



# Prognostic Model Development with Missing Labels

## A Condition-Based Maintenance Approach Using Machine Learning

Patrick Zschech · Kai Heinrich · Raphael Bink ·  
Janis S. Neufeld

Received: 25 June 2018 / Accepted: 19 March 2019 / Published online: 1 April 2019  
© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2019

**Abstract** Condition-based maintenance (CBM) has emerged as a proactive strategy for determining the best time for maintenance activities. In this paper, a case of a milling process with imperfect maintenance at a German automotive manufacturer is considered. Its major challenge is that only data with missing labels are available, which does not provide a sufficient basis for classical prognostic maintenance models. To overcome this shortcoming, a data science study is carried out that combines several analytical methods, especially from the field of machine learning (ML). These include time-domain and time–frequency domain techniques for feature extraction, agglomerative hierarchical clustering and time series clustering for unsupervised pattern detection, as well as a recurrent neural network for prognostic model training. With the approach developed, it is possible to replace decisions that were made based on subjective criteria with data-driven decisions to increase the tool life of the milling machines. The

solution can be employed beyond the presented case to similar maintenance scenarios as the basis for decision support and prognostic model development. Moreover, it helps to further close the gap between ML research and the practical implementation of CBM.

**Keywords** Condition-based maintenance · Predictive maintenance · Prognostics · Big data analytics · Data science study · Machine learning · Unsupervised learning · Missing labels

### 1 Introduction

The maintenance function plays a central role in today's industrial value creation as it helps manufacturing companies to remain productive and competitive. It aims at ensuring a plant's functionality and environmental safety while keeping costs and resources at a low level to operate profitably (Peng et al. 2010; Muchiri et al. 2011). To adequately meet such superior objectives, a decision must be made regarding when necessary maintenance actions should be carried out. For this purpose, condition-based maintenance (CBM) has emerged in recent years as a proactive decision-making strategy, observing a system's health condition to determine the time and type of intervention (Jardine et al. 2006). As such, it is possible to reduce the uncertainty of maintenance actions and avoid unnecessary work by taking actions only when there is evidence of abnormal behavior (Peng et al. 2010).

The implementation of CBM is greatly promoted by the ubiquitous use of advanced information and communication technologies (ICT) that simplify the collection of large and multifaceted data, often referred to as 'big data' (Zschech 2018; Bumblauskas et al. 2017; Meeker and

---

Accepted after two revisions by the editors of the special issue.

---

P. Zschech (✉) · K. Heinrich  
Lehrstuhl für Wirtschaftsinformatik, Business Intelligence  
Research, TU Dresden, 01062 Dresden, Germany  
e-mail: patrick.zschech@tu-dresden.de

K. Heinrich  
e-mail: kai.heinrich@tu-dresden.de

R. Bink  
Herzogenrath R&D Centre, Saint-Gobain, 52134 Herzogenrath,  
Germany  
e-mail: raphael.bink@saint-gobain.com

J. S. Neufeld  
Lehrstuhl für BWL, Industrielles Management, TU Dresden,  
01062 Dresden, Germany  
e-mail: janis\_sebastian.neufeld@tu-dresden.de

Hong 2014). According to Manyika et al. (2011), for example, nearly two exabytes of newly generated data were estimated in the manufacturing sector alone in 2010, ranging from production status and equipment utilization data to records of tool and machinery condition monitoring (CM). CM data are of particular interest for maintenance purposes because they provide a basis for system health assessment using modern sensor technology with the capability of measuring a multitude of parameters at high frequencies. Thus, it is possible to continuously monitor health indicators in real time to trigger rapid actions in the case of undesirable changes and also to collect large historical data to identify patterns for anticipatory prognostic models that predict events before they occur (Meeker and Hong 2014; Bousdekis et al. 2015; Jardine et al. 2006).

However, it cannot always be assumed that the big data assets generated and stored within ICT-driven manufacturing settings provide all necessary input required for adequate decision support. There are several reasons why critical information might be lacking in real industrial application scenarios. In the case of critical machines, for example, the aim is to avoid failures and faults through strictly short maintenance intervals. As a result, no thresholds and tolerance limits are known or can be observed that provide labels to mark necessary points of intervention. In addition, sensors, which are able to describe physical health conditions directly (e.g., crack size, state of wear, etc.), are rarely used. Moreover, due to the pressure to use plants efficiently, it is often not possible to carry out test runs that go beyond the limits of safe conditions. Consequently, possible data observations might be truncated before the actual end of life, and thus, interesting events to describe fault patterns are not recorded (Susto et al. 2015; Leturiondo et al. 2017). Overall, such circumstances can be characterized by the absence of a prospective target variable on which to build the prognosis. This problem, henceforth called missing labels, can be seen as a major hurdle in the development of adequate prognostic models (Gouriveau et al. 2013).

To address this problem and show how it is possible to provide maintenance decision support in this unfortunate situation, a solution approach is developed by conducting a data science study based on a real-world case of a German car manufacturer with an imperfect maintenance situation. Multiple analytical methods are applied, especially from the field of machine learning (ML). The challenge here was to support the decision-making process of a wear-induced replacement of a milling machine by predicting the remaining useful life (RUL) when no labels are present in the dataset due to individual risk preferences and poor available information.

The major contributions of this paper can be summarized as follows:

- A classification of prognostic CBM approaches and, in particular, different label situations is provided and serves as a systematization to structure the field and position future research.
- To the best of our knowledge, this is one of the first attempts to address the problem of missing labels in the context of building prognostic decision models. Therefore, the paper provides a novel solution for a known and relevant problem within the interdisciplinary domain of CBM-based prognostics and ML.
- For the studied real-world problem, a suitable combination of methods for feature extraction as well as model building via unsupervised and supervised learning is identified by a comparison of several analytical approaches.
- Finally, the data science study demonstrates how implicit empirical knowledge of machine operators, which is only latently available in recorded data assets, can be made tangible for better decision support.

The rest of this paper is organized as follows. In Sect. 2, the conceptual background of prognostic CBM approaches is structured, narrowing the scope for which a new solution is provided. For this purpose, a systematization of different label situations is contributed, and existing RUL prediction approaches from related work are identified. Section 3 provides the context of the maintenance case and describes the proposed conceptual solution to overcome the challenges depicted. In Sect. 4, the structure of the applied data science study is outlined, and the implementation consisting of multiple steps is demonstrated. Finally, the results are discussed, a conclusion is drawn, and an outlook for further research is presented in Sect. 5.

## 2 Conceptual Background

### 2.1 Prognostic Approaches in Condition-Based Maintenance

CBM approaches generally consist of two central components: (i) diagnostics dealing with fault detection, isolation and identification when any abnormality occurs and (ii) prognostics dealing with RUL prediction of operating machines using suitable indicators before malfunctions occur. As such, prognostics (also known as ‘predictive maintenance’) can be considered more efficient for achieving zero downtime performance, while diagnostics are still required when fault prediction fails and a fault occurs (Jardine et al. 2006). Prognostic solutions in CBM can be classified in different ways, depending on the type of data and knowledge available and the methods applied (Zschech 2018). For this reason, there are several review

papers summarizing existing work in the field from slightly different perspectives (e.g., Jardine et al. 2006; Dragomir et al. 2009; Peng et al. 2010; Si et al. 2011; Veldman et al. 2011; Ahmad and Kamaruddin 2012; Bousdekis et al. 2015; Elattar et al. 2016; Vogl et al. 2016). Basically, it can distinguish between (i) physical models, (ii) knowledge-based models and (iii) data-driven models, comprising statistics and ML, while the focus of this article is primarily on ML.

Physical models are usually built on a thorough understanding of physical mechanisms (e.g., specific degradation laws), whereas knowledge-based approaches, such as expert systems or fuzzy logic, try to simulate human thinking (Dragomir et al. 2009; Peng et al. 2010; Elattar et al. 2016). Data-driven approaches, in contrast, use collected data observations to identify and model relationships that can be used for RUL predictions on new data observations. For this purpose, statistical approaches model the conditional distribution of time lapses to failure given the history of CM data using several approaches such as hidden Markov models (HMM) or estimation of stochastic process parameters. While these approaches provide useful estimations of RUL and risk quantification of the solutions, they rely heavily on underlying assumptions of distributions or underlying processes from the Wiener or Gamma family. In most cases, these assumptions cannot be verified as fulfilled due to the bias introduced by truncated processes (e.g., sometimes a tool is changed before RUL is reached) in practice (Wang and Christer 2000; Si et al. 2011).

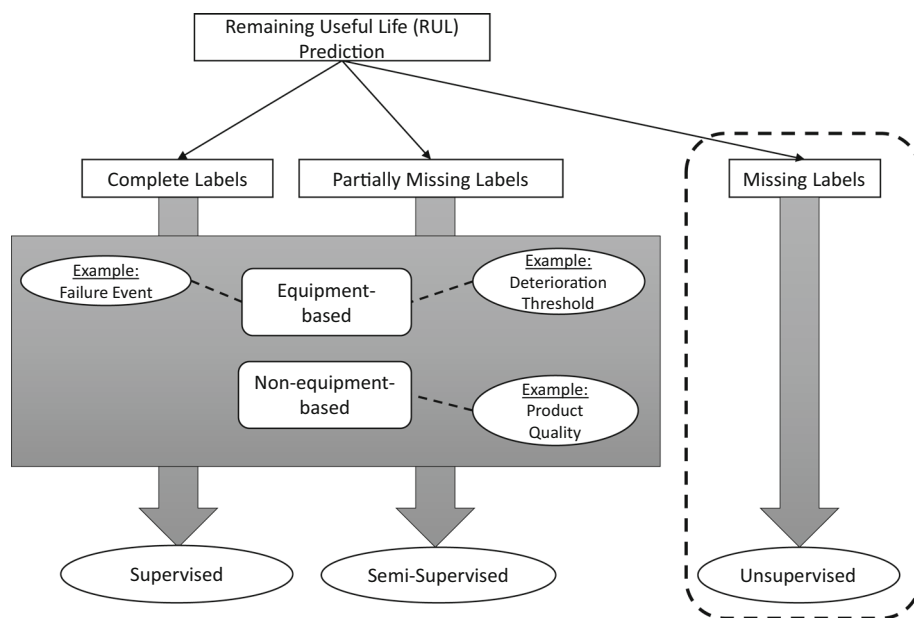
To overcome this issue, methods from the field of ML can be applied (Breiman 2001). If abundant data are available, ML methods have the advantage of learning hidden relations about system behavior that are difficult to directly measure with sensors due to internal processes such as wear and tear, where the inferential process does not require any, or only weak, assumptions due to the validation mechanism of sample splitting embedded in statistical learning theory (Vapnik 1999; Hansen 2000; Rinaldo et al. 2016). In this case, comprehensive system knowledge is not required because ML algorithms such as artificial neural networks (ANN), support vector machines and decision trees are able to determine complex, non-linear relationships between high-dimensional CM data and the RUL of a system (Peng et al. 2010; Elattar et al. 2016; Vogl et al. 2016). However, a key requirement is the availability of representative training data that reflect all symptomatic behavior of the system, from normal and faulty operations to degradation patterns under certain operating conditions (Dragomir et al. 2009; Tian et al. 2010; Elattar et al. 2016). Therefore, a further distinction must be made between different levels of data availability determining the solution approach to be applied.

## 2.2 Systematization of Different Label Situations and Related Work

In general, two prerequisites must be given for the training of prognostic ML models: (i) feature variables to describe the input data and (ii) output data to label the target variable to be predicted (Gouriveau et al. 2013). Focusing on CBM-based models, the input is given by CM data, which, in a broad sense, comprise any data having a connection with the RUL prediction, such as monitored conditions, degradation signals, operational data or performance records. Possible sources include, for example, pressure, temperature, vibration, moisture, humidity, loading, speed or oil analysis data (Si et al. 2011). Considering the target variable, different label situations can occur that can be classified into (i) complete labels, (ii) partially missing labels and (iii) missing labels (cf. Fig. 1).

A situation of complete labels occurs when all data observations for each cycle reflect the actual end of useful life. Clearly, the definition of ‘useful life’ depends on the individual expectations of a machine owner (Si et al. 2011); thus, different labeling strategies are feasible. In the case of run-to-failure (R2F) policies, for example, it seems reasonable to consider failure events to label the end of useful life (Susto et al. 2015). Another option is the application of predefined tolerance limits or deterioration thresholds of the CM variables, which can be specified, for example, by equipment vendors, domain experts or simulation and test runs. Beyond such equipment-based labels, it is also possible to consider alternative labels from closely related manufacturing functions, such as quality control or yield management, to measure the quality of produced items or other individually defined performance indicators (e.g., material utilization, productivity) to determine the definition of ‘useful life’ (e.g., Muchiri et al. 2011; Si et al. 2011; Cheng et al. 2018; Choudhary et al. 2009). Overall, complete labels can be considered an ideal basis for prognostic model development since ML algorithms can be readily implemented in a supervised learning fashion. For this reason, most existing work focuses on this area, as demonstrated by recent contributions based on real-world scenarios (e.g., Susto et al. 2015; Cline et al. 2017; Ullah et al. 2017). In most cases, however, supervised approaches are developed on the basis of synthetic data, such as the C-MAPSS datasets as a frequently used example (Saxena and Goebel 2008). These datasets were explicitly created for data-driven model development based on an R2F scenario, where Ramasso and Saxena (2014) identified more than seventy contributions proposing different solutions. The results also showed a high ratio of ML methods, with different types of ANN being most commonly used, such as multilayer perceptron (MLP) or recurrent neural networks (RNNs). This was also confirmed by other

**Fig. 1** Differentiation of label situations and reference to ML approaches for RUL prediction



prognostic survey papers with broader review scopes (e.g., Peng et al. 2010; Jardine et al. 2006), where ANNs were dominantly used for supervised learning.

The situation of partially missing labels, on the other hand, is given when only a part of all observed cycles is marked with relevant target information. This may be the case, for example, if the product quality is used for labeling, but no quality measurements are available for all manufactured products. In such situations, semi-supervised learning methods can be applied, which use unlabelled data together with labeled data for labeling purposes. Possible solutions were proposed by Yuan and Liu (2013) and Zhao et al. (2011). Another reason could be that in some cases, maintenance actions are carried out well before the occurrence of critical events such as failures; thus, the observations do not only consist of full cycles, but also include suspensions truncating the data records. In this context, Tian et al. (2010) proposed a solution approach to demonstrate how both failure histories and suspension histories can be used in combination for model training.

However, the worst possible starting point for prognostic model development occurs when no labeling information is available at all. To achieve zero downtime, critical assets are usually not allowed to fail, which results in missing event data. Therefore, defining CM thresholds is often a challenging task (Tian et al. 2010; Susto et al. 2015). Depending on the type of CM setting used, not all types of CM data, such as vibration or pressure, are capable of directly describing the underlying state of a system, which in turn could create the necessity for obtaining additional event data (Si et al. 2011). In addition, manual labeling of CM data can be considered expensive due to the efforts required to integrate field knowledge of experienced

human annotators (Zhao et al. 2011; Yuan and Liu 2013). If, on the other hand, non-equipment-based measures such as product quality are used for labeling, it is not always guaranteed that this information can be directly assigned to the corresponding machine operations. Furthermore, quality inspections often require considerable effort or are difficult to integrate into existing manufacturing processes, especially in complex, hierarchical settings. Overall, these circumstances lead to a situation of missing labels without appropriate prediction targets. This can only be addressed with the help of unsupervised learning methods that seek to identify hidden structures and patterns without any target specifications (Susto et al. 2015). However, existing approaches that apply such methods for labeling purposes in the context of CBM-based prognostics are rather scarce. Some solutions apply self-organizing feature maps (SOFM), a specific type of unsupervised ANN to learn structures from highly deviating, non-linear data for the purpose of detecting malfunctions and degradation indicators (e.g., Jämsä-Jounela et al. 2003; Huang et al. 2007). Another approach was proposed by Baruah et al. (2006), which combined principal component analysis (PCA) for dimensionality reduction with an unsupervised clustering technique to identify and prospectively predict different operating modes of equipment behavior. However, these approaches are not sufficient in situations where, due to a lack of missing labeling information, it cannot be assessed whether maintenance actions were performed too late or too early, hindering the development of a prognostic model for decision support. In addition, the missing label information in the problem class described in this paper is caused by the absence of missing links to direct or indirect equipment wear indicators, as all maintenance actions to

prevent wear and tear are based on the individual experiences of the human machine operators. This is where the current study contributes a new solution approach.

### 3 Description of the Case and Solution Development

#### 3.1 Current Maintenance Situations and Challenges

The current case describes the production scenario of a German car manufacturer facing an inefficient maintenance strategy for a frequently used machine tool. Specifically, the scenario concerns the production of a ball hub that is to be installed in a drive shaft in subsequent assembly steps to enable the power transmission from the gearbox to the wheels within the vehicle. Due to the central location, the importance for customer safety, tight specifications and a highly complex manufacturing procedure, the company made the strategic decision to keep the production of the ball hub in-house instead of procuring it externally. For this reason, it is of the utmost importance for the competitiveness of the company that the production runs as efficiently as possible.

The process for manufacturing the ball hub comprises several sub-steps, such as forging, case hardening and milling. The sub-step of milling is of particular importance because it is carried out by a machine tool with replaceable milling tools and is, therefore, subject to permanent maintenance actions that are performed by human operators. The milling machine is responsible for providing the ball hub with six ball raceways. This procedure involves various components, such as the main spindle, a milling spindle, a tilting table and several machine axes, which are all equipped with sensors to collect comprehensive data about the executed machine operations. Due to material processing, the tools of the milling machine are subject to natural wear and tear over their entire lifetime, which results in increasing wear marks on the cutting edge of the milling tools. This effect also has a considerable impact on the quality of the manufactured product, since at a certain level, the tools show such a high extent of wear that they are considered to be damaged and thus adversely affect the milling result. Therefore, maintenance actions in the sense of various smaller corrections have to be carried out to reduce wear and tear until the milling tool finally needs to be replaced. This type of maintenance is also known as imperfect maintenance, where the health of a system is not always restored to its “as good as new” conditions (Cheng et al. 2018).

In the past, such maintenance actions were initiated on the basis of subjective decisions of the machine operators by taking a milled product from the production process and measuring it in a checking fixture. The measured deviations

from the nominal properties of the product are then used to determine the extent of wear. At the same time, the checking fixture specifies necessary parameter corrections required to reduce the wear effect. However, the crucial decision as to whether a worn milling tool must be replaced by a new tool due to increased damage symptoms, or whether imperfect corrections are sufficient, must be made by the machine operator at his or her discretion based on visual tool inspections. In general, the principle applies that the longer a milling tool is in use, the higher the extent of wear, and thus, the risk of rejects for a manufactured ball hub increases. At the same time, however, the longer tools remain in use, the lower the tooling costs are due to fewer replacements for the same number of produced parts. This results in a trade-off between tooling costs and impaired product quality. Simultaneously, the damage of tools is not only subject to natural wear processes, since it is assumed that other accompanying factors, such as externally caused vibrations, faulty tool installations or even dirt particles, additionally influence the course of wear and tear and thus accelerate the occurrence of damage.

In a situation of complete information, this decision problem could be solved, for example, by considering all necessary constraints (i.e., tooling costs, tool condition, level of product quality) and determining the best time for tool replacements (e.g., Cheng et al. 2018). In the current setting, however, such information is not available. Despite extensive data records on machine behavior, no thresholds of adequate condition indicators are known or have been specified. Similarly, the influential factors that are assumed to affect the occurrence of accelerated tool damage are either not explicitly confirmed or they cannot be captured adequately in order to use them for an improved maintenance policy. Moreover, appropriate indicators are missing to assess the quality of milled products. This is due to the fact that the quality of a processed ball hub can only be determined at a very late stage in the entire manufacturing process, which is why it is not directly traceable to a particular milling tool. Thus, the machine operators’ replacement decisions are exclusively based on their perception during the visual tool inspections, their empirical knowledge and their individual risk preferences. Consequently, less experienced machine operators with a more risk-averse attitude tend to replace tools well before the actual end of useful life, while risk-affine machine operators tend to carry out late replacements risking impaired product quality. Overall, this leads to inefficient use of resources, which is why a solution approach for better decision support is required.

### 3.2 Conceptualization of the Solution Approach

To overcome the problem of subjective decisions, a solution is proposed that aims to provide machine operators with a model that reflects the course of wear and tear and is therefore suitable for predicting the tools' RUL. This makes it possible to check whether a critical limit has been reached or whether a tool can still be used. The development of such a prognostic model takes advantage of the available data assets recorded during machine operations to extract useful information by means of ML methods. In particular, the idea was to extract the implicit knowledge of those operators who made correct decisions in the past leading to longer tool life. This should allow the entire workforce to benefit from the empirical knowledge of more experienced machine operators, using the prognostic model as a tool for communication and a reference point for avoiding individual risk preferences.

However, since no direct labels were provided due to missing CM thresholds and non-traceable results from quality control, such required information to assess the quality of an operator's decision first had to be extracted from the available data records to ensure adequate model training. In other words, it was first necessary to separate "good decisions" from "bad decisions" based on latently available information hidden in historical observations about executed tool replacements. Therefore, the problem space was conceptualized using two orthogonally related dimensions. The first dimension refers to the time when a tool replacement was carried out and is therefore closely related to the expression of a tool's useful life. Here, it can be determined whether a replacement was performed at an early or a late stage by considering all information that reflects the utilization of a particular tool over time, such as produced quantities or the amount of executed corrections. However, to determine whether an early or a late replacement is justified, it was also necessary to consider the state of the tool. Therefore, the second dimension refers to the condition, distinguishing between damaged and undamaged milling tools. Even if this information is not directly available in the data, it is reasonable to assume that a critical damage pattern must also be reflected within the

recorded CM values of one of the milling machine's components. By separating the two levels in both dimensions, a 4-field matrix can be set up as illustrated in Table 1. Based on this matrix, it is possible to differentiate between the following four types of tool replacements due to subjective decisions:

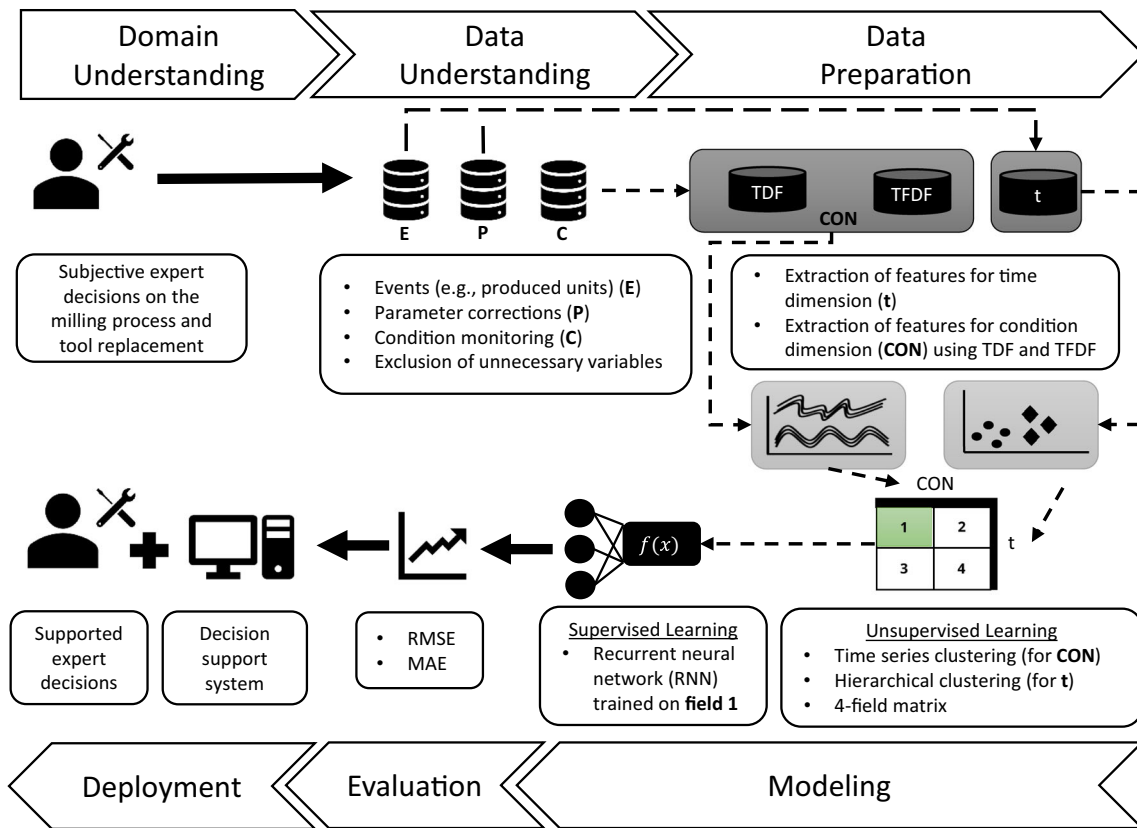
- *Type 1* represents undamaged tools that have been replaced correctly at a late time, implying an efficient use of resources.
- *Type 2* represents damaged tools that have not been replaced in time, leading to impaired product quality.
- *Type 3* represents undamaged tools that have been replaced too early, resulting in high tool costs and truncated data for model training.
- *Type 4* represents damaged tools that have been replaced correctly at an early time, also corresponding to an efficient use of resources.

To provide machine operators with better support for future decisions and to ensure more resource-efficient tool replacements of types 1 and 4, two analytical models are required: A diagnostic model continuously checks on the basis of condition indicators whether a milling tool shows any signs of impending damage. If this is the case, it will be replaced. If this is not the case, a prognostic model trained on "type 1" observations is used to estimate the RUL of the tool since "type 1" observations represent tools that were correctly replaced at a late time, implying that these replacements are close to the actual end of useful life based on empirical knowledge of more experienced machine operators.

In the following, the scope is primarily limited to the differentiation of the recorded tool observations into the four types presented above and the development of a prognostic model. For the first task, methods from the field of unsupervised ML were used to detect structures that can be used for labeling purposes, while for the latter task the results were applied in combination with supervised ML to develop a prognostic model that can predict the RUL of the milling tools in productive use. The development of a diagnostic model, on the other hand, is only partially addressed in this article. The identification of hidden

**Table 1** 4-field matrix for the distinction of different tool replacements

	Condition	
	Tool undamaged	Tool damaged
Time		
Replacement late	Type 1 – GOOD (efficient tool usage, type 3 prevented)	Type 2 – AVOID (impaired product quality)
Replacement early	Type 3 – AVOID (high tool costs)	Type 4 – GOOD (efficient tool usage, type 2 prevented)



**Fig. 2** Overall solution approach with reference to case characteristics and applied methods

structures based on unsupervised learning can be used to detect, isolate and describe faults. However, for the establishment of a comprehensive diagnostic model, deeper system knowledge and a more profound consideration are required, such as differentiating between different failure modes (e.g., continuous deterioration vs. intermittent effects). For this reason, it was not explicitly considered at this stage and is subject to further research.

#### 4 Data Science Study

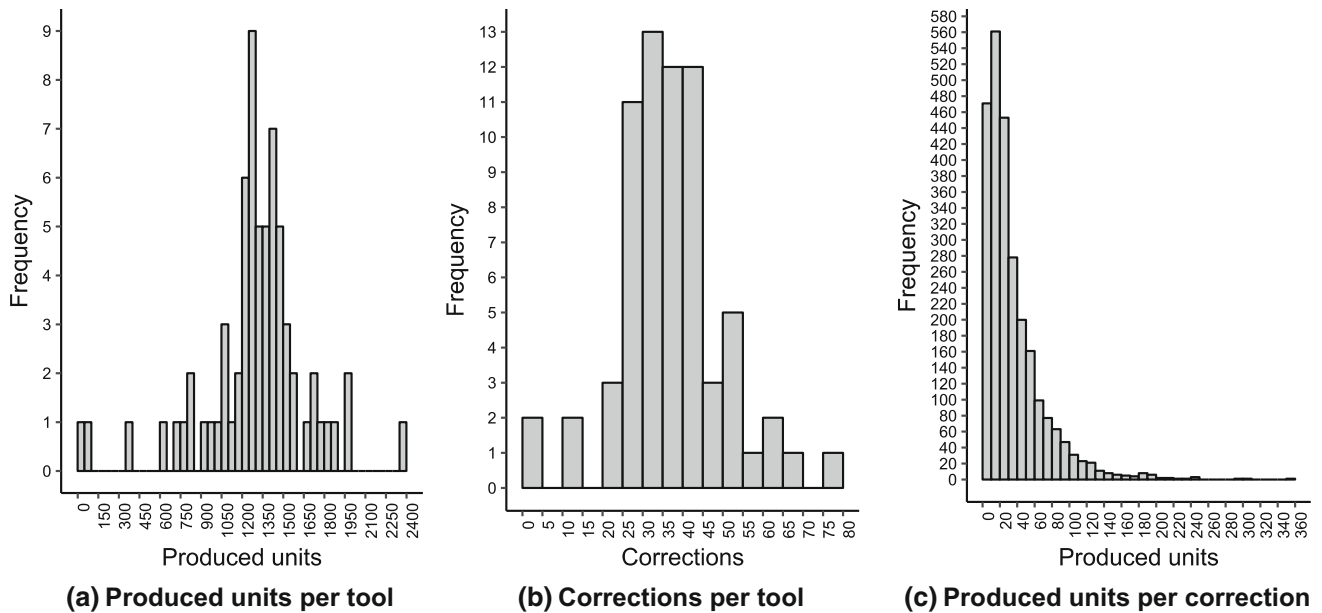
For the implementation of the conceptually derived solution, a data science study was carried out consisting of multiple analytical methods. While all the programming was performed with the statistical software R, the methodical procedure was based on a systematic process model from the field of ‘knowledge discovery and data mining’ (KDDM). In particular, a six-step approach served as a guideline consisting of the following phases: (i) domain understanding, (ii) data understanding, (iii) data preparation, (iv) modeling, (v) evaluation, and (vi) deployment. These steps can be considered a common foundation of multiple existing KDDM process models as identified by Kurgan and Musilek (2006) and thus serve as

a basis to guide the implementation of data science projects in research and practice. While the first phase of establishing an understanding of the domain problem has already been described thoroughly in the previous section, the remaining five steps are presented below. Figure 2 illustrates the overall approach and refers to the characteristics of the case and the analytical methods applied. Further details are described in each corresponding step.

##### 4.1 Data Understanding

The case study partner provided a representative dataset reflecting the process behavior and the operations of the milling machine at execution time. Data records were made available in the form of distributed, structured text files with a total size of 9 gigabytes. Those text files contained the following information:

- Comprehensive machine messages on executed operations, including event data for logging the number of produced ball hubs, as well as the time of maintenance interventions in terms of tool replacements and smaller parameter corrections.
- Control data for tracing the extent of executed parameter corrections.



**Fig. 3** Distributions of the event data

- Fine-granular sensor data of the different components involved in the milling machine.

The above-mentioned entries were recorded with a uniform timestamp to ensure correct mapping between different data entities. In the following, the most important variables regarding event data, control data and sensor data are briefly described to provide a better data understanding.

Screening the event data, the provided dataset contained information on an output volume of 88,125 produced ball hubs. During the processing of these parts, a total of 67 tool replacements and 2551 parameter corrections were recorded. Thus, on average, 1315.3 units per tool were produced, and 38.07 parameter corrections per tool were carried out. In addition, an average of 34.55 units were produced between two parameter corrections. The distribution of those properties, especially the observed variance measured in produced units as depicted in Fig. 3a, indicated the potential for further improvement (cf. Fig. 3a–c).

Considering the parameter corrections, it was possible to distinguish between different control variables, such as milling speed, feed rate, roll radius, angle of attack, system level and diameter corrections. However, with the exception of the system level and the diameter corrections, the remaining control variables could be neglected for further analysis, as they remained constant over time except for a few outliers. After conducting a deeper exploratory analysis, it was also necessary to exclude the system level corrections as an influencing variable due to their low variance. This could also be confirmed by the experience of the domain experts. Thus, the focus for further analysis was exclusively on the diameter corrections. This variable

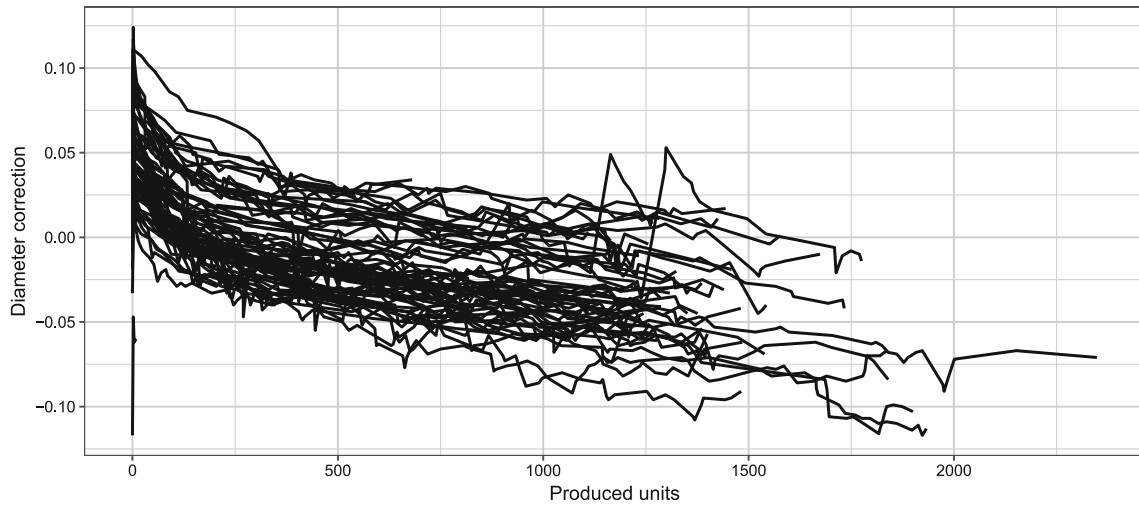
usually followed a typical pattern over the course of a tool lifecycle, as depicted in Fig. 4. After a new tool setup, a large correction of the diameter is carried out because the material of the new tool has not yet worn out. During the course of the milling process, the diameter is increasingly adjusted into the negative range to correct the distance to the workpiece due to the impact of wear and tear.

In addition to the event and control data, extensive sensor data of the milling machine were recorded in terms of condition indicators, such as various workload variables for machine axes and spindles. Of particular importance was the variable “load.axis.c”, which describes the load of the milling spindle. Figure 5 shows exemplary courses of the variable for five units, all produced with the same milling tool but at different points in time. The six recurring sections are due to the ball raceways of the ball hub, and the negative peak at the beginning of the process is caused by a routine removal of milled chips. Since this is a highly standardized procedure, each milling cycle has the same duration. However, when comparing the five milling cycles depicted, different levels can be observed due to the impact of wear and tear. Hence, this variable reflects an increasing change within the machine condition over time and thus could be used as an indicator of the RUL prediction of the milling tools.

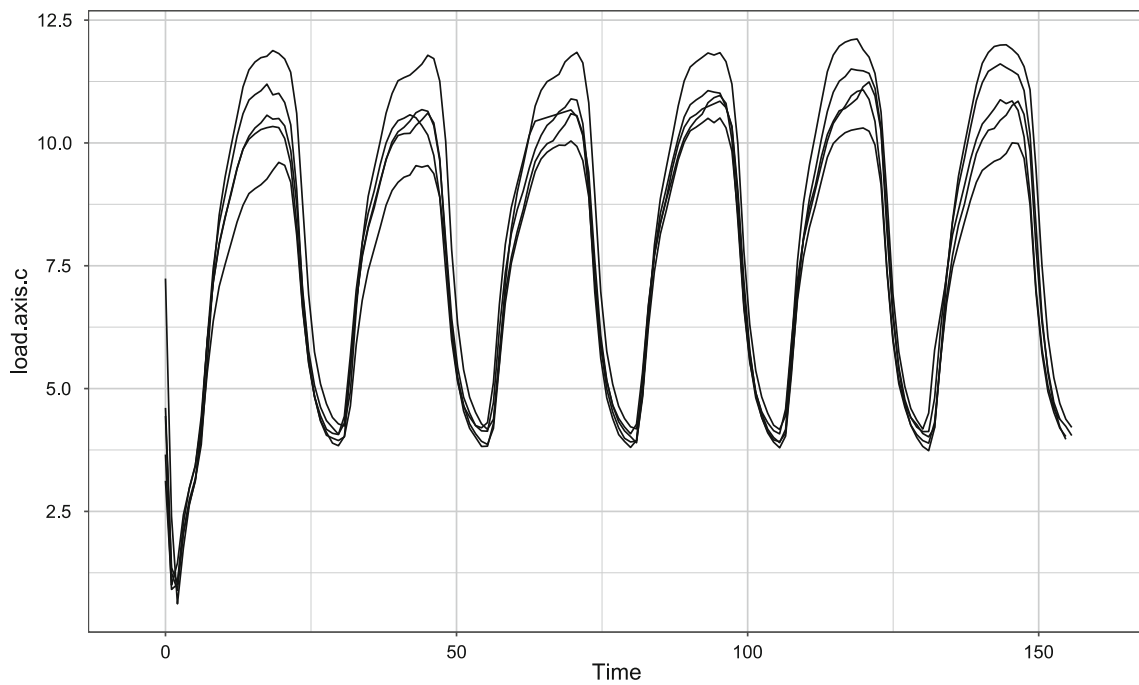
#### 4.2 Data Preparation

In the next phase, the data basis was pre-processed so that it could be used for the subsequent modeling task. Following the 4-field matrix derived from the domain





**Fig. 4** Course of the diameter corrections over tool lifecycles



**Fig. 5** Exemplary courses for the load of the milling spindle during the milling process

understanding, representative data variables for the two dimensions *time* and *condition* had to be selected and prepared accordingly.

For the *time dimension*, two types of variables were used to represent early or late tool replacements. The first variable was the number of produced ball hubs for each individual milling tool. This variable, called “produced.quantity”, could directly be derived from the recorded event data. The second variable was an additionally derived feature, called “cumulated.corrections”. This feature was created based on the parameter

corrections by calculating the cumulative, maximal possible diameter corrections per tool. Thus, it was possible to reflect the residual material capacity of the milling tools, which is increasingly reduced by wear and tear during the milling process.

For the *condition dimension*, the given sensor data were used, focusing on the previously described variable “load.axis.c”. Since this variable contained time series data for each milling cycle in terms of 155 measuring points (cf. Fig. 5), further pre-processing was required to use it for subsequent modeling tasks. For this purpose, several

techniques for feature extraction were applied. Thus, it was possible to reduce dimensionality, remove noise and extract informative properties from the time series. In particular, this included the extraction of time-domain features (TDF) and time–frequency domain features (TFDF) (Goyal and Pabla 2015). Moreover, the time series were previously tested for stationarity by a Dickey-Fuller test (Dickey and Fuller 1979) since TDFs can only be used for stationary signals. Eventually, the following eight TDFs were extracted: peak value, root mean square, standard deviation, kurtosis value, crest factor, clearance factor, impulse factor and shape factor (Galar et al. 2012). For the extraction of the TFDFs, a short-time Fourier transform was used, which can also be applied to non-stationary signals (Aghabozorgi et al. 2015). The features generated in this way can be understood as partial frequency bands. These are less intuitively interpretable than the TDFs, but they represent alternative latent information that may deliver valuable inputs for later predictive modeling purposes. In a final pre-processing step, the dimensionality of all extracted features was reduced. This was done by maintaining only those features with high explanatory power. In the case of the TDFs, the peak value, kurtosis value and crest factor were selected using pairwise correlation analysis, while in the case of TFDFs, only those frequency bands were kept that followed a clearly recognizable trend over the entire lifetime of the milling tools.

### 4.3 Modeling

After data preparation, the actual modeling step was carried out. Due to the conceptualized solution approach above, this step included two subsequent tasks: (i) application of unsupervised ML to detect structural patterns within the data observations and to assign them into the previously described 4-field matrix, and (ii) application of supervised ML to train a prognostic model based on those observations that led to the “right decisions” (i.e., longer tool lifetimes) in the past. For the first task, individual clustering techniques (Everitt et al. 2011) were applied for both the time dimension and the condition dimension, and the second task was implemented using RNNs (Williams 1995).

#### 4.3.1 Unsupervised Learning: Application of Clustering Techniques

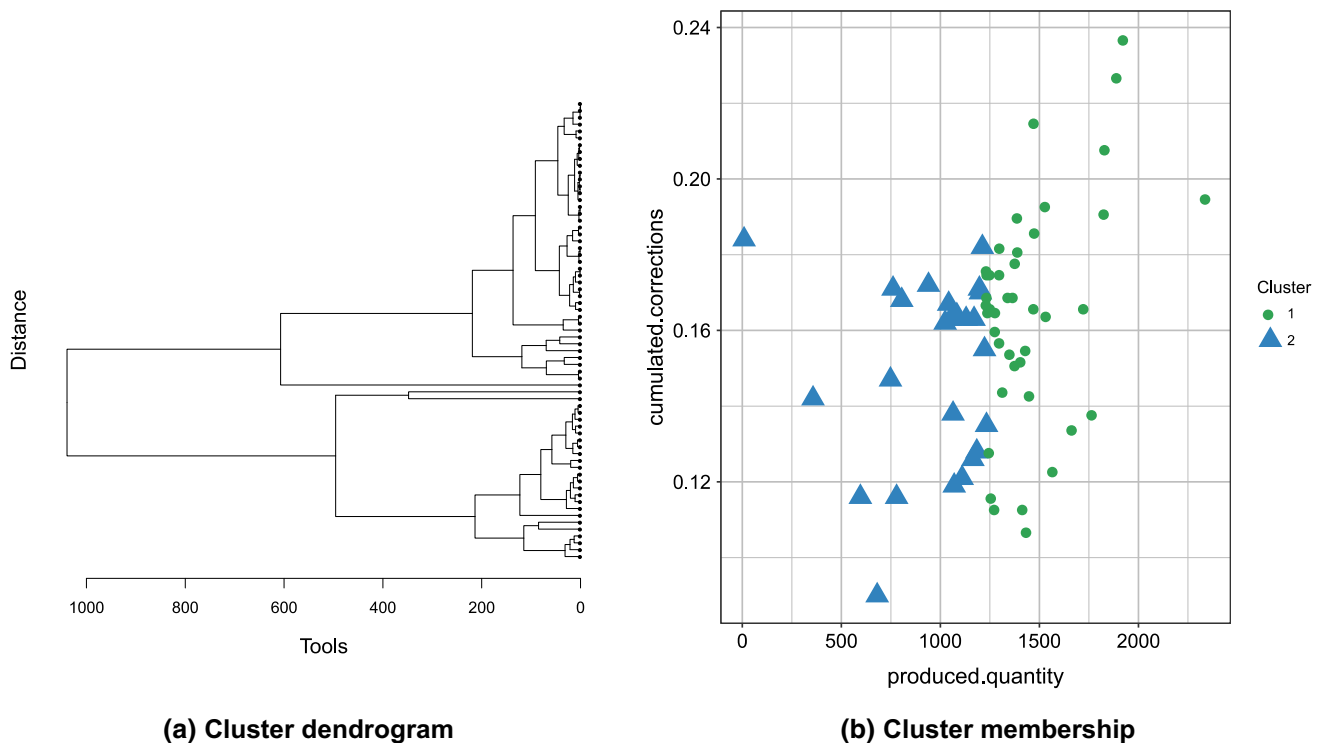
For clustering the time dimension based on the two variables “cumulated.corrections” and “produced.quantity”, an agglomerative hierarchical method was applied using a WPGMC approach (weighted pair-group method using centroid) (Sneath and Sokal 1973). With this approach, clusters are merged together in the order in which the

fusion leads to the smallest increase in variance, while the centroids of the clusters are evaluated equally to prevent the dominance of larger clusters. The choice of a hierarchical method is motivated by the need for an exploratory application where it is possible to generate different partitions based on hierarchical structures. The use of the WPGMC method is supported by its ability to create homogeneous groups, which facilitates the interpretation of the clusters. In addition to the hierarchical method, two more cluster methods were applied for comparison purposes: k-means and partitioning around medoids (PAM) as an implementation of the k-medoid algorithm (Van der Laan et al. 2003). For the evaluation, the three different algorithm classes and their implementations were compared by employing the common metrics: connectivity as a measure of cluster connectedness as well as Dunn index and silhouette width as combined measures of cluster compactness and cluster separation (Brock et al. 2008). The comparison results are displayed in Table 2. The results confirm that the WPGMC implementation of hierarchical clustering yields the best results in terms of all three cluster properties, as reflected by the minimum connectivity and maximum Dunn and silhouette values. It is also confirmed that  $k = 2$  is the optimal number of clusters in that case.

Figure 6a, b depict the cluster results after applying the hierarchical clustering. In summary, partitioning with two larger clusters can be identified, where their fusion leads to the greatest increase in variance as demonstrated within the dendrogram in Fig. 6a at a distance level of 600–1000. Cluster 1 represents tool replacements at a late time. It also includes observations just below the average tool lifetime of 1315.3 units. Cluster 2, however, can be interpreted as a cluster of early replacements and comprises observations with a rather short tool life, as depicted in Fig. 6b.

**Table 2** Cluster algorithm evaluation for different numbers of clusters  $k$

Cluster algorithm	Metric	Number of clusters ( $k$ )		
		2	3	4
Hierarchical (WPGMC)	Connectivity	3.0290	7.7238	11.8857
	Dunn	0.2157	0.1360	0.1360
	Silhouette	0.6286	0.5835	0.5201
k-means	Connectivity	4.6313	12.6583	15.3857
	Dunn	0.0115	0.0365	0.0297
	Silhouette	0.4414	0.6239	0.5773
k-medoid (PAM)	Connectivity	7.8619	11.7758	12.9710
	Dunn	0.0138	0.0580	0.0290
	Silhouette	0.5990	0.6053	0.5544

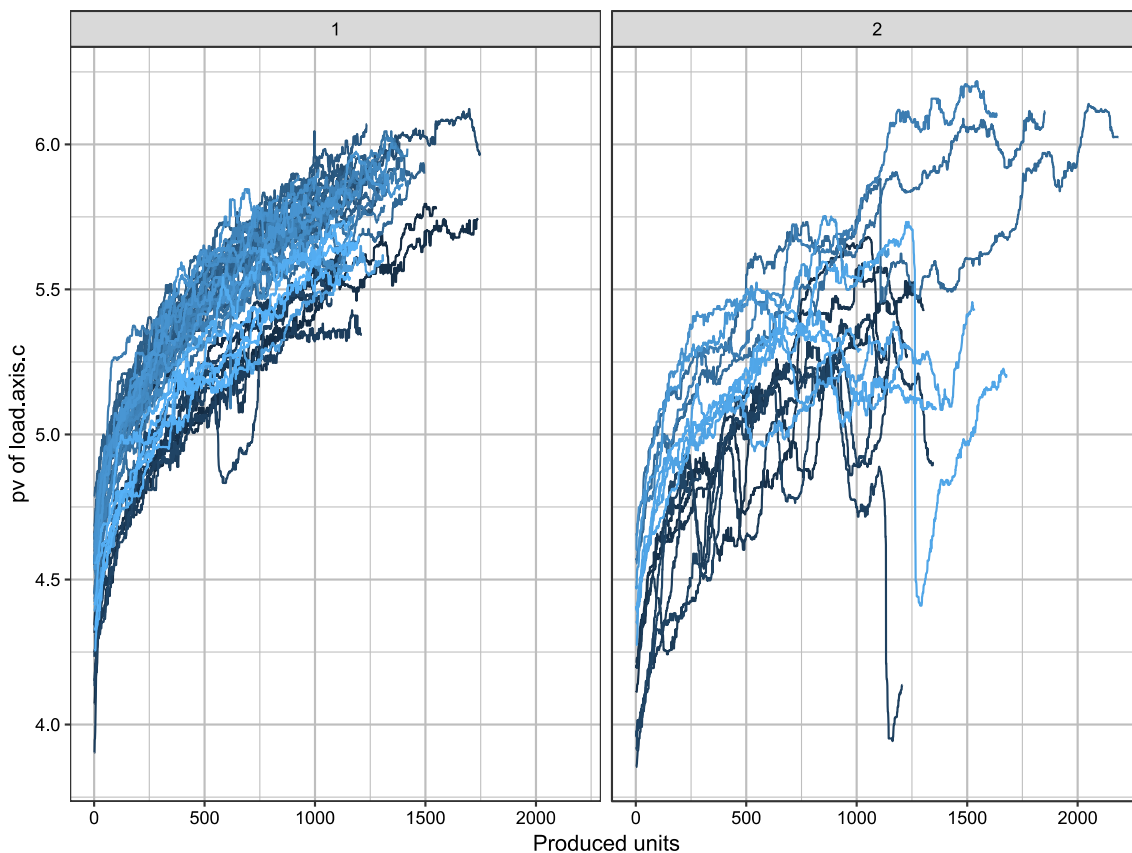


**Fig. 6** Cluster analysis results of the time dimension

When clustering the condition dimension, the aim was to distinguish between observations in which a flawless course over a tool's lifetime could be observed and observations in which a tool was subject to damage due to a critical level of wear and tear. For the latter case, it can further be assumed that any tool damage also shows a noticeable reflection within the recorded CM data of the milling machine in the form of remarkable changes or fault-specific signatures. For this purpose, the broad spectrum of extracted features of the variable "load.axis.c" was used to examine their temporal progression over the entire lifetime of each tool to detect patterns that could be useful for distinguishing temporal sequences of damaged tools from undamaged tools. While in this step the majority of features could immediately be excluded because the corresponding sequences were either too homogeneous or too heterogeneous to each other, the remaining features were examined in more detail. This was done by visually comparing individual sections across the entirety of all sequences as well as by applying the ultimately used cluster methods described below. In addition, experienced machine operators were consulted to integrate their domain knowledge and to collectively discuss observed particularities. As a result, the feature "peak value" (pv) was selected as the most relevant feature for clustering the condition dimension because it was suitable to reflect remarkable changes during the temporal progression of the

machine conditions, while simultaneously it could be used to identify two characteristic groups of sequences.

For the computational determination of the different clusters, the following two approaches were used. In the first iteration, dynamic time warping (DTW) was applied since this algorithm is able to reveal similarities between temporal sequences that may vary in speed (Aghabozorgi et al. 2015). This appeared to be useful because the sequences had different lengths due to varying tool lifetimes. However, no valid clusters could be detected with this approach because DTW suffers from the principle that simple geometric shapes are similar to all forms. In other words, sequences with strong oscillations but different occurrence times were assigned to more constant sequences instead of sorting all oscillating sequences into a common cluster. Therefore, the intensity of the oscillations was used for clustering in a second iteration. Thus, the sequences were first transferred into stationary series without a trend component by subtracting the median across all sequences. Subsequently, the median absolute deviation (MAD) served as a suitable indicator for measuring the intensity of the oscillations. Sequences with a lower MAD were assigned to cluster 1, and all other sequences were grouped into cluster 2. Therefore, after testing different values, the final MAD threshold was set to 0.2, where the overall variance within both clusters was minimized while keeping the number of items in each cluster in balance to avoid dominant clusters. The resulting



**Fig. 7** Cluster analysis results of the condition dimension

clusters are shown in Fig. 7, where characteristic progressions are recognizable for both groups of sequences.

#### 4.3.2 Assignment to the 4-Field Matrix

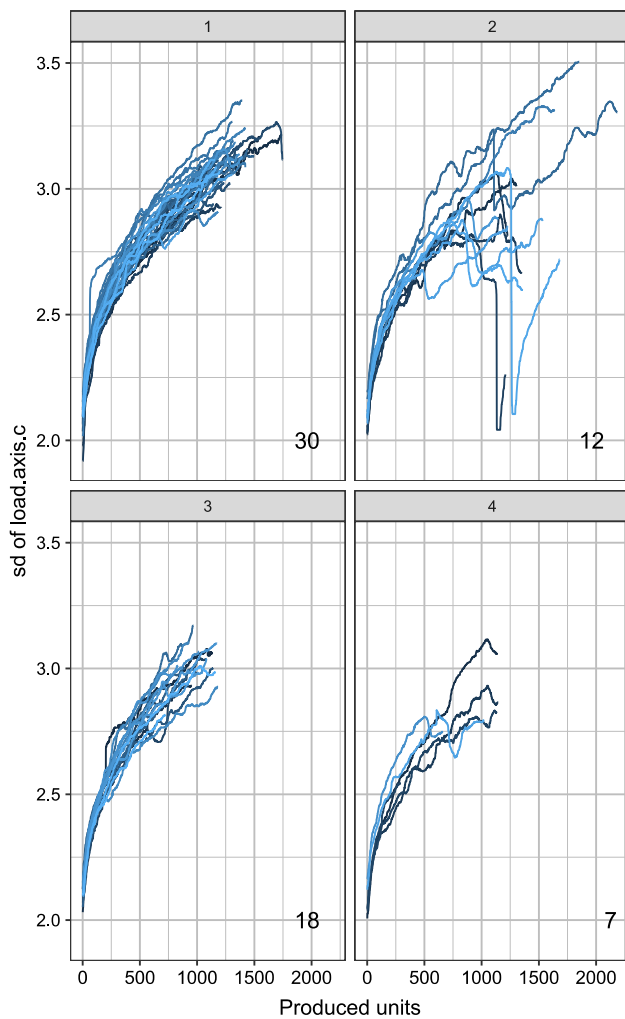
In the next step, the cluster results of both dimensions were used to relate them orthogonally to each other and assign the resulting four subsets to the respective quadrants of the 4-field matrix. Thus, by applying the clusters from the time dimension, tool replacements at a late time were assigned to types 1 and 2, while earlier replacements were assigned to types 3 and 4. Likewise, the two clusters from the condition dimension were used to assign sequences with lower oscillations to types 1 and 3 in the sense of undamaged tools, while sequences with stronger oscillations were assigned to types 2 and 4 in the sense of damaged tools. Figure 8 displays the distribution of all observations among each field, with 30 tool replacements of type 1, 12 of type 2, 18 of type 3, and 7 of type 4.

After assigning the 67 tool replacements, the sequences of all observations were further examined with respect to their temporal progression using the broad variety of extracted features. Therefore, it was possible to identify systematic peculiarities from different perspectives.

Specifically, in the sequences of the feature “standard deviation” (sd), noticeable turning points were observed within the fields of types 2 and 4 (cf. Fig. 8, right side). In the case of type 4, the turning points occurred shortly before the tool replacements, implying that these replacements were correctly carried out at an early stage due to remarkable signs of impending damage. In the case of type 2, however, the turning points were observed well before the end of the tool life, indicating that an intervention was performed too late despite the occurrence of damage signs. Considering these results, the “type 1” observations were particularly interesting for the next modeling step because they reflected “right decisions” of the machine operators, which led to longer lifetimes for undamaged milling tools.

#### 4.3.3 Supervised Learning: Development of a Prognostic Model

In the final stage of the modeling phase, a prognostic model had to be developed. For this purpose, all observations of the 30 “type 1” sequences, each consisting of about 1250–1750 points, were selected and used in the form of training and test data to learn a model and assess its predictive accuracy (Shmueli and Koppius 2011). Based on



**Fig. 8** Assignment of the empirical observations to the 4-field matrix

these observations, appropriate training vectors with a suitable target variable had to be built. Since a single-step prediction is usually not sufficient to apply the model for operational use and to initiate maintenance measures at an early stage (Khawaja et al. 2005), a multi-level target variable with multiple forecasting horizons was created. In particular, the target variable comprised RUL values (expressed in produced quantities) for three forecasting horizons, anticipating the remaining lifetime in 35, 175 and 350 milling cycles. The initial size of 35 was chosen because the first diameter corrections were carried out after 34.55 units on average in the past (cf. Sect. 4.1). The other time horizons were motivated by the necessary preparation time for possible maintenance actions according to the domain experts. For the construction of the training vectors, random sections of the 30 sequences were extracted to consider different prediction times for the operational use of the model. Each section included input data from 150 previously conducted milling cycles as well as the

corresponding values of the target variable. Furthermore, the training vectors were divided into training and test data in a ratio of 80:20 to avoid overfitting and to ensure comparability of the models (Cartella et al. 2015). Since two different sets of input data were available due to the different feature extraction approaches (TDF and TFDF), two separate models were trained. For the model training, RNNs were applied. Such neural networks use internal memories and are capable of mapping complex, non-linear relationships of multivariate time series, whereby it is also possible to learn hidden states within the data structures (Williams 1995). This is relevant since a large number of hidden conditions can be assumed in the course of continuous wear and tear due to the milling procedure. Another advantage of RNNs is that the internal memory of such networks is able to retain the time-related dependencies of previously performed process executions (Heimes 2008). Specifically, the networks were trained in 350 epochs by using a learning rate of 0.02 and a network architecture with 20 hidden layers. The optimization algorithm was based on stochastic gradient descent with a batch size of 1, no pre-training, no added bias and a learning weight decay set at 1, which results in equal weighting over the epoch optimization process.

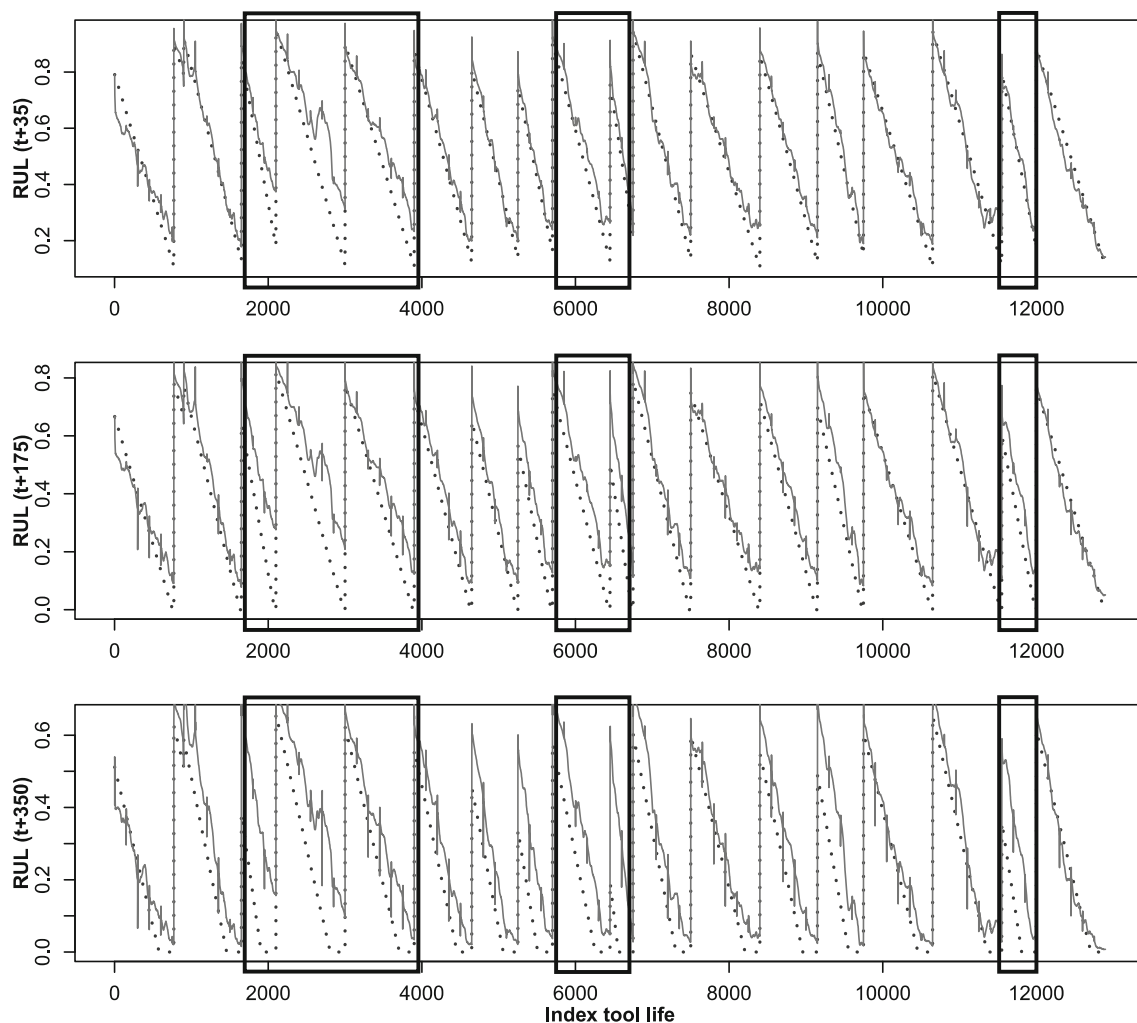
#### 4.4 Evaluation

For the evaluation and comparison of both prognostic models, two performance metrics were applied for each individual forecasting horizon: the mean absolute error (MAE) and the root mean squared error (RMSE) (Pan et al. 2014). Both metrics are generic and commonly used measures for numerical outcomes to assess the predictive performance of an empirical model based on the out-of-sample data (Shmueli and Koppius 2011). The results are summarized in Table 3, where it can be seen that the TDF model performs best across both metrics since MAE and RMSE are smaller for all forecasting horizons.

Considering the interpretation for practical assessments, the MAE usually has the advantage that it can be intuitively interpreted since it uses the same scale as the data being measured. It must be noted, however, that the target variable has been scaled to an interval between 0 and 1,

**Table 3** Assessment of the model performance

Model	MAE			RMSE		
	t + 35	t + 175	t + 350	t + 35	t + 175	t + 350
TDF	0.0468	0.0457	0.0441	0.0635	0.0627	0.0608
TFDF	0.0539	0.0572	0.0566	0.0720	0.0739	0.0712



**Fig. 9** Comparison of the actual tool life (dotted line) with the RUL estimated by the trained prognostic model (continuous line)

which results in low levels as shown in Table 3. Therefore, the MAE needs to be reconverted to the original scale of the RUL values by multiplying the MAE by the maximum tool lifetime of 1747 produced units for “type 1” observations. Thus, the best performing model TDF underestimates or overestimates the tool lifetime by an average of 82, 80 and 77 units for the prediction horizons  $t + 35$ ,  $t + 175$  and  $t + 350$ . Considering an average lifetime of 1315.3 units (cf. Sect. 4.1), the overall prediction error can be regarded as relatively small.

#### 4.5 Deployment

In addition to evaluating the model’s predictive performance, it was further examined which advantage the model would provide if it was deployed for operational use within the production process. This could be achieved by using the

prognostic model to predict the RUL for “type 3” observations in which tool replacements were carried out too early. By comparing the actual tool lifetime with predicted RUL values estimated by the prognostic model, it was possible to quantify the unused service life (measured in produced units) of early changed milling tools. As illustrated in Fig. 9, considerably higher predicted tool lifetimes could be observed for several milling tools (marked by frames). In approximately one-third of all “type 3” observations, a decision about the parameter “tool replacement yes/no” based on the derived prognostic model would have led to an extension of the tool lifetimes. Thus, in terms of the available data, it would have been possible to produce 6340 additional ball hubs within the period under consideration, which corresponds to savings of approximately 4–5 milling tools for a total number of 67 tools. That would be 6–7% of the cost savings expected

from the usage of the prognostic model, and this assessment does not even include the potential savings from preventing “type 2” replacements, which were carried out too late. However, such an additional consideration would also require an adequate diagnostic model, which was explicitly excluded from the scope of this research, as well as further information about costs caused by produced rejects due to damaged milling tools, which could not be provided by the case study partner for reasons of confidentiality.

## 5 Conclusion and Outlook

The contribution of this paper is the development of a solution approach to overcome the situation of missing labels in the context of CBM-based prognostics and ML. Therefore, it offers a novel solution for a known and relevant problem within an established application domain and can be positioned in the improvement quadrant within the knowledge contribution framework proposed by Gregor and Hevner (2013). To draw from a solid understanding of the problem context, prior research was considered by providing a systematization to characterize different label situations while referring to the existing body of knowledge. The systematization can further be used to structure the field and position future work.

Subsequently, for the solution development, a data science study on a real-world scenario of a German automotive manufacturer was carried out facing the challenge of improving an imperfect maintenance situation that was so far based on subjective decisions due to missing replacement criteria. Based on the study results, it was demonstrated how techniques from the field of ML could be used to retrieve information that was only latently available in vast amounts of maintenance-related data. In particular, hidden threshold values were revealed, which usually require profound knowledge about the internal physical processes of a system. Moreover, it was then possible to extract the information whether past maintenance actions were carried out in a risk-averse or risk-affine manner, even though those maintenance actions were not explicitly audited and assessed retrospectively. As a result, a prognostic decision support model was developed that is capable of replacing decisions that have previously been made on the basis of individual risk preferences of the human machine operators. As discussed before, the proposed approach addresses the worst possible problem constellation in a non-synthetic, real-world application where no discernible clues towards a wear-induced replacement were available, and the replacement was solely based on rules from empirical knowledge with no comprehensible foundation. The benefits are potential savings that result from

both the reduction in impaired product quality by preventing tool replacements at a late time and the reduction in tool costs by preventing early replacements.

The proposed approach is sufficiently generic to be applied in other cases where machine tools are subject to continuous wear and tear, such as those involved in cutting, grinding, drilling, polishing or similar operations, since only data types were used that are expected to be recorded by default in industry. This includes (i) event data such as produced quantities, cycle times and tool lifetimes, (ii) control data and machine configurations in terms of applied parameter corrections, and (iii) CM data in terms of measurable variables that reflect the observable machine behavior at a certain point in time. The extraction of useful knowledge despite a poor information situation (i.e., missing CM thresholds, truncated data, missing connection to product quality), which is commonly encountered in industrial practice, illustrates the potential of the developed solution approach. Manufacturing companies not only save expensive investments in additional sensor technology and/or inspection systems but also avoid the redesign of their existing production processes, which would be necessary to make the required latent information explicitly measurable.

Nevertheless, the present contributions also have some limitations, especially with regard to evaluation. In the literature, prognostic models are often developed under experimental laboratory conditions with synthetically generated datasets, where an assessment of the model quality can readily be carried out (Dragomir et al. 2009). In the studied real-world setting of missing labels, however, where no ground truth is accessible, an evaluation approach to ensure reliability and validity can only be fully achieved by involving coherence of expert knowledge on the extracted label thresholds and by testing the feasibility in real process executions. Although the results were discussed with responsible machine operators in each individual development step, the overall approach has not yet been applied under proper conditions. Therefore, it is crucial to carry out a comprehensive evaluation design in a future research project, where the prognostic model is implemented in operational use to see whether it indeed leads to longer tool lifetimes. Current barriers in this context include, for example, multi-layered approval processes of the car manufacturer due to organizational policies that hinder such an application in real operations. The risks of the approach involve the accuracy of the prognostic model and also the overall scope of the decision support system, as it must provide solid statements in all possible constellations and exceptions during runtime that do not endanger the stability of the process. For this purpose, a prototype assistance system is planned for development in which the machine operators should continue to make decisions primarily based on their subjective experiences

but can now indicate whether their decision could be influenced positively by the RUL prediction in terms of longer tool life and, if not, what the reason there would be for a deviation. A similar approach is sought with the extracted thresholds for diagnostic purposes, where the operators can use the model results to assess whether the visually perceived signs of wear and tear correspond to the thresholds identified by the model and vice versa.

However, before introducing a prototype system in operational processes, several more investigations are planned. First and foremost, this includes the use of a larger sample size for model development. Although a representative and rich dataset with more than 88,000 produced units was provided, only 67 tool replacements must be considered relatively small. Therefore, the aim is to collect a broader dataset to evaluate the current results and extend them with further insights. Moreover, the focus of this contribution was to demonstrate the feasibility of the overall solution approach within an industrial setting and thus present a sequence of analytical method combinations. In further research, it is worthwhile to consider each step individually in more detail and apply different alternative approaches for more comprehensive benchmarking purposes. As such, the current approaches can serve as baseline models to examine which variations in terms of modified algorithms and parameter fine-tuning can improve the quality of the results.

## References

- Aghabozorgi S, Shirkhorshidi AS, Wah TY (2015) Time-series clustering – a decade review. *Inform Syst* 53:16–38
- Ahmad R, Kamaruddin S (2012) An overview of time-based and condition-based maintenance in industrial application. *Comput Ind Eng* 63(1):135–149
- Baruah P, Chinnam RB, Filev D (2006) An autonomous diagnostics and prognostics framework for condition-based maintenance. In: *International joint conference on neural networks*, pp 3428–3435
- Bousdekis A, Magoutas B, Apostolou D, Mentzas G (2015) Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance. *J Intell Manuf* 115:1225–1250
- Breiman L (2001) Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Stat Sci* 16(3):199–231
- Brock G, Pihur V, Datta S, Datta S (2008) *clValid: an R package for cluster validation*. *J Stat Softw* 25(4):1–22
- Bumblauskas D, Gemmill D, Igou A, Anzengruber J (2017) Smart maintenance decision support systems (SMDSS) based on corporate big data analytics. *Expert Syst Appl* 90:303–317
- Cartella F, Lemeire J, Dimiccoli L, Sahli H (2015) Hidden semi-markov models for predictive maintenance. *Math Probl Eng* 2015:278120
- Cheng GQ, Zhou BH, Li L (2018) Integrated production, quality control and condition-based maintenance for imperfect production systems. *Reliab Eng Syst Safe* 175:251–264
- Choudhary AK, Harding JA, Tiwari MK (2009) Data mining in manufacturing: a review based on the kind of knowledge. *J Intell Manuf* 20(5):501–521
- Cline B, Niculescu RS, Huffman D, Deckel B (2017) Predictive maintenance applications for machine learning. In: *Reliability and maintainability symposium (RAMS), 2017 Annual*. IEEE, pp 1–7
- Dickey DA, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit root. *J Am Stat Assoc* 74(366):427–431
- Dragomir OE, Gouriveau R, Dragomir F, Minca E, Zerhouni N (2009) Review of prognostic problem in condition-based maintenance. In: *European control conference. IEEE control systems society*. Budapest, pp 1587–1592
- Elattar HM, Elminir HK, Riad AM (2016) Prognostics: a literature review. *Complex Intell Syst* 2(2):125–154
- Everitt BS, Landau S, Leese M, Stahl D (2011) *Cluster analysis*. Wiley Series in Probability and Statistics, New York
- Galar D, Kumar U, Lee J, Zhao W (2012) Remaining useful life estimation using time trajectory tracking and support vector machines. *J Phys: Conf Ser* 364:1–10
- Gouriveau R, Ramasso E, Zerhouni N (2013) Strategies to face imbalanced and unlabelled data in PHM applications. *Chem Eng Trans* 33:115–120
- Goyal D, Pabla BS (2015) Condition based maintenance of machine tools – a review. *CIRP J Manuf Sci Technol* 10:24–35
- Gregor S, Hevner AR (2013) Positioning and presenting design science research for maximum impact. *MIS Q* 37(2):337–355
- Hansen BE (2000) Sample splitting and threshold estimation. *Econometrica* 68(3):575–603
- Heimes FO (2008) Recurrent neural networks for remaining useful life estimation. In: *International conference on prognostics and health management*, pp 1–6
- Huang R, Xi L, Li X, Liu CR, Qiu H, Lee J (2007) Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. *Mech Syst Signal Process* 21(1):193–207
- Jämsä-Jounela SL, Vermasvuori M, Endén P, Haavisto S (2003) A process monitoring system based on the Kohonen self-organizing maps. *Control Eng Pract* 11(1):83–92
- Jardine AK, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech Syst Signal Process* 20(7):1483–1510
- Khawaja T, Vachtsevanos G, Wu B (2005) Reasoning about uncertainty in prognosis: a confidence prediction neural network approach. In: *Annual meeting of the north american fuzzy information processing society*, pp 7–12
- Kurgan LA, Musilek P (2006) A survey of knowledge discovery and data mining process models. *Know Eng Rev* 21(1):1–24
- Leturiondo U, Salgado O, Ciani L, Galar D, Catelani M (2017) Architecture for hybrid modelling and its application to diagnosis and prognosis with missing data. *Measurement* 108:152–162
- Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH (2011) *Big data: the next frontier for innovation, competition, and productivity*. Technical Report McKinsey Global Institute
- Meeker WQ, Hong Y (2014) Reliability meets big data: opportunities and challenges. *Qual Eng* 26(1):102–116
- Muchiri P, Pintelon L, Gelders L, Martin H (2011) Development of maintenance function performance measurement framework and indicators. *Int J Prod Econ* 131(1):295–302
- Pan Y, Er MJ, Li X, Yu H, Gouriveau R (2014) Machine health condition prediction via online dynamic fuzzy neural networks. *Eng Appl Artif Intel* 35:105–113



- Peng Y, Dong M, Zuo MJ (2010) Current status of machine prognostics in condition-based maintenance: a review. *Int J Adv Manuf Tech* 50(1):297–313
- Ramasso E, Saxena A (2014) Performance benchmarking and analysis of prognostic methods for CMAPSS datasets. *Int J Progn Health Manag* 5(2):1–15
- Rinaldo A, Wasserman L, G'Sell M, Lei J (2016) Bootstrapping and sample splitting for high-dimensional, assumption-free inference. *ArXiv Preprint: [arxiv:1611.05401](https://arxiv.org/abs/1611.05401)*
- Saxena A, Goebel K (2008) C-MAPSS data set. NASA Ames Prognostics Data Repository
- Shmueli G, Koppius OR (2011) Predictive analytics in information systems research. *MIS Q* 35(3):553–572
- Si XS, Wang W, Hu CH, Zhou DH (2011) Remaining useful life estimation – a review on the statistical data driven approaches. *Eur J Oper Res* 213(1):1–14
- Sneath PHA, Sokal RR (1973) *Numerical taxonomy. The Principles and Practice of Numerical Classification*. Freeman, San Francisco
- Susto GA, Schirru A, Pampuri S, McLoone S, Beghi A (2015) Machine learning for predictive maintenance: a multiple classifier approach. *IEEE Trans Ind Inform* 11(3):812–820
- Tian Z, Wong L, Safaei N (2010) A neural network approach for remaining useful life prediction utilizing both failure and suspension histories. *Mech Syst Signal Process* 24(5):1542–1555
- Ullah I, Yang F, Khan R, Liu L, Yang H, Gao B, Sun K (2017) Predictive maintenance of power substation equipment by infrared thermography using a machine-learning approach. *Energy* 10(12):1–13
- Van der Laan M, Pollard K, Bryan J (2003) A new partitioning around medoids algorithm. *J Stat Comput Simul* 73(8):575–584
- Vapnik VN (1999) An overview of statistical learning theory. *IEEE Trans Neural Netw* 10(5):988–999
- Veldman J, Wortmann H, Klingenberg W (2011) Typology of condition based maintenance. *J Qual Maint Eng* 17(2):183–202
- Vogl GW, Weiss BA, Helu M (2016) A review of diagnostic and prognostic capabilities and best practices for manufacturing. *J Intell Manuf* pp 1–17
- Wang W, Christer AH (2000) Towards a general condition based maintenance model for a stochastic dynamic system. *J Oper Res Soc* 51(2):145–155
- Williams RJ (1995) Adaptive state representation and estimation using recurrent connectionist networks. In: Miller WT, Sutton RS, Werbos PJ (eds) *Neural networks for control*. MIT Press, Cambridge, pp 97–114
- Yuan J, Liu X (2013) Semi-supervised learning and condition fusion for fault diagnosis. *Mech Syst Signal Process* 38(2):615–627
- Zhao X, Li M, Xu J, Song G (2011) An effective procedure exploiting unlabeled data to build monitoring system. *Expert Syst Appl* 38(8):10199–10204
- Zschech P (2018) A taxonomy of recurring data analysis problems in maintenance analytics. In: *Proceedings of the 26th European conference on information systems*