Maintenance-Management in Light of Manufacturing 4.0

Michele Urbani, Dario Petri, Matteo Brunelli,
and Mikael Collan

1   INTRODUCTION

During the last decade, the manufacturing industry has gone through a deep transformation with the digitalization of processes, the arrival of the Internet of Things, the spread of artificial intelligence (AI) in daily practices, and the ubiquitous presence of data—thanks to the cloud technologies lifting the efficiency of manufacturing systems to a new level. Notwithstanding these radical changes, the manufacturing industry still has a strong dependence on maintenance, a field that is still considered to be a *necessary evil* by most managers, but without which plants and equipment will not remain safe and reliable. The importance of

M. Urbani (✉) • D. Petri • M. Brunelli
Department of Industrial Engineering, University of Trento, Trento, Italy
e-mail: michele.urbani@unitn.it; dario.pertri@unitn.it; matteo.brunelli@unitn.it

M. Collan
School of Business and Management, LUT University, Lappeenranta, Finland
e-mail: mikael.collan@lut.fi

maintenance-management as part of tangible asset-management is clearly inscribed within modern international industry standards [1], where asset-management is defined as "the coordinated activity of an organization to realize value from assets". Maintenance-management takes care of physical assets with the aim of minimizing their life-cycle cost and achieving stated business objectives. Depending on the specific sector of industry, maintenance takes different forms—its most elementary form involves simple operations and inspections of and on machines, while the most cutting-edge applications include intelligent maintenance control-systems capable of predicting the remaining useful life (RUL) of components and triggering maintenance activities automatically when needed. Moreover, some companies are adopting more holistic approaches to maintenance, aimed at improving the efficiency of the whole productive unit. Such approaches are called *total productive maintenance* (TPM) [2] and they aim at improving the quality of products, developing corporate culture, and enhancing the attention to safety and environment.

The popularization of Industry 4.0 paradigm around the year 2011 represented a new starting point for the manufacturing industry after the financial crisis of 2008. Asset-management and maintenance-management of physical equipment underwent a transformation: real-time monitoring of working conditions became very common due to decreasing cost of sensor technology (IoT devices), thus making possible the development of new technologies such as Virtual Factories and Digital Twins (DTs) of machines and processes. The digital replication of the physical environment allows the optimization of processes already during the design phase and the optimization of running processes during the production phase. Real-time monitoring of assets and the direct control of processes remotely has became a part of the new paradigm of manufacturing; with respect to maintenance, diagnostics and prognostics of equipment are spreading into daily practices and a new stream of research is contributing to the development of these technologies.

In this chapter we illustrate some of the connections between modern manufacturing (Manufacturing 4.0) and maintenance-management, present shortly the evolution of maintenance-methodologies starting from early models until today and summarizing the most important concepts relevant to the field including a discussion of how the digital twin concept may become an important issue for maintenance-management.

2   MAINTENANCE-MANAGEMENT: AN OVERVIEW

Maintenance-management is nowadays a fundamental function in most industry. In its traditional form, maintenance is aimed at ensuring that a system performs its function in a safe and efficient manner. Due to information technology (IT) development, maintenance-management has seen a significant evolution within its best practices: the classical methods for maintenance-planning and scheduling have been integrated and improved by technologies such as the Internet of Things, cloud computing, and artificial intelligence.

Engineering systems often have a complex structure, with a limited number of dedicated resources and strict requirements on safety and on performance—under these circumstances maintenance is an issue that needs to be handled in a systematic way. A clear strategy for maintenance must be defined, where components of a system to be maintained should be documented and listed according to priority, then a set of rules for the daily management of operations must be drafted. The set of rules that are used to coordinate maintenance tasks are typically called a *maintenance-policy*. As basic example, maintenance-policies for lifts and elevators that typically depend on country-wise regulations and that state that maintenance must be carried out on regular intervals, such as "every twelve months", which is then the rule that triggers a maintenance intervention that is aimed at avoiding sudden failures of the system. The above discussed types of interventions that are carried out before a failure has taken place are called *preventive* and they may range from simple inspections to the replacement of broken components. Maintenance actions undertaken after a failure are called *corrective* and they typically consist of the replacement and/or the repair of failed components. Usually corrective actions are more expensive than preventive, but when this is not the case it is sometimes possible to let a system *run to failure* that is, a system is left un-serviced until it fails, or until its fails and its failure is detected. Non-critical system components with a steady failure rate are often let run to failure.

Implementing preventive maintenance-policy typically requires more in terms of analysis, than a corrective policy—it requires information about the state of the maintained system such as information about the degradation level of system components. Depending on the information available, preventive maintenance-policies can be time-, or condition-based.

Time-Based Maintenance

Time-based or predetermined, as they are also called, maintenance-policies were the first approach adopted to effectively manage maintenance. In these types of policies maintenance actions are scheduled to take place on predefined times, according to set intervals of duration t_M, or upon failure (whichever occurs first). The aim of the policies is to preventively maintain the asset through shorter, but planned downtimes and by doing so avoiding longer and more expensive corrective maintenance actions. In this way the asset availability increases and consequences of failure can most often be avoided.

Scheduling of activities can be organized according to block-based- or age-based approaches. Block-based approaches schedule maintenance actions at constant time intervals, regardless of the asset operating time. The block-based approach is commonly used, when several assets of the same class (a block) are in (constant) use simultaneously. Age-based, or runtime, models are applied, when asset degradation and failures depend on the cumulative load exposure. Since the active age of a mechanical component has a strong correlation with the physical wear, or fatigue, of a component the maintenance of mechanical systems is often managed according to the age of system components. Asset age can be measured by using the working time of a machine as proxy, or in other ways, such as by observing the number of kilometres travelled or by the number of take-offs or landings, as can be done with aircraft. Approaches that combine more than one proxy for component states are also possible. Literature is ripe with research on time-based approaches for maintenance-optimization, we refer the interested reader to see the review by Wang [3]. It is worth to mention that time-based maintenance-policies carry a risk of over-maintenance, as some of the performed actions may not be necessary, on the other hand, time-based policies cannot weed-out failures, when component-deterioration happens at a non-standard pace—these are clear handicaps, when compared to condition-based policies. In fact, when the cost-risks of a time-based policy, or the costs of over-maintenance, are too high, condition-based maintenance may represent a feasible alternative.

Condition-Based Maintenance

Experience shows that failures can occur independently of the asset age, but at the same time most of these undesired events give some sort of

warning about the fact that they are about to occurring—thanks to the presence of such symptoms an early detection of fault occurrence is possible. This means that preventive actions can be taken, if the signals and symptoms of impending failures are understood, this is the fundamental concept that underpins condition-based maintenance. According to condition-based maintenance-policies maintenance actions are initiated by performance of a system reaching a trigger-level, typically determined by monitoring one or more indicators (sensors) of the maintained system. This means that maintenance is not done based on a predetermined schedule, but actions are taken based on observed, evidence-based deterioration of system performance that signals impending (component) failure and as such on only-when-needed basis.

A prerequisite for condition-based maintenance-policies (CBM) is that the there is objective *monitoring* of the system state in place—the monitoring should be carried out in a non-invasive way and it is typically achieved by using sensors. Monitoring can be scheduled or continuous and the output from monitoring is a set of observations (indicators, failure precursors) that describe the capacity of a system to perform its function. A typical example of a failure precursor is the vibration frequency of a rotating machine—shift in the frequency is a clear indication of a change in the working conditions. As a rule of thumb used in CBM, once enough data has been gathered, thresholds on the monitored feature-values are established to more reliably identify degraded asset performance—a comparison between the system-state and the thresholds is used to track the system health. With knowledge about the system health and history-based thresholds a decision about maintenance-scheduling can be made in a way that actions are performed only when needed and as a result both the probability of failure and the overall cost of maintenance can be optimized.

3   MORE ABOUT CONDITION-BASED MAINTENANCE

Setting up condition-based maintenance is a process and it can be divided roughly into three main steps. Condition-based maintenance assumes that objective monitoring of the system is possible, which means that *acquisition of data* about the system state is in place. Sensors that measure issues such as material cracking, corrosion, vibration, and change in electrical resistance are the types of information that are usable from the point of view of understanding the system state—one must also remember that these issues depend on the operating and the environmental conditions,

such as the frequency of use, ambient temperature, and humidity. It is typical that a monitored system must be equipped with sensors, signal conditioning and digitizing components that are typically already embedded in new modern machines. We emphasize the importance of sensors, because they are a core technology needed for the implementation of the Manufacturing 4.0 paradigm in maintenance—they are the bond that connects machines into networks and they allow the realization of the Internet of Things.

Based on the data collected the features that explain and describe the state of the system and allow determining whether maintenance is necessary must be estimated. Features can be difficult to observe directly (by observing the system), but by exploiting data and a priori knowledge of the system *feature extraction* can be made easier. The quality of a feature is determined by its capacity to represent the system state, in order to achieve a better state representation, usually a set of features is used—the more clearly different system states can be distinguished from each other the better the condition of the system can be described. In practice finding the correct features or sets of features that allow high failure detection capability and a low false alarm probability are problems that can be solved by specific methods created for feature-selection and for information fusion. Improvement in feature-selection methods has been fuelled by the great interest analytics and AI have received in recent times. One must remember that sudden changes in the operative and environmental conditions may render features that work well under normal conditions imprecise—this is why the best modern systems may use different sets of features for different operating conditions and are able to change the feature sets used "on the fly", when conditions change.

Once the data acquisition and feature extraction processes are ready *condition monitoring* can be effectively performed. Monitoring is the last step prior to the definition of the maintenance-strategy that is forming the set of rules that aids managers in taking maintenance decisions.

The main goal of condition monitoring is to provide fault-recognition, which typically foresees three sub-goals: (1) fault detection, aimed at identifying if a fault or the degradation of a component occurred; (2) fault isolation, that identifies the damaged component among many others; and (3) fault identification, aimed at determining the nature, extent, and severity of the isolated fault. In the following we look at these issues in more detail.

Fault Detection

The task of fault detection is to identify the presence of abnormal working conditions in a system by leveraging the information from the system history and information that can be learned from actual data. Typically a benchmark that defines the "normal" working conditions of the system is needed—the normal conditions depend on the task that the system is carrying out and on the environment surrounding the system. Because of different environments a system may have several normals—each normal will have a "profile" that is a set of features that defines it. Another thing is the extraction of profiles for different fault-states, such as "healthy", "degraded", and "faulty". The state of the system can be compared to the different profiles and this allows one to understand the state of the system and to predict the failure. Typically one will want to see several system states that precede the "failed" state, because the more states there are the finer is the information about the system state and better one can predict what will happen next. The comparison of the observed system state and the normal state can be done by different means, two examples of usable modelling techniques for this purpose are the auto associative kernel regression (AAKR) [4] and principal component analysis (PCA) [5] for the identification of the state and subsequently a statistical test is applied to identify the extent to which the state of the system differs from a normal condition. Typically used tests include the threshold based approach, Q statistics, and the Sequential Probability Ratio Test (SPRT) [6]. When the state of the system is known an action is taken (not taken) depending on the recommendations described for each state—the recommendations are drafted by using fault diagnosis techniques.

In order to clarify how fault detection works, we provide a simple example of condition monitoring. We assume that the state of a system is represented by a single feature $x(t)$. We define two thresholds considered important for the component. In Figure 1, the first threshold x_W identifies a warning-level, while threshold x_F identifies the failure of the component. When the value of $x(t)$ surpasses level x_W, an alarm is triggered, and a preventive action can be undertaken to prolong the life of the component, or to change it, to avoid incurring a sudden failure. The curve representing the behaviour of $x(t)$ is known as the Performance/Failure curve and it expresses the evolution of the system-feature as a function of either calendar time or system age time. A realistic mathematical model of $x(t)$ will also include the uncertainty related to the estimated quantity, which in

Fig. 1   A performance/
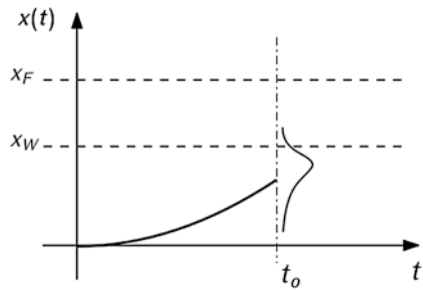failure curve for a
generic system

Fig. 1 is represented by a generic probability distribution. The importance
of modelling the degradation of a component using a random variable is
represented by the possibility to express the result using a probability that
is, a degree of belief about the triggering event.

Fault Diagnosis

Fault diagnosis is isolating and identifying the fault and typically means
identifying the cause, this means identifying which component in a system
is degrading among many possible components and to determine the
nature, the extent, and the severity of the fault. Isolating and identifying
the fault are sometimes overlapping and not always clearly separable. Fault
diagnostics means most often solving a classification problem—any given
set of measurements from the system can be matched to a single compo-
nent if sufficient data is available for training a machine learning classifica-
tion algorithm. In cases where data is abundant algorithms can even spot
specific conditions within components and provide a credible probability
of a failure event. Many techniques are good for this task, the interested
reader may find an extensive review about modern fault diagnostics tech-
niques applied to rotary machines in [7], where the authors describe
both the fundamental principles behind adopted AI algorithms and pres-
ent numerous application examples. As a caveat about AI-based tech-
niques one must observe that where there is no data, or data is very
incomplete, machine learning algorithms cannot be used—in such cases
suitable data must first be collected. In the cases of very rare faults diagno-
sis is difficult and diagnostics performance for them is typically poor.

   The performance of condition-based maintenance systems is only as
good as the system in place and there is uncertainty associated with the

outputs (alarms) from these systems. Uncertainty is caused by a number of things, some were already mentioned above such as the operating conditions and the environment, but others like production tolerances also affect the reliability of CBM system—because of tolerances two nominally identical machines may have a different wear. Due to this inherent inaccuracy the output from CBM systems is most often expressed as a probability or an interval. We refer the reader interested in deepening their knowledge in maintenance and maintenance optimization to read the review by De Jonge and Scarf [8].

4   Prognostics and Health Management—Towards Industry 4.0

Thanks to the availability of cheap networked sensors the monitoring and maintenance of systems is undergoing a fast and deep change. In the past, manual collection of maintenance-relevant data made the processing slow and unreliable—today technology allows abundant collection of data often in real-time. This profound change has caused the attention of maintenance systems development to move towards maintenance process-optimization. The new generation of production systems that are "smart" and networked has been labelled as Cyber Physical Production Systems (CPPS)—important to maintenance, they offer the possibility to perform real-time monitoring and accurate analysis of the degradation of critical components. This means that the long stream of research carried out on condition-based maintenance can now be exploited for its full potential—this change has given rise to the term Prognostics and Health Management (PHM), which can be said to be the cutting-edge approaches to predictive maintenance born within the last two decades. Keeping in mind that PHM is part of the same continuum with CBM and that the two cannot be sharply separated, it can be said that PHM aims higher than the "traditional CBM" and uses more advanced tools to get there.

The higher goals of PHM include, for example, optimization of maintenance-planning, reduction of downtimes, just in time spare parts provision, energy consumption optimization, minimization of raw material use and of pollution—all in all the focus is on increasing profitability through "better maintenance". PHM means effectively the same thing that is meant when the term Predictive Maintenance is used in common parlance. A fundamental prerequisite for a well-functioning predictive

maintenance system is the high quality of information that is used as an input into the system. This is true for both the real-time operation of the system as it is true for the information that is needed to construct or teach the system to be able to operate reliably—the information needed typically includes operating and maintenance histories, prior knowledge about system failure modes, resource constraints, and mission requirements. The information is used in tuning complex models the architecture of which may include numerous machine learning sub-systems and that require top of the line know-how. This means that these systems are expensive and they can be constructed only for systems that either merit such costs from the point of view of safety or that are business-critical and can economically justify the expenses.

In prognostics and health-management systems the system status received as input from condition monitoring is used to create an estimate of the system degradation state, which is used together with the P/F curve, or by using a classification-based architecture, to determine the distance between the current degradation level and a failure threshold (health-margin). The idea of the modern systems is to not only identify the cause of the fault but also to predict any secondary failures that may occur and to forecast the system health evolution as reliably as possible. Prognostics is considered the "holy grail" of PHM systems [9], because diagnostics has a retrospective approach to failure that consists of identifying and quantifying failures that have already occurred, while prognostics is about forecasting and as such, if successful means that the remaining useful life (RUL) of components can be accurately predicted. This will happen simply by being able to accurately estimate the end of life of a component and calculating the time to the end of life—the more accurate this ability is, the more precise can any optimizations performed based on it, including just in time deliveries of spare parts and maintenance scheduling become. The difference between high accuracy and medium accuracy can mean great savings in cases, where multiple systems are maintained and costs associated with maintenance are high. Another important issue is to know how much in advance a prognostics system can (accurately) predict the failure time—in fact, the relative RUL estimation accuracy and the prognostic horizon are key performance parameters of PHM systems.

In the literature, three types of approaches to prognostics have been identified, namely (1) experience-based approaches, which exploit historical information of a similar components; (2) model-based approaches, which make use of a physical fault model, and; (3) data-driven approaches,

which are mainly based on AI-techniques. We propose the interested reader to explore model-based and data-driven approaches by reading the book by Kim et al. [10].

Digital Twins and Their Connection to Maintenance

According to recent literature on maintenance and industrial management [11, 12] prognostics and health management systems be viewed as an examples of cyber-physical systems (CPS). The idea of CPS started to spread in the beginning of the 2010's, when NASA published their Modelling, Simulation, Information Technology & Processing Roadmap [13]—the document delineated the intention to integrating all the available physical and virtual technologies, the context back then was aeronautics. In essence the idea is that of a digital replica of a physical asset and it was called a Digital Twin (DT) and defined as "*an integrated multi-physics, multi-scale, probabilistic simulation of a vehicle or system that uses the best available physical models, sensor updates, fleet history, etc. to mirror the life of its flying (sic) twin. It is ultra-realistic and may consider one or more important and independent vehicle systems*". What makes this interesting from the point of maintenance is that predictive maintenance was one of the first fields of application of the DT concept, together with the check of mission requirements and a more transparent life-cycle view. The DT concept was subsequently extended to the manufacturing industry and the term Cyber-Physical Production System (CPPS) was coined to indicate the specific application area. A CPPS is composed of a physical part, a virtual part (the DT), and a stream of data between the two [14]. The DT strives to hold a perfect real-time synchronization between the physical and the virtual worlds, the physical part sends data to the virtual model, and the virtual part reproduces the physical system with ultra-high fidelity. As this is the case, historical data stored can be used together with real-time sensory information from the physical system in order to run, e.g., simulations and to optimize the production process virtually and then transmit "orders" to the physical system in order to optimize the way it functions. Theoretically the CPPS can harness the interaction between the virtual and the physical parts in order to create a continuously improving system. Digital twins are a clear way to remedy the typical problems of data collection, organization, and exploitation widespread in the context of production systems.

In fact, digital twins start to look like the key to reaching solutions for the problems of fitting together the best practices in engineering design and in process control. The advantages of adopting the DT concept seem cover the whole of product lifecycle that is, production design, manufacturing, and service providing are all immersed in the realm of DT [14]. In the design phase, if realized with a sophisticated digital model, issues that have to do with the maintainability of the production system can perhaps be addressed already on the drawing board—this may include the instrumentation of the system for best possible diagnostics and prognostics. During the production life of the production system the DT can perhaps assist in production planning, resource management, and procurement that can be optimized also with regards to predicted downtimes due to maintenance. The DT may run failure prediction algorithms in real-time so that users can be notified when the system state changes and in cases of imminent failure. It seems feasible to say that there is clear potential for maintenance systems development based on the digital twin concept.

5   CONCLUSION

Maintenance has always been a part of the management of production systems and it has become a craft of its own, the early mathematical models for maintenance management were based on the notion of optimizing the interval between maintenance activities in order to minimize downtime and the maintenance related costs. This type of maintenance management systems may still exist in cases, where preventive maintenance is the norm and the systems maintained are "old school" and not instrumented with sensors.

The modern approach of maintenance management is based on condition-based maintenance, which in the early days was more expensive than time-based maintenance management and thus reserved to high-risk and high-cost applications. Today the price of sensors and instrumentation is considerably low, which has made condition-based maintenance the leading way of handling maintenance management. Improvement of maintenance policies has created competitive advantages for companies that have been able to adopt them successfully and therefore a shift to modern maintenance management approaches is occurring in many companies. Automation of industrial facilities, such as the increasing use of robotics, improves productivity and safety, but it also increases the technological complexity of industrial assets and means a higher dependence on

production systems—this accentuates the role of effective and efficient maintenance.

Key Industry 4.0 technologies, such as artificial intelligence and Internet-of-Things, enable the implementation of very effective maintenance policies at an affordable cost and have paved the way for better diagnostic and prognostic systems, which can be said to be the backbone of what is typically called predictive maintenance. These systems are able to make fault-prediction even more accurate than what is possible with traditional condition-based maintenance methods and therefore offer a possibility for even further savings through better optimization. Predictive maintenance most importantly is a forward looking approach to maintenance, where traditionally the policies have been based on after-the-fact optimization.

The concept of digital twin is interesting from the point of view of maintenance management, as it is based on the idea of having a highly accurate real-time virtual model of a physical system that are "conversing" with one another. In effect, this is a concept that is not very far away from the ideal maintenance management system in terms of the information exchange between a production system and the maintenance management system. The digital twin, as it is used in the lifecycle management of products today is already opening avenues for many issues that are relevant to making maintenance better—looking forward there is potential for much more, specifically in terms of using digital twins in a maintenance focused way.

Getting back to the real-world, one must observe that the choice of maintenance management systems and policies is always constrained by the economic and technical realities surrounding the maintained systems. In this respect, predictive maintenance is at the start of a road that may lead at some point to something that resembles a digital twin—one thing is for sure, the Industry 4.0 paradigm and what we already can see beyond it will change maintenance management.

References

1. ISO 55000, "Asset Management—Overview, Principles, and Terminology." 2014.
2. S. Nakajima, *Introduction to TPM: Total Productive Maintenance*. Productivity Press, Inc., 1988.

3. H. Wang, "A survey of maintenance policies of deteriorating systems," *Eur. J. Oper. Res.*, vol. 139, no. 3, pp. 469–489, 2002.

4. P. Baraldi, F. Di Maio, P. Turati, and E. Zio, "Robust signal reconstruction for condition monitoring of industrial components via a modified Auto Associative Kernel Regression method," *Mech. Syst. Signal Process.*, vol. 60, pp. 29–44, 2015.

5. P. Baraldi, G. Gola, E. Zio, D. Roverso, and M. Hoffmann, "Expert Systems with Applications A randomized model ensemble approach for reconstructing signals from faulty sensors," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 9211–9224, 2011.

6. F. Di Maio, P. Baraldi, E. Zio, and R. Seraoui, "Fault detection in nuclear power plants components by a combination of statistical methods," *IEEE Trans. Reliab.*, vol. 62, no. 4, pp. 833–845, 2013.

7. R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, 2018.

8. B. de Jonge and P. A. Scarf, "A review on maintenance optimization," *Eur. J. Oper. Res.*, 2019.

9. J. B. Coble and J. W. Hines, "Prognostic algorithm categorization with PHM challenge application," *2008 Int. Conf. Progn. Heal. Manag. PHM 2008*, 2008.

10. N.-H. Kim, D. An, and J.-H. Choi, *Prognostics and Health Management of Engineering Systems.* Springer, 2017.

11. E. Negri, L. Fumagalli, and M. Macchi, "A Review of the Roles of Digital Twin in CPS-based Production Systems," *Procedia Manuf.*, vol. 11, no. June, pp. 939–948, 2017.

12. W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, "Digital Twin in manufacturing: A categorical literature review and classification," *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016–1022, 2018.

13. NASA, "Modeling, Simulation, information Technology & Processing Roadmap—Technology Area 11," *Natl. Aeronaut. Sp. Adm.*, p. 27, 2010.

14. F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, "Digital twin-driven product design, manufacturing and service with big data," *Int. J. Adv. Manuf. Technol.*, vol. 94, no. 9–12, pp. 3563–3576, 2018.