6

Prognostics and Health Management in Railways

Pierre Dersin, Allegra Alessi, Olga Fink, Benjamin Lamoureux, and Mehdi Brahimi

CONTENTS

6.1 Introduction and Historical Perspective ..............................................................79
6.2 Links between PHM and Reliability, Availability, and Maintainability .......... 81
6.3 ISO 17359 and Other Standards ........................................................................ 82
6.4 Data Acquisition Challenge .............................................................................. 84
6.5 Performance Indicators ..................................................................................... 84
   6.5.1 Data Acquisition KPIs ................................................................................ 85
   6.5.2 Data Processing KPIs ................................................................................ 85
   6.5.3 Detection KPIs .......................................................................................... 85
   6.5.4 Diagnostics KPIs ....................................................................................... 85
   6.5.5 Prognostics KPIs ....................................................................................... 86
   6.5.6 Advisory Generation KPIs ........................................................................ 86
6.6 Integration and Human–Machine Interface ....................................................... 86
   6.6.1 Questions and Issues for Hardware Integration ....................................... 86
   6.6.2 Questions and Issues for Software Integration ......................................... 87
6.7 Importance of PHM in the Railway Industry .................................................... 88
6.8 Case Studies ....................................................................................................... 88
   6.8.1 Railway Turnouts ...................................................................................... 88
   6.8.2 Bogies ....................................................................................................... 89
   6.8.3 Infrastructure ........................................................................................... 92
6.9 IT Challenges ..................................................................................................... 93
6.10 Challenges and Future Perspectives ................................................................. 94
6.11 Conclusions ...................................................................................................... 95
Acknowledgment .................................................................................................... 96
References ............................................................................................................... 96

6.1 Introduction and Historical Perspective

Industrial maintenance has evolved considerably over the last 70 years. Roughly speaking, one could say that the first generation, until approximately 1950, was characterized by a purely “corrective” perspective, i.e., failures led to repairs, and then came the second generation, characterized by scheduled overhauls and maintenance control and planning tools (roughly the period of 1950–1980). From the 1980s onward, the notion of condition-based maintenance (CBM) gained ground. With the turn of the twenty-first century, a great interest in “predictive maintenance” has emerged, along with the concept of prognostics and health management (PHM). A number of rail companies and original equipment manufacturers now have a PHM department.
Let us backtrack and take a closer look at those concepts and try to understand this evolution.

First of all, what is maintenance? The standard IEC 60300-3-14 defines maintenance as “the combination of all technical and administrative actions, including supervision actions, intended to retain an item in, or restore it to, a state in which it can perform a required function.” It is divided into two broad categories: preventive, i.e., maintenance operations taking place before a failure, and corrective, after a failure, as illustrated in Figure 6.1.

As far as preventive maintenance is concerned, there are two main categories: scheduled (or systematic) maintenance, which takes place at fixed intervals (defined by time, distance, or other measures of usage), and CBM, which takes place when the condition of the asset warrants it. Maintenance costs (corrective and preventive) of complex systems such as railways represent a considerable percentage of asset life cycle cost (several times the initial investment), and, therefore, there is considerable economic pressure to reduce them. At the same time, the direct and indirect costs of a service-affecting failure (SAF) can be considerable (in terms of lost missions, punctuality, and service disruptions in the case of railways). Therefore, there is substantial incentive to reduce their frequency of occurrence.

The prevalent belief in the 1950s and 1960s was that every item on a piece of complex equipment had a right age at which complete overhaul was needed to ensure safety and reliability. This attitude had led to the conclusion that the new Boeing 747 would require 4 hours of maintenance for each hour of flight because it contained many more parts to be maintained than in previous models! To escape this conclusion, which defied common sense and would have meant that the new aircraft was economically not viable, a task force was set up in 1960 by the US Federal Aviation Administration. The task force discovered that scheduled overhauls have little effect on the overall reliability of a complex item unless the item has a dominant failure mode and that there are many items for which there is no effective form of scheduled maintenance. This effort led to the Maintenance Steering Group (MSG), which was able at the same time to reduce maintenance costs of

![Figure 6.1](image)

**FIGURE 6.1**
Classification of maintenance types according to IEC 60300-3-12. (Courtesy of International Electrotechnical Commission, Geneva, Switzerland, 2004.)
civilian aircraft and increase their reliability. Eventually, in 1978, the success of the MSG led to the “reliability-centered maintenance” (RCM) methodology, pioneered by the US Department of Defense. RCM is a process used to determine the maintenance requirements of an asset in its operating context (see standard IEC 60300-3-11). It is based on functional failure modes and effects analyses (FMEAs) and systematic cost-performance trade-off analyses, which lead to the definition of the appropriate maintenance policy. A key concept part of RCM is the “P–F curve,” where P denotes a point in time when a progressive deterioration in the maintained asset is detected and F denotes the point in time when failure takes place if no maintenance action is performed after that detection. CBM can be based on periodic inspections (whose periodicity must be adapted to the speed of evolution of the failure mechanisms) or on continuous monitoring.

The notion of predictive maintenance is a special case of CBM: CBM infers from the condition of the asset that there is an impending failure, i.e., a failure will take place if no action is taken, and predictive maintenance intends to predict, in a probabilistic sense, after how much time the failure will take place if no action is taken. That time is called the remaining useful life (RUL). It is, of course, very much dependent on the assumed future load profile of the asset.

Since the first decade of the new millennium, several factors have favored the effective development and deployment of CBM and predictive maintenance: the appearance of low-cost wireless sensors and the expansion of the Internet of things, the significant progress of advanced multivariate statistics and machine intelligence methods, and the ability to manage and store very large data sets (big data) and to perform concurrent processing at affordable costs.

The expression “prognostics and health management” refers to the methods, means, and tools (both hardware and software) that are used to perform diagnostics, health assessment, and prognostics of industrial assets.

The digitalization of railways is a major trend (as, for instance, illustrated at InnoTrans 2016), and PHM both benefits from that context and influences it.

6.2 Links between PHM and Reliability, Availability, and Maintainability

The concepts of reliability engineering and PHM are closely related. However, they are often not considered jointly in practical applications. There is a great potential to improve the reliability and availability of the system and to reduce the maintenance costs by integrating both approaches already starting in the concept phase and to develop designs for reliability and for PHM and by continuously updating and improving the implemented sensors, models, algorithms, and the decision support systems during the operating phase.

Traditional reliability analysis typically relies on time-to-failure data to estimate the lifetime distributions and to evaluate the reliability of the “average” asset operated under “average” operating conditions. With the highly customized components and subsystems, and very different operating conditions of the components within the fleet, the traditional analysis provides a good estimate for the maintenance decisions on the aggregated level, but not on the level of an individual component. With the increased reliability of the critical systems, full time-to-failure trajectories are often available only for noncritical systems, as critical system failures are rare events. Additionally, the average behavior of the systems does not sufficiently
decrease the temporal uncertainty of the required maintenance actions. With the increased availability of advanced sensors and measuring devices, communication networks, and computer processing power, PHM approaches enable the assessment of the reliability of a component or a system under its actual application conditions and its actual system health state.

The main difference between the reliability and the PHM approaches can be seen in the change of the perspective from a fleet or a set of components operated under diverse conditions to the perspective of an individual component with its unique characteristics and unique operating conditions. Both are complementary.

PHM can be considered as a holistic approach to an effective and efficient system health management of a component or an overall system. PHM integrates the detection of an incipient fault, its isolation, the identification of its root cause, and prediction of the RUL. The system health management goes one step beyond the predictions of failure times and supports optimal maintenance and logistic decisions by considering the available resources, the operating context, and the economic consequences of different faults.

While reliability considerations are important during design for taking the design decisions, during the operation, on the contrary, early fault detection, diagnostics, and the prediction of the RUL at the level of the individual component combined with an uncertainty in the prediction become more important. By aggregating the predicted RUL of the components within the fleet, the reliability can be considered by quantifying the mean residual life and using it for CBM decision making. Indeed, the mean residual life is the mathematical expectation of the RUL. It is derived from the reliability function or the failure rate.

While PHM methods are not able to improve the inherent reliability of a component, they are able to avoid operational consequences of a fault and thereby preserve or extend the operational reliability of the asset. PHM also influences the operational availability by reducing the frequency of occurring failures affecting the operational reliability and reducing the time for fault finding and improving the preparedness of the maintenance organization and by improving the maintainability of the system.

PHM can therefore be considered as an enabler for improved operational reliability and availability.

Conversely, reliability, availability, and maintainability (RAM) engineering is a powerful tool for PHM methodology: FMEAs, or their extension failure modes, mechanisms, and effects analysis, make it possible to pinpoint the key causes of SAFs and thereby determine some of the assets that stand most to benefit from PHM; and they constitute the first step toward identifying the physical variables to be monitored.

6.3 ISO 17359 and Other Standards

Given the impact of PHM technologies on any process, it is necessary to ensure the interoperability of various systems as well as a standardized approach to apply a PHM technology to any process. Having a standard which is adopted throughout the same field is beneficial, as it provides a generally accepted technical specification which enables different parties across different domains to interact and understand the practices adopted. For this reason, a dedicated safety unit already exists: the European Railway Agency, which monitors all rail safety in the European Union and develops common technical standards and approaches to safety, promoting interoperability. In the United Kingdom, the Rail Safety and Standard Board is an independent and nonprofit company which aims to improve the
performance of the railway in Great Britain and defined the rules and standards for operating the network. It is also important to underline that as the number of standards for the same topic increases, their effectiveness and usefulness significantly diminish: a series of competing standards go against the very nature and aim of the standard itself.

While a PHM standard per se does not yet exist, the International Organization for Standardization (ISO) has developed several guidelines for condition monitoring and diagnostics of machinery, for which ISO 17359 (Condition Monitoring and Diagnostics of Machines—General Guidelines) is the parent document, containing references to prognostics as well. In the same manner, the Institute of Electrical and Electronics Engineers Standards Development Committee has written a PHM standard, P1856, which is undergoing approval process at the time of writing.

Both ISO 17359 and P1856 introduce a series of capabilities requirements that are outlined in the following phases:

- Data acquisition
- Data processing
- Detection
- Diagnostics
- Prognostics
- Advisory generation

Each of these phases is defined as necessary and fundamental for the overall application of the PHM process to a system. These steps pertain to the actual application of PHM to a system and do not delve on matters of feasibility, cost-effectiveness, and business scenarios for the implementation of PHM. While it is true that the standards are meant to cover thoroughly the PHM process, a consideration of such matters is strongly advised before diving into the technical aspects in a real application. These are briefly addressed in the informative section of the P1856 standard, as preliminary considerations.

The first step is data acquisition, which consists of identifying the right technology and methods to acquire the data from the system on which PHM will be applied. This step implicitly requires a previous evaluation of the failure modes of the system as a whole, to define the critical elements which will be the key to a successful PHM application. The critical elements are often those which fail more often, cause the most severe failures, cause the longest downtimes, or cost the most to replace. The decision of which parts of the system constitute the elements which require monitoring needs to be taken from a hybrid of RAM and safety approach and cost–benefit analysis. Once these key elements are identified, the parameters and data to be measured and recorded can be defined. The acquisition of these data is not necessarily straightforward, as the railway industry has static infrastructure (rails and signaling) as well as moving assets (rolling stock), and the acquisition of data from both needs to be coherent and coordinated into the same platform.

The second step is data processing, where the data acquired from the system are treated to extract the relevant information, pertinent to producing a PHM process. This is a step common to any use of data, as rarely is the raw signal acquired ready for the analysis. The acquisition of these data is not necessarily straightforward, as the railway industry has static infrastructure (rails and signaling) as well as moving assets (rolling stock), and the acquisition of data from both needs to be coherent and coordinated into the same platform.

The third step is detection, where signals are taken and analyzed to evaluate whether the behavior of the monitored system is indicating any incipient degradation mechanisms. If this is the case, then the PHM system detects that there is an anomaly in the functioning of the system and triggers the evaluation of the next steps.
The fourth step is diagnostics, where the anomaly which has been detected is analyzed, and its location, severity, and root cause are identified. This information is vital, as it differentiates a PHM process from a CBM process, which merely detects the condition of a system.

The fifth step is prognostics, which is the study and prediction of the future evolution of the health of the system being monitored, to estimate the RUL of the system. By calculating these values, it is possible to define the optimum time for maintenance intervention, to maximize the use of the assets and plan the maintenance actions at a time when the impact on productivity and availability is the lowest.

The sixth and final step is advisory generation, also called health management, where the data collected and evaluated in the previous phases are combined with the information of the system process and the results of the processing are readily available for use. This information can be sent to a visualization interface or integrated in the current maintenance planning platform.

While ISO 17359 and P1856 offer a good framework for reference on how to apply and work with PHM, the application to the railway sector still lacks a well-defined standard. Other industries, such as the aerospace industry, are regulated by the plethora of standards on maintenance and health management systems, which are drafted and released by the Society of Automotive Engineers (SAE) Aerospace Propulsion System Health Management Technical committee. In the coming years, because of the increased awareness toward PHM and its contributions, as well as the increase in importance of the railway sector, it is expected that standards well defining these practices will be released.

However, while a series of standards for PHM exists, there is still a lack of harmonization and coordination in the overall field, which often causes confusion. For this reason, it is imperative that when proceeding with the development of PHM technology, the user ensures to comply with the standard that they deem most appropriate.

6.4 Data Acquisition Challenge

Whatever “analytics” are used to process data collected from the field, the output will not be of great value if the input data are of poor quality.

The reliability of the entire acquisition chain, including sensors, data loggers, and communication channels, is of paramount importance, and adequate noise filtering must be performed. In addition, the frequency of acquisition (both how often data are acquired and with what frequency signals are sampled) must be tailored to the type of signal processing method, which the analyst intends to use and to the time constant of the degradation processes. Additional considerations are presented in Section 6.6 (“Integration and Human–Machine Interface”).

6.5 Performance Indicators

In the industry, performance indicators or key performance indicators (KPIs) are nowadays the fundamental quantitative values on which managers rely to justify the success or
failure of projects. In the past decade, the issue of defining KPIs for PHM has been raised in several papers. For example, the works of Saxena et al. [1], Roemer and Byington [2], and Dzakowic and Valentine [3] can be cited. For PHM, six types of KPIs are defined accordingly to the main PHM steps.

### 6.5.1 Data Acquisition KPIs

The data acquisition KPIs quantify how well the sensors, data acquisition board, wires, and network infrastructure perform in faithfully translating physical quantity into analog values, digitalizing them and sending them as fast as possible while minimizing the loss of information. Usually, engineers rely on values such as accuracy, sampling rate, resolution, and bandwidth to give a short list.

### 6.5.2 Data Processing KPIs

The data processing KPIs quantify how well the features represent the raw data. In the case of PHM, a set of good features must principally show (a) a very good sensitivity to degradations and (b) a minimum correlation between the components. To quantify the first aspect, many KPIs issued from the domain of sensitivity analysis can be used, for example, the Fisher criterion [4], the Sobol indices [5], or the Morris method [6]. The L1 term of the Lasso [7] can also be used for that matter. For the second aspect, the best indicator of the correlation between features is the covariance matrix.

### 6.5.3 Detection KPIs

The detection KPIs quantify how well the anomaly detection performs. The most common approach is to compute the confusion matrix, as explained by Wickens [8]. Usually, for industrial purposes, the KPIs of interest are the false alarm ratio and the good detection ratio. One can also use a receiver operating characteristics (ROC) curve, which can help in tuning the detection threshold in function of the true positives/false positives ratios targeted. The area under the curve, or similarly the Gini coefficient [9] of the ROC, can also be used as a direct KPI of the detection performance. In the field of big data, some other KPIs, derived from the confusion matrix, are also commonly used, for example, accuracy, prevalence, negative/positive likelihood ratios, and diagnostics odds ratio.

### 6.5.4 Diagnostics KPIs

The diagnostics KPIs quantify how well the classification of the degradation type performs. A common way of quantifying this type of performance is to compute the classification matrix for the degradation types. At this point, the reader must be warned that a necessary condition for that is to have labeled data. The classification matrix gives, for each set of degradation type label, the results of the classification process. A perfect classification would have all the values in the diagonal of the matrix. In the case of PHM, industry practitioners are usually interested not only in the type of degradation occurring in their system but also (and sometimes more) in its localization. Indeed, many systems are now modular, according to the principle of line replaceable unit (LRU), and identifying which LRU is degraded is the only necessary information needed for maintenance. In this case, a localization matrix can be used instead of (or complementarily to) the classification one [10].
6.5.5 Prognostics KPIs

The prognostics KPIs quantify how well the assessment of the RUL performs. Some authors have recently focused their research on this topic, such as Saxena et al., who gives in their study [11] a list of metrics that can be used to quantify both the online performance of prognostics and the uncertainty in predictions. For example, the online performance can be assessed from the RUL online precision index and the dynamic standard deviation that quantify, respectively, the precision of RUL distributions by quantifying the length of the 95% confidence bounds and the stability of predictions within a time window. The uncertainty can be evaluated by using the β-criterion that specifies the desired level of overlap between the predicted RUL probability density function and the acceptable error bounds.

6.5.6 Advisory Generation KPIs

The advisory generation KPIs quantify how well the predictive maintenance itself performs. Indeed, because it is the last layer of the PHM process, its requirements are directly given in terms of service reliability and availability, for example, the rate of SAFs, the service availability, the time to repair or, even better, if it is possible, the difference between these values taken without and with PHM with respect to the total cost of the PHM system.

6.6 Integration and Human–Machine Interface

Once the sensors and the PHM algorithms have been validated in laboratory conditions, the integration consists of pushing the hardware and software to a real-life field environment. Multiple questions are then to be answered.

6.6.1 Questions and Issues for Hardware Integration

Is the equipment compliant with railway standards and does it necessitate safety case reopening? Indeed, as soon as integration is addressed, the issue of compliance with the domain standards is raised. These standards can impose, for example, electromagnetic and vibration resistance for sensors or specific shielding for wires. Intrusivity is also a very important factor for integration: if the sensor is intrusive, it is very likely that a safety case will have to be reopened and significant additional costs are then to be expected.

Does the equipment necessitate dedicated processing/storage unit? Of course, as long as you stay in laboratory conditions, the processing and storage capabilities are not an issue, as they can be easily and quickly scaled to the need. However, in a field environment, these resources can be limited, especially in a location where the empty spaces are rare, such as in bogies. For some cases, the already installed processing unit may be enough to process the measurements, for example, in local control units for heating, ventilation, and air-conditioning, but most of the time, a dedicated unit will have to be designed and integrated in the train architecture.

Does the equipment necessitate dedicated communication network? The trains of the new generation are generally equipped with a remote diagnostics system that sends the train status variables (speed, Global Positioning System coordinates, etc.) and some discrete events (mainly fault codes) to a cloud server in which dynamic availability analysis
are performed. However, this system uses 4G routers, which are often not designed to handle PHM-related data in addition to the events. In some cases, it can be necessary to design a dedicated communication system to download data from the train, for example, a Wi-Fi or Ethernet connection in the depot.

What is the reliability of the installed hardware? PHM hardware includes sensors, wires, and processing units which have their very own reliability properties. Thus, even if neither the safety nor the service reliability is impacted, a failure of one of the PHM data acquisition chain components will trigger some corrective maintenance that will necessitate workload and spare parts. Because this hardware is generally multiplied by the number of a given subsystem occurrence, a low reliability can have such a significant impact that operators prefer to shut down the system. Hence, performing reliability analysis before the integration of the hardware is of paramount importance.

### 6.6.2 Questions and Issues for Software Integration

Does the communication network constrain the bandwidth? In case that a dedicated communication system for PHM cannot be designed, and if the bandwidth of the existing one is limited, it is not an option to send sampled physical quantities from sensors. In this case, the main solution is to perform the data manipulation (features extraction) onboard and to send only the features to the PHM off-board server that runs the next processing steps. The issue of the onboard/off-board split is very important for PHM integration because it dimensions the onboard computation and storage needs. Usually, if the bandwidth permits, sending the raw sampled measurements to be able to modify the features extraction logic as often as needed is preferred.

Do the algorithms necessitate a specific safety integrity level (SIL) and do some parts of my process need to be recoded in another language? Whatever the onboard/off-board split, there will always be a part of the PHM algorithms onboard, at least to trigger the acquisition according to specific operating modes of the system (example: launch acquisition of the bogie acceleration if the speed of the train is above 100 km/hour) because storing all the measurements during the whole mission is not wanted for obvious volumetry issues. In some cases explained before, the feature extraction logic can also be performed onboard. For each of those embedded tasks, it has to be checked as early as possible in the design process if an SIL is required (in case that it interacts with a safety-related unit) or if a specific language is required.

What are the different machines/servers needs for the information technology (IT) architecture? For a fully instrumented fleet, the volume of collected data per day and the computational power needed to process them can be enormous. In the future, it will be more and more standard to use a cloud-based architecture. The key for success at this level is to ensure that the architecture is not only dimensioned for the current data but also scalable for future evolutions. Nowadays, several companies are providing cloud-based service, under the form of raw infrastructure (infrastructure as a service [IaaS]), platform (platform as a service [PaaS]), or high-level software (software as a service [SaaS]). The PaaS is a compromise between the high complexity (coding and networking) of the IaaS and the low flexibility of the SaaS.

What is the global availability of the IT architecture? Similarly to the hardware reliability, a study of the IT architecture availability has to be performed to assess the service level agreement (SLA) that can be sold to the end user. Most of the time, this SLA is in the scope of the platform provider.
Another aspect of the integration is the interface with the end user, the so-called human–machine interface. Beyond the purely network and coding language issues, the human–machine interface is the emerging part of the iceberg and has to be (a) compliant with the client requirements, i.e., functionalities are tuned with respect to the needs; (b) user friendly and aesthetic; and (c) scalable and highly reliable. According to ISO 13374, the minimal information that a PHM human–machine interface shall contain is the current value of the state (healthy on degraded), the degradation type identified, the health index, and the RUL, and this for each asset of the fleet. Ideally, the possibility to visualize the history of these values should be given to the end user.

### 6.7 Importance of PHM in the Railway Industry

Investments and projects based on monitoring assets and their functionalities have significantly increased in numbers in the past years in the railway industry. This is due both to the significant cost of scheduled maintenance on railway assets and infrastructure and to the impact of a failure on the overall functioning of the railway industry. Failures of components of the railway industry, in terms of both infrastructure (e.g., deformed rail, failure to move a turnout, signaling failure) and assets (e.g., traction faults, bogie failure), cause delays, which are often costly and can cause severe accidents.

According to the European Commission, passenger and freight transportation in Europe is expected to, respectively, double and triple in 2020 [12]. This increase in demand and, consequently, in service frequency, greater axle loads, and an overall increased attention to the railway industry cannot be simply met by building new infrastructure, which is often expensive and not time effective. To meet this challenge, it is pivotal to ensure the safety and reliability of the current network, which will become of critical importance in the coming years. This, together with the high costs of scheduled maintenance and planned overhauls, which often turn out to be not necessary, has pushed the industry to seek more advanced methods to monitor the real-time state of the assets.

Implementing a monitoring strategy on assets allows for early identification of failures, thereby increasing railway availability and safety while reducing maintenance costs.

### 6.8 Case Studies

#### 6.8.1 Railway Turnouts

Railway turnouts are devices which enable trains to be directed from one track to another, by mechanically moving a section of railroad track. They are composed of a motor which generates power, which is used in turn to move sections of track to the desired position. Their functioning is central to allowing a functional infrastructure and an undisturbed flow of traffic. When a turnout is faulty and therefore unable to move to the desired position, it forces the network operator to have to reroute the trains when possible, or halt the traffic and send a team to repair the turnout as rapidly as possible to recommence the normal flow of operations.