Chapter 4
Harnessing AI for Reliability and Maintenance

Pierre Dersin

Introduction

To keep the key critical systems of today's complex world operating smoothly and cost-effectively, reliability and asset management have become priorities for industry and governments. Accordingly, more than ever, close attention is paid to ensuring reliability performance and optimizing maintenance and asset management policies.

The last two decades have witnessed the triumph of the digital transformation and, in particular, the "Internet of things" which, through the combination of cost-effective sensors and efficient communication infrastructure, allows many industrial items of equipment to communicate in real time physical magnitudes that can be used to estimate their health condition.

This evolution, combined with the fast-paced development of advanced data processing algorithms ("analytics"), including the vast area called artificial intelligence (AI), has spurred the emergence of a discipline named Prognostics and Health Management (PHM). At the same time, it is revolutionizing reliability engineering.

Back in the 1940s until the 1960s, reliability engineering evolved as a full-fledged scientific discipline under the pressure of the Cold War and the space race- and, not surprisingly, many prominent actors were found in the United States (e.g., Barlow and Proschan 1965) and the Soviet Union (e.g. Gnedenko). They built on important pre-WWII work such as that of W. Weibull in Sweden (Weibull 1939), as well as the foundations of reliability-related statistics by the likes of Fisher (1922) and Cox (1972a). A body of methodologies was soon constituted—necessarily, with simplifying assumptions.

P. Dersin (✉)
Eumetry SaS, Louveciennes, France
e-mail: pierre.dersin@ltu.se

Luleå University of Technology, Luleå, Sweden

P. V. Varde et al. (eds.), *Advances in Risk-Informed Technologies*, Risk, Reliability and Safety Engineering, https://doi.org/10.1007/978-981-99-9122-8_4

The situation in those days is best characterized by scarcity of data, and primitive computation tools. Indeed the beginnings of reliability theory were contemporaneous with the first computers, which utilized vacuum tubes and occupied an entire room. As late as the late 1970s, computer programs—even in a place like MIT-still had to be typed on punch cards and brought to an intimidating "computer center" which would (in the best case) deliver the outputs the following day. Many countries still had manually operated analog telephone exchanges—so much for telecommunication. Under those circumstances, emphasis was placed on simplified models: statistical independence, stationarity, exponential time to failures, and the like. Metrics of interest addressed average characteristics at population level, such as MTTF or MTBF.

One looks back in awe at the incredible technological successes which were achieved under those conditions—including landing the first man on the Moon, in 1969. In parallel, maintenance practice has evolved from 'run to failure" purely corrective maintenance to preventive maintenance and, in the 1960s, spurred by the aeronautics industry, the RCM (Reliability-Centered Maintenance) methodology which links reliability engineering (or, more precisely, RAMS) with maintenance needs, and stresses the notion of functional maintenance, i.e., maintenance plans determined by the need to keep fulfilling the function. In that context, condition-based maintenance increasingly has become part of the landscape.

In the 1950s, the expression "artificial intelligence" (AI) appeared, with the 1956 Dartmouth seminar organized by J. Mc Carthy. In its most general definition, AI aims at replicating human reasoning automatically. The first direction of investigation was symbolic AI, whereby special languages (such as LISP) were created to manipulate symbolic logic. This approach had limited success in the form of rule-based expert systems, for medical diagnostics for instance. But the unreasonable expectations they had raised were crushed, and led to disillusionment in the 1980s.

In parallel, another approach was pursued, with the research on *machine learning*, which can be described as the field of study that gives computers the ability to learn without being explicitly programmed (see e.g., Alpaydin 2014).While, traditionally, computers are given a model and inputs, and apply the explicit instructions of the model to the inputs in order to generate outputs, instead, with machine learning, the computer is given inputs and outputs and is asked to find a model that could have generated the given outputs from the given inputs. Once it has found that model, it can then apply it to new inputs.

Today, AI today is often understood to mean ML but actually ML is just a subset of AI (Fig. 4.1).

Numerous machine learning methods have appeared, which essentially fall into two main categories: supervised learning, whereby the computer is trained on a set of 'labelled" input data and is given outputs corresponding to each input data set (for instance, the output word "cat" corresponds to input images of cats); and unsupervised learning, whereby the computer is just given input data and has to detect patterns somehow (for instance, through clustering). An important and growing area is also reinforcement learning, which provides feedback so that 'good 'decisions

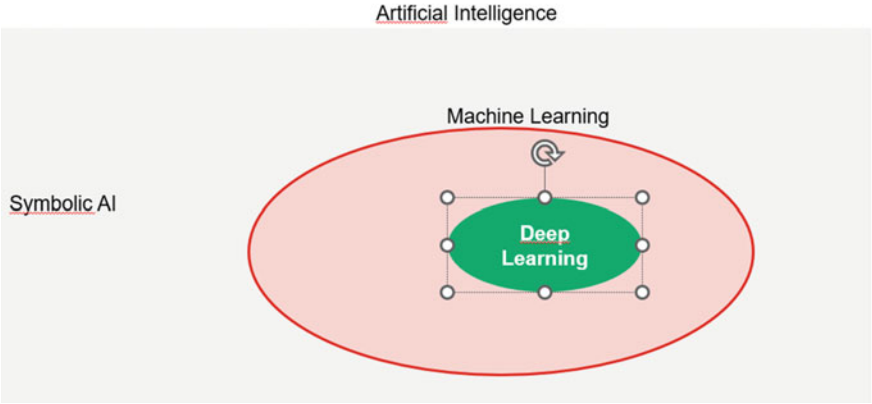


Fig. 4.1 Artificial intelligence (AI)

are rewarded and bad decisions are penalized, and in that way the algorithm learns from experience and improves over time. The catalog of ML method is huge and growing; some of the better known ones include support vector machines (SVM), KNN (K-nearest neighbors), various regression methods, random forests, ensemble methods, etc.

Generally, in PHM, signals acquired by sensors that monitor an asset are summarized by so-called ‘features’, which form the basis for constructing health indicators and performing anomaly detection, fault diagnostics, and if possible prognostics. Features are traditionally engineered by domain experts but, with ML, there are methods that permit automatic feature learning.

One family of methods that has had great success recently, i.e. in the last decade or so, after knowing its ups and downs since the 1950s (the “down” periods have been called “AI winters”) is that of artificial neural networks (ANN) (Fig. 4.2). ANNs have been inspired by biological neural networks. They incorporate two fundamental components: neurons (represented by nodes) and synapses (represented by links). Each node computes a nonlinear function of a weighted sum of inputs. The choice of the nonlinear functions is part of the “network architecture”. The values of the weights result from an optimization. For instance, in the supervised learning case, the weights are determined from a set of inputs in order to minimize some distance between the generated outputs and the target outputs. In general, a “loss function” is minimized to determine the best weights.

The ability of neural networks to learn any nonlinear function is characterized by “universal approximation theorems” (Hornik et al. 1989). This echoes somehow the intuition of MIT mathematician Norbert Wiener who, in one of his epochal books (Wiener 1964), wrote that “a learning machine operates with nonlinear feedback”.

When the neural network contains more than one inner layer, it is called a deep network, and machine learning using such a network is called *deep learning*. Various neural network architectures have been introduced and used successfully. For instance the convolutional neural network (CNN) is ideal for image processing

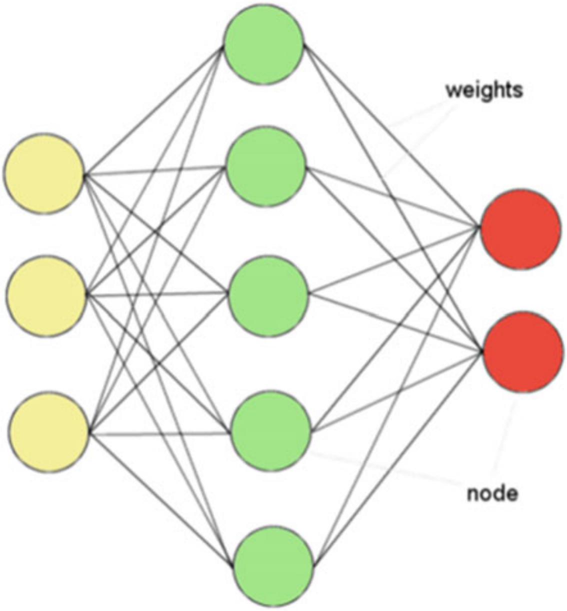


Fig. 4.2 Artificial neural network

in grids. Recurrent neural networks (RNN) include feedback loops and are quite suitable for automatic language translation for instance, as they can keep the memory of sequences of events. Recently, graph neural networks (GNN) have been successfully used for problems which can be framed in a graph structure.

Although the concept of neural network goes back to nearly three quarters of a century, their recent success (largely after 2012) is explained by the following trends (Fink et al. 2020):

- Advanced efficient algorithms, such as backpropagation. Many are now available freely on the Internet;
- The availability of huge amounts of data;
- The availability of affordable high-performance hardware (such as graphics processing units, GPU, for parallel computing) which makes it possible to process huge amounts of data very fast.

As a result, nowadays AI/machine learning and in particular deep learning, is permeating many fields of human activity. Particularly impressive have been the recent successes in image processing (the ImageNet, a database with more than one million labelled images), in machine translation, and, most recently, in Large Language Models (the ubiquitous Chat GPT and its competitors). The progresses in industry have been so far a little slower due to the particular challenges that need to be met (see e.g., Karim et al. 2023), but it is only a matter of time before industrial AI becomes widespread. In particular, since a survey paper on "potential and opportunities of deep learning for PHM" (Fink et al. 2020) appeared in 2020,

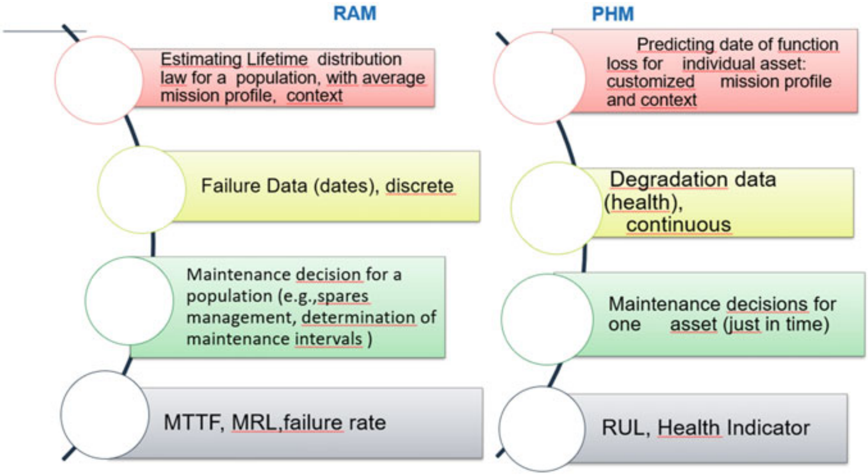

RAM
Estimating Lifetime distribution law for a population, with average mission profile, context
Failure Data (dates), discrete
Maintenance decision for a population (e.g.,spares management, determination of maintenance intervals)
MTTF, MRL,failure rate
PHM
Predicting date of function loss for individual asset: customized mission profile and context
Degradation data (health), continuous
Maintenance decisions for one asset (just in time)
RUL, Health Indicator

Fig. 4.3 Classical RAM (reliability-availability-maintainability) versus PHM (prognostics & health management)

applications have been multiplying in various industrial fields including railways, aerospace and electric power generation and transmission.

What are the implications for reliability engineering?

As a general statement, one could say that what AI enables, which traditional approaches could not, is the consideration of individual items beyond just population averages, and taking into account very precisely all the context variables that influence an item's degradation and failures. Figure 4.3 illustrates in that respect the main differences between classical RAM (Reliability-Availability-Maintainability) and PHM. Also, there is a key difference between traditional statistics and machine learning, as expressed by Breiman ("the two cultures", Breiman 2001): traditional statistics necessarily postulate an a priori probability model to describe the data, while machine learning explores the data without a priori.

The "ML revolution" offers tremendous opportunities in reliability engineering, which have only barely begun to be exploited. The field of reliability engineering (or more generally RAMS) and PHM (prognostics & health management) both are benefitting from ML and are actually coalescing into one single discipline.

The subject is vast and, in a lecture such as this, our modest goal is to give a few examples to try and illustrate the great potential of AI-ML as applied to reliability engineering and maintenance; while, at the same time, stressing the need for continuity and complementarity between traditional methods and AI-based ones. Necessarily, a number of important aspects had to be overlooked, and this is by no means an exhaustive survey of "AI for Reliability Engineering and Maintenance".

Design for Reliability and Reliability Prediction

Design for reliability is an important engineering activity, which can benefit from AI. For instance, in dealing with locomotives that operate in harsh conditions, it is important to understand the impact of various operating conditions on mechanical stresses, so that those stresses can be reduced, and reliability improved. This application has been treated recently (Gauthier 2022) with pattern recognition algorithms based on muti-layer feed-forward networks which are able to evidence very strong correlations between certain physical variables and mechanical stresses, with very small error probabilities.

Those algorithms have then been implemented in on-board computers, so that the axle force is adapted in real time according to the sensed operational conditions, in order to limit the mechanical stresses.

In traditional reliability engineering, models have been devised to represent the impact of various operational conditions, or stresses, on reliability. It is well known that environmental factors such as temperature or humidity impact reliability; but also, given a particular environment, asset mission profile will play a role as well.

Those factors, sometimes called covariates, have been incorporated in proportional hazard models, the best-known one being the Cox model (Cox 1972b):

$$\lambda(t; S) = \lambda_0(t)e^{\sum_{i=1}^{i=n} \beta_i S_i} \tag{4.1}$$

where λ denotes the failure rate (sometimes called hazard rate), S_i $(i = 1 \ldots . n)$ denote the various stresses, and the coefficients β_i are to be estimated statistically.

In the Cox model (4.1), the failure rate is assumed to be proportional to a base failure rate $\lambda_0(t)$, and it further assumed that the coefficient of proportionality (i) is independent of time; (ii) depends linearly on the stresses. Those are of course simplifying assumptions, which are not necessarily verified in practice. For instance, they do not allow for modeling stresses that vary with time. Some more complex models have been introduced, whereby the coefficient can be a nonlinear function of the stresses, possibly time-dependent.

Recently, reliability researchers at Ford Motor Co. (Li et al. 2022) have introduced an AI-ML based method, inspired by machine translation. They have designed a special type of RNN, with an "attention mechanism"-the main idea behind which is to weigh all outputs of hidden states to dynamically highlight relevant features of the input data.

The goal is to exploit the ability of neural networks to learn highly complex, nonlinear functions. To do so, they adopt the viewpoint of translating time series, just as, in machine translation, a time series of written or spoken words in a given source language is translated into a time series of words in another, target language.

The source language here is the language describing asset status (i.e. the stresses, or observed features) at various points in time, and the target language is the language describing failure probability at various points in future, i.e. the reliability function, also called survival function (whose knowledge is equivalent to that of the failure rate). This is why the model is called a survival model, and it is a "deep survival" model because it relies on a deep (i.e. multi-layer) neural network.

This AI algorithm, call seq2surv2 (for "sequence-to-survival"), is able to make individual predictions on each asset based on that asset's individual exposure to stresses and operating conditions over time, which is much more powerful than the traditional reliability engineering approach which deals only with average population behaviors in static (i.e. non-time-varying) environments.

The learning scheme contains two aspects: feature extraction and survival function prediction. The architecture follows an encoder-decoder structure (see e.g., Doersch 2016); the encoder maps the input sequence to a latent state vector; the decoder generates the output sequence from the latent state vector. The method has been tested satisfactorily on the NASA C-MAPSS open data set (Arias et al. 2021), with excellent performance results. It must be emphasized though that the method is a "black box". At this stage, the predictions are not traced to identified failure modes or root cause analysis (such endeavors might be addressed in future).

Transfer Learning for Diagnostics and Prognostics

Data-driven machine learning models usually require (i) sufficiently many labeled data (so that the algorithms can be trained in a supervised way); (ii) identical distributions of data in the training set and the test set.

Particularly in industrial applications for diagnostics and prognostics, at least one of those two conditions is usually not fulfilled. The data on which an algorithm is trained often does not contain enough labelled data: concretely, a number of failure data have no identified root cause or failure code associated with them. Or, there are simply not enough failure data. Or the data on which the algorithms has been trained (the training set) are not really representative of the data to which it will be applied (the test set). Then, if some knowledge has been gained previously on systems that resemble the system under study, it is useful to transfer from those other systems whatever has been learned on them. This is the idea of transfer learning. For instance, one could transfer to one type of machine (e.g., an engine) what has been learned on a different machine (say, another engine type).Or, knowledge acquired in one context can be transferred to a different context: for instance, from the knowledge of a reliability function in accelerated stress conditions, derive the reliability function under standard, normal operating conditions. There are several methods for accomplishing transfer learning, an area of research which is progressing fast (see e.g., Yao et al. 2023) for an extensive survey).

In broad terms, they fall into the following three categories: (1) Model-based and parameter-based methods; (2) Feature-matching-based methods; (3) Adversarial adaptation methods.

(1) *Model-based and parameter-based methods*

Those methods take advantage of pre-trained models and adapt some of the model parameters to the new conditions. For instance, when the model is a deep neural network, weights from another application are used as initial weights, which cut the training time considerably. Thus the key idea is fine-tuning pre-trained parameters to the new application context.

(2) *Feature-matching-based Methods*

The key idea underlying those methods is to reduce the feature distribution difference across domains via feature transformation. The goal is to draw source-domain features and target-domain features closer to each other, so as to facilitate transfer (i.e. classification can be performed fairly easily in the target domain from features extracted from the data in the source domain).

There is a feature extraction step and a domain adaptation step.

(3) *Adversarial Adaptation Methods*

Those methods exploit a modified version of the General Adversarial Network (GAN) (Goodfellow et al. 2014). The generative network extracts features, and the discriminative network is used to tell the differences between source and target features. The goal is to learn features of one domain in such a way that the discriminator cannot distinguish them from features of the other domain. Domain-adversarial neural networks (DANN) belong to that category.

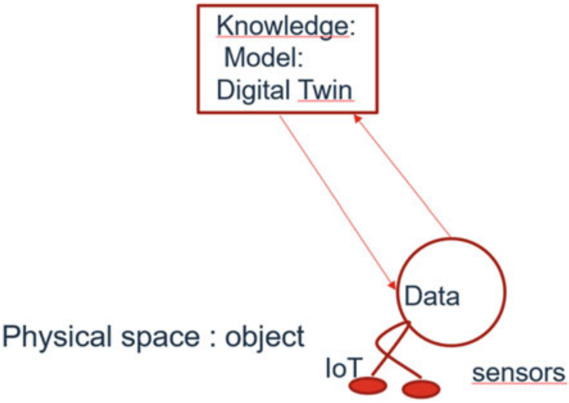
Industrial applications have been reported, for instance to turbines (see e.g., Michau and Fink 2019; Wang et al. 2019).

So far, there have been more applications to diagnostics than to prognostics. In any case, transfer learning has become a key enabler of advanced PHM systems and holds promises for reliability engineering (more generally RAMS) as well.

Combining Physics and Data

As illustrated in the previous sections, data-driven methods benefit from a number of recent breakthroughs and can be very efficient. However, as also mentioned, they suffer from a number of shortcomings. On the other hand, a number of physical degradation phenomena are well described by known physical laws, and it seems natural to use that knowledge when it exists, instead of just blindly processing data without paying attention to their meaning. Therefore, it seems that hybrid PHM, i.e., the joint use of physics knowledge with the processing of acquired field data, is a promising way forward. Physics-based and data-driven algorithms are complementary. A purely physics-based algorithm is limited by the need for a detailed knowledge of model

Fig. 4.4 Digital twin and physical space



parameters, some of which can vary over time with changing environmental conditions and evolving mission profile. In hybrid PHM systems, physical parameters are continually updated as new data is being acquired, as illustrated in Fig. 4.4.

The multi-physics model is sometimes called a digital twin.

A digital twin however is more than just a model. Just as any model, it is an abstraction of reality, i.e. it does not include all the details but only the parameters that are essential to the function being studied; at the same time, it ideally contains all relevant information relating to the history of the physical system; i.e., design changes, maintenance history, etc. A digital twin should accompany the physical system throughout its life time. Initially invented by NASA back in the Apollo program days, the concept has known considerable success recently, in PHM and RAMS applications in particular. Definitions vary and no unique standardized definition has emerged yet.

For instance, an example of definition is: “A Digital Twin is an integrated multi-physics, multi-scale simulation of a complex product which uses available models and information updates (such as sensor measurements, procurement and maintenance actions, configuration changes etc.) to mirror an asset during its entire lifecycle” (Karim et al. 2023).

IEEE distinguishes several classes of digital twins (IEEE 2020): the ‘digital model’, where changes in the physical object must be manually carried over to the digital twin; the ‘digital shadow’, where changes in the physical object are automatically carried over to the twin, and the full digital twin where changes occur automatically in both directions. Most digital twins in existence correspond to IEEE’s ‘digital shadow’ concept.

A fairly recent illustrative example (Staino et al. 2018) is provided by the filter of an HVAC (heating-ventilation-air conditioning) system installed on tramway cars. HVAC performs a crucial function in warm climates. The filters tend to clog with dust accumulation, up to the point where the function of air exchange is no longer adequately performed. To avoid such a ‘failure’, filters are replaced preventively, and filter providers tend to be conservative in their recommendations for a replacement period. To avoid unnecessary replacements while avoiding loss of function, a tramway manufacturer (Alstom) has put in place a preventive maintenance strategy relying on the concept of digital twin. A physical law (Darcy's law) describes material accumulation. The model parameters are continually updated by means of pressure sensors: increasing clogging results in the need for a higher pressure differential upstream and downstream from the filter in order to achieve the same air flow through the filter. Continuous measurement of that pressure difference is the basis for the definition of a health indicator and the dynamic estimation of the filter's remaining useful life, i.e. the time left until reaching an unacceptable level of clogging.

This technique has led to a very accurate filter RUL prediction and it was evidenced that, under normal operating conditions, filter replacement periodicity could be halved without harm, with respect to supplier's recommendations.

That example is comparatively simple, in that there is a single failure mode (clogging) and the complete physical model is readily available (although not elementary at all) and lends itself well to the construction of a simple health indicator.

In general, a full physical model of all relevant degradations is not available.

Several approaches have been proposed to combine physics knowledge with data-driven methods (e.g., Arias-Chao et al. 2019).

The general idea is to use physics-based models to guide the discovery of useful machine-learning models, what is sometimes called "physics-informed machine learning" (Huber et al. 2023).

One promising approach (Arias et al. 2022) consists of estimating unobservable parameters from system dynamics (physics) and sensor readings. Those parameters encode the health condition of system components. They are then input into a deep neural network, along with sensor readings and physical model responses, to generate a prognostics model. This approach has been validated on a standard dataset, the 'Commercial Modular Aero-Propulsion System Simulation' (CMAPPS). It falls into the general category of "physics informed neural networks" (PINN), an active area of research.

A potential benefit of those hybrid approaches is to leverage the advantages of physics-based models and data-driven ones. Clearly, one should carefully avoid inheriting drawbacks from the two methods.

Quantifying and Managing Uncertainty

Reliability engineers have long been accustomed to dealing with uncertainty, typically by attaching a confidence interval to reliability, maintainability or availability estimates instead of just providing point estimates. In AI-based methods, especially when using neural networks, it is only recently that attention has been paid to uncertainty quantification. However, it is extremely important because several sources of uncertainty typically impact detection, diagnostics and prognostics metrics. Those include epistemic uncertainty, i.e., how much is unknown about the model (for instance the value of some model parameters), and what is sometimes called aleatory uncertainty, related to measurement errors and to variability in mission profile. Especially when estimating RUL (remaining useful life), which is affected among other by the future mission profile, providing confidence bounds is indispensable for risk management (Dersin 2023). In purely physics-based models, uncertainty quantification can be based on analytical methods such as first-order reliability methods (FORM) and first-order Taylor expansion of RUL based on the state equation (Sankararaman et al. 2014). For machine-learning based algorithms, more—complex methods have to be considered. A recent comprehensive tutorial (Nemani et al. 2023) surveys state-of-the-art methods, which include, among other, Gaussian process regression, Bayesian neural networks, and neural network ensembles. They lead to the notion of 'uncertainty-aware machine learning'.

Combining AI with Traditional Reliability Engineering

As already pointed out, we believe AI can enhance classical reliability engineering and that there should be a strong synergy between classical approaches and AI-based ones.

Let us illustrate this idea. Recently, this author introduced the use of a time transformation or time warping to describe the time evolution of equipment of system degradation (Dersin 2023). The time warping is in a one-to-one correspondence with the reliability function. It has the property that, in the transformed time, the mean residual life (MRL), i.e., the expectation of the RUL, is a linear function (Fig. 4.5).

The transformation (denoted g) also leaves invariant the first-and second-order moments of the time-to-failure distributions. In the transformed time, the slope of the MRL can be derived explicitly in terms of the TTF's coefficient of variation (i.e. the ratio of mean to standard deviation)-as in (4.2):

$$k = \frac{1 - \left(\frac{\sigma}{\mu}\right)^2}{1 + \left(\frac{\sigma}{\mu}\right)^2} \tag{4.2}$$

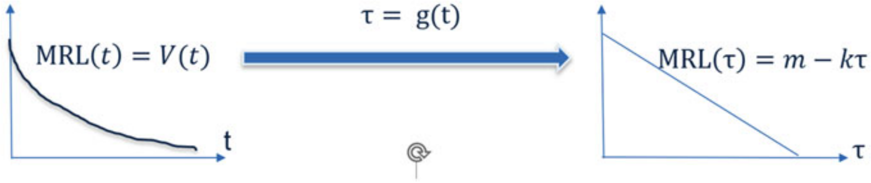


Fig. 4.5 Time transformation

where

$$\mu = \text{MTTF} \tag{4.3}$$

$$\sigma^2 = E\left[(TTF - \mu)^2\right] \tag{4.4}$$

That slope parameter *k* (a dimensionless quantity) is therefore an invariant of the time transformation g. It characterizes the speed of degradation somehow.

As the class of time-to-failure distributions with a MRL linear in time enjoys useful properties, among other an explicit formulation of the RUL confidence interval, those properties can be translated into equivalent properties for the initial distribution, using the inverse mapping g^{-1}. It is thus possible to derive explicit confidence intervals, and also bounds on the average time derivative of the RUL.

Now, identifying the time warping function can be performed by means of classical statistical methods (curve fitting) (Dersin 2023) but, if individual asset conditions, which typically evolve in time, must be taken into account, machine learning-including deep learning-algorithms are probably preferable because they make it possible to take into account a much larger number of parameters, dynamically.

Toward Optimized Dynamic Maintenance

The goal of predictive maintenance is to avoid failures as much as possible, while, at the same time, keeping the maintenance costs reasonable. In order to explicitly manage risks, one can impose an upper bound on the probability that RUL(*t*) is lower than the time to the next inspection. Let the next inspection occur at time $t + s^*$. Then the constraint is

$$P\left[\text{RUL}(t) < s^*\right] < \alpha \tag{4.5}$$

In that way, s^* can be determined explicitly, as a function of *t*, using the time warping g(), by the method described in the previous section.

One can then select the value of the risk α in such a way as to minimize the sum of the total expected cost of maintenance (including preventive and corrective maintenance) and the expected cost of failures (Dersin 2023). More preventive maintenance will mean fewer failures, and therefore the optimal solution is a function of the relative costs (the cost of failure includes failures operational impacts, such as the costs of cancelled flights or of train delays, or lost production).

Since, in practice, environment and mission profile conditions, and therefore stresses, evolve with time, the estimates of the time warping g(.) and the 'degradation speed' k must be updated dynamically. They will also be influenced by the various maintenance operations that are carried out over time. Accordingly, a dynamic decision support process, described by the iterative procedure of Fig. 4.6, can be put in

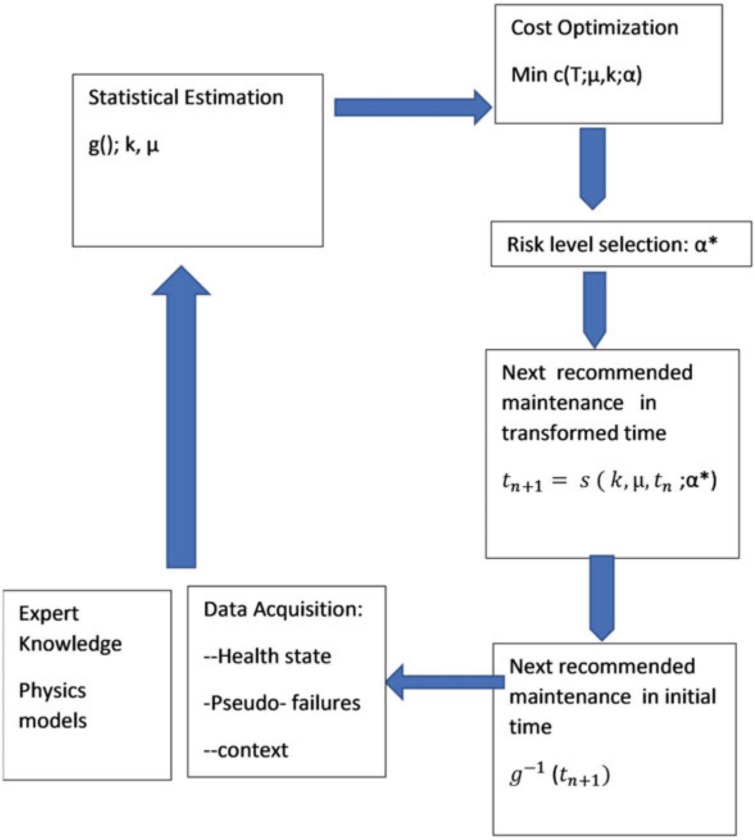


Fig. 4.6 Dynamic risk-based Predictive Maintenance. μ = MTTF. The cost function per unit of time is denoted c. (reproduced from (Dersin 2023) by kind permission of Taylor & Francis)

place. Identification of the g transformation and estimation of mean μ and variance (or mean and slope k) can also rely, not just on acquired field data, but on physical laws and expert knowledge as well. In our view at least, this process can only be a decision support tool and the ultimate decision maker is human.

Conclusions, Opportunities and Challenges

In this brief survey, it was only possible to scratch the surface of the burgeoning field of AI and its potential benefits for reliability engineering and maintenance management. For instance, reinforcement learning, natural language processing, large language models, are all about to revolutionize the field to some extent.

One of the challenges of AI in industrial applications (so-called 'Industrial AI') is interpretability, or explainability: how can a domain expert be convinced that the decisions recommended by a black box (typically a neural network) are justified? That would seem to require something like a 'white box' approach. Clearly, physics-based models are by definition more-easily explainable than purely data-driven ones. But as seen earlier, purely physics-based models are rarely available. The whole field of "explainable AI" (XAI), is therefore receiving increased attention (see (Arrieta et al. 2019) for a recent survey). For instance, SHAP (SHapley Additive exPlanations)—(Strumbelj et al. 2014; Ribeiro et al. 2016) is a game-theoretic approach to explain the output of any machine learning model, by determining the contributions of individual features on the algorithm's decisions. A methodology for focusing on causality rather than just correlation, named Causal Inference (Pearl 2009), is also part of that effort towards explainable AI.

In all algorithms, data quantity and quality is an important concern. Lack of data can sometimes be overcome by data augmentation, and plethora of data can be addressed by pre-processing to eliminate unnecessary or redundant data. Data quality management is a key activity in itself. Algorithms for tracking and eliminating corrupted or contaminated data do exist (see e.g., Ulmer et al. 2023). Another challenge, not the least one, is cybersecurity and data ownership. The more algorithms reside on the cloud, the more this subject comes to the fore (see e.g., Kour et al. 2022).

Industrial AI was the subject of a recent conference sponsored by Luleå University of Technology (IAI 2023), featuring among other the "AI Factory" (Karim 2022, 2023), a collaborative platform that allows multiple industrial partners to share only the data they need to share—in particular by bringing models to data rather than the opposite. That approach has shown its merit in railway applications, and is being generalized to other fields. Last but not least, in this era of climate change and emphasis on sustainable development, a key challenge is the energy consumption of algorithms-frugal AI is the corresponding 'buzzword". The notion combines that of energy frugality and low data consumption.

A good overview of PHM challenges can be found in (Zio 2022). A comprehensive treatment of state-of-the-art Reliability Engineering techniques can be found in (Birolini 2017) and (Nachlas 2017).

In conclusion, in spite of the very real challenges that still exist, especially in industrial applications of AI, we are of the opinion that reliability and maintenance engineers stand to benefit enormously from the potential of AI; and that, at the same time, it would be a mistake to believe that reliability engineering will "dissolve into AI".

References

Alpaydin E (2014) Introduction to machine learning. MIT Press, Cambridge, MA

Arias-Chao M, Adey D, Fink O (2019) Knowledge-induced learning with adaptive sampling variational autoencoders for open set fault diagnostics. arXiv:1912.12502v1

Arias Chao M, Kulkarni C, Goebel K, Fink O (2021) Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics. Data 6(1):5

Arias Chao M, Kulkarni C, Goebel K, Fink O (2022) Fusing physics-based and deep learning models for prognostics. Reliabil Eng Syst Saf 217:107961

Arrieta et al (2019) Explainable artificial intelligence. arXiv:1910.10045

Barlow RE, Proschan F (1965) Mathematical theory of reliability. Wiley, New York

Birolini A (2017) Reliability engineering: theory & practice, 8th edn. Springer

Breiman L (2001) Statistical modeling; the two cultures (with comments and a rejoinder by the author). Stat Sci 16(3):199–231

Cox DR (1972a) The analysis of multivariate binary data. Appl Stat 113–120

Cox DR (1972b) Regression models and life-tables (with discussion). J R Stat Soc Ser B 34:187–202

Dersin P (2023) Modeling remaining useful life dynamics in reliability engineering. CRC Press, Taylor & Francis

(2020) Digital transformation. White Paper of the IEEE Digital Reality Initiative. DigitalReality@ieee.org

Doersch C (2016) Tutorial on variational auto-encoders. arXiv:1606.05908v2

Fink O, Wang Q, Svensén M, Dersin P, Lee W-J, Ducoffe M (2020) Potential, challenges and future directions for deep learning in prognostic and health management applications. In: Engineering applications of artificial intelligence, vol 92

Fisher RA (1922) On the mathematical foundations of theoretical statistics. Philos Trans R Soc Lond Ser A 222:594–604

Gauthier S (2022) Concrete applications of machine learning in railways. In: Proceedings of the ESREL 2022

Goodfellow IJ et al (2014) Generative adversarial nets. In: Proceedings of the international conference on neural information processing systems (NIPS 2014), pp 2672–2680

Hornik J, Stinchcombe M, White H (1989) Multi-layer feed-forward networks are universal approximators. Neural Netw 2:359–366

Huber LG, Palmé T, Chao MA (2023)Physics-informed machine learning for predictive maintenance: applied use-cases. In: 2023 10th IEEE Swiss conference on data science (SDS), Zurich, Switzerland, pp 66–72. https://doi.org/10.1109/SDS57534.2023.00016

Karim R, Galar D, Kumar U (2023) AI factory: theory, applications, case studies. Taylor & Francis, CRC Press

Karim A, Dersin P, Galar D, Kumar U, Jarl H (2022) AI factory: a framework for digital asset management. In: Proceedings of the ESREL 2022

Kour R, Patwardhan A, Thaduri A, Karim R (2022) A review of cybersecurity in railways. Proc Inst Mech Eng Part F: J Rail Rapid Transit

Li X, Krivtsov V, Arora K (2022) Attention-based deep survival model for time-series data. Reliabil Syst Saf 217

Michau G, Fink O (2019) Domain adaptation for one-class classification: monitoring the health of critical systems under limited information. arXiv:1907.09204v2

Nachlas J (2017) Reliability engineering-probabilistic models and maintenance methods, 2nd edn. Taylor & Francis, CRC Press

Nemani V et al (2023) Uncertainty quantification in machine learning for engineering design and health prognostics: a tutorial. arXiv:2305.0493

Pearl J (2009) Causal inference in statistics: an overview. Stat Surv 3:96–146. https://doi.org/10.1214/09-SS05

Ribeiro M, Sameer S, Carlos Guestrin C (2016) *LIME*: “Why should I trust you?: Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM

Sankararaman S, Daigle MJ, Goebel K (2014) Uncertainty quantification in remaining useful life prediction using first-order reliability methods. IEEE Trans Reliab 63(2):603–619

Staino A, Abou-Eïd R, Dersin P (2018) A Monte-Carlo approach for prognostics of clogging process in HVAC filters using a hybrid strategy—A real case study. In: Proceedings of the IEEE Conference on PHM

Strumbelj E, Igor Kononenko I (2014) Shapley sampling values. Explaining prediction models and individual predictions with feature contributions. Knowl Inf Syst 41(3):647–665

Ulmer M, Zgraggen J, Goren-Huber L (2023) A generic fully unsupervised framework for machine-learning-based anomaly detection. In: Proceedings of the ESREL 2023

Wang Q, Michau G, Fink O (2019) Domain-adaptive Transfer learning for fault diagnostics. arXiv:1905.06004v1

Weibull W (1939) A statistical theory of the strength of materials. In: Proceedings of the Swedish Royal Institute of Engineering Research, p 153

Wiener N (1964) Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications. MIT Press

Yao S et al (2023) A survey of transfer learning for machinery diagnostics and prognostics. Artif Intell Rev 56:2871–2922

Zio E (2022) Prognostics and health management (PHM): where are we and where do we (need to) go in theory and practice. Reliabil Eng Syst Saf 218(Part A)