherence metric (Mimno, Wallach, Talley, Leenders, & McCallum) is given by Equation 1:

$$C_{Umass} = \sum_{i=2}^{N} \sum_{j=1}^{i-1} \log \frac{D(w_i, w_j)+1}{D(w_j)} \qquad (1)$$

Where, in a document D, D(wi, wj) is the document frequency of the set of words wi and wj and D(wj) is the document frequency of only wj. When fitting LDA models, a CUmass value close to zero is desirable.

Considering the constitution and size of the corpus of the two periods analyzed, the following criteria were established in the LDA algorithm:

- Reduce the search to 4 topics due to the small size of the corpus and guarantee an intertopic distance that avoids the creation of topic clusters.

- Use 30 relevant terms per topic. Although there is the possibility of increasing redundancy in the terms that define the topics, this redundancy makes it possible to accentuate the validity of the name attributed to the topic.

- Prepare the documents belonging to the corpus. Considering that LDA is an ML technique, the algorithm must include specific routines to eliminate inaccuracies, duplications, redundancies and inconsistencies during data processing (Sharma, 2023). In this process, the use of stopwords is restricted. All these steps were supported by the NLTK library (Phand & Chakkarwar), with the most common words, such as prepositions and numbers, being eliminated. In order not to lose the expression “4.0”, “4.0” was replaced by "fourdotzero" in the texts. There is a routine that textualizes the acronyms with the information present in the articles and eliminates any text that is in parentheses. The final parts with additional information from the editors of the articles is removed. All words became singular. Only nouns were considered.

- Run a routine with 1000 simulations to create n-grams for the LDA model. The selection criterion for the most favorable simulation is the CUmass measure.

The methodology is summarized in figure 1.

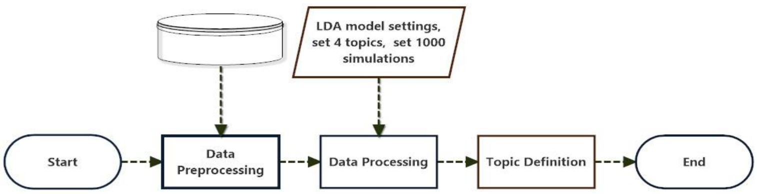
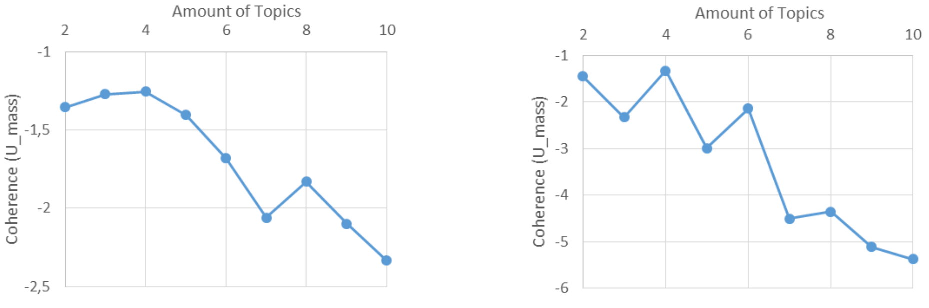


Figure 1. Proposed methodology for the automatic identification of the major trends in current maintenance policies.

4. Results

Figure 2 shows the plot of the CUmass coherence scores. The best results of 1000 simulations confirms that the selection of the number of 4 topics gives the value of coherence CUmass closest to zero as required, for each collection.



a) b)

Figure 2. Topic coherence index plot for data between a) 2000 and 2018 and b) 2019 and 2023.

The results, in Figure 3 to 10, present intertopic distance maps that show the 4 topics and the most relevant terms, considering λ=0.6. The Principal Component (PC) analysis projection compiles the similarity, allowing to obtain the intertopic distance (Chuang, Ramage, Manning, & Heer, 2012). PC1 embodies the most relevant variation in the data, while PC2 represents the second most relevant variation in the data. Although they represent the same data set, PC2 is independent of PC1. PC analysis allows the visualization of topics in a two-dimensional orthogonal graph, simplifying subsequent interpretation.

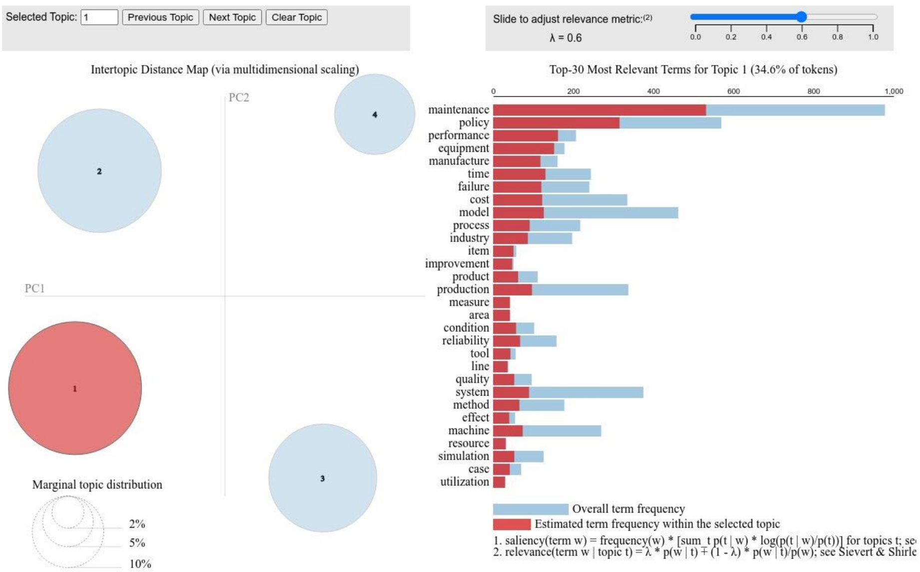


Figure 3. Intertopic Distance Map and the list of the Top-30 most relevant terms in topic 1 (2000-2018).

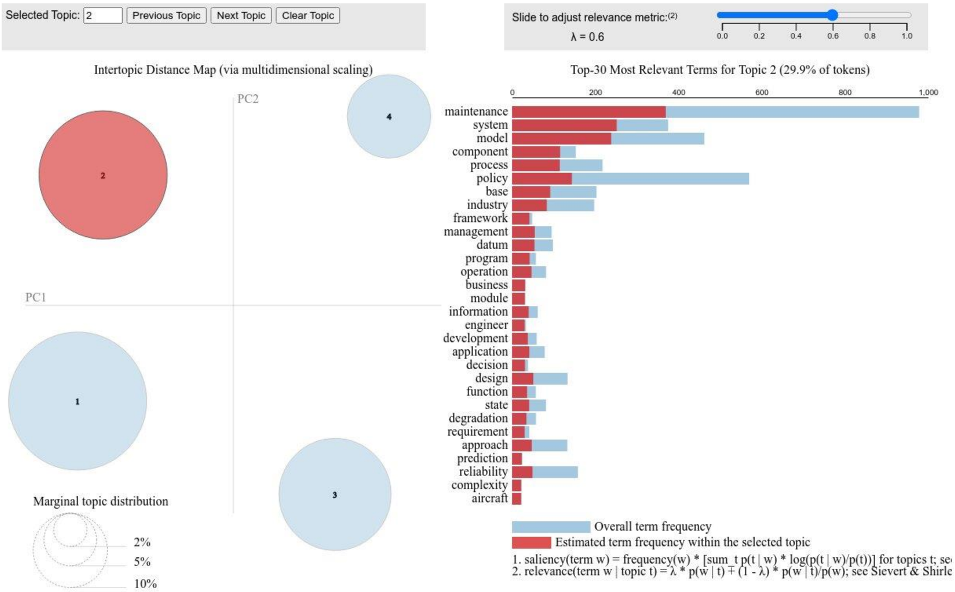


Figure 4. Intertopic Distance Map and the list of the Top-30 most relevant terms in topic 2 (2000-2018).

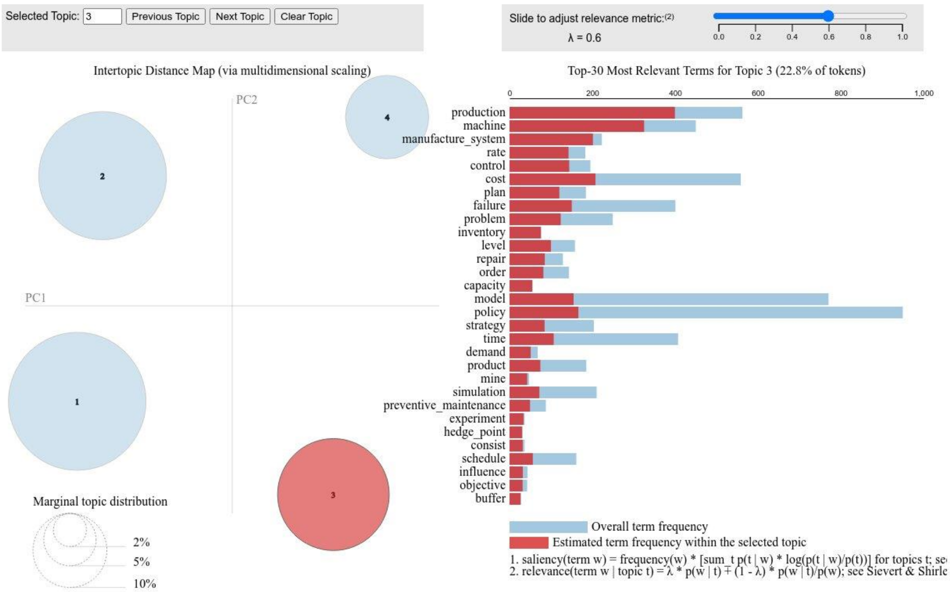


Figure 5. Intertopic Distance Map and the list of the Top-30 most relevant terms in topic 3 (2000-2018).

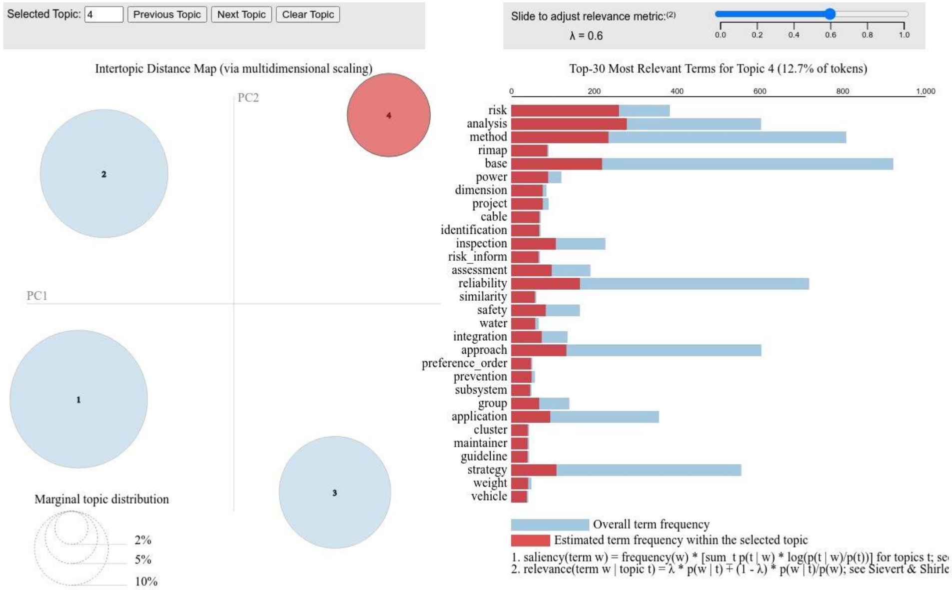


Figure 6. Intertopic Distance Map and the list of the Top-30 most relevant terms in topic 4 (2000-2018).

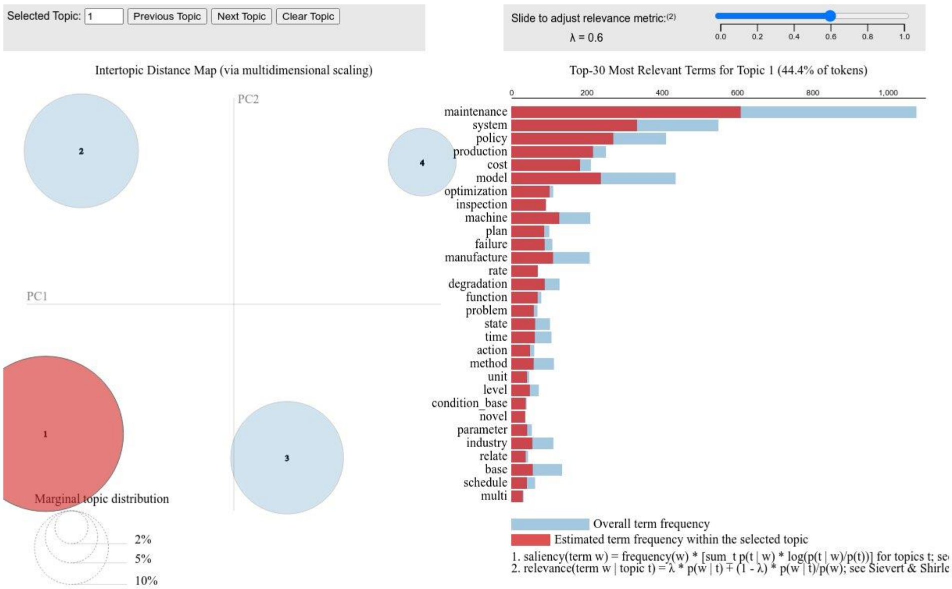


Figure 7. Intertopic Distance Map and the list of the Top-30 most relevant terms in topic 1 (2019-2023).

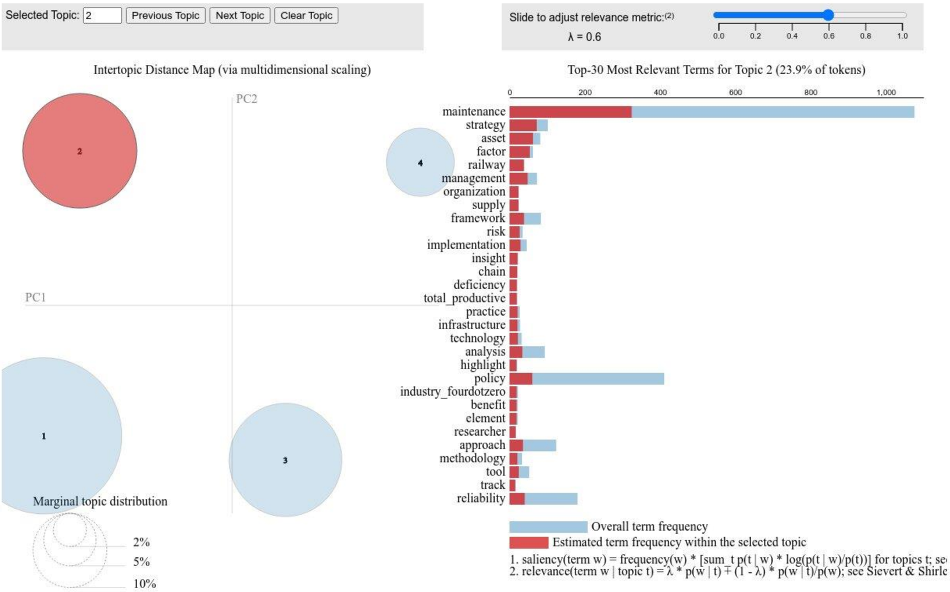


Figure 8. Intertopic Distance Map and the list of the Top-30 most relevant terms in topic 2 (2019-2023).

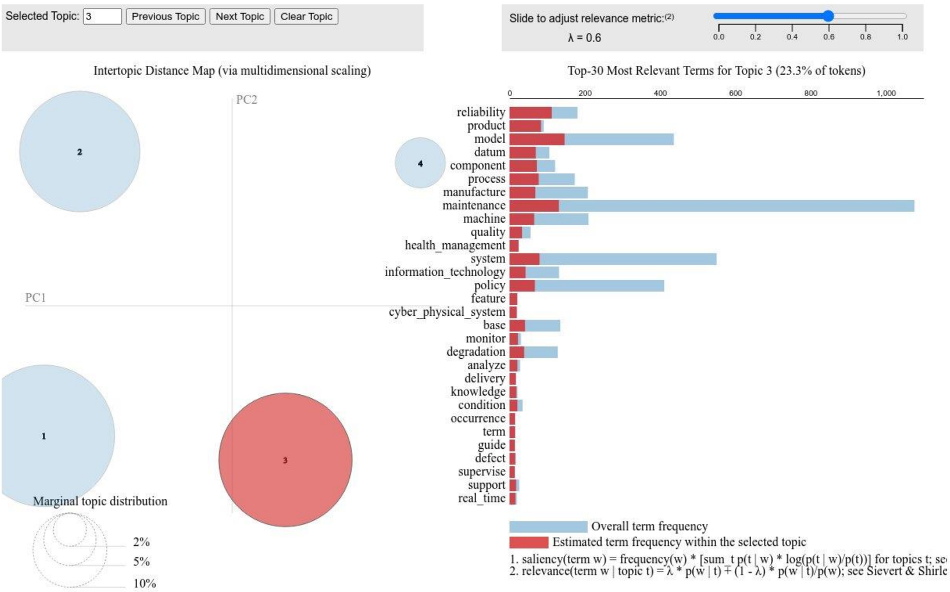


Figure 9. Intertopic Distance Map and the list of the Top-30 most relevant terms in topic 3 (2019-2023).

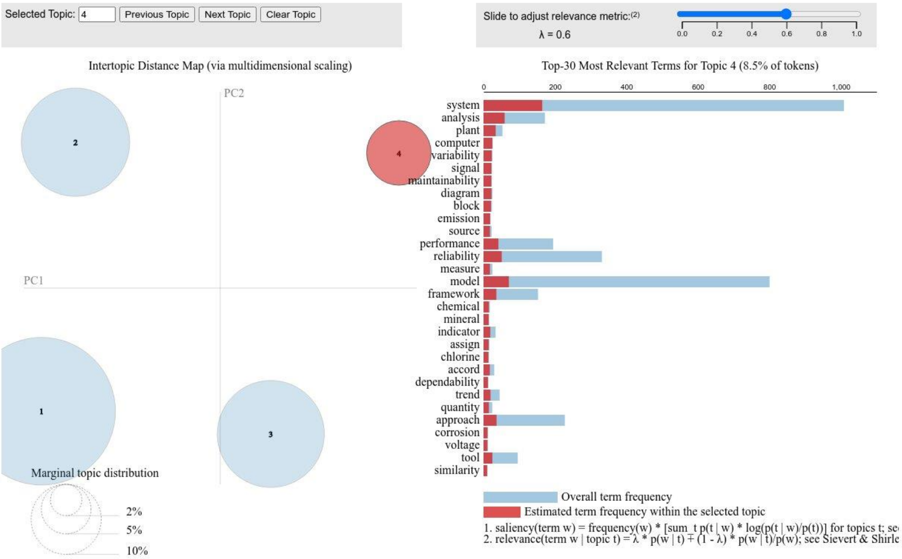


Figure 10. Intertopic Distance Map and the list of the Top-30 most relevant terms in topic 4 (2019-2023).

Starting with the collection of articles between 2000 and 2018, table 1 shows the 4 topics, and the 30 most relevant terms for each topic. The topic definition is constructed by the experts taking in the account corresponded relevant terms and their influences.

The table 2 shows the 4 topics, and the 30 most relevant terms for each topic, for the collection of articles between 2019 and 2023.

Table 3 shows the topics defined for the 2 period analysed.

Table 1. Topics and 30 most relevant terms from 2000 to 2018.

Topics (2000-2018)	30 most relevant terms
T1 – Production system, reliability, and conditional maintenance	[“maintenance”, “policy”, “performance”, “equipment”, “manufacture”, “time”, “failure”, “cost”, “model”, “process”, “industry”, “item”, “improvement”, “product”, “production”, “measure”, “area”, “condition”, “reliability”, “tool”, “line”, “quality”, “system”, “method”, “effect”, “machine”, “resource”, “simulation”, “case”, “utilization”]
T2 - Predictive maintenance, engineering and operations management	[“maintenance”, “system”, “model”, “component”, “process”, “policy”, “base”, “industry”, “framework”, “management”, “datum”, “program”, “operation”, “business”, “module”, “information”, “engineer”, “development”, “application”, “decision”, “design”, “function”, “state”, “degradation”, “requirement”, “approach”, “prediction”, “reliability”, “complexity”, “aicraft”]
T3 – Planning, inventory control and preventive maintenance	[“production”, “machine”, “rate”, “control”, “cost”, “plan”, “failure”, “problem”, “inventory”, “level”, “repair”, “order”, “capacity”, “model”, “policy”, “strategy”, “time”, “demand”, “product”, “mine”, “simulation”, “preventive_maintenance”, “experiment”, “hedge_point”, “consist”, “schedule”, “influence”, “objective”, “buffer”]
T4 – Risk assessment and integration of maintenance strategies	[“risk”, “analysis”, “method”, “rimap”*, “base”, “power”, “dimension”, “project”, “cable”, “identification”, “inspection”, “risk_inform”, “assessment”, “reliability”, “similarity”, “safety”, “water”, “integration”, “approach”, “preference_order”, “prevention”, “subsystem”, “group”, “application”, “cluster”, “maintainer”, “guideline”, “strategy”, “weight”, “vehicle”]

* “EU project RIMAP [Risk Based Inspection and Maintenance Procedures for European Industry (2000)] a new European Guideline for optimized risk based maintenance and inspection planning of industrial plants” (Bareiß et al., 2004).

Table 2. Topics and 30 most relevant terms from 2019 to 2023.

Topics (2019-2023)	30 most relevant terms
T1 – Condition-based maintenance, inspection planning and schedule optimization.	[“maintenance”, “system”, “policy”, “production”, “cost”, “model”, “optimization”, “inspection”, “machine”, “plan”, “failure”, “manufacture”, “rate”, “degradation”, “function”, “problem”, “state”, “time”, “action”, “method”, “unit”, “level”, “condition_base”, “novel”, “parameter”, “industry”, “relate”, “base”, “schedule”, “multi”]
T2 – Total productive maintenance, Industry 4.0 and reliability	[“maintenance”, “strategy”, “asset”, “factor”, “railway”, “management”, “organization”, “supply”, “framework”, “risk”, “implementation”, “insight”, “chain”, “deficiency”, “total_productive”, “practice”, “infrastructure”, “technology”, “analysis”, “highlight”, “policy”, “industry_fourdotzero”, “benefit”, “element”, “researcher”, “approach”, “methodology”, “tool”, “track”, “reliability”]
T3 – Cyber-physical systems, health management and information technologies	[“reliability”, “product”, “model”, “datum”, “component”, “process”, “manufacture”, “maintenance”, “machine”, “quality”, “health_management”, “system”, “information_technology”, “policy”, “feature”, “cyber_physical_system”, “base”, “monitor”, “degradation”, “analyze”, “delivery”, “knowledge”, “condition”, “occurrence”, “term”, “guide”, “defect”, “supervise”, “support”, “real_time”]
T4 – Systems reliability and maintainability performance	[“system”, “analysis”, “plant”, “computer”, “variability”, “signal”, “maintainability”, “diagram”, “block”, “emission”, “source”, “performance”, “reliability”, “measure”, “model”, “framework”, “chemical”, “mineral”, “indicator”, “assign”, “chlorine”, “accord”, “dependability”, “trend”, “quantity”, “approach”, “corrosion”, “voltage”, “tool”, “similarity”]

Table 3. Topics defined for the period from 2000 to 2018 and for the period from 2019 to 2023.

Topic	**Topic definition for the period from 2000 to 2018**	**Topic definition for the period from 2019 to 2023**
T1	Production system, reliability, and conditional maintenance	Condition-based maintenance, inspection planning and schedule optimization.
T2	Predictive maintenance, engineering and operations management	Total productive maintenance, Industry 4.0 and reliability
T3	Planning, inventory control and preventive maintenance	Cyber-physical systems, health management and information technologies
T4	Risk assessment and integration of maintenance strategies	Systems reliability and maintainability performance

Given the topics that define trends in maintenance policies in the period from 2019 to 2023, comparing with the period from 2000 to 2018, we observe that:

1) Maintainability implies design for maintenance;
2) Industry 4.0, cyber-physical systems and information technologies imply the use of IIoT, distributed computing, cloud, edge and fog architectures, i.e., smart production;
3) Integration of inspection planning, in addition to the “classic” integration of programming/ operation management with maintenance. This development is a step towards a kind of “holistic” maintenance.

5. Conclusion

The terms generated by the LDA analyses show that cost, reliability, risk, failure and performance, among others, are present in the both periods studied, remaining relevant issues in current maintenance policies. The word optimization only became relevant in the most recent period of 2019-2023, resulting from the opportunities that emerged through the adoption of cyber-physical systems and industry 4.0 technologies that improve PHM performance, increasing the efficiency and flexibility of organizational systems.

The appearance of TPM in topic 2, referring to the period from 2019 to 2023, shows that old methodologies and technologies are re-emerging as viable solutions to reduce the uncertainties surrounding production systems.

The results of applying LDA are in line with the literature review in chapter 2. It shows that LDA is a valid alternative to traditional literature review techniques that use statistics and require an enormous expenditure of time to analyze a large amount of information. Furthermore, with LDA, the prejudice of researchers, whose biases can lead to a discriminatory selection of articles, is significantly eliminated. In the proposed application LDA algorithm, we seek to ensure that bias only appears in the synthesis of topic construction transparently.

The methodology and algorithm applied favour data quality by removing texts' acronyms that could become redundant elements and sources of bias in the LDA analysis. The emergence of the RIMAP acronym in topic 4, referring to the period from 2000 to 2018, reveals that the algorithm doesn't automatically replace all acronyms due to the differences of the documents that are part of the corpus. Although it does not seem that this gap has significantly harmed the qualification of the topics, for future work, we intend to improve the algorithm in terms of data processing and the evaluation of the parameters that are input to LDA, considering the size and constitution of the corpus.

Acknowledgement: The project is funded by the FCT - Fundação para a Ciência e Tecnologia through the R&D Units Project Scope: UIDB/00319/2020, and EXPL/EME-SIS/1224/2021.

References:

Arsovski, Z., Pavlovic, M., & Arsovski, S. (2008). Improving the quality of maintenance processes using information technology. *International Journal for Quality Research*, *2*(2), 101–114.

Bajestani, M. A., Banjevic, D., & Beck, J. C. (2014). Integrated maintenance planning and production scheduling with Markovian deteriorating machine conditions. *International Journal of Production Research*, *52*(24), 7377–7400.

Bareiß, J., Buck, P., Matschecko, B., Jovanovic, A., Balos, D., & Perunicic, M. (2004). RIMAP demonstration project. Risk-based life management of piping system in power plant Heilbronn. *International Journal of Pressure Vessels and Piping*, *81*(10-11), 807–813.

Baykasoğlu, A., Madenoğlu, F. S., & Hamzadayı, A. (2020). Greedy randomized adaptive search for dynamic flexible job-shop scheduling. *Journal of Manufacturing Systems*, *56*, 425–451.

Ben-Daya, M., & Rahim, M. A. (2000). Effect of maintenance on the economic design of x-control chart. *European Journal of Operational Research*, *120*(1), 131–143.

Biggio, L., & Kastanis, I. (2020). Prognostics and health management of industrial assets: Current progress and road ahead. *Frontiers in Artificial Intelligence*, *3*, 578613.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, *19*(2), 171–209.

Cheng, Y., Zhang, Y., Ji, P., Xu, W., Zhou, Z., & Tao, F. (2018). Cyber-physical integration for moving digital factories forward towards smart manufacturing: a survey. *The International Journal of Advanced Manufacturing Technology*, *97*, 1209–1221.

Chuang, J., Ramage, D., Manning, C., & Heer, J. (2012). Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the SIGCHI conference on human factors in computing systems*.

Cui, W., Lu, Z., Li, C., & Han, X. (2018). A proactive approach to solve integrated production scheduling and maintenance planning problem in flow shops. *Computers & Industrial Engineering*, *115*, 342–353.

Dakkak, B., Irhirane, E. H., & Bounit, A. (2019). Support tool to assist the choice of skills for maintenance intervention based on experimental design. *International Journal for Quality Research*, *13*(3), 705–720.

Fink, O., Wang, Q., Svensén, M., Dersin, P., Lee, W.-J., & Ducoffe, M. (2020). Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence*, *92*, 103678.

Gao, K. Z., Suganthan, P. N., Chua, T. J., Chong, C. S., Cai, T. X., & Pan, Q. K. (2015). A two-stage artificial bee colony algorithm scheduling flexible job-shop scheduling problem with new job insertion. *Expert Systems with Applications*, *42*(21), 7652–7663.

Hajej, Z., Nidhal, R., Anis, C., & Bouzoubaa, M. (2020). An optimal integrated production and maintenance strategy for a multi-wind turbines system. *International Journal of Production Research*, *58*(21), 6417–6440.

Harmonosky, C. M., & Robohn, S. F. (1991). Real-time scheduling in computer integrated manufacturing: a review of recent research. *International Journal of Computer Integrated Manufacturing*, *4*(6), 331–340.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet Allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, *78*, 15169–15211.

Kamalraj, R., Verma, V., & Joshi, M. (2023). Machine Learning based grid safety assessment through simulation of unexpected contingencies during maintenance. *Proceedings on Engineering Sciences*, *5*(S1), 86–92.

Kim, N.-H., An, D., & Choi, J.-H. (2017). Prognostics and health management of engineering systems. *Switzerland: Springer International Publishing*.

Köksal, G., Batmaz, I., & Testik, M. C. (2011). A review of data mining applications for quality improvement in manufacturing industry. *Expert Systems with Applications*, *38*(10), 13448–13467.

Ladj, A., Varnier, C., & Tayeb, F. B.-S. (2016). IPro-GA: an integrated prognostic based GA for scheduling jobs and predictive maintenance in a single multifunctional machine. *IFAC-PapersOnLine*, *49*(12), 1821–1826.

Lee, J., Bagheri, B., & Kao, H.-a. (2014). Recent advances and trends of cyber-physical systems and big data analytics in industrial informatics. In *International proceeding of int conference on industrial informatics (INDIN)*.

Lee, J., Lapira, E., Bagheri, B., & Kao, H.-a. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters*, *1*(1), 38–41.

Liao, W., Pan, E., & Xi, L. (2010). Preventive maintenance scheduling for repairable system with deterioration. *Journal of Intelligent Manufacturing*, *21*, 875–884.

Manupati, V. K., Panigrahi, S., Ahsan, M., Lahiri, S., Chandra, A., Thakkar, J. J., . . . Varela, M. L. R. (2019). Estimation of manufacturing systems degradation rate for residual life prediction through dynamic workload adjustment. *Sādhanā*, *44*, 1–23.

Matyas, K., Nemeth, T., Kovacs, K., & Glawar, R. (2017). A procedural approach for realizing prescriptive maintenance planning in manufacturing industries. *CIRP Annals*, *66*(1), 461–464.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. Optimizing semantic coherence in topic models. In *Proceedings of the 2011 conference on empirical methods in natural language processing*.

Mobley, R. K. (2002). *An introduction to predictive maintenance*. Amsterdam: Elsevier Science.

Mourtzis, D., & Vlachou, E. (2018). A cloud-based cyber-physical system for adaptive shop-floor scheduling and condition-based maintenance. *Journal of Manufacturing Systems*, *47*, 179–198.

Moutinho, J., & Oliveira, F. (2015). Optimizing maintenance processes on customer site in a decentralized organization based on multi-site teams. *International Journal for Quality Research*, *9*(1), 141–150.

Nemeth, T., Ansari, F., Sihn, W., Haslhofer, B., & Schindler, A. (2018). PriMa-X: A reference model for realizing prescriptive maintenance and assessing its maturity enhanced by machine learning. In L. Wang (Chair), *51ST CIRP CONFERENCE ON MANUFACTURING SYSTEMS*.

Pan, E., Liao, W., & Xi, L. (2010). Single-machine-based production scheduling model integrated preventive maintenance planning. *The International Journal of Advanced Manufacturing Technology*, *50*, 365–375.

Phand, S. A., & Chakkarwar, V. A. Enhanced sentiment classification using geo location tweets. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)* (pp. 881–886). IEEE.

Pinciroli, L., Baraldi, P., & Zio, E. (2023). Maintenance optimization in Industry 4.0. *Reliability Engineering & System Safety*, 109204.

Putnik, G. D., Manupati, V. K., Pabba, S. K., Varela, L., & Ferreira, F. (2021). Semi-Double-loop machine learning based CPS approach for predictive maintenance in manufacturing system based on machine status indications. *CIRP Annals*, *70*(1), 365–368.

Putnik, G. D., Varela, M. L. R., Carvalho, C., Alves, C., Shah, V., Castro, H., & Ávila, P. (2015). Smart objects embedded production and quality management functions. *International Journal for Quality Research*, *9*(1), 151–166.

Rastgar, I., Rezaeian, J., Mahdavi, I., & Fattahi, P. (2023). A novel mathematical model for Integration of Production Planning and Maintenance Scheduling. *International Journal of Industrial Engineering and Management*, *14*(2), 122-137.

Rasztorf, R., Urbaniak, M., & Zimon, D. (2023). The role of quality in building relationships with MRO (Maintenance, Repair and Operating) suppliers. *International Journal for Quality Research*, *17*(3), 769–778.

Rødseth, H., Schjølberg, P., & Marhaug, A. (2017). Deep digital maintenance. *Advances in Manufacturing*, *5*, 299–310.

Santos, A. C. J., Cavalcante, C. A. V., & Wu, S. (2022). Maintenance policies and models: A bibliometric and literature review of strategies for reuse and remanufacturing. *Reliability Engineering & System Safety*, 108983.

Sezer, E., Romero, D., Guedea, F., Macchi, M., & Emmanouilidis, C. (2018). An industry 4.0-enabled low cost predictive maintenance approach for SMEs. In *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)* (pp. 1–8). IEEE.

Sharma, R. K. (2023). Improving quality of predictive maintenance through machine learning algorithms in industry 4.0 environment. *Proceedings on Engineering Sciences*, *5*(1), 63–72.

Shivajee, V., Singh, R. K., & Rastogi, S. (2019). Manufacturing conversion cost reduction using quality control tools and digitization of real-time data. *Journal of Cleaner Production*, *237*, 117678.

Shrivastava, M., Singhal, P., & Bhuvana, J. (2023). Integrating sensor data and machine learning for predictive maintenance in industry 4.0. *Proceedings on Engineering Sciences*, *5*(S1), 55–62.

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*.

Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). Exploring topic coherence over many models and many topics. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*.

Sun, X., & Geng, X.-N. (2019). Single-machine scheduling with deteriorating effects and machine maintenance. *International Journal of Production Research*, *57*(10), 3186–3199.

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582.

Vrignat, P., Kratz, F., & Avila, M. (2022). Sustainable manufacturing, maintenance policies, prognostics and health management: A literature review. *Reliability Engineering & System Safety*, *218*, 108140.

Wang, H. (2002). A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, *139*(3), 469–489.

Xu, L. D., He, W., & Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics*, *10*(4), 2233–2243.

Zandieh, M., Khatami, A. R., & Rahmati, S. H. A. (2017). Flexible job shop scheduling under condition-based maintenance: improved version of imperialist competitive algorithm. *Applied Soft Computing*, *58*, 449–464.

Zhai, S., Riess, A., & Reinhart, G. Formulation and solution for the predictive maintenance integrated job shop scheduling problem. In *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)* (pp. 1–8). IEEE.

Zhang, X., Liao, H., Zeng, J., Shi, G., & Zhao, B. (2021). Optimal condition-based opportunistic maintenance and spare parts provisioning for a two-unit system using a state space partitioning approach. *Reliability Engineering & System Safety*, *209*, 107451.

Zheng, P., Wang, H., Sang, Z., Zhong, R. Y., Liu, Y., Liu, C., . . . Xu, X. (2018). Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives. *Frontiers of Mechanical Engineering*, *13*, 137–150.

Zlatić, M. (2019). TPM – Total Productive Maintenance. *Proceedings on Engineering Sciences*, *1*(2), 581–590.

Zonta, T., Da Costa, C. A., da Rosa Righi, R., Lima, M. J. de, da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. *Computers & Industrial Engineering*, *150*, 106889.

Goran D. Putnik
Department of Production and Systems Engineering, University of Minho, Portugal
ALGORITMI Research Centre /LASI, University of Minho, Portugal
putnikgd@dps.uminho.pt
ORCID 0000-0003-3378-6866

Leonilde Varela
Department of Production and Systems Engineering, University of Minho, Portugal
ALGORITMI Research Centre /LASI, University of Minho, Portugal
ORCID 0000-0002-2299-1859

Pedro Pinheiro
ALGORITMI Research Centre /LASI, University of Minho, Portugal
ORCID 0000-0001-7849-4109

Sara Maia
ALGORITMI Research Centre /LASI, University of Minho, Portugal

Cátia Alves
ALGORITMI Research Centre /LASI, University of Minho, Portugal;
2Ai – School of Technology, IPCA, Portugal
LASI – Associate Laboratory of Intelligent Systems, Guimarães, Portugal
ORCID 0000-0001-6445-5475