Creation of Publicly Available Data Sets for Prognostics and Diagnostics Addressing Data Scenarios Relevant to Industrial Applications

Simon Hagmeyer[1], Fabian Mauthe[2] and Peter Zeiler[3]

[1,2,3]*Esslingen University of Applied Sciences, Goeppingen, 73037, Germany*

simon.hagmeyer@hs-esslingen.de
fabian.mauthe@hs-esslingen.de
peter.zeiler@hs-esslingen.de

ABSTRACT

For a successful realization of prognostics and health management (PHM), the availability of sufficient run-to-failure data sets is a crucial factor. The sheer number of given data points holds less importance than the full coverage of the potential state space. However, full coverage is a major challenge in most industrial applications. Among other things, high investment and operating costs as well as the long service life of many technical systems make it difficult to acquire complete run-to-failure data sets. Consequently, in industrial applications data sets with specific deficiencies are frequently encountered. The development of appropriate methods to address such data scenarios is a fundamental research issue. Therefore, the purpose of this paper is to provide facilitation for this research. Accordingly, the paper starts by specifying the value and availability of data in PHM. Subsequently, criteria for characterizing data sets are defined independent of the actual PHM application. The criteria are used to identify typical data scenarios with specific deficiencies that possess significant relevance for industrial applications. Thereafter, the most comprehensive overview of data sets suitable for PHM and currently publicly accessible is provided. Thereby, not all previously identified data scenarios with their specific deficiencies are addressed by at least one data set. A program is established for the aforementioned facilitation of further research. One objective of the program is to create data sets reflecting these data scenarios using a test bench. First, possible applications and their degradation processes to be studied on the test bench are briefly characterized. Thereby, the final decision to select filtration as a test bench application is argued. Subsequently, the test bench created is introduced, including a description of the functional concept, pneumatic layout and components involved, as well as the filter media and test dusts employed. Typical run-to-failure trajectories are illustrated. Thereafter, the data set published under the name *Preventive to Predictive Maintenance* is presented. Additionally, a schedule for future releases of data sets on further industry-relevant data scenarios is sketched.

1. INTRODUCTION

Key tasks in PHM are the detection of faults, the diagnosis of which component in a system causes the current fault condition, the health assessment, and the prognosis of the remaining useful life (RUL). The health information acquired accordingly serves as a foundation for the health management. Approaches to address the prognosis and diagnosis are often divided into data-driven methods, model-based methods, and hybrid methods (Atamuradov, Medjaher, Dersin, Lamoureux, & Zerhouni, 2017).

Each approach requires or at least benefits from the availability of sufficient data. Data-driven methods originate from the domains of statistics and machine learning (ML). The fundamental characteristic of these methods is that the modeling of system behavior is done entirely mathematically or rather statistically. Structural understanding of the behavior being modeled is not acquired. In order to use data-driven methods in a purposeful way, all relevant areas of the state space should be reasonably covered by available data (Javed, Gouriveau, & Zerhouni, 2017). This involves at least that several run-to-failure data are available for each fault mode of the system (Uckun, Goebel, & Lucas, 2008). Depending on the application, different levels of production quality, operating conditions, and other additional aspects also hold importance.

Much more than data-driven methods, (physical) model-based methods require a thorough understanding of the mechanism involved in the degradation process. Even if theoretically no data is required for the creation of such a

physical model, it can be essential for a model-based approach when it comes to actually applying PHM. First, insufficiently precise known parameter values of a physical model can be reduced to a narrower range by means of data (An, Choi, & Kim, 2013). Second, the basis of scientific work is to assess whether a theory or model can be disproven by the data obtained from experiments (Popper, 1963). Testing based on concrete process data assists in identifying conditions that have hitherto been insufficiently modeled. Hybrid methods – which stem from the combination of data-driven and physical models – accordingly also require the availability of data. Thus, for each of the approaches, plentiful data is required.

However, one major drawback of PHM is that the objective of extensive data usually implies immense effort in terms of cost and time. Data of the assessed system in good condition is often plentiful available, but simply having several seamless run-to-failure data sets for each fault mode usually corresponds to a practically impossible amount of effort (Hemmer, Klausen, van Khang, Robbersmyr, & Waag, 2019; Pillai, Kaushik, Bhavikatti, Roy, & Kumar, 2016). Therefore, when PHM is applied in industrial domains, data scenarios are encountered that do not comply with ideal conditions. Instead, the available data often contains deficiencies that cannot be eliminated. For example, this could be missing parts within the life cycle recording, caused by a failure of the measurement system or the data handling. There could also be records on the beginning of the life cycle missing due to retrofitting of sensors. Another typical case is records that do not extend to the point of failure. There are a variety of reasons for this, such as premature replacement due to preventive maintenance, a very long life of a highly reliable system or not being able to operate the system up to its failure point due to the resultant harm (Chao, Kulkarni, Goebel, & Fink, 2021). Thus, examinations of such deficiencies and data scenarios are performed within this paper.

For PHM, there are various methodologies, functional architectures, and frameworks that describe how to apply PHM. This often involves trying to distinguish the most promising approach for an application, whether it is data-driven, model-based, hybrid, or classical reliability engineering. There are also many architectural schemes for integrating PHM into data processing of a manufacturing plant, vehicle, or building, for instance. Representative examples and reviews can be found in Aizpurua and Catterson (2015), Elattar, Elminir, and Riad (2016), Aizpurua and Catterson (2016), and Atamuradov, Medjaher, Dersin, Lamoureux, and Zerhouni (2017). Thus far, these frameworks and other types of guidance do not account for the deficiencies and their data scenarios mentioned before.

In PHM research, there are already several papers that focus on developing suitable methods specifically for a particular data scenario; for example, Xu, Baraldi, Al-Dahidi, and Zio (2019), Cannarile, Baraldi, and Zio (2019), and Wiese, Pedersen, Nadimi, and Herp (2020). However, a general categorization of data scenarios detached from the particular application is missing. Furthermore, such data scenarios are only partially and rather coincidentally represented by publicly available data sets. Consequently, there is also a lack of individual data sets that specifically reflect the data scenarios.

The main purpose of this paper is to contribute to the research on addressing PHM-specific data scenarios in industrial applications. Therefore, criteria for the assessment of data sets are defined in section 2. Based on these criteria, typical data scenarios with strong relevance to industrial applications are identified in section 3. Subsequently, the most comprehensive overview of publicly available data sets that are suitable for PHM research thus far is given in section 4. Here, some data sets are identified to represent specific data scenarios. Nonetheless, given that not all data scenarios with strong relevance are represented by these publicly available data sets, a program is established to provide data sets for all relevant scenarios. The data is generated by means of a filtration test bench, which is introduced in section 5, along with the first data set already generated. Additionally, future releases of further data sets on other data scenarios are sketched. The paper ends in section 6 with conclusions on the content and an outlook. Overall, the paper aims to provide:

- Criteria for the qualitative assessment of PHM data sets

- The most comprehensive overview on publicly available PHM data sets thus far

- Publicly available data sets relevant for industrial applications

- A foundation for further research on developing or adapting PHM methods to specific data scenarios

2. DEFINITION OF CRITERIA FOR AN ASSESSMENT OF DATA SETS

As stated in the previous section, having data available is an integral part of implementing PHM. Thereby, a distinction can be made between four essential tasks relating to the condition of a technical system, which serve as the basis for health management. The following categorization of the four tasks is based on Jia, Huang, Feng, Cai, and Lee (2018):

- **Fault detection/anomaly detection:** Detect a fault state/anomaly of a technical system without knowing the root cause. This results in a binary classification problem with the states *fault* or *no fault*.

- **Diagnosis:** Assign one or more causes to a detected fault state.

- **Health assessment:** Assess the state of health or the current risk of failure of a system based on its current condition.

- **Prognosis:** Predict the future state of health or RUL.

1. History provided by the data	2. Data acquisition	3. Data on degradation

1.1 Data history range
  - Run-to-failure/threshold
  - End of life not recorded
  - Start during life cycle
  - No history, only a current status
1.2 System with maintenance
1.3 Lifetime criterion specified
	2.1 Recording
	  - Continuous
	  - Periodic, equidistant
	  - Random
	  - Condition-dependent
	2.2 Measurement properties
	  - Sampling rate
	  - Time information
	    • Total operating time
	    • Number of the current record without absolute time indication
		3.1 Data on
		  - Parameters affecting the degradation
		  - Health indicators
		3.2 Information on degradation
		3.3 Prior knowledge on:
		  - Degradation features
		  - Degradation process
		3.4 Reproducibility of the data

4. System information	**5. Quality & quantity of data**	**6. Data source**

4.1 Operating condition
  - Constant
  - Deterministic
  - Stochastic
4.2 Failure modes of the system
  - Number from failure modes
  - Number of components
4.3 Motivation for PHM adoption
  - Safety
  - Availability
  - Profitability
4.4 Redundancy exists, what kind
	5.1 Coverage of state space
	  - Relevant state space
	    • Complete/partial
	  - Extended state space
	5.2 Data volume
	  - Total number of data
	  - Number of data per state
	5.3 Data quality
	  - Systematic measurement error/bias
	  - Random measurement errors/noise
		6.1 Source
		  - Simulation
		  - Laboratory experiments
		    • Accelerated/not accelerated
		  - Real application
		    • Selected test specimen
		    • Random test specimen

Table 1. Collection of criteria for the qualitative assessment of data sets at PHM.

Which of the four tasks is to be performed, as well as what characteristics the data itself possesses are fundamental for utilizing a data set in PHM. In the following, industry-typical data scenarios are outlined based on their application independent data characteristics. This is primarily illustrated using the prognosis as an example task, but could also be applied to the other tasks, such as diagnosis. It is done by defining a set of criteria with six main criteria and their respective sub-criteria, as shown in Table 1. There are already various collections of criteria for the evaluation of data sets, which mostly come from the general data analysis domain. Examples of such collections are Kahn, Strong, and Wang (2002), Wang, Storey, and Firth (1995), and Merino, Caballero, Rivas, Serrano, and Mario (2016). Those focus on general data quality aspects, such as data format consistency, data security, and ease of manipulation. However, the authors are not aware of any having such a PHM focus, as is required for the present objective. Hence, criteria are designed by the authors. These are only intended to provide a generally valid description of typical data scenarios for PHM in industry, taking into account the main application independent characteristics. It should be noted that the individual sub-criteria are not prioritized or further detailed, which is beyond the scope of this paper. In the following, the six main criteria with their sub-criteria are briefly explained:

- The **history provided by the data** contains the sub-criteria of data history range, system with maintenance,

and lifetime criterion. The sub-criterion of data history range describes the part of the run-to-failure trajectory that is covered by the data. The second sub-criterion considers a system with maintenance and subsequent restart. The third sub-criterion involves the lifetime criterion of the system.

- **Data acquisition** analyses the time characteristics of the measurement series. The first sub-criterion of recording distinguishes based on the time intervals between recorded data during the service life. The second sub-criterion covers the measurement properties of the data recording. Part of this is the sampling rate as well as the absolute time information on the data points.

- The third main criterion of **data on degradation** looks in detail at the state of information on the degradation process of the system. The first sub-criterion assesses whether the available data includes parameters affecting the degradation, such as mileage, operating hours and whether measurements also provide information about the health; for instance, particle concentration in oil. Next, it is evaluated whether the measurements directly reflect the health respective degradation or, as is usually the case, indirectly via a required data processing (Si, Wang, Hu, & Zhou, 2011). The third sub-criterion takes into account the prior knowledge about features that enable designing health indicators and prior knowledge

on the overall degradation process. The fourth sub-criterion judges the reproducibility of data or present effects.

- In the fourth main criterion, **information** on the overall system are considered. The first sub-criterion of operating condition characterizes the sequence of operating conditions during the service life of a system. The second sub-criterion focuses on the number of failure modes and the number of components that can trigger a failure of the system. The third sub-criterion considers the cause of the intended adoption of PHM. The fourth sub-criterion examines possible redundancies of components.

- The fifth main criterion includes the **quality and quantity of the data** at hand. The first sub-criterion assesses the availability of data within the state space. The second sub-criterion covers the volume of data available. The total number of data and if can be assigned the number of data points per state are taken into account.

- The sixth main criterion looks at the **source** from which the data is derived. The data can be generated by simulation, laboratory, test bench experiments or real application.

In order for a data set to offer the greatest possible added value, sufficient knowledge on the existing data scenario is necessary. Developing a dependable PHM application without knowledge on the system considered and its available data is rather impossible in the authors' point of view. The level of knowledge about a planned PHM application can be assessed, for example, in using this set of criteria.

3. IDENTIFICATION OF DATA SCENARIOS RELEVANT FOR INDUSTRIAL APPLICATIONS

Specific data scenarios that typically occur with the same general data set characteristics in industrial application hold strong importance for the widespread implementation of PHM (Weiss & Brundage, 2021). These data scenarios – which involve deficiencies that require their own methodological adaptation – are therefore addressed. It is already evident from the combinatorics of the criteria that there are a large number of data scenarios. Hence, the focus is on relevant data scenarios related to individual criteria. The identification and categorization of data scenarios is based on the criteria collection. The selection of the data scenarios that are considered as relevant and listed below is based on typical causes of industrial applications, which yield to these scenarios with their data characteristics. Therefore, in the following, each data scenario is illustrated based on a description of its various causes. The assessment of industry-typical causes is based on the experience of the authors.

Data history range (criterion 1.1) examines the degradation trajectories available for developing a prognosis application. It is analyzed how these are distributed over the service life of the technical system. These trajectories can be run-to-failure or run-to-threshold data. In case of run-to-failure, data is available from the start to the failure throughout the entire life of the system. For run-to-threshold, data is available up to a defined threshold, where the system is also considered to be failed.

Degradation trajectories are best suited for developing a prognosis application when the actual failure and RUL information exists. Data up to the point of system failure often originate from life cycle tests carried out specifically for this purpose. However, in actual applications, particularly valuable systems are rarely operated up to the point of actual system failure (Cannarile, Baraldi, & Zio, 2019). There are several reasons for the frequent occurrence of not recording the end of life. These include, as partially mentioned, predetermined replacement due to preventive maintenance. Predetermined maintenance intervals remain widespread in industrial domains, which makes such data frequently present (Chao, Kulkarni, Goebel, & Fink, 2021; Widodo & Yang, 2011). Furthermore, a long service life of a highly reliable system can lead to the same type of data. Associated with this is the case of a short overall product development schedule, where life testing or field deployment may not have sufficiently progressed at the time of PHM implementation for a large number of failures to be present. However, it may also plainly be the case that the system cannot be operated up to its failure point due to the resultant harm (Chao, Kulkarni, Goebel, & Fink, 2021). Hence, despite a large number of life cycles recorded, the development of the prognosis application is made more difficult as the actual RUL information is not available.

The next typical data scenario in industry is a data history that starts during the life cycle. For systems with long service lives, such as machine tools, it is often the case that they are retrofitted with sensors and degradation recording is started during the service life. Thus, the progression of the health in an initial part of the service life is not covered by the data and is therefore unknown.

In accordance with the description above, two relevant data scenarios are identified:

a) Data without the end of life being recorded
b) Data with start of the data history during the life cycle

Criterion 2.1 considers the **recording** of the data over time in the context of the overall service life. In continuous recording, the measurements are logged concurrently as the system operates and are available seamless throughout its service life. If such form of recording is not performed or is not possible, gaps in the recording will occur. These can occur either equidistant or randomly resp. stochastically distributed.

An example that causes such gaps is the use of defined test cycles for monitoring. These always follow the same sequence, after a period of use. This procedure is also referred to as active inspection or active monitoring. Such a scenario occurs, for instance, in the case of machining centers that manufacture products with different levels of complexity. Here, the operating time varies depending on the machined product and job. Thus, possible test cycles can only be performed between machining steps or after finishing a job at the earliest (Tobon-Mejia, Medjaher, & Zerhouni, 2012). This results in degradation measurements with inconsistent time spacing. It is similar to a random sample from the data of the continuous recording. The same data characteristic occurs when the objective is to reduce the amount of data. In industrial applications, it is often attempted to keep the amount of unmodified data transmitted and stored as small as reasonably possible (Omri, Al Masry, Mairot, Giampiccolo, & Zerhouni, 2021). This can result in equidistant or randomly distributed data batches, albeit which do not originate from a defined test interval with a specified load.

In addition to the previous time-influenced recordings, condition or event-based recordings are also relevant in practice and likewise lead to gaps in the data set. Here, the acquisition is started when a defined threshold value is exceeded; for example, in terms of load, fault mode or damage gradient. This can also be seen as a recording trigger (Zhu, Nostrand, Spiegel, & Morton, 2014). It should be noted that the measurement can be continuous but the data is only recorded, i.e. stored, when a changed state is detected. By recording in this manner, the amount of data can be reduced by avoiding recording when there is little or no change of state and thus information.

Hence, the following two data scenarios are identified:

c) Data with random recording
d) Data with condition-dependent recording

The **measurement properties (criterion 2.2)** take into account the time information linked to the data. For example, this affects the ability to assign a data point to the overall life. This could be knowledge on the total current operating time or the current cycle count. The length of the service life thus far is usually an important input feature in the prognosis on its own. For example, a distinction can be made between similar appearing run-in effects at the beginning of the service life and degradation effects at the end.

Systems for which no overall time information is available can be seen, for example, in self-sufficient systems that are not continuously powered. Therefore, the corresponding data scenario is identified as:

e) Data batches with total/with partial/without time information

Regarding the criterion of **redundancy (criterion 4.4)**, the aspect of workload sharing is considered, for example. This means similar subsystems that work in parallel and are all active during operation. If one subsystem fails, the remaining subsystem(s) also takes over its workload. As illustrated by this example, redundancies in a system already influence maintenance strategies that do not take into account the current system state, namely corrective maintenance and preventive maintenance (Dong, Liu, & Du, 2019; Mendes, Coit, & Ribeiro, 2014). However, for maintenance strategies such as predictive maintenance, redundancy is particularly important. The load distribution among intact subsystems as well as the degradation state of the partner system(s) significantly influence the RUL of each subsystem. Furthermore, there can be sudden changes in the loads as well as the RUL if one subsystem fails. This research topic is highly relevant to the industry and at the same time represents a field of research with few PHM-related studies thus far. The basis for such research could be a data set that reflects this data scenario. The resulting data scenario is:

f) Data on the life of redundant subsystems containing the failure of subsystem(s)

In the context of PHM, the **state space (criterion 5.1)** represents all possible loads, fault modes, wear rates, operating states, etc. of the technical system. It needs to be assessed to what extent the relevant part of the state space – as it can also be encountered in the actual application – is covered by the data. Partial coverage of the state space occurs, for example, when not all fault modes are present in the data used for the development of a prognosis application.

As argued in section 1, it is not possible to generate actual run-to-failure records for most industrial applications of PHM, as the service life is a few years of operation. Possible solutions to the problem of no actual records being accessible are accelerated testing under laboratory conditions or, to some extent, simulations tailored to the degradation process. Even if the basic degradation mechanism is to be maintained, such data is referred to an extended state space in the criteria catalog, as they do not occur in the actual application. In the case of accelerated testing, this leads to load conditions that are significantly increased to enhance degradation. The challenge in developing a prognosis for this kind of scenario is to accomplish the carry-over to the actual application with regular load conditions.

Based on the described data characteristics, there are two data scenarios of practical relevance:

g) Data sets with partial coverage of the relevant state space
h) Presence of training data that is to be mapped to the extended state space

The **data quality (criterion 5.3)** is significantly affected by the sensors, sensor location, and peripherals used. This yields a different bias and different random noise (Atamuradov, Medjaher, Dersin, Lamoureux, & Zerhouni, 2017). If technical systems are in the market with different sensors and their positioning, for example at the customer's request, this results in varying data qualities. This typically occurs when cost-benefit analyses are carried out in the course of product development to identify the most suitable components for different customers from a techno-economic perspective. A diagnostic or prognostic model that has been developed based on a significantly different data quality must take this into account. This applies especially for the management of uncertainties. The corresponding data scenario is therefore:

i) Run-to-failure data with differing measurement bias and noise

4. EVALUATION BASED ON AN OVERVIEW OF PUBLICLY AVAILABLE DATA SETS

Large run-to-failure data sets are hold strong value. The companies and institutions that possess such data sets therefore have clear reasons not to provide free access to their data sets (Saxena, Goebel, Simon, & Eklund, 2008). The few publicly available data sets therefore hold strong value to the research community. They enable the mutual empirical benchmarking of the many different methods at PHM (Ramasso & Saxena, 2014). Furthermore, they offer researchers the opportunity to demonstrate a concrete application of their presented methods. Hence, as a basis for further research on data sets with specific data scenarios, this section provides an overview of the currently publicly available data sets and highlights different aspects of the overview. Subsequently, the data sets from the overview are examined for compliances with the identified data scenarios from section 3.

4.1. Scope of the Data Set Overview

The current data set overview differs from previous overviews such as Jia, Huang, Feng, Cai, and Lee (2018) and Huang, Di, Jin, and Lee (2017), because in addition to typical PHM sources, as they are stated in Lei et al. (2018) and Ochella and Shafiee (2019), many more sources and platforms for possible data sets are taken into account. By typical sources, we refer to the PHM Data Challenges of the PHM Society (PHM Society, 2021) and the Prognostics Center of Excellence of the National Aeronautics and Space Administration (NASA) (Prognostics Center of Excellence - Data Repository, 2021). Moreover, the scope of the overview

is extended to research institutions and comparable organizations that deal with the aforementioned PHM tasks. Another beneficial feature is the recency of this overview. In the light of Industry 4.0 and the associated availability of data, ML approaches are becoming increasingly common in technical systems and production facilities (Dalzochio, et al., 2020; Cachada, et al., 2018). This enhances the interest in data sets derived from technical systems, as these are also needed and used in other research areas. Therefore, artificial intelligence (AI) or ML dedicated data repositories are considered in the overview as additional platforms. By covering research institutions and comparable organizations as well as AI and ML dedicated data repositories, the overview is the most comprehensive so far.

An important constraint of the overview is that data sets that do not consider the degradation state but only the process quality (often the quality of a manufactured product) are not included.

4.2. Overview and its Analysis

Table 3 in the appendix lists in detail the 70 PHM-related data sets contained in the overview. The table also contains the publishing authors or institutions and the URL where the respective data set can be obtained. A summary of the distribution of the 70 data sets among their sources and platforms is given in Figure 1. With a total of 29 data sets, the PHM Society and NASA continue to have a major impact on the provision of data sets. The data sets from these sources are widely used in the PHM-related literature and are most commonly used as benchmarks to validate newly developed PHM methods (Lei, et al., 2018; Kim, An, & Choi, 2016). This is advantageous for developing PHM methods as it provides a reference for the performance of a method. For the industrial application of PHM, this can be a disadvantage as the same use cases and thus also the same data scenarios are considered.

Additional sources and platforms include the online community Kaggle (Kaggle: Your Machine Learning and Data Science Community, 2021) and the repository of the University of California at Irvine (UCI) (UCI Machine Learning Repository, 2021). Especially for the online community Kaggle, it can be observed that data sets are added here with major frequency compared to the previous sources, thus providing the most recent data sets. The sources of the remaining 22 data sets, which are referred to as *Others* (PHM 16 /ML 6) in Figure 1, stem from other online communities, the various research institutions and comparable organizations. The division of their focus on PHM or general ML can be made rather less precisely. Again, some of the data sets are very recent in terms of their publication date.
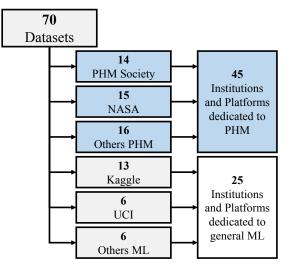
Figure 1. Depiction of the number of data sets provided by each platform.

Data sets from PHM Challenges and the NASA are generally well documented in accordance with scientific standards, which also applies to various other data sets from PHM-related institutions, as can be seen in Saxena, Goebel, Simon, and Eklund (2008), for instance. Especially regarding other sources and platforms, the quality of the information on the data should not be neglected when using them. Nevertheless, many data sets with a logical structure and valuable description are also included here. Data sets that have no actual description at all are not part of this overview, as their usability and therefore value is clearly limited.

4.3. Status of the Representation of Data Scenarios

In the following, the data scenarios from the overview are examined for compliances with the identified data scenarios from section 3. The descriptions of the data sets were considered to estimate their respective data scenarios. First, the data scenarios for which there is a corresponding data set are specified.

Data with start of the data history during the life cycle (b): This data scenario is already considered in the *NASA - Turbofan engine degradation simulation data set*, which is often used as a benchmark in PHM research. The data set involves initial wear at the start of the simulated turbines' service life trajectories (Saxena, Goebel, Simon, & Eklund, 2008). This can be considered as a start during the lifetime.

Data sets with partial coverage of the relevant state space (g): The data set that shares this scenario is *Data Challenge PHM Soc. 2020 Europe - Filtration System*. The training data only contains run-to-failure data sets with particles of size $45 - 53\,\mu m$ and $63 - 75\,\mu m$. On the other hand, the test data only examines predictions for particles of size $53 - 63\,\mu m$. Hence, data on this particle size are not available for modeling or training. This results in incomplete coverage of the relevant state space. Nonetheless, an extrapolation where

the test data comprised one of the boundary particle sizes would be even more challenging.

Further data scenarios cannot be identified based on the data set descriptions. The following seven data scenarios are therefore designated as not covered thus far:

- Data without the end of life being recorded (a)
- Data with random recording (c)
- Data with condition-dependent recording (d)
- Data batches with total/with partial/without time information (e)
- Data on the life of redundant subsystems containing the failure of subsystem(s) (f)
- Presence of training data which is to be mapped to the extended state space (h)
- Run-to-failure data with differing measurement bias and noise (i)

For data scenario (a), it is noted that there are data sets for which no exact point of failure is defined. However, these data sets do not contain any test data for which the end of life or the RUL value is specified. The objective of the data scenario to make the most use of incomplete degradation trajectories for the effective development of a RUL prediction is thus not addressed.

For data scenario (h), the following still needs to be considered. In the overview, no equivalent data set that reflects these characteristics is found. While there are data sets such as *IEEE PHM Data Challenge 2012 - FEMTO Bearing Data Set* and *NASA - IGBT* that stem from accelerated life tests, there is no test data under regular load conditions. As such, these data sets do not address the carry-over challenge in this scenario, but rather contain a PHM task under special load conditions.

The 70 data sets listed in Table 3 in the appendix often originate from specific tests or simulations. Consequently, they generally have a significantly higher degree of completeness than is usually the case in real applications and do not sufficiently represent the previously mentioned data scenarios. Among other things, one main task in PHM research is therefore still to generate realistic data sets that require purposeful treatment of deficiencies and data characteristics.

5. ESTABLISHED PROGRAM TO GENERATE DATA SETS FOR RELEVANT DATA SCENARIOS

In order to facilitate research regarding the most relevant data scenarios, a program is established. In accordance with the conclusion of the previous section, one main objective of this program is to generate and publicly provide data sets reflecting the data scenarios. For this purpose, an application with its degradation process is selected as data source and the corresponding test bench is introduced.

5.1. Selection Process of the Test Bench Application

The selection of the application with its degradation process, which is studied on the test bench, takes place in two stages. The result of an initial pre-selection of possibilities is four applications, which are described below with the central characteristics of their degradation process. They are all fundamentally suitable to be used as studied application. For the final selection, special focus is placed on research aspects of PHM. Based on these aspects, the final selection is explained subsequently.

1. Filtration: Almost all industries depend on filtration as a process step, such as the food, power generation, chemical industries, etc. The health of a filter has a considerable influence on the filter performance itself, but also on the efficiency of the respective overall industrial process of which it is part (Sparks & Chase, 2015). The degradation and the resulting maintenance process of a filter are primarily caused by its loading. There are additional failure mechanisms of minor relevance such as aging, fatigue or chemical damaging (Sparks & Chase, 2015; Saarela, Hulsund, Taipale, & Hegle, 2014). Here, only the separation of solids and fluids (here: liquids, gases) is examined, which is the filtration type with the highest industrial significance. For instance, the global industrial air filtration market in 2019 is estimated at USD 3.19 billion (Statista Research Department, 2019). Consequently, filtration has also been studied in PHM; for example, in Eker, Camci, and Jennions (2016), Sreenuch, Khan, and Li (2015), Saarela, Hulsund, Taipale, and Hegle (2014), and Skaf, Eker, and Jennions (2017). The main focus of these works is placed on the filtration of solids from liquids, more precisely the fuel filtration of vehicles.

2. Gear failure: As with filtration, almost all industries depend on gears and the assurance of their reliable operation. Hence, degradation of gears – especially in the context of condition diagnosis – has already been thoroughly investigated within PHM (Sharma & Parey, 2016). Common fault modes for gears are scuffing, pitting, spalling, abrasive wear, and cracking (Sharma & Parey, 2016; Cubillo, Perinpanayagam, & Esperon-Miguez, 2016). The underlying causes of such failures can be divided into three categories: operational conditions (overload, fatigue, etc.), lubricant (lubricant failure, contamination, etc.), and assembly (misalignment, tight fit, etc.) (Cubillo, Perinpanayagam, & Esperon-Miguez, 2016). The most common measurement techniques for monitoring gears are measurement of vibration, acoustic emission, temperature, oil condition, and torque fluctuation (Večeř, Kreidl, & Šmíd, 2005; Kundu, Darpe, & Kulkarni, 2020).

3. Mechanical stress to the attachment of electronic components: Nowadays, most technical systems include electronic components, which is why an in-depth study of their degradation process has already taken place in PHM research. A typical failure mechanism is the fatigue of their attachment, such as wire bond, solder leads, and bond pads. This is primarily due to temperature and motion factors. Material stresses can arise from temperature gradients or thermal load cycles when a thermal expansion mismatch exists. Vibrations and shocks are the main source of stress by motion (Gu, Barker, & Pecht, 2007; Cheng, Raghavan, Gu, Mathew, & Pecht, 2018, S. 63ff. 75).

4. Wear of stamping dies: Stamping is a widely-applied type of sheet metal processing method. Stamping is a highly complex process with many influencing parameters such as elastic and plastic deformations, lubrication, the dynamic and static behavior of the matrix (Ge, Du, Zhang, & Xu, 2004). In particular, the increasingly relevant processing of high-strength steels is responsible for high wear rates and the resulting defective production (Shanbhag, Rolfe, Arunachalam, & Pereira, 2018). The main causes of wear are adhesive and abrasive wear. The degradation due to wear occurs often highly stochastically and rapidly progresses, which makes fixed maintenance times unsuitable. There are various approaches for capturing the degradation state, such as measurements of stamping force, strain, acoustic emission, and vibration (Huang & Dzulfikri, 2021). A test bench to investigate wear of the stamping die can be undertaken, as Ge, Du, Zhang, and Xu (2004) and Shanbhag, Pereira, Voss, Ubhayaratne, and Rolfe (2019) show in a reduced scale.

The four applications considered feature degradation processes that are representative of PHM deployments. The final decision on the test bench application is based on aspects that particularly address the provision of data relevant to PHM research. These aspects are physical modeling of the degradation process, variety of operating conditions, testing effort, and research potential on system-level PHM. From the authors' perspective, filtration is consistently ranked among the best in all aspects. Therefore, filtration is selected as the test bench application and its properties with respect to these aspects are described in the following. As mentioned above, filtration studies in PHM have mainly focused on the separation of liquid and solid particles to date. However, here the separation of gas and solid particles is now applied. Therefore, the application with its generated data significantly differs from previous test setups and especially from the data set *Data Challenge PHM Soc. 2020 Europe - Filtration System* in the overview.

Physical modeling of the degradation process: Degradation processes of various technical systems are highly complex. Consequently, physical modeling of them is only possible at all with great effort. A distinct feature of filter loading for PHM research is that it is one of the few failure mechanisms with significant complexity, besides crack propagation, for which several physical models are nevertheless described in the literature; for example, Eker, Camci, and Jennions (2016), Abdolghader, Brochot, Haghighat, and Bahloul (2018), Chikhi, Clavier, Laurent, Fichot, and Quintard (2016), Thomas, Penicot, Contal,

Leclerc, and Vendel (2001), Bergman, et al. (1978), and Novick, et al. (1990).

Variety of operating conditions: The possibility to have a wide range of operating conditions is given for filtration. This involves variation of the filter media, test particles, concentration of particles in the fluid, flow rate, flow angle, etc. It holds particular value that the mentioned condition changes require no or only minor additional setup effort when performing run-to-failure tests.

Testing effort: Run-to-failure test cycles of filters can be performed automatically in less than one hour. The maintenance time between tests is also less than one hour and primarily comprises cleaning contaminated components of the test bench. Consumables of the tests are simply test dust, filter media, and pressurized air.

Research potential on system-level PHM: One area of PHM that still holds strong research potential today is *system-level PHM* (Lei, et al., 2018). It features system-wide fault identification, an examination of the interactions of failure mechanisms, and, as a result of system-wide condition diagnosis and prognosis, a more comprehensive health management. Data sets on this extend beyond the objective of this paper but also hold strong interest to the research community. The proposal to use a filtration application for such studies has already been introduced by Niculita, Irving, and Jennions (2012). Such a fluidic system is highly modular and can be further extended by peripheral components and their failures. This includes increased system complexity due to line branching, clogging of lines, leakage at joints and lines, degradation of pump or compressor, stuck valves, etc.
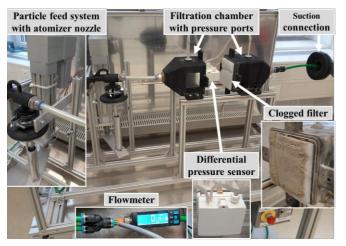

Figure 2. Front side of the filtration test bench with its particle feed system, sensors, open filtration chamber, and a clogged filter (Hagmeyer, Mauthe, Dutt, & Zeiler, 2021).

5.2. Air Filtration Test Bench

In the following, the test bench created in accordance with the selection is introduced. The test bench enables data generation by performing run-to-failure tests on the filtration process of gas or, in this case, specifically compressed air. The front of the test bench with its essential components, namely the particle feeder and the filtration chamber, is shown in Figure 2. When testing filter media, a general distinction is made between pressure and suction operation. The fluid flow through the filter can be achieved by increasing the pressure in front of the filter or reducing the
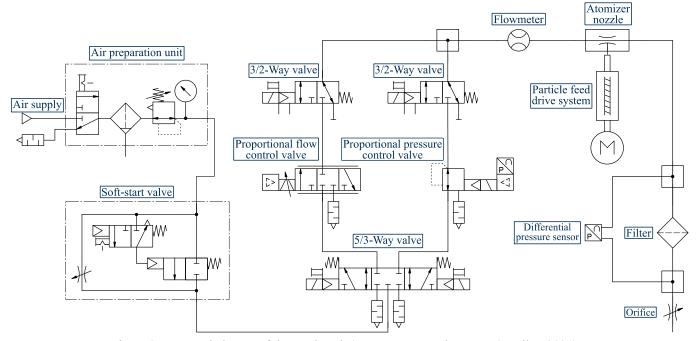

Figure 3. Pneumatic layout of the test bench (Hagmeyer, Mauthe, Dutt, & Zeiler, 2021).

pressure behind the filter compared to the ambient pressure. In this case, an external pressurized air supply is used to employ the pressure operation principle. The pneumatic layout of the test bench can be seen in Figure 3. The air preparation unit and the soft-start valve shown in it serve exclusively to ensure proper functioning of the system but do not affect the test procedure. By means of the corresponding valves, the filter loading can be carried out in a flow- or pressure-controlled mode. The connected 5/3- and 3/2-way valves enable selection of the respective control mode. The filter loading process is recorded by a flowmeter and a differential pressure sensor. A controlled amount of dust is introduced into the compressed air stream by using an atomizer nozzle and a particle feed drive system. The components behind the 3/2-way valves are those essential for testing, which are all shown in Figure 2. In the following, a further insight into the key features of the test bench and the test procedure is provided.

Filter media: The test objects are filter pads made of cut flat material, which are fixed in the filtration chamber. The effective filter area is hereby squared and $(78.3\ mm)^2$ in size. Figure 2 also depicts a loaded filter after finishing a life test. In addition to the effective filter area, particularly important properties for testing are the nominal volume flow per unit area, the maximum volume flow per unit area, the differential pressure in the unloaded state, the maximum differential pressure and the matching of filter class and particle distribution utilized. For the physical modeling of the filter loading, additional information regarding the filter medium is required. For fibrous filters, this includes the packing density, the fiber diameter and the pad thickness (Song, Park, & Lee, 2006).

Test particles and particle feeding system: Standardized test dusts with a defined distribution function in terms of particle size and chemical composition are used for filter loading. Here, Arizona Test Dust is used in accordance with the ISO 12103-1 standard in particle sizes from A2 to A4. Representative differential pressure trajectories of the three particle sizes are shown in Figure 4. Size type A2 is the test dust with the smallest mean particle size and A4 the dust with the largest. As is common in filtration, the smaller particles result in a higher differential pressure for the same loading mass (Song, Park, & Lee, 2006).

The dispersion of the dust particles into the compressed air takes place by means of a nozzle for powdered solids, based on the Venturi principle. The test dust is placed in a cylinder and fed to the nozzle by an electrically-operated spindle drive, as can be seen in Figure 3. By moving the spindle drive, the quantity of particles introduced into the air stream per time is controlled independently from the flow rate and allows setting load profiles.
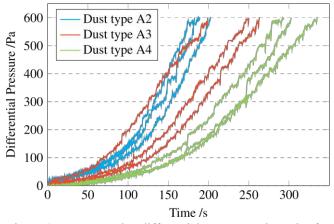

Figure 4. Representative differential pressure trajectories for filter loading with three different particle sizes.

Sensors: During life tests, the volume flow rate, the velocity of the particle feed drive, and the differential pressure across the filter are recorded. The flow sensor is located in front of the particle nozzle. The reason for this is the underlying measuring principle of thermal cooling by the flowing fluid. This is considerably influenced in its measuring accuracy by the contamination of the pressurized air. However, the sensor arrangement in front of the nozzle causes the flow rate to be falsified. Therefore, characteristic curves are measured and used for compensation. The differential pressure across the filter can be measured up to 2500 Pa, which is considerably above the regular operating range of the tested filters. The pressure ports in the filtration chamber are designed to minimize the influence on the differential pressure reading by the air flowing past the ports.

Control and user interface: The control and data acquisition is undertaken by a programmable logic controller (PLC). The user interface (UI) is based on Node-Red and communicates with the PLC via OPC UA, which is a protocol for industrial machine-to-machine communication. The life cycle tests are automated, whereby the test conditions and the termination criteria are defined in the UI.

As with any scientific experiment, an effort is made to reduce the influence of external disturbances on the measurement result. For this purpose, the nozzle and the filtration chamber are always cleaned between two test runs. The spindle of the particle feeding system is always in the same bottom end position at the beginning of a test and the cylinder is filled with the same amount of dust. However, variations in measurement results, which are inherent to the filter application, are an essential part of its use in studies. Such fluctuations are common in PHM applications and a source of uncertainty for PHM tasks (Goebel, et al., 2017). For instance, filter media made of randomly oriented non-woven fiber material are tested. The orientation of fibers when the filter media is manufactured results in varying filtration characteristics among tested filter pads (Chase, Beniwal, & Venkataraman, 2000).

5.3. Data Set: Preventive to Predictive Maintenance

A data set has been published by the name *Preventive to Predictive Maintenance*. This data set reflects the data scenario *Data without the end of life being recorded (a)*. Information on its availability is provided in section 5.4. The data set is based on the transition from an application of preventive maintenance to an application of predictive maintenance, with the data deficiencies that it entails. It is an issue with strong industrial relevance (Selcuk, 2017). The fundamental difference between the two forms of maintenance is illustrated in Figure 5. Preventive maintenance takes action when a predefined threshold of time units, work cycles, etc. is reached. When determining the threshold value, factors such as the failure costs, wasted life, as well as the lead time of a maintenance action need to be taken into account. (Wang, Chu, & Wu, 2007). Since the actual state of the system under consideration is not incorporated, life is given away under a mild load, and under excessive load failure can still occur, as depicted in Figure 5a. On the other hand, predictive maintenance determines the current condition and predicts at least the RUL. This shall enable improved management of the entire maintenance process, a higher usage rate of the available life and a reduction of unplanned downtimes (Selcuk, 2017).

The data set mimics already having run-to-threshold data sets at hand. However, due to the fixed maintenance periods while recording the data, these service lives are at most as long as the maintenance interval. The only times when the system failure is known is when lifespan is shorter than the maintenance interval. All other service lives are right censored. The challenge of this data scenario is to make the most use of the right censored service lives for the effective development of a RUL prediction. The utilization of right censored data in statistical lifetime models has already been thoroughly investigated in the related discipline of reliability engineering (Yang, 2007). Nonetheless, in the case of PHM, the research on censored data is not as extensive. There are only a limited number of papers on this subject; for instance, Widodo and Yang (2011), TV, Diksha, Malhotra, Vig, and Shroff (2019), and Chi, Lin, Chen, and Huang (2020).

The training data provided contains 50 predominantly censored service lives. However, the test data contains 50 randomly censored service lives for which the corresponding RUL is also given. The aim is to use the vast but censored training data to develop a RUL prediction for the test data and present solutions for addressing the data scenario. The data set incorporates a much more detailed description on the test configurations used, so that model-based and hybrid approaches are also feasible. Among others, this involves the physical properties of the filter (filter type, filter area, fiber diameter, degree of filling, filter thickness) and the properties of the test dust (distribution of particle size, density).
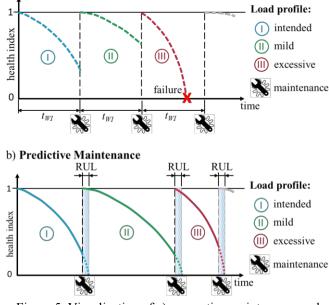

Figure 5. Visualization of a) preventive maintenance and b) predictive maintenance (Hagmeyer, Mauthe, Dutt, & Zeiler, 2021).

5.4. Publication of Further Data Sets for Additional Data Scenarios

One main objective of the program is to publish data sets for data scenarios where no data sets are available thus far. Table 2 lists the planned publication date of further data sets regarding their data scenarios. Besides the core objective of encouraging research on data scenarios, the data sets could also be used directly for PHM in related filtering applications, for example, as part of the research topic transfer learning (Moradi & Groth, 2020).

Data scenario	Publication date
Data without the end of life being recorded (a)	Already published
Data with random recording (c)	12/2021
Run-to-failure data with differing measurement bias and noise (i)	6/2022
Data batches with total/with partial/ without time information (e)	10/2022
Presence of training data which is to be mapped to the extended state space (h)	5/2023
Data on the life of redundant subsystems containing the failure of subsystem(s) (f)	2023-2024

Table 2. Schedule for planned releases of publicly available data sets on data scenarios.

In addition to the data set, a detailed description of the respective experimental condition is provided. The description includes sufficient information to allow physical modeling to be included in the solution approach for the given data scenario. The data sets and their experiment descriptions are made available on Kaggle for public use under the license type CC BY 4.0. The first data set *Preventive to Predictive Maintenance* is already released. The publishing account and its URL are:

- Prognostics @ HSE
- https://www.kaggle.com/prognosticshse

The data of each test run from the test bench will only be part of one data set. New test runs will be performed for each data set, preferably also under different operating conditions. The data sets to be published are therefore from different populations. This ensures that by including the data set of another data scenario, no inference is possible for resolving the respective deficiencies and data characteristics.

6. CONCLUSIONS AND OUTLOOK

6.1. Conclusions

Data is the basis for a purposeful implementation of PHM. However, in industrial applications, this data typically involve specific characteristics and deficiencies. The defined criteria collection provides an approach to systematically assess specific data characteristics and deficiencies regarding PHM. The data scenarios identified by means of the criteria catalog hold strong relevance and addressing them appropriately is an important research issue in PHM.

The ensuing overview of publicly available data sets leads to two main conclusions. First, data sets appropriate for PHM research are not only provided by typical sources such as the PHM Data Challenges of the PHM Society; rather, there are also platforms on general ML that increasingly provide suitable data sets. These platforms are not covered by any previous data set overview, making this overview the most comprehensive one thus far. Second, the data sets listed in the overview rarely cover industry-relevant data scenarios, but instead often represent seamless records of specific tests and simulations.

The established program to facilitate research on addressing data scenarios therefore has as one of its main objectives to provide data sets representing these data scenarios. The selection process of the test bench application revealed that filtration is the most suitable application for generating appropriate data. The test bench created and introduced features run-to-failure trajectories of filter loading that show the general pattern of an increasing degradation rate, which is representative of many PHM applications. A data set on one data scenario has already been published, while the schedule of further publications is also outlined. Overall, this

paper provides a significant step forward in the research on addressing industry-relevant data scenarios.

6.2. Outlook

Based on the identified industry-relevant data scenarios, a methodology is to be developed within the scope of the introduced program. There are already methodologies for PHM, but none that address ways of dealing with such data scenarios. The purpose of this methodology is to recommend solutions for relevant data scenarios that feature data deficiencies.

From the authors' perspective, there are a number of potential research issues to further advance the state of research on the topic of this paper. Those issues extend beyond the scope of the mentioned program.

The key aspect here is how to address different data scenarios. This aspect is one of the inherent features that distinguishes PHM as a specific field of application from general research on data-driven, hybrid, and model-based methods. In PHM, there are already studies dedicated to particular data scenarios, but there remains substantial research potential.

This research on data scenarios can be backed up by the creation of further data sets, in addition to what is planned here. Having a selection of data sets for the same data scenario where one set then proves to be the most appropriate benchmark through the use by the research community is a preferable way forward. In addition, a wide-scale industry study could provide a quantitative statement on the relevance of different data scenarios.

APPENDIX

No.	Data Set Designation and Reference to data origin	URL for Data Download
1	**Data Challenge PHM Soc. 2008 - Turbofan** NASA Prognostics Center of Excellence	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#phm08_challenge
2	**Data Challenge PHM Soc. 2009 - Gearbox Fault Detection** PHM Society, Gearbox fault detection data set, 2010	https://c3.nasa.gov/dashlink/resources/997/
3	**Data Challenge PHM Soc. 2010 - CNC milling machine cutters** X. Li[1], B.S. Lim[1], J.H. Zhou[1], S. Huang[1], S.J. Phua[1], K.C. Shaw[1], and M.J. Er[2] [1]Singapore Institute of Manufacturing Technology, 71 Nanyang Drive, Singapore 638075 [2]School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798	https://www.phmsociety.org/competition/phm/10
4	**Data Challenge PHM Soc. 2011 - Anemometer Fault Detection** Creators Unknown	https://www.phmsociety.org/competition/phm/11/problem
5	**Data Challenge PHM Soc. 2013** Creators Unkown	https://www.phmsociety.org/events/conference/phm/13/challenge
6	**Data Challenge PHM Soc. 2014** Creators Unknown	https://www.phmsociety.org/events/conference/phm/14/data-challenge
7	**Data Challenge PHM Soc. 2015 - Plant Fault Detection** Creators Unknown	https://www.phmsociety.org/events/conference/phm/15/data-challenge
8	**Data Challenge PHM Soc. 2016 - Semiconductor CMP** Crystec Technology Trading GmbH	https://www.phmsociety.org/events/conference/phm/16/data-challenge
9	**Data Challenge PHM Soc. 2017 - Bogie Vehicle** Creators Unknown	https://www.phmsociety.org/events/conference/phm/17/data-challenge
10	**Data Challenge PHM Soc. 2018 - Ion Mill in Waver Manufactoring** A dataset shared by Kai Goebel (NASA)	https://www.phmsociety.org/events/conference/phm/18/data-challenge
11	**Data Challenge PHM Soc. 2019 - Fatigue Cracks** Creators Unknown	https://www.phmdata.org/2019datachallenge/
12	**Data Challenge PHM Soc. 2020 Europe - Filtration System** Eker, O. F., Camci, F., & Jennions, I. K. (2016). Physics-based prognostic modelling of filter clogging phenomena. Mechanical Systems and Signal Processing, 75, pp. 395-412.	http://phmeurope.org/2020/data-challenge-2020
13	**Data Challenge PHM Soc. 2021 Europe: SCARA-robot** Swiss Centre for Electronics and Microtechnology (CSEM)	https://phm-europe.org/data-challenge
14	**PHM Data Challenge 2021 - Turbofan 2** M. Chao, C.Kulkarni, K. Goebel and O. Fink (2021). "Aircraft Engine Run-to-Failure Dataset under real flight conditions", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA (additional validation data compared to data set no. 29)	https://data.phmsociety.org/2021-phm-conference-data-challenge/
15	**NASA - Milling Data set** A. Agogino and K. Goebel. BEST lab, UC Berkeley. "Milling Data Set ", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#milling
16	**NASA - Li-ion Battery Aging Datasets** B. Saha and K. Goebel. "Battery Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#battery
17	**NASA - Bearing Data Set** J. Lee, H. Qiu, G. Yu, J. Lin, and Rexnord Technical Services. IMS, University of Cincinnati. "Bearing Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#bearing
18	**NASA - Turbofan engine degradation simulation data set** A. Saxena and K. Goebel. "Turbofan Engine Degradation Simulation Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#turbofan

No.	Data Set and Reference to data origin	URL for Data Download
19	**NASA - CFRP Composites Data Set** Abhinav Saxena, Kai Goebel, Cecilia C. Larrosa, and Fu-Kuo Chang "CFRP Composites Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#composites
20	**NASA - IGBT** J. Celaya, Phil Wysocki, and K. Goebel "IGBT Accelerated Aging Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#igbt
21	**NASA - Trebuchet** B. Morton. Sentient Corporation. "Trebuchet Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#trebuchet
22	**NASA - Randomized Battery Usage Data Set** B. Bole, C. Kulkarni, and M. Daigle "Randomized Battery Usage Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#batteryrnddischarge
23	**NASA - Capacitor Electrical Stress** J. Renwick, C. Kulkarni, and J Celaya "Capacitor Electrical Stress Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#escapacitor
24	**NASA - MOSFET Thermal Overstress Aging** J. R. Celaya, A. Saxena, S. Saha, and K. Goebel "MOSFET Thermal Overstress Aging Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#mosfet
25	**NASA - Capacitor Electrical Stress 2** J. Celaya, C. Kulkarni, G. Biswas, and K. Goebel "Capacitor Electrical Stress Data Set - 2", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#eoscapacitor
26	**NASA - HIRF Battery** C. Kulkarni, E. Hogge, C. Quach and K. Goebel "HIRF Battery Data Set", NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository), NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#hirfbatterytests
27	**NASA - Small Satellite Power Simulation** C. Kulkarni and A. Guarneros "Small Satellite Power Simulation Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#smallsat
28	**NASA - Algae Raceway Data Set** Brad Bebout, Leslie Profert-Bebout, Erich Fleming, Angela Detweiler, and Kai Goebel "Algae Raceway Data Set", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#algae
29	**NASA - Turbofan engine degradation simulation data set 2** M. Chao, C.Kulkarni, K. Goebel and O. Fink (2021). "Aircraft Engine Run-to-Failure Dataset under real flight conditions", NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA	https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#turbofan-2
30	**IEEE PHM Data Challenge 2012 - FEMTO Bearing Data Set** FEMTO-ST Institute, Besançon, France	https://github.com/wkzs111/phm-ieee-2012-data-challenge-dataset
31	**IEEE PHM Data Challenge 2014 - Fuel Cell** FCLAB Research Federation (FR CNRS 3539, France)	https://repository.lboro.ac.uk/articles/dataset/IEEE_2014_Data_Challenge_Data/3518141
32	**Kaggle - Air pressure system failures in Scania trucks** Scania CV AB - Stockholm	https://www.kaggle.com/uciml/aps-failure-at-scania-trucks-data-set
33	**Kaggle - CNC Mill Tool Wear** System-level Manufacturing and Automation Research Testbed (SMART) at the University of Michigan	https://www.kaggle.com/shasun/tool-wear-detection-in-cnc-mill#experiment_03.csv
34	**Kaggle - E-coating ultrafiltration maintenance** Tinsley Bridge Limited, Sheffield, United Kingdom	https://www.kaggle.com/boyangs444/process-data-for-predictive-maintenance
35	**Kaggle - Genesis Demonstrator** Institut für industrielle Informationstechnik (inIT) der Technischen Hochschule Ostwestfalen-Lippe (IMPROVE)	https://www.kaggle.com/inIT-OWL/genesis-demonstrator-data-for-machine-learning

No.	Data Set and Reference to data origin	URL for Data Download
36	**Kaggle - Microsoft Azure Predictive Maintenance** Azure AI Notebooks for Predictive Maintenance	https://www.kaggle.com/arnabbiswas1/microsoft-azure-predictive-maintenance?select=PdM_telemetry.csv
37	**Kaggle - One Year Industrial Component Degradation** Institut für industrielle Informationstechnik (inIT) der Technischen Hochschule Ostwestfalen-Lippe (IMPROVE)	https://www.kaggle.com/inIT-OWL/one-year-industrial-component-degradation
38	**Kaggle - Predict maintenance** Creators Unkown	https://www.kaggle.com/c/predictive-maintenance1/overview
39	**Kaggle - Predictive Maintenance Hydraulics System** ZeMA gGmbH, Eschberger Weg 46, 66121 Saarbrücken	https://www.kaggle.com/mayank1897/condition-monitoring-of-hydraulic-systems?select=description.txt
40	**Kaggle - Preventive to Predictive Maintenance** Hagmeyer, Simon; Mauthe, Fabian; Dutt, Marcel; Zeiler, Peter "Preventive to Predictive Maintenance Dataset", Hochschule Esslingen - University of Applied Sciences: Research Department Reliability Engineering and Prognostics and Health Management, Göppingen, Germany, DOI: 10.34740/kaggle/dsv/2339298	https://www.kaggle.com/prognosticshse/preventive-to-predicitve-maintenance
41	**Kaggle - Production Plant Data for Condition Monitoring** IMPROVE research project	https://www.kaggle.com/inIT-OWL/production-plant-data-for-condition-monitoring
42	**Kaggle - Pump** Kaggle User: UnknownClass	https://www.kaggle.com/nphantawee/pump-sensor-data
43	**Kaggle - Sensor Fault Detection** Schneider-Electric, France	https://www.kaggle.com/arashnic/sensor-fault-detection-data
44	**Kaggle - Versatile Production System** Institut für industrielle Informationstechnik (inIT) der Technischen Hochschule Ostwestfalen-Lippe (IMPROVE)	https://www.kaggle.com/inIT-OWL/versatileproductionsystem
45	**UCI - Condition Based Maintenance of Naval Propulsion Plants Data Set** 1: DIBRIS - University of Genoa 2: School of Marine Science and Technology, Newcastle University	http://archive.ics.uci.edu/ml/datasets/Condition+Based+Maintenance+of+Naval+Propulsion+Plants
46	**UCI - Condition monitoring of hydraulic systems Data Set** ZeMA - Zentrum für Mechatronik und Automatisierungstechnik gemeinnützige GmbH Saarbrücken	https://archive.ics.uci.edu/ml/datasets/Condition+monitoring+of+hydraulic+systems
47	**UCI - Mechanical Analysis 1990** University of Torino, Italy	http://archive.ics.uci.edu/ml/datasets\Mechanical+Analysis
48	**UCI - Robot Execution Failures Data Set** Universidade Nova de Lisboa, Monte da Caparica, Portugal	https://archive.ics.uci.edu/ml/datasets\Robot+Execution+Failures
49	**UCI - Steel Plates Faults** Semeion Research Center of Sciences of Communication - Rome, Italy	https://archive.ics.uci.edu/ml/datasets/steel+plates+faults
50	**UCI - AI4I 2020 Predictive Maintenance Data Set** Matzka, S. (2020). "Explainable Artificial Intelligence for Predictive Maintenance Applications," *Proceedings of the Third International Conference on Artificial Intelligence for Industries (AI4I)*, pp. 69-74, doi: 10.1109/AI4I49448.2020.00023.	https://archive.ics.uci.edu/ml/datasets/AI4I+2020+Predictive+Maintenance+Dataset
51	**AIDAR Lab - Air Compressor** Nishchal K. Verma, R. K. Sevakula, S. Dixit and Salour A. (2016). Intelligent Condition Based Monitoring using Acoustic Signals for Air Compressors, *IEEE Transactions on Reliability*, vol. 65, no. 1, pp. 291-309.	http://www.iitk.ac.in/idea/datasets/
52	**AIDAR Lab - Drill Bit** Nishchal K. Verma, R. K. Sevakula, S. Dixit and Salour A. (2015). Data Driven Approach for Drill Bit Monitoring, *Reliability Digest*, pp. 19-26.	http://www.iitk.ac.in/idea/datasets/
53	**Aramis - Data Challenge ESREL2020PSAM15** Francesco Cannarile (Aramis Srl, Italy), Michele Compare (Aramis Srl, Italy and Politecnico di Milano, Italy), Piero Baraldi (Politecnico di Milano, Italy), Zhe Yang (Politecnico di Milano, Italy), Enrico Zio (Politecnico di Milano, Italy and MINES ParisTech, France).	https://aramis3d.com/innovation-challenges/
54	**Backblaze - Hard Drive Stats** Backblaze, Inc.	https://www.backblaze.com/b2/hard-drive-test-data.html

No.	Data Set and Reference to data origin	URL for Data Download
55	**Bearing Data Center - Fan and Bearing** Case Western Reserve University, USA	https://csegroups.case.edu/bearingdatacenter/pages/download-data-file
56	**GitHub - Predictive Maintenance using PySpark** GitHub User: linya9191	https://github.com/Azure/PySpark-Predictive-Maintenance
57	**Mendeley - Brushless DC motor** Mazzoleni, Mirko; Scandella, Matteo; Previdi, Fabio; Pispola, Giulio (2019). Data for: First endurance activity of a Brushless DC motor for aerospace applications - REPRISE project, Mendeley Data, V2, doi: 10.17632/m58bdhy2df.2	https://data.mendeley.com/datasets/m58bdhy2df/2
58	**Mendeley - Diesel Engine Faults** Denys Pestana-Viana - Federal Center of Technological Education Celso Suckow da Fonseca (CEFET-RJ), Rio de Janeiro, Brazil	https://data.mendeley.com/datasets/k22zxz29kr/1
59	**MFPT - Condition Based Maintenance Fault** Data Assembled and Prepared on behalf of MFPT by Dr Eric Bechhoefer, Chief Engineer, NRG Systems	https://www.mfpt.org/fault-data-sets/
60	**OpenEI - Evaluation of Building Fault** LBNL, PNNL, NREL, ORNL and ASHRAE RP-1312 (Drexel University)	https://data.openei.org/submissions/910
61	**OpenEI - Gearbox Fault Diagnosis** OpenEI User: Yogesh Pandya	https://data.openei.org/submissions/623
62	**ResearchGate - Driveline Unbalanced Shaft** Giacomo Barbieri and David Sanchez-Londono and Laura Cattaneo and Luca Fumagalli and David Romero	https://www.researchgate.net/publication/341220576_Dataset_A_Case_Study_for_Problem-based_Learning_Education_in_Fault_Diagnosis_Assessment
63	**SDOL - Diagnostics 101 bearing data** Kim, S., An, D., Choi, J-H. (2020). Diagnostics 101: A Tutorial for Fault Diagnostics of Rolling Element Bearing Using Envelope Analysis in MATLAB. *Applied Sciences*. 10(20):7302. doi:10.3390/app10207302	https://www.kau-sdol.com/bearing
64	**SDOL - HS Gear** Creators Unknown	https://www.kau-sdol.com/kaug
65	**SDOL - KAU Gear** Creators Unknown	https://www.kau-sdol.com/kaug
66	**Uni Lulea - Vibration Signals Wind Turbines** Luleå University of Technology: Martin del Campo Barraza, Sergio (Department of Computer Science, Electrical and Space Engineering, Embedded Internet Systems Lab); Sandin, Fredrik (Department of Computer Science, Electrical and Space Engineering, Embedded Internet Systems Lab); Strömbergsson, Daniel (Department of Engineering Sciences and Mathematics, Machine Elements)	http://ltu.diva-portal.org/smash/record.jsf?pid=diva2%3A1244889&dswid=-6104
67	**Uni Paderborn KAt - Bearing Damage** University Paderborn Kat	https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-datacenter/data-sets-and-download/
68	**Virkler** Virkler, D. A., Hillberry, B. M. and Goel, P. K. (1979). The Statistical Nature of Fatigue Crack Propagation. Journal of Engineering Materials and Technology 101, 148–153.	https://rdrr.io/github/SimoneHermann/hierRegSDE/man/Virkler.html
69	**WASEDA - Fault Detection and Classification** De Bruijn, B., Nguyen, T. A., Bucur, D., & Tei, K. (2016). Benchmark datasets for fault detection and classification in sensor data. In A. Ahrens, O. Postolache, & C. Benavente-Peces (Eds.), *SENSORNETS 2016 - Proceedings of the 5th International Confererence on Sensor Networks* (pp. 185-195). (SENSORNETS 2016 - Proceedings of the 5th International Confererence on Sensor Networks). SciTePress. https://doi.org/10.5220/0005637901850195	https://waseda.pure.elsevier.com/en/publications/benchmark-datasets-for-fault-detection-and-classification-in-sens
70	**Zenodo - Predictive maintenance dataset** Huawei Munich Research Center	https://zenodo.org/record/3653909#.YAmTTBYxkcQ

Table 3. Overview of publicly available datasets for research on PHM, sorted by the providing data platform. The designation of the data sets is divided into two parts consisting of the source or platform and the underlying technical application.

REFERENCES

Abdolghader, P., Brochot, C., Haghighat, F., & Bahloul, A. (2018). Airborne nanoparticles filtration performance of fibrous media: A review. *Science and Technology for the Built Environment, 24*(6), pp. 648-672. doi:10.1080/23744731.2018.1452454

Aizpurua, J. I., & Catterson, V. M. (2015). Towards a Methodology for Design of Prognostic Systems. *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, *7*. Coronado, USA. doi:10.36001/phmconf.2015.v7i1.2646

Aizpurua, J. I., & Catterson, V. M. (2016). ADEPS: A Methodology for Designing Prognostic Applications. *Proceedings of the European Conference of the Prognostics and Health Management Society.* Bilbao, Spain. doi:10.36001/phme.2016.v3i1.1585

An, D., Choi, J.-H., & Kim, N. H. (2013). Prognostics 101: A tutorial for particle filter-based prognostics algorithm using Matlab. *Reliability Engineering & System Safety, 115*, pp. 161-169. doi:10.1016/j.ress.2013.02.019

Atamuradov, V., Medjaher, K., Dersin, P., Lamoureux, B., & Zerhouni, N. (2017). Prognostics and Health Management for Maintenance Practitioners - Review, Implementation and Tools Evaluation. *International Journal of Prognostics and Health Management, 8*(3). doi:10.36001/ijphm.2017.v8i3.2667

Bergman, W., Taylor, R. D., Miller, H. H., Bierman, A. H., Hebard, H. D., daRoza, R. A., & Lum, B. Y. (1978). Enhanced filtration program at LLL. A progress report. *Proceedings of the 15th Nuclear Air Cleaning Conference.* Boston, USA.

Cachada, A., Moreira, P. M., Romero, L., Barbosa, J., Leitno, P., Gcraldcs, C. A., . . . Moreira, A. H. (2018). Maintenance 4.0: Intelligent and Predictive Maintenance System Architecture. *Proceedings of the IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, (pp. 139–146). Turin, Italy. doi:10.1109/ETFA.2018.8502489

Cannarile, F., Baraldi, P., & Zio, E. (2019). An evidential similarity-based regression method for the prediction of equipment remaining useful life in presence of incomplete degradation trajectories. *Fuzzy Sets and Systems, 367*, pp. 36-50. doi:10.1016/j.fss.2018.10.008

Chao, M. A., Kulkarni, C., Goebel, K., & Fink, O. (2021). Aircraft Engine Run-to-Failure Dataset under Real Flight Conditions for Prognostics and Diagnostics. *Data, 6*(1). doi:10.3390/data6010005

Chase, G. G., Beniwal, V., & Venkataraman, C. (2000). Measurement of uni-axial fiber angle in non-woven fibrous media. *Chemical Engineering Science, 55*(12), pp. 2151-2160. doi:10.1016/S0009-2509(99)00510-2

Cheng, S., Raghavan, N., Gu, J., Mathew, S., & Pecht, M. (2018). Physics-of-failure approach to PHM. In *Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things.* Hoboken, USA: JohnWiley and Sons Ltd.

Chi, Z., Lin, J., Chen, R., & Huang, S. (2020). Data-driven approach to study the polygonization of high-speed railway train wheel-sets using field data of China's HSR train. *Measurement, 149*. doi:10.1016/j.measurement.2019.107022

Chikhi, N., Clavier, R., Laurent, J.-P., Fichot, F., & Quintard, M. (2016). Pressure drop and average void fraction measurements for two-phase flow through highly permeable porous media. *Annals of Nuclear Energy, 94*, pp. 422-432. doi:10.1016/j.anucene.2016.04.007

Cubillo, A., Perinpanayagam, S., & Esperon-Miguez, M. (2016). A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery. *Advances in Mechanical Engineering, 8*(8), pp. 1–21. doi:10.1177/1687814016664660

Dalzochio, J., Kunst, R., Pignaton, E., Binotto, A., Sanyal, S., Favilla, J., & Barbosa, J. (2020). Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. *Computers in Industry, 123*, p. 103298. doi:10.1016/j.compind.2020.103298

Dong, W., Liu, S., & Du, Y. (2019). Optimal periodic maintenance policies for a parallel redundant system with component dependencies. *Computers & Industrial Engineering, 138*, p. 106133. doi:10.1016/j.cie.2019.106133

Eker, O. F., Camci, F., & Jennions, I. K. (2016). Physics-based prognostic modelling of filter clogging phenomena. *Mechanical Systems and Signal Processing, 75*, pp. 395-412. doi:10.1016/j.ymssp.2015.12.011

Elattar, H. M., Elminir, H. K., & Riad, A. M. (2016). Prognostics: a literature review. *Complex & Intelligent Systems, 2*, pp. 125-154. doi:10.1007/s40747-016-0019-3

Ge, M., Du, R., Zhang, G., & Xu, Y. (2004). Fault diagnosis using support vector machine with an application in sheet metal stamping operations. *Mechanical Systems and Signal Processing, 18*(1), pp. 143 -159. doi:10.1016/S0888-3270(03)00071-2

Goebel, K., Daigle, M., Saxena, A., Sankararaman, S., Roychoudhury, I., & Celaya, J. (2017). *Prognostics:*

The science of making predictions. CreateSpace Independent Publishing Platform.

Gu, J., Barker, D., & Pecht, M. (2007). Prognostics implementation of electronics under vibration loading. *Microelectronics Reliability, 47*(12), pp. 1849 - 1856. doi:10.1016/j.microel.2007.02.015

Hagmeyer, S., Mauthe, F., Dutt, M., & Zeiler, P. (2021). Preventive to predictive maintenance. https://www.kaggle.com/prognosticshse/preventive -to-predicitve-maintenance, DOI: 10.34740/kaggle/dsv/2339298.

Hemmer, M., Klausen, A., van Khang, H., Robbersmyr, K., & Waag, T. (2019). Simulation-driven Deep Classification of Bearing Faults from Raw Vibration Data. *International Journal of Prognostics and Health Management, 10*(4). doi:10.36001/ijphm.2019.v10i4.2615

Huang, B., Di, Y., Jin, C., & Lee, J. (2017). Review of data-driven prognostics and health management techniques: lessions learned from PHM data challenge competitions. *Proceedings of the Conference Machine Failure Prevention Technology.* Virginia Beach, USA.

Huang, C.-Y., & Dzulfikri, Z. (2021). Stamping Monitoring by Using an Adaptive 1D Convolutional Neural Network. *Sensors, 21*(1). doi:10.3390/s21010262

Javed, K., Gouriveau, R., & Zerhouni, N. (2017). State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels. *Mechanical Systems and Signal Processing, 94*, pp. 214–236. doi:10.1016/j.ymssp.2017.01.050

Jia, X., Huang, B., Feng, J., Cai, H., & Lee, J. (2018). Review of PHM Data Competitions from 2008 to 2017. *Proceedings of the Annual Conference of the PHM Society, 10.* Philadelphia, USA. doi:10.36001/phmconf.2018.v10i1.462

Kaggle: Your Machine Learning and Data Science Community. (2021). Retrieved 06 17, 2021, from https://www.kaggle.com/

Kahn, B. K., Strong, D. M., & Wang, R. Y. (2002). Information quality benchmarks: product and service performance. *Communications of the ACM, 45*(4), pp. 184-192. doi:10.1145/505248.506007

Kim, N.-H., An, D., & Choi, J.-H. (2016). *Prognostics and health management of engineering systems: An introduction.* Berlin, Heidelberg: Springer.

Kundu, P., Darpe, A. K., & Kulkarni, M. S. (2020). A review on diagnostic and prognostic approaches for gears. *Structural Health Monitoring.* doi:10.1177/1475921720972926

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing, 104*, pp. 799–834. doi:10.1016/j.ymssp.2017.11.016

Mendes, A. A., Coit, D. W., & Ribeiro, J. L. (2014). Establishment of the optimal time interval between periodic inspections for redundant systems. *Reliability Engineering & System Safety, 131*, pp. 148–165. doi:10.1016/j.ress.2014.06.021

Merino, J., Caballero, I., Rivas, B., Serrano, M., & Mario, P. (2016). A Data Quality in Use model for Big Data. *Future Generation Computer Systems, 63*, pp. 123-130. doi:10.1016/j.future.2015.11.024

Moradi, R., & Groth, K. (2020). On the application of transfer learning in prognostics and health management. *Proceedings of the Annual Conference of the PHM Society, 12.* virtual. doi:10.36001/phmconf.2020.v12i1.1300

Niculita, O., Irving, P., & Jennions, I. K. (2012). Use of COTS Functional Analysis Software as an IVHM Design Tool for Detection and Isolation of UAV Fuel System Faults. *Proceedings of the Annual Conference of Prognostics and Health Management Society, 4.* Minneapolis, USA. doi:10.36001/phmconf.2012.v4i1.2116

Novick, V., Higgins, P., Dierkschiede, B., Abrahamson, C., Richardson, W., Monson, P., & Ellison, P. (1990). Efficiency and mass loading characteristics of a typical HEPA filter media material. *Proceedings of the 21st DOE/NRC Nuclear Air Cleaning Conference.* San Diego, USA.

Ochella, S., & Shafiee, M. (2019). Artificial Intelligence in Prognostic Maintenance. *Proceedings of the 29th European Safety and Reliability Conference (ESREL)*, (pp. 3424–3431). Hannover, Germany. doi:10.3850/978-981-11-2724-3_0188-cd

Omri, N., Al Masry, Z., Mairot, N., Giampiccolo, S., & Zerhouni, N. (2021). Towards an adapted PHM approach: Data quality requirements methodology for fault detection applications. *Computers in Industry, 127.* doi:10.1016/j.compind.2021.103414

PHM Society. (2021). *The Prognostics and Health Management Society .* Retrieved 06 16, 2021, from https://phmsociety.org/

Pillai, P., Kaushik, A., Bhavikatti, S., Roy, A., & Kumar, V. (2016). A Hybrid Approach for Fusing Physics and Data for Failure Prediction. *International Journal of Prognostics and Health Management, 7*(4). doi:10.36001/ijphm.2016.v7i4.2463

Popper, K. (1963). *Conjectures and refutations : The growth of scientific knowledge.* London, UK: Routledge and Kegan Paul.

Prognostics Center of Excellence - Data Repository. (2021). Retrieved 06 16, 2021, from https://ti.arc.nasa.gov/tech/dash/groups/pcoe/progn ostic-data-repository/

Ramasso, E., & Saxena, A. (2014). Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets. *International Journal of*

Prognostics and Health Management, 5(2). doi:10.36001/ijphm.2014.v5i2.2236

Saarela, O., Hulsund, J. E., Taipale, A., & Hegle, M. (2014). Remaining Useful Life Estimation for Air Filters at a Nuclear Power Plant. *Proceedings of the European Conference of the PHM Society.* Nantes, France. doi:10.36001/phme.2014.v2i1.1493

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. *Proceedings of the International Conference on Prognostics and Health Management. Denver, USA*. doi:10.1109/PHM.2008.4711414

Selcuk, S. (2017). Predictive maintenance, its implementation and latest trends. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 231*(9), pp. 1670-1679. doi:10.1177/0954405415601640

Shanbhag, V. V., Rolfe, B. F., Arunachalam, N., & Pereira, M. P. (2018). Investigating galling wear behaviour in sheet metal stamping using acoustic emissions. *Wear, 414*, pp. 31-42. doi:10.1016/j.wear.2018.07.003

Shanbhag, V., Pereira, M., Voss, B., Ubhayaratne, I., & Rolfe, B. (2019). Developing smart multi-sensor monitoring for tool wear in stamping process. *Proceedings of the IOP Conference Series: Materials Science and Engineering 651*. Enschede, Netherlands. doi:10.1088/1757-899X/651/1/012085

Sharma, V., & Parey, A. (2016). A Review of Gear Fault Diagnosis Using Various Condition Indicators. *Procedia Engineering, 144*, pp. 253-263. doi:10.1016/j.proeng.2016.05.131

Si, X.-S., Wang, W., Hu, C.-H., & Zhou, D.-H. (2011). Remaining useful life estimation – A review on the statistical data driven approaches. *European Journal of Operational Research, 213*(1), pp. 1–14. doi:10.1016/j.ejor.2010.11.018

Skaf, Z., Eker, O. F., & Jennions, I. K. (2017). System Component Degradation: Filter Clogging in a UAV Fuel System. *Proceedings of the Asia-Pacific Conference of the PHM Society*, (pp. 334-342). Jeju, Republik Korea.

Song, C. B., Park, H. S., & Lee, K. W. (2006). Experimental study of filter clogging with monodisperse PSL particles. *Powder Technology, 163*(3), pp. 152-159. doi:10.1016/j.powtec.2006.01.016

Sparks, T., & Chase, G. (2015). *Filters and filtration handbook* (6th ed.). Stanford: Elsevier Science.

Sreenuch, T., Khan, F., & Li, J. (2015). Particle Filter with Operational-Scalable Takagi–Sugeno Fuzzy Degradation Model for Filter-Clogging Prognosis. *Journal of Aerospace Information Systems, 12*(5), pp. 398-412. doi:10.2514/1.I010385

Statista Research Department. (2019). Size of the global industrial air filtration market between 2014 and 2025. https://www.statista.com/statistics/593079/global-industrial-air-filtration-market-size/: Retrieved April 12, 2021.

Thomas, D., Penicot, P., Contal, P., Leclerc, D., & Vendel, J. (2001). Clogging of fibrous filters by solid aerosol particles Experimental and modelling study. *Chemical Engineering Science, 56*(11), pp. 3549-3561. doi:10.1016/S0009-2509(01)00041-0

Tobon-Mejia, D. A., Medjaher, K., & Zerhouni, N. (2012). CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks. *Mechanical Systems and Signal Processing, 28*, pp. 167–182. doi:10.1016/j.ymssp.2011.10.018

TV, V., Diksha, Malhotra, P., Vig, L., & Shroff, G. (2019). Data-driven Prognostics with Predictive Uncertainty Estimation using Ensemble of Deep Ordinal Regression Models. *International Journal of Prognostics and Health Management, 10*(4). doi:10.36001/ijphm.2019.v10i4.2612

UCI Machine Learning Repository. (2021). Retrieved 06 17, 2021, from https://archive.ics.uci.edu/ml/index.php

Uckun, S., Goebel, K., & Lucas, P. J. (2008). Standardizing research methods for prognostics. *Proceedings of the International Conference on Prognostics and Health Management.* Denver, USA. doi:10.1109/PHM.2008.4711437

Večeř, P., Kreidl, M., & Šmíd, R. (2005). Condition Indicators for Gearbox Condition Monitoring Systems. *Acta Polytechnica, 45*(6), pp. 35-43. doi:10.14311/782

Wang, L., Chu, J., & Wu, J. (2007). Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process. *International Journal of Production Economics, 107*(1), pp. 151-163. doi:10.1016/j.ijpe.2006.08.005

Wang, R. Y., Storey, V. C., & Firth, C. P. (1995). A Framework for Analysis of Data Quality Research. *IEEE Transactions on Knowledge and Data Engineering, 7*(4), pp. 623-640. doi:10.1109/69.404034

Weiss, B., & Brundage, M. (2021). Measurement and Evaluation for Prognostics and Health Management (PHM) for Manufacturing Operations – Summary of an Interactive Workshop Highlighting PHM Trends. *International Journal of Prognostics and Health Management, 12*(1). doi:10.36001/IJPHM.2021.v12i1.2653

Widodo, A., & Yang, B.-S. (2011). Machine health prognostics using survival probability and support vector machine. *Expert Systems with Applications, 38*(7), pp. 8430-8437. doi:10.1016/j.eswa.2011.01.038

Wiese, B., Pedersen, N. L., Nadimi, E. S., & Herp, J. (2020). Estimating the Remaining Power Generation of Wind Turbines–An Exploratory Study for Main Bearing Failures. *Energies, 13*(13). doi:10.3390/en13133406

Xu, M., Baraldi, P., Al-Dahidi, S., & Zio, E. (2019). Fault Prognostics in Presence of Event-Based Measurements. *Proceedings of the 29th European Safety and Reliability Conference (ESREL)*, (pp. 1187-1193). Hannover, Germany. doi:10.3850/978-981-11-2724-3 0372-cd

Yang, G. (2007). *Life Cycle Reliability Engineering.* New York: John Wiley & Sons.

Zhu, J., Nostrand, T., Spiegel, C., & Morton, B. (2014). Survey of Condition Indicators for Condition Monitoring Systems. *Proceedings of the Annual Conference of the PHM Society*, *6.* Fort Worth, USA. doi:10.36001/phmconf.2014.v6i1.2514

BIOGRAPHIES

Simon Hagmeyer works as a research assistant in the Reliability Engineering and Prognostics and Health Management research group at Esslingen University of Applied Science, Germany. He is currently pursuing his PhD in cooperation with the University of Stuttgart. He received his B.Eng. and his M.Sc. degree in Mechatronics from Reutlingen University of Applied Science, Germany. His research interest includes diagnostics and prognostics, data-driven methods, hybrid methods and data science.

Fabian Mauthe works as a research assistant in the Reliability Engineering and Prognostics and Health Management research group at Esslingen University of Applied Science, Germany. He is currently pursuing his PhD in cooperation with the University of Stuttgart. He received his B.Sc. degree in Mechanical Engineering and Mechatronics and his M.Sc. degree in Mechanical Engineering from Furtwangen University of Applied Science, Germany. His research interest includes diagnostics and prognostics, condition monitoring, data-driven methods and machine learning.

Peter Zeiler is professor of the Mechanical and Systems Engineering faculty at Esslingen University of Applied Sciences in Germany, where he is head of the Reliability Engineering and Prognostics and Health Management research group. He graduated in Mechanical Engineering with the academic degree Dipl.-Ing. He was awarded the academic degree Dr.-Ing. from the University of Stuttgart. He then worked as a team leader in the R&D at the Robert Bosch GmbH. Thereafter, he became head of the departments for Reliability Engineering and for Drive Technology at the Institute of Machine Components at the University of Stuttgart. He joined Esslingen University of Applied Sciences in 2017. He is also committed to the Verein Deutscher Ingenieure (VDI) (English: Association of German Engineers). At the latter, he is chairman of the Safety and Reliability Advisory Board as well as the technical committees for Prognostics and Health Management and for Monte-Carlo-Simulation. Furthermore, he is conference chair of the German conference Fachtagung Technische Zuverlässigkeit (TTZ) (English: Technical Reliability Conference). His research interest is in the areas of prognostics and health management, methods for calculating reliability under operating conditions, as well as modeling and simulation of reliability and availability for complex systems.