



HAL
open science

Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice

Enrico Zio

► To cite this version:

Enrico Zio. Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice. *Reliability Engineering and System Safety*, 2022, 218, pp.108119. 10.1016/j.ress.2021.108119 . hal-03907690

HAL Id: hal-03907690

<https://minesparis-psl.hal.science/hal-03907690v1>

Submitted on 5 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Prognostics and Health Management (PHM):**
2 **where are we and where do we (need to) go in theory and practice**
3 Enrico Zio

4
5
6 *MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France enrico.zio@minesparitech.fr*
7 *Energy Department, Politecnico di Milano, Milan, Italy, enrico.zio@polimi.it*

8
9 **ABSTRACT**

10
11 We are performing the digital transition of industry, living the 4th industrial
12 revolution, building a new World in which the digital, physical and human dimensions are
13 interrelated in complex socio-cyber-physical systems. For the sustainability of these
14 transformations, knowledge, information and data must be integrated within model-based and
15 data-driven approaches of Prognostics and Health Management (PHM) for the assessment and
16 prediction of structures, systems and components (SSCs) evolutions and process behaviors, so
17 as to allow anticipating failures and avoiding accidents, thus, aiming at improved safe and
18 reliable design, operation and maintenance.

19 There is already a plethora of methods available for many potential applications and
20 more are being developed: yet, there are still a number of critical problems which impede full
21 deployment of PHM and its benefits in practice. In this respect, this paper does not aim at
22 providing a survey of existing works for an introduction to PHM nor at providing new tools or
23 methods for its further development; rather, it aims at pointing out main challenges and
24 directions of advancements, for full deployment of condition-based and predictive
25 maintenance in practice.

26 **Keywords:** *Prognostics and Health Management (PHM), predictive maintenance, Recurrent*
27 *Neural Networks (RNNs), Reservoir Computing (RC), Generative Adversarial Networks*
28 *(GANs), Deep Neural Networks (DNNs), Optimal Transport (OT)*

29
30 **NOMENCLATURE**

AAKR	Auto-Associative Kernel Regression
AANN	Auto-Associative Neural Networks
ADNN	Adjacency Difference Neural Network
AE	Auto-Encoder
AE-	
GAN	Auto-Encoder aided GAN
ALE	Accumulated Local Effect
ANNs	Artificial Neural Networks
ARM	Association Rule Mining
ARMA	Auto-Regressive Moving Average
BN	Bayesian Network
CatAAE	Categorical Adversarial Autoencoder
CBM	Condition-Based Maintenance
CDT	Cumulative Distribution Transform
CNN	Convolutional Neural Network
CVNN	Complex Valued Neural Network
DAE	Denosing Auto Encoder

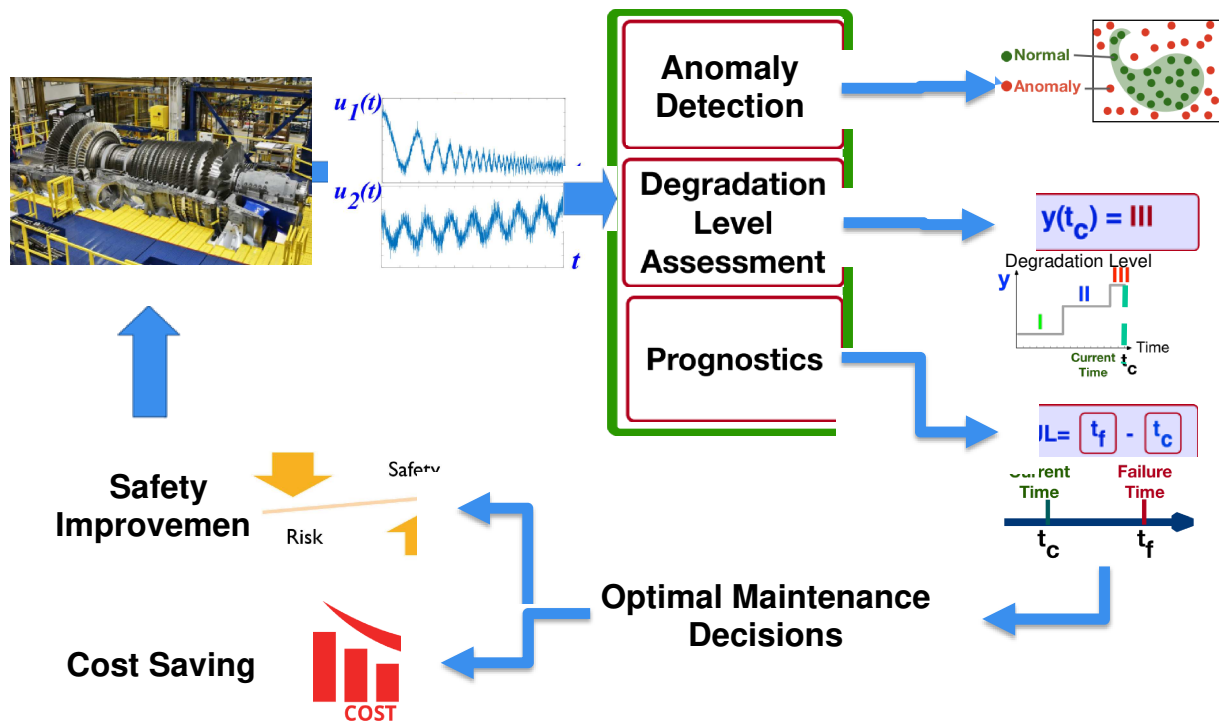
DBN	Deep Belief Network
DL	Deep Learning
DNNs	Deep Neural Networks
DT	Decision Trees
ELM	Extreme Learning Machine
EM	Expectation Maximization
EMD	Earth Mover's distance
ESNs	Echo-State Networks
FCM	Fuzzy C-Means
FFT	Fast Fourier Transform
GANs	Generative Adversarial Networks
GLRT	Generalized Likelihood Ratio Test
GRU	Gated Recurrent Unit
HI	Health Indicator
HMM	Hidden Markov Model
ICE	Individual Conditional Expectation
ICT	Information and Communication Technology
IoTs	Internet of Things
KD	Kantorovich distance
KF	Kalman Filtering
KNN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
LIME	Local Interpretable Model Explanation
LS	Least Square
LSTM	Long Short Term Memory
MAR	Missing At Random
ML	Machine Learning
MODE	Multi-Objective Differential Evolution
NPPs	Nuclear Power Plants
OC-SVM	One Class-Support Vector Machine
OT	Optimal Transport
OTT	Optimal Transport Theory
PCA	Principle Component Analysis
PDP	Partial Dependence Plot
PF	Particle Filtering
PFSA	Probabilistic Finite State Automation
PHM	Prognostics and Health Management
PPIs	Prognostic Performance Indicators
RC	Reservoir Computing
RF	Random Forest
RNNs	Recurrent Neural Networks
RUL	Remaining Useful Life
RVM	Relevance Vector Machine
SA	Sensitivity Analysis

- SaNSDE Self-adaptive Differential Evolution with Neighborhood Search
- SC Spectral Clustering
- SOM Self-Organizing Map
- SPRT Sequential Probability Ratio Test
- SSCs Structures, Systems and Components
- STPN Spatio-Temporal Pattern Network
- SVM Support Vector Machine
- TOPSIS Technique for Order of Preference by Similarity to Ideal Solution

31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

1. INTRODUCTION

Prognostics and Health Management (PHM) is a computation-based paradigm that elaborates on physical knowledge, information and data [1] of structures, systems and components (SSCs) operation and maintenance, to enable detecting equipment and process anomalies, diagnosing degradation states and faults, predicting the evolution of degradation to failure so as to estimate the remaining useful life (Figure 1). The outcomes of the PHM elaboration are used to support condition-based and predictive maintenance decisions for the efficient, reliable and safe operations of SSCs [2]–[5]. In fact, the capability of deploying these maintenance strategies provides the opportunity of setting efficient, just-in-time and just-right maintenance strategies: in other words, providing the right part to the right place at the right time. This opportunity is big because doing this would maximize the production profits and minimize all costs and losses, including asset ones [6]. As a result, in the past decade PHM research and development has intensified, both in academia and industry, involving various disciplines of mathematics, computer science, operations research, physics, chemistry, materials science, engineering, etc. [7], [8].



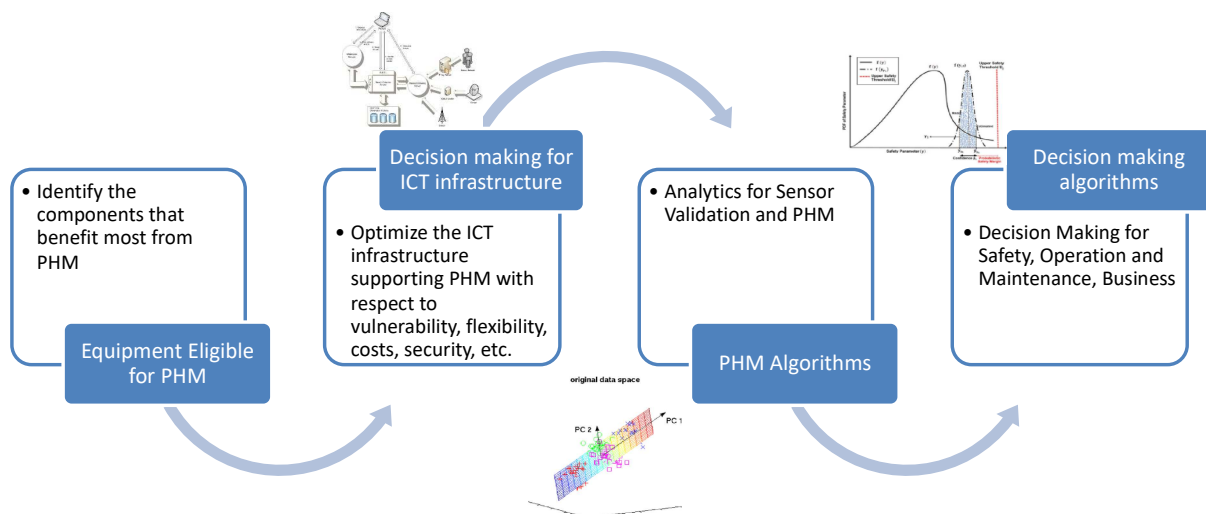
50
51
52
53

Figure 1. PHM tasks. The data collected from industrial component sensors feeds three major PHM tasks: fault detection (anomaly detection), fault diagnostics (degradation

level assessment) and fault prognostics (remaining useful life prediction). The successful deployment of PHM provides solid foundations for the optimal maintenance decisions, and thus improve the safety of industrial SSCs while reducing cost.

For making reliability and safety decision using PHM outcomes in practice, identifying, understanding and quantifying the impacts and benefits that the development of a PHM system can have on the health management of a SSC is necessary (e.g. avoid unexpected catastrophic failures, reduce maintenance frequency, optimize spare parts and storage, optimize resources, etc.). Then, the practical implementation of PHM includes data acquisition to enable detection, diagnostics and prognostics tasks, and maintenance decision making [9] (Figure 1). The supporting PHM development framework (Figure 2) and its requirements must, then, be properly defined to perform well in real industrial scenarios [9]–[11]. Given the increasing complexity, integration and informatization of modern engineering SSCs, PHM can no longer be an isolated addition for supporting maintenance but must be closely linked to the other structure, power, electromechanical, information and communication technology (ICT), control parts of the systems. Then, PHM must be included at the beginning of the system conceptualization, and carried through its design and development in an integrated framework capable of satisfying the overall operation and performance requirements [12], [13].

Finally, for the use of PHM in practice, the question of which methods to use is fundamental. For example, referring, in particular, to the prognostic task of PHM, the prediction capability of a prognostic method refers to its ability to provide trustable predictions of the Remaining Useful Life (RUL), with the quality characteristics and confidence level required for making decisions based on such predictions. Indeed, this heavily influences the decision makers' attitude toward taking the risk of using the predicted RUL outcomes to inform their decisions [14]. The choice of which method to use is typically driven by the data available and/or the physics-based models available, and the cost-benefit considerations related to the implementation of the PHM system. A set of Prognostic Performance Indicators (PPIs) must be used to guide the choice of the approach to be implemented, within a structured framework of evaluation. These PPIs measure different characteristics of a prognostic approach and need to be aggregated to enable a final choice of prognostic method, based on its overall performance [15]. For this reason, various performance metrics have been defined to enable the evaluation of the performance of PHM methods [16]. These metrics are needed to guide the PHM system development (Figure 2).



89
90

91 *Figure 2 PHM development framework for informed decision-making*
92

93 Up to now, for the maturation of PHM, the main efforts have been mainly devoted to
94 the development of hardware (i.e., Internet of Things (IoTs), smart meters, etc. [17]–[20] and
95 software for tracking the health state of monitored equipment (e.g., data analytics, platforms
96 for IoT interconnection and clouding for computing, etc. [21]–[23]). On the other hand, the
97 full deployment of PHM in practice involves other aspects, including design (e.g. the use of
98 smart components may lead to different reliability allocation solutions), and impacts various
99 work units involved in maintenance decisions and actuations (e.g., workers can use smart
100 systems, maintenance engineers can analyze big data), including the supporting logistics
101 (spare parts availability and warehouse management can be driven by the PHM results) [17].

102 In this paper, we present some main challenges for the development of PHM in
103 practice, corroborated by practical examples, and associate to some of them the developments
104 of Recurrent Neural Networks (RNNs), Reservoir Computing (RC), Generative Adversarial
105 Networks (GANs), Deep Neural Networks (DNNs), Optimal Transport Theory (OTT), as
106 potential directions to successfully address them.
107

108 **2. CHALLENGES TO PHM IN PRACTICE** 109

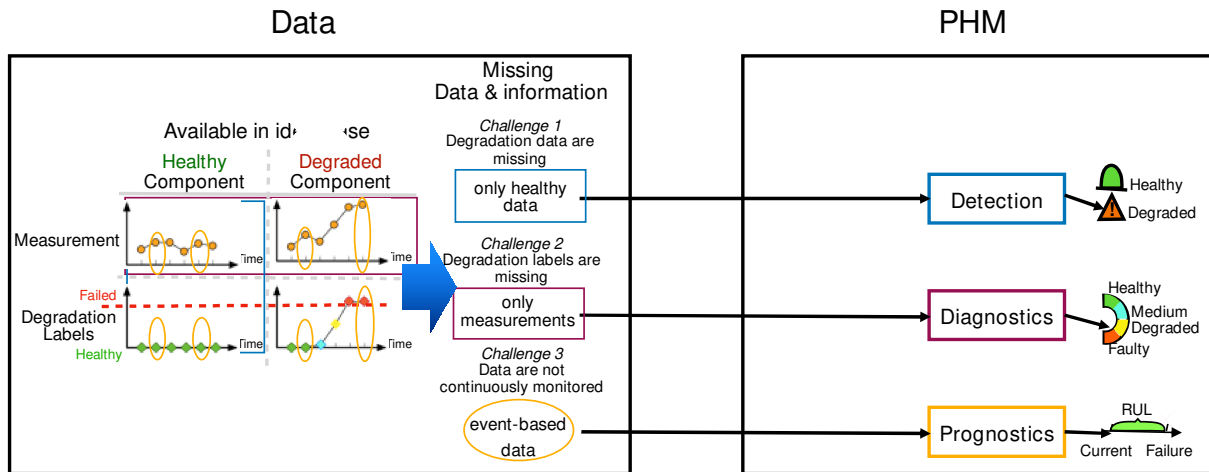
110
111 Main challenges to the deployment of PHM in practice still remain, coming from
112 different sides:

- 113 • the physics of the problem
- 114 • the data available
- 115 • the requirements of the solutions.

116 The challenges related to the physics of the problem derive from the complexity of the
117 SSCs degradation processes, which are not completely known, dynamic and highly non-
118 linear, and hence their understanding, characterization and modelling are difficult.

119 The challenges related to the data relate to multiple aspects (Figure 3):

- 120 • the many anomalies in the real data collected in the field (including missing
121 data and erroneous data from malfunctioning sensors)
- 122 • the scarcity and incompleteness of data recognizably related to the state of
123 degradation of the SSC of interest (labelled patterns)
- 124 • the difficulty of managing and treating big data, with a large variety of signals
125 collected by sensors of different types
- 126 • the changing operational and environmental conditions which affect the data
127 used to train the PHM models and calibrate their parameters, and on which the
128 models are applied.
129



130
131
132 *Figure 3 PHM challenges from data. Fault detection is affected by the challenge of*
133 *missing data and erroneous data from malfunctioning sensors; fault diagnostics is affected by*
134 *the challenge of missing labels relating to the state of degradation of the SSCs; fault*
135 *prognostics is affected by the challenge that data are collected in event-based scenarios*
136 *because of the difficulty of managing and treating big data.*
137

138 The challenges related to the requirements of the PHM solutions come from the
139 multiple objectives that they must achieve, depending on the applications. The obvious ones
140 are accuracy and precision, quantified with defined performance indicators and measured
141 against the decisions that they support: in some cases, very high accuracy and precision is
142 required to be able to take confident decisions (e.g. of stopping a system upon an alert of fault
143 detection, of replacing a component upon a fault diagnosis, of anticipating or postponing a
144 scheduled maintenance based on accurate remaining useful life predictions); in other cases,
145 accuracy and precision need not be so high, and may be compromised for other objectives.
146 For example, transparency, explainability and interpretability of PHM models are attributes of
147 particular interest, if not demanded, for decision making in safety-critical applications, for
148 which they may also be a regulatory prerequisite. Also, PHM as a data-dependent enabling
149 technology for smart condition-based and predictive maintenance has issues regarding
150 security. Indeed, the technological network supporting PHM is made of devices,
151 communication technologies and various protocols, so that security issues regarding
152 availability, data integrity, data confidentiality and authentication exist. As these issues
153 hamper operational efficiency, robustness and throughput, they must be adequately addressed.

154 Finally, an enveloping challenge to the deployment of PHM in practice comes from
155 the fact that the PHM tasks of fault detection, diagnostic and prognostic are inevitably
156 affected by various sources of uncertainty, such as incomplete knowledge on the present state
157 of the equipment, randomness in the future operational usage profile and future evolution of
158 the degradation of the equipment, inaccuracy of the PHM model and uncertainty in the values
159 of the signal measurements used by the PHM model to elaborate its outcomes, etc. Therefore,
160 any outcome of a PHM model should be accompanied by an estimate of its uncertainty, in
161 order to confidently take robust decisions based on such outcome, considering the degree of
162 mismatch between the PHM model outcomes and the real values.

163 As these issues hamper operational inefficiency, robustness and throughput, they must
164 be adequately addressed.

165 With specific reference to data-driven methods and models for the tasks of fault
166 detection, fault diagnostics and failure prognostics in PHM, the next section addresses some

167 of the above challenges with the focus on advanced methods that are proving as promising for
168 their solution.

169

170 **3. ADVANCING METHODS OF FAULT DETECTION, FAULT DIAGNOSTICS** 171 **AND FAILURE PROGNOSTICS FOR MEETING THE CHALLENGES OF** 172 **PHM IN PRACTICE**

173

174 Methods of fault detection, fault diagnostics and failure prognostics within the PHM
175 framework are continuously being developed and advanced, and applications to various SSCs
176 are being deployed, supported by the technology of sensors and monitoring systems, the
177 techniques of data analytics, image processing and text mining, mostly based on the Artificial
178 Intelligence (AI) and Machine Learning (ML) paradigms, and the computational power [24].
179 The objective of fault detection is to recognize abnormalities/anomalies in SSCs behavior. The
180 objective of fault diagnostics is to identify the SSCs degradation states and the causes of
181 degradation. Prognostics aims at predicting the SSCs Remaining Useful Life (RUL), i.e. the
182 time left before it will no longer be able to perform its intended function. Fault detection and
183 diagnostics, and failure prognostics are the enablers of condition-based and predictive
184 maintenance, which offers major opportunities for Industry 4.0 and smart SSCs, as they can
185 allow reducing failures, increasing SSCs usage, and reducing operation and maintenance
186 costs, with tangible benefits of reduction of production downtime, risk and asset losses, and
187 consequent increase of production profit [24].

188 A number of challenges still remain, arising from the complexity of the physics which
189 PHM is addressed to in practice, from the data available and from the requirements to the
190 PHM solutions for practical applications. In this Section, we go through some of these
191 challenges, to see where we stand, and where we are going and need to go.

192

193 *3.1 Data Challenges*

194

194 *3.1.1 Fault detection*

195 As mentioned above, fault detection is the PHM task which aims at identifying the
196 presence of abnormalities/anomalies during the operation of a SSC. While such
197 abnormalities/anomalies are commonly referred to as faults in certain disciplines, such as
198 energy and mechanical engineering, the term damage is commonly used in some other
199 disciplines such as structural engineering. In practical applications, fault/damage detection is
200 challenging because it is necessary to assess the presence of the fault/damage based on signals
201 of physical variables measured during the SSC operation and such process is complicated by
202 the various sources of uncertainty that can render the signal processing extremely difficult.

203 Fault detection methods are classified as model-based and data-driven [25]. Model-
204 based methods use first principles and physical laws to describe the physical phenomena and
205 processes of interest [26][27][28]. For example, [26] builds a model of the behavior of a rotor
206 using the finite element method and successfully applies it to fault detection. [27] introduces a
207 model-based fault detection and isolation technique for manufacturing machinery based on a
208 defined relationship between a fault signal and observer theory. [28] presents a two-level
209 Bayesian approach based on the use of Hidden Markov Model (HMM) and Expectation
210 Maximization (EM) to detect early faults in a milling machine. However, the practical
211 application of model-based methods is limited by the difficulty of developing accurate
212 mathematical models of the processes and behaviors of complex modern SSCs [29].

213 For this reason, data-driven fault detection methods are more popular than model-
214 based ones, as they rely only on data for the recognition of anomalous patterns attributable to
215 faults [30][31][32][33][34][35]. For example, [30] develops a fault detection method for
216 power generation systems, by combining Principle Component Analysis (PCA) for feature

217 extraction and Random Forest (RF) for fault behavior pattern learning. Support Vector
218 Machine (SVM) techniques are introduced to detect faults considering concept drift in nuclear
219 power plants [31], and to detect faults in high speed train brake systems in case of highly
220 imbalanced data [35]. Neural Network based approaches attract attention in fault detection,
221 e.g. [32] combines a set of Artificial Neural Networks (ANNs) through Bayesian statistics for
222 heavy-water nuclear reactor fault detection and uncertainty quantification, [34] uses ANN to
223 detect false alarms in wind turbines for reliability centered maintenance, [33] introduces a
224 Recurrent Neural Network (RNN) with optimized hyperparameters for the detection of
225 software faults.

226 These methods can be divided in those which rely on one-class classification models
227 and those which use residuals, i.e., the differences between the real measurements and the
228 reconstructed values of the signals in normal conditions, to identify the normal/abnormal
229 conditions [36].

230 The former require training of a one-class classification model on signal
231 measurements collected from both normal (healthy) and abnormal/anomalous (faulty)
232 conditions of SSCs. However, in practical applications, faults are rare and the data have
233 manifold distributions embedded in high-dimensional spaces. Distributions with non-smooth
234 densities and the curse of dimensionality of the data in the long-term multivariate time series
235 collected from sensors on real industrial SSCs, can cause model overfitting and render
236 difficult the empirical reconstruction of the data distribution, which, therefore, leads to
237 unsuccessful detection of abnormal/anomalous conditions in SSCs behavior. These technical
238 issues hamper the successful deployment of one-class classification methods for fault
239 detection in practical applications. The need is, then, to develop methods able to detect
240 anomalous (faulty) conditions given data in normal conditions, and to deal with the manifold
241 distribution and large dimensionality of real data. In this direction, Generative Adversarial
242 Networks (GANs) are an interesting perspective as they can be used to reproduce complex
243 distributions, e.g. manifolds [37], [38]. An example is given in the work by [39], which
244 proposes an Auto-Encoder aided GAN (AE-GAN) model for the detection of
245 abnormal/anomalous conditions in the behavior of a SSC, in which the generator of the GAN
246 and an auxiliary encoder form an AE module, and the reconstruction error generated by the
247 AE is used as score to detect abnormalities/anomalies in the SSC behavior. Adaptive noise is
248 added on the data and AdaBoost ensemble learning is adapted to integrate the AE-GANs
249 applied to detect anomalies in each small time slice of the long-term multivariate time series
250 collected by the sensors [40]. Furthermore, this work derives a lower bound of Jensen-
251 Shannon divergence between generator distribution and normal data distribution as an
252 objective to optimize the AE-GANs hyperparameters; by probing, the optimization works
253 without test data, as commonly needed by other methods. Extensive experiments are
254 conducted on real industrial datasets to demonstrate the usefulness of the developed Adaboost
255 ensembled AE-GAN method for abnormality/anomaly detection in practice.

256 Residual-based fault detection methods rely on the use of normal-conditions (healthy)
257 data, only [41]. These methods reconstruct the values of the signals expected in normal
258 conditions and use the residuals, i.e., the differences between the real measurements and the
259 reconstructed signals, to identify the normal/abnormal conditions. Examples of residual-based
260 methods include Auto-Associative Kernel Regression (AAKR) [42]–[44], Principal
261 Component Analysis (PCA) [45], One Class-Support Vector Machine (OC-SVM) [46], and
262 Artificial Neural Networks (ANNs) [47]. The empirical model, fitted to the data so as to
263 provide accurate signal reconstructions, plays an essential role in the above procedure.
264 However, its training may require a large amount of healthy data collected under various
265 operating conditions [48]. Besides, different choices of the reconstruction model may yield
266 different detection results [49].

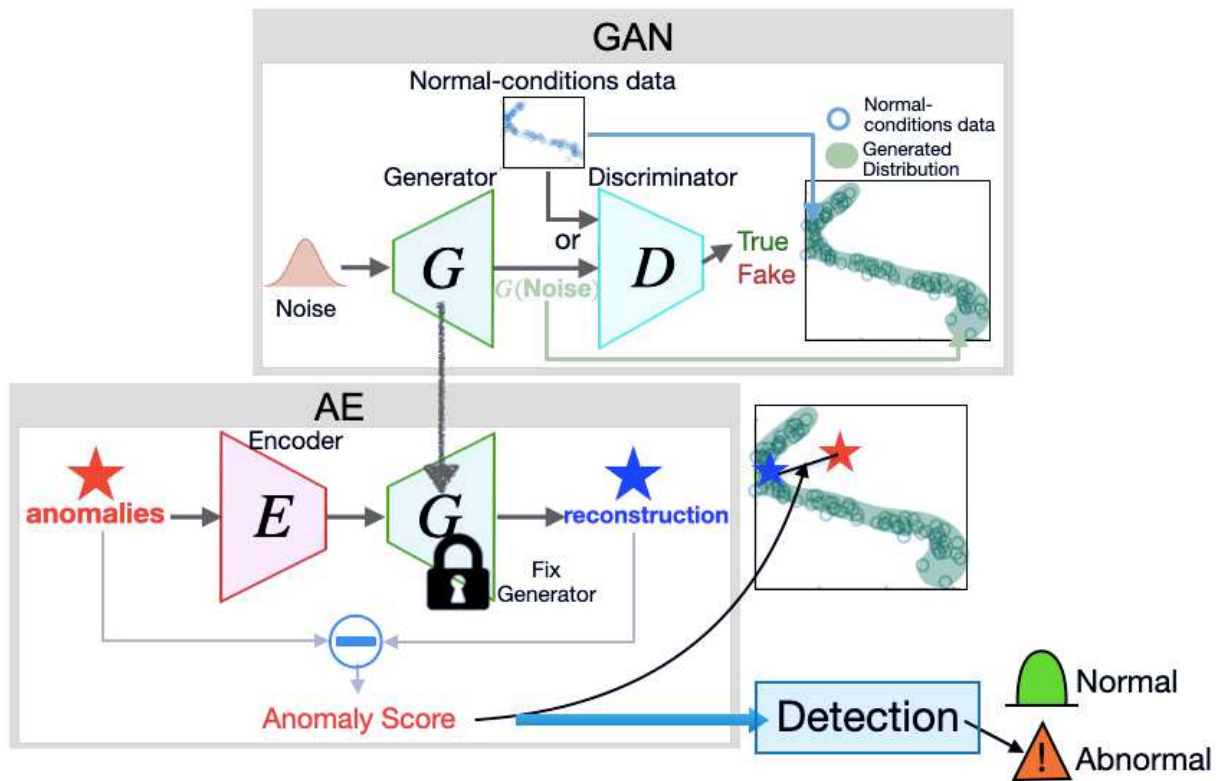
267 Eventually, the detection of an abnormal condition is confirmed by considering
268 whether the obtained residuals exceed a threshold or by statistical tests. For example, [43]
269 uses the Sequential Probability Ratio Test (SPRT) on the residuals obtained from an AAKR
270 model; [50] applies T2- and Q-statistics of the PCA residuals to detect damages in structures;
271 [51] establishes a statistical hypothesis model in the residual subspace of PCA transform, to
272 detect and isolate sensor faults based on a Bayesian formulation and the generalized
273 likelihood ratio test (GLRT). Notice that, although these methods assume a certain
274 distribution of the residuals, most distributions of real-world data may be a priori unknown or
275 may not actually follow the assumed distributions [52].

276 Another challenge of fault detection lies in the data pre-processing [53] to extract
277 features providing the information useful for enabling the detection. Various pre-processing
278 techniques, such as Fast Fourier Transform (FFT) [54], Continuous Wavelet Transform [55],
279 Mathematical Morphology [56], have been applied to raw signals, and the processed
280 outcomes have been fed to fault detection [57]. The quality of the features selected by pre-
281 processing strongly impacts the detection results, but unfortunately there is no universal rule
282 for choosing the optimal pre-processing method.

283 Recently, transport-related methods are being considered for applications in PHM.
284 They have already been successfully employed in other domains [58], involving signal and
285 image processing [59], computer vision [60], machine learning and statistics [61], [62].
286 Commonly used optimal transport distances include Wasserstein distance (or Kantorovich
287 distance) [63] and Earth Mover's distance (EMD) [64]. Wasserstein distance has proved a
288 promising statistic for the nonparametric two-sample test [65].

289 In the PHM area, [66] has studied the bearing diagnostics problem using EMD
290 combined with dynamical system reconstruction. [67] has used a PCA scheme combined with
291 the Kantorovich distance (KD) for fault detection in the process industry. [68] has developed a
292 method of OT in which the abnormality score is built using the Wasserstein distance and has
293 verified its performance considering the detection of abnormal conditions in bearings. The
294 method differs from other state-of-the-art methods for fault detection, since it directly deals
295 with raw signals and does not require the use of signal reconstruction methods or feature
296 extraction; it is also distribution-free, i.e., it does not require to formulate any a priori
297 hypothesis on the distribution of the data. The basic idea behind the method is to generate an
298 abnormality score, based on Wasserstein distance, to quantify the dissimilarity between the
299 probability distributions of the currently monitored and healthy data. The Cumulative
300 Distribution Transform (CDT) [69] is used to find the univariate Optimal Transport (OT)
301 solution. The method has been applied to a real bearing dataset and successfully compared
302 with two other fault detection methods of literature: a Z-test based method [70] and a PCA-
303 based method for signal reconstruction, combined with the Q-statistic for residual analysis
304 [71]. The Adaboost ensembled AE-GAN method mentioned earlier [39] can also be adapted
305 for application to normal-conditions data only. The generator of the GAN and the auxiliary
306 encoder form the AE module, and the reconstruction error generated by the AE is used as the
307 score to detect abnormalities/anomalies in SSC behavior. For the abnormality/anomaly
308 detection, it is assumed, as usual, that the probability distribution of the abnormal/anomalous-
309 conditions data is significantly different from that of the normal-conditions data: as the
310 generator can only reproduce the distribution of the normal-conditions data, the AE always
311 successfully reconstruct the normal data but fails to reconstruct the abnormal ones. So, any
312 test sample processed through the AE-GAN is declared anomalous if the AE reconstruction
313 error is larger than a certain predefined threshold. For dealing with the high dimensionality of
314 the data, again, an ensemble framework can be used. Non-overlapped sliding time windows
315 are introduced to partition the multivariate time series and a separate data sample for each
316 time window is analyzed by AE-GAN for abnormal/anomaly detection. Finally, the AdaBoost

317 algorithm is used to aggregate the abnormality/anomaly detection results for each time
 318 window. The GAN-based method for addressing the challenge of missing fault data in fault
 319 detection is shown in Figure 4.
 320



321
 322 *Figure 4 Illustration of GAN-based method in fault detection w.r.t. the challenge of*
 323 *missing fault data [39]. GAN-based method is a type of distribution reconstruction method*
 324 *which reproduces the normal-conditions data distribution by the Generator and uses an extra*
 325 *Encoder to form an Auto-Encoder, which can obtain anomaly scores (reconstruction errors)*
 326 *to distinguish whether samples are anomalous or not.*
 327

328 Table 1 summarizes the fault detection techniques, with specific regard to the
 329 challenge of missing fault data.

330 Table 1 Fault detection techniques with regard to the challenge of missing fault data.

Model-based	Data-driven			
	Supervised learning (Limitation: need both healthy and fault condition data)	One-class classification (Advantage: addressing the challenge of missing fault data)		
		Residual-based (Limitation: need to apply statistical test)	Transport-related	Distribution reconstruction-based
Finite Element Method: [26] rotor crack diagnostics	RF: [30] power generators fault detection	AAKR: [42] non-linear multimode processes fault detection [43] reactor coolant pump fault detection [44] power plant	EMD: [66] bearing fault detection	GAN-based: [39] high-speed train automatic door

		fault detection	
Observer theory: [27] rotor fault detection	SVM: [31] early fault detection of numerical case [35] high-speed train brake fault detection	PCA: [45] air handling unit fault detection	Kantorovich Distance: [67] tank heater simulation case fault detection
HMM: [28] mechanical equipment early fault detection	ANN: [32] heavy-water reactor early fault detection [34] wind turbine false alarm detection	OC-SVM: [46] building air conditioning system fault detection	Wasserstein Distance: [68] bearings fault detection
	RNN: [33] software fault detection	ANN: [47] wind turbine gearbox fault detection	CDT: [69] numerical case

331

332

3.1.2 Fault diagnostics

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

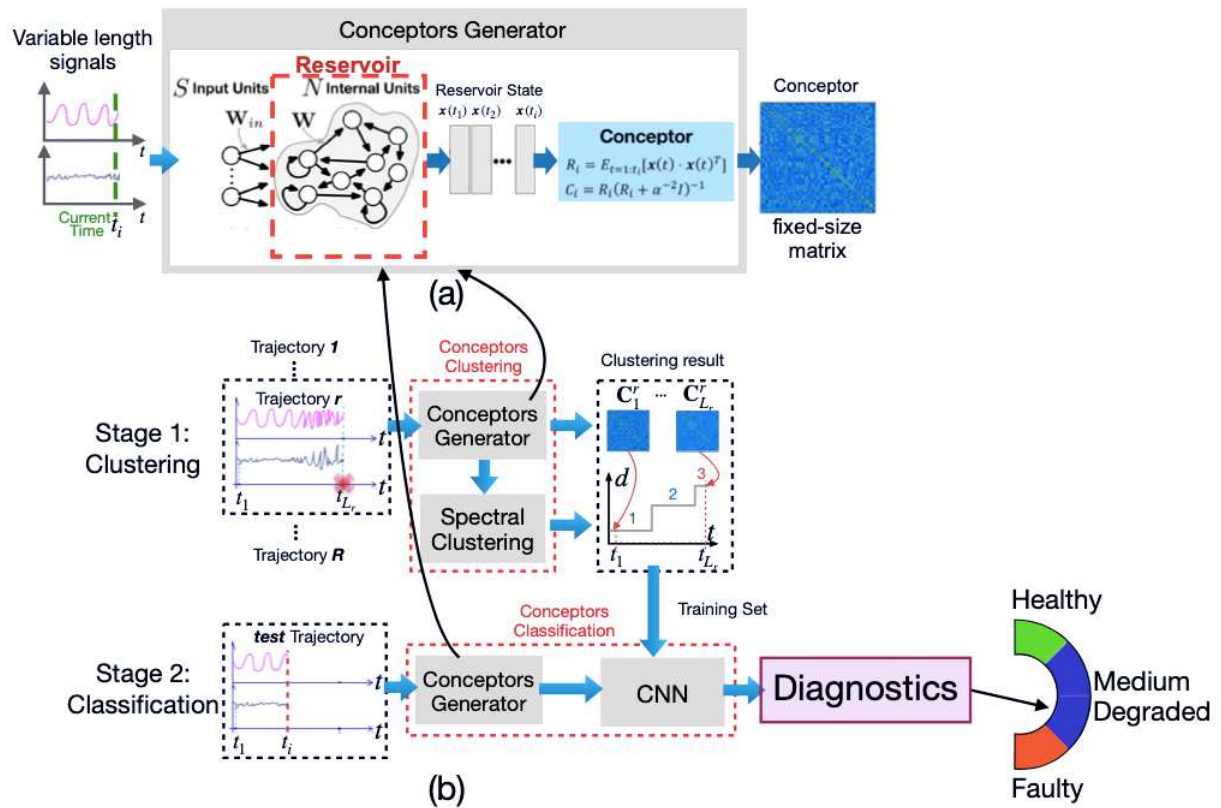
361

Fault diagnostics requires data analytics capable of identifying the equipment fault state, mode, location and other characteristics of interest, based on monitored signals (temperature, pressure, current, acceleration, etc.). As for the detection task previously discussed, in practical applications it also suffers from the presence of uncertainty coming from the processing of data of the measured signals. Common approaches make use of historical operational data to build empirical classifiers capable of discriminating different classes from the data, for fault diagnostics. Different classification techniques, such as Complex Valued Neural Network (CVNN) [72], Deep Belief Network (DBN) [73], Bayesian Network (BN) [74], Decision Trees (DT) [75], Linear Discriminant Analysis (LDA) [75], K-Nearest Neighbor (KNN) [75], Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) [76][75][77], have been successfully used in applications of different industrial and civil sectors [78]. These methods rely on supervised learning of labelled data, which, however, are rarely available in practice, so that their real application is limited [79]: the real application calls for unsupervised learning of unlabeled data.

Unsupervised learning is an important topic in machine learning for time series segmentation [80], [81] and pattern recognition [21], [82], [83]. In fault diagnostic applications, it is used to provide abstract representations of the raw measurement data and obtain various clusters representing healthy and faulty conditions [22], [84]–[86]. In the work of [22], a Categorical Adversarial Autoencoder (CatAAE) has been proposed for unsupervised learning aimed at fault diagnostics of rolling bearings. In the work of [84], a diagnostic methodology based on unsupervised Spectral Clustering (SC) combined with fuzzy C-means (FCM) has been developed for identifying groups of similar shutdown transients performed by a nuclear turbine. In [85], Self-Organizing Map (SOM) has been used for clustering and identifying degradation states of a railway-signal system. In [86], a methodology combining k-means and Association Rule Mining (ARM) has been developed to mine failure data and diagnose interconnections between failure occurrences in wind turbines. Representation learning can disentangle the different explanatory factors of variation behind the data, making it easier to extract and organize the discriminative information when building fault diagnostic models [87]–[92]. In traditional unsupervised methods for fault diagnostics, the features are

362 extracted applying ad hoc signal processing techniques to the collected signals, e.g. Fourier
363 spectral analysis and Wavelet transformations [93]. The processing is heavily dependent on a
364 priori knowledge and diagnostic expertise [22], [94], and can be quite time consuming and
365 labor-intensive [88]. Since representation learning is adaptively capable of learning features
366 from raw data, it can constitute an excellent a priori choice for the development of diagnostic
367 techniques. In the work of [95], an unsupervised sparse filtering method based on a two-layer
368 neural network is used to directly learn features from mechanical vibration signals. In the
369 work of [96], a Spatio-Temporal Pattern Network (STPN) based on Probabilistic Finite State
370 Automation (PFSA) and Markov machines is proposed to represent temporal and spatial
371 structures for fault diagnostics in complex systems. However, these conventional
372 representation learning methods cannot capture long-term temporal dependencies in the time
373 series and they typically require high computational complexity.

374 From the above, it is seen that traditional fault diagnostic approaches typically require
375 the acquisition of signal measurements from SSCs whose true degradation state is known.
376 However, to acquire such labelled data is a difficult, expensive and labor-intensive task.
377 Furthermore, streaming data collected in online-monitored SSCs have long-term temporal
378 dependencies. However, unsupervised learning methods have a hard time dealing with long-
379 term time dependencies, because these dependencies are limited by the size of the sliding time
380 window which can be used for the analysis. Then, there is a need for advancements in the
381 methods to estimate the degradation level at a given time on the basis of a few run-to-failure
382 trajectories with long-term temporal dependencies and for which the true degradation state is
383 unknown. In the work of [97], for example, a two-stage method for unsupervised learning is
384 proposed for fault diagnostic applications, inspired by the idea of representing temporal
385 patterns by a mechanism of neurodynamical pattern learning, called Conceptor. Considering a
386 reservoir, i.e. a randomly generated and sparsely connected RNN [98], Conceptors can be
387 understood as filters characterizing the geometries of the temporal states of the reservoir
388 neurons in the form of square matrices [99], achieving a direction-selective damping of high-
389 dimensional reservoir states [100]. The proposed method develops in two stages. In the first
390 stage, the Conceptors extracted from the training run-to-failure degradation trajectories are
391 clustered into several non-overlapped time series segments representing different degradation
392 levels. In the second stage, the Conceptors and corresponding labels obtained in the first-stage
393 clustering are used to train a Convolutional Neural Network (CNN) for real-time diagnosing
394 the SSC degradation level. The CNN receives in input the Conceptors extracted from the
395 reservoir states at the current time, which contain information about the long-term evolution
396 of the SSC degradation, and the difference between the Conceptors extracted at the present
397 and previous time steps, which contains information about the short-term degradation
398 variation. The proposed method has been applied to two literature case studies concerning
399 bearings fault diagnostics. The results show satisfactory accuracy and efficiency of the
400 method. The Reservoir computing-based method for addressing the challenge of missing
401 labels of degradation state is shown in Figure 5.



402
 403 *Figure 5 Illustration of Reservoir Computing-based method in fault diagnostics w.r.t.*
 404 *the challenge of missing labels of degradation state [97]. a) Conceptor Generator converts*
 405 *the variable-length signals into a fixed-size Coceptor matrix by using reservoir computing.*
 406 *The Conceptor matrix is a filtered correlation matrix of reservoir states, which decreases the*
 407 *impact of less important features in the degradation dynamics, i.e. noise and operation*
 408 *conditions. b) stage 1: the combined use of Coceptors matrix and spectral clustering can*
 409 *obtain the pseudo-labels in the run-to-failure trajectories, stage 2: CNN trained by Coceptors*
 410 *with associated pseudo-labels is, then, used for fault diagnostics.*
 411
 412

413 Table 2 summarizes the fault diagnostics techniques, with specific regard to the
 414 challenge of missing labels for degradation states.

415 Table 2. Fault diagnostics techniques with regard to the challenge of missing labels for
 416 degradation state.

Supervised classifiers (Limitation: need degradation state labels)	Unsupervised approaches		
	Unsupervised learning (Limitation: depend on diagnostic expertise for feature extraction)	Representation learning	
		Neural Network- based (Limitation: difficult to capture long-term temporal dependencies)	Reservoir Computing-based
CVNN: [72] railway track turnouts degradation level assessment	CatAAE: [22] rolling bearing fault diagnostics	Unsupervised Sparse Filtering Neural Network: [95] motor bearing fault diagnostics	Conceptor + SC: (Conceptor as a representation capable of encoding long- term temporal
DBN: [73] aircraft engine	SC + FCM: [84] Nuclear Power		

health state classification	Plant (NPP) steam turbine transient identification		<i>dependencies</i> [97] degradation level assessment of bearings
BN: [74] failure type classification of water distribution system	SOM: [85] fault mode classification for railway monitoring equipment	STPN: [96] fault severities diagnostics of wind turbine	
DT: [75] anomaly (mud) diagnostic on wind turbine blade	k-means + ARM: [86] failure cause and weather condition correlation diagnostics of wind turbine system		
LDA: [75] anomaly (mud) diagnostic on wind turbine blade			
KNN: [75] anomaly (mud) diagnostic on wind turbine blade			
ANNs: [101] fault diagnostics of bearings			
SVMs: [75] anomaly (mud) diagnostic on wind turbine blade [76] bearing defects diagnostics [77] pipe failure prediction in water supply networks			

417

418

3.1.3 Fault prognostics

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

Prognostics is concerned with the prediction of the future evolution to failure of the state of a SSC. It involves the processing of data to predict the future degradation of the SSC structural and functional attributes, based on which to estimate the SSC failure probability and RUL. The prognostic outcomes are used for the health management of the SSC, which seeks to use the prognosis to decide on and actuate operational actions and maintenance interventions. To the uncertainties coming from the use of the data available from the sensors, like for the detection and diagnosis tasks, prognostics adds further challenges related to the future evolution of the usage profile and operational environment, whose uncertainties affect the degradation state evolution. This makes it practically impossible to precisely predict the future evolution of the SSC state of health and it is necessary to account for the different sources of uncertainty that affect prognostics, within a systematic framework for uncertainty quantification and management [102].

Prognostics is dependent on the available knowledge, information and data on the process of degradation. There may be situations in which a sufficient quantity of run-to-failure data has been collected during the life of the SSCs, and these can be used to develop empirical (data-driven) models. In other cases, the degradation mechanism is known and a physics-

435 based model is available. On these bases, prognostics approaches can be grouped into three
436 categories: (i) model-based, (ii) data-driven and (iii) hybrid:

- 437 • (i) Model-based approaches use physics-based degradation models to predict the
438 future evolution of the SSCs degradation state and infer the time at which the
439 degradation will reach the failure threshold. These approaches have been applied with
440 success in various practical cases, e.g., to pneumatic valves [103], Li-Ion batteries
441 [104], the residual heat removal subsystem of a nuclear power plant [105], and
442 structures subject to fatigue degradation [106]. In the case of complex SSCs, subject to
443 multiple and competing degradation mechanisms, accurate physics-based models are,
444 however, often not available.
- 445 • (ii) Data-driven approaches directly extract from the data the degradation law for SSCs
446 RUL prediction [107]. Such approaches include conventional numerical time series
447 techniques, as well as AI intelligence and data mining algorithms, such as similarity-
448 based [108] and regression-based methods [107]. A variety of AI techniques, such as
449 Convolutional Neural Network (CNN) [109][110], Denoising Auto Encoder (DAE)
450 [111], Long Short Term Memory (LSTM) [112][109][113][114], Gated Recurrent Unit
451 (GRU) [115], SVM [116], Adjacency Difference Neural Network (ADNN) [117], are
452 applied to RUL estimation of different industrial systems and components. The
453 performances of data-driven approaches depend on the quantity and informative
454 quality of the data available to develop the predictive models.
- 455 • (iii) Hybrid approaches combine, all the available sources of knowledge, information
456 and data. They bring the advantages of both model-based and data-driven methods.
457 Specifically, they can integrate the robustness and interpretability of model-based
458 methods with the specificity and accuracy of data-driven methods. For instance, [118]
459 combines Kalman Filtering (KF) with data-driven approaches, [119] integrates the
460 Health Indicator (HI) and regression model, [120] combines Relevance Vector
461 Machine (RVM) and Particle Filtering (PF), and [121] integrates a physical model and
462 the Least Square (LS) method to estimate RUL of a variety of industrial equipment.

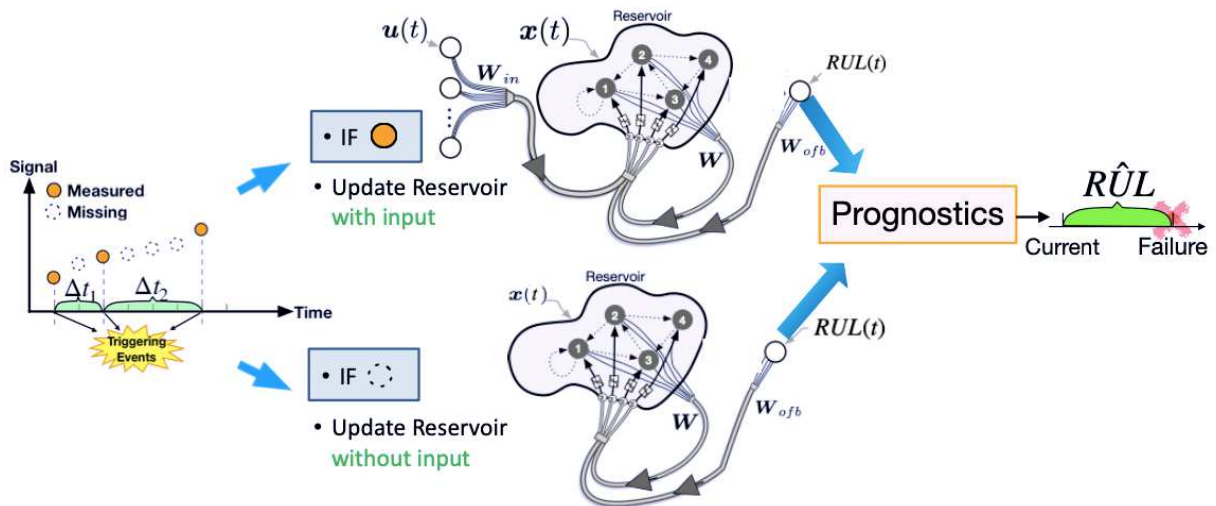
463 Traditional fault prognostic methods face the challenge of dealing with incomplete and
464 noisy data collected at irregular time steps, e.g. in correspondence of the occurrence of
465 triggering events in the system. For example, for monitoring the degradation and failure
466 processes of bearings in large turbine units, signal measurements collection (e.g., vibration
467 signals measured by eddy current displacement sensors measuring the radial vibration of the
468 rotor at both ends, the axial vibration of the rotor, and sensors measuring the unit rotating
469 speed) is only triggered by abnormal behaviors of the units, such as large environmental noise
470 and anomalous vibration behavior. These “snapshot” datasets are often encountered in
471 industrial applications, dominated by the necessity of cost saving in storing and managing the
472 databases, and of reducing energy consumption and bandwidth resources. Since failure events
473 are rare, event-based datasets are dominated by missing measurements, where the values of all
474 signals are missing at the same time. With these characteristics, traditional methods for
475 missing data management, e.g. case deletion, imputation [122]–[125] and maximum
476 likelihood estimation [126], are difficult to apply. For instance, since case deletion methods
477 discard patterns whose information is incomplete, they are not useful in case of event-based
478 datasets where a pattern is either present or absent for all signals [126]. Imputation
479 techniques, which are based on the idea that a missing value of a signal can be replaced by a
480 statistical indicator of the probability distribution generating the data, such as the signal mean
481 value [127] or a value predicted by a multivariable regression model, have been shown
482 inaccurate in case of large fractions of missing values in the dataset [128]. Maximum
483 Likelihood methods use the available data to identify the values of the probability distribution
484 parameters with the largest probability of producing the sample data. They typically require

485 the Missing At Random (MAR) assumption, i.e. the probability of having a missing value is
486 not dependent on the missing values[127],[129], which is not met in event-based datasets.

487 Few research works have considered fault prognostics in presence of missing data. A
488 model based on Auto-Regressive Moving Average (ARMA) and Auto-Associative Neural
489 Networks (AANN), has been developed for fault diagnostics and prognostics of water process
490 systems with incomplete data [130]. An integrated Extreme Learning Machine (ELM)-based
491 imputation-prediction scheme for prognostics of battery data with missing data [125] and an
492 hybrid architecture of physics-based and data-driven approaches have been proposed to deal
493 with missing data in a rotating machinery prognostic application [131]. In the medical field, a
494 Bayesian simulator has been used to generate missing data for developing prognostic models
495 [132] and a Multiple Imputation approach has been embedded within a prognostic model for
496 assessing overall survival of ovarian cancer in presence of missing covariate data [133].
497 Notice that all these methods are based on the two successive steps of missing data
498 reconstruction and prediction.

499 Then, advancements and new methods are still needed to enable predicting the RUL of a
500 SSC on the basis of measurements collected only when triggering events occur, such as SSC
501 faults or extreme operational conditions, and providing an estimate of the uncertainty
502 affecting the RUL prediction. As an example, [134] has developed a method based on Echo-
503 State Networks (ESNs) to directly predict the RUL of a SSC without requiring to reconstruct
504 the missing data. ESNs are considered because of their ability of maintaining information
505 about the input history inside the reservoir states. The main difficulty is that, contrarily to the
506 typical applications of ESNs, the time intervals at which the data become available are
507 irregular. Two different strategies have been considered to cope with the event-based data
508 collection. In one strategy, the ESN receives an input pattern only when an event occurs. The
509 pattern is formed by the measured signals and the time at which the event has occurred. In a
510 second strategy, the reservoir states are excited at each time step. If an event has occurred, the
511 reservoir states are excited both by the previous reservoir states and the measured signals,
512 whereas, if an event has not occurred, they are excited only by the previous reservoir states.
513 By so doing, the connection loops in the reservoir allow reconstructing the SSC dynamic
514 degradation behavior at those time steps in which events do not occur. Multi-Objective
515 Differential Evolution (MODE) algorithm based on a Self-adaptive Differential Evolution
516 with Neighborhood Search (SaNSDE) [135] is used to optimize the ESN hyper-parameters.
517 The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [136] is,
518 then, used to select the optimal solution from the obtained Pareto solutions. Furthermore, a
519 bootstrap aggregating (Bagging) ensemble method is applied to improve the RUL prediction
520 accuracy and estimate the RUL prediction uncertainty. Given that ESNs cannot be fed by
521 random sequences of patterns, the traditional Bagging sampling mechanism used to create the
522 bootstrap training sets has been modified. In the proposed solution, the bootstrap training sets
523 are obtained by concatenating entire run-to-failure trajectories, randomly sampled with
524 replacement. The benefits of the proposed methods are shown by application to the prediction
525 of the RUL of a sliding bearing of a turbine unit. The ESN-based one-step RUL prediction
526 method for the challenge of missing data, i.e., event-based measurements, is shown in Figure
527 6.

528



529
 530 *Figure 6 Illustration of ESN-based method in fault prognostics w.r.t. the challenge of missing*
 531 *data, i.e. event-based measurements [134]. The input neurons of ESN are excited to update the*
 532 *reservoir state when measurements are available (events are triggered), whereas the input*
 533 *neurons are canceled if data are missing (no events occur) and the reservoir is only updated*
 534 *by the reservoir state at the previous time step and the target signal, which force the reservoir*
 535 *to learn from the historical degradation pattern and the target signal evolution pattern.*
 536

537 Table 3 summarizes fault prognostics techniques, with specific regard to the challenge
 538 of missing data, i.e. event-based measurements.
 539

540 Table 3. Fault prognostics techniques with regard to the challenge of missing data, i.e. event-
 541 based measurements.

Traditional fault prognostics (Limitations: cannot deal with missing data, i.e. event-based measurements)			Fault prognostics in presence of missing data	
Model-based	Data-driven	hybrid	Conventional type of missing data (Limitations: difficult to deal with event-based measurements)	Event-based measurement
[103] fault prognostics of pneumatic valves [104] prognostics and health monitoring of Lithium-ion battery [105] fault	CNN: [109] RUL estimation for bearing [110] RUL estimation for turbofan engine	KF and data-driven approaches: [118] RUL estimation for aircraft bleed valve	Missing at random	ESN-based one-step RUL prediction without requiring to reconstruct the missing data: [134] RUL and uncertainty estimation of bearing
	DAE: [111] RUL estimation for	HI and regression model:	Missing not at random	

prognostics in residual heat removal subsystem	centrifugal pumps	[119] RUL estimation for bearings	Missing data prediction by Quasi-Newton Optimization; Prognostics by AANN and ARMA: [130] prognostics for wastewater plant
	LSTM: [109] RUL estimation for bearing [112] RUL estimation for turbofan engine [113] RUL estimation for turbofan engine [114] RUL estimation for turbofan engine	RVM and PF: [120] RUL estimation for Lithium-ion battery	Missing at extreme operating condition
		physical model and LS: [121] RUL estimation for Lithium-ion battery	Missing data generation by physical model; Prognostics by physical model and data-driven: [131] prognostics for bearing
	GRU: [115] RUL estimation for turbofan engine		
	SVM: [116] RUL estimation for aircraft engine		
	ADNN: [117] RUL estimation for aircraft engine		

542
543
544
545
546
547
548
549
550
551
552
553
554
555
556

3.2 Challenges from requirements on practical solutions

3.2.1 Interpretability of models

The ability to correctly interpret a PHM model’s output, be it the detection of a fault, its diagnosis or prognosis, is extremely important, and particularly so in safety-critical applications like those concerning the high-risk systems and processes of the chemical, nuclear, aerospace industries, to name a few. It allows understanding of the state of the system or process being modeled and supports analytic reasoning and prescriptive decision making to intervene (or not) and how. It also engenders appropriate trust by the analyst, providing insights on how the model works. The importance of this is such that in some applications, simple models (e.g., even linear models) are preferred for their ease of interpretation, even if they may be less accurate than complex ones. Yet, currently the growing availability of big data for PHM has increased the benefits of using complex models for achieving accuracy, at the expenses of model intelligibility. This brings to the forefront the need of a trade-off between accuracy of the model and interpretability of its output. A wide variety of different

557 methods have been recently proposed to address this issue, but an understanding of how these
558 methods relate and when one method is preferable to another is still lacking.

559 Most models and algorithms for PHM are developed and trained to maximize
560 accuracy, neglecting interpretability and causality. Accounting for these aspects may, indeed,
561 lead to a loss in performance but would enhance their safe, reliable and robust use both in
562 terms of undesired biases and uncertainty reduction. Understanding why a PHM model makes
563 a certain prediction can be as crucial as the prediction's accuracy in many applications.
564 However, the highest accuracy for large modern datasets is often achieved by complex models
565 that even experts struggle to interpret, such as ensemble or deep learning models, creating a
566 tension between accuracy and interpretability [137]. Some general attributes sought for in the
567 interpretability of PHM models are:

- 568 • fairness: no discrimination in algorithm decisions, which could come from bias
569 in the collected data
- 570 • robustness: small changes in input should not cause big changes in output
- 571 • causality: causal relations are picked up from the model and rendered explicit
- 572 • quantifiable reliability of outcomes and predictions.

573 The awareness of the relevance of transparency, explainability and interpretability of PHM
574 models is growing as a need and a requirement, particularly for supporting decision making in
575 safety critical systems, for which it may also be a regulatory prerequisite. For example, in
576 Nuclear Power Plants (NPPs), there is still resistance to the deep penetration of digital I&C
577 systems and PHM, because of the difficulty of testing performance under all postulated
578 conditions, on one side, and guaranteeing reliability based on transparent understanding and
579 interpretation, on the other side. The decision making related to tasks of control, operation,
580 maintenance and safety of NPPs, which have traditionally relied on procedures and expert
581 evaluation and judgment, are gradually being assisted by intelligent machines (i.e. software
582 algorithms) for PHM, developed and trained on the basis of big and customized data: how far
583 and how it can be permitted in safety-critical systems that require licensing depends also on
584 the possibility of interpreting the causality of their output.

585 For the modelling approaches to PHM based on learning from data, one issue lies in
586 possible biases in the training set that are, then, not present in the test set or contain patterns
587 undesired with respect to the test data, and may be unknown to the user of the trained model
588 output. In this sense, achieving robustness in PHM models is fundamental and one way to
589 proceed is to try to design inherently interpretable models, i.e. so as to exclude all undesired
590 features that are not causally related to the outcome. By examining interpretable models:

- 591 • features or functions capturing quirks in the data can be noted and excluded, thereby
592 avoiding related harm in the successive use of the model output, and the understanding
593 of the phenomena analyzed
- 594 • knowledge can be extracted, in terms of the interactions among the inputs and how
595 they determine the output
- 596 • an evaluation of the reliability of the PHM outcomes can be performed
- 597 • some limited extrapolation can be possible, with the aim of gaining knowledge on
598 unexplored scenarios.

599 Methodologies are used to gain interpretability in a model by looking at the importance of the
600 different input features in determining the model outputs. A distinction is made between
601 model-specific and model-agnostic methodologies for evaluating feature importance. An
602 interesting example of the former is the “attention mechanism” for Neural Networks applied
603 in Prognostics, where importance values are assigned to specific input subsets [138], [139].

604 As the name implies, model-agnostic feature importance evaluation methodologies can
605 in principle be used for any model. Local approaches are used for online applications and
606 global approaches for offline applications. Local measures focus on the contribution of

607 features to a specific outcome instance, whereas global measures take all outcomes into
608 account.

609 The Local Interpretable Model Explanation (LIME) method aims at explaining
610 individual outputs and can be applied to any learning model [140]. Instead of training a global
611 surrogate model, LIME focuses on training local surrogate models to explain individual
612 model outputs. The method works by building for each output instance of interest a local-
613 interpretable model that approximates the original, complex model. Each model output
614 instance is, then, explained by an “explainer-model” that highlights the symptoms that are
615 most relevant to it. With this information about the rationale behind the model, the analyst is
616 now empowered to trust the model output—or not – for her/his decisions and consequent
617 actions.

618 The idea behind LIME is quite intuitive and it is based on the fact that one can probe
619 the model as many times as desired, by feeding the input data points and retrieving the
620 corresponding outputs of the model. The goal of this is to understand why the learning model
621 gave a certain output. The LIME tests are local sensitivity tests performed in a way to explore
622 what happens to the output when the inputs are locally varied by small perturbations. By so
623 doing, a new dataset is generated, consisting of permuted input samples and corresponding
624 model outputs. For example, the new samples can be created by perturbing each feature
625 individually, drawing from a normal distribution with mean and standard deviation taken from
626 the feature values. On this new dataset, LIME builds and trains the interpretable explainer-
627 model, which is weighed by the proximity of the sampled instances to the instance of interest.
628 The interpretable model should give a good approximation of the original model outputs
629 locally, but it does not have to be a good global approximation of the original model itself.
630 Mathematically, the interpretable explainer model for instance x is the (simple) model g (e.g. a
631 linear regression model) that results as solution of the optimization problem that minimizes
632 the loss function L (e.g. the mean squared error) measuring how close the explanation output
633 of g is to the output of the original model f (e.g. a neural network), while the model
634 complexity $\Omega(g)$ is kept low (e.g. as few features as possible):

$$635 \text{explanation}(x) = \underset{g \in G}{\operatorname{argmin}} L(f, g, \pi_x) + \Omega(g) \quad (1)$$

636 where G is the family of possible explainer models, for example all possible linear regression
637 models, and the proximity measure π_x defines how large is the neighborhood around instance
638 x that is considered for the explanation. In practice, LIME only optimizes the loss part and the
639 user controls the model complexity by $\Omega(g)$, e.g. by selecting by forward and backward
640 feature selection methods the maximum number of features that the linear regression model
641 may use.

642 The procedure for interpreting locally the complex original model is, then:

- 643 i) select the instance of interest x for which an explanation of the original complex model
644 outcome $f(x)$ is needed
- 645 ii) perturb the input data and get the original model output values for these new data
646 samples
- 647 iii) weigh the new samples according to their proximity to the instance of interest
- 648 iv) train a weighed, interpretable model on the new dataset generated in *ii*)
- 649 v) explain the local output of the interpretable model g .

650 LIME has been applied for the interpretation of machine learning models in applications
651 of medical diagnostics [141]. In a recent study about early Parkinson detection, LIME has
652 been used to highlight the features determining the healthy/disease decision of a ML classifier
653 of images of the brain: LIME allows highlighting the super-pixels mostly determining the
654

657 classification in healthy or disease states; experts can, then, focus on the super-pixels selected
658 with LIME to interpret and explain the basis for the decision by the ML algorithm, and choose
659 to accept or refuse it.

660 Shapley values also can be used to assess local features importance [142]. Although they
661 can be used to explain which feature(s) contribute most to a specific model output, Shapley
662 values are not designed to answer the “what would happen if” questions that LIME’s local
663 explainer models are designed for. They come from game theory and are designed to construct
664 a fair payout scheme for the players in a game. Suppose one could look at all possible
665 combinations of (a subset of) players in a team replaying a game and observe the resulting
666 team score. One could, then, assign each player of the team a portion of the total payout based
667 on its average added value across all possible subteams to which it was added to play the
668 game repeatedly. Such individual payout is the player’s Shapley value and gives the only
669 payout scheme that is proven to be:

- 670 • efficient: the sum of the Shapley values of all players should sum up to the total
671 payout
- 672 • symmetric: two players should get the same payout if they add the same value in all
673 team combinations
- 674 • dummy-sensitive: a player should get a Shapley value of zero if it never improves a
675 subteam’s performance when it is added
- 676 • additive: in case of a combined payout (say we add two game bonuses), the combined
677 Shapley value of a player across the games is the sum of the individual game’s
678 Shapley values; this criterion has no relevant analogy in the context of model
679 interpretability.

680 In the “game” of our interest for PHM model interpretability, the players are models with
681 different features subsets and they get the same payout mechanism introduced above. The
682 team score in this context is the performance measure of a (sub)model built on a given feature
683 subset. The total payout is the difference between a base value — output of the null model —
684 and the actual output. This difference is, then, divided over all features in accordance to their
685 relative contribution.

686 Obviously looking at all possible subsets of features is computationally prohibitive in
687 most realistic models with many features. Instead, Shapley value approximations can be
688 computed based on sampling of features.

689 Other model-agnostic methodologies are based on Sensitivity Analysis (SA), which has
690 been widely applied to models used in various areas, such as nuclear risk assessment [143],
691 industrial bioprocessing [144] and climate change [145]. Indeed, a main application of SA is
692 for identifying the input quantities most responsible of a given output variation [146]. Both
693 local and global approaches to SA have been developed. Local approaches identify the critical
694 input features as those whose variation leads to the most variation in the output. One practical
695 approach for such identification consists in perturbing one single input at a time with small
696 variations around its nominal value, while maintaining the others set at their respective
697 nominal values. The analysis is intrinsically local and the resulting indication can be
698 considered valid for the characterization of the model response around the nominal values.
699 The possibility of extending the results of the analysis to draw global considerations on the
700 model response over the whole input variability space depends on the model itself: if the
701 model is linear or mildly non-linear, then the extension may be possible; if the model is
702 strongly non-linear and characterized by sharp variations, the analysis is valid only locally.
703 Typical local approach techniques are those based on Taylor’s differential analysis and on the
704 one-at-a-time simulation, in which the input features are varied one at a time while the others
705 remain set at their nominal values [146].

706 In those situations (often encountered in practice) in which models are non-linear and non-
707 monotone, the results provided by a local analysis may have limited significance. For this
708 reason, global approaches to SA have been developed. In these approaches, the focus is
709 directly on the uncertainty distribution of the output, which contains all the information about
710 the variability of the model response, with no reference to any particular value of the input
711 (like in the local approaches, where reference is made to the nominal values). The two
712 principal characteristics of the global approaches are somewhat opposite to those of the local
713 ones: 1) the account given to the whole variability range of the input features (and not only to
714 small perturbations around the nominal values); 2) the focus on the effects resulting from
715 considering also the variation of the other uncertain features (instead of keeping them fixed to
716 their nominal values). Many global analysis methods have been developed [146]. The high
717 capabilities of these methods are paid by a very high computational cost.

718 Another direction to build interpretability into PHM models and algorithms is by
719 integrating prior physical knowledge in the learning models, for providing improved
720 performance and achieving interpretability. This is a promising approach for inducing
721 interpretability into the learning models and different approaches have been proposed where
722 the physical knowledge can be introduced at different levels of the learning process, including
723 in the training data and in the training algorithm [147][148][149][150].

724 To aid the interpretation of the model, there exists also a suite of methods for the
725 visualization of the relations between input and output. The Partial Dependence Plot (PDP)
726 shows the marginal effect that features have on the output provided by the model [151].
727 Intuitively, we can interpret the partial dependence as the expected target response as a
728 function of the input features of interest. A partial dependence plot can show whether the
729 relationship between the output and a feature is linear, monotonic or more complex. For
730 example, when applied to a linear regression model, partial dependence plots always show a
731 linear relationship. The computation of partial dependence plots is intuitive: the partial
732 dependence function at a particular feature value represents the average output if we force all
733 data to assume that value for the feature. If the feature for which the PDP is computed is not
734 correlated with the other features, then the PDP perfectly represents how the feature
735 influences the output on average. In the uncorrelated case, the interpretation is clear: the PDP
736 shows how the average output changes when a given feature is changed. The interpretation is
737 more complicated when features are correlated. Also, PDPs are easy to implement and the
738 calculations to obtain them have a causal interpretation which aids model understanding: one
739 intervenes on a feature and measures the corresponding change in the output. By doing so,
740 one analyzes the causal relationship between the feature and the output in the model, and the
741 relationship is causal for the model whose outcome is explicated as a function of the features.
742 However, there are several disadvantages in PDPs. Due to the limits of human perception, the
743 number of features in a partial dependence function must be small (usually, one or two) and,
744 thus, the features considered must be chosen among the most important ones. Some PDPs do
745 not show the feature distribution. Omitting the distribution can be misleading, because one
746 might overinterpret regions with almost no data. This problem is easily solved by showing a
747 rug (indicators for data points on the x -axis) or a histogram. Also, heterogeneous effects might
748 be hidden because PDPs only show the average marginal effects. Suppose that for a feature,
749 half of the input data has a positive correlation with the output (the larger the feature value the
750 larger the output value) and the other half has a negative correlation (the smaller the feature
751 value the larger the output value): then, PDP could be a horizontal line, since the effects of
752 both halves of the dataset could cancel each other out and one would, then, conclude that the
753 feature has no effect on the output. In other words, whereas the PDPs are good at showing the
754 average effect of the target features, they can obscure a heterogeneous relationship created by
755 interactions.

756 When interactions are present, the Individual Conditional Expectation (ICE) plot can be
 757 used to extract more insights [152]. An ICE plot shows the dependence between the output
 758 and an input feature of interest. However, unlike a PDP, which shows the average effect of the
 759 input feature, an ICE plot visualizes the dependence of the output on a feature for each sample
 760 separately, with one line per sample. Again, due to the limits of human perception, only one
 761 input feature of interest is supported by ICE plots. On the other hand, in ICE plots it might not
 762 be easy to see the average effect of the input feature of interest. Hence, it is recommended to
 763 use ICE plots alongside PDPs: they can be plotted together.

764 Finally, the assumption of independence is the biggest issue with PDPs. It is assumed that
 765 the features for which the partial dependence is computed are not correlated with other
 766 features. One solution to this problem is Accumulated Local Effect (ALE) plots that work
 767 with the conditional instead of the marginal distribution (Apley et al., 2020). ALE plots are a
 768 faster than and unbiased alternative to PDPs. Based on the conditional distribution of the
 769 features, they calculate differences in outputs instead of averages. ALE plots are unbiased,
 770 which means they still work when features are correlated, and are faster to compute than
 771 PDPs. The interpretation of ALE plots is also clear: conditional on a given value, the relative
 772 effect on the output due to changing the feature value can be read from the ALE plot. Even
 773 though ALE plots are not biased in case of correlated features, interpretation remains difficult
 774 when features are strongly correlated. Because if they have a very strong correlation, it only
 775 makes sense to analyze the effect of changing both features together and not in isolation. This
 776 disadvantage is not specific to ALE plots, but a general problem of strongly correlated
 777 features. Table 4 summarizes the investigated approaches for interpreting the PHM models.

778

779 Table 4 Summary of model interpretability approaches

Approaches	Characteristic	Reference
LIME	Focuses on training local surrogate models to explain individual model outputs	[140]
Shapley value	Uses game theory to construct a fair payout scheme for the player (input features) to obtain features importance	[137]
Sensitivity analysis-based	Identifies the input quantities most responsible of a given output variation	[143][144][145][146]
Prior physical knowledge-based	Builds interpretability into PHM models and algorithms by integrating prior physical knowledge in the learning models	[147][148][149][150]
PDP	Shows the average marginal effect that features have on the output provided by the model	[151]
ICE	ICE plots visualize the dependence of the output on a feature, for each sample separately	[152]

780

781

782

3.2.2 Security of models

783

784

785

786

Applications of PHM methods for condition-based and predictive maintenance rely on the exchange and elaboration of data. The models and algorithms used are technological elements of larger socio-human-technical systems that must be engineered with safety and security in mind. They are increasingly used in support of high-value decision-making

787 processes in various industries, where the wrong decision may result in serious consequences.
788 The underlying models and algorithms are largely unable to discern between malicious input
789 and benign anomalous data. On the contrary, they should be capable of discerning
790 maliciously-introduced data from benign “Black Swan” events. In particular, the learning
791 models and algorithms should reject training data with negative impacts on results. Otherwise,
792 learning models will always be susceptible to gaming by attackers. The specific danger is that
793 an attacker will attempt to exploit the adaptive aspect of a learning model to cause it to fail
794 and produce errors: if the model misidentifies an hostile input as benign, the hostile input is
795 permitted through the security barrier; if it misidentifies a benign input as hostile, the good
796 input is rejected [153]. The adversarial opponent has a powerful weapon: the ability to design
797 training data that cause the learning model to produce rules that misidentify inputs. To avoid
798 this, the models and algorithms used for PHM must have built-in forensic capabilities [154].
799 These should enable a form of intrusion detection, allowing engineers to determine the exact
800 point in time that an output was given by the model, what input data influenced it and whether
801 or not that data was trustworthy. The data visualization capabilities for the interpretation of
802 the relations between model input features and model output discussed in the previous
803 subsection 3.2.1 show promise to help engineers identify and resolve root causes for these
804 complex issues. Also, specific solutions are required in the areas of Authentication, Input
805 Validation and Denial of Service.

806

807

3.2.3 *Uncertainty*

808

809 Uncertainty is intrinsically present in the PHM tasks of detection, diagnostics and
810 prognostics, and may adversely affect their outcomes, so to lead to an imprecise assessment of
811 the state and prediction of the behavior of such systems, which could lead to wrongly
812 informed system health management decisions with possibly costly, if not catastrophic,
813 consequences. For practical deployment, it is necessary to be able to estimate the uncertainty
814 and confidence in the outcomes of detection, diagnostics and prognostics activities, for
815 quantifying the risk associated to the PHM decision-making on the operation of engineering
816 systems. Yet, in spite of the recognition of the importance of uncertainty in PHM [155], work
817 is still needed to concretely address the impact of uncertainty on the different PHM tasks and
to effectively manage it.

818

819 The challenge comes from the fact that there are different sources of uncertainty that
820 affect PHM, whose interactions are not fully understood and, thus, it is difficult to
821 systematically account for them in the PHM tasks. While some sources are internal to the
822 SSC, others are external, and all must be accounted for in the different activities of PHM.
823 There is aleatory uncertainty in the physical behavior of the SSC and epistemic uncertainty in
824 the model of it (developed based on sensors data or physic-based or based on a hybrid
825 combination of both data and physics) and the associated parameters. As mentioned earlier,
826 there is uncertainty in the sensors measurements and in their processing tools. For the
827 prognostic task of PHM, there is also uncertainty on the future SSC operation profile and state
evolution.

828

829 Given the relevance of uncertainty in the PHM tasks, it becomes necessary to develop
830 systematic frameworks for accounting for such uncertainty in practical applications, in order
831 to enable the robust verification and validation of the solutions developed, with respect to the
832 requirements for their use for decision-making and their contribution to the risk involved in
833 such decisions. Such frameworks must enable the systematic identification, representation,
834 quantification and propagation of the different sources of uncertainty, so that any PHM
835 outcome is provided also with its uncertainty, which needs to be considered for robust
decision-making [156].

836 Focusing specifically on data-driven methods for PHM, the challenge of quantifying
837 the uncertainty in PHM outcomes has rarely been addressed and mostly with ensemble
838 approaches, which can become computationally burdensome, and are highly dependent on
839 how the individual models are developed and how their outcomes are aggregated
840 [157][158][159][160][134][161]. Recently, Bayesian neural networks and variational
841 inference have been used in PHM, for accounting of uncertainty [162][163]. Also, the
842 combination of neural networks and gaussian processes are being considered as a promising
843 direction for providing PHM outcomes equipped with the needed estimates of the associated
844 uncertainty [164].

845

846 4. CONCLUSIONS

847

848 PHM has become a fashionable area of research and development, due to its promises
849 of enabling condition-based and predictive maintenance, which can be game-changers for the
850 production performance, reliability and safety of industrial businesses. Then, many academic
851 words have been and are developed, and several applications have been attempted, with a
852 more or less significant degree of success. These have been facilitated by the availability of
853 numerous and large data sets, of affordable computational hardware to train the models, of
854 freely available software to implement the models in a reliable and relatively straightforward
855 manner. Yet, quite some work still needs to be done to increase the significance of PHM
856 impacts on industry, due to a number of theoretical and practical issues that still require an
857 effective solution. These come from different perspectives, related to the physics of the
858 problem itself, the nature and type of data, the requirements of the solutions. As for the
859 physics of the problem, it is undoubtful that the SSCs degradation processes in practice are
860 most of the times quite complex and dependent on a large number of parameters and
861 mechanisms, which are dynamic and highly non-linear, and not completely known.

862 But much of the problem comes from the data and the extraction of informative
863 content for the fault detection, diagnostics and prognostics tasks of PHM. Managing and
864 treating the big condition-monitoring data collected by the sensors and comprised of a large
865 variety of heterogeneous signals is not an easy task and the data are often anomalous, scarce,
866 incomplete and unlabeled. Furthermore, they are collected under changing operational and
867 environmental conditions during the life of the SSC.

868 Surely, for the effectiveness of extracting informative content from data, undoubtedly
869 Deep Learning (DL) has contributed a great leap by incorporating feature engineering in the
870 process of learning of the models, for automatic processing of big and heterogeneous
871 condition monitoring data and extraction of features relevant for the application. Encouraging
872 results have been obtained already in fault detection and diagnostics, whereas Prognostics
873 remains still a challenge for DL.

874 Other of the above challenges are being addressed with sophisticated advancements
875 which need to be, then, effectively deployed in practice. These include: Recurrent Neural
876 Networks for PHM applications, and their transformation into images so as to exploit the
877 powerful methods of image processing (including the novel Convolutional Neural Networks
878 (CNNs), particularly for fault detection and diagnostics; signal reconstruction methods

879 (including Auto-Encoders) of unsupervised and semi-supervised learning for fault detection
880 and diagnostics, and for degradation state prediction, to cope with the frequent practical cases
881 of unlabeled data; Optimal Transport (OT) methods and unsupervised adaptation techniques
882 to cope with the problem that the test data distribution may be a different distribution (or
883 evolve to a different distribution) than that of the training data, with the consequence that the
884 trained data-learned model may perform poorly on the test data.

885 An issue of particular relevance for the prognostic task of PHM is the proper treatment
886 of the uncertainty in the data and, then, in the models. Several sources of uncertainty exist in
887 practice, as the models are inevitably only representations of the real relationships between
888 input and output, the measured data are inevitably noisy due to measurement errors, and the
889 future operational and environmental profiles of the SSCs are not known. All these
890 uncertainties affect the predictions of the future degradation and failure of the SSCs. With
891 respect to the uncertainty issue in PHM, frameworks are being developed for a probabilistic
892 treatment of the RUL of SSCs: given the potentially costly and catastrophic consequences
893 associated with the decisions that are made based on the PHM outcomes, it is absolutely
894 necessary to provide also estimates of the uncertainty alongside the predictions. For
895 example, frameworks are being developed by Bayesian neural networks and deep gaussian
896 processes.

897 An issue which is arising with the data-driven models and algorithms used for PHM is
898 that they lack interpretability, which reduces trust in their use particularly for safety-critical
899 applications. This leads to the need to find ways for improving transparency and
900 interpretability for a clearer understanding of what the model predicts and how, and finally for
901 building trust on its use. Methods for injecting physical information in learning models, post-
902 hoc sensitivity approaches and visualization techniques are being studied to provide
903 interpretability from different perspectives, including explaining the learned input-output
904 relation representations, explaining the individual model outputs, explaining the way the
905 output is produced by the model.

906 Strong concerns are also arising with respect to the security of PHM models for real
907 applications, in particular for safety-critical ones. PHM is increasingly used to support
908 maintenance decision-making processes in various high-value/high-risk industries, where the
909 wrong decision may result in serious consequences. The methods and models used perform
910 exchange and elaboration of data, and must then be secure to reject training data with negative
911 impacts on the results of decision-making.

912
913 **Acknowledgments**

914 The author is grateful to Dr. Mingjing Xu for helping in the preparation and revision of the
915 paper, and to the four referees whose comments and questions have helped to greatly improve
916 the work.

917

918 **REFERENCES**

- 919 [1] E. Zio, "Some challenges and opportunities in reliability engineering," *IEEE*
 920 *Transactions on Reliability*, 2016, doi: 10.1109/TR.2016.2591504.
- 921 [2] S. T. Kandukuri, A. Klausen, H. R. Karimi, and K. G. Robbersmyr, "A review of
 922 diagnostics and prognostics of low-speed machinery towards wind turbine farm-level
 923 health management," *Renewable and Sustainable Energy Reviews*. 2016, doi:
 924 10.1016/j.rser.2015.08.061.
- 925 [3] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A
 926 systematic review from data acquisition to RUL prediction," *Mechanical Systems and*
 927 *Signal Processing*. 2018, doi: 10.1016/j.ymssp.2017.11.016.
- 928 [4] M. Tahan, E. Tsoutsanis, M. Muhammad, and Z. A. Abdul Karim, "Performance-based
 929 health monitoring, diagnostics and prognostics for condition-based maintenance of gas
 930 turbines: A review," *Applied Energy*. 2017, doi: 10.1016/j.apenergy.2017.04.048.
- 931 [5] D. Wang, K. L. Tsui, and Q. Miao, "Prognostics and Health Management: A Review of
 932 Vibration Based Bearing and Gear Health Indicators," *IEEE Access*, 2017, doi:
 933 10.1109/ACCESS.2017.2774261.
- 934 [6] M. Compare, P. Baraldi, and E. Zio, "Challenges to IoT-Enabled Predictive
 935 Maintenance for Industry 4.0," *IEEE Internet of Things Journal*, 2020, doi:
 936 10.1109/JIOT.2019.2957029.
- 937 [7] C. L. Gan, "Prognostics and Health Management of Electronics: Fundamentals,
 938 Machine Learning, and the Internet of Things," *Life Cycle Reliability and Safety*
 939 *Engineering*, 2020, doi: 10.1007/s41872-020-00119-y.
- 940 [8] H. M. Elattar, H. K. Elminir, and A. M. Riad, "Prognostics: a literature review,"
 941 *Complex & Intelligent Systems*, 2016, doi: 10.1007/s40747-016-0019-3.
- 942 [9] I. K. Jennions, O. Niculita, and M. Esperon-Miguez, "Integrating IVHM and asset
 943 design," *International Journal of Prognostics and Health Management*, 2016.
- 944 [10] J. R. Dumargue, T., Pugeon, J. R., & Massé, "An approach to designing PHM systems
 945 with systems engineering," in *European conference of the Prognostics and health*
 946 *management society.*, 2016.
- 947 [11] M. Sharp and B. A. Weiss, "Hierarchical modeling of a manufacturing work cell to
 948 promote contextualized PHM information across multiple levels," *Manufacturing*
 949 *Letters*, 2018, doi: 10.1016/j.mfglet.2018.02.003.
- 950 [12] D. Han, J. Yu, Y. Song, D. Tang, and J. Dai, "A distributed autonomic logistics system
 951 with parallel-computing diagnostic algorithm for aircrafts," in *AUTOTESTCON*
 952 *(Proceedings)*, 2019, doi: 10.1109/AUTEST.2019.8878478.
- 953 [13] L. Yang, Q. Sun, and Z. S. Ye, "Designing mission abort strategies based on early-
 954 warning information: Application to UAV," *IEEE Transactions on Industrial*
 955 *Informatics*, 2020, doi: 10.1109/TII.2019.2912427.
- 956 [14] F. Di Maio, P. Turati, E. Z. PHM Society European Conference 2016, "Prediction
 957 capability assessment of data-driven prognostic methods for railway applications,"
 958 *phmsociety.org*.
- 959 [15] Z. Zeng, F. Di Maio, E. Zio, and R. Kang, "A hierarchical decision-making framework
 960 for the assessment of the prediction capability of prognostic methods," *J Risk and*
 961 *Reliability*, vol. 231, no. 1, pp. 36–52, Feb. 2017, doi: 10.1177/1748006X16683321.
- 962 [16] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, "Metrics for Offline Evaluation
 963 of Prognostic Performance," *International Journal of Prognostics and Health*
 964 *Management*, 2010.
- 965 [17] D. Kwon, M. R. Hodkiewicz, J. Fan, T. Shibutani, and M. G. Pecht, "IoT-Based
 966 Prognostics and Systems Health Management for Industrial Applications," *IEEE*
 967 *Access*, 2016, doi: 10.1109/ACCESS.2016.2587754.

- 968 [18] J. Siryani, B. Tanju, and T. J. Eveleigh, "A Machine Learning Decision-Support
969 System Improves the Internet of Things' Smart Meter Operations," *IEEE Internet of
970 Things Journal*, 2017, doi: 10.1109/JIOT.2017.2722358.
- 971 [19] L. Winnig, "GE's big bet on data and analytics," *Proc. MIT Sloan Manag. Rev.*, 2016.
- 972 [20] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge Computing: Vision and
973 Challenges," *IEEE Internet of Things Journal*, 2016, doi: 10.1109/JIOT.2016.2579198.
- 974 [21] K. Lin, J. Lu, C. S. Chen, J. Zhou, and M. T. Sun, "Unsupervised Deep Learning of
975 Compact Binary Descriptors," *IEEE Transactions on Pattern Analysis and Machine
976 Intelligence*, 2019, doi: 10.1109/TPAMI.2018.2833865.
- 977 [22] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng, and W. Jiang, "Unsupervised fault
978 diagnosis of rolling bearings using a deep neural network based on generative
979 adversarial networks," *Neurocomputing*, 2018, doi: 10.1016/j.neucom.2018.07.034.
- 980 [23] L. Chen, G. Xu, Y. Wang, and J. Wang, "Detection of weak transient signals based on
981 unsupervised learning for bearing fault diagnosis," *Neurocomputing*, 2018, doi:
982 10.1016/j.neucom.2018.07.004.
- 983 [24] K. Goebel, *Prognostics: The Science of Making Predictions*, 1st ed. CreateSpace
984 Independent Publishing Platform, 2017.
- 985 [25] E. Zio, "Computational methods for reliability and risk analysis," in *Computational
986 methods for reliability and risk analysis*, World Scientific Publishing Company, 2009.
- 987 [26] A. S. Sekhar, "Model-based identification of two cracks in a rotor system," *Mechanical
988 Systems and Signal Processing*, 2004, doi: 10.1016/S0888-3270(03)00041-4.
- 989 [27] S. L. Jeong Haedong, Bumsoo Park, Seungtae Park, Hyungcheol Min, "Fault detection
990 and identification method using observer-based residuals," *Reliability Engineering &
991 System Safety*, vol. 184, pp. 27–40, 2019.
- 992 [28] Duan, Chaoqun, Viliam Makis, and Chao Deng. "A two-level Bayesian early fault
993 detection for mechanical equipment subject to dependent failure modes." *Reliability
994 Engineering & System Safety* 193 (2020): 106676.
- 995 [29] H. P. Wan and Y. Q. Ni, "Bayesian multi-task learning methodology for reconstruction
996 of structural health monitoring data," *Structural Health Monitoring*, 2019, doi:
997 10.1177/1475921718794953.
- 998 [30] J. A. A. J. et al. Quintanilha, Igor M., Vitor RM Elias, Felipe B. da Silva, Pedro AM
999 Fonini, Eduardo AB da Silva, Sergio L. Netto, "A fault detector/classifier for closed-
1000 ring power generators using machine learning," *Reliability Engineering & System
1001 Safety*, no. 107614, 2021.
- 1002 [31] M. Marseguerra, "Early detection of gradual concept drifts by text categorization and
1003 Support Vector Machine techniques: The TRIO algorithm," *Reliability Engineering &
1004 System Safety*, vol. 129, pp. 1–9, 2014.
- 1005 [32] and E. P. Tolo, Silvia, Xiang Tian, Nils Bausch, Victor Becerra, T. V. Santhosh,
1006 Gopika Vinod, "Robust on-line diagnosis tool for the early accident detection in
1007 nuclear power plants," *Reliability Engineering & System Safety*, vol. 186, pp. 110–119,
1008 2019.
- 1009 [33] Hu, Q. P., Min Xie, Szu Hui Ng, and Gregory Levitin. "Robust recurrent neural
1010 network modeling for software fault detection and correction prediction." *Reliability
1011 Engineering & System Safety* 92, no. 3 (2007): 332-340.
- 1012 [34] Marugán, Alberto Pliego, Ana María Peco Chacón, and Fausto Pedro García Márquez.
1013 "Reliability analysis of detecting false alarms that employ neural networks: A real case
1014 study on wind turbines." *Reliability Engineering & System Safety* 191 (2019): 106574.
- 1015 [35] J. Liu, Y. F. Li, and E. Zio, "A SVM framework for fault detection of the braking
1016 system in a high speed train," *Mechanical Systems and Signal Processing*, vol. 87, no.
1017 October 2016, pp. 401–409, 2017, doi: 10.1016/j.ymsp.2016.10.034.

- 1018 [36] M. Ahmed, A. Naser Mahmood, and J. Hu, "A survey of network anomaly detection
1019 techniques," *Journal of Network and Computer Applications*. 2016, doi:
1020 10.1016/j.jnca.2015.11.016.
- 1021 [37] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S. K. Ng, "MAD-GAN: Multivariate
1022 Anomaly Detection for Time Series Data with Generative Adversarial Networks," in
1023 *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial
1024 Intelligence and Lecture Notes in Bioinformatics)*, 2019, doi: 10.1007/978-3-030-
1025 30490-4_56.
- 1026 [38] H.-J. Xing and W.-T. Liu, "Robust AdaBoost based ensemble of one-class support
1027 vector machines," *Information Fusion*, vol. 55, no. August 2019, pp. 45–58, 2019, doi:
1028 10.1016/j.inffus.2019.08.002.
- 1029 [39] M. Xu, P. Baraldi, X. Lu, F. Cannarile, and E. Zio, "Anomaly Detection for Industrial
1030 Systems using Generative Adversarial Networks," in *4th International Conference on
1031 System Reliability and Safety (ICSRS 2019)*, 2019.
- 1032 [40] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class AdaBoost," *Statistics and Its
1033 Interface*, 2009, doi: 10.4310/sii.2009.v2.n3.a8.
- 1034 [41] S. Joe Qin, "Data-driven fault detection and diagnosis for complex industrial
1035 processes," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 2009, doi:
1036 10.3182/20090630-4-ES-2003.0408.
- 1037 [42] J. Yu, J. Yoo, J. Jang, J. H. Park, and S. Kim, "A novel hybrid of auto-associative
1038 kernel regression and dynamic independent component analysis for fault detection in
1039 nonlinear multimode processes," *Journal of Process Control*, 2018, doi:
1040 10.1016/j.jprocont.2018.05.004.
- 1041 [43] F. Di Maio, P. Baraldi, E. Zio, and R. Seraoui, "Fault detection in nuclear power plants
1042 components by a combination of statistical methods," *IEEE Transactions on
1043 Reliability*, 2013, doi: 10.1109/TR.2013.2285033.
- 1044 [44] P. Baraldi, F. Di Maio, P. Turati, and E. Zio, "Robust signal reconstruction for
1045 condition monitoring of industrial components via a modified Auto Associative Kernel
1046 Regression method," *Mechanical Systems and Signal Processing*, 2015, doi:
1047 10.1016/j.ymsp.2014.09.013.
- 1048 [45] S. Li and J. Wen, "A model-based fault detection and diagnostic methodology based on
1049 PCA method and wavelet transform," *Energy and Buildings*, 2014, doi:
1050 10.1016/j.enbuild.2013.08.044.
- 1051 [46] K. Yan, Z. Ji, and W. Shen, "Online fault detection methods for chillers combining
1052 extended kalman filter and recursive one-class SVM," *Neurocomputing*, 2017, doi:
1053 10.1016/j.neucom.2016.09.076.
- 1054 [47] P. Bangalore, S. Letzgus, D. Karlsson, and M. Patriksson, "An artificial neural
1055 network-based condition monitoring method for wind turbines, with application to the
1056 monitoring of the gearbox," *Wind Energy*, 2017, doi: 10.1002/we.2102.
- 1057 [48] P. F. Odgaard, B. Lin, and S. B. Jorgensen, "Observer and data-driven-model-based
1058 fault detection in power plant coal mills," *IEEE Transactions on Energy Conversion*,
1059 2008, doi: 10.1109/TEC.2007.914185.
- 1060 [49] C. Yang, J. Liu, Y. Zeng, and G. Xie, "Real-time condition monitoring and fault
1061 detection of components based on machine-learning reconstruction model," *Renewable
1062 Energy*, 2019, doi: 10.1016/j.renene.2018.10.062.
- 1063 [50] L. E. Mujica, J. Rodellar, A. Fernández, and A. Güemes, "Q-statistic and t2-statistic
1064 pca-based measures for damage assessment in structures," *Structural Health
1065 Monitoring*, 2011, doi: 10.1177/1475921710388972.

- 1066 [51] H.-B. Huang, T.-H. Yi, and H.-N. Li, "Sensor Fault Diagnosis for Structural Health
1067 Monitoring Based on Statistical Hypothesis Test and Missing Variable Approach,"
1068 *Journal of Aerospace Engineering*, 2017, doi: 10.1061/(asce)as.1943-5525.0000572.
- 1069 [52] S. Seo and P. D. Gary M. Marsh, "A review and comparison of methods for detecting
1070 outliers in univariate data sets," *Department of Biostatistics, Graduate School of Public
1071 Health*, 2006.
- 1072 [53] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and
1073 prognostics implementing condition-based maintenance," *Mechanical Systems and
1074 Signal Processing*, vol. 20, no. 7. pp. 1483–1510, 2006, doi:
1075 10.1016/j.ymssp.2005.09.012.
- 1076 [54] B. Wang, D. Tang, Q. Yue, J. Zhou, and N. Deonauth, "Study on nonlinear dynamic
1077 characteristics inherent in offshore jacket platform using long-term monitored response
1078 of ice-structure interaction," *Applied Ocean Research*, 2018, doi:
1079 10.1016/j.apor.2017.12.009.
- 1080 [55] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Fault diagnosis of ball bearings using
1081 continuous wavelet transform," *Applied Soft Computing*, vol. 11, no. 2, pp. 2300–2312,
1082 2011, doi: 10.1016/j.asoc.2010.08.011.
- 1083 [56] A. S. Raj and N. Murali, "Early classification of bearing faults using morphological
1084 operators and fuzzy inference," *IEEE Transactions on Industrial Electronics*, 2013,
1085 doi: 10.1109/TIE.2012.2188259.
- 1086 [57] W. Caesarendra and T. Tjahjowidodo, "A review of feature extraction methods in
1087 vibration-based condition monitoring and its application for degradation trend
1088 estimation of low-speed slew bearing," *Machines*. 2017, doi:
1089 10.3390/machines5040021.
- 1090 [58] S. Kolouri, S. R. Park, M. Thorpe, D. Slepcev, and G. K. Rohde, "Optimal Mass
1091 Transport: Signal processing and machine-learning applications," *IEEE Signal
1092 Processing Magazine*, 2017, doi: 10.1109/MSP.2017.2695801.
- 1093 [59] P. Li, Q. Wang, and L. Zhang, "A novel earth mover's distance methodology for image
1094 matching with gaussian mixture models," in *Proceedings of the IEEE International
1095 Conference on Computer Vision*, 2013, doi: 10.1109/ICCV.2013.212.
- 1096 [60] J. Rabin, S. Ferradans, and N. Papadakis, "Adaptive color transfer with relaxed optimal
1097 transport," in *2014 IEEE International Conference on Image Processing, ICIP 2014*,
1098 2014, doi: 10.1109/ICIP.2014.7025983.
- 1099 [61] G. Montavon, K. R. Müller, and M. Cuturi, "Wasserstein training of restricted
1100 boltzmann machines," in *Advances in Neural Information Processing Systems*, 2016.
- 1101 [62] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel
1102 two-sample test," *Journal of Machine Learning Research*. 2012.
- 1103 [63] M. Cuturi and A. Doucet, "Fast computation of Wasserstein barycenters," in *31st
1104 International Conference on Machine Learning, ICML 2014*, 2014.
- 1105 [64] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proceedings of
1106 the IEEE International Conference on Computer Vision*, 2009, doi:
1107 10.1109/ICCV.2009.5459199.
- 1108 [65] A. Ramdas, N. G. Trillos, and M. Cuturi, "On wasserstein two-sample testing and
1109 related families of nonparametric tests," *Entropy*, 2017, doi: 10.3390/e19020047.
- 1110 [66] S. S. Y. Ng, J. Cabrera, P. W. T. Tse, A. H. Chen, and K. L. Tsui, "Distance-based
1111 analysis of dynamical systems reconstructed from vibrations for bearing diagnostics,"
1112 *Nonlinear Dynamics*, 2015, doi: 10.1007/s11071-014-1857-4.
- 1113 [67] S. Kammammettu and Z. Li, "Change point and fault detection using Kantorovich
1114 Distance," *Journal of Process Control*, 2019, doi: 10.1016/j.jprocont.2019.05.012.

- 1115 [68] B. Wang, P. Baraldi, X. Lu, and E. Zio, "Fault detection based on optimal transport
1116 theory," in *Proceedings of ESREL 2020 – PSAM 15*, 2020.
- 1117 [69] S. R. Park, S. Kolouri, S. Kundu, and G. K. Rohde, "The cumulative distribution
1118 transform and linear pattern classification," *Applied and Computational Harmonic
1119 Analysis*, 2018, doi: 10.1016/j.acha.2017.02.002.
- 1120 [70] M. Panda and P. M. Khilar, "Distributed soft fault detection algorithm in wireless
1121 sensor networks using statistical test," in *Proceedings of 2012 2nd IEEE International
1122 Conference on Parallel, Distributed and Grid Computing, PDGC 2012*, 2012, doi:
1123 10.1109/PDGC.2012.6449816.
- 1124 [71] D. Garcia-Alvarez, "Fault detection using Principal Component Analysis (PCA) in a
1125 Wastewater Treatment Plant (WWTP)," in *62th International Student's Scientific
1126 Conference*, 2009.
- 1127 [72] O. Fink, E. Zio, and U. Weidmann, "Predicting component reliability and level of
1128 degradation with complex-valued neural networks," *Reliability Engineering and
1129 System Safety*, vol. 121, pp. 198–206, Jan. 2014, doi: 10.1016/j.res.2013.08.004.
- 1130 [73] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health
1131 state classification," *Reliability Engineering and System Safety*, vol. 115, pp. 124–135,
1132 Jul. 2013, doi: 10.1016/j.res.2013.02.022.
- 1133 [74] K. Tang, D. J. Parsons, and S. Jude, "Comparison of automatic and guided learning for
1134 Bayesian networks to analyse pipe failures in the water distribution system," *Reliability
1135 Engineering and System Safety*, vol. 186, pp. 24–36, Jun. 2019, doi:
1136 10.1016/j.res.2019.02.001.
- 1137 [75] A. Arcos Jiménez, C. Q. Gómez Muñoz, and F. P. García Márquez, "Dirt and mud
1138 detection and diagnosis on a wind turbine blade employing guided waves and
1139 supervised learning classifiers," *Reliability Engineering and System Safety*, vol. 184,
1140 pp. 2–12, Apr. 2019, doi: 10.1016/j.res.2018.02.013.
- 1141 [76] M. M. Manjurul Islam and J. M. Kim, "Reliable multiple combined fault diagnosis of
1142 bearings using heterogeneous feature models and multiclass support vector Machines,"
1143 *Reliability Engineering and System Safety*, vol. 184, pp. 55–66, Apr. 2019, doi:
1144 10.1016/j.res.2018.02.012.
- 1145 [77] A. Robles-Velasco, P. Cortés, J. Muñuzuri, and L. Onieva, "Prediction of pipe failures
1146 in water supply networks using logistic regression and support vector classification,"
1147 *Reliability Engineering and System Safety*, vol. 196, p. 106754, Apr. 2020, doi:
1148 10.1016/j.res.2019.106754.
- 1149 [78] Y. Xu, Y. Sun, J. Wan, X. Liu, and Z. Song, "Industrial Big Data for Fault Diagnosis:
1150 Taxonomy, Review, and Applications," *IEEE Access*, 2017, doi:
1151 10.1109/ACCESS.2017.2731945.
- 1152 [79] P. Baraldi, F. Di Maio, M. Rigamonti, E. Zio, and R. Seraoui, "Unsupervised clustering
1153 of vibration signals for identifying anomalous conditions in a nuclear turbine," *Journal
1154 of Intelligent and Fuzzy Systems*, 2015, doi: 10.3233/IFS-141459.
- 1155 [80] R. Arn, P. Narayana, B. Draper, T. Emerson, M. Kirby, and C. Peterson, "Motion
1156 Segmentation via Generalized Curvatures," *IEEE Transactions on Pattern Analysis
1157 and Machine Intelligence*, 2018, doi: 10.1109/TPAMI.2018.2869741.
- 1158 [81] F. Cakir, K. He, S. A. Bargal, and S. Sclaroff, "Hashing with Mutual Information,"
1159 *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, doi:
1160 10.1109/tpami.2019.2914897.
- 1161 [82] Y. Zheng, S. Li, R. Yan, H. Tang, and K. C. Tan, "Sparse Temporal Encoding of
1162 Visual Features for Robust Object Recognition by Spiking Neurons," *IEEE
1163 Transactions on Neural Networks and Learning Systems*, 2018, doi:
1164 10.1109/TNNLS.2018.2812811.

- 1165 [83] C. Cao, Y. Huang, Y. Yang, L. Wang, Z. Wang, and T. Tan, "Feedback Convolutional
1166 Neural Network for Visual Localization and Segmentation," *IEEE Transactions on*
1167 *Pattern Analysis and Machine Intelligence*, 2019, doi: 10.1109/TPAMI.2018.2843329.
- 1168 [84] P. Baraldi, F. Di Maio, M. Rigamonti, E. Zio, and R. Seraoui, "Clustering for
1169 unsupervised fault diagnosis in nuclear turbine shut-down transients," *Mechanical*
1170 *Systems and Signal Processing*, vol. 58, pp. 160–178, 2015, doi:
1171 10.1016/j.ymsp.2014.12.018.
- 1172 [85] C. Bian, S. Yang, T. Huang, Q. Xu, J. Liu, and E. Zio, "Degradation state mining and
1173 identification for railway point machines," *Reliability Engineering and System Safety*,
1174 vol. 188, pp. 432–443, Aug. 2019, doi: 10.1016/j.res.2019.03.044.
- 1175 [86] M. Reder, N. Y. Yürüşen, and J. J. Melero, "Data-driven learning framework for
1176 associating weather conditions and wind turbine failures," *Reliability Engineering and*
1177 *System Safety*, vol. 169, pp. 554–569, Jan. 2018, doi: 10.1016/j.res.2017.10.004.
- 1178 [87] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new
1179 perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013,
1180 doi: 10.1109/TPAMI.2013.50.
- 1181 [88] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An Intelligent Fault Diagnosis Method
1182 Using Unsupervised Feature Learning Towards Mechanical Big Data," *IEEE*
1183 *Transactions on Industrial Electronics*, 2016, doi: 10.1109/TIE.2016.2519325.
- 1184 [89] W. Yang, Y. Shi, Y. Gao, L. Wang, and M. Yang, "Incomplete-data oriented multiview
1185 dimension reduction via sparse low-rank representation," *IEEE Transactions on Neural*
1186 *Networks and Learning Systems*, 2018, doi: 10.1109/TNNLS.2018.2828699.
- 1187 [90] M. Zhang, N. Wang, Y. Li, and X. Gao, "Deep Latent Low-Rank Representation for
1188 Face Sketch Synthesis," *IEEE Transactions on Neural Networks and Learning*
1189 *Systems*, 2019, doi: 10.1109/TNNLS.2018.2890017.
- 1190 [91] C. Li, M. Z. Zia, Q. H. Tran, X. Yu, G. D. Hager, and M. Chandraker, "Deep
1191 Supervision with Intermediate Concepts," *IEEE Transactions on Pattern Analysis and*
1192 *Machine Intelligence*, 2019, doi: 10.1109/TPAMI.2018.2863285.
- 1193 [92] O. Costilla-Reyes, R. Vera-Rodriguez, P. Scully, and K. B. Ozanyan, "Analysis of
1194 Spatio-Temporal Representations for Robust Footstep Recognition with Deep Residual
1195 Neural Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*,
1196 2019, doi: 10.1109/TPAMI.2018.2799847.
- 1197 [93] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for
1198 industrial process monitoring," *IEEE Transactions on Industrial Electronics*. 2014, doi:
1199 10.1109/TIE.2014.2301773.
- 1200 [94] H. Zhu, L. Lu, J. Yao, S. Dai, and Y. Hu, "Fault diagnosis approach for photovoltaic
1201 arrays based on unsupervised sample clustering and probabilistic neural network
1202 model," *Solar Energy*, 2018, doi: 10.1016/j.solener.2018.10.054.
- 1203 [95] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An Intelligent Fault Diagnosis Method
1204 Using Unsupervised Feature Learning Towards Mechanical Big Data," *IEEE*
1205 *Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016, doi:
1206 10.1109/TIE.2016.2519325.
- 1207 [96] T. Han, C. Liu, L. Wu, S. Sarkar, and D. Jiang, "An adaptive spatiotemporal feature
1208 learning approach for fault diagnosis in complex systems," *Mechanical Systems and*
1209 *Signal Processing*, 2019, doi: 10.1016/j.ymsp.2018.07.048.
- 1210 [97] Xu, M., P. Baraldi, and E. Zio. "Fault diagnostics by conceptors-aided clustering." *30th*
1211 *European Safety and Reliability Conference, ESREL 2020 and 15th Probabilistic Safety*
1212 *Assessment and Management Conference, PSAM 2020.*

- 1213 [98] M. Lukoševičius and H. Jaeger, "Reservoir computing approaches to recurrent neural
1214 network training," *Computer Science Review*, vol. 3, no. 3, pp. 127–149, 2009, doi:
1215 10.1016/j.cosrev.2009.03.005.
- 1216 [99] H. Jaeger, "Using Conceptors to Manage Neural Long-Term Memories for Temporal
1217 Patterns," *Machine Learning*, vol. 18, pp. 1–43, 2016.
- 1218 [100] Qian, Guangwu, and Lei Zhang. "A simple feedforward convolutional conceptor neural
1219 network for classification." *Applied Soft Computing* 70 (2018): 1034-1041.
- 1220 [101] B. Samanta and K. R. Al-Balushi, "Artificial neural network based fault diagnostics of
1221 rolling element bearings using time-domain features," *Mechanical Systems and Signal
1222 Processing*, vol. 17, no. 2, pp. 317–328, 2003, doi: 10.1006/mssp.2001.1462.
- 1223 [102] S. Sankararaman and K. Goebel, "Uncertainty in Prognostics and Systems Health
1224 Management," *International Journal of Prognostics and Health Management*, 2015,
1225 doi: 10.36001/ijphm.2015.v6i4.2319.
- 1226 [103] Daigle, Matthew J., and Kai Goebel. "A model-based prognostics approach applied to
1227 pneumatic valves." *International journal of prognostics and health management* 2, no.
1228 2 (2011): 84-99.
- 1229 [104] J. Zhang and J. Lee, "A review on prognostics and health monitoring of Li-ion
1230 battery," *Journal of Power Sources*. 2011, doi: 10.1016/j.jpowsour.2011.03.101.
- 1231 [105] J. Liu and E. Zio, "System dynamic reliability assessment and failure prognostics,"
1232 *Reliability Engineering and System Safety*, vol. 160, pp. 21–36, Apr. 2017, doi:
1233 10.1016/j.ress.2016.12.003.
- 1234 [106] J. M. W. Brownjohn, A. de Stefano, Y. L. Xu, H. Wenzel, and A. E. Aktan, "Vibration-
1235 based monitoring of civil infrastructure: Challenges and successes," *Journal of Civil
1236 Structural Health Monitoring*, 2011, doi: 10.1007/s13349-011-0009-5.
- 1237 [107] X. S. Si, W. Wang, C. H. Hu, and D. H. Zhou, "Remaining useful life estimation - A
1238 review on the statistical data driven approaches," *European Journal of Operational
1239 Research*. 2011, doi: 10.1016/j.ejor.2010.11.018.
- 1240 [108] Y. Liu, X. Hu, and W. Zhang, "Remaining useful life prediction based on health index
1241 similarity," *Reliability Engineering and System Safety*, vol. 185, pp. 502–510, May
1242 2019, doi: 10.1016/j.ress.2019.02.002.
- 1243 [109] X. Li, W. Zhang, and Q. Ding, "Deep learning-based remaining useful life estimation
1244 of bearings using multi-scale feature extraction," *Reliability Engineering and System
1245 Safety*, vol. 182, pp. 208–218, Feb. 2019, doi: 10.1016/j.ress.2018.11.011.
- 1246 [110] X. Li, Q. Ding, and J. Q. Sun, "Remaining useful life estimation in prognostics using
1247 deep convolution neural networks," *Reliability Engineering and System Safety*, 2018,
1248 doi: 10.1016/j.ress.2017.11.021.
- 1249 [111] R. He, Y. Dai, J. Lu, and C. Mou, "Developing ladder network for intelligent
1250 evaluation system: Case of remaining useful life prediction for centrifugal pumps,"
1251 *Reliability Engineering and System Safety*, vol. 180, pp. 385–393, Dec. 2018, doi:
1252 10.1016/j.ress.2018.08.010.
- 1253 [112] A. Listou Ellefsen, E. Bjørlykhaug, V. Æsøy, S. Ushakov, and H. Zhang, "Remaining
1254 useful life predictions for turbofan engine degradation using semi-supervised deep
1255 architecture," *Reliability Engineering and System Safety*, vol. 183, pp. 240–251, Mar.
1256 2019, doi: 10.1016/j.ress.2018.11.027.
- 1257 [113] K. T. P. Nguyen and K. Medjaher, "A new dynamic predictive maintenance framework
1258 using deep learning for failure prognostics," *Reliability Engineering and System Safety*,
1259 vol. 188, pp. 251–262, Aug. 2019, doi: 10.1016/j.ress.2019.03.018.
- 1260 [114] Z. Shi and A. Chehade, "A dual-LSTM framework combining change point detection
1261 and remaining useful life prediction," *Reliability Engineering and System Safety*, vol.
1262 205, p. 107257, Jan. 2021, doi: 10.1016/j.ress.2020.107257.

- 1263 [115] J. Chen, H. Jing, Y. Chang, and Q. Liu, "Gated recurrent unit based recurrent neural
1264 network for remaining useful life prediction of nonlinear deterioration process,"
1265 *Reliability Engineering and System Safety*, vol. 185, pp. 372–382, May 2019, doi:
1266 10.1016/j.ress.2019.01.006.
- 1267 [116] P. J. García Nieto, E. García-Gonzalo, F. Sánchez Lasheras, and F. J. De Cos Juez,
1268 "Hybrid PSO-SVM-based method for forecasting of the remaining useful life for
1269 aircraft engines and evaluation of its reliability," *Reliability Engineering and System
1270 Safety*, vol. 138, pp. 219–231, Jun. 2015, doi: 10.1016/j.ress.2015.02.001.
- 1271 [117] Z. Zhao, Bin Liang, X. Wang, and W. Lu, "Remaining useful life prediction of aircraft
1272 engine based on degradation pattern learning," *Reliability Engineering and System
1273 Safety*, vol. 164, pp. 74–83, Aug. 2017, doi: 10.1016/j.ress.2017.02.007.
- 1274 [118] M. Baptista, E. M. P. Henriques, I. P. P. de Medeiros, J. P. P. Malere, C. L. L.
1275 Nascimento, and H. Prendinger, "Remaining useful life estimation in aeronautics:
1276 Combining data-driven and Kalman filtering," *Reliability Engineering and System
1277 Safety*, vol. 184, pp. 228–239, Apr. 2019, doi: 10.1016/j.ress.2018.01.017.
- 1278 [119] W. Ahmad, S. A. Khan, M. M. M. Islam, and J. M. Kim, "A reliable technique for
1279 remaining useful life estimation of rolling element bearings using dynamic regression
1280 models," *Reliability Engineering and System Safety*, vol. 184, pp. 67–76, Apr. 2019,
1281 doi: 10.1016/j.ress.2018.02.003.
- 1282 [120] Y. Chang and H. Fang, "A hybrid prognostic method for system degradation based on
1283 particle filter and relevance vector machine," *Reliability Engineering and System
1284 Safety*, vol. 186, pp. 51–63, Jun. 2019, doi: 10.1016/j.ress.2019.02.011.
- 1285 [121] A. Downey, Y. H. Lui, C. Hu, S. Laflamme, and S. Hu, "Physics-based prognostics of
1286 lithium-ion battery using non-linear least squares with dynamic bounds," *Reliability
1287 Engineering and System Safety*, vol. 182, pp. 1–12, Feb. 2019, doi:
1288 10.1016/j.ress.2018.09.018.
- 1289 [122] I. Eekhout, R. M. de Boer, J. W. R. Twisk, H. C. W. de Vet, and M. W. Heymans,
1290 "Missing data: a systematic review of how they are reported and handled.,"
1291 *Epidemiology (Cambridge, Mass.)*, 2012, doi: 10.1097/EDE.0b013e3182576cdb.
- 1292 [123] M. Ranjbar, P. Moradi, M. Azami, and M. Jalili, "An imputation-based matrix
1293 factorization method for improving accuracy of collaborative filtering systems,"
1294 *Engineering Applications of Artificial Intelligence*, 2015, doi:
1295 10.1016/j.engappai.2015.08.010.
- 1296 [124] Y. S. CH Cheng, CP Chan, "A novel purity-based k nearest neighbors imputation
1297 method and its application in financial distress prediction," *Engineering Applications of
1298 Artificial Intelligence*, vol. 81, pp. 283–299, 2019.
- 1299 [125] R. Razavi-Far, S. Chakrabarti, M. Saif, and E. Zio, "An integrated imputation-
1300 prediction scheme for prognostics of battery data with missing observations," *Expert
1301 Systems with Applications*, 2019, doi: 10.1016/j.eswa.2018.08.033.
- 1302 [126] A. N. Baraldi and C. K. Enders, "An introduction to modern missing data analyses,"
1303 *Journal of School Psychology*, 2010, doi: 10.1016/j.jsp.2009.10.001.
- 1304 [127] A. R. T. Donders, G. J. M. G. van der Heijden, T. Stijnen, and K. G. M. Moons,
1305 "Review: A gentle introduction to imputation of missing values," *Journal of Clinical
1306 Epidemiology*, 2006, doi: 10.1016/j.jclinepi.2006.01.014.
- 1307 [128] D. Vergouw *et al.*, "The search for stable prognostic models in multiple imputed data
1308 sets," *BMC Medical Research Methodology*, 2010, doi: 10.1186/1471-2288-10-81.
- 1309 [129] J. Honaker and G. King, "What to do about missing values in time-series cross-section
1310 data," *American Journal of Political Science*, vol. 54, no. 2, pp. 561–581, 2010, doi:
1311 10.1111/j.1540-5907.2010.00447.x.

- 1312 [130] H. Xiao, D. Huang, Y. Pan, Y. Liu, and K. Song, "Fault diagnosis and prognosis of
1313 wastewater processes with incomplete data by the auto-associative neural networks and
1314 ARMA model," *Chemometrics and Intelligent Laboratory Systems*, 2017, doi:
1315 10.1016/j.chemolab.2016.12.009.
- 1316 [131] U. Leturiondo, O. Salgado, L. Ciani, D. Galar, and M. Catelani, "Architecture for
1317 hybrid modelling and its application to diagnosis and prognosis with missing data,"
1318 *Measurement: Journal of the International Measurement Confederation*, 2017, doi:
1319 10.1016/j.measurement.2017.02.003.
- 1320 [132] A. Marshall, D. G. Altman, P. Royston, and R. L. Holder, "Comparison of techniques
1321 for handling missing covariate data within prognostic modelling studies: A simulation
1322 study," *BMC Medical Research Methodology*, 2010, doi: 10.1186/1471-2288-10-7.
- 1323 [133] T. G. Clark and D. G. Altman, "Developing a prognostic model in the presence of
1324 missing data: An ovarian cancer case study," *Journal of Clinical Epidemiology*, 2003,
1325 doi: 10.1016/S0895-4356(02)00539-5.
- 1326 [134] M. Xu, P. Baraldi, S. Al-Dahidi, and E. Zio, "Fault prognostics by an ensemble of Echo
1327 State Networks in presence of event based measurements," *Engineering Applications of
1328 Artificial Intelligence*, 2020, doi: 10.1016/j.engappai.2019.103346.
- 1329 [135] Z. Yang, K. Tang, and X. Yao, "Self-adaptive differential evolution with neighborhood
1330 search," in *2008 IEEE Congress on Evolutionary Computation, CEC 2008*, 2008, doi:
1331 10.1109/CEC.2008.4630935.
- 1332 [136] Yoon, K. Paul, and Ching-Lai Hwang. *Multiple attribute decision making: an
1333 introduction*. Sage publications, 1995.
- 1334 [137] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model
1335 predictions." In *Proceedings of the 31st international conference on neural information
1336 processing systems*, pp. 4768-4777. 2017.
- 1337 [138] Y. Chen, G. Peng, Z. Zhu, and S. Li, "A novel deep learning method based on attention
1338 mechanism for bearing remaining useful life prediction," *Applied Soft Computing
1339 Journal*, 2020, doi: 10.1016/j.asoc.2019.105919.
- 1340 [139] C. Liu, L. Zhang, J. Niu, R. Yao, and C. Wu, "Intelligent prognostics of machining
1341 tools based on adaptive variational mode decomposition and deep learning method with
1342 attention mechanism," *Neurocomputing*, 2020, doi: 10.1016/j.neucom.2020.06.116.
- 1343 [140] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should i trust you?' Explaining the
1344 predictions of any classifier," in *Proceedings of the ACM SIGKDD International
1345 Conference on Knowledge Discovery and Data Mining*, 2016, doi:
1346 10.1145/2939672.2939778.
- 1347 [141] P. R. Magesh, R. D. Myloth, and R. J. Tom, "An Explainable Machine Learning Model
1348 for Early Detection of Parkinson's Disease using LIME on DaTSCAN Imagery,"
1349 *Computers in Biology and Medicine*, 2020, doi: 10.1016/j.combiomed.2020.104041.
- 1350 [142] E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual
1351 predictions with feature contributions," *Knowledge and Information Systems*, 2014,
1352 doi: 10.1007/s10115-013-0679-x.
- 1353 [143] T. Aven and E. Zio, "Some considerations on the treatment of uncertainties in risk
1354 assessment for practical decision making," in *Reliability Engineering and System
1355 Safety*, 2011, doi: 10.1016/j.ress.2010.06.001.
- 1356 [144] J. M. P. King, N. J. Titchener-Hooker, and Y. Zhou, "Ranking bioprocess variables
1357 using global sensitivity analysis: A case study in centrifugation," *Bioprocess and
1358 Biosystems Engineering*, 2007, doi: 10.1007/s00449-006-0109-5.
- 1359 [145] A. Saltelli and S. Tarantola, "On the Relative Importance of Input Factors in
1360 Mathematical Models," *Journal of the American Statistical Association*, 2002, doi:
1361 10.1198/016214502388618447.

- 1362 [146] E. Borgonovo and E. Plischke, "Sensitivity analysis: A review of recent advances,"
1363 *European Journal of Operational Research*. 2016, doi: 10.1016/j.ejor.2015.06.032.
- 1364 [147] A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided Neural Networks
1365 (PGNN): An Application in Lake Temperature Modeling." arXiv preprint
1366 arXiv:1710.11431.
- 1367 [148] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural
1368 networks: A deep learning framework for solving forward and inverse problems
1369 involving nonlinear partial differential equations. *Journal of Computational Physics*,
1370 378, 686-707.
- 1371 [149] Tipireddy, R., Perdikaris, P., Stinis, P., & Tartakovsky, A. (2019). A comparative study
1372 of physics-informed neural network models for learning unknown dynamics and
1373 constitutive relations. arXiv preprint arXiv:1904.04058.
- 1374 [150] L. Von Rueden *et al.*, "Informed Machine Learning-A Taxonomy and Survey of
1375 Integrating Knowledge into Learning Systems." arXiv preprint arXiv:1903.12394.
- 1376 [151] J. H. Friedman, "Greedy function approximation: A gradient boosting machine,"
1377 *Annals of Statistics*, 2001, doi: 10.1214/aos/1013203451.
- 1378 [152] A. Goldstein, A. Kapelner, J. Bleich, and E. Pitkin, "Peeking Inside the Black Box:
1379 Visualizing Statistical Learning With Plots of Individual Conditional Expectation,"
1380 *Journal of Computational and Graphical Statistics*, 2015, doi:
1381 10.1080/10618600.2014.907095.
- 1382 [153] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical
1383 black-box attacks against Deep learning systems using adversarial examples," *ASIA
1384 CCS 2017 - Proceedings of the 2017 ACM Asia Conference on Computer and
1385 Communications Security*, 2016.
- 1386 [154] M. Barreno, B. Nelson, A. D. Joseph, and J. D. Tygar, "The security of machine
1387 learning," *Machine Learning*, 2010, doi: 10.1007/s10994-010-5188-5.
- 1388 [155] S. Sankararaman, S. Mahadevan, and M. E. Orchard, "Uncertainty in PHM,"
1389 *International Journal of Prognostics and Health Management*, vol. 6, no. 4, Nov.
1390 2020, doi: 10.36001/ijphm.2015.v6i4.2289.
- 1391 [156] R. Flage, T. Aven, E. Zio, and P. Baraldi, "Concerns, Challenges, and Directions of
1392 Development for the Issue of Representing Uncertainty in Risk Assessment," *Risk
1393 Analysis*, vol. 34, no. 7, pp. 1196–1207, 2014, doi: 10.1111/risa.12247.
- 1394 [157] P. Baraldi, R. Razavi-Far, E. Z.-R. E. & S. Safety, and undefined 2011, "Classifier-
1395 ensemble incremental-learning procedure for nuclear transient identification at different
1396 operational conditions," *Reliability Engineering & System Safety* 96.4 (2011): 480-488.
- 1397 [158] J. Liu, E. Z.-A. S. Computing, and undefined 2016, "A SVR-based ensemble approach
1398 for drifting data streams with recurring patterns," *Applied Soft Computing* 47 (2016):
1399 553-564.
- 1400 [159] S. Al-Dahidi, F. Di Maio, P. Baraldi, and E. Zio, "A locally adaptive ensemble
1401 approach for data-driven prognostics of heterogeneous fleets," *Proceedings of the
1402 Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 231,
1403 no. 4, pp. 350–363, 2017, doi: 10.1177/1748006X17693519.
- 1404 [160] Zhang, D., Baraldi, P., Cadet, C., Yousfi-Steiner, N., Bérenguer, C., & Zio, E. (2019).
1405 An ensemble of models for integrating dependent sources of information for the
1406 prognosis of the remaining useful life of proton exchange membrane fuel cells.
1407 *Mechanical Systems and Signal Processing*, 124, 479-501.
- 1408 [161] Deng, Y., Di Bucchianico, A., & Pechenizkiy, M. (2020). Controlling the accuracy and
1409 uncertainty trade-off in RUL prediction with a surrogate Wiener propagation model.
1410 *Reliability Engineering & System Safety*, 196, 106727.

- 1411 [162] Peng, W., Ye, Z. S., & Chen, N. (2019). Bayesian deep-learning-based health
1412 prognostics toward prognostics uncertainty. *IEEE Transactions on Industrial*
1413 *Electronics*, 67(3), 2283-2293.
- 1414 [163] Benker, M., Furtner, L., Semm, T., & Zaeh, M. F. (2020). Utilizing uncertainty
1415 information in remaining useful life estimation via Bayesian neural networks and
1416 Hamiltonian Monte Carlo. *Journal of Manufacturing Systems*.
- 1417 [164] Biggio, L., Wieland, A., Chao, M. A., Kastanis, I., & Fink, O. (2021). Uncertainty-
1418 aware Remaining Useful Life predictor. *arXiv preprint arXiv:2104.03613*.
1419
1420