
PHM and Predictive Maintenance

1.1. Anticipative maintenance and prognostics

1.1.1. *New challenges and evolution of the maintenance function*

1.1.1.1. *Industrial maintenance*

According to the standard EN 13306 (2001), maintenance can be defined as a “*combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function*” [EN 01]. It also includes a set of actions of troubleshooting, repairing, controlling and verifying physical equipment, and it should contribute to the improvement of industrial processes. From the traditional viewpoint, the maintenance function guarantees dependability characteristics of equipment, in particular its availability. Therefore, globally it aims to understand the failure mechanisms and to act accordingly in order to ensure that the system (the good) can perform the function that it has been conceived for. However, the missions of the maintenance function are no longer limited to the implementation of means to ensure the “goods service”. Different requirements in terms of quality, safety and costs have emerged, and the challenges and prerogatives of the maintenance function have evolved in the last 20 years.

1.1.1.2. *Challenges and prerogatives of the maintenance function*

The challenges of the maintenance function can be discussed from different points of view. First of all, as industrial equipment becomes more and more complex, it requires greater competence in maintenance. Furthermore, the company evolves within a strongly competitive environment and the financial concerns are very prominent. Therefore, maintenance doesn't escape the cost reduction rule. At another level, for some years, the industrial managers have been facing more significant environmental and social constraints. It is not sufficient anymore to be content with

technical and economic performance, but it becomes necessary, or even compulsory, to take into account the environmental “constraints”: a factory produces waste, pollutes, and contributes to the greenhouse effect, etc. This is coupled with the respect for human dignity, which constitutes a social constraint. The latter aspects have recently led to drafting of legislative texts that strongly encourage companies to include the notion of sustainable development in their strategy. The concrete result is the pursuit of a triple performance, where business performance of course remains essential, but is also complemented by new human/social and environmental requirements. The prerogatives of the maintenance function have thus been studied, and it has had to evolve with regard to the growing challenges:

- It aims to increment the equipment availability while reducing the direct exploitation costs (technical and economic).
- It has to ensure a safe operation of equipment, namely avoiding accidents which can be judged as detrimental to the environment (environmental).
- It is responsible for satisfactory work conditions and for human safety (social).

1.1.1.3. *Evolution of the maintenance function*

In view of the ever-growing requirements, maintenance costs have rapidly risen in recent years. For example, it is estimated that in the USA, the maintenance costs amounted to \$200 billion in 1979, and that they have seen a growth of about 10 to 15% in the following years [BEN 04]. However, an important part of this maintenance cost could be avoided: poor planning leads to waste of supplementary maintenance hours, perhaps on equipment without a major role in production continuity. This increase of costs alone doesn't justify the need to reconsider the traditional maintenance approaches. First of all, production systems evolve continuously and newer technologies have appeared, thanks to automation (as machines could ensure production without human intervention). Next, companies seek to rapidly adapt the production quantity and quality in relation to variations in client's demand, which requires a high level of flexibility for the industrial equipment. Therefore, although maintenance activity is today considered an activity in its own right, the companies do not hesitate anymore to outsource it in order to benefit from the strong core competencies of the service providers. This evolution is due to a large extent to the development of sciences and technologies of information and communication (STIC). At another level, in the last few years, new maintenance architectures have appeared. One of the most recent is that of s-maintenance (“s” for semantics). This maintenance incorporates the concept of e-maintenance, and it is based on the principle of sharing and generating knowledge, formalized by ontology [KAR 09b]. However, before the development of maintenance architecture aiming to reduce the distance between these actors, it is the maintenance strategies themselves that evolve. Indeed, today the maintainers wish to go beyond the static maintenance (without an anticipation of the evolution of the equipment state), and to

implement more “dynamic” maintenance strategies. The following section is dedicated to the analysis of this evolution.

1.1.2. Towards an anticipation of failure mechanisms

1.1.2.1. Cartography of maintenance forms

Before the 1960s, the main mission of a company’s maintenance service was the intervention on broken equipment in order to repair it as soon as possible. This kind of maintenance, known as corrective, has been complemented gradually by an approach that anticipated failure mechanisms, that is, by a maintenance carried out before the failure occurred. These two vast kinds of maintenance – corrective and preventive – present certain variations described below. Figure 1.1 shows their global structure.

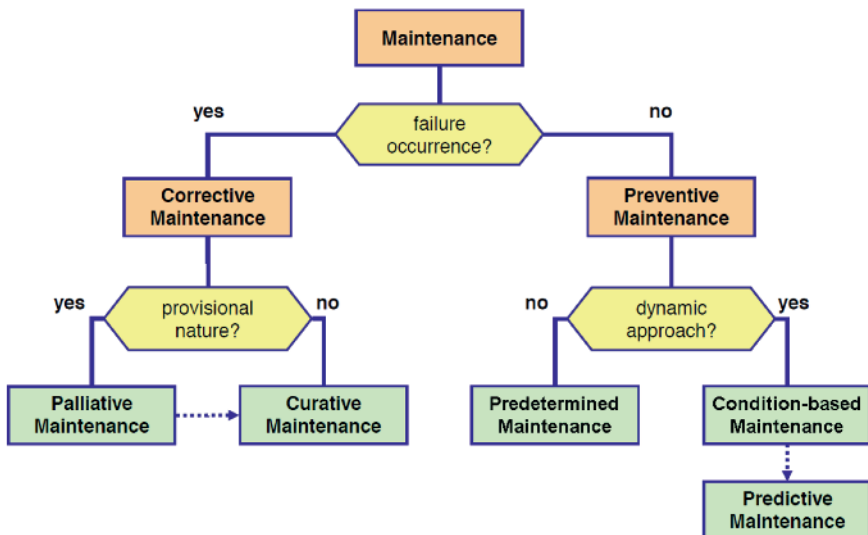


Figure 1.1. Forms of maintenance according to the standard EN 13306 (2001) [EN 01]. For the color version of this figure, see www.iste.co.uk/zerhouni1/phm.zip

1.1.2.2. Corrective and preventive maintenances

The standard EN 13306 (2001) defines corrective maintenance as a “*maintenance carried out after fault recognition and intended to put an item into a state in which*

it can perform a required function.” [EN 01]. This kind of maintenance is generally suitable in case of equipment for which:

- the consequences of the breakdown are not critical,
- the repairs are easy and does not require a lot of time and
- the investment costs are low.

We can distinguish two forms of corrective maintenance. When the intervention of the maintenance is provisional, we refer to it as “palliative maintenance”. If the works are definitive, we refer to “curative maintenance”.

Preventive maintenance aims to reduce the risks of a failure occurring. The standard EN 13306 (2001) defines it as a “*maintenance carried out at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure or the degradation of the operation of an item.*” [EN 01]. When maintenance intervention is performed at fixed and predefined intervals of time, the term “predetermined maintenance” is used. This kind of maintenance is triggered following a schedule (hours of work, kilometers completed, etc.), and is achieved by periodically replacing the parts, without a prior inspection and whatever the degradation state of the goods. Predetermined maintenance can lead to overcare, that is, an excess of useless interventions, and thus financial wastes for the company. In order to remedy this, other forms of preventive maintenance have appeared, based on the monitoring of the actual state of the goods: condition based and predictive maintenances.

1.1.2.3. Condition based and predictive maintenances

Condition-based maintenance is defined as “*a preventive maintenance based on performance and/or parameter monitoring and the subsequent actions*”. This maintenance strategy is thus based on real-time analysis of data of the industrial equipment (for instance vibrations, temperature, etc.). It aims to detect anomalies in the operation of industrial machinery: the discovery of changes in their characteristics prefigures a future failure in the short term. Condition-based maintenance takes into account the usage conditions of the equipment better than the traditional predetermined maintenance. This said, it does not allow designing the maintenance policy with certainty: the occurrence date of the failure remains uncertain. Predictive maintenance aims to remedy this lack of knowledge. It is defined as “*a condition based maintenance carried out following a forecast derived from the analysis and evaluation of the significant parameters of the degradation of the item.*” The underlying idea is to project into the future the current state of the good, in order to estimate the operating time before failure. Therefore, predictive maintenance is more dynamic. It takes into the account the current conditions of the equipment and tries to foresee the good’s state evolution in time. As maintenance interventions are planned with precision beforehand, predictive maintenance saves money substantially, and it has

been the subject of growing attention for some years now. There are numerous benefits expected from it:

- Reduction of the number of breakdowns
- Increased reliability of production processes
- Improvement of personnel safety and of company image
- Reduction of periods of inactivity for the equipment (costly)
- Increment of the performance of the company.

The implementation of a predictive maintenance policy is based on deployment of a key process targeted at determining the future states of the monitored system: the “industrial prognostics”. The next section is dedicated to this concept.

1.2. Prognostics and estimation of the remaining useful life (RUL)

1.2.1. What is it? Definition and measures of prognostics

Many definitions for the term “prognostics” have been proposed in the literature [BYI 02, ENG 00, HEN 09a, HES 05, JAR 06, LEB 01, LUO 03, MUL 05, PRO 03b, SIK 11, VAC 06, WU 07, ZIO 10a]. The different meanings stem essentially from the career and the applicatory sensitivity of the authors. However, although without a total consensus, prognostics can be defined as proposed by the ISO committee:

Standard ISO 13381 (2004). *The aim of prognostics is the “estimation of time to failure and risk for one or more existing and future failure modes”* [ISO 04].

We can highlight one key feature. The very concept of failure¹ implies that prognostics should be based on evaluation criteria, whose limits depend on the monitored system and on performance targets. In other words, prognostics implies not only that we should be able to project into the future the behavior of a system, but also that we should be capable of identifying the health state at each instant, taking into account the chosen mission criteria. As a consequence, there is no unique set of evaluation metrics that would be appropriate for any prognostics application [ORC 10, SAN 15, SAX 08a, SAX 09, SAX 10, VAC 06]. However, we can distinguish two classes of measures.

▷ *Prognostics measures.* The main goal of prognostics is to provide information that helps in making correct decisions. Therefore, an initial set of metrics is that which

¹ EN 13306. *Failure: termination of the ability of an item to perform a required function.* [EN 01].

quantifies the risks incurred by the monitored system. This kind of metric corresponds to prognostic measures among which the main one is Time To Failure - TTF or *Remaining Useful Life (RUL)*. We also need to construct a *confidence* measure in order to indicate the degree of certainty of the RUL. As an example, let us consider the left part of the Figure 1.2 in which, for the sake of simplicity, degradation is considered to be a one-dimensional quantity. The RUL can be defined as the time between the current instant t_c (after the detection of the failure; t_D), and the instant where the degradation will reach the failure threshold (t_f):

$$RUL = t_f - t_c \quad [1.1]$$

▷ *Performance measures of the prognostic system.* It is necessary to be able to judge the quality of the prognostics as well, in order to decide the adequate actions. To that end, we can construct several indicators: the performance measures of the prognostic system. The main measures highlighted in the literature are “timeliness”, “precision” and “accuracy”. These metrics cannot be detailed here, but a clear explanation can be found in [GOE 05, VAC 06]. In any case, they represent a measure of the distance between a set of RUL estimates and a set of exact values of RUL (cf. right part of Figure 1.2).

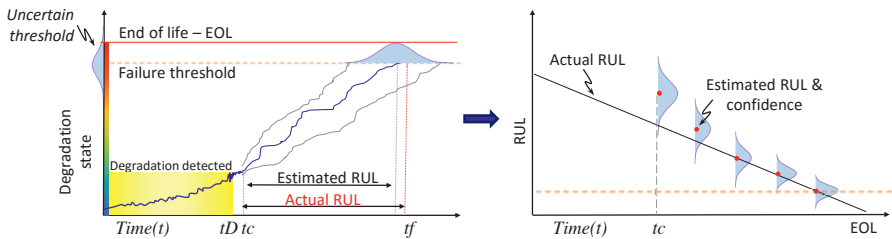


Figure 1.2. Illustration of prognostic process. For the color version of this figure, see www.iste.co.uk/zerhouni1/phm.zip

At this point, we need to remember that the prognostic process is globally stabilized, yet inherently uncertain. Furthermore, it raises some evaluation problems (how can it be qualified/quantified?).

1.2.2. How? Prognostic approaches

1.2.2.1. A taxonomy of prognostic approaches

During the last decade, several tools and methods of failure prognostics have been proposed, and literature is rife with papers that have aimed (in part) at

defining a classification of prognostic approaches [DRA 09, GOR 09, HEN 09b, JAR 06, KOT 06, LEB 01, LEE 14, PEC 08, PEN 10, SI 11, SIK 11, TOB 12b, VAC 06, VAN 09, ZIO 12, ZIO 10a]. It seems that prognostic methods generally differ according to the type of application considered, while the implemented tools depend mainly on the nature of available data and knowledge. Furthermore, these methods and tools can be grouped in a limited number of approaches. The following classification usually reaches a good consensus within the PHM community (Figure 1.3):

- Prognostics based on a physical model, that is, physics-based
- Prognostics guided by data, that is, data-driven
- Hybrid prognostics.

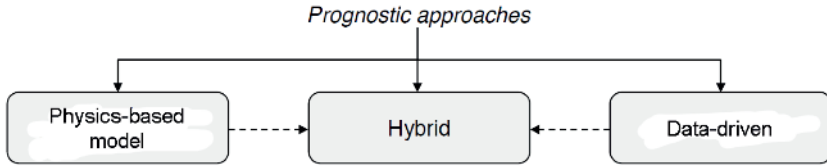


Figure 1.3. *Taxonomy of prognostic approaches*

▷ *Physics-based prognostics.* Methods based on a physical model require the construction of a dynamic model representing the behavior of the system and integrating the degradation mechanism (mainly by models of fatigue, wear, or corrosion), whose evolution is modeled by a deterministic law or by a stochastic process [CHE 04, CHO 11, FAN 11, GUC 11, HON 13, KAC 04, LI 05, LI 00b, LUO 03, PEC 09, PEC 10, PEY 07, QIU 02, UCK 08, WAN 10]. These methods usually offer more precise results than those of the two other approaches. Furthermore, they have the advantage of being interpretable; model parameters are related to physical quantities of the system, and a degradation systematically induces parametric deviations. However, the greatest inconvenience of this kind of approach resides in the fact that for real systems it is difficult, or even impossible, to obtain a dynamic model integrating the degradation mechanisms in an analytic form (because of complexity and diversity of the physical phenomena). Moreover, a model built for a specific application is very difficult to transpose to another physical system, even of the same kind. Therefore, their field of application is restricted.

▷ *Data-driven prognostics.* These approaches are based on the exploitation of monitoring data, which is processed in order to extract the characteristics that reflect the behavior of the system and its degradation. These quantities are further employed to train predictive models of the current and future states of the system, and thus to provide an estimation of the RUL. Undeniably, this is the most developed category of

approaches, with works based on the exploitation of neural networks, neuro-fuzzy systems, and their variations [CHI 04, DRA 10, EL 11, GOU 12, HUA 07, JAV 14a, MAH 10, RAM 14, TSE 99, WAN 01, WAN 07, WAN 04], probabilistic methods (Bayesian networks, Markov models and their derivatives) [BAR 05b, CAM 10, DON 07, DON 08, MED 12, MOS 13b, MUL 05, SER 12, SER 13, TOB 12a, TOB 11a, TOB 12b], stochastic models [BAR 10, BAR 05a, GRA 06, LE 12, LE 13, LOR 13], state space and filtering models (Kalman, particle filters) [AN 13, BAR 12, CAD 09, ORC 05, PHE 07, SAX 12, SIK 11, SWA 99], regression tools [BEN 15, HON 14b, KHE 14, LEE 06b, NIU 09, WU 07, YAN 04, ZIO 10b], or combinations of methods [BAR 13b, BAR 13c, BAR 12, HU 12, JAV 12, RAM 10]. These approaches do not require an analytical model of behavior and failure of the system; therefore they are relatively simple to deploy. The user is exempted from building complex models and instead exploits data gathered *in-situ*. The development of sensors and monitoring systems, combined with growing computing performances, offer remarkable capacities for processing, analysis, and learning, and thus they facilitate the implementation of this approach. On the other hand, data-driven prognostics loses accuracy as the learned models deviate from the real behavior of the system. Therefore, it constitutes a compromise between applicability and precision.

▷ *Hybrid prognostics.* A hybrid prognostic method is an integration of a physical behavioral model and a data-driven approach. Usually, we distinguish two classes of hybrid prognostics (Figure 1.4). When a physical model (even an empirical one) can be established, a data-driven approach is used to estimate and predict the nonobservable parameters of the model. In this case, we speak of “series approaches” [BAR 13a, BAR 13d, DAL 11, DON 14, FAN 15, HU 15, JOU 14, MED 13, OLI 13, ORC 10, PEC 10, PEY 07, ZIO 11]. An approach called “parallel” (or “fusion”) consists of combining the output of the physical model with the output of a data-driven tool in order to reconstruct a global output. In such cases, the data-driven tool is generally used to estimate and to predict the phenomena that are not explained and thus not modeled [CHE 09, HAN 95, KUM 08, MAN 13, PEC 10, THO 94]. Hybrid approaches show good estimation and prediction performances. Moreover, they help to build a good model of the uncertainties. On the other hand, they can be very costly in terms of computing resources, and they are limited by the need of a physical model of degradation mechanisms.

1.2.2.2. *Synthesis and remarks*

Classification of the prognostic approaches is not an end in itself, and of course the boundaries between the classes are not hermetic. As an example, a Bayesian network can be used to generate a dynamic model of a system (model-based approach). For this purpose, it is possible to use a set of algorithms that learn the structure and

network parameters (data-driven approach). Furthermore, particle filters are based on the expression of a model of hidden states, updated with the help of observations provided by sensors; it is sometimes qualified as a data-driven tool or a model-based tool. Of course, the message here is not that the classification of prognostic approaches is wrong, but rather that it has to be considered with caution. It appears that there is no universal prognostic approach and that the choice of an adequate technique depends on the classical constraints that limit the applicability of the tools:

- Possibility of measurements and availability of data record;
- Availability of engineering models or of physical knowledge of the phenomena;
- Dynamics and complexity of the real systems;
- Variability of operating conditions and/or of the mission profile;
- Implementation constraints (precision, computing time, etc.), etc.

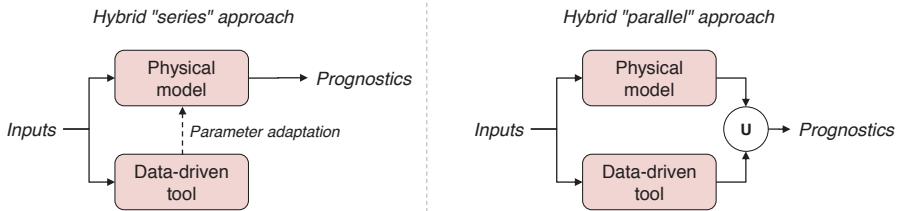


Figure 1.4. *Hybrid prognostic approaches*

1.3. From data to decisions: the PHM process

1.3.1. Detection, diagnostics and prognostics

Industrial monitoring and maintenance involve different business processes with a purpose of maintaining the system in operating conditions at the lowest cost. So, we frequently speak of fault detection, failure diagnostics, choice of control and/or mitigation actions (preventive or corrective), and of planning of these actions over time. Figuratively, these steps correspond to the necessity, first of all, to “perceive” certain phenomena, then, to “understand” them, and, finally, to “act” as a consequence. This said, as we already mentioned, another approach (complementary but not exclusive) consists not in understanding *a posteriori* a phenomenon, which has just manifested (a failure), but, rather, in trying to anticipate their occurrence in order to proceed accordingly to protective actions. This is the goal of “failure

prognostics”. The relative placement of “detection”, “diagnostic” and “prognostic” processes is represented in Figure 1.5a. From a point of view of phenomenology, their complementarity can be explained as follows [GOU 11] (Figure 1.5b):

- Detection aims to identify the operating mode of the system and its state.
- When a failure occurs, diagnostics isolates and identifies the component that stopped functioning (from effects to causes).
- Prognostics aims to predict the future states of the system (from causes to effects).

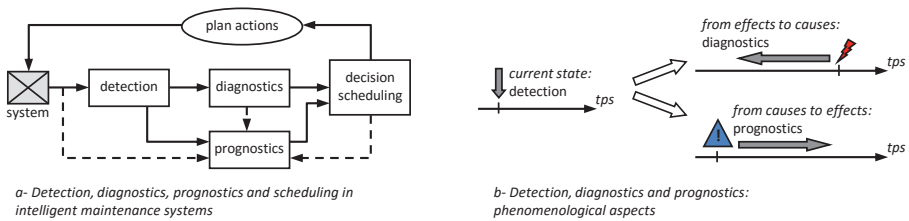


Figure 1.5. Complementarity of detection, diagnostic, and prognostic activities [GOU 11]

1.3.2. CBM Architecture and PHM process

▷ *CBM Architecture.* Business processes connected to the failure expertise listed above complement each other perfectly. Furthermore, it appears that prognostics should not be treated in an isolated way.

– Beforehand: anticipation of failures can be achieved only if the degradation mechanisms are correctly comprehended (data acquisition, feature extraction, detection, diagnostics).

– Afterwords: in practice, RUL is not an end in itself. On the contrary, its estimation should lead to adequate reaction policies (decision analysis).

Therefore, prognostics deployment requires a set of tasks (modules) that are commonly grouped under the term “CBM system” (condition-based maintenance). To structure it, a great number of works have appeared [BEN 03b, BYI 02, DJU 03, GAU 11, LEB 01, LEE 06b, MUL 08a, PRO 03a, PRO 03b, VAC 06]. The differences focus essentially on the form of software architecture to be implemented (local, distributed, in modules, etc.), and the standard distributed by MIMOSA group turns out to be unifying [MIM 98]: the OSA/CBM (Open System Architecture for Condition Based Maintenance — let us remark that this standard has been

normalized [ISO 06]). This architecture consists of seven functional levels, which can be considered as sequential or distributed (Figure 1.6).

1) *Data acquisition*. This module provides the system with digital data acquired from sensors or transducers (acquisition, back-up, and securing). The operator (exploiter or maintainer) can also enter basic data (intervention type, causes, date, duration, etc.).

2) *Data processing*. The signals issued from sensors are processed in order to extract the features that suggest the presence of an anomaly, the beginning of a degradation, and, in the long term, that represent the state evolution of the monitored system over time.

3) *Condition assessment (detection)*. The detection module compares real-time data (extracted features) with some expected or known values; it has to be capable of generating alerts depending on prefixed thresholds, based on criteria of performance, security, etc.

4) *Diagnostics*. On the basis of the detected state, this module determines whether the monitored system or component is degraded or not, and it suggests probable causes of failure (identification and localization). The diagnostic module requires a thorough knowledge of system's components, their interactions, and their operating and environmental conditions.

5) *Prognostics*. This module relies on data issued from the previous modules, in order to predict the future state of the monitored system or component, and to give an estimation of the time to failure (RUL). This prediction needs to know the current state of the system, and to extrapolate future utilization conditions.

6) *Decision analysis*. Its main function is to recommend control/maintenance actions such that the system can function until the accomplishment of its mission. This phase is based in particular on RUL estimates.

7) *Presentation*. This module receives the information from all the previous modules for online or further usage. It can be built in the form of a HMI (human machine interface).

▷ *What about PHM?*. While the OSA/CBM structure has been defined from the perspective of a computerized implementation of an intelligent maintenance system, it nevertheless reveals a set of elementary processes necessary to properly comprehend and manage failure mechanisms. Furthermore, today, this decomposition is still used to describe indifferently the process of prognostics and health management (PHM). Besides, PHM is a term for which no consensual definition has been proposed. According to CALCE Center, one of the most active groups of research in this domain, PHM is “*the means to predict and protect the integrity of equipment and complex systems, and avoid unanticipated operational problems leading to mission performance deficiencies, degradation, and adverse*

effects to mission safety". In other words, the field of application of PHM is not limited to industrial maintenance, but the elementary processes, which it resorts to, are the same as those of OSA/CBM decomposition. The main difference lies in the nature of the decisions to be taken: PHM is considered to be more general and can be applied to any kind of activity.

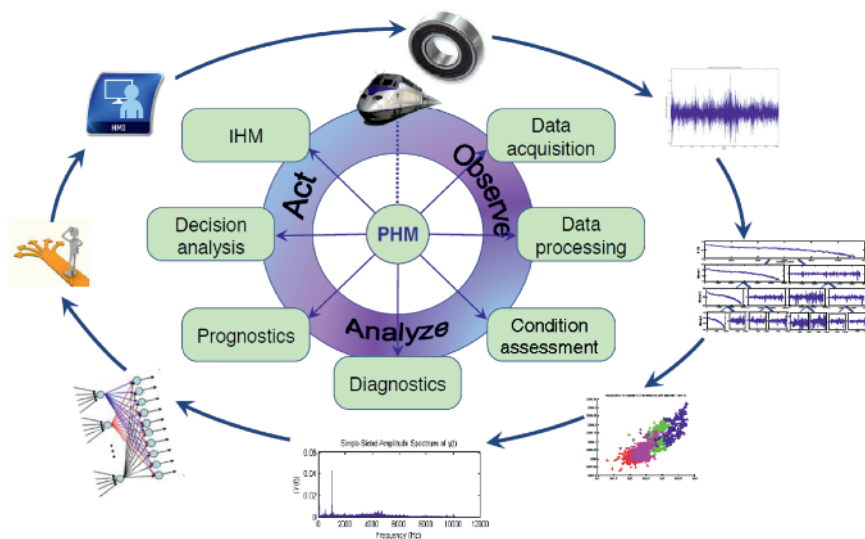


Figure 1.6. PHM cycle as an adaptation of the OSA-CBM architecture. For the color version of this figure, see www.iste.co.uk/zerhouni1/phm.zip

1.4. Scope of the book

As has been described above, the PHM process covers a wide spectrum of activities; it would be intricate to tackle them in a single work. For this reason, three of the seven levels of PHM shown in Figure 1.6 are not treated here. First of all, failure “diagnostics” (layer 4) is today a mature discipline and it is well developed both on scientific and on industrial levels. Therefore, this aspect is not discussed here; the interested reader can refer to the following reference works: [ISE 97, ISE 05, JAR 06, MED 05, SAM 08, VEN 05]. On the other hand, as was mentioned in the introduction, the “decision” and the underlying “data architecture” phases (PHM layers 6 and 7) are the subject of a second book (From PHM concept to predictive maintenance Part 2 - Knowledge, traceability and decision). Thus, we do not develop them here.

The following chapters are articulated as follows, in relation to the PHM layers.

- Chapter 2: data acquisition (PHM layer 1),
- Chapter 3: data processing (PHM layer 2),
- Chapters 4 and 5: health assessment and failure prognostics (PHM layers 3 and 5). Two families of approaches are introduced.

