Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen
©2025 ESREL SRA-E 2025 Organizers. *Published by* Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P7412-cd

Overview and Analysis of Publicly Available Degradation Data Sets for Tasks within Prognostics and Health Management

Fabian Mauthe, Luca Steinmann, Moritz Neu, and Peter Zeiler

Institute for Technical Reliability and Prognostics (IZP), Esslingen University of Applied Sciences, Germany.
E-mail: fabian.mauthe@hs-esslingen.de, luca.steinmann@hs-esslingen.de, moritz.neu@hs-esslingen.de, peter.zeiler@hs-esslingen.de

The effectiveness of Prognostic and Health Management (PHM) methods relies on degradation data that reflect the health of engineering systems over time. In particular, publicly available degradation data sets are of high value. Despite their importance, these data sets are rarely discussed comprehensively in the literature, with existing overviews often limited in scope and detail. As a result, the search process for suitable data sets is often very time-consuming for users of PHM methods. Therefore, this work provides a comprehensive overview of 98 publicly available degradation data sets and conducts a novel, detailed, PHM-specific analysis. In order to carry out the analysis, a taxonomy is developed to categorize and classify the data sets based on defined PHM-specific aspects. The resulting taxonomy classifies the data sets across 38 applications in 11 domains. The analysis provides a comprehensive overview and entry point for selecting and using the available data sets. It shows that almost half of the data sets are from the domains of electrical and mechanical components, with battery and bearing applications being the most common. However, the analysis also reveals that the number of data sets is limited for many applications.

Keywords: PHM, prognostics and health management, degradation data set, data set overview, degradation signal, fault detection, diagnosis, prognosis.

1. Introduction

Central to the efficacy of Prognostics and Health Management (PHM) methods is the acquisition and analysis of degradation data, which reflect the health condition of Engineering Systems (ESs) over time. Such degradation data provide crucial information on degradation processes, failure modes, and performance trends. However, the availability of such suitable data is a challenge, both in practice and in research (Zio, 2022). Therefore, publicly available degradation data sets are of high value. They enable the development and empirical benchmarking of PHM methods and can be used to demonstrate the practical applications of the introduced methods (Zöller et al., 2023; Ramasso and Saxena, 2014).

Although publicly available degradation data sets are of significant importance, the literature addresses this subject to a limited extent. Overviews often consider a restricted range of sources and platforms or only specific applications, resulting in constrained summaries (Soualhi et al., 2023; Su and Lee, 2024). Analyses usually emphasize the amount of data or data quality, overlooking specific PHM aspects (Jourdan et al., 2021). As a result, the search process for suitable data sets is often very time-consuming for users of PHM methods.

Therefore, the objective of this work is to provide a comprehensive overview of publicly available degradation data sets and conduct a novel, detailed, PHM-specific analysis. This PHM-specific analysis includes aspects that are relevant for users of PHM methods and data sets, respectively. These aspects include the application being considered in the data set, the domain from which the application originates, the task addressed by the data set, and the signal types included in the data set.

For this purpose, an existing overview of publicly available degradation data sets is revised and extended to a total number of 98 data sets, resulting in one of the most comprehensive overviews in the context of PHM so far. A taxonomy is developed based on the PHM-specific aspects in order to categorize and classify these data sets accordingly. Based on this taxonomy, the PHM-specific analysis of the data sets is performed.

The contributions of this work are as follows:

(1) An extended overview of publicly available degradation data sets[a].
(2) A taxonomy for the classification of these degradation data sets according to their applications, domains, PHM tasks, and included signal types.
(3) A PHM-specific analysis of these data sets focusing on relevant and practical aspects for users of PHM methods and data sets.

In the following, Section 2 introduces related work. Section 3 describes the revision and extension of the data set overview. This overview is the basis for the PHM-specific analysis, which is presented in Section 4. Section 5 concludes the results of this work.

2. Related Work

In PHM, a data set typically addresses a certain task regarding diagnostics and prognostics. Based on the works of Hagmeyer et al. (2021), Zio (2022), and Su and Lee (2024), these tasks can be subdivided as follows:

- **Fault detection / anomaly detection**: Detect a fault state or anomaly in an ES without considering its root cause. This results in a binary classification problem with the state fault or no fault.
- **Diagnosis**: Assign one or more causes to a detected fault state. This results in a classification with the number of classes depending on possible causes.
- **Health assessment**: Assess the state of degradation or the current risk of failure of an ES based on its current condition, regardless of whether an error has already occurred. The risk level is specified as a continuous value.
- **Prognosis**: Prediction of future degradation behavior or the current Remaining Useful Life (RUL) of an ES.

This list of tasks can also be interpreted as a sequential framework for implementing a comprehensive PHM application, starting with fault detection and ending with prognosis (Zio, 2022). The following example of rolling bearing failure illustrates this: The fault detection recognizes an abnormal vibration behavior. The diagnosis identifies the root cause, such as a specific bearing component. Health assessment evaluates the fault's impact on the bearing's health, while prognosis predicts the RUL of the bearing.

Publicly available data sets facilitate a wide range of the aforementioned research activities dealing with these tasks. Su and Lee (2024), Soualhi et al. (2023), Ochella et al. (2022), and Jia et al. (2018) provide overviews of such publicly available degradation data sets. However, their works are limited due to the consideration of only specific topics and the typical platforms and sources of degradation data sets, such as the PHM Data Challenges of the PHM Society and the Center of Excellence of the National Aeronautics and Space Administration (NASA). Significantly more platforms and sources for data sets are included in the overviews of the works of Hagmeyer et al. (2021) and Jourdan et al. (2021). In these overviews, the scope is extended to research institutions and comparable organizations dealing with the above mentioned tasks, as well as to artificial intelligence and machine learning dedicated data repositories. This includes, among others, the online community Kaggle, the University of California Irvine (UCI) machine learning repository, or GitHub. Taking into account this wide scope of platforms and sources, Hagmeyer et al. (2021) compiled a list of 70 publicly available degradation data sets, one of the most comprehensive overviews in the context of diagnostics and prognostics so far.

In the given overviews of publicly available data sets, aspects such as the amount of data, data quality, or data acquisition details are mainly analyzed (Su and Lee, 2024; Soualhi et al., 2023). However, no consistent PHM-specific analysis can be found in these overviews. The presence of such an analysis would assist users in identifying appropriate data sets more easily, thereby reducing the time-consuming nature of the search process.

[a]Due to the page limit, the complete data set overview can be found here: https://arxiv.org/abs/2403.13694v2

3. Publicly Available Data Sets

Due to regular data challenges and several research activities of the PHM community in recent years, new data sets are continuously being published. Therefore, the overview by Hagmeyer et al. (2021) is critically reviewed and updated using a similar procedure, i.e., the identical scope of platforms and sources of degradation data sets is considered. As a result, three data sets are removed as they no longer clearly address the tasks of diagnostics and prognostics, and 31 are being added. The overview is extended to 98 data sets in total. In accordance with Hagmeyer et al. (2021), data sets that focus solely on process quality without including the degradation state are excluded from consideration. In the following, all the added data sets are listed alphabetically:

(1) 4TU - Motor Current and Vibration Monitoring Dataset (Sietze Bruinsma et al., 2024)
(2) 4TU - Lifecycle ageing tests on commercial 18650 Li ion cell (Trad, 2021)
(3) Calce - Battery Data Repository (CALCE, 2025)
(4) GitHub - XJTU-SY Bearing Datasets (Wang et al., 2020)
(5) Kaggle - Bearings with Varying Degradation Behaviors (Mauthe et al., 2025)
(6) Kaggle - Condition Data with Random Recording Time (Mauthe et al., 2022)
(7) Kaggle - Prognosis based on Varying Data Quality (Mauthe et al., 2022)
(8) Mendeley - Battery Degradation Dataset (Fixed Current Profiles & Arbitrary Uses Profiles) (Lu, 2021)
(9) Mendeley - Data for: Accelerated Cycle Life Testing and Capacity Degradation Modeling of LiCoO2-graphite Cells (Diao, 2019)
(10) Mendeley - HUST Bearing (Nguyen Thuan and Hoang Si Hong, 2023)
(11) Mendeley - Long-term Dynamic Durability Test Dataset for Single Proton Exchange Membrane Fuel Cell (Zuo et al., 2021)
(12) Mendeley - NMC cell 2600 mAh cyclic aging data (Burzyński and Kasprzyk, 2021)
(13) Mendeley - Run-to-Failure Vibration Dataset of Self-Aligning Double-Row Ball Bearings (Gabrie lli et al., 2024)
(14) NASA - Accelerated Battery Life Testing (Fricke et al., 2023)
(15) NIST - Robot Arm Position Accuracy (NIST, 2017)
(16) Oxford - Oxford Battery Degradation Dataset (Birkl, 2017)
(17) PHM Data Challenge 2022 - Rock Drills (PHM Society Data Repository, 2022)
(18) PHM Data Challenge 2022 Europe - PCB Production line (Giordano and Trevisan, 2022)
(19) PHM Data Challenge 2023 - Gearbox (PHM Society Data Repository, 2023b)
(20) PHM Data Challenge 2023 Asia Pacific - Experimental Propulsion System (PHM Society Data Repository, 2023a)
(21) PHM Data Challenge 2024 - Helicopter Turbine Engines (PHM Society Data Repository, 2024)
(22) PHM IEEE Data Challenge 2023 - Planetary Gearbox (ICPHM 2023, 2023)
(23) PHM Society - Electromechanical Ball Screw Drive (PHM Society Data Repository, 2023c)
(24) Toyota Research Institute - Battery Cycle Life (Severson et al., 2019)
(25) UBFC - AMPERE Detection and diagnostics of rotor and stator faults in rotating machines (Soualhi et al., 2023c)
(26) UBFC - LASPI Detection and diagnostics of gearbox faults (Soualhi et al., 2023b)
(27) UBFC - METALLICADOUR Detection and diagnostics of multi-axis robot faults (Soualhi et al., 2023a)
(28) Zenodo - Ball bearings subjected to time-varying load and speed conditions (Javanmardi et al., 2024)
(29) Zenodo - Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation (Jiangong Zhu et al., 2022)
(30) Zenodo - Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning (Zhang et al., 2020)
(31) Zenodo - TBSI Sunwoda Battery Dataset (Shengyu, 2024)

In terms of amount, completeness, and descriptions, all newly added 31 data sets are of high quality. The extended overview, including all 98 data sets and the respective sources and platforms, is available on arXiv[b] (Mauthe et al., 2025). In the future, the authors intend to regularly review the overview and update it accordingly.

4. Data Set Analysis

Based on the updated and expanded overview of publicly available degradation data sets, a detailed analysis is carried out. This analysis is intended to shorten the time-consuming search for suitable data sets for its users. The focus of this analysis

[b] Data set overview: https://arxiv.org/abs/2403.13694v2

is on PHM-specific aspects. As a remark, detailed considerations and analyses of individual data sets in terms of, e.g., data volume or data acquisition are beyond the scope of this work. In the following, the PHM-specific aspects are defined and applied to develop a taxonomy. This taxonomy is used to classify the data sets. The analysis is then carried out based on these classified data sets.

4.1. *Taxonomy for degradation data sets*

In the context of PHM, users consider specific crucial aspects when selecting the appropriate degradation data sets. Typically, users look for data sets that align with their particular requirements and objectives. As a result, the following PHM-specific aspects are defined:

- **Application**: The application within a data set refers to the object of consideration from which the data originate. Thus, an application can represent a single component, such as a bearing, or an entire system, such as a production line.
- **Domain**: The domain from which the application originates. The domains associated with the previous examples would then be, for example, mechanical components or production systems.
- **Task**: The task, which is addressed by the corresponding data set. A distinction is made among the tasks described in Section 2: fault detection, diagnosis, health assessment, and prognosis.
- **Signal**: The types of signals included in the data set. This refers to signals that offer insight into the health state of the specific application or can be derived from it and used within PHM methods.

These PHM-specific aspects serve as the basis for the creation of a taxonomy that is tailored to the classification of the data sets in a PHM-specific manner. This taxonomy categorizes all data sets based on their application and domain of origin, the task addressed by the data set, and the relevant signals contained in the data set.

The results using this taxonomy for the 98 publicly available data sets are provided in Table 1 in Section 4.4 and in Table 2 in Appendix A. The respective assignment of the PHM-specific aspects (application, domain, task, and signal) to the individual data sets can be found in the overall overview of the data sets on arXiv[c] (Mauthe et al., 2025).

4.2. *Domains and application*

Applying the presented taxonomy, the 98 publicly available degradation data sets are classified into 11 domains (see Table 2). The electrical component domain contains with 23 the most data sets, followed by the mechanical component domain with 22 and the drive technology domain with 11 assigned data sets. Except for the domain building, the remaining domains in Table 2 are assigned between at least four and a maximum of eight data sets. Moreover, there are five data sets with unknown applications. As the respective data sets can be used to apply PHM methods, they are taken into account and assigned to the unknown domain.

In total, 38 different applications are included in the 11 domains, covering a variety of different applications and degradation processes, respectively. However, as these applications are spread across 98 data sets, the number of data sets available per application is limited. This can be seen in a more in-depth consideration of the applications in Table 2. With 15 data sets each, battery and bearing are the applications most frequently considered. Both originate from the two main domains electrical component and mechanical component. The application with the second most data sets is the production line, which occurs only five times. The problem of limited data sets per application is highlighted by looking at the domain mechatronic system, which contains the most different applications (seven), with only one data set for each application present. Overall, only for seven applications are three or more data sets available.

4.3. *Considered tasks within data sets*

The diagnostic and prognostic tasks of the data sets[d], as introduced in Section 2, are shown in the

[c]Data set overview: https://arxiv.org/abs/2403.13694v2

[d]Certain data sets may be assigned to several tasks. As these tasks build on each other, the highest-ranking task is assigned to a respective data set. Therefore, for each data set, only one task is assigned.

third column of Table 2. With 46 assigned data sets, the prognosis task is the most represented, followed by the diagnosis with 33 entries, whereas fault detection is assigned to 16. The transition from the task of diagnosis to the task of health assessment, as well as the transition to the task of prognosis, are often not distinct steps. For this reason, there are only three data sets that explicitly address the task of health assessment.

Nearly half of the 46 data sets for prognosis belong to the electrical component domain, with 20 data sets overall, 14 of which consider battery applications. The drive technology and mechanical component domain contain the second highest number of data sets, with six each. Bearing is the second most common application for prognosis (six). Aircraft engine and filtration each have four data sets, making them the third most frequent applications. In total, 16 distinct applications are available for prognosis.

The mechanical component domain contains the most data sets for diagnosis, with 12 out of 33 data sets. The robotic and mechatronic system domains contain five data sets each, while the production system domain contains three data sets. The bearing application is the most frequently used for diagnosis with seven data sets, followed by articulated robot applications with four and both gear and production line applications with three data sets each. It is noteworthy that 20 different applications are considered across the 33 data sets for the diagnosis task, resulting in a high variety of applications.

For the fault detection task, the domains of mechanical component and production system are the most extensive, with four data sets each. With regard to the applications considered, there is a large variety, as only two applications, bearing and production line, are considered in two data sets each. In total, the 16 available data sets are distributed between 12 different applications.

4.4. *Signals within data sets*

Table 1 summarizes the 19 signals used in at least two of the given data sets. Note that Table 1 only lists signals that are measured directly. Values calculated from these, such as capacity or power, are excluded. Nevertheless, if only calculated signals are included in a data set, they are still listed. Also summarized are signals of a comparable nature, such as speed, including rotation and velocity.

Temperature appears most frequently, occurring in 32 data sets, followed by vibration with 31 data sets, and electrical signals (current and voltage) with 29 and 23 data sets, respectively. These signals are typically measured parameters within electrical and mechanical component domains, as well as in battery and bearing applications. Therefore, they are most prevalent in the analyzed data sets. Speed, pressure, and anonymized

Table 1. Signals used in at least two of the given data sets. Sorted by the sum of signal occurrences across the tasks of fault detection (FD), diagnosis (D), health assessment (HA), and prognosis (P).

Signal	**FD**	**D**	**HA**	**P**	**Sum across tasks**
Temperature	4	8	0	20	32
Vibration	4	15	0	12	31
Current	1	9	0	19	29
Voltage	2	4	0	17	23
Speed	4	11	0	4	19
Pressure	0	5	1	10	16
Anonymized(1)	3	1	2	4	10
Flow rate	0	1	0	8	9
Torque	2	5	0	1	8
Operating condition(2)	0	1	0	6	7
Position	2	4	0	1	7
Acoustic emission	0	2	0	2	4
Force	0	3	0	1	4
Unknown(3)	2	0	0	1	3
Acceleration(4)	1	1	0	0	2
Capacity	0	0	0	2	2
Humidity	0	1	0	1	2
Inspection data	0	2	0	0	2
Power	1	1	0	0	2

(1): signal values changed/transformed, (2): varying operating conditions during the life, (3): signal (or sensor) type unknown (4): in terms of movement, not vibration signals are also present in 19, 16, and ten data sets, respectively. However, anonymized signals permit specific observations only within a data set or application, limiting their utility for general analysis. With regard to the availability of data sets per signal, there is a limitation similar to the applications. Of the 19 signals listed, eight are contained only in four or fewer data sets.

In relation to the task, the signal occurrences are similar. Importantly, temperature and current signals are also included in some mechanical component data sets, particularly in bearings and gears; therefore they are crucial in diagnosis tasks.

5. Conclusion

The increase to 98 data sets reflects the dynamics of development in the PHM research field on the one hand and the need for a PHM-specific analysis of the degradation data sets required for its development on the other. Therefore, the taxonomy developed and the PHM-specific analysis carried out in this work serve as an entry point when searching for suitable data sets. This analysis assists users in identifying appropriate data sets more easily, thereby reducing the time-consuming search process. For example, when searching for a data set related to a prognosis task that incorporates a vibration signal, a user can refer to the provided analysis to identify appropriate data sets.

The analysis shows that almost half of the data sets (45 out of 98) are from the domains of electrical and mechanical components, with battery and bearing applications being the most common, highlighting the limited number of data sets for other applications. The prognosis task is gaining popularity, as reflected in the growing number of recent battery data sets. Diagnosis and fault detection mainly focus on mechanical components. While bearing applications offer a balanced task distribution and sufficient data sets, battery applications dominate in prognosis data sets. The types of signals, such as vibration, current, temperature, and voltage, are typical measured parameters for battery and bearing applications, or within the respective domain. Therefore, these four signals are also the most common in the analyzed data sets.

During the analysis of the 98 publicly available data sets, it became evident that the newer data sets, particularly the 31 newly added, are well-documented with all essential aspects identifiable. In contrast, older data sets sometimes lack adequate documentation. Clear and thorough documentation is essential for the optimal usability of the data sets.

The presented analysis facilitates further investigation of the data sets. For example, data sets can be specifically analyzed and categorized depending on respective tasks or applications. In the future, the authors intend to carry out such an extended investigation.

Acknowledgement

The research findings presented in this paper are funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant no. 514247199 and Project-ID 528745080 – FIP 68.

References

Birkl, C. (2017). Oxford battery degradation dataset 1.

Burzyński, D. and L. Kasprzyk (2021). Nmc cell 2600 mah cyclic aging data.

CALCE (18.01.2025). Battery data — center for advanced life cycle engineering.

Diao, W. (2019). Data for: Accelerated cycle life testing and capacity degradation modeling of licoo2-graphite cells.

Fricke, K., R. Nascimento, M. Corbetta, C. Kulkarni, and F. Viana (2023). Prognosis of li-ion batteries under large load variations using hybrid physics-informed neural networks. *Annual Conference of the PHM Society 15*(1), 1–12.

Gabrie lli, A., M. Battarra, E. Mucchi, and G. Dalpiaz (2024). Physics-based prognostics of rolling-element bearings: The equivalent damaged volume algorithm. *Mechanical Systems and Signal Processing 215*, 111435.

Giordano, D. and M. Trevisan (2022). Github - phme-datachallenge/data-challenge-2022.

Hagmeyer, S., F. Mauthe, and P. Zeiler (2021). Creation of publicly available data sets for prognostics and diagnostics addressing data scenarios relevant to industrial applications. *IJPHM 12*(2), 1–20.

ICPHM 2023 (2023). Icphm 2023.

Javanmardi, A., O. K. Aimiyekagbon, A. Bender, J. K. Kimotho, W. Sextro, and E. Hüllermeier (2024). Remaining useful lifetime estimation of bearings operating under time-varying conditions. *PHM Society European Conference 8*(1), 9.

Jia, X., B. Huang, J. Feng, H. Cai, and J. Lee (2018). Review of phm data competitions from 2008 to 2017. *Annual Conference of the PHM Society 10*(1), 1–10.

Jiangong Zhu, Yixiu Wang, Yuan Huang, R. Bhushan Gopaluni, Yankai Cao, Michael Heere, Martin J. Mühlbauer, Liuda Mereacre, Haifeng Dai, Xinhua Liu, Anatoliy Senyshyn, Xuezhe Wei, Michael Knapp, and Helmut Ehrenberg (2022). Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation.

Jourdan, N., L. Longard, T. Biegel, and J. Metternich (2021). Machine learning for intelligent maintenance and quality control: A review of existing datasets and corresponding use cases. In D. Herberger and M. Hübner (Eds.), *CPSL 2021*. Hannover:publish-Ing.

Lu, J. (2021). Battery degradation dataset (fixed current profiles & arbitrary uses profiles).

Mauthe, F., E. M. Bakir, T. Scheerer, and P. Zeiler (2022). Prognosis based on varying data quality.

Mauthe, F., C. Braun, J. Raible, P. Zeiler, and M. F. Huber (2025). Overview of publicly available degradation data sets for tasks within prognostics and health management. *arXiv 2403.13694v2*, 1–14.

Mauthe, F., S. Hagmeyer, and P. Zeiler (2025). Bearings with varying degradation behaviors.

Mauthe, F., J. Schmid, and P. Zeiler (2022). Condition data with random recording time.

Nguyen Thuan and Hoang Si Hong (2023). Hust bearing: a practical dataset for ball bearing fault diagnosis.

NIST (2017). Degradation measurement of robot arm position accuracy — nist.

Ochella, S., M. Shafiee, and F. Dinmohammadi (2022). Artificial intelligence in prognostics and health management of engineering systems. *Engineering Applications of Artificial Intelligence 108*, 104552.

PHM Society Data Repository (09.08.2024). Phm north america 2024 conference data challenge - phm society data repository.

PHM Society Data Repository (2022). 2022 phm conference data challenge - phm society data repository.

PHM Society Data Repository (2023a). Phm asia-pacific 2023 conference data challenge - phm society data repository.

PHM Society Data Repository (2023b). Phm north america 2023 conference data challenge - phm society data repository.

PHM Society Data Repository (2023c). Servomotor-driven ballscrew mechanism degradation data set - phm society data repository.

Ramasso, E. and A. Saxena (2014). Performance benchmarking and analysis of prognostic methods for cmapss datasets. *IJPHM 5*(2), 1–15.

Severson, K. A., P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, and R. D. Braatz (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy 4*(5), 383–391.

Shengyu, T. (2024). Tbsi sunwoda battery dataset.

Sietze Bruinsma, Rinze Geertsma, Richard Loendersloot, and Tiedo Tinga (2024). Nln-emp - motor current and vibration monitoring dataset for various faults in an e-motor-driven centrifugal pump.

Soualhi, M., A. Soualhi, K. T. P. Nguyen, K. Medjaher, C. Guy, and R. Hubert (2023). Open heterogeneous data for condition monitoring of multi faults in rotating machines used in different operating conditions. *IJPHM 14*(2), 1–16.

Soualhi, M., A. Soualhi, T.-P. K. Nguyen, K. Medjaher, G. Clerc, and H. Razik (2023a). Ampere: Detection and diagnostics of rotor and stator faults in rotating machines.

Soualhi, M., A. Soualhi, T.-P. K. Nguyen, K. Medjaher, G. Clerc, and H. Razik (2023b). Laspi: Detection and diagnostics of gearbox faults.

Soualhi, M., A. Soualhi, T.-P. K. Nguyen, K. Medjaher, G. Clerc, and H. Razik (2023c). Metallicadour: Detection and diagnostics of multi-axis robot faults.

Su, H. and J. Lee (2024). Machine learning approaches for diagnostics and prognostics of industrial systems using open source data from phm data challenges. *IJPHM 15*(2), 1–26.

Trad, K. (2021). Lifecycle ageing tests on commercial 18650 li ion cell @ 25°c and 45°c.

Wang, B., Y. Lei, N. Li, and N. Li (2020). A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE Transactions on Reliability 69*(1), 401–412.

Zhang, Y., Q. Tang, Y. Zhang, J. Wang, U. Stimming, and A. A. Lee (2020). Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning.

Zio, E. (2022). Prognostics and health management (phm): Where are we and where do we (need to) go in theory and practice. *Reliability Engineering & System Safety 218*, 108119.

Zöller, M.-A., F. Mauthe, P. Zeiler, M. Lindauer, and M. F. Huber (2023). Automated machine learning for remaining useful life predictions. In *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2907–2912. IEEE.

Zuo, J., H. Lv, D. Zhou, Q. Xue, L. Jin, W. Zhou, D. Yang, and C. Zhang (2021). Long-term dynamic durability test datasets for single proton exchange membrane fuel cell. *Data in brief 35*, 106775.

Appendix A. Classification of Publicly Available Data Sets

Table 2. Classification of the 98 publicly available data sets based on their domain, the respective application, and the addressed PHM task (fault detection (FD), diagnosis (D), health assessment (HA), and prognosis (P)). An application marked with an asterisk (*) entails only simulated data sets and additionally the respective amount (n) per task ($*^{n \cdot FD/D/HA/P}$) if it entails partly simulated data sets.

Domain	Application	FD	D	HA	P	Sum across domain
Building	Building		1			1
Drive technology	Aircraft engine*				4	11
	Electric motor	1	1		1	
	Diesel engine		1			
	Helicopter engine	1				
	Propulsion system*			1	1	
Electrical component	Battery		1		14	23
	Capacitor				2	
	Fuel cell				2	
	Sensor	1	1			
	Transistor				2	
Manufacturing process	Planarization system				1	6
	Drilling		1			
	Electrophoresis painting				1	
	Milling	1			2	
Material	Aluminum plate				2	4
	Polymer composite				1	
	Steel plate		1			
Mechanical component	Anemometer	1				22
	Bearing$*^{(1 \cdot P)}$	2	7		6	
	Bogie		1			
	Gear	1	3			
	Shaft		1			
Mechatronic system	Air compressor		1			7
	Air pressure system	1				
	Electromechanical ball screw*		1			
	Electromechanical device		1			
	Elevator				1	
	Hydraulic system		1			
	Rock drill		1			
Process technology	Filtration				4	6
	Pump	1	1			
Production system	Ion mill etching tool				1	8
	Log data	1				
	Production line	2	3			
	Shrink-wrapper	1				
Robotic	Articulated robot		4			5
	Linear robot		1			
Unknown	—	2		2	1	5
all		16	33	3	46	**98**