Laura Cattaneo, Adalberto Polenghi & Marco Macchi (2022) A framework to integrate novelty detection and remaining useful life prediction in Industry 4.0-based manufacturing systems, International Journal of Computer Integrated Manufacturing, 35:4-5, 388-408, DOI: 10.1080/0951192X.2021.1885062

A FRAMEWORK TO INTEGRATE NOVELTY DETECTION AND REMAINING USEFUL LIFE PREDICTION IN INDUSTRY 4.0-BASED MANUFACTURING SYSTEMS

The capability to predict the behaviour of machines is nowadays experiencing a tremendous growth of interest within Industry 4.0-based manufacturing systems. The route to this end is not straightforward when Run-To-Failure (RTF) data are poorly available or not available at all, thus a strategy must be properly defined. In this proposal, assuming no RTF data, a novelty detection is combined with random coefficient statistical modelling for Remaining Useful Life (RUL) prediction. This approach is formalized by means of a reference framework extending the ISO 13374 – OSA-CBM standards. The framework guides the integration of novelty detection and RUL prediction finally implemented in the scope of a Flexible Manufacturing Line part of the Industry 4.0 Lab *(the full name of the Lab with institution details is removed to guarantee anonymity)*.

Keywords: Industry 4.0, predictability, statistical modelling, condition monitoring, prognostics

1. Introduction

The capability to predict the behaviour of machines is nowadays showing a tremendous growth of interest within Industry 4.0-based manufacturing systems. Indeed, predictability is a major characteristic for future smart factories built upon Cyber-Physical Systems (CPSs) (Napoleone, Macchi, and Pozzetti 2020), where machines are monitored to predict their Remaining Useful Life (RUL) (Penas et al. 2017; Xu, Xu, and Li 2018). The ability to monitor the machine degradation should be built in CPSs to foster RUL prediction (Lee, Bagheri, and Kao 2015), which leverages on Prognostics Health Management (PHM) as underlying engineering discipline.

As a matter of fact, PHM provides the knowledge background for building advanced maintenance systems (Guillén et al. 2016). It is generally understood as the process of determining the current state of a system in view of reliability and forecast of its future state (Pellegrino et al. 2016), based on the detection and interpretation of Condition Monitoring (CM) data and degradation signals for RUL prediction (Jardine, Lin, and Banjevic 2006; Si et al. 2011). However, the route to this end is not straightforward. Based on authors' industrial experience it is remarked that, when initiating a machine CM process, companies typically lack both of historical dataset and of Run-To-Failure (RTF) data, due to newly commissioned machines, high reliability of machines or poor history of failure records. Physics-based approaches represent a first answer, but they are hard to establish for complex systems (Sikorska, Hodkiewicz, and Ma 2011). Artificial Intelligence (AI) approaches leverage on high quantity and quality of data, which impacts also their training (Cho et al. 2018). Thus, statistical approaches are still more used for machine prognostics at the state of the art (Lei et al. 2018).

The proposed strategy relies on the CM process to model the healthy state and to allow the detection of abnormal behaviours when they appear in machine life. With no RTF data, only available CM data are exploited through a novelty detection (ND) method. The detection of a novelty means the machine has started experimenting a degradation and algorithms for AHI definition (Asset Health Index, or Indicator) and RUL prediction are triggered to forecast the health state evolution. In this work, a random coefficient model is adopted to this end.

This strategy is formalized through a reference framework extending the ISO 13374 – OSA-CBM standards. It is based on two constitutive parts: a process model formalizing and integrating the PHM, and a data model specifying the machine (generically defined as asset) decomposition, its functioning, related information sources and formalizing a library of algorithms for ND-RUL integration. The data model is introduced since it enables full exploitation of CPS potentialities (Garetti, Fumagalli, and Negri 2015). The framework is suited for guiding the integration of ND and RUL prediction in the scope of a Flexible Manufacturing Line within the Industry 4.0 Lab *(the full name of the Lab with details and institution is removed to guarantee anonymity)*.

The paper is organized as follows. Section 2 presents the ISO 13374 and OSA-CBM as foundations of the adopted research methodology. Section 3 illustrates the findings of the literature review used to look after RUL prediction models in the Industry 4.0-like context. Section 4 regards the deployment of the framework, presenting both process and data models to integrate ND and RUL prediction. Section 5 presents the Proof of Concept (PoC) implemented in the FML of the Industry 4.0 Lab. Finally, section 6 reports conclusion and further research directions.

2. ISO 13374 and OSA-CBM as foundations of the research methodology

The PHM process starts from the raw data collection and ends exploiting information for decision-making in maintenance, logistics, engineering design, etc. (Guillén et al. 2016). In the maintenance scope, PHM provides the knowledge background for building advanced maintenance systems where CBM (Condition Based Maintenance) acts as enabler since it allows to monitor the machines' conditions. Nonetheless, the two terms are complementary and CBM/PHM represents an empowered CBM, essential for today's complex systems in industry (Vachtsevanos et al. 2006).

Amongst PHM-related standards applicable to manufacturing systems, the ISO 13374 represents a relevant source (Vogl, Weiss, and Donmez 2014). In the ISO 13374-1, a set of six functionality levels, intended as phases to be implemented in a CBM system, is defined, as well as their inputs and outputs, as summarised in Figure 1.

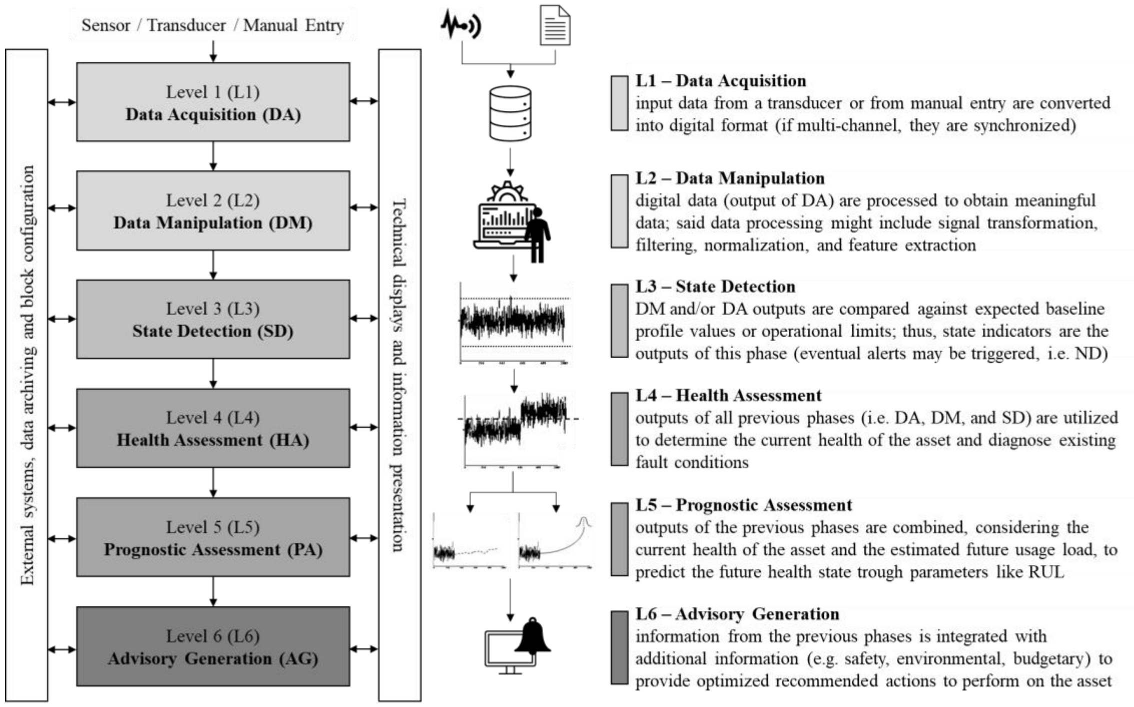


Figure 1. ISO 13374-1 levels and their explanation.

The ISO 13374 has been adopted within the OSA-CBM, which is the first-ever CBM-oriented technological-independent framework to design ICT solutions where the information content is separated from the technical interfaces (Bever 2007). Hence, OSA-CBM is conceived to provide the structure of a CBM implementation in the form of an architecture aimed at fully exploiting the technological deployment.

All in all, the ISO 13374 and the OSA-CBM represent relevant sources for the implementation of the CBM/PHM by specifying information and data along the process. As raw data input, *"sensors, transducer, manual entry"* are needed to unleash algorithms to provide the final support to decision-making. Complementarily, the importance of work histories and operational data is remarked to assess machinery health. These represent a starting point in regard to data to be considered when starting a PHM process.

When dealing with no RTF data, which is the assumption of this work, some functionality levels of the OSA-CBM are limited in their application, like the SD, where typically only the health state could be known, or the PA, where prediction could not be built on proper historical RTF data interpolation.

3. RUL prediction models: overview and literature review

Predicting the RUL is one of the main challenges in PA, along with determining the future trends of AHI. As such, several models have been developed, especially in Industry 4.0 manufacturing systems where predictability is a key capability. Thus, this section firstly provides an overview of the main RUL prediction models; then, a literature review explores prior scientific literature in the scope of this research.

3.1 Overview on RUL prediction models

Nowadays several models are available to establish a proper RUL prediction according to the current industrial problem (Gao et al. 2015) and relying on both local and cloud infrastructures (Caggiano 2018). These models could be grouped into four homogeneous families (Lei et al. 2018): physics-based models, statistical models, AI-based models, and hybrid models.

3.1.1 Physics-based models

Physics-based models, also called model-based approaches (Cubillo, Perinpanayagam, and Esperon-Miguez 2016), allow describing the health of a system by relying on engineering knowledge of the failure mechanism (Sikorska, Hodkiewicz, and Ma 2011). These models are built upon sets of equations describing the system of interest, including any physical property as material properties or stress levels. This allows describing the dynamics and degradation by applying random loads, and evaluating the effect on the RUL (Luo et al. 2003). Moreover, there could exist different formulations of the equations according to the domain of application and the complexity the modellers want to reach (Beden, Abdullah, and Ariffin 2009). On the whole, physics-based models represent the hardest way towards RUL forecasting (Sikorska, Hodkiewicz, and Ma 2011).

3.1.2 Statistical models

Statistical models determine the RUL of the machine by fitting available data, either through a pre-defined function or through interpolation by adaptable models, using both stochastic and random coefficient models. As result, a conditional probability density function (PDF) is built, able to predict the machine RUL (Si et al. 2011). These models effectively manage variability related to machine, measurement and other (Lei, Li, and Lin 2016). It makes statistical models particularly suitable for RUL prediction and some examples are reported in Table 1.

Model	Description	Reference
Autoregressive model	The model assumes the future state of the machine being linearly correlated with previous observations and random errors.	(Saha, Goebel, and Christophersen 2009)
Random coefficient model	The model describes the stochasticity of the degradation processes by introducing random coefficients, usually normally distributed, which allows providing a PDF of RUL.	(Chen and Tsui 2013)
Wiener process model	Wiener process model is a good choice for modelling non-monotonic degradation processes and assumes that the future state is only a function of the current state.	(Si et al. 2013)
Gamma process model	Gamma process model assumes that the accumulated damage in different time intervals is modelled with independent random variables following gamma distributions. Gamma process model can also describe a time-variant degradation process.	(Le Son, Fouladirad, and Barros 2016)
Inverse Gaussian process model	This model assumes that the degradation occurs with independent increments that follow an inverse Gaussian distribution. It is used for monotonic degradation patterns and random factors can be incorporated to model variability.	(Pan, Liu, and Cao 2016)
Markov model	This model describes a stochastic degradation with independent increments. It lies in the assumption that it is possible to determine the health state directly, also those which are hidden (HMM - Hidden Markov Models).	(Liu et al. 2015)
Proportional hazards model	This model assumes that the hazard rate of a system is made by the combination of a baseline hazard and covariate functions. It integrates the event data and CM data which make it achieve a high level of accuracy.	(Pham, Yang, and Nguyen 2012)

Table 1. Examples of statistical models.

3.1.3 AI-based models

AI-based models rely on the observation of a phenomenon to discover underlying degradation pattern of the machine (Cattaneo and Macchi 2019). They could be broadly categorised into supervised and unsupervised models. AI-based models are not forced to respect any physical (in terms of mechanical explanation of the degradation) or statistical (in terms of compliance with specific models associated with a PDF) property. Thus, they particularly fit the case of complex systems. Table 2 provides examples of these models.

Model	**Description**	**Reference**
Artificial Neural Networks *(supervised)*	They are the most extended AI technique for RUL prediction. ANNs can model processes with complex non-linear relationships. They tend to be quite accurate if they are trained with the right (large) amount of good quality data (as historical datasets). Their low transparency is one of their limitations.	(Tian 2012)
Decision Tree Method *(supervised)*	This method is composed by a set of decision rules. Visually it is like a tree, each node is a decision point where two attributes of observations are compared. Each node branches into two paths towards the leaf nodes, which provides a classification of all the instances.	(Yeo and Grant 2018)
Support Vector Machine *(supervised)*	Relying on supervised learning, given a collection of labelled data sets a SVM training algorithm builds a model that assigns non-labelled examples to one of the categories created during the training stage.	(Huang et al. 2015)
Cluster Analysis *(unsupervised)*	These techniques allow to subdivide the initial dataset into two or more subsets according to a similarity criterion. The definition of the similarity (or dissimilarity) is central in these techniques. A proper selection of this criterion allows the separation of the observations into subsets, each of them having different properties.	(Flath and Stein 2018)

Table 2. Examples of AI-based models.

3.1.4 Hybrid models

To overcome the limitations of each model family, hybrid solutions are also implemented combining two or more models of different natures. Examples of these applications are different, as (Di Maio, Tsui, and Zio 2012) or (Zemouri and Gouriveau 2010) that combine AI-based and statistical models.

3.2 Review of RUL prediction models in Industry 4.0-based manufacturing systems

RUL prediction models are cornerstones of Industry 4.0-based manufacturing systems, built upon CPSs, defined as smart embedded and networked systems (Lee, Bagheri, and Kao 2015) which allow connecting the physical and the virtual worlds to enable advanced monitoring and controlling capabilities. Digital Twin (DT) is a core part of CPSs, representing the virtual counterpart of the physical machine for enhanced simulation for different purposes (Negri, Fumagalli, and Macchi 2017), no more restricted to design, but including the machine's operations (Polenghi, Fumagalli, and Roda 2018). Also, DT fuses the data from multiple sources, thus boosting RUL prediction in production environments. A literature review is useful to shed light on current RUL prediction models, considering DT as a relevant part in Industry 4.0 CPS-based smart factories.

To properly organize the literature review, a three-step methodology is adopted, whose details are specified in Figure 2. For the sake of analysis, articles have been categorized according to four dimensions: industrial sector, level of analysis, simulation type and RUL model. Simulation type is included as it is a distinctive element of a DT (Negri, Fumagalli, and Macchi 2017). The results are summarized in Table 3.

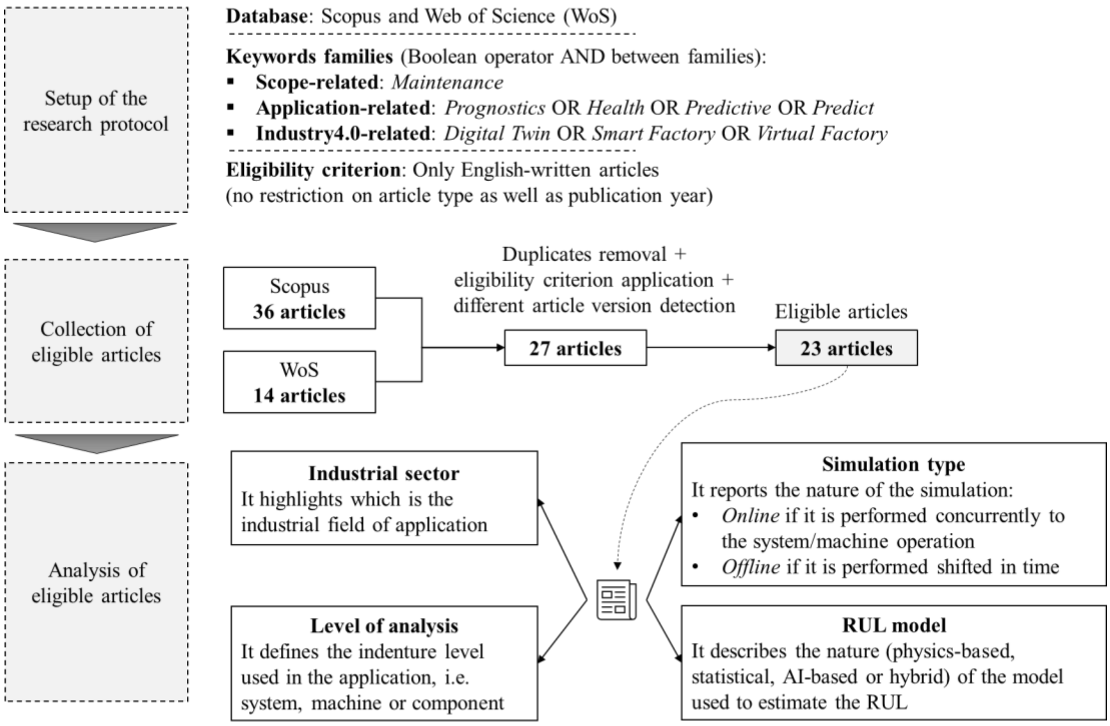


Figure 2. Literature review process.

Reference	Industrial sector	Level of analysis	Simulation type	RUL model
(Kumar, Chinnam, and Tseng 2019)	Manufacturing	-	-	S
(Raman and Hassanaly 2019)	Aerospace	Fleet/System	Online	-
(Tygesen et al. 2019)	Petrochemical	System (OGPS)	-	H: AI + P
(Cho et al. 2018)	Manufacturing	Machine (Milling)	Online	AI
(Sivalingam et al. 2018)	Wind farm	Machine (Wind turbine)	-	S
(Liu et al. 2018)	Manufacturing	Component	Online	S
(Qi and Tao 2018)	General	-	STD	-
(Luo et al. 2018)	Manufacturing	Machine (Milling)	Online	AI
(Tao et al. 2018)	Wind farm	Machine (Wind turbine)	STD, Online	H: AI + P
(Patnaik and Wu 2018)	Aerospace	Component	Online	H: AI/S + P
(Zaccaria et al. 2018)	Aerospace	Fleet	STD, Online	H: AI + P
(Rúbio, Dionísio, and Torres 2018)	Manufacturing	Component (Electric motor)	Online	AI
(Vathoopan et al. 2018)	Manufacturing	Component	Online	-
(Lee et al. 2018)	General	-	-	-
(Shubenkova et al. 2018)	Automotive	Fleet	Online	S
(Liu, Meyendorf, and Mrad 2018)	General	-	-	-
(Moyne and Iskandar 2017)	Semiconductor	-	-	S
(Li et al. 2017)	Aerospace	Component	-	H: P + S
(Kraft and Kuntzagk 2017)	Aerospace	Fleet	Online	P
(Boutrot et al. 2017)	Petrochemical	System (OGPS)	Online	H: AI/S + P
(Li and Roy 2015)	Manufacturing	Machine (Picking machine)	STD, Online	H: AI + S
(Tuegel 2012)	Aerospace	System (Airframe)	-	S
(Glaessgen and Stargel 2012)	Aerospace	-	-	-

OGPS: Offshore Oil and Gas Production Structure, STD: Simulated Training Data, H: Hybrid, S: Statistical, P: Physics-based, AI: AI-based

Table 3. Eligible documents analysis.

As observations from the literature findings.

- Manufacturing sector shows different applications of RUL prediction especially for machine tools, mainly performed at machine level; examples are milling machines (Cho et al. 2018) and drilling machines (Kumar, Chinnam, and Tseng 2019).
- AI-based and statistical models are the most widely discussed to predict the RUL, either applied alone or hybridized (results coherent with (Lei et al. 2018)).
- Hybrid models are widely used, especially in the form physics-based model joint with one of the other two types; typically, this is done as support if RTF data are low in number or not available, for example in (Li et al. 2017; Boutrot et al. 2017).
- AI-based models show an increasing application due to their ability in tackling heterogeneous data for machine behaviour forecasting (Luo et al. 2018); they also rely on physics-based models to define failures, underlining their need to rely on data for a successful RUL prediction.

Concluding, the availability of data appears as the major cornerstone for the application of AI-based models, limiting their performance if not or few RTF data are present. As (Cho et al. 2018) notices: *"available data limits application of machine learning algorithms and thus supervised learning and semi-supervised learning are not available for data analytics"*. Therefore, when no historical or RTF data are available, statistical models actually result to be the only feasible way (Hsieh et al. 2012) for monitoring the current state of the asset and predict its future state; this is possible by assuming a certain statistical function derived from sources like technical or scientific literature.

4. Deployment of the proposed framework to integrate ND and RUL prediction

The proposed framework aims to integrate ND and RUL prediction in context of Industry 4.0-based manufacturing systems, assuming no availability of RTF data and, therefore, statistical modelling as the key capability.

The framework extends the ISO 13374 – OSA-CBM in three functionality levels, SD, HA and PA, and with an additional preparatory phase, as illustrated in Figure 3, inspired by (Guillén et al. 2016). It is composed of two constitutive parts, namely the process model and the data model. The former aims at providing guidelines from a process viewpoint, focused on the interesting ISO 13374 – OSA-CBM levels; while the latter provides support in decomposing the asset of interest in its components and understanding its functioning, in identifying the information sources, and in structuring a library of possible alternative algorithms to be implemented. Being the basis of the framework, process and data models are thought to formalize the strategy for CBM/PHM implementation, through guidelines and supporting models tailored to the specific issue of no RTF data.

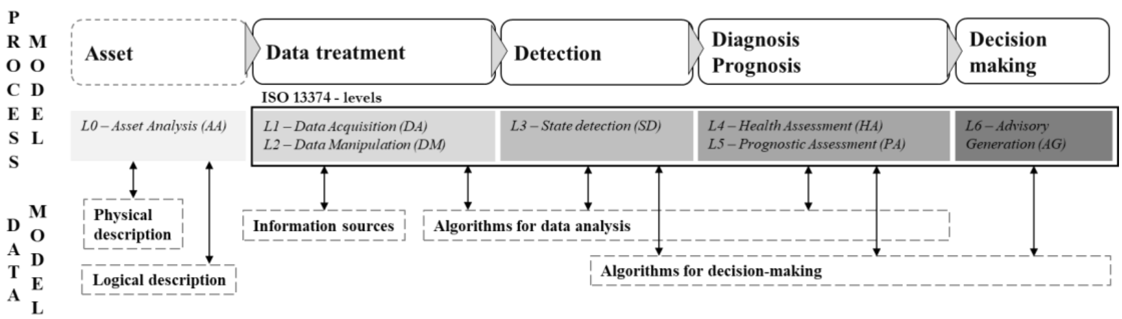


Figure 3. Proposed framework: process and data models.

4.1 Process model

The process model explains how to integrate ND and RUL when RTF data are not available by retracing the levels, from L0 to L6:

- L0 is introduced even if not present in the ISO 13374; it aims at exploiting all existing prior knowledge about the asset of interest in the system;
- L3, L4 and L5 focus on the guidelines to integrate ND and RUL prediction;
- L1, L2 and L6 are discussed to complete the whole process viewpoint.

4.1.1 L0 – Asset Analysis

Preliminary fundamental activities for PHM entail the selection of the most critical asset/s in the portfolio, its/their decomposition and functioning. To this end, standards are available:

- ISO 13329-2 includes an equipment audit, aiming at decomposing the asset, and a reliability and criticality audit, exploring relative failure modes and criticalities;
- ISO 14224 offers support in properly decomposing the asset;
- ISO 13306 offers an overview of the appropriate terminology to be used.

After having defined the physical structure (as asset decomposition), it is needed to (i) identify all modes in which the components could fail, the corresponding causes, and effects and (ii) prioritize the assets, and/or their failure modes, on which to focus the implementation of the PHM. The prioritization is supported by PHA (Process Hazards Analysis), which is a set of tools built upon risk management principles; amongst them, FMECA (Failure Modes, Effects and Criticality Analysis) is widely used, as described in IEC 60812.

4.1.2 L1 and L2 – Data Acquisition and Manipulation

The acquisition/collection of raw data about/from the asset is the target of DA level. Data may be of two kinds (Jardine, Lin, and Banjevic 2006): event data, related to the occurrences the asset went through in its life, and CM data, namely the measurements related to the asset conditions/states. DA is not investigated according to the scope of this research, but relevant references are (Ćwikła 2014) and the ISO 13xxx family entitled "Condition monitoring and Diagnostics of machines".

Then, DM is devoted to processing the digital data to obtain meaningful data. It is the most demanding activity since it is a combination of both analytic methods and engineering knowledge (Wilder-James 2016). DM encompasses three main sub-steps:

1. Data pre-processing, whose purpose is checking integrity and consistency of collected raw data, smoothing and eliminating the noise that might characterize them, and coping with missing values or errors;
2. Feature extraction from the acquired signals (Leturiondo et al. 2017), in different domains of interest (time, frequency, time-frequency);
3. Feature selection, aimed at selecting only the features that properly describe the degradation process of the asset under analysis.

More details may be found in (Alasadi and Bhaya 2017) and (García et al. 2016), which includes Big Data.

4.1.3 L3 – State detection

The identification of relevant features prepares for understanding the machine state/s. Firstly, the working regimes must be identified. Indeed, a machine usually works under dynamic working conditions, but most of the existing CM approaches focus on a single operating condition. Consequently, the asset health state is not properly assessed, and detection performance is unsatisfactory (Wang et al. 2019). The most common models addressing this issue are: dynamic multi-state models, which describe the state transition probability (Li et al. 2019; Tao, Zio, and Zhao 2018), and clustering algorithms (Wang et al. 2019).

Once defined the working regimes, the SD aims at developing a model for the asset whose RTF data are not available, and so features values describing the unhealthy state, labelled as "abnormal state", are not present. Coping with this, the proposed framework implements a ND algorithm, which aims at recognizing that test data differ to some extent from data available in the training dataset (Pimentel et al. 2014). In this optics, SD represents a one-class classification problem, where a single group of data (corresponding to the healthy state), must be differentiated from all other possibilities.

Different ND approaches are available, such as probabilistic, distance-based, and reconstruction-based. In this work, the former is considered and it is assumed that the healthy class is well sampled and follows a statistical distribution (Pimentel et al. 2014).

Each new sample from the machine could be compared with these pre-defined distributions upon establishing a proper novelty threshold, i.e. a control limit that separates the healthy class to the abnormal one. In the followed probabilistic ND approach, the novelty threshold is defined as $z(x) = y_{abnormal}$ such that every new feature value is considered "normal" if $P(value_{feature}) < y_{abnormal}$ or "abnormal" otherwise (being P(x) the cumulative probability associated with p(x), where p(x) is the PDF of the corresponding distribution).

Concluding, the SD level is particularly relevant to exploit the data analysis aimed at detecting novelties in different working regimes.

4.1.4 L4 – Health Assessment

A novel behaviour is experienced when the feature exceeds the novelty threshold $y_{abnormal}$ and so the machine is working abnormally. At least three states could be identified, where $y_{feature}$ is the measured value of the feature and y_{faulty} is the threshold for the faulty state:

- **Healthy state** ($y_{feature} < y_{abnormal}$): in this state the machine is behaving healthy and the RUL is not needed to be predicted, thus saving computational power especially useful for scaled up ND applications;

- **Abnormal state** ($y_{feature} \geq y_{abnormal}$ & $y_{feature} < y_{faulty}$): in this state, the machine is still able to perform its function, but a novelty arises; the event $y_{feature} \geq y_{abnormal}$ triggers the RUL prediction;
- **Faulty state** ($y_{feature} \geq y_{faulty}$): in this case, the machine is not anymore able to deliver the function it is designed for.

While $y_{abnormal}$ could be defined through a probabilistic approach, the faulty threshold y_{faulty} is harder to set, since faults have not been experienced yet. As such, if the failure mechanism physics is not known, y_{faulty} could be fixed by comparing similar machines owned by the company. Anyhow, the selection is dependent on the company risk aversion to asset failures (Polenghi et al. 2019): y_{faulty} closer to $y_{abnormal}$ is dictated by very stringent product/process requirements and quality, while a larger y_{faulty} implies more relaxed business or operational constraints.

4.1.5 L5 – Prognostic Assessment

Once fixed abnormal and faulty threshold, the PA involves the prediction of the RUL through the projection in time of interesting AHI. The AHI is a significant feature, extracted and selected during L2-DM, that is used for prognostic analyses purposes (Niu, Yang, and Pecht 2010). Given the unavailability of RTF data, the degradation model of the AHI must be assumed, and the proposed framework relies on statistical models.

As anticipated, the degradation model is triggered only if the AHI overcomes the abnormal state threshold $y_{abnormal}$. Consequently, the RUL is calculated as the time left before the AHI will cross the faulty state threshold y_{faulty}. In mathematical terms, the RUL (l_k) at the current time t_k can be expressed as in Equation 1 (Lei et al. 2018):

$$l_k = \inf\left(l : x(l + t_k) \geq y_{faulty}\right) \quad \text{------------Eq. 1}$$

where inf (·) represents the inferior limit of a variable, $x(l + t_k)$ is the state at $l + t_k$ with $l \geq 0$. This formulation is applicable with no RTF data since it does not require previous failure records to be used, but only the definition of the faulty threshold. Analogously, Equation 2 expresses the RUL by considering the probability of failure (P_f) obtained by integrating the PDF (f_{RUL}):

$$\int_0^t f_{RUL}(t)dt = P_f(t) \text{ -----------Eq. 2}$$

4.1.6 L6 – Advisory Generation

The levels L3, L4, and L5 support the maintenance decision-making and, by leveraging on available information, different decisions could be taken:

- *No maintenance action is needed*: all monitored values are within the healthy state bounds ($y_{feature} < y_{abnormal}$);
- *Reactive action*: the monitored value exceeds the threshold of the faulty state ($y_{feature} \geq y_{faulty}$); so, the machine is likely to fail and its function should be restored through a corrective maintenance action;
- *Proactive action*: the monitored value is between the abnormal and the faulty state ($y_{feature} \geq y_{abnormal}$ & $y_{feature} < y_{faulty}$); so, a maintenance action should be scheduled in the next *n* period/s according to the results of the prognostics algorithm/s, i.e. based on the predicted AHIs or RUL.

4.2 Data model

The data model, formalised in UML (Unified Modelling Language) as recommended by the ISO 19505-1, complements the process perspective. The data model defines the needed classes and relationships in the PHM process. As such, it empowers the process model in two ways:

- At levels L0 (AA) and L1 (DA), it supports asset decomposition, functioning and identification of sensors and information sources for DA;
- At level L2 (DM), it supports data processing since it provides a library of algorithms to be implemented in levels L3 (SD), L4 (HA) and L5 (PA).

The classes of the data models mainly come from international standards in line with the need of adopting a general and shared vocabulary.

4.2.1 L0, L1 – Physical and logical description of the asset and the information sources

The decomposition of the asset and its functioning is essential for the PHM development and the data model offers support to this end by (i) enabling 1-to-1 relation between the measured variables and its component/s, thus identifying relevant information sources, and (ii) tracking the asset operating states, both in terms of production function and healthiness.

The decomposition of the asset and its functioning is reported in Figure 4. It integrates several international standards, like ISO 14224, ISO 13306, ISO 13372, and scientific literature.

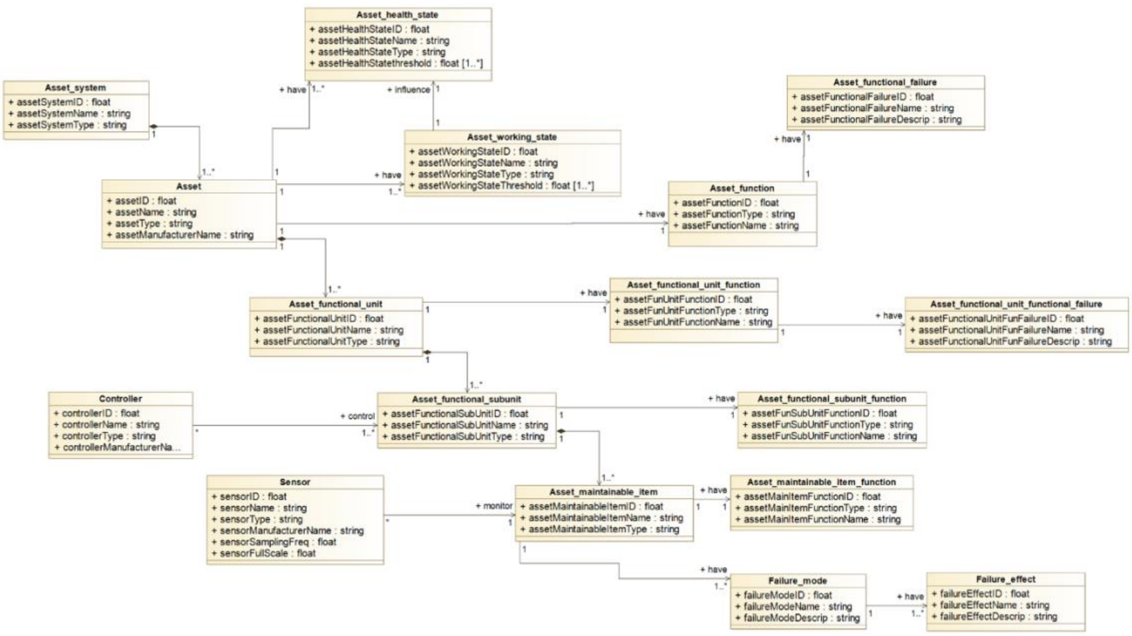


Figure 4. Data model for L0 and L1.

More in details:

- The **Asset** class is central and is defined as an *item, thing or entity that has potential or actual value to an organization* (ISO 55000 2014), which is comparable with “item” in ISO 13306. However, various definitions are available. Thus, instances of **Asset** are, for example, pumps, turning or milling machines.

- One or more **Asset**(s) belongs to an **Asset_system**, which is defined as a group of **Asset** that share the same functional goal or that are homogeneous in their technology, like assembly systems realising a complete product or a turning machines department in a job shop, respectively.
- The **Asset** is decomposed in one or more **Asset_functional_unit**, which enables the **Asset**(s) to work, e.g. lubrification unit, cooling unit, or machining unit. The **Asset_functional_unit** is further decomposed into **Asset_functional_subunit** and, furthermore, into **Asset_maintainable_item**. The former is a subunit needed for the functional unit to work, e.g. a cooling unit needs a refrigerator subunit and a distribution subunit to work properly; the latter is the item which is maintained in the subunit, e.g. cooler, coupling, filter.

Relating to the hierarchical decomposition of the asset, it is also important to understand, as prior knowledge, the functional failures the asset or its functional units may undergo and the failure mode/s of the maintainable item. To this regard, the main reference is ISO 13372.

- An **Asset** has an **Asset_function**, which is the required function it must perform by design. This could be of different types, like manufacturing, washing, assembly, etc. If the **Asset_function** cannot be performed an **Asset_functional_failure** exists, e.g. the turning machine is no more able to realise the turned metal product.
- Likely, the **Asset_functional_unit** has an **Asset_functional_unit_function**, which may have an **Asset_functional_unit_functional_failure**. For example, the machining unit of a turning machine, whose function is to remove metal from the raw metal product, is no more able to perform it.

- The functional failures may come from the occurrence of a **Failure_mode** that is the way in which the failure manifests in the **Asset_maintainable_item**. Each **Failure_mode** generates a **Failure_effect**. For example, the electro-spindle whose header is worn out and the cut is no more performed, leads to downtime as failure effect.

When the **Asset** is in service, its operations should be considered as well. In this regard, international standards offer few supports in defining asset operations and states if not in general terms. The following states are considered:

- **Asset_health_state** defines the healthy, abnormal, or faulty state in which the **Asset** could be;
- **Asset_working_state** describes the set of conditions under which the asset must work, e.g. the setup of a specific part program on a milling machine that defines the cutting speed and the feeding rate; this **Asset_working_state** influences the **Asset_health_state** since some programs may be more critical for the **Asset** rather than others, inducing different speeds for the degradation processes and the subsequent state transitions from healthy to abnormal and faulty states.

Finally, the information sources must be identified and the ISO 13379-1, ISO 17359, ISO 15531-1, ISO 13374-1 provide some references.

- **Sensor** is a class representing an item that has the precise goal of measuring a specific quantity, e.g. temperature or vibration; it is essential to provide CM data for the PHM.
- Other data may come from the **Controller**(s) of the **Asset_functional_unit**, such as PLC, Programmable Logic Controller, or CNC, Computerized Numerical Control, that generate CM data as well.
- Information systems on the shopfloor, like MES (Manufacturing Execution System) or CMMS (Computerized Maintenance Management System), provide

event data. They consider all maintenance-related occurrences the asset underwent, like installation, breakdown, overhaul, but also production-related occurrences, like production plans and part programs. These may be relevant for defining the **Asset_working_state**.

4.2.2 L2, L3, L4 and L5 – Algorithms for data analysis

The data model offers support also to the sub-steps of the DM phase, that are data pre-processing, feature extraction and feature selection, for which a library of algorithms is required:

- firstly, an **Algorithm_preprocessing** is needed for different reasons, such as data cleaning, data fusion and other issues;
- secondly, an **Algorithm_feature_extraction** needs to be applied to derive, from the original dataset of measured variables, a set of **Feature**;
- finally, an **Algorithm_dimensionality_reduction** aims at reducing the **Feature** space by identifying those features that mainly describe the variability, which are then selected for use in next levels (**Significant_feature**).

It is worth remarking that an inheritance of the class **Significant_feature** is the class **Asset_Health_Index**. Even though the two may seem the same concept at a first sight, the extant literature outlines a distinction:

- **Significant_feature** highlights the set of significant features able to describe the behaviour of the **Asset_maintainable_item** and the used term is compliant with previous CBM-related literature for SD and HA (Pimentel et al. 2014; Vachtsevanos et al. 2006);
- **Asset_Health_Index** is adherent in the meaning but different in the scope of usage; according to the extant literature, the term "health indicator" (or other

equivalents such as AHI) is used when the goal of the analysis is the PA (Niu, Yang, and Pecht 2010).

Therefore, in the data model, the **Significant_feature** is both used by the **State_detection_analysis** and the **Diagnosis_analysis**, while the **Asset_Health_Index** is used for **Prognosis_analysis**, which is the core of the proposed framework. In PA, a set of **Algorithm** elaborates the **Asset_Health_Index**, which was previously identified starting from the set of **Significant_feature** adopted during the **State_detection_analysis** according to a ND method.

Looking specifically at PA, the analyses could rely on a library of **Algorithm**, encompassing the following families: **AI_model**, **Statistical_model**, and **Physics_model**, with corresponding inheritances as summarised in Figure 5.

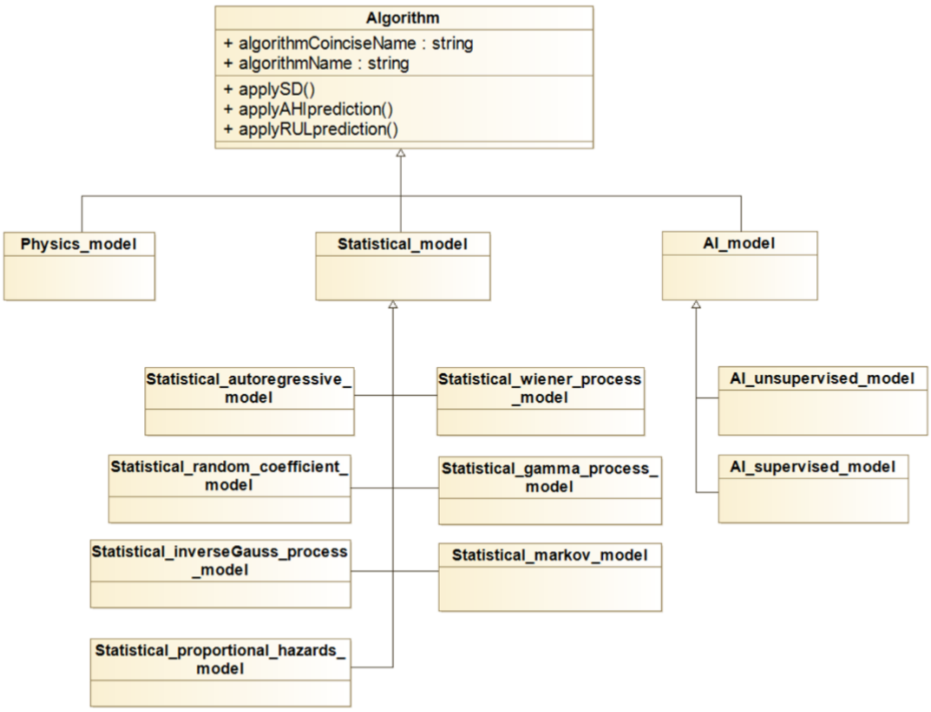


Figure 5. Data model: class Algorithm and its inheritances.

Theoretically, each of the above models may be applied, but their implementation is constrained by the availability of data (e.g. with or without RTF data).

5. Proof of Concept in an Industry 4.0-based Flexible Manufacturing Line

The framework is used in a PoC to guide the integration of ND and RUL in the context of an FML in the Industry 4.0 Lab *(the full name of the Lab with details and institution is removed to guarantee anonymity)*, which is a controlled environment offering industry-like challenges for operation and maintenance management. According to the results of a FMECA study, the drilling station is resulted as the most critical asset, whose details could be found in (Cimino, Negri, and Fumagalli 2019).

5.1 L0 – Asset Analysis

The drilling station is thus targeted, and its hierarchical decomposition and functioning are represented as data model instances in Figure 6 where, for the sake of simplicity, the class attributes are avoided.

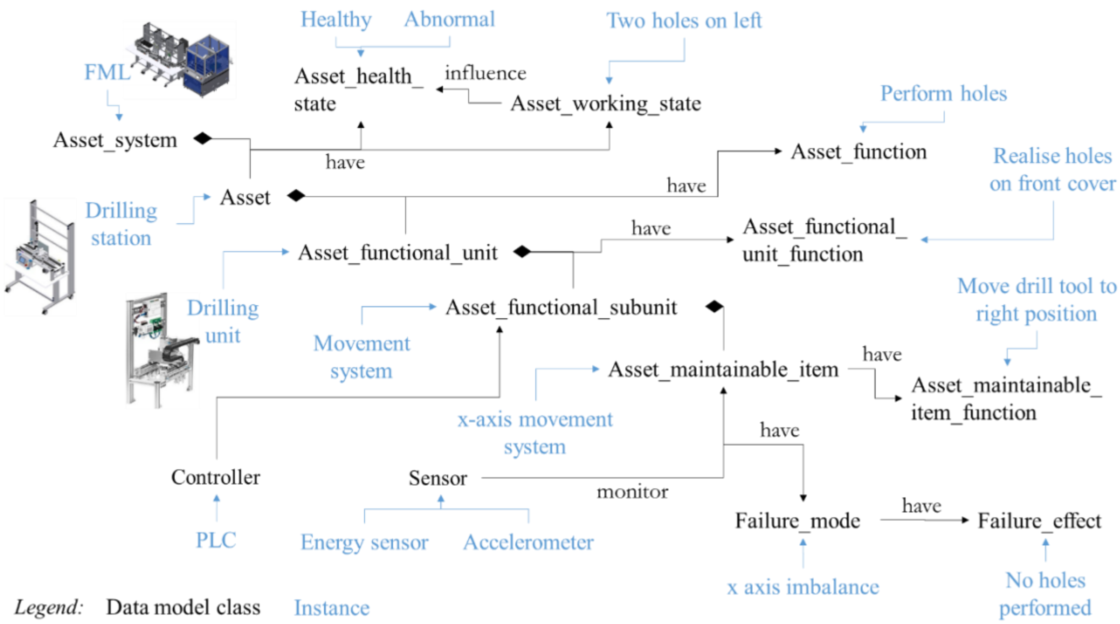


Figure 6. Data model instantiation for asset decomposition and functioning.

The acceleration provides useful information in case of drilling axis imbalance since this failure mode is related, through prior knowledge, to a vibration analysis, often used in the literature for rotating component (Heng et al. 2009).

5.2 L1 – Data Acquisition

The DA relies on the ICT architecture of the Industry 4.0 Lab, structured as in Figure 7.

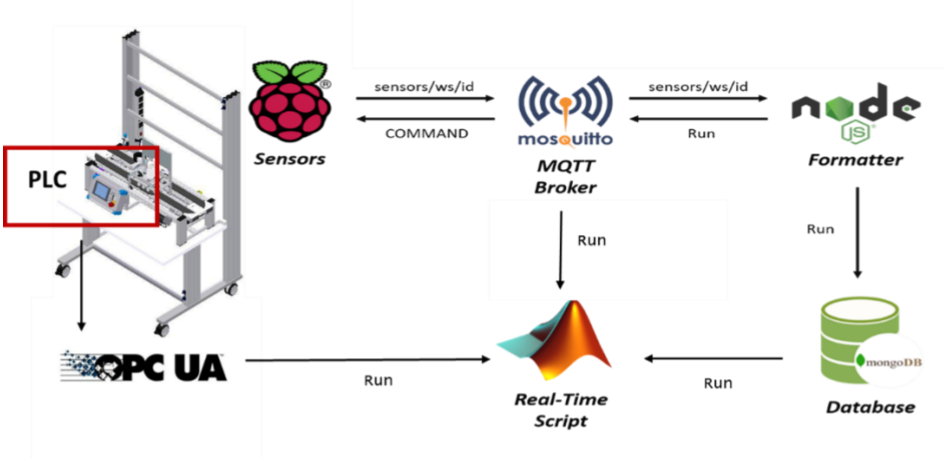


Figure 7. Industry 4.0 Lab ICT architecture.

The PLC of the drilling station provides signals mainly referred to the operational state. The MATLAB model establishes a real-time connection with the client-server through OPC-UA (Open Platform Communications Unified Architecture) protocol, a standard M2M (machine-to-machine) communication protocol according to IEC 62541-100. The MATLAB model then calculates the possible state: *Working*, *Idle*, *Error*, *Emergency* and *Energy Saving* (Cimino, Negri, and Fumagalli 2019). The working state triggers the acquisition of accelerations, which are measured through a Raspberry Pi sensor installed on the drilling axis (Figure 8), with a sampling frequency of 200 Hz.

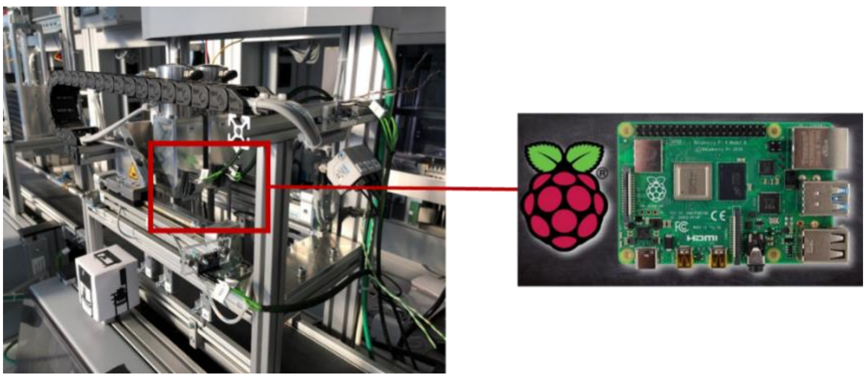


Figure 8. Raspberry Pi installed on drilling axes.

On average, each drilling cycle, making two holes on left (Figure 6), lasts for 11 seconds, thus collecting 2200 acceleration values then sent, through a MQTT protocol, to a predefined server that stores them in a non-relational document database, namely MongoDB. For each cycle, also date and timestamp are saved.

5.3 L2 – Data Manipulation

In Figure 9, the instantiation of the data model reports all classes and instances related to levels L2 to L5, discussed in the remainder.

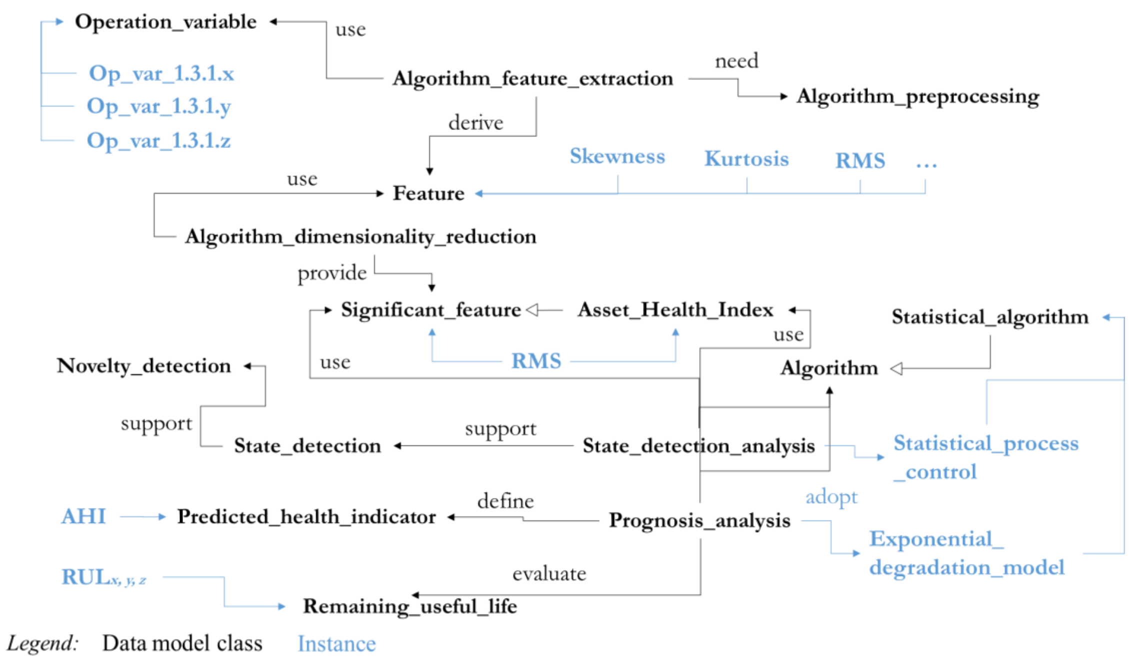


Figure 9. Data model instantiation for L2, L3, L4 and L5.

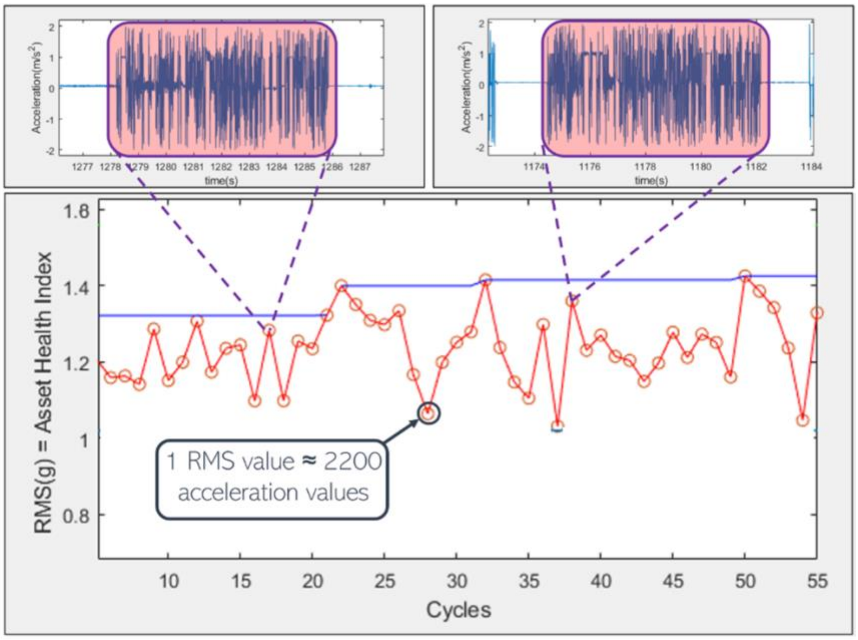
The RMS (Root Mean Square) is selected among the possible time-based statistical features of the acceleration signal, since it has a particular physical meaning. It characterizes the amount of dissipated energy during the working process due to vibrations (Večeř, Kreidl, and Šmíd 2005). In case damage occurs to the drilling axis, due to an imbalance, the vibration level increases and, as the damage progresses, the dissipated energy increases too, directly impacting on the RMS values.

Thus, the RMS of the axis is selected as both significant feature and AHI to evaluate the health of the drilling station. Thus, it is evaluated for all three axes, x, y, and z, as described in Equation 3 and 4.

$$RMS_{acc_j} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \qquad j = \{x, y, z\} \text{ ---------------Eq. 3}$$

$$AHI_j = RMS_j \quad \forall j = \{x, y, z\} \text{-----------Eq.4}$$

When a cycle is operated, a single RMS point, for each axis, summarises 2200 acceleration values, as cleared out in Figure 10.



RMS measured in *g*, that is the gravitational acceleration, almost equal to 9.81m/s^2

Figure 10. RMS real-time evaluation.

5.4 L3 – State Detection

According to the ND approach described in the framework, it is necessary to determine what level of RMS indicates the machine healthy state.

Since no RTF are available, an off-line production campaign of 100 equal workpieces is realized, measuring accelerations signals; these result from drilling working conditions that do not change in time, as in this PoC a unique kind of process/product is performed along the line. The RMS values are then calculated via MATLAB and they are assumed to describe the healthy state. According to the probabilistic ND approach, a Gaussian distribution is assumed for the RMS values on the three axes and a normality test is performed, with a significance level of 0.005. Figure 11 shows the probability plots.

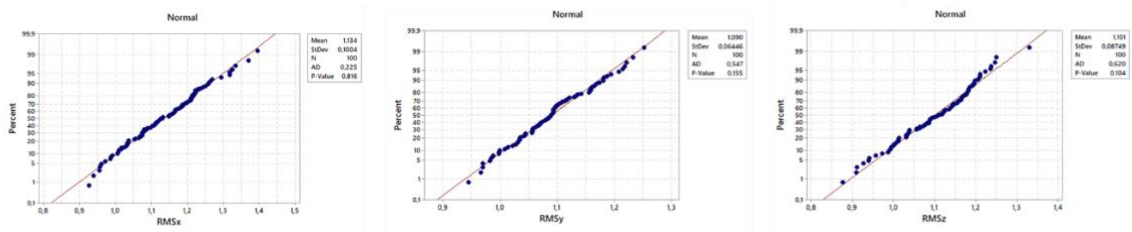


Figure 11. Probability plots for RMS values on the three axes.

The test is satisfied since the RMS value lie close to the straight line representing the normality condition (Johnson and Wichern 2002). Therefore, the healthy state is properly modelled through a Gaussian distribution and every new sample, collected in real-time, can be compared with the healthy class to see if it belongs to it or not. This is not critical for on-line monitoring since the comparison takes in total around 3 seconds, including OPC-UA connection and RMS evaluation, on a common laptop (i7-8th generation 3,5 GHz, 16 GB of memory). It guarantees a real-time computation with respect to the time requirements of the operations (each working cycle lasts around 11 seconds, a ND is computed in the same time scale of one cycle; in the FML under study this enables to avoid to process a next workpiece, if needed).

The comparison is made by defining a novelty threshold. For the Gaussian distribution, the novelty threshold is defined at a distance of three standard deviations (σ) from the mean (μ), i.e. $y_{abnormal}$ = P(μ + 3σ) (Grubbs 1969). Thus, every new RMS is labelled "healthy" if P(RMS) < $y_{abnormal}$ or "abnormal" otherwise, where $y_{abnormal}$ = P(RMS^{UP}) = P(μ + 3σ). This assumption is confirmed by previous work (Fumagalli et al. 2019), and RMS^{UP} can be described by Equation 5.

$$RMS_j^{UP} = \mu_j + 3\sigma_j \qquad j = \{x, y, z\} \text{----------Eq.5}$$

Such distance, regardless of μ and σ, always corresponds to $y_{abnormal}$ = P(RMS^{UP}) = P(μ + 3σ) = 0.9987. This means that a "healthy" RMS falls inside the healthy boundary limit in the 99.87 % of the cases. It is worth remarking that only the upper bound is calculated due to the physical meaning of RMS.

5.5 L4 – Health assessment

By classifying the RMS, the asset could be assessed if in a healthy or abnormal state. Yet, it is still needed to determine if the asset is faulty. Figure 12 summarises the three states in which the drilling station could be.

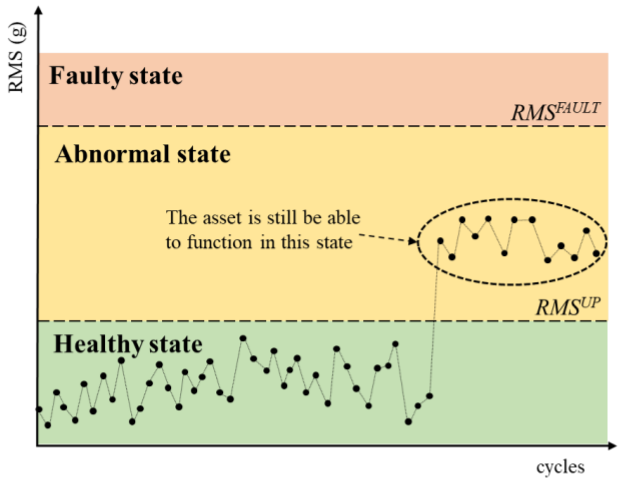


Figure 12. States of the drilling station.

While the RMS^{UP} has been previously set up, the threshold RMS^{FAULT} could not be fixed a priori since no previous RTF data are available, so the behaviour of the asset in the faulty state is unknown. In this PoC, a constant threshold for RMS^{FAULT} is fixed according to (Nectoux et al. 2012) and the faulty state starts when the RMS exceeds by four times RMS^{UP} (RMS^{FAULT} = 4RMS^{UP}).

According to the proposed framework, the RUL is not predicted unless the RMS enters in the abnormal state, allowing to save computational power for scaled up applications that may include several variables from the asset.

5.6 L5 – Prognostic assessment

Being the states defined, the prognostic assessment could start. The future behaviour of the AHI, i.e. RMS, is not known due to the lack of RTF data and its evolution must be assumed from scientific literature or company expertise. In this PoC, statistical methods are applied due to their wide use in case of missing RTF data. It results that the exponential trend of the RMS is a promising approximation of the real trend (Nectoux et al. 2012). More precisely, to model this exponential growth, the selected method is the Exponential Degradation Model (EDM) (Gebraeel et al. 2005), belonging to the family of random coefficient statistical methods, expressed in Equation 6 and Table 4.

$$\widehat{AHI}(t) = \widehat{RMS}(t) = \vartheta + \theta\, e^{\left(\beta t + \epsilon(t) - \frac{\sigma^2}{2}\right)} \text{--------------Eq.6}$$

Term	Definition
$\widehat{AHI}(t)$	Asset Health Index estimation for time *t*, in this case equal to the maximum of the RMS in the three directions *x*, *y*, *z*
ϑ	Intercept terms considered as a constant
θ	$\ln(\theta) \sim N(\mu_\theta, \sigma_\theta^2)$ – log-normal random variable
β	$\beta \sim N(\mu_\beta, \sigma_\beta^2)$ – random variable
$\epsilon(t)$	$\epsilon(t) \sim N(\mu = 0, \sigma^2)$ – noise term

Table 4. EDM terms definition.

In the beginning, when few data are available, the model is set up with an arbitrary prior distribution, described through high variance (i.e. the variances of θ and β are fixed equal to 10^6). As CM data becomes available, the model is updated, and the approximating function is adjusted accordingly. This is performed during the operation of the asset. The algorithm updates automatically the EDM random coefficient at the end of each working cycle allowing for better real-time performance of degradation process description and RUL prediction. Figure 13 reports an example for the RMS on the *z*-axis.

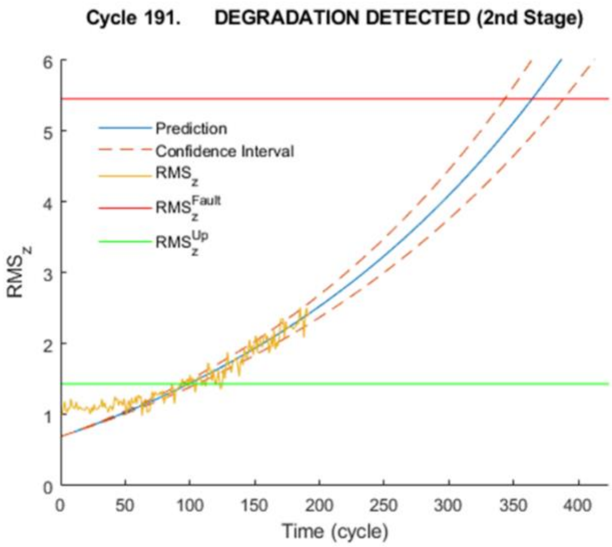


Figure 13. Example of real-time application in the drilling station for z-axis.

When the axis enters in the abnormal state, then the predicted trend of the $\widehat{AHI}$ (RMS) is plotted, together with an alert note "Degradation Detected" and the actual cycle number. In Figure 13, also the confidence interval (CI) is plotted (orange dashed line) for the predicted $\widehat{AHI}$. As expected, the CI is wide as the first abnormality is detected and narrows down as drilling station operates since the EDM random coefficients are updated based on new CM data, as shown in Figure 14: the prediction becomes more accurate as the cycles advance.

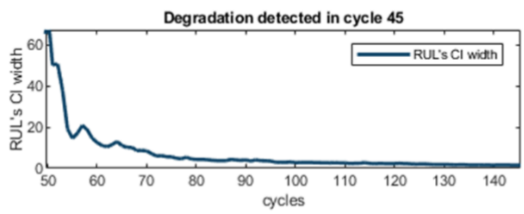


Figure 14. CI narrowing as drilling station operates.

Furthermore, it is possible to estimate the RUL since the statistical method provides a PDF, called $f_{\mathrm{RUL}}(t)$. The probability of failure P_f can be calculated according to Equation 2 (from the initial instant to the current time instant) and it serves as a basis for L6.

5.7 L6 – Advisory generation

The previous developments allow establishing an advisory generation based on the real-time values of the $\widehat{AHI}$s, which are RMS values of the three axes in our PoC. Therefore, the monitoring activity supports three actions:

- *No maintenance action* is required if $P(RMS_i) < P(RMS_i^{UP})$
- *Reactive action*: the drilling station must be stopped, and corrective maintenance should take place if $P(RMS_i) > P(RMS_i^{FAULT})$;
- *Proactive action* is based on the predicted $\widehat{AHI}$ and must take place in t^{days} as highlighted by the prognosis algorithm.

The t^{days} is evaluated by comparing the maximum allowed probability of failure P_f^{max} (or, equivalently, a minimum acceptable reliability R^{min}) and the current P_f. A conversion from the number of cycles t^{cycles} to t^{days} may be required, as expressed in Equation 7 and 8, completed by Table 5.

$$t^{cycles} = \left\{t : \int_o^t f_{RUL} dt = P_f^{max} = 1 - R^{min}\right\} \text{ ---------Eq.7}$$

$$t^{days} = \frac{t^{cycles}}{\overline{UR} \times PC_{daily}} \text{ ------Eq.8}$$

Term	Definition
t^{days}	Time units remaining until $R \le R^{min}$ [days]
t^{cycles}	Time units remaining until $R \le R^{min}$ [cycles]
$\overline{UR}$	Average utilization rate of the drill [%]
PC_{daily}	Daily production capacity of the drill [cycles/day]

Table 5. Terms to transform cycles in days.

In this PoC, the minimum acceptable reliability R^{min} is set to 0.98 since the station is fundamental to the overall work of the FML.

6. Conclusions

This research work investigates the integration of ND and RUL prediction in Industry

4.0-based manufacturing systems. In such advanced systems, real-time monitoring and predictability are relevant characteristics. Nevertheless, to implement the prediction of failures some limits typically exist in practice, when no historical and/or RTF data are available. This happens with newly commissioned or highly reliable machines, or due to poor recording activity. In these cases, the RUL prediction is particularly challenged since it is not possible to forecast future trend by looking at previous machine history.

The developed framework aims at extending the functionality levels proposed by the ISO 13374 and OSA-CBM standards, taken as reference, to improve particularly SD, HA and PA. It aims at integrating ND and RUL prediction through statistical modelling of the healthy state of the machine. The use of statistical models is indeed supported by the literature review. In the specific implementation of the PoC in a laboratory environment, the framework is built on EDM. It enables the real-time update of its coefficients, allowing the model to be more precise as long as more CM data are available. All this is built on top of a decomposition and understanding of the asset, as a preparatory phase.

Overall, the following contributions are claimed by the proposed framework:

- the framework provides a specific strategy for CBM/PHM implementation, tailored to the specific issue of no RTF data; in particular, it provides guidelines and supporting models both from the process and data viewpoints;
- the framework sets the implementation strategy based on industry-related knowledge, adopting a general and shared vocabulary rooted in several international standards, like the ISO 13374 and 13379-1, among others;
- the framework leverages on a library of algorithms (physics-based, statistical, and AI-based), each one with specific modelling capabilities; thus, for a certain algorithm to be applied, the case-related characteristics and boundary conditions (as availability of RTF data) should be evaluated;

- the framework implies assets working under dynamic working conditions; indeed, the proposed modelling of the asset working regimes, useful for ND for asset healthiness assessment, provides a seed to further explore in challenging manufacturing contexts with varying production mixes and processing requirements.

It is of particular interest the case when new products or processes are introduced due to varying production mixes. The approach suggested in this work may be adopted and further extended to cope with this variability: a change of working conditions is identified through ND algorithms (at the SD level) and changes of degradation model and future trend in the AHI are adjustable through the Random Coefficient modelling (at PA level). Future research is also required to extend the framework with higher-level analysis to drive the PHM process, especially linking the PHM process with asset prioritization based on risk management decisions. Finally, a prospective work may regard to connect this research with the investigation on Collaborative Prognostics of industrial assets. Exploiting this new theory would enable to manage the lack of RTF data through clustering the behaviour of assets taken from an entire fleet of assets, also considering their different manufacturing contexts (i.e. with varying production mixes and processing requirements). This will enable an application at scale of AI-based models, as the few or no RTF data collected by the single asset are now surpassed by the different premises of single collaborative assets present in the fleet.

References

Alasadi, S.A., and W.S. Bhaya. 2017. Review of Data Preprocessing Techniques in Data Mining. *Journal of Engineering and Applied Sciences* 12, no. 16: 4102–4107.

Beden, S.M., S. Abdullah, and A.K. Ariffin. 2009. Review of Fatigue Crack Propagation Models for Metallic Components. *European Journal of Scientific Research* 28, no. 3: 364–397.

Bever, K. 2007. Understanding MIMOSA's Open Standards for On-Board Health Assessment and Enterprise Application Integration. In *AIAA Infotech@ Aerospace 2007 Conference and Exhibit*, 2925.

Boutrot, J., Y. Giorgiutti, F. Rezende, and S. Barras. 2017. Reliable and Accurate Determination of Life Extension for Offshore Units. In *Offshore Technology Conference*. Offshore Technology Conference.

Caggiano, A. 2018. Cloud-Based Manufacturing Process Monitoring for Smart Diagnosis Services. *International Journal of Computer Integrated Manufacturing* 31, no. 7 (July 3): 612–623. https://doi.org/10.1080/0951192X.2018.1425552.

Cattaneo, L., and M. Macchi. 2019. A Digital Twin Proof of Concept to Support Machine Prognostics with Low Availability of Run-To-Failure Data. *IFAC-PapersOnLine* 52, no. 10: 37–42. https://doi.org/10.1016/j.ifacol.2019.10.016.

Chen, N., and K.L. Tsui. 2013. Condition Monitoring and Remaining Useful Life Prediction Using Degradation Signals: Revisited. *IIE Transactions* 45, no. 9: 939–952.

Cho, S., G. May, I. Tourkogiorgis, R. Perez, O. Lazaro, B. de la Maza, and D. Kiritsis. 2018. A Hybrid Machine Learning Approach for Predictive Maintenance in Smart Factories of the Future. In *IFIP International Conference on Advances in Production Management Systems*, 311–317. Springer.

Cimino, C., E. Negri, and L. Fumagalli. 2019. Review of Digital Twin Applications in Manufacturing. *Computers in Industry* 113: 103–130. https://doi.org/10.1016/j.compind.2019.103130.

Cubillo, A., S. Perinpanayagam, and M. Esperon-Miguez. 2016. A Review of Physics-Based Models in Prognostics: Application to Gears and Bearings of Rotating Machinery. *Advances in Mechanical Engineering* 8, no. 8: 1–21.

Ćwikła, G. 2014. Methods of Manufacturing Data Acquisition for Production

Management-a Review. In *Advanced Materials Research*, 837:618–623. Trans Tech Publ.

Flath, C.M., and N. Stein. 2018. Towards a Data Science Toolbox for Industrial Analytics Applications. *Computers in Industry* 94: 16–25.

Fumagalli, L., L. Cattaneo, I. Roda, M. Macchi, and M. Rondi. 2019. Data-Driven CBM Tool for Risk-Informed Decision-Making in an Electric Arc Furnace. *International Journal of Advanced Manufacturing Technology*: 1–14.

Gao, R., L. Wang, R. Teti, D. Dornfeld, S. Kumara, M. Mori, and M. Helu. 2015. Cloud-Enabled Prognosis for Manufacturing. *CIRP Annals* 64, no. 2: 749–772. http://www.sciencedirect.com/science/article/pii/S000785061500150X.

García, S., S. Ramírez-Gallego, J. Luengo, J.M. Benítez, and F. Herrera. 2016. Big Data Preprocessing: Methods and Prospects. *Big Data Analytics* 1, no. 1: 9.

Garetti, M., L. Fumagalli, and E. Negri. 2015. Role of Ontologies for CPS Implementation in Manufacturing. *Management and Production Engineering Review* 6, no. 4: 26–32.

Gebraeel, N.Z., M.A. Lawley, R. Li, and J.K. Ryan. 2005. Residual-Life Distributions from Component Degradation Signals: A Bayesian Approach. *IiE Transactions* 37, no. 6: 543–557.

Glaessgen, E., and D. Stargel. 2012. The Digital Twin Paradigm for Future NASA and US Air Force Vehicles. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA*, 1818.

Grubbs, F.E. 1969. Procedures for Detecting Outlying Observations in Samples. *Technometrics* 11, no. 1: 1–21.

Guillén, A.J., A. Crespo, M. Macchi, and J. Gómez. 2016. On the Role of Prognostics and Health Management in Advanced Maintenance Systems. *Production Planning* *and Control* 27, no. 12: 991–1004. http://dx.doi.org/10.1080/09537287.2016.1171920.

Heng, A., S. Zhang, A.C.C. Tan, and J. Mathew. 2009. Rotating Machinery Prognostics: State of the Art, Challenges and Opportunities. *Mechanical Systems and Signal Processing* 23, no. 3: 724–739.

Hsieh, Y.-S., F.-T. Cheng, H.-C. Huang, C.-R. Wang, S.-C. Wang, and H.-C. Yang. 2012. VM-Based Baseline Predictive Maintenance Scheme. *IEEE Transactions on Semiconductor Manufacturing* 26, no. 1: 132–144.

Huang, H.-Z., H.-K. Wang, Y.-F. Li, L. Zhang, and Z. Liu. 2015. Support Vector Machine Based Estimation of Remaining Useful Life: Current Research Status and Future Trends. *Journal of Mechanical Science and Technology* 29, no. 1: 151–163.

ISO 55000. 2014. Asset Management — Overview, Principles and Terminology. *BSI Standards Publication*.

Jardine, A.K.S., D. Lin, and D. Banjevic. 2006. A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance. *Mechanical Systems and Signal Processing* 20, no. 7: 1483–1510.

Johnson, R.A., and D.W. Wichern. 2002. *Applied Multivariate Statistical Analysis*. Prentice hall Upper Saddle River, NJ.

Kraft, J., and S. Kuntzagk. 2017. Engine Fleet-Management: The Use of Digital Twins From a MRO Perspective. In *ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition*. American Society of Mechanical Engineers Digital Collection.

Kumar, A., R.B. Chinnam, and F. Tseng. 2019. An HMM and Polynomial Regression Based Approach for Remaining Useful Life and Health State Estimation of Cutting Tools. *Computers & Industrial Engineering* 128: 1008–1014.

Lee, G.-Y., M. Kim, Y.-J. Quan, M.-S. Kim, T.J.Y. Kim, H.-S. Yoon, S. Min, D.-H.

Kim, J.-W. Mun, and J.W. Oh. 2018. Machine Health Management in Smart Factory: A Review. *Journal of Mechanical Science and Technology* 32, no. 3: 987–1009.

Lee, J., B. Bagheri, and H.A. Kao. 2015. A Cyber-Physical Systems Architecture for Industry 4.0-Based Manufacturing Systems. *Manufacturing Letters* 3: 18–23. http://dx.doi.org/10.1016/j.mfglet.2014.12.001.

Lei, Y., N. Li, L. Guo, N. Li, T. Yan, and J. Lin. 2018. Machinery Health Prognostics: A Systematic Review from Data Acquisition to RUL Prediction. *Mechanical Systems and Signal Processing* 104: 799–834.

Lei, Y., N. Li, and J. Lin. 2016. A New Method Based on Stochastic Process Models for Machine Remaining Useful Life Prediction. *IEEE Transactions on Instrumentation and Measurement* 65, no. 12: 2671–2684.

Leturiondo, U., O. Salgado, L. Ciani, D. Galar, and M. Catelani. 2017. Architecture for Hybrid Modelling and Its Application to Diagnosis and Prognosis with Missing Data. *Measurement* 108: 152–162.

Li, C., S. Mahadevan, Y. Ling, S. Choze, and L. Wang. 2017. Dynamic Bayesian Network for Aircraft Wing Health Monitoring Digital Twin. *Aiaa Journal* 55, no. 3: 930–941.

Li, N., N. Gebraeel, Y. Lei, L. Bian, and X. Si. 2019. Remaining Useful Life Prediction of Machinery under Time-Varying Operating Conditions Based on a Two-Factor State-Space Model. *Reliability Engineering & System Safety* 186: 88–100.

Li, Y., and U. Roy. 2015. Challenges in Developing a Computational Platform to Integrate Data Analytics with Simulation-Based Optimization. In *ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers Digital Collection.

Liu, Q., M. Dong, W. Lv, X. Geng, and Y. Li. 2015. A Novel Method Using Adaptive Hidden Semi-Markov Model for Multi-Sensor Monitoring Equipment Health Prognosis. *Mechanical Systems and Signal Processing* 64: 217–232.

Liu, Y.-Y., M.-H. Hung, Y.-C. Lin, C.-C. Chen, W.-L. Gao, and F.-T. Cheng. 2018. A Cloud-Based Pluggable Manufacturing Service Scheme for Smart Factory. In *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, 1040–1045. IEEE.

Liu, Z., N. Meyendorf, and N. Mrad. 2018. The Role of Data Fusion in Predictive Maintenance Using Digital Twin. In *AIP Conference Proceedings*, 1949:20023. AIP Publishing LLC.

Luo, J., A. Bixby, K. Pattipati, L. Qiao, M. Kawamoto, and S. Chigusa. 2003. An Interacting Multiple Model Approach to Model-Based Prognostics. In *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483)*, 1:189–194. IEEE.

Luo, W., T. Hu, W. Zhu, and F. Tao. 2018. Digital Twin Modeling Method for CNC Machine Tool. In *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, 1–4. IEEE.

Di Maio, F., K.L. Tsui, and E. Zio. 2012. Combining Relevance Vector Machines and Exponential Regression for Bearing Residual Life Estimation. *Mechanical Systems and Signal Processing* 31: 405–427.

Moyne, J., and J. Iskandar. 2017. Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing. *Processes* 5, no. 3: 39.

Napoleone, A., M. Macchi, and A. Pozzetti. 2020. A Review on the Characteristics of Cyber-Physical Systems for the Future Smart Factories. *Journal of Manufacturing Systems* 54: 305–335.

Nectoux, P., R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier. 2012. PRONOSTIA: An Experimental Platform for Bearings Accelerated Degradation Tests. In .

Negri, E., L. Fumagalli, and M. Macchi. 2017. A Review of the Roles of Digital Twin in CPS-Based Production Systems. *Procedia Manufacturing* 11, no. June: 939–948. http://linkinghub.elsevier.com/retrieve/pii/S2351978917304067.

Niu, G., B.S. Yang, and M. Pecht. 2010. Development of an Optimized Condition-Based Maintenance System by Data Fusion and Reliability-Centered Maintenance. *Reliability Engineering and System Safety* 95, no. 7 (July): 786–796.

Pan, D., J.-B. Liu, and J. Cao. 2016. Remaining Useful Life Estimation Using an Inverse Gaussian Degradation Model. *Neurocomputing* 185: 64–72.

Patnaik, P., and X. Wu. 2018. Linking MRO to Prognosis Based Health Management Through Physics-of-Failures Understanding. In *ASME Turbo Expo 2018: Turbomachinery Technical Conference and Exposition*. American Society of Mechanical Engineers Digital Collection.

Pellegrino, J., M. Justiniano, A. Raghunathan, and B.A. Weiss. 2016. *Measurement Science Roadmap for Prognostics and Health Management for Smart Manufacturing Systems*.

Penas, O., R. Plateaux, S. Patalano, and M. Hammadi. 2017. Multi-Scale Approach from Mechatronic to Cyber-Physical Systems for the Design of Manufacturing Systems. *Computers in Industry* 86: 52–69.

Pham, H.T., B.-S. Yang, and T.T. Nguyen. 2012. Machine Performance Degradation Assessment and Remaining Useful Life Prediction Using Proportional Hazard Model and Support Vector Machine. *Mechanical Systems and Signal Processing* 32: 320–330.

Pimentel, M.A.F., D.A. Clifton, L. Clifton, and L. Tarassenko. 2014. A Review of

Novelty Detection. *Signal Processing* 99: 215–249. http://dx.doi.org/10.1016/j.sigpro.2013.12.026.

Polenghi, A., L. Fumagalli, and I. Roda. 2018. Role of Simulation in Industrial Engineering: Focus on Manufacturing Systems. In *IFAC-PapersOnLine*, 51:496–501. Elsevier B.V. https://doi.org/10.1016/j.ifacol.2018.08.367.

Polenghi, A., I. Roda, M. Macchi, and P. Trucco. 2019. Risk Sources Affecting the Asset Management Decision-Making Process in Manufacturing: A Systematic Review of the Literature. In *Advances in Production Management Systems. Production Management for the Factory of the Future. APMS 2019. IFIP Advances in Information and Communication Technology*, ed. F. Ameri, K. Stecke, G. von Cieminski, and D. Kiritsis, 566:274–282. Springer, Cham.

Qi, Q., and F. Tao. 2018. Digital Twin and Big Data towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. *IEEE Access* 6: 3585–3593.

Raman, V., and M. Hassanaly. 2019. Emerging Trends in Numerical Simulations of Combustion Systems. *Proceedings of the Combustion Institute* 37, no. 2: 2073–2089.

Rúbio, E.M., R.P. Dionísio, and P.M.B. Torres. 2018. Industrial IoT Devices and Cyber-Physical Production Systems: Review and Use Case. In *International Conference on Innovation, Engineering and Entrepreneurship*, 292–298. Springer.

Saha, B., K. Goebel, and J. Christophersen. 2009. Comparison of Prognostic Algorithms for Estimating Remaining Useful Life of Batteries. *Transactions of the Institute of Measurement and Control* 31, no. 3–4: 293–308.

Shubenkova, K., A. Valiev, V. Shepelev, S. Tsiulin, and K.H. Reinau. 2018. Possibility of Digital Twins Technology for Improving Efficiency of the Branded Service System. In *2018 Global Smart Industry Conference (GloSIC)*, 1–7. IEEE.

Si, X.-S., W. Wang, C.-H. Hu, M.-Y. Chen, and D.-H. Zhou. 2013. A Wiener-Process-

Based Degradation Model with a Recursive Filter Algorithm for Remaining Useful Life Estimation. *Mechanical Systems and Signal Processing* 35, no. 1–2: 219–237.

Si, X.-S., W. Wang, C.-H. Hu, and D.-H. Zhou. 2011. Remaining Useful Life Estimation–a Review on the Statistical Data Driven Approaches. *European Journal of Operational Research* 213, no. 1: 1–14.

Sikorska, J.Z., M. Hodkiewicz, and L. Ma. 2011. Prognostic Modelling Options for Remaining Useful Life Estimation by Industry. *Mechanical Systems and Signal Processing* 25, no. 5: 1803–1836.

Sivalingam, K., M. Sepulveda, M. Spring, and P. Davies. 2018. A Review and Methodology Development for Remaining Useful Life Prediction of Offshore Fixed and Floating Wind Turbine Power Converter with Digital Twin Technology Perspective. In *2018 2nd International Conference on Green Energy and Applications (ICGEA)*, 197–204. IEEE.

Le Son, K., M. Fouladirad, and A. Barros. 2016. Remaining Useful Lifetime Estimation and Noisy Gamma Deterioration Process. *Reliability Engineering & System Safety* 149: 76–87.

Tao, F., M. Zhang, Y. Liu, and A.Y.C. Nee. 2018. Digital Twin Driven Prognostics and Health Management for Complex Equipment. *CIRP Annals* 67, no. 1: 169–172.

Tao, T., E. Zio, and W. Zhao. 2018. A Novel Support Vector Regression Method for Online Reliability Prediction under Multi-State Varying Operating Conditions. *Reliability Engineering & System Safety* 177: 35–49.

Tian, Z. 2012. An Artificial Neural Network Method for Remaining Useful Life Prediction of Equipment Subject to Condition Monitoring. *Journal of Intelligent Manufacturing* 23, no. 2: 227–237.

Tuegel, E. 2012. The Airframe Digital Twin: Some Challenges to Realization. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials*

Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA, 1812.

Tygesen, U.T., K. Worden, T. Rogers, G. Manson, and E.J. Cross. 2019. State-of-the-Art and Future Directions for Predictive Modelling of Offshore Structure Dynamics Using Machine Learning. In *Dynamics of Civil Structures, Volume 2*, 223–233. Springer.

Vachtsevanos, G., F. Lewis, M. Roemer, A. Hess, and B. Wu. 2006. *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Ed. Jhon Wiley and Sons. Wiley Hoboken.

Vathoopan, M., M. Johny, A. Zoitl, and A. Knoll. 2018. Modular Fault Ascription and Corrective Maintenance Using a Digital Twin. *IFAC-PapersOnLine* 51, no. 11: 1041–1046.

Večeř, P., M. Kreidl, and R. Šmíd. 2005. Condition Indicators for Gearbox Condition Monitoring Systems. *Acta Polytechnica* 45, no. 6.

Vogl, G.W., B.A. Weiss, and M.A. Donmez. 2014. Standards for Prognostics and Health Management (PHM) Techniques within Manufacturing Operations. In *Annual Conference of the Prognostics and Health Management Society 2014*, 576–588. National Institute of Standards and Technology Gaithersburg United States.

Wang, H., H. Wang, G. Jiang, J. Li, and Y. Wang. 2019. Early Fault Detection of Wind Turbines Based on Operational Condition Clustering and Optimized Deep Belief Network Modeling. *Energies* 12, no. 6: 984.

Wilder-James, E. 2016. Breaking down Data Silos. *Harvard Business Review*.

Xu, L. Da, E.L. Xu, and L. Li. 2018. Industry 4.0: State of the Art and Future Trends. *International Journal of Production Research* 56, no. 8: 2941–2962.

Yeo, B., and D. Grant. 2018. Predicting Service Industry Performance Using Decision Tree Analysis. *International Journal of Information Management* 38, no. 1: 288–

300.

Zaccaria, V., M. Stenfelt, I. Aslanidou, and K.G. Kyprianidis. 2018. Fleet Monitoring and Diagnostics Framework Based on Digital Twin of Aero-Engines. In *ASME Turbo Expo 2018: Turbomachinery Technical Conference and Exposition*. American Society of Mechanical Engineers Digital Collection.

Zemouri, R., and R. Gouriveau. 2010. Towards Accurate and Reproducible Predictions for Prognostic: An Approach Combining a RRBF Network and an Autoregressive Model. *IFAC Proceedings Volumes* 43, no. 3: 140–145.