SPIM

Thèse de Doctorat

école doctorale **sciences pour l'ingénieur et microtechniques**

UFC UNIVERSITÉ DE FRANCHE-COMTÉ

Remaining useful life estimation of critical components based on Bayesian approaches

Ahmed MOSALLAM

SPIM

Thèse de Doctorat

école doctorale **sciences pour l'ingénieur et microtechniques**

UFC UNIVERSITÉ DE FRANCHE-COMTÉ

N° X X X

THÈSE présentée par

Ahmed MOSALLAM

pour obtenir le

Grade de Docteur de
l'Université de Franche-Comté

Spécialité : **Automatique**

Remaining useful life estimation of critical components based on Bayesian approaches

Soutenue le 18 Décembre 2014 devant le Jury :

Antoine GRALL	Rapporteur	Professeur à l'Université de Technologie de Troyes, France
Piero BARALDI	Rapporteur	Maître de Conférences à l'Ecole Polytechnique de Milan
Mustapha OULADSINE	Examinateur	Professeur à l'Université de Marseille, France
Nasr Eddine BERRACHED	Examinateur	Professeur à l'Université des Sciences et Technologies d'Oran, Algérie
Jayant SEN GUPTA	Invité	Ingénieur de Recherche, Airbus Group Innovations, France
Noureddine ZERHOUNI	Directeur de thèse	Professeur à l'ENSMM, France
Kamal MEDJAHER	Co-encadrant de thèse	Maître de Conférences à l'ENSMM, France

Contents

List of Figures

List of Tables

Notations

$\overline{X}$	mean value
σ	standard deviation
$cov(X, Y)$	covariance measure
$p(w)$	prior probability
$p(D\|w)$	likelihood function
$p(w\|D)$	posterior probability
$p(D)$	probability density function
R_{XY}	Pearson's correlation coefficient
ρ	Spearman's rank correlation coefficient
$H(X)$	information entropy
$H(X, Y)$	joint entropy
$I(X, Y)$	mutual information
$SU(X, Y)$	symmetrical uncertainty
$r(t)$	residual function
$imf_i(t)$	intrinsic mode function
Q	clustering quality measure
$\mathcal{N}(0, \sigma_n^2)$	normal distribution
z_t	sensor observation at time t

Abbreviations

AE	Acoustic Emission
AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Average Probability Index
AR	Auto Regression
ARMAX	Auto Regressive Moving Model with Exogenous
C-MAPSS	Commercial Modular Aero-Propulsion System Simulation
CBM	Condition-based Maintenance
CI	Computational Intelligence
CMF	Combined Mode Function
CMMS	Computerized Maintenance Management Systems
CTM	Constrained Topological Mapping
EKF	Extended Kalman Filter
EMD	Empirical Mode Decomposition
EOL	End Of Life
FMEA	Failure Modes and Effects Analysis
FT	Fourier Transforms
FTA	Fault Tree Analysis
GA	Genetic Algorithm
GLM	General Linear Model
GMM	Gaussian Mixture Models
GP	Gaussian process

GPR	Gaussian Process Regression
HHT	Hilbert-Haung Transform
HI	Health Indicator
HMM	Hidden Markov Model
ICA	Independent Component Analysis
IMF	Intrinsic Mode Function
k-NN	k-Nearest Neighbor
LDR	Levinson-Durbin Recursion
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
OM	Operating Modes
PCA	Principal Component Analysis
PDF	Probability Density Function
PF	Particle Filter
PHA	Preliminary Hazard Analysis
PHM	Prognostics and health management
PLS	Partial Least Square
PM	Predictive Maintenance
PSD	Power Spectral Density
RUL	Remaining Useful Life
SOM	Self Organised Map
STFT	Short-Time Fourier Transform
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TTF	Time To Failure
WPD	Wavelet Packet Decomposition

1

General introduction

The industrial revolution was a period in which fundamental developments occurred in almost every aspect of our lives. It thoroughly changed the old ways of production from depending on human/animal power to automated industrial systems. *Industrial system* is a general term that describes the process of utilizing different types of control systems to manage the behavior of machine(s) or sub-system(s) in modern factories, trains, airplanes and many other complex systems. Due to the advances in science and technology, industrial systems are becoming more efficient and have increased the average income and safety of the population. Industrial systems, however, can suffer from unplanned stops due to break downs of machines and subsystems. Unplanned stops can cause catastrophic consequences such as loss of lives, environmental contamination and high financial costs.

There are several examples when unplanned stops have resulted in disasters and accidents with extensive losses. For example, on May 1979, McDonnell Douglas DC-10-10 aircraft lost control and crashed due to breakdown of an engine. The accident happened shortly after the takeoff from O'Hare international airport in Chicago. According to the analysis, the leading cause of the engine breakdown was due to improper maintenance. All 271 passengers and crew on board, as well as two people on the ground were killed (Figure 1.1a). On April 2010, an explosion occurred on the Deepwater Horizon oil rig in the Gulf of Mexico. The main cause of the explosion was a faulty cement sealing, followed by a failure of blowout preventing equipment. The failures allowed the release of gas and subsequent ignition of hydrocarbons. Eleven workers were killed, 16 injured and a massive offshore oil spill leaked in the Gulf of Mexico (Figure 1.1b). On January 2013 two major Japanese airlines, All Nippon Airways (ANA) and Japan Airlines (JAL), announced that they were grounding flights for their fleets of Boeing 787 Dreamliner aircraft. The decision took place after multiple lithium-ion battery warning followed by a burning smell incidents including emergency landings. The grounding costed ANA over $1.1 million a day (Figure 1.1c).

Needless to say, achieving high reliability and availability of industrial systems is a crucial task and requires serious efforts for preventing unplanned stops. This can be

[1]Images courtesy: Wikipedia.org

(a) DC-10-10 aircraft minutes before crash.

(b) Deepwater horizon explosion.

(c) 787 aircrafts grounded at Tokyo airport.

Figure 1.1: Consequences of unplanned stops.[1]

done through adopting efficient maintenance activities, such as measuring and making adjustments to detect and correct problems before they become severe and shut down the industrial system. Effective maintenance was shown to increase the reliability and availability by offering greater utilization of any facility of industrial systems and reducing costs through managing work and downtime. Many types of maintenance strategies have been developed over the last decades according to the requirements of different industrial systems. Due to recent development of sensor and monitoring technology, Condition-Based Maintenance (CBM) has emerged as a promising maintenance strategy. It uses visual inspection and sensor data to assess condition of machinery. CBM replaces the predefined maintenance tasks with only the necessary ones, based on the equipment condition. In this way, CBM strategy reduces maintenance costs while increasing efficiency by performing maintenance actions only when there is evidence of abnormal behavior. Recently, CBM+ strategy is proposed to deal with the new re-

quirements in the maintenance domain. Such requirements necessitates predicting the system health condition in the future and take decisions accordingly. *CBM+* can be defined as an updated maintenance concept that emphasizes prognostics or predictive capabilities, assessment of the material condition and estimation of the remaining useful life at any time during a system or component's life. Moreover, *Prognostics and Health Management* (PHM) is a set of advanced diagnostic, prognostic, and health management capabilities and data products that enables and supports CBM+. PHM research attracts significant research interest due to the need for prediction and decision models, which are important concepts for performing efficient CBM+ strategy.

Performing PHM for a whole industrial system, however, is challenging and still quite difficult in practice. Instead, component-oriented PHM approaches are more feasible. Such approaches are based on identifying, monitoring and maintaining the critical subsystems or components in the industrial systems. Component-oriented PHM approaches consist of seven main tasks, namely: 1) data acquisition, 2) data processing, 3) fault detection, 4) diagnostic, 5) prognostics, 6) decision making and 7) human machine interface. In particular, prognostics research has recently attracted a lot of research interest due to the need of predictive models and will be the main focus of this thesis.

1.1 Positioning of the research

Generally, prognostics approaches can be categorized into: 1) model-based (physics of failure) approach, 2) data-driven approach and 3) hybrid approach. Model-based approach constructs mathematical models of the desired critical components by the means of state-space models and dynamic ordinary or partial differential equations. This approach can be very reliable if accurate models are built. On the other hand, such models require extensive experimentation and model verification. Data-driven approach can be used when the first principles of the system operation are complex such that developing of accurate physics of failure model is not feasible. Such approach employs artificial intelligence and machine learning models to characterize the degradation behavior of the monitored components. Finally, hybrid approach combines physics of failure and data-driven approaches to leverage the advantages from both approaches. Hybrid approach can be complicated to develop as well as computationally expensive.

Data-driven approaches can be further sub categorized into cumulative degradation and direct RUL mapping prognostics approaches. Cumulative degradation prognostics approach can be performed by modeling the damage propagation evolution of the desired component using sensor data. This model can be used later for health assessment and Remaining Useful Life (RUL) estimation. Direct RUL mapping prognostics approach includes the use of different models to map the relation between the input sensor data and the required quantity, such as RUL. Such models can be used later to directly estimate the RUL. In this thesis, we propose a novel data-driven prognostic method based on direct RUL mapping prognostics approach. To do this, health indicators are extracted from the raw monitoring signals, which may have originated from a single or multiple sensors aggregated to represent the degradation evolution over time. The

relations between such health indicators and RUL are mapped using different data-driven models.

1.2 Research questions

The large volume of data gathered continuously from monitoring critical components has created challenges to interpret such data in order to anticipate the future breakdowns. Most large industries have specialized engineers skilled in the use of high technology monitoring equipment and have earned some special certification in the field of maintenance. Nevertheless, it is still hard to take immediate decisions and predict the system failure beforehand. The need for computer systems that constantly record data, monitor the status, provide health assessment by measuring the degradation level and estimate the RUL of different critical components is particularly important for increasing reliability while decreasing the maintenance costs. One of the main challenges for estimating the RUL is that sensor signals acquired from the critical components are usually hidden by noise. Hence, it is very challenging to process such signals and to extract information, if any at all, about the degradation behavior of the desired component. Different approaches have been proposed for extracting such information from different component that can represent degradation modeling. However, it is still difficult to find universal approach that can be used for different applications. Furthermore, it is still challenging to construct health indicators directly from processed sensor signals or features to infer the health status of the desired component. Another problem is the prediction uncertainty associated with the prediction models due to the variation of the End Of Life (EOL) time. EOL can differ for two components made even by the same manufacturer and operating under the same conditions. Therefore, proposed models should include such uncertainties and represent them in a probabilistic form. The challenges for component-based data-driven PHM can be summarized as follows:

1. Building data-driven models require historical data that represent the failure mechanism.
   - Where to place the sensors for acquiring condition monitoring data?
   - What parameters should be monitored?
   - How to choose monitoring sensors and data acquisition methods?
2. Sensor data is usually noisy, imprecise and incomplete.
   - How to extract relevant information from the sensor data?
   - How to construct new signals that represent the component deterioration?
3. The resulting historical data depicts huge variation due to the stochastic nature of the degradation phenomena, unforeseen future loads and variable environmental conditions.
   - How to model such data?

- How to represent the uncertainty of the data?
- Does the model cover all phenomena associated with degradation mechanism?
- What applications can be modeled?

1.3 Formalization and tools

Data-driven prognostic approaches use empirical models to learn the degradation mechanism from monitoring data. Such models map the relation between the system state variables, namely input, internal and output variables without explicit knowledge of the physical behavior of the monitored component. Empirical models can be divided into two main overlapping groups, namely Computational Intelligence (CI) and Machine Learning (ML) models. CI group includes models that mimic the nature, such as neural networks and fuzzy systems. ML include approaches that learn from experience and can enhance its performance over time, such as similarity based approaches and Bayesian based approaches.

In the thesis, different ML algorithms are used to perform prognostics for critical components. First, a non linear similarity based approach is applied to select a smaller subset of the input sensor data. Such subset contains information about the degradation evolution over time. Then, different Bayesian based algorithms, such as discrete Bayesian filter, k-NN classifier and Gaussian process regression, are used to assess the health status and to estimate the RUL of the monitored component. Bayesian approaches offer strong framework that can represent the uncertainty about the estimation in a probabilistic form. The probabilistic representation of the estimations can be useful for later steps such as decision making.

1.4 Global assumptions

Data-driven prognostics approaches build on learning the degradation mechanism from sensor data acquired from the critical components. This data should contain information about health degradation evolution over time, otherwise, the prognostics models will not efficient. Extracting such data from the critical components is a challenging step towards building reliable prognostics models. System experts have to study the industrial system to identify monitoring level, critical components, the parameters to be measured and the sensors to be used for data acquisition. These tasks are essential for extracting sensor data that will be the basis of building reliable models. Thus, the efficiency of the proposed prognostics models can be affected by the quality of sensor data. The global assumptions considered in this thesis can be summarized as follows:

1. The proposed method is dedicated to critical components and not for a whole system.
2. The critical components, the monitoring parameters and monitoring sensors are identified previously by system experts.

3. The sensors used for data acquisition are not faulty.
4. Sensor data are multidimensional, non stationary and non linear time series signals.
5. Data acquisition of historical data stops when the monitored component reaches its EOL condition (run to failure).
6. No maintenance intervention took place during the data acquisition process.
7. The degradation of the monitored components develops gradually over time due to incipient faults (drift-like).

1.5 Contributions of the thesis

We propose a novel data-driven approach for health state assessment and direct RUL estimation of critical components. The approach learns the relation between acquired sensor data and EOL to estimate the current health state and RUL of the monitored component. The proposed method is composed of two phases, namely offline and online.

In the offline phase, the method starts by looking for "interesting" variables in the form of non-random relationships among measured sensor signals, or features derived from signals. The assumption is that, information about the degradation of a critical component can be extracted from the relationships between signals of that component. The selected variables are then compressed, using Principal Component Analysis (PCA), into compact form. Then, Empirical Mode Decomposition (EMD) algorithm is applied to extract monotonic trends that represent the degradation of the critical component over time. Next, statistical features are extracted from the trends to represent each trend in a compact form through the time. Finally, such features are used to construct different Health Indicators (HI) of the monitored component.

In the online phase, new sensor data goes through the same steps as the offline signals. Then, discrete Bayesian filter is applied in order to estimate the current health status and represent the uncertainty about the new data in a probabilistic form. Finally, the new acquired trends are then compared to the trends learned during the offline phase first by using k-Nearest Neighbor (k-NN) algorithm. If the posterior probability of the selected class is less than a certain threshold, then the method uses Gaussian process regression (GPR) to approximate the closest correct group. In both cases, the method associates a probability value to the decision to represent the uncertainty about it. The method is demonstrated on real data sets namely, bearings, turbofan engines and batteries. The results demonstrate the efficiency of the method in finding important relationships and using them for predicting the RUL with high efficiency. The contributions of this thesis can be summarized as follows (Figure 1.2):

1. Selecting/extracting informative variables/features to represent the relation between the input signals and their EOL values.
2. Constructing different health indicators that represent the degradation evolution of the monitored components.

3. Integrating different machine learning algorithms, such as k-NN, GPR and histogram filter, to learn the degradation evolution, assess the health status of the components and estimate their RUL.

4. Representing the uncertainty about the estimations in a probabilistic form.

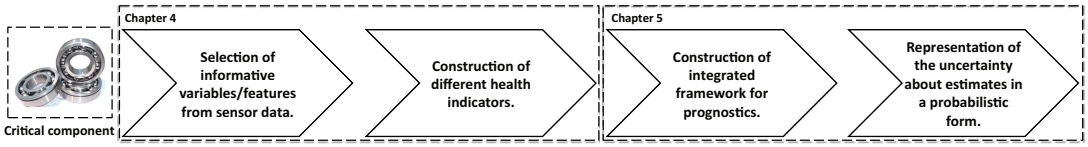


Figure 1.2: Overall scheme of the contributions.

1.6 Publications

The contributions presented in this thesis were reported in the following articles.

I) **Journal papers**

1 **Mosallam, A.**, Medjaher, K., Zerhouni, N., (2014), Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. Journal of Intelligent Manufacturing. Pages 1-12, DOI: 10.1007/s10845-014-0933-4.

2 **Mosallam, A.**, Medjaher, K., Zerhouni, N., (2014) Time Series Trending for Condition Assessment and Prognostics, Journal of Manufacturing Technology Management, Vol. 25 Issue: 4, Pages: 550-567.

3 **Mosallam A.**, Medjaher K., Zerhouni N., (2013), Bayesian approach for remaining useful life prediction, Chemical Engineering Transactions, 33, Pages: 139-144, DOI: 10.3303/CET1333024.

4 **Mosallam, A.**; Medjaher, K.; and Zerhouni, N. (2013). Nonparametric time series modelling for industrial prognostics and health management. The International Journal of Advanced Manufacturing Technology: Volume 69, Issue 5, Pages: 1685-1699. DOI: 10.1007/s00170-013-5065-z.

II) **Conference papers**

1 **Mosallam A.**, Medjaher K., Zerhouni N., Component based data-driven prognostics for complex systems: Methodology and applications. (Submitted on November 19th, 2014)

2 **Mosallam A.**, Medjaher K., Zerhouni N., (2014), Integrated Gaussian Process and Bayesian Framework for Remaining Useful Life Prediction. IEEE International Conference on Prognostics and Health Management, PHM'2014. Pages: 1-6.

3 **Mosallam, A.**; Medjaher, K.; and Zerhouni, N. (2012). Unsupervised trend extraction for prognostics and condition assessment. In A-MEST'12, 2nd IFAC workshop on Advanced Maintenance Engineering Service and Technology , Pages: 97-102. Seville, Spain.

1.7 Outlines of the thesis

In addition to the general introduction, this thesis is composed of six chapters.

Chapter 2 - briefly introduces the main maintenance strategies that have been developed over the years. PHM approaches are then introduced as a key role process for enabling maintenance activities. Finally, state of the art of different PHM approaches are presented with main focus on data-driven methods.

Chapter 3 - discusses the first step towards building a PHM process, in particular the methods used to identify critical components for an industrial system. Then, the selection method of the parameters that can represent degradation behavior of a monitored component is presented. Finally, selection of the sensors used to monitor the degradation parameters and signal acquisition and pre-processing are presented.

Chapter 4 - presents an overview of different signal analysis approaches for extracting information about the degradation behavior, such as variable selection, feature extraction and dimensionality reduction. Then, a method for HI construction, from multidimensional sensor data, is presented. HI can be used to assess the health status of the monitored component and predict the RUL.

Chapter 5 - presents a novel data driven prognostics method for health assessment and RUL estimation of critical components. The method uses different machine learning algorithms, such as k-NN and Gaussian process regression to map the relation between sensor data and degradation behavior. The method deduces the health status using discrete Bayesian filter applied on the online HI. Finally, the results of applying the proposed method on different real life applications, namely bearings, Lithium-ion batteries and turbofan is presented.

Chapter 6 - concludes the research work developed in this thesis and discusses the perspectives and future work.

2

Towards prognostics and health management

"Minds are the ideal way to express complexity, energy density, increasing specialization, expanding diversity - all in one system. Mindedness is what evolution produces. Mindedness is what technology wants, too."

– Kevin Kelly

2.1 Introduction

Increasing profitability is at the top of industries management concerns. This is being driven by reducing the downtime costs, through improving products quality and increasing the availability and productivity of the systems (Table 2.1). Effectively addressing

Table 2.1: Estimations for downtime costs [1–3].

Industry	Average downtime costs per hour
Railway downtime penalty	$ 4000
Forest products	$7,000
Food processing	$30,000
Petroleum and chemical	$87,000
Metal casting	$100,000
Automotive	$200,000

the challenges of reaching such goals involves having efficient maintenance strategies

[4, 5]. Therefore, maintenance is becoming a very essential process in modern industry. The cost of maintenance in general can be up to 20%-30% for the chemical industry and 40%-50% for the mining industry of the total budget. It is estimated that more than $300 billion are spent on plant maintenance and operations by U.S. industry each year, and that 80% of this is spent to correct the chronic failure of machines, systems, and human errors. Moreover, the size of the maintenance group in an industrial organization varies over a range of 5%-10% of the total operating force. Efficient maintenance strategy, however, reduces the probability of machine breakage which can reduce the downtime cost by 40%-60% [6, 7]. Thus, preventing failures through maintenance actions is crucial for safety, reliability and economy. Importance of maintenance increases significantly and there is a constant need to increase its efficiency. To do that, Prognostics and health managements (PHM) research is linking the failure mechanisms modeling and decision making research with recent maintenance strategies. PHM approaches have become a key enabler to achieve efficient maintenance tasks. In this chapter, we briefly introduce different maintenance strategies that have been developed over the years. PHM approaches are then introduced as a key role process for enabling maintenance activities. Finally, state of the art of different PHM approaches is presented with main focus on data processing and data-driven prognostics approaches.

2.2 Maintenance

Maintenance is defined as the 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 [8–10]. Maintenance tasks vary according to working environments, which include e.g. visual inspection, testing, measurement, changes of consumables (greasing, lubrication, oil filters), adjustment, repair, upkeep, replacement of parts, servicing, oil sampling, lubrication, re-tightening of the bolts, cleaning, fault detection, fault diagnosis and so on [11, 12]. Maintenance can be categorized into three main types, namely corrective, time-based and condition-based maintenance.

2.2.1 Corrective maintenance

It is the easiest approach for maintaining an asset. In this strategy, the equipment is allowed to run until failure and then the failed part is repaired or replaced. No actions are taken to maintain the equipment, however, in some cases temporary repairs may be made in order to return equipment to temporary operation, with permanent repairs put of until a later time [13]. The advantage of this approach is that it minimizes the maintenance manpower by keeping equipment running until failure without interruption for maintenance. In this way, this strategy fully utilizes the service life of the component. However, the downside of this approach includes unpredictable production size and increased maintenance costs to repair failures. The cost of repairs might also increase upon the failure of a secondary device that is associated with failure of the primary failure. Labor cost associated with repair will probably be higher than normal because

the failure will most likely require more extensive repairs than would have been required if the piece of equipment had not been run to failure.

2.2.2 Time-based maintenance

This type of maintenance relies on the estimated probability that the equipment will fail in the specified time [14]. Actions are performed on a time-based schedule regardless of the machine health condition. Maintenance activities may include equipment lubrication, parts replacement, adjustment and inspection for signs of deterioration during the inspection. The advantages of time-based maintenance are reduced breakdown frequency of equipment and increased service life. On the other hand, the disadvantage is the need to interrupt production at scheduled intervals to perform the maintenance and the service life of the component is not fully utilized.

2.2.3 Condition-based maintenance

Under CBM, different kinds of sensors are used to measure the physical condition of equipment. When the equipment degradation status reaches a specified level, maintenance actions are performed to restore the equipment to the normal condition. Quite recently, more challenging requirements have emerged in the maintenance domain, which require predicting the system health condition in the future and take decisions accordingly. This kind of maintenance is known as CBM+. It defines the needed maintenance tasks in the future based on equipment current condition [15, 16]. In this way, the equipment is taken out of service only when direct evidence exists that deterioration has taken place. On the downside, CBM/CBM+ require increased investment in monitoring equipment and training for engineers. The advantages of CBM/CBM+ are the reduction of maintenance costs while increasing efficiency by performing maintenance actions only when there is evidence of abnormal behavior.

Identifying an appropriate maintenance strategy for a specific industry is not a simple task. Maintenance experts have to define the best maintenance strategy to adopt for each equipment or system in the plant [17], based on two main factors: 1) measurable factors: such as (cost, productivity, availability, system functions, failure modes and identified maintenance requirements and tasks) 2) immeasurable factors: such as (safety, comfort). Several approaches are noted in the literature for selecting appropriate maintenance strategy [18, 19]. In [20] a method based on multi-criteria decision making (MCDM) approach is proposed. The method utilizes the application of decision making theory to maintenance with particular attention to multiattribute utility theory. Analytical Hierarchy Process (AHP) is proposed in [21], where the authors considered only four maintenance criteria: cost, reparability, reliability and availability. Reliability Centered Maintenance (RCM) methodology is probably the most widely used technique [22]. It is defined as a process used to determine what must be done to ensure that any physical asset continues to do whatever its users want it to do in its present operating context [23–27]. RCM is widely used in many industrial fields, such as steel plants [28], railway

networks [29, 30], wind turbine industry [31]. Table 2.2 depicts the advantages and disadvantages of the three main maintenance strategies.

CBM+ is the newest maintenance strategy and required in different industries for many reasons. CBM+ minimizes the costs of maintenance, improves operational safety, and reduces the quantity and severity of system failures [32, 33]. Furthermore, it can facilitate the planing of the future maintenance tasks by predicting the failures of the different equipment. In order to allow industrial systems to shift from traditional maintenance strategies to CBM+, Prognostics and health management (PHM) approaches are used [35]. PHM approaches have become a key enabler to achieve CBM+ goals [34]. PHM is an emerging research field that links studies of data processing, fault detection, diagnostic, failure mechanisms and decision making to CBM+ [36]. PHM attracts significant interest due to the need for prognostics and decision models, which are important concepts for performing efficient CBM+ strategy [37].

2.3 Prognostics and health management

PHM can be defined as the research and engineering disciplines that form the essential foundations required by CBM/CBM+. It uses information provided by monitoring data and maintenance engineers to identify the current health status of the monitored system and calculate the time left for stable operation to plan the required maintenance action(s). PHM is a compilation of seven main tasks depicted in Figure 2.6.

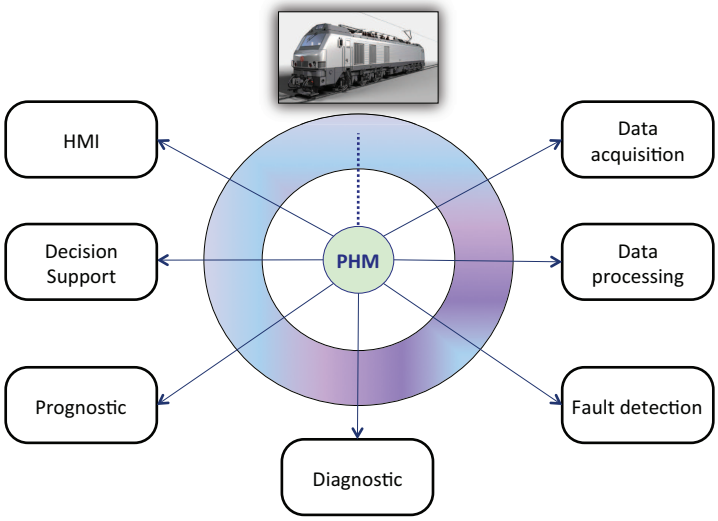


Figure 2.1: PHM tasks.

Table 2.2: Comparison of the advantages and disadvantages of the three maintenance strategies.

Strategy	Advantage	Disadvantage
Corrective maintenance	• Low cost. • Less staff. • Component service life is fully utilized.	• Unplanned downtime. • High costs of repair or replacement of equipment. • Sever consequences.
Time-based maintenance	• Increased availability. • Reduced downtime. • Cost savings.	• Probable unplanned downtime. • Labor intensive. • Unneeded maintenance. • Component service is not fully utilized.
CBM/CBM+ maintenance	• Increased availability. • Reduced downtime. • Decreased parts and labors cost. • Optimal utilization of component service life.	• Increased investment in monitoring equipment. • Difficulty to cope with dynamic planning of the maintenance actions.

2.3.1 Data acquisition

It is the process of gathering signals from measurement sources, such as sensors attached to critical components/subsystems, and digitizing the signals for storage, analysis, and presentation on Personal Computers (PC). This process is an essential step for implementing PHM algorithms and can affect the quality of the final decisions. Generally, data collected from the critical component can be categorized into two main types: event data and condition monitoring data.

1. Event data: include qualitative information about the monitored component such as description of installation, breakdown, overhaul, causes etc., and the description of what was done to fix the failure and the severity of the repair.

2. Condition monitoring data: are the measurements related to the health condition/state of the physical asset. Condition monitoring data are very versatile. They can be vibration data, acoustic data, oil analysis data, temperature, pressure, moisture, humidity, weather or environment data, etc.

2.3.2 Data processing

Data processing lays a solid cornerstone to building reliable data driven models. Processing the data before modeling can enhance the performance of the model. It aims at converting raw sensor data into usable information. In practice, raw sensor signals are usually very complex and information about degradation process of the monitored component is not always available. Processing raw sensor data is therefore required before building degradation models. Generally, data processing methods can be divided into two main tasks, namely pre-processing and data analysis. The main goals of can be summarized as follows:

1. Enhancing the noisy raw sensor data.

2. Providing more information by understanding and/or visualizing the underlying process that generated the data.

3. Enhance degradation models performance by reducing the effect of curse of dimensionality.

4. Providing computationally effective models by reducing the measurements size.

2.3.3 Fault detection

Fault detection is the process of determining that a problem has occurred to the monitored component. The problem can be seen as any change of the monitored component from the normal state to a new abnormal state. Figure 2.2 shows an example of the fault detection process. When the system behavior lies in the nominal region, the system is considered normal. On the other hand, if the sensor data is outside the nominal region, the system is considered faulty.

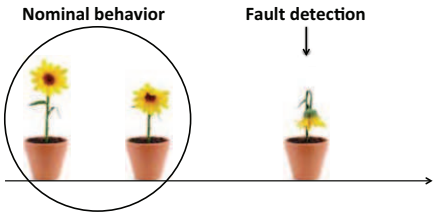


Figure 2.2: Illustration of fault detection.

2.3.4 Diagnostics

Diagnostics is the process of determining the fault type, size, location and cause identification. When a fault occurs, it affects the corresponding event and therefore the output signal. Diagnostic process starts analyzing the input signals to diagnose the fault. The final diagnostic decision is taken based on data base of previously defined fault types (Figure 2.3).

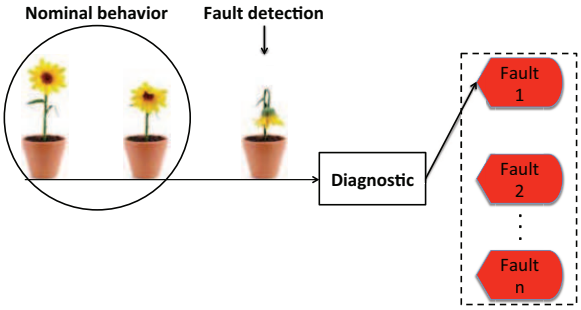


Figure 2.3: Illustration of diagnostic.

2.3.5 Prognostics

It is the process of estimating the remaining time left for a system or a component before failure. Prognostics is used by industry to manage risks that result from unexpected equipment failure. So far, it is still based on the experience of maintenance engineers. However, human decision making is not always sufficiently reliable when dealing with complex equipment. Therefore, over recent years a significant amount of research has been undertaken to develop models that can be used to reduce industry's dependence on individuals. This can be done by performing health assessment and estimation of RUL for the monitored components to plan the required maintenance actions in advance (Figure 2.4).

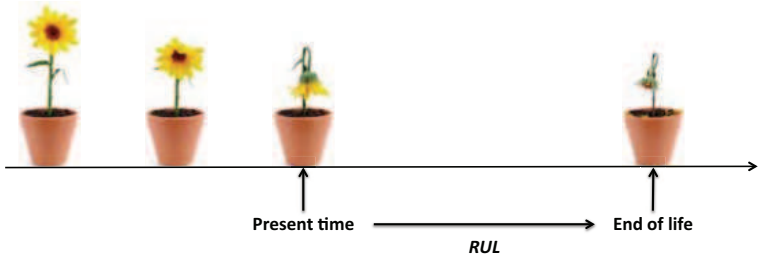


Figure 2.4: Illustration of prognostics.

2.3.6 Decision support and human machine interface

All the results generated by the previous processes can not be used directly to take maintenance decisions. Maintenance prioritization is crucial and important to reduce unnecessary maintenance activities, especially when availability of maintenance resources are limited. Decision support is the process of using all the information gathered about the monitored system status to choose the optimal maintenance actions. It includes scheduling techniques to properly plan the maintenance activities. For example, the machine degradation information, i.e. predicted RUL, the estimated health status and corresponding uncertainties, produced by this method can be used as an input for maintenance decision making routine. Decision-making routine considers both machine degradation information and system structure to assist the plant manager in making a dynamic maintenance plan based not only on the optimization of single component/subsystem plan, but also on the global scheduling of whole system for optimized maintenance prioritization. Human machine interface is a medium that handles the interaction between the PHM system and the user. It also handles the interactions between different PHM layers.

2.4 Literature review

In this section, literature review on the main PHM tasks is presented. The emphasis of the review is on data processing and prognostics algorithms; as both tasks are the core of this thesis work. Furthermore, fault detection and diagnostics will be briefly reviewed. Finally, decision support and human machine interface tasks are out of the scope of this work and will not be presented in this review.

2.4.1 Data processing

Data processing is defined as the collection and manipulation of items of data to produce meaningful information [38]. Processing sensor data is usually required before the modeling step to provide signals that are reasonably robust against different variations that might affect the raw data [39]. Also, data processing approaches can be performed

to analyze the data for better understanding. Data processing approaches vary according to the application and there is no universal method to process raw data to extract information. Instead, one has to choose the appropriate method(s) according to the problem in hand and the target. Approaches used to process data to construct fault detection models might be different than the approaches required to build fault diagnostics or prognostics models. As shown in figure 2.5, different objects can be constructed from the same material but each one will be used for different applications. State of the art on data processing for different applications, namely fault detection, diagnostic and prognostic, are presented hereafter.

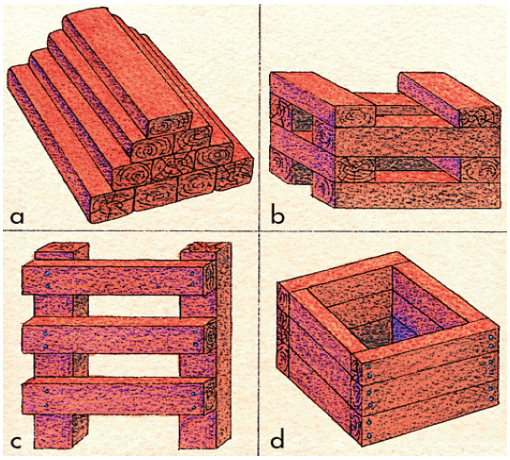


Figure 2.5: Data processing approaches are problem dependent.[1]

Data processing for fault detection

To build fault detection models, it is important to process the input data in a way that represents the normal behavior. For each new data the distance to the normal behavior should be calculated to decide if this is normal or abnormal.

For example, an online anomaly detection algorithm that is not based on any learning algorithm has been proposed in [77]. The algorithm sets different operation modes once it starts running on N observations. Any change on the number of the clusters or the behavior of the new data in an already defined cluster will be considered as suspicious behavior. The proposed method is divided into two main parts, initialization and monitoring. For initialization, 27 features were extracted, however, the authors did not specify exactly all the features. Then, data standardization has been performed by applying unit variance and mean centering. Principal component analysis (PCA) is used for data dimensionality reduction. Finally, the method estimates the operating modes by using greedy expectation maximization clustering algorithm. For monitoring part, first feature extraction has been applied. Then, calculating of the online data and

[1]Images courtesy: facpub.stjohns.edu/~wolfem/4322/chapter1/sld003.htm

updating the modeled mean and variance of the old data-set have been performed. Data standardization for the online readings was performed as in the initialization part and also dimensionality reduction. Then, the method updates operating modes clusters by using Mahalanobis distance measure. Finally, tracking OM by using evolving Takagi-Sugeno model in order to predict the dynamics of the data within each cluster.

Furthermore, PCA is used as tool to construct a reference model for fault detection in [78]. The proposed method is divided into three parts, feature extraction, modeling and deviation detection. For feature extraction, 16 signals have been measured from moving gate-type incinerator. Then, the model of the normal state is constructed using PCA and Partial Least Squares (PLS). Finally, T square and Q statistics have been proposed to detect faults. The results show that both PCA and PLS performed fairly well in fault detection.

Also, a method for unsupervised change detection and health monitoring for Diesel engines is proposed in [83]. The method is based on building a model using Independent Component Analysis (ICA). Probabilistic outlier detection algorithm has been also proposed for anomalies detection.

Another method to construct fault detection model based on Hidden Markov Model (HMM) is proposed by [79]. The method decomposes the input signal into Intrinsic Mode Function(s) (IMF) using Empirical Mode Decomposition (EMD), then Combined Mode Function (CMF) is applied to mix neighboring IMFs to obtain the best signal. Feature extraction is then performed using Fourier Transform (FT) on the acquired signal. Finally, the method uses HMM to build a model of the normal condition of the gearbox and Average Probability Index (API) is constructed as an index for machinery health status.

Gaussian mixture model is used for fault detection for electrical machines in [82]. The method learns the normal behavior over the time and detects any changes between the signals due to degradation of the system. The proposed method selects interesting features from the measured signals from the system using pairwise similarity measure algorithm and uses Gaussian mixture models for relation description. Finally, a distance measure between different signals was calculated as an indicator for the system deviation.

In like manner, two methods for fault detection are proposed in [81]. The first approach uses entropy analysis over the entire set of sensors at once to detect anomalies that have broad system-wide impact. It starts by smoothing and normalizing time-series data for each sensor. Then, sampling the time-series values using uniformly sized bins. Finally, Shannon Entropy is computed for each time step as a reference model for the system. The second approach uses automated clustering of sensors combined with intra-cluster entropy analysis to detect anomalies and faults that have more local impact. For each of the n sensors, the method computes the Pearson correlation between each pair of sensors and form an n by n distance matrix. Then, the method clusters the sensors by performing a graph-partitioning of the adjacency graph and performs a time-windowed correlation within each sensor cluster. Finally, the entropy of the m discrete values has been computed to provide the cluster entropy for the system. The strength points in this work were no parameters to tune and it is easy to implement. However, the entropy

does not differentiate between anomalies or noise which makes it difficult for the method to detect different types of faults.

Finally, a visual tool based on Self Organized Map (SOM) for fault detection and monitoring temporal evolution of aircraft engines health is presented in [80]. The environmental variables and engine effects are removed from rough measurements using general linear model. The residuals of the regression are used after that for training SOM which shows the evolution of the motor status.

Data processing for fault diagnostics

Generally, diagnostics models are built to represent different types of faults for the monitored component. It is therefore important to choose discriminant features to represent each fault.

For example, a multidimensional diagnostics approach for mechanical systems has been presented in [84]. In this work, vibration signals were acquired from diesel engine, heavy fan and rubbing blades in a turbo-set to validate the approach. The authors applied mean centering and normalization for the signals and then using Singular Value Decomposition (SVD), the multidimensional data set were reduced to a lower dimension. Finally, health evolution indexing and fault diagnostics has been proposed by choosing the most informative SVD indexes. The same authors also proposed PCA instead of SVD for its computational efficiency in [85].

Temporal models are also proposed to construct diagnostics models. For example, HMM and auto regressive moving model with exogenous input (ARMAX) are used for diagnostics as proposed in [87]. In this work, sixteen features are extracted from force signals acquired from cutter milling machine and the best group of features is selected. Finally, the authors used ARMAX and HMM to build different models for different tool wears.

In like manner, a method based on Artificial Neural Network (ANN) and support vector machine (SVM) for gear fault diagnostics is presented in [90]. The proposed method is based on selecting important features from a larger features set. The selection process shows increased classification accuracy. Finally, the authors used data-sets for 9 different gear fault classes. For each class 40 measures were recorded. Finally, the performance of SVM versus ANN are compared. The comparison shows that SVM outperforms ANN.

SOM is proposed in [86] and [92] for fault diagnostics. In the first work, Kurtosis and line integral of acceleration signal have been extracted from bearings vibration signals for different faults. Then a SOM is trained using the extracted features. EMD was also proposed in many works as it returns smooth monotonic like signals. On the later work, the authors propose an empirically derived equation that governs the proper network size for efficient diagnostics. The method shows that, it is possible to monitor and identify different range of faults if the size of the SOM is chosen judiciously.

Similarly, a method for diagnostics is presented in [91]. The method builds the diagnostics model using Naive Bayes classifier. In this method, an unsupervised feature selection method for deciding the optimal depth for Wavelet Packet Decomposition

(WPD) is performed. The result shows that, using feature selection to determine the depth for WPD led to a model with the same accuracy as a model built using a much deeper transform.

Selecting a smaller set of the training features is shown to increase the efficiency of diagnostics models. For example, a method for diagnostics and tool state recognition is presented in [88]. This work shows that selection of a smaller set of features yield in more effective results. The proposed method is divided into three parts, namely feature extraction, feature selection and tool state learning. For feature extraction, 13 features are extracted from Acoustic Emission (AE) signals acquired from a computer numerical controlled (CNC) machine. Feature selection is performed using automatic relevance determination to select features which appear to have more potential use. Finally, the tool state learning has been conducted by Support Vector Machine (SVM).

In similar fashion, feature extraction was performed for fault diagnostics in [89]. In this work, sixteen features were extracted from force sensors attached to a cutting machine. Furthermore, the performance of four classification algorithms was compared. Then, 3 features were automatically selected from the extracted features using Genetic Algorithm (GA).

Data processing for fault prognostics

Prognostics models require special data analysis techniques. The idea is to monitor how the system degradation evolves over time. Then, one can model the relation between the rate of the change and the health status of the system or the end of life value.

As an example, a method for prognostics and health assessment has been proposed in [93]. The method starts by extracting 16 time and frequency features from double suction pump vibration signals. The method then uses PCA to merge features and to project the multidimensional features vector into a compact indicator. Finally, fault threshold has been calculated using Best Efficiency Point.

Also, a method for prognostics and trend analysis using modified SOM has been proposed in [97]. The authors propose unequal scaling method for improving the performance of SOM. It shows that the SOM outperforms Constrained Topological Mapping (CTM) on estimation of an unknown function with multiple indicators.

Furthermore, an integrated framework for fault detection, diagnostics and prognostics using HMM is presented in [94]. The proposed framework starts the data preprocessing by using frame blocking, frequency spectral analysis and noise filtering. Then, the dimensionality of the data set is reduced by using PCA. Next, the health status estimator has been built using HMM and HI interpolation by using Paris's formula.

Different feature selection methods are proposed to select features that can represent the degradation evolution. For example, a single hidden semi-markov model for prognostics is proposed in [95]. In this work seven signals, three force, three vibration and one acoustics have been acquired from CNC milling machine. Then sixteen statistical features were extracted from the three force acquired signals. Then, a wavelet feature extraction approach from the force signals is applied to vibration and AE signals. Finally,

by using Fisher's discriminant ratio and Gaussian Mixture Model clustering algorithm, the important features have been selected.

In the same way, a semi-supervised feature selection algorithm for prognostics is presented in [96]. The method extracts sixteen features from the raw force signals acquired from cutting tools. Then, the authors applied SVD on the feature space to select the most dominant components. K-means clustering was applied on the feature space using n clusters and the set of closest m features to the clusters centroids have been identified. Finally, multiple regression model is applied for the selected set of m features. The authors reported the quality of the regression results using different sets of selected features and also against different feature selection approaches.

Particle Filter (PF) is widely used for fault progression modeling and estimation specially for nonlinear systems, [98]. The PF dose not assume a general analytic form for the state space Probability Density Function (PDF). The Extended Kalman Filter (EKF) is the most popular solution to the recursive nonlinear state estimation problem. However, the desired PDF is approximated by a Gaussian, which may have significant deviation from the true distribution causing the filter to diverge. In contrast, for the PF approach, the PDF is approximated by a set of particles representing sampled values from the unknown state space, and a set of associated weights denoting discrete probability masses. The particles are generated and recursively updated from a nonlinear process model that describes the evolution in time of the system under analysis, a measurement model, a set of available measurements and an estimate of the state PDF. Furthermore, PF is proposed in [99] and [100] for RUL estimation.

To conclude, it can be seen that the processing of sensory signals differs according to the task to be done, i.e. fault detection, diagnostics or prognostics. To the authors knowledge, there are no proposed methods that can be used to select/extract features that represent the behavior of the monitored component without strong initial assumptions. Such features, can be used to represent critical components' health evolution over time to build reference models that could be used for health status assessment and RUL estimation.

2.4.2 Fault detection

Fault detection is the process of determining whether a fault has occurred in the monitored component [40]. There are two main approaches for doing fault detection, namely model based and data-driven methods [41].

- Model based approaches: these methods are based on a physical model derived from the principle laws of physics. They require performing large number of experiments in the lab, prior to production, to find characteristics that can be used to detect a fault. It requires a lot of experimental work and time which can be expensive. Model based approaches can be built to represent the most frequent malfunctions in the systems. Many model based approaches have been proposed in the literature such as [42–46]. Model based methods can be further classified into three main approaches, namely

fault detection based on process models [59], fault detection of control loops [47] and state observer [48].

- Data-driven approaches: another way to do fault detection is by using data driven approaches to build models about normal behavior [49]. The methods measure the new sensor data and compare it to the available data model [50]. If the new data is not within the normal boundaries, the system is deemed faulty [51]. Methods for data driven fault detection utilizes machine learning models to represent the nominal behavior of the system [52].

2.4.3 Diagnostics

Fault diagnostic is defined as the process of determining fault characteristics such as kind, size, location, and time of detection. Usually, diagnostic process follows fault detection and the main task is to classify the fault. Many diagnostic approaches have been proposed in the literature such as [53–55] and can be divided into three main approaches, namely model based, data driven and expert system approaches.

- Model based approaches: these approaches are based on physical laws which represent the relationships between different system parameters. These methods are used to generate the rule that represent the current fault [56]. The generated rule is then passed to a rule-based inference system to deduce the exact diagnostic.

- Data driven approaches: these approaches are based on building models based on the sensor data. These methods utilize machine learning and Artificial Intelligence (AI) methods to build such models [57]. The rule is generated by measuring the distance between the new data pattern and the data models.

- Expert system approaches: these approaches are not based on physical or data driven models. They are based on representing the knowledge of experienced operators and system engineers [58]. The knowledge can be represented by using one of the following representations: 1) rules, 2) frames, 3) predicate logic or 4) directed graphs.

2.4.4 Prognostics

According to Oxford dictionary, prognostics is an advance indication of a future event. A number of different definitions of prognostics have been proposed in the literature. Collectively, some of these definitions are presented in Table 2.3.

The definition proposed in [63] offers the most all-encompassing description of prognostics and thus will be adopted in this work. We assume that the degradation of a system results from one or more incipient faults after operating the system or by increasingly stressing the system. Furthermore, RUL refers to the time left for the system, starting from the current time, after the initiation of the degradation. The observed RUL appears to be random and can be modeled as stochastic process [75]. Therefore, uncertainty bounds or confidence intervals should be applied and accompany RUL estimation [76]. Prognostics is in the core of PHM field lately due to the increased demand

Table 2.3: Various definitions of prognostics.

Author	Definition
ISO	Estimation of time to failure and risk for one or more existing and future failure modes [62].
Goebel	Estimation of the remaining useful life (RUL) of a component (or a system) based on its current health state and knowing its future operating conditions [63].
Jardine	Algorithm which predicts how much time is left before a failure (or more) occurs given the current machine condition and past operation profile [64].
Engel	The capability to provide early detecting of the precursor and/or incipient fault condition of a component, and to have the technology and means to manage and predict the progression of this fault condition to component failure [65].
Hess	Predictive diagnostics, which includes determining the remaining life or time span of proper operation of a component [66].
Wu	The prediction of future health states and failure modes based on current health assessment, historical trends and projected usage loads on the equipment and/or process [67].
Luo	Failure prognosis involves forecasting of system degradation based on observed system condition [68].
Brotherton	The ability to assess the current health of a part for a fixed time horizon or predict the time to failure [69].
Katipamula	Addresses the use of automated methods to detect and diagnose degradation of physical system performance, anticipate future failures, and project the remaining life of physical systems in acceptable operating state before faults or unacceptable degradation of performance occur [70].
Lewis	Prediction of when a failure may occur. To calculate the remaining useful life on an asset [71].
Smith	The capability to provide early detection and isolation of precursor and/or incipient fault condition to a component or sub-element failure condition, and to have the technology and means to manage and predict the progression of this fault condition to component failure [72].
Baruah	Prognostics builds upon the diagnostic assessment and are defined as the capability to predict the progression of this fault condition to component failure and estimate the RUL [73].
Heng	The forecast of an asset's remaining operational life, future condition, or risk to completion [74].

from industry for reliable prediction algorithms. This work will particularly focus on prognostic methods and their applications. In general, prognostics can be classified in model based, data-driven and hybrid methods and will be explained hereafter (Figure 2.6).

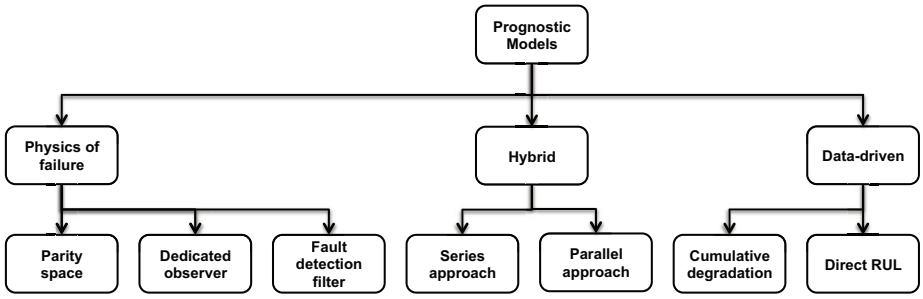


Figure 2.6: Summary of prognostic methods.

A) **Model based prognostics methods**

These methods represent domain knowledge of the monitored system and how it fails, in order to predict the RUL, using physics based models which are derived from first principles. One way to do that is by studying the physics of failure of the monitored system and represent it by a physical model. Physics of failure can be represented by measurable indicators which can be related to the fault progression such as crack propagation, corrosion rate, stress (Microscopic or Macroscopic) and fatigue (Low or high cycles). These indicators are then modeled using dynamic ordinary differential or partial differential equations [103]. These equations can be solved with Lagrangian or Hamiltonian dynamics. Another category of model based methods applies state-space models [102] to represent the nominal behavior of the critical system. By using state-space model of the critical system, the actual input to the system, measurements and the noise, the system health status can be described. This process is known as the residual generation and can be performed using parity space, dedicated observer or fault detection filter.

- Parity space: in this approach the difference between the model output and the system output, given similar input, is calculated. This difference is known as residual and a fault is detected when the residual exceeds a certain threshold. Once a fault is detected, dedicated fault progression models can be used to predict the RUL of the monitored system (Figure 2.7).
- Dedicated observer: in this approach the output of the system is reconstructed and the hidden states are estimated using both the input(s), the output(s) of the real system and a bank of observers to generate residuals of the system. Dedicated fault progression models can be used to predict the RUL after detecting fault in the system (Figure 2.8).

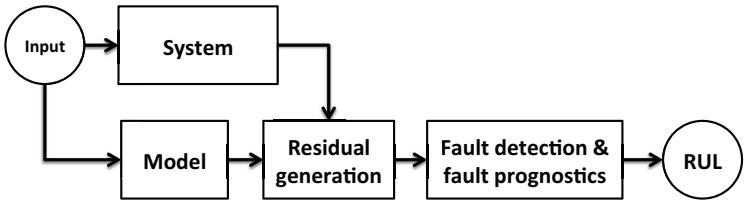


Figure 2.7: Residual generation in parity space approach.

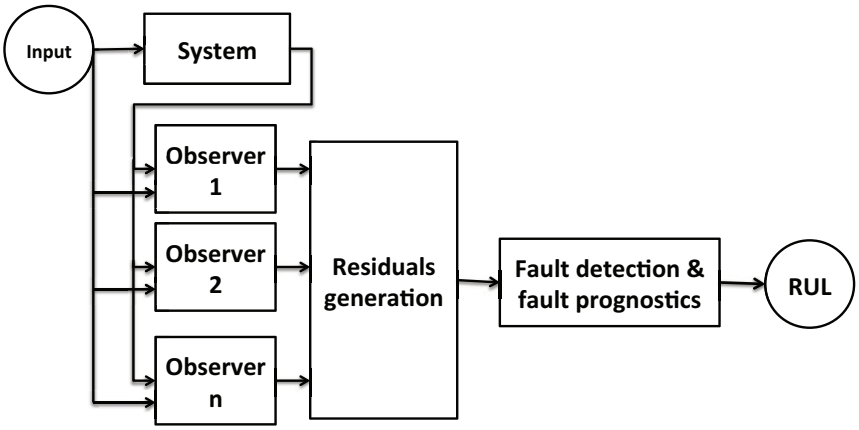


Figure 2.8: Residual generation in dedicated observer approach.

- Fault detection filter: in this approach a model is built to represent a specific parameter which reflects a certain fault. This model can be used to detect this fault once it happens and isolate it. Then, specific prognostics models can be used to estimate the remaining useful life (Figure 2.9).

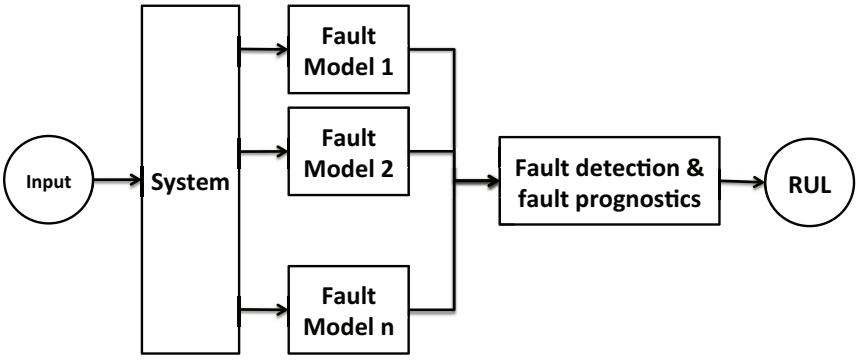


Figure 2.9: Residual generation in fault detection filter approach.

Model based methods will be very reliable once the model is built [105]. However, they require deep understanding of the physical mechanism of the failure, extensive experimentation, expert knowledge, and model verification which might be difficult in case of complex systems [104].

B) **Data-driven prognostics methods**

They are applied when the first principles of the system operation are complex such that developing an accurate physics of failure model is not feasible. Data-driven methods employ empirical models to learn from data recorded over the entire operating process of the desired critical components. In this way, the recorded historical data contains the degradation behavior, and the task is to infer this knowledge using only the data without using the first principles [106, 107]. Such methods employ pattern recognition and machine learning techniques to characterize the desired critical components' degradation behavior [108]. There are two main approaches to build data driven models.

- Cumulative degradation prognostics: in this approach, empirical models are used to map the degradation evolution of the desired system [126–128]. These models are later used to estimate the new system health status. After knowing the new system's current health status, the RUL can be predicted based on the expected future behavior (Figure 2.10). For example, different regression models have been proposed in the literature to deal with data-driven RUL prediction problem such as the auto regressive model and the multivariate adaptive regression splines [109–114]. A drawback of using regression methods is that when available component degradation history is incomplete the extrapolation may lead to large errors [129]. There have been more interests lately on various types of neural networks and neural-fuzzy systems [115–125]. However, these methods generate black box models and it is difficult to select the structure of the network [130]. Similarity-based methods can also be used to build prognostics models and they are less complex compared to neural network models. For example, a similarity-based method based on linear regression to construct offline degradation models is proposed in [129]. The method measures the similarity between test instance and offline models and the selected offline instance is used for RUL prediction. The RUL probability density of the test instance is estimated from the multiple local predictions using the kernel density estimation method. The main problem with this method is the manual selection of the informative sensor data. Another similarity-based method that utilizes k-NN and belief function theory to estimate the health and from that deduce the RUL of turbofan engines is proposed in [130]. The authors manually annotate the health status of the offline data sets and then the method predicts the RUL when the degradation level reaches a predefined alarm threshold.
- Direct RUL mapping prognostics: in this approach empirical models are also employed to build RUL models. However, these approaches directly map the relation between sensor data and the corresponding EOL value without the need to estimate the health status and from that estimate the RUL of the monitored component

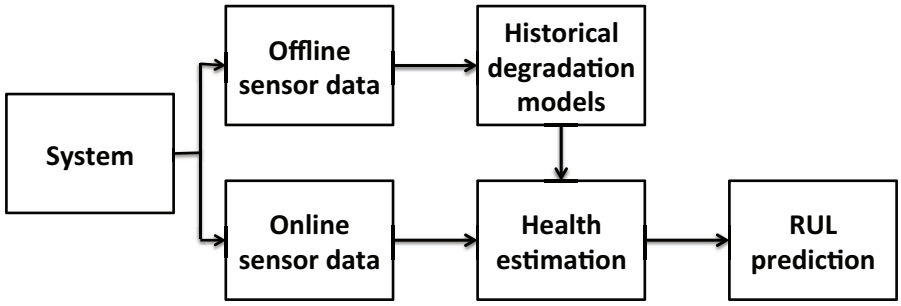


Figure 2.10: Cumulative degradation based prognostics.

(Figure 2.11). To do this, health indicators are extracted from the raw monitoring signals, which may have originated from single sensor or from a number of sensors aggregated to represent the degradation evolution over time. Although this approach is relatively easy to implement, there are few published examples in the literature [135].

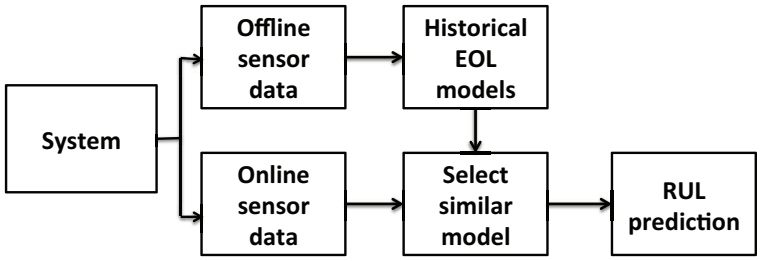


Figure 2.11: Direct RUL mapping approach.

Data driven methods can build prognostics models faster with less costs compared to other approaches. However, the main disadvantage of data driven methods is that it mainly depends on the historical data acquired from the critical systems over many run to failure iterations.

C) **Hybrid prognostics methods**

They represent the group of prognostics methods that integrate physics of failure models with data-driven models, such as online parameter updating methods, to increase the models efficiency. Indeed, the increased complexity of critical systems can complicate the process of developing model based methods. For instance, it is necessary sometimes to make strong assumptions about the system to simplify the models. Also, model based methods do not put in consideration unforeseeable conditions that might affect the system. These shortcomings can lead to huge uncertainty in the models and consequently decrease the models efficiency. There are two major types of hybrid methods, namely, series approach and parallel approach [131].

- Series approach: in this approach, the parameters that are correlated with fault progression are identified. These parameters can be measured using specific sensors from the system such as vibration, acoustic emission, load and so on [132]. The value of some model parameters can be then estimated from these signals using, for example, recursive estimators [133].

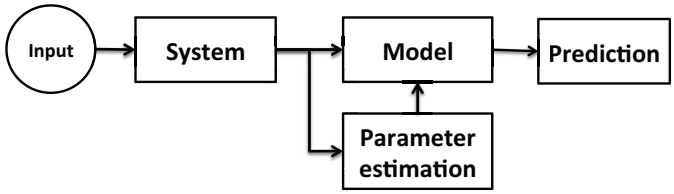


Figure 2.12: Series approach.

- Parallel approach: in this approach, data driven models are used to model processes which can not be modeled by first principles models. When in operation, both residual generated from first principle models and residual generated by data driven model are combined to get prediction with reduced error [134].

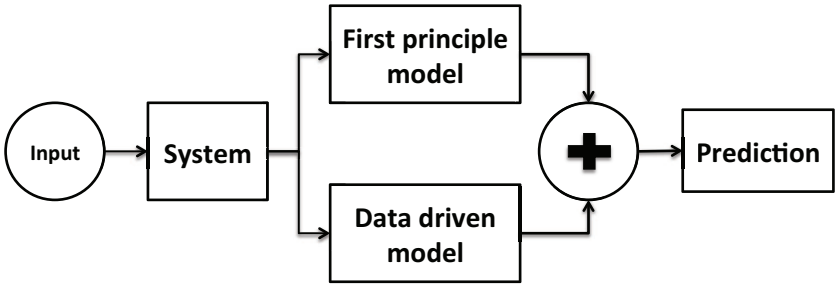


Figure 2.13: Parallel approach.

Hybrid approaches combine model-based and data-driven approaches to leverage the advantages from both approaches. In this way, such approaches can increase the efficiency of the physical models by using data driven models. On the other hand, hybrid approaches require understanding of the physical mechanism of the failure and expert knowledge about system's most informative sensor signals to estimate model parameters which might be difficult to achieve for complex systems. A summary of the advantages and disadvantages of the three prognostics methods is presented in Table (2.5).

Model and hybrid based prognostics approaches can be used when understanding of the first principles of a system operation is comprehensible. This means that, the monitored system is not complex such that developing an accurate model and verification are not expensive. Alas, most of the modern industrial systems are becoming more

Table 2.4: Advantages and disadvantages of the three prognostics methods.

Method	Advantage	Disadvantage
Model-based	Very reliable and precise once the model is built	Require deep understanding of the physical mechanism of the failure, extensive experimentation, or expert knowledge, and model verification
Data-driven	Can build models faster with less costs compared to other approaches	Depends on the historical data acquired from the systems over many run to failure iterations
Hybrid	The model performance can be enhanced based on the online data	Require understanding of the physical mechanism of the failure and expert knowledge about system's most informative sensor signals to estimate model parameters

complex and consist of multiple components with multiple failure modes. Therefore, understanding of all potential physics of failures and their interactions is almost impossible. Thus, data-driven prognostic approaches are becoming popular in the industry and research due to their intuitive nature, fast developmental cycle and the advances of modern sensor systems as well as data storage and processing technologies. These approaches are mainly based on building empirical models using massive sensory data with less requirement of knowing inherent system failure mechanisms. Recent developments in the area of artificial intelligence and machine learning algorithms have greatly expanded the capabilities of empirical modeling of the degradation process. Empirical modeling builds on analyzing the sensor signals acquired from the monitored component to find the relation between the system state variables, namely input, internal and output variables without explicit knowledge of the physical behavior of the monitored component. In general, empirical models represent a large advances in two main overlapping fields, namely Computational Intelligence (CI) and Machine learning and will be explained in more details hereafter.

I) **Computational intelligence**

It is a branch of modern AI research field, which includes approaches that mimics the nature to model complex real world problems such as, neural networks and fuzzy systems.

- Neural networks: they are computational models that mimic human brain functions. Human brains are made up of approximately 100 billion neurons, which are connected to other neurons and communicate with them via electrochemical

signals. Each neurons continuously receives signals at its inputs and then sums them up in some way. If the result is greater than specific threshold, the neuron outputs a signal along the axon (Figure 2.14).

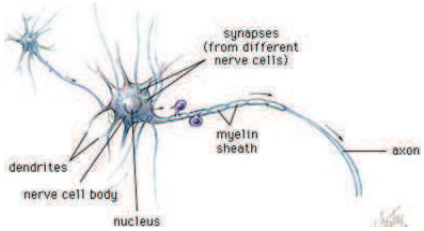


Figure 2.14: Human brain neural.

An artificial neuron is simply an electronically modeled biological neuron. The number of the neurons that should be used to construct a particular model is a problem dependent. There are many different ways of connecting artificial neurons together to create a neural network, but the most common is called a feed-forward network. A neuron can have any number of n inputs which can be represented as $x_1, x_2, x_3...x_n$. Each input connected to the neuron has its own weight associated with it, which can take a negative or positive value and can be represented as $w_1, w_2, w_3...w_n$. As each input enters the input, it's multiplied by the associated weight. The neuron then sums all these new input values:

$$a = \sum_{i=1}^{n} X_i W_i \tag{2.1}$$

If the result is greater than a threshold, the neuron outputs a signal and if the result is less than one the neuron outputs zero. Figure 2.15 shows an example of an artificial neural network with three inputs.

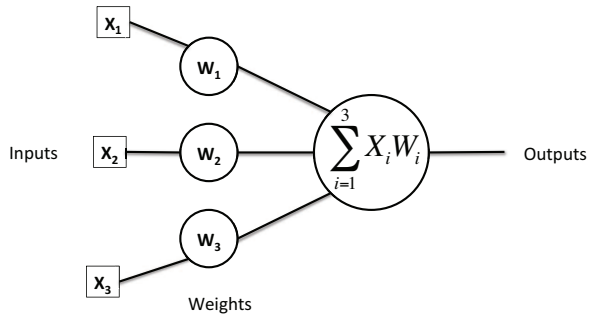


Figure 2.15: Artificial neural network with three input nodes.

- Fuzzy rule-based systems: they use fuzzy logic for inference. Fuzzy logic is based on fuzzy set theory in which binary set membership has been extended to

include partial membership ranging between 0 and 1. Fuzzy sets have gradual transitions between defined sets, which allow for the uncertainty associated with these concepts to be modeled. After defining each model variable with a series of overlapping fuzzy sets, the mapping of inputs to outputs can be expressed as a set of IF-THEN rules, which can be entirely specified from expert knowledge, or from data. One drawback of fuzzy models is that they are prone to a rule explosion. When the number of variables or fuzzy sets per variable increases, there is an exponential increase in the number of rules, which makes it difficult to specify the entire model from expert knowledge alone. The fuzzy sets and rules are referred to as the fuzzy model knowledge-base. Crisp inputs to the model are first fuzzified via knowledge-base, and a fuzzy inference engine is then used to process the rules in parallel via a fuzzy inference procedure such as max-min or max-product operations. The fuzzy solution surface resulting from the execution of the rule-base is defuzzified to produce the system output(s) (Figure 2.16).

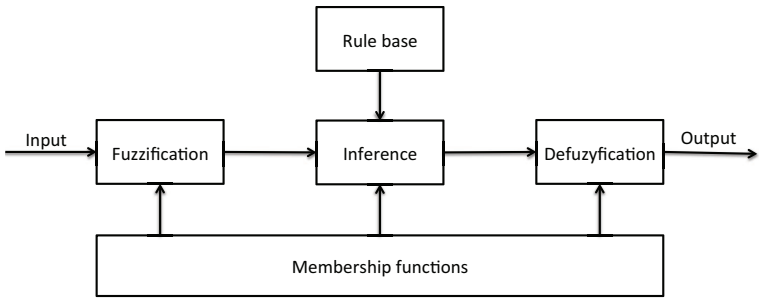


Figure 2.16: Fuzzy rule-based system.

II) **Machine learning**

It is a sub-field of computer science and statistics, which includes models that can learn from data and enhance its performance over time. It can be broadly divided into two main models, namely Bayesian based and similarity based models.

- Bayesian based learning: they use machine learning algorithms that are based on Bayes' theorem. Bayes' theorem plays a central role in machine learning algorithms and there are different efficient algorithms that perform inference and learning. The main idea behind using Bayes' theorem is that it converts a prior probability into a posterior probability by incorporating the evidence provided by the observed data. The idea of Bayes' theorem is better explained by an example. Assume a time series D acquired from a monitored component. The real process that generate the time series can be defined as w. The probability of a the real process w before observing any new data is $p(w)$. The effect of the newly observed data D can be expressed as the conditional probability $p(D|w)$. The effect of the new observed data D on the probability of real process value

$p(w)$ can be represented as $p(w|D)$. Bayes' theorem is used to calculate such probability as follows:

$$p(w|D) = \frac{p(D|w)p(w)}{p(D)} \tag{2.2}$$

where, $p(w)$ is a prior probability for the process w, $p(D|w)$ is the probability density for D given w also known as likelihood function and $p(w|D)$ is the posterior probability for process w, finally, the denominator $p(D)$ is the probability density for D, which is a normalization constant, to ensure that the posterior distribution on the left-hand side is a valid probability density and integrates to one. It can be expressed in terms of the prior distribution and the likelihood function such as:

$$p(D) = \int p(D|w)p(w)dw \tag{2.3}$$

The value of the posterior probability can be used as a prior if knew data is observed. The process can continue in an iterative manner, while updating the probability of the process and hence the knowledge of the real increases by time (Figure 2.17).

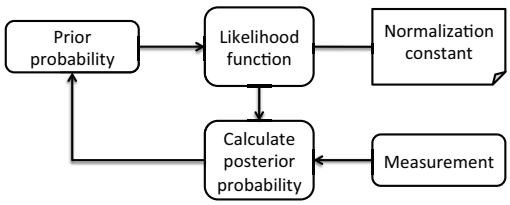


Figure 2.17: General scheme of Bayes theory.

- Similarity based learning: they are based on answering the questions of how to compare examples. If a model can compare two examples and determine whether they are semantically similar or dissimilar, the subsequent machine learning tasks would become trivial. Similarity based learning is the task of learning a distance function over objects. Usually, a metric or distance function has to obey four axioms, namely non-negativity, identity of indiscernible, symmetry and triangle inequality. The algorithm would only require to learn one known example from each category during the training phase and in the testing phase the algorithm has to group all similar examples in one category. For example, in classification settings, one would only require one labeled example per class and could then, during test-time, categorize all similar examples with the same class-label. An analogous reduction applies to regression if a continuous estimate of the degree of similarity were available. Many ML algorithms apply this approach such as, Support Vector Machines (SVM), Gaussian Processes (GP), k-nearest neighbors (kNN) and k-means. The main challenge in such approach is the choice of the similarity measure. Many similarity measures are proposed in the literature,

such as Pearson's correlation, Spearman's rank correlation and symmetrical uncertainty.

i. Pearson's correlation: it is a commonly used similarity measure for calculating the sample linear correlation coefficient for two variables X and Y as follows:

$$R_{XY} = \frac{S_{XY}}{\sqrt{S_{XX}}\sqrt{S_{YY}}} = \frac{\sum_{j=1}^{M}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{j=1}^{M}(X_i - \overline{X})^2}\sqrt{\sum_{j=1}^{M}(Y_i - \overline{Y})^2}} \tag{2.4}$$

where M is the number of the observations. The correlation has a magnitude bounded between -1 and $+1$. The value $+1$ means complete linear association between two variables and -1 means also linear association but with negative direction. The value of R remains unchanged if the measurements of the variables X and Y are changed linearly.

ii. Spearman's rank correlation: this method measures the statistical dependence between two variables by evaluating how good the relationship between those two variables can be indicated using a monotonic function defined by:

$$\rho = 1 - \frac{6\sum d^2}{N(N^2 - 1)} \tag{2.5}$$

where N is the number of observations and d is the distance between observations rank.

iii. Symmetrical uncertainty: is an information theory based method for variable comparison. In this section we review some of the fundamental concepts of information theory and then show how those concepts can be used towards assessing relationship between variables. The information entropy of a random variable X that takes on possible values in the domain $X = \{x1, x2, ...xn\}$ is defined by:

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \tag{2.6}$$

The joint entropy of two random variables X and Y is defined by:

$$H(X, Y) = -\sum_{x,y \in X,Y} p(x, y) \log p(x, y) \tag{2.7}$$

The mutual information between two random variables X and Y with respective domains X and Y is defined by:

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{2.8}$$

The mutual information is a symmetric measure that quantifies the mutual dependence between two random variables, or the information that X and Y share. It measures how much knowing one variable reduces the uncertainty about the other. The mutual information measures the information shared by two variables, and thus, their similarity. The mutual information is a non negative quantity upper bounded by both the entropies $H(X)$ and $H(Y)$, i.e. $I(X,Y) \leq min\ \{\ H(X)\ ,\ H(Y)\ \}$. If we want to use the mutual information as a similarity measure, its value has to be normalized. The normalized version of the mutual information is called symmetrical uncertainty defined by:

$$SU(X,Y) = 2\frac{I(X,Y)}{H(X) + H(Y)} \tag{2.9}$$

A feature Y is regarded more similar to feature a X than to a feature Z, if $SU(X,Y) > SU(Z,Y)$. Furthermore, SU is normalized to the range $[0,1]$ with the value 1 indicating that knowledge of the value of either variables completely predicts the value of the other variable and the value 0 indicating that X and Y are independent. In addition, it still treats a pair of features symmetrically.

A similarity measure method can be suitable for a certain application and not for others. It is therefore important to learn the method explicitly for each specific application.

To conclude, Table 2.5 presents the advantages and disadvantages of the four data-driven models. Unlike the other models, Bayesian approaches have a natural way of representing the uncertainty in a probabilistic form. This property is paramount for estimating the current health state and estimating the RUL of a critical component. The RUL appears to be random and can be modeled as stochastic process [75]. Bayesian approaches can represent uncertainty about RUL estimation. Furthermore, building Bayesian models does not require understanding the system behavior and it can be used to model multidimensional dynamic systems. Therefore, in this thesis we particularly focus on Bayesian models for building data driven prognostics method. Once the model is built and trained, it can be tested using independent data that was never used to train the model to determine how good the developed model is in generalizing unseen data. The goal of the testing part is to minimize the error between the observed output of the monitored component and the predicted output from the data-driven model (Figure 2.18). Three factors should be considered when choosing a data-driven model:

- System complexity due to increased conventionality and system dynamics.

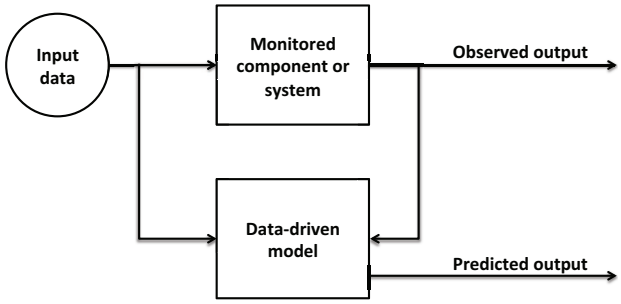


Figure 2.18: General scheme of a data-driven model.

- Knowledge of the system behavior.
- Uncertainty representation.

In this work, we propose a data-driven prognostic method based on Bayesian approaches for direct remaining useful life estimation. The method builds on two main phases, namely online and offline. In the offline phase, the method starts by looking for "interesting" variables in the form of non-random relationships among measured sensor signals, or features derived from signals. The assumption is that, information about the wear of a critical component can be extracted from the relationships between signals of the monitored component. The selected variables are then compressed, using PCA, into compact form. Then, EMD algorithm is applied to extract monotonic trends that represent the degradation of the critical component through time. Next, statistical features are extracted from the trends to represent each trend in a compact form through the time. Finally, such features are used to construct different HIs of the monitored component. In the second phase, the method uses new data that has never been used in the training. The method applies the same steps to extract health indicators from the same variables proposed by the variable selection part in the offline phase. The similarity between the online health indicators and the offline ones is measured and the most similar offline signal is used as a health predictor for the new component. The tools used for similarity measure represent the uncertainty about the decision in a probabilistic form. The method is summarized in Figure 2.19. The assumptions taken in this work can be summarized as follows:

1. The method can only be applied to critical components, which are already identified by the system expert.
2. Historical data should contain degradation evolution of the critical component over time.
3. Historical data should contain sufficient number of training instances to build representative models of the desired critical component's behavior.

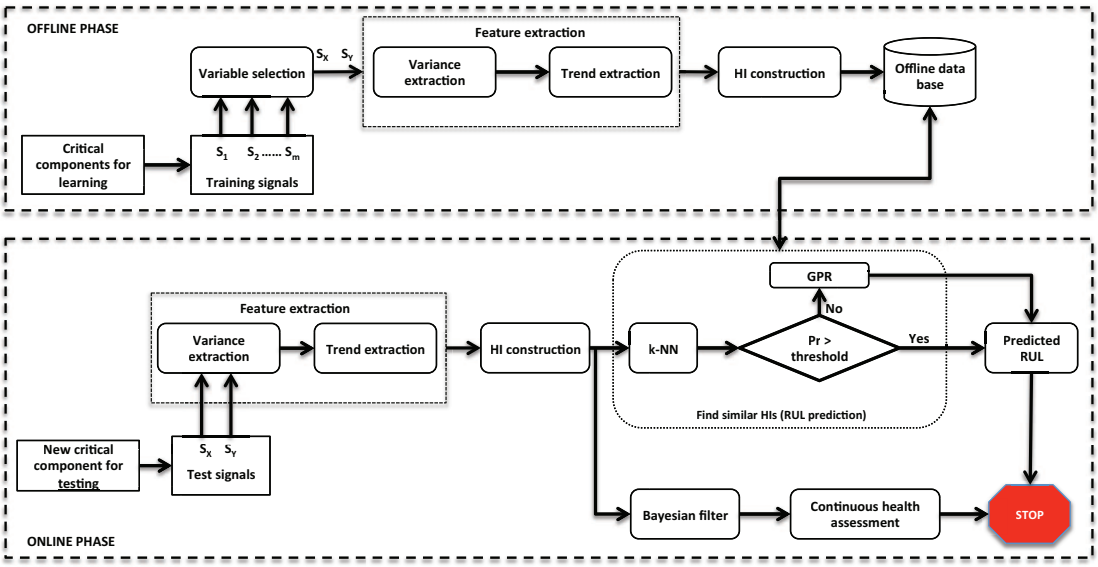


Figure 2.19: Overall scheme of the proposed method.

4. The predicted RUL values will span between the values available in the offline data sets.

2.5 Conclusion

In this chapter, we presented a state of the art research on prognostics and health management for enabling efficient maintenance strategies. Performing efficient maintenance for industrial systems can increase the reliability and availability while reducing the costs. Therefore, many types of maintenance strategies have been developed over the last decades. The most recent maintenance strategy, CBM+, emphasizes on prognostics to estimate the condition and remaining useful life of the monitored components.

Prognostics approaches can be realized using three main approaches, namely physics-based, hybrid and data-driven. The later group, in contrast to the first two, involves employing empirical models that are not derived from the physical process of the degradation. Due to the complexity of nowadays industrial systems, data driven prognostics approaches are getting increased interest. One reason is that such approaches build models which can learn the degradation behavior from the sensor data without the need of physical knowledge. In this way, building a data driven model will not be as expensive as building a physical model which in contrast needs a lot of experimental work and validation. Another reason is that the data driven model can adapt its performance with time. Therefore it can be used, with some changes, if the system is upgraded.

Empirical models can be broadly divided in two main groups, namely computational intelligence and machine learning models. Machine learning models can be further divided in two main models, such as similarity based models and Bayesian models. Bayesian based models are shown to be promising as they can be used to model multivariate and dynamic systems with unknown degradation behavior. Most importantly, Bayesian models represent the uncertainty in a probabilistic form which can be very important for later decision making step. However, Bayesian models require a signal processing step to reduce the dimensional of the input signals and to reveal hidden degradation structures.

In this thesis, we chose ML algorithms to build data driven prognostics method for critical components. First, a non linear similarity based approach is applied to select a smaller subset of the input sensor data. Such subset contains information about the degradation evolution over time. Then, different Bayesian based algorithms, such as discrete Bayesian filter, k-NN classifier and Gaussian process regression, are used to assess the health status and to estimate the RUL of the monitored component. Bayesian approaches offer strong framework that can represent the uncertainty about the estimation in a probabilistic form. The probabilistic representation of the estimations can be useful for later steps such as decision making.

In the next chapter, we present the methods used to identify critical components for an industrial system. Then, the selection process of the parameters that can represent degradation behavior of a monitored component is presented. Finally, we describe the approaches used for the selection of the sensors used to monitor the degradation parameters and signal acquisition and pre-processing approaches.

Table 2.5: Advantages and disadvantages of the four data-driven models.

Method	Advantage	Disadvantage
Neural networks	• Understanding of the system behavior is not required. • Can be used to model multivariate, dynamic systems.	• Difficulty of choosing the network structure. • Pre-processing is required. • Cannot provide uncertainty representation.
Fuzzy logic	• Interpret-able models. • Inputs can be imprecise, noisy or incomplete.	• Requires expert's knowledge and heuristics.
Bayesian methods	• Understanding of the system behavior is not required. • Represent the uncertainty in a probabilistic form. • Can be used to model multivariate, dynamic systems.	• Pre-processing is required.
Similarity based methods	• Understanding of the system behavior is not required. • Easy to implement.	• Require significant training data. • Pre-processing is required.

3

Identification of critical components and data acquisition

"The role of a museum of modern art is to make a good selection and identify what we believe to be the coming movements, and that requires taste."
– David Rockefeller

3.1 Introduction

This chapter presents an overview of the procedures that are required before constructing degradation models. The efficiency of such models depends on the quality of the available historical data. Extracting such data from the industrial system is a challenging step due to the increased complexity of modern industrial systems and due to the error and the noise that might affect the acquired signals. Therefore, system experts have to study the system to decide whether the monitoring level is system level or component level. Performing system level monitoring, however, is still challenging and quite difficult in practice. Instead, component level monitoring approaches are more feasible. To do that, system experts have to identify the critical components that need to be monitored using hazard analysis methods. Then, the experts should select the parameters that hold information about the degradation. Such parameters should be be extracted from the selected component. To extract such parameters, experts have to choose specific sensors to be placed on the selected component and perform data acquisition. Finally, pre-processing can be performed due different sources of noise that might affect the acquired signals.

In this chapter we present two main approaches to identify critical components in an industrial system, namely quantitative and qualitative methods. Then, we show how to identify the parameters which should be monitored on the critical component. Finally, different pre-processing approaches that are required to enhance the input signal quality and to remove the outliers are presented.

3.2 Identification of critical components

Deciding the monitoring level in an industrial system to acquire the monitoring data constitutes an essential step towards building reliable PHM algorithms. Traditionally, PHM approaches used control and performance data to infer fault signatures because they can provide useful information relating to behaviors of critical system being monitored. However, with increased complexity of the industrial systems, fault signatures are more complex and require additional PHM specific features to be measured from other components in the system which carry information about the fault behavior. Monitoring industrial systems can be done on two different levels, namely system level or component level.

1. System level: system level sensor placement is used with large-scale systems consisting of multiple components or/and subsystems where the fault propagates through several components.

2. Component level: components that show high failure rate are considered critical and should be monitored.

For example, building a PHM algorithm for avionics systems is paramount due to the increased risk of accidents and costs of the maintenance. Performing PHM for a whole airplane, however, is challenging and still quite difficult in practice. Instead, component-oriented PHM approaches build on identifying critical subsystems or components in the systems to be monitored and maintained individually (Figure 3.1). One of the critical components in an avionic system can be turbofan engines. It has been reported in the literature that many accidents took place due to a failure of the turbofan engine. United airlines flight 232 was a scheduled flight from Colorado, to Chicago. On July 19, 1989, the flight crash-landed in Sioux City, Iowa, after suffering catastrophic failure of its tail-mounted engine, which led to the loss of all flight controls. Of the 296 people on board, 111 died in the accident. American airlines flight 191 crashed on May 25, 1979, moments after takeoff from Chicago. Investigators found that as the jet was beginning its takeoff rotation, engine number one on the left wing separated and flipped over the top of the wing. All passengers and crew on board were killed, along with two people on the ground. It is the deadliest aviation accident to occur on U.S. soil. Another critical component in the avionic system is the lithium-ion battery. For example, in January 2013, the Federal Aviation Administration (FAA) ordered all U.S. based airlines to ground their Boeing 787s until determine what modifications are needed to reduce the electrical system risk of the battery overheating or catching fire. The focus of the review was on the safety of the lithium-ion batteries made of lithium cobalt oxide (LiCoO2). Bearing is also an important part of many subsystems in any avionic system. It can be found in moving part such motors, doors, gears, and so on. For example, Polish airlines flight 5055 crashed in Warsaw, Poland on 1987 due to disintegration of an engine shaft due to faulty bearings. Therefore, such critical components should be identified and monitored to avoid such catastrophic consequences.

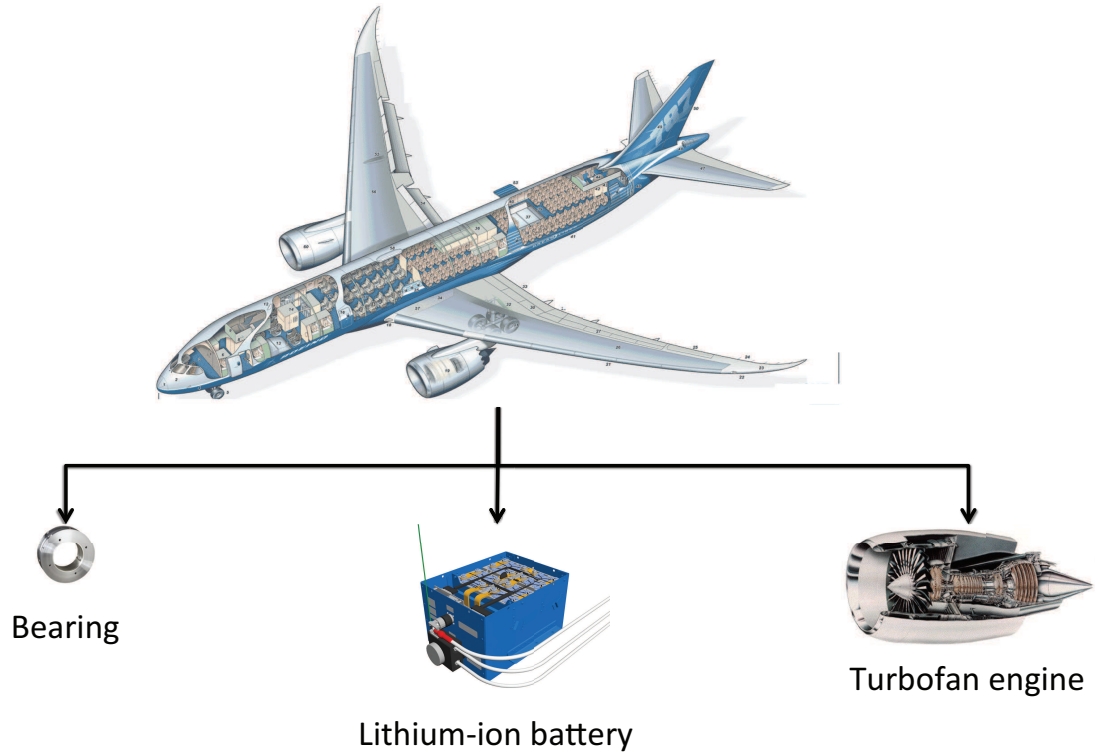


Figure 3.1: Example of critical components in a commercial airplane.

One way to identify critical components in an industrial system is by using hazard analysis [136]. Hazard analysis is a methodology to estimate the likelihood that a condition or event might happen, which could lead to an undesirable circumstance [138]. The result of a hazard analysis for a desired system is a list of all possible hazards that could result from a failed component or subsystem and their likelihood. Components with high failure rate and their failure can lead to catastrophic consequences are considered critical and should be monitored. A successful hazard analysis requires sufficient technical knowledge about the desired system and appropriate hazard analysis methodology. There are many hazard evaluation techniques which complement rather than supplant the others (Figure 3.2). Each technique approaches the system in a different way. Therefore, there is no one technique that is suitable for all situations. Generally, hazard evaluation can be divided into two main techniques, namely qualitative and quantitative.

3.2.1 Qualitative approaches

Qualitative approaches are non-mathematical techniques. They depend on experts sound judgment of the available data to identify and evaluate the potential accident scenarios in sufficient details to make a reasonable judgment of risks. Qualitative tech-

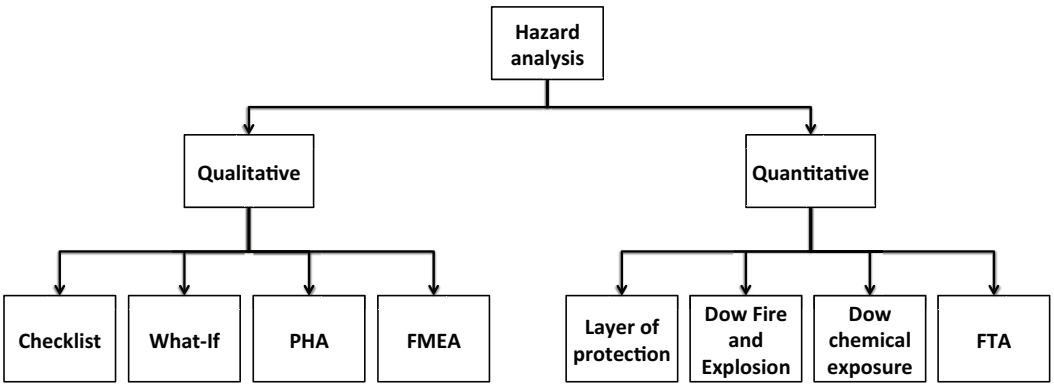


Figure 3.2: Summary of hazard analysis discussed in this section.

niques commonly employ qualitative terms such as low or high to describe the hazard introduced by a specific event. Such techniques are useful when insufficient information is available to develop a detailed hazard assessment or when the relationships between the various system processes cannot be precisely represented. If the risk is not clear, the accident scenario is identified in a qualitative terms and can be later analyzed using quantitative techniques. There are many qualitative approaches for hazard analysis.

Checklist

It is one of the simplest hazard analysis techniques. It produces a detailed list, written from experience of safety professionals, of steps for a system or operator to perform [137]. Checklist approach usually used to assess the status of the system or operation compared to the norms. Table 3.1 depicts and example of an electrical hazard checklist.

What-If analysis

It is a structured method performed by experienced review team using brainstorming approach to determine hazardous situations and evaluate the likelihood and consequences of those situations [137]. The review team starts at the lowest level of component and asks questions posed in the form of "What if" utilizing a form similar to one illustrated in Table3.2. Subsequently, the team answers each question in the list along with calculates the likelihood of the hazard, estimates the consequences and specifying recommendations if there is a need for additional action or study. The basic What - IF analysis steps are as follows:

1. Define the objectives and scope of the analysis.
2. Conduct the questioning.
3. Document the results.

Table 3.1: Example of electrical hazard checklist.

Electrical hazard	Hazard number(s)	Location	Comments
Electrical service and panels			
Wiring			
Outlets			
Receptacles			
Fixtures			
Switches			
Appliances			
General inspection required			
Other			

Table 3.2: Sample of What-IF analysis worksheet.

What If?	Answer	Likelihood	Consequences	Recommendations
What if the cooling water stops?	Overheating	Possible	Serious	Cooling water flow switch that shuts down process when it gets below certain threshold

4. Track the hazards until eliminated or controlled.

Preliminary hazard analysis

The Preliminary Hazard Analysis (PHA) is an initial activity in hazard analysis process during the early stages to identify the high level hazards and provides the foundation for future analyses based on the best available data [143]. It is often conducted to identify non-trivial system hazards during *design and development* phase, but may also be applied in the *concept definition* phase, before the system has been physically designed using a team of safety personnel associated with the design of that system. PHA produces a tabular inventory of the system hazards and assessment of their remaining

risk after countermeasures with a qualitative delineation of their predicted effectiveness. PHA analysis can be performed through the following steps:

1. Identification of valuable resources or potential targets to be protected, such as personnel, facilities, equipment, environment, etc.
2. Identification of the acceptable level of risks.
3. Definition of the physical boundaries of the system to be assessed.
4. Identification and assessment of the known hazards, their causes, effects and probabilities.
5. Risk assessment of each hazard.
6. Categorization of each identified risk as acceptable or unacceptable. Development of countermeasures if the risk is not acceptable.
7. Re-evaluation of the risk with new countermeasures installed and determination if they introduce new hazards.

The general PHA steps are summarized in Figure 3.3.

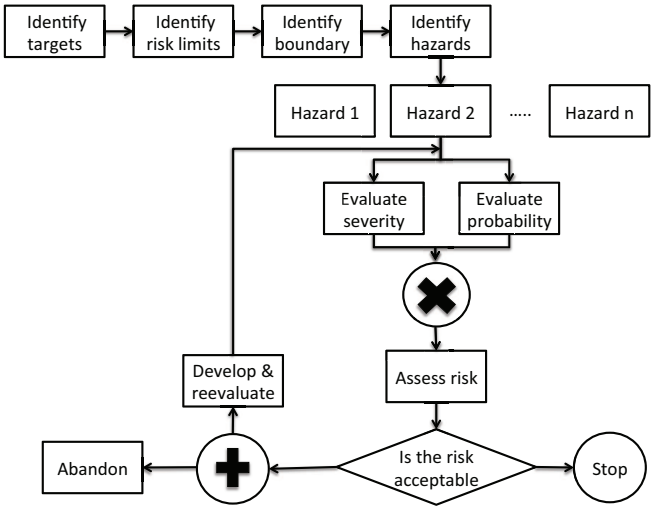


Figure 3.3: General scheme of preliminary hazard analysis.

Failure Mode and Effects Analysis (FMEA)

It is a systematic technique for analyzing component failure and documenting the resulting effect on system performance [145]. FMEA is performed by a team of experts whom thoroughly analyze product design or manufacturing processes early in the product development process. Its objective is to find product drawbacks before it gets to the

customer. Also, It can be used as a guide to the development of a complete set of actions that will reduce the risk associated with the industrial system and its components to an acceptable level. It can be applied to electrical or mechanical systems that consist of many unreliable components or subsystems such as instrument transmitters, controllers, valves and pumps. FMEA process can be performed by the following steps:

1. Identify the steps required to perform target process.
2. Specify all components that perform each step.
3. Define how each component can fail.
4. Determine the consequences of each component failure.
5. Rank the components by frequencies of failure and the likelihood of harm.

Table 3.3 presents an example of a FMEA procedure for a car rear speaker installation process at an automobile assembly line. In the example, three potential failure modes have been presented such as installing front speaker instead of rear speaker, connection cables are not covered and the speaker does not fit in rear speaker dedicated location. As can be seen from the table, the second failure mode has the highest risk performance number and therefore it has the highest priority for monitoring.

Table 3.3: Example of a FMEA for a rear speaker installation.

Process name: left front seat belt install							
Failure mode	Cause	Effect	Fault detection	Severity	Frequency	Detection	Risk preference no.
Install front speaker	Human error	Unclear sound	Sound inspection	2	3	4	24
Cables are not covered	Human error	Short circuit	Visual inspection	9	2	8	144
Speaker does not fit	Product defect	Unstable speaker	Visual inspection	5	4	3	60

3.2.2 Quantitative approaches

Quantitative techniques provide statistical evaluations of the risk of a specific scenario and can be used if the probability and consequence of events are available. Such techniques have the advantage of providing numbers that can be used to express outcomes or qualitative categories such as high, medium, and low. Below are some of the most used quantitative approaches.

Layer of protection analysis

It is based on the information developed by a qualitative hazard method to quantitatively evaluate risks [146]. It can be applied when a scenario is too complicated or the consequence is too severe for the qualitative analysis to make a sound judgment based on the available information.

Dow Fire and Explosion Index

It is a risk ranking approach developed by Dow chemical company [139]. It gives a relative index to the risk of a scenario due to expected resulting fires and explosions using material characteristics and process data.

Dow chemical exposure index

It provides rating to the hazard which can affect personnel within work area due to chemical release incidents by addressing five types of factors which can influence the effects of chemical release. These factors are, toxicity, volatile portion of material which could be released, distance to vulnerable areas and molecular weight of the material [140].

Fault tree analysis (FTA)

It is originally developed by Bell Labs for the US Air Force and was later adopted and extensively applied by the Boeing Company and other industries later on. It is defined as a graphic model of the possible factors within a system that can lead to an undesirable event [141]. FTA represents the factors and events using standard logic symbols. probability of the undesirable event can be calculated by propagating the numerical probabilities of occurrence of the factors through the model. Starting with the final event, at the top of the model, the possible causes of that event are identified at the next lower levels using Boolean logic gates. Statistical values can be assigned to each end point on a branch allowing the calculation of risk quantitatively (Figure 3.4). FTA however can be misleading if the top event is not clearly defined. Also, the human factor failures are hard to model.

To conclude, selecting an appropriate hazard analysis technique is not an easy task and often seems more of an experience than a science [147]. The main advantage of qualitative approaches is that they provide analysis of the desired system at much less time

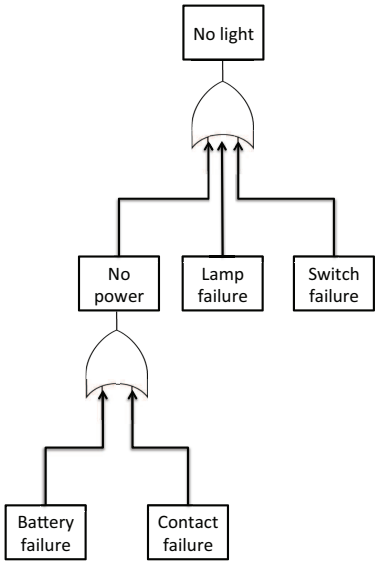


Figure 3.4: Example of Fault tree analysis.

and expense than quantitative approaches. On the other hand, when more information arise, quantitative analysis can be conducted for high risk hazards to gain more precise knowledge. Table3.4 summarizes set of attributes which can be used to characterize the quantitative and qualitative approaches.

Table 3.4: Summary of the general characteristics of qualitative and quantitative Methods.

	Attribute	Qualitative	Quantitative
1	Cost	Lower	Higher
2	Difficulty	Lower	Higher
3	Complexity	Lower	Higher
4	Data	Less Detailed	More Detailed
5	Technical Expertise	Lower	Higher
6	Time Required	Lower	Higher

3.3 Selection of physical parameters to monitor

After locating the desired component, system expert chooses the appropriate physical parameters to monitor. These parameters are chosen on the basis of experience gathered from dealing with such systems (Table 3.5). Quantities such as position, speed, acceler-

Table 3.5: General parameters which can be monitored in critical components [148].

Category	Parameter
Thermal	Temperature, heat flux, heat dissipation
Electrical	Voltage, current, resistance, inductance, capacitance, dielectric constant, charge, polarization, electric field, frequency, power, noise level, impedance
Mechanical	Length, area, volume, velocity or acceleration, mass flow, force, torque, stress, strain, density, stiffness, strength, angular, direction, pressure, and acoustic intensity or power, acoustic spectral distribution
Chemical	Chemical, species concentration, gradient, re-activity, mess, molecular weight
Humidity	Relative humidity, absolute humidity
Optical	Intensity, phase, wavelength, polarization, reflection,, transmittance, refractive index, distance, vibration, amplitude and frequency
Magnetic	Magnetic field, flux density, magnetic moment, permeability, direction, distance, position, flow

ation, torque, vibration, temperature and strain are studied for long time and chosen to monitor mechanical systems. For example, the cause vibration in different machines can be linked to fault progression. Accurate monitoring of the vibration using appropriate sensors is therefore required to monitor health status of such machines. Table 3.6 depicts an example of possible parameters which can be used to characterize failure mechanism for critical components.

3.4 Sensor selection

After locating the desired component and the parameters that represent failure propagation, the system expert chooses the appropriate sensors to record data from such component. Sensors are chosen on the basis of experience gathered from dealing with

Table 3.6: Examples of physical parameters which can be monitored in the desired component.

Component	Parameter
Bearing	Temperature, vibration and acoustics
Lithium-ion batteries	Charge and discharge voltage, charge and discharge current, temperature, voltage and battery impedance
Turbofan engine	Temperature at fan inlet, pressure at fan inlet, physical fan speed, physical core speed and demanded fan speed

such systems. A sensor is a device that receives a stimulus and responds with electrical signal (Figure 3.5).

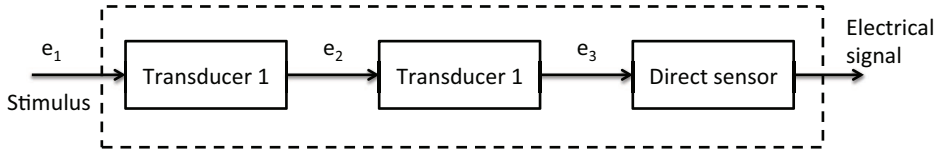


Figure 3.5: Basic structure of a general sensor.

Various sensors, such as micro-sensors, ultrasonic sensors, acoustic emission sensors, etc., have been designed to collect different types of data. Wireless technologies, such as Bluetooth, have provided an alternative solution to cost-effective data communication (Figure 3.6).

The criteria of selecting senors for monitoring a system should take in consideration six aspects.

1. Parameters: parameters to be monitored are selected based on their ability to represent degradation and/or past knowledge of the system experts.
2. Reliability: is the probability that a sensor will function without failure over a specified time or a number of uses. The sensor's useful life must be much longer than the estimated lifespan of the components/subsystems it is intended to monitor.
3. Accuracy: the highest deviation value of the value represented by the sensor from the correct value.
4. Span: representing the dynamic range of the quantities being monitored.
5. Resolution: the smallest change of the stimulus which can be sensed.

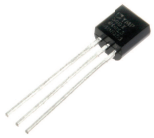
(a) Rotational velocity sensor.

(b) Torque sensor.

(c) Accelerometer.

(d) Acustic sensor.

(e) Optical sensor.



(f) Temperature sensor.

Figure 3.6: Example of different commercial sensors.

6. Characteristics properties: such as size, weight, cost, wired or wireless.

7. Cost: monitoring solution should be affordable.

Once the sensors are fixed and the system is operating, the system expert start collecting data from such system for processing tasks.

3.5 Data acquisition

Maintenance information systems, such as Computerized Maintenance Management Systems (CMMS), enterprise resource planning systems, etc., have been developed for data storage and handling. Collection of event data usually requires manual data entry to the information systems. With the rapid development of computer and advanced sensor technologies, data acquisition facilities and technologies have become more powerful and less expensive, making data acquisition for PHM implementation more affordable and feasible. One point the authors would like to make is that event data and condition monitoring data are equally important in PHM. However, in this thesis we consider only condition monitoring data.

Figure 3.7 depicts a basic scheme of data acquisition system which is composed of two main components.

3.5.1 Signal conditioning

It is the process of transforming sensor signals into suitable forms for data acquisition device. These signals can be analog or digital and measured by sensors attached to critical components which measure physical phenomena. This process is applied in situations such as dealing with high voltages, noisy environments, extreme high and low

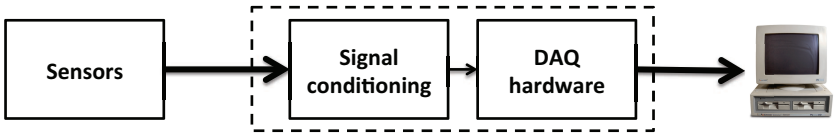


Figure 3.7: Data acquisition system.

signals, or simultaneous signal measurement to transform the sensor data into suitable forms for later processing. In this way signal conditioning can maximize the accuracy of a system and guarantees safety of the constitutive devices. Signal conditioning can include the following processes.

1. Amplification: is the process of boosting the level of input signal, if it is very small in magnitude, to a higher level which match the range of the data acquisition instrument.

2. Attenuation: is the opposite process of amplification and used when the amplitude of input signal is very high and should be diminished to a lower range suitable for later data acquisition instrument.

3. Filtering: of unwanted noise within a certain frequency range or to prevent signal aliasing when a signal is under sampled. Different kind of filters can be used to remove such frequencies, such as low-pass or high-pass filters.

4. Validation: of the data by applying sanity check, handling missing data and outlier removal.

3.5.2 Data acquisition hardware

In this step the input sensor signals are transformed into digital data by dedicated equipment/cards so that they can be stored on computers. This process introduces some errors to the signals known as quantization error, which can be reduced by calibration of the output signal.

3.6 Data pre-processing

It constitutes an initial and important step in the processing of raw data from any sensor system. It is defined as the process of manipulating an input signal to be suitable for the next stage of processing. Data pre-processing is not used to extract features or reduce dimensions of the raw signals. It is used as a preparation step to enhance the input signal quality and to remove the outliers. In this way, raw signal pre-processing reduces the computational complexity and prepare the signal for better analysis in the later step. There are many data pre-processing approaches noted in the literature such as, handling missing data, noise reduction, normalization and smoothing [149].

3.6.1 Handling missing data

The first step after recording the sensor data acquired from monitored component is to review this data to check if there are any missing values. Some observations can be missing from a stream of sensory data due to equipment malfunction or a problem in the data collection or recording mechanisms (Figure 3.8). Missing data problem is common and can have a significant effect on the conclusions that can be drawn from the data [150, 151].

Raw sensor data				Time
S_1	S_2	S_3	S_4	
Val_{11}	Val_{21}	Val_{31}	Val_{41}	T_1
Val_{12}	X	Val_{32}	Val_{42}	T_2
Val_{13}	X	Val_{33}	Val_{43}	T_3
Val_{14}	Val_{24}	Val_{34}	X	T_4
Val_{15}	Val_{25}	Val_{35}	Val_{45}	T_5
Val_{16}	Val_{26}	X	Val_{46}	T_6
X	Val_{27}	Val_{37}	Val_{47}	T_7
Val_{18}	Val_{28}	Val_{38}	X	T_8
Val_{19}	Val_{29}	X	Val_{49}	T_9
Val_{110}	Val_{210}	Val_{310}	Val_{410}	T_{10}

X Missing data
Val_{xx} Sensory data at specific time

Figure 3.8: Example of a file contains sensor data with missing values.

Examples of methods which can be used to insert the missing data include:

1. Replacement the missing data with zero.
2. Linear regression to interpolate the missing data.
3. Nearest neighbors interpolation using specific n numbers of neighbors.

In general, this step is used to replace missing values due to any error in the data acquisition step and is not in the scope of this work.

3.6.2 Noise reduction

Raw signals collected from critical components are generally contaminated by noise. Noise is defined as any random and irregular fluctuations that might perturb a raw signal, but are not part of it and tend to obscure it. Noise can be introduced by many kinds of sources, such as sensors, transmitting lines, imperfect instruments, quantization noise, interfering natural phenomena and so on. The process of reducing the noise level in a signal is known as noise reduction and it is often required before the raw data is analyzed. There are different kinds of noise which can be associated with specific signals. Table 3.7 depicts a short list of some signals and the noise associated with them.

One way to reduce the noise is by applying linear filtering to the raw signals. The output from such filters is linear function of the input. Many linear filters can be found

Table 3.7: Examples of different kinds of noise and the signals they might be associated with.

Signal	Noise	Description
Digital images	Salt and pepper noise	Dark and bright pixels in bright and dark regions respectively.
Electronics	Impulse noise	Instantaneous short peak due to sudden flaw in the system
Radio signals	White noise	Signal with constant power spectral density.
Video signals	Snow noise	Electromagnetic signals generated by cosmic microwave radiation
Electromagnetic signals	Pink noise	Signal with inversely proportional frequency to the raw signal frequency.

in the literature such as low-pass filter, high-pass filter, band-pass filter and band-stop filter. Another way to denoise raw signals is by applying nonlinear filters. The output from such filters is nonlinear function of the input. Examples of this type of filters include median filter and entropy filter. Choosing specific filter depends on the prior knowledge of the system and the nature of possible noise that might superimpose the generated signals (Figure 3.9).

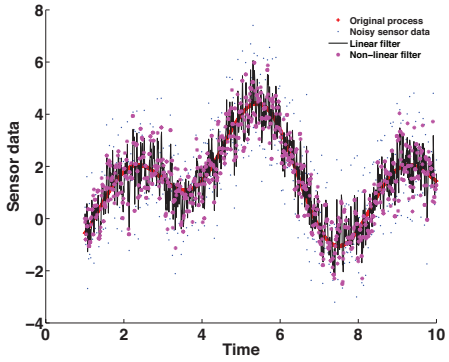
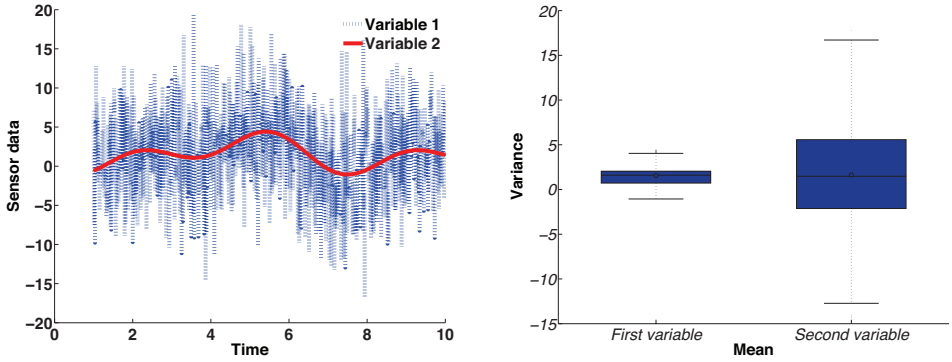


Figure 3.9: Example of linear and nonlinear noise reduction.

To conclude, this step is usually used when the data contains specific kind of noise which does not contribute to the characterization of the degradation mechanism of the monitored component.

3.6.3 Standardization

Is the process of regulating data set to have zero mean and unit variance. Each stream of data acquired from a certain sensor, which measures specific physical quantity, can be considered as random variable. This random variable will have different numerical ranges compared to other variables acquired from different sensors that measure other physical quantities (Figure 3.10).



(a) Example of two sensor data acquired from different sensors.

(b) The boxplot for both sensor data.

Figure 3.10: Example of the variability between two signals from two different sensors.

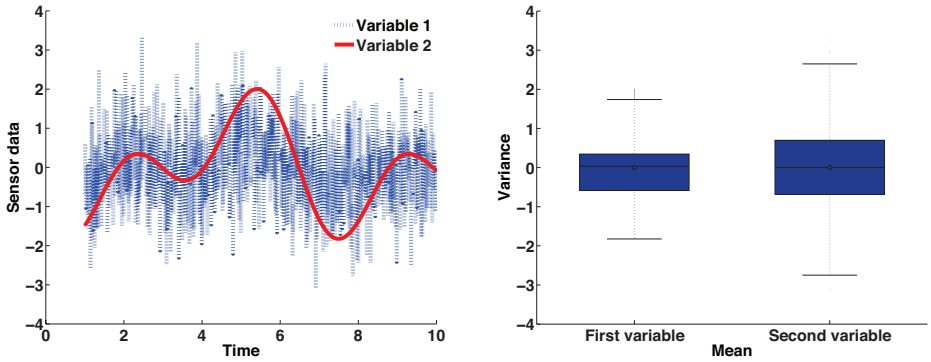
One way to standardize raw data, is by applying unit variance (UV) scaling and mean centering. For each variable, one calculates the standard deviation S_n and forms the scaling weights by taking the inverse of each standard deviation $1/S_n$. Then, each variable is multiplied by the term $1/S_n$. By using this multiplication with the inverse of the standard deviation, it ensures that each scaled variable has a unit variance and such as:

$$\sigma_n^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_{in} - \mu_i)^2 = 1, \forall i \tag{3.1}$$

Mean centering is then performed by calculating the average value of each variable and then subtracting it from the data such as:

$$\mu_n = \frac{1}{m}\sum_{i=1}^{m} x_{in} = 0, \forall n \tag{3.2}$$

The result of this process is a variable with adjusted values with zero mean and unit variance and can be compared to another normalized variable measured from a different sensor (Figure 3.11). Standardization is usually applied to raw data to facilitate the comparison between different raw sensor data measuring different physical quantities.



(a) Example of two normalized sensor data acquired from different sensors.

(b) The boxplot for both sensor data.

Figure 3.11: Example of normalizing two signals from two different sensors.

3.6.4 Smoothing

It is the process of removing the unnecessary oscillations associated with the raw signal by modifying its data points so that points with higher values than neighboring points are reduced and points with lower values are increased. The resulting signal is more smooth with reduced noise level and outliers. This can be done either by parametric fitting methods if the true process of generating the data is known or by non parametric fitting tools otherwise. Figure 3.12 depicts an example of smoothing a noisy signal using

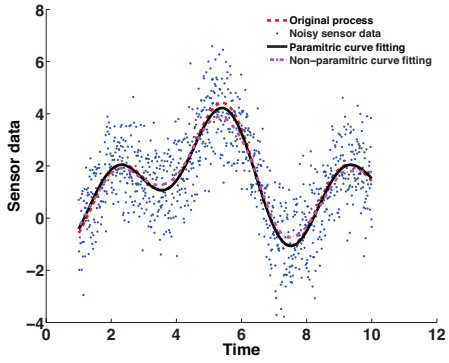


Figure 3.12: Example of smoothing a raw signal by applying curve fitting.

parametric and non-parametric curve fitting. The parametric fitting shows small error given that the true process of generating the data is known. On the other hand, the non-parametric fitting can also lead to small error but it needs careful selection of the parameters. Finally, smoothing can be applied when the oscillation of the input data is not necessary for building the degradation models.

In summary, choosing particular pre-processing method, to be applied on the raw signals, requires knowledge of data acquisition process and visual inspection of the raw signals. The results of the data collected from the monitored component are not always ready to be analyzed directly due to missing values, noise and outliers. Sensor data in this status might not be so helpful for building accurate models. Pre-processing step prepares the data for later analysis with reduced noise and less number of outliers. The pre-processed data are then ready for more analysis.

3.7 Conclusion

In this chapter, we presented the steps required to extract monitoring data sets. Extracting informative raw data is a very important step in building reliable prognostics models. The first step to do that is to identify the critical components that need to be monitored using hazard analysis methods. Selecting an appropriate hazard analysis technique, however, is not an easy task and often seems more of an experience than a science.

Furthermore, we presented two main approaches for doing hazard analysis, namely qualitative and qualitative methods. The main advantage of qualitative approaches is that they provide analysis of the desired system at much less time and expense than quantitative approaches. On the other hand, when more information arise, quantitative analysis can be conducted for high risk hazards to gain more precise knowledge. Then, we presented how the system expert should choose the appropriate sensors to record data from such component based on several criteria. Finally, pre-processing can be performed due different sources of noise that might affect the acquired signals. Choosing particular pre-processing method, to be applied on the raw signals, requires knowledge of data acquisition process and visual inspection of the raw signals. The results of the data collected from the monitored component are not always ready to be analyzed directly due to missing values, noise and outliers. Sensor data in this status might not be so helpful for building accurate models. Pre-processing step prepares the data for later analysis with reduced noise and less number of outliers. The pre-processed data are then ready for more analysis, which will be the main focus of the next chapter.

4

Data analysis and health indicators construction

"Birds are indicators of the environment. If they are in trouble, we know we'll soon be in trouble."

– Roger Tory Peterson

4.1 Introduction

In this chapter a method for data analysis of raw sensor signals is presented. The main goal of the proposed method is to transform run-to-failure multidimensional raw sensory data into more comprehensible form, such as health indicators (HI) (Figure 4.1). Such HIs can show the degradation evolution of the monitored component over time. Also, HIs are used by machine learning algorithms to assess the current health status and to estimate the RUL. The main challenge is to extract relevant information from the sensor

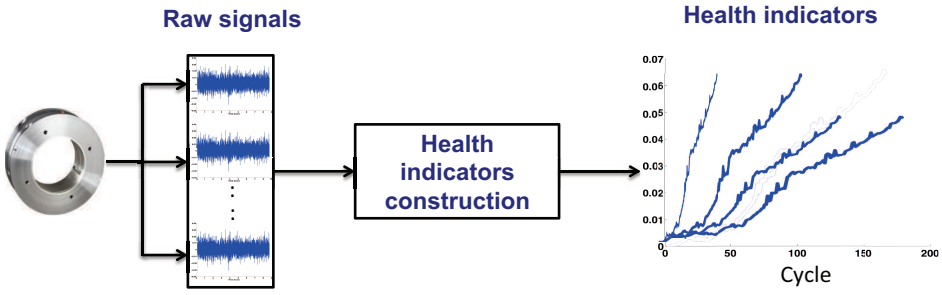


Figure 4.1: The main goal of proposed health indicators construction method .

data such that it can represent the component deterioration over time. This is because sensor data are usually multidimensional and obscured by noise.

In this chapter, we first present an overview of different data analysis approaches for extracting information about the degradation behavior, such as variable selection, feature extraction and dimensionality reduction. Then, we present a method for HI construction, from multidimensional sensor data. The assumptions taken into account in this work can be as follows:

1. The method can only be applied to critical components which are already identified by the system expert.

2. The input to the proposed method are multidimensional, non stationary and non linear time series sensory run-to-failure data acquired from the monitored component.

3. Historical data set ends when the monitored component reaches its EOL condition.

4.2 Data analysis literature review

The performance of machine learning algorithms can be enhanced when the set of features are uniquely describing the degradation level of the monitored component. Various data analysis techniques have been developed to process raw data to extract useful information for further degradation modeling. Unfortunately, there is no a general rule about how to choose a particular approach. Furthermore, data analysis phase mainly depends on the type of the application and the nature of the available sensor data [152–154]. Data analysis approaches can be broadly divided in three overlapping categories, namely, feature extraction, feature selection and feature reduction (Figure 4.2).

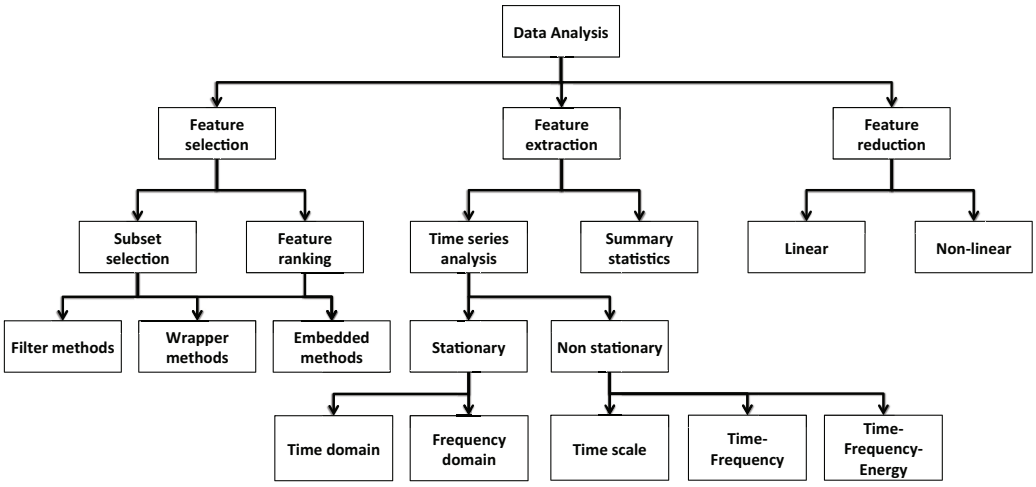


Figure 4.2: Summary of data analysis techniques.

4.2.1 Feature extraction approaches

Using raw data for modeling is usually inefficient and may even obscure interpretation of the process behind generating the data. Thus, it is essential to transform the raw data into a comprehensible form, which can be used to build reliable models. The process of transforming the input data to another informative form using linear or nonlinear functions is known as feature extraction [160]. The generated features can be multivariate or uni-variate. The main task of feature extraction is to define a set of characteristics which will meaningfully represent the information that is important for analysis and modeling. The objective of feature extraction can be summarized in three main points:

1. Remove irrelevant information from the input data.
2. Summarize the data to reduce the execution time and to avoid curse of dimensionality.
3. Represent the data with unique signatures.

Different approaches have been proposed for extracting features such as mean, variance, multi-exponential function, curve fitting, discrete wavelet transform and discrete Fourier transform. However, selecting an appropriate approach is mainly problem specific. Feature extraction categories can be classified in two different groups, such as summary statistics and time series analysis.

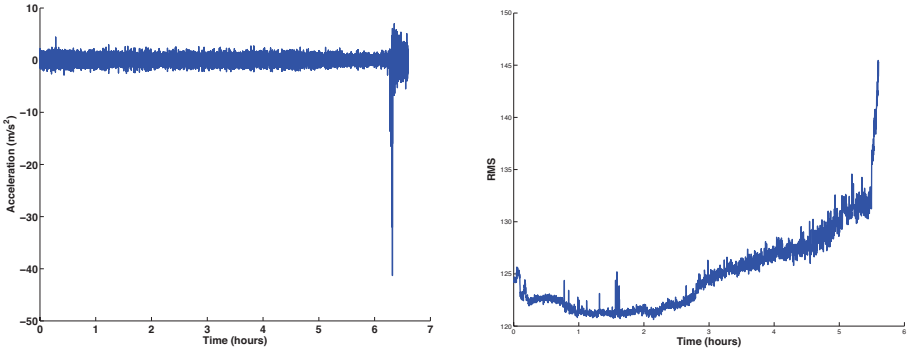
Summary statistics

Gives a quick and simple description of the data by returning a uni-variate summary of the overall input data. The idea is to summarize a set of observations to deliver a large amount of information by using one simple value without loosing the generality. To illustrate how summary statistics can be applied in PHM context, suppose the following example. An acceleration signal acquired from a critical component has been sampled at a specific regular interval. The sampled data, at each one second, are then saved in one file (Figure 4.3a). As can be seen from the figure, it is difficult to see any changes on the signal over the time apart from some abrupt changes towards the end of the signal. If one feature is used to represent each file, there will be less data to show and more knowledge might be available. Figure 4.3b shows the result of extracting root mean square from each file. As can be seen, the extracted signal shows a change over the time which might give better understanding compared to the original data.

In this way, summary statistical methods enables:

1. Reduction of the dimensionality of the input data.
2. Interpretation the input data to gain a better understanding.

Furthermore, the most commonly used types of summary statistics are measures of location and they are summarized in Table 4.1.



(a) Hidden structure.

(b) Uncovered relation between variables.

Figure 4.3: Example of how feature extraction can reveal hidden structures in the data.

Time series analysis

Monitoring data points acquired over time from critical components can be descried as time series signal. Time series signals are usually highly oscillating and obscured by noise. A time series $\{X_t\}$ is said to be strictly stationary if for any finite sequence of integers $t_1, ...t_k$ and shift h the distribution of $(X_{t1}, ...X_{tk})$ and $(X_{t1+h}, ...X_{tk+h})$ are the same. In contrast non-stationary time series can simply be defined as signal that is not stationary [155]. Usually, it is very important to analyze the signal for better interpretation of the generating process using time series analysis methods. Time series analysis can be then defined as the process of extracting meaningful statistics and internal structure of a given time series signal. Methods for time series analyses may be divided according to the nature of the time series signal into two classes, namely stationary and non-stationary time series analysis.

- Stationary times series analysis: methods for stationary time series analyses can be divided into two main categories:

  (a) Time domain methods: are the most common processing approaches. They are based on analyzing the input signal in the time domain. One of the simplest time series models is the auto-regressive models $AR(p)$ of order p in which the current output is a linear combination of the past p outputs plus a white noise input. The weights on the p past outputs minimize the mean-square prediction error of the auto-regression:

  $$X_t = b_1 X_{t-1} + ... + b_p X_{t-p} + \epsilon_t \tag{4.1}$$

  where, X_t is the current value, $X_{t-1} + ...X_{t-p}$ are immediate p past values and ϵ_t is the white noise. The estimated parameters can be used as features to identify the corresponding time series signal (Figure 4.4).

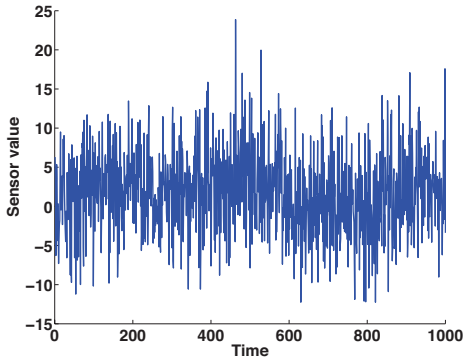
  (b) Frequency domain methods: they are based on analyzing the input signal in the frequency domain. Such methods show how much of the input time series signal

Table 4.1: Example of summary statistic features extracted from raw signals

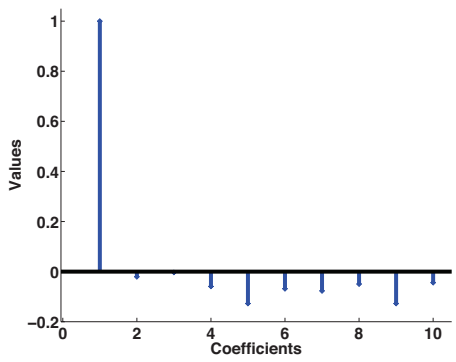
Feature	Formula
Mean	$\mu_n = \frac{1}{m}\sum_{i=1}^{m} x_{in}$
Variance	$\sigma_n^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_{in} - \mu_i)^2$
Root mean square	$\sqrt{\frac{1}{n}(x_1^2 + x_2^2 ... + x_n^2)}$
Kurtosis	$\frac{E(x-\mu)^4}{\sigma^4}$
Skewness	$\frac{\sum_{i=1}^{N}(x_i-\bar{x})^3}{(N-1)\sigma^3}$
Peak-to-Peak	$mean(upper_{pks}) + mean(lower_{pks})$
Maximum peak value	$max(findpeaks(signal))$
Mutual information	$I(X,Y) = H(X) + H(Y) - H(X,Y)$
Entropy	$H(X) = -\sum_{x\in X} p(x)\log p(x)$
Line integral	$i = \sum_{i=0}^{n} abs(x_{i+1} - x_i)$
Autoregressive model	$x_t = c + \sum_{i=1}^{p}\phi_i x_{t-i} + \epsilon_t$
Energy	$e = \sum_{i=0}^{n} x_i^2$

falls within each given frequency band over a range of frequencies. A common way of extracting features based on frequency domain analysis is Fourier transform representation [166]. The output of this transformation is a representation of the frequency content of the input signal. Frequencies that contain the interesting amplitude can be selected and the rest of the frequencies can be ignored. Selected frequencies can be used as set of features which can represent the input signal (Figure 4.5).

- Non-stationary times series analysis: in reality, stationary signals are approximation of the real life time series. Usually monitoring signals are non stationary time series, which require time-variant analysis techniques. For example, the traditional time domain or frequency domain analysis can not describe the changes in the frequency domain with the time. The past few years witnessed major developments in the domain of time series analysis. Many new time-variant analysis methods have been proposed

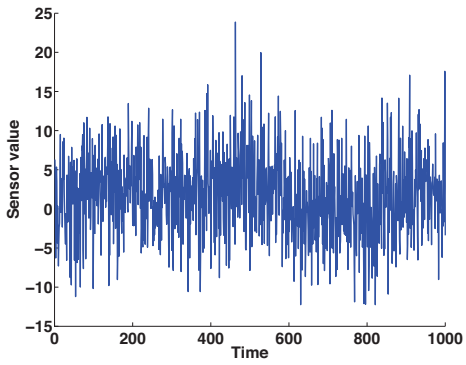


(a) Input time series.

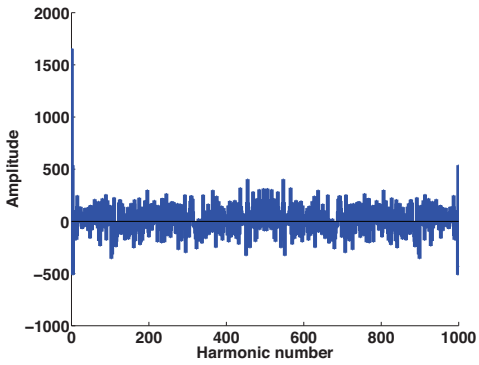


(b) Paramters of auto-regressive model of order 10.

Figure 4.4: Example of how AR can be used for feature extraction.



(a) Input time series.



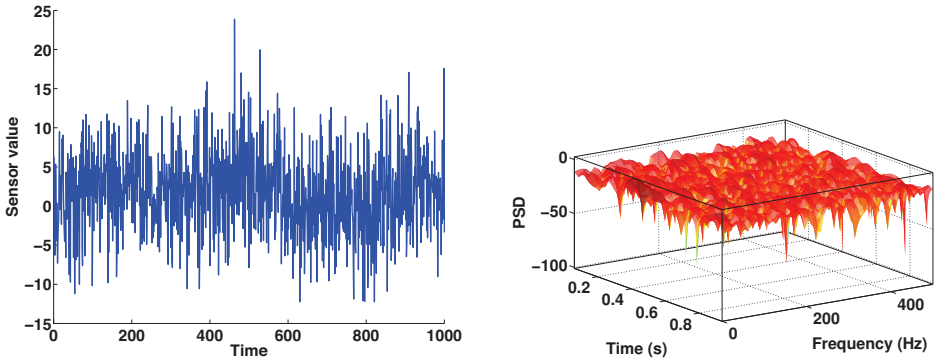
(b) Amplitude of the discrete Fourier transform for the input time series.

Figure 4.5: Example of how Fourier transform can be used for feature extraction.

and can be classified into three main categories, namely time-frequency, time-scale and time-frequency-energy methods.

(a) Time-frequency: these techniques provide a bridge between the two representations, time and frequency, to provide temporal and spectral information in the same time. This approach analyzes the input time series signal in both the time and frequency domains simultaneously. The idea is to analyze the frequency content of the input time series within a fixed size window which is moving along the input time series over time. A common way to perform time-frequency on time series is know as Short-Time Fourier Transform (STFT) [167]. In this method, the input signal is divided into many parts using a sliding fixed size window that moves along the time axis. Fourier analysis is then applied for each window. The resulting transformation is a function of time and frequency. In this way

STFT can represent how the frequency is changing over the time which can be a discriminative feature. Figure 4.6 shows the result of applying STFT to a time series signal.



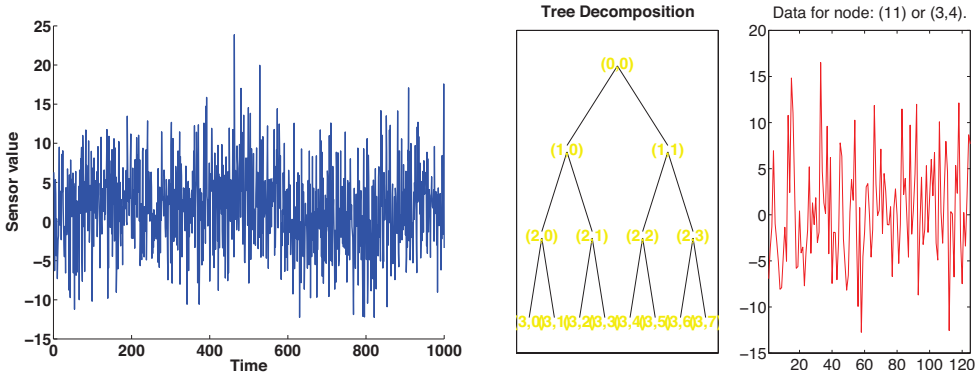
(a) Input time series.

(b) PSD of the discrete short fast Fourier transform for the input time series.

Figure 4.6: Example of how short time Fourier transform can be used for feature extraction.

(b) Time-scale: time series signal can be represented as overlapped basis functions localized in time. In this case, these basis functions can be used to represent different frequency contents by scaling them with respect to time. Analyzing a signal in this way is know as time-scale signal decomposition. One way to do that is called wavelets decomposition, which can approximate time varying non-stationary signals in a better way than the Fourier transform and can easily detect local features in the input time series signal. Wavelet Packet Decomposition (WPD) is one of the most popular time-scale analysis approaches [168]. WPD is a generalization of wavelet decomposition that offers a richer range of possibilities for signal analysis. It allows analysis of the input signal at different resolution levels. WPD divides each signal into a component containing two parts, namely, detail part and approximation part. Both the detail and approximation coefficients are then decomposed. These coefficients can be used as features that represent specific time series signal (Figure 4.7).

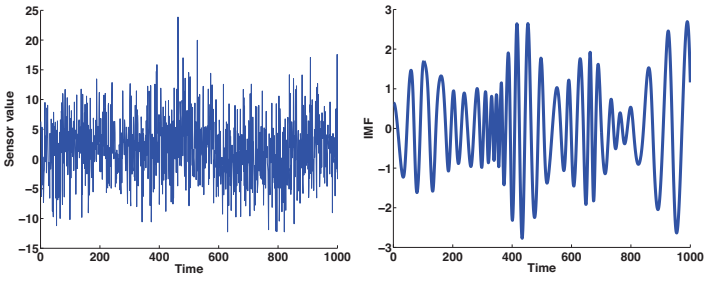
(c) Time-Frequency-Energy: it is a category for time series analysis in which the input signal is represented with frequency-time-energy distribution. This method is known as Hilbert-Haung Transform (HHT) which performs Empirical decomposition of the input signal (EMD) [169]. This method is suitable for processing nonlinear non-stationary time series signals. It is composed of two main procedures. First, the EMD is applied to decompose the input time series into all possible Intrinsic Mode Functions (IMF). The decomposition process stops when the residual of the input time series is monotonic or constant signal. Then, Hilbert transform is applied to the IMFs to extract the instantaneous frequency and amplitudes. The residual signal, instantaneous frequency or instantaneous



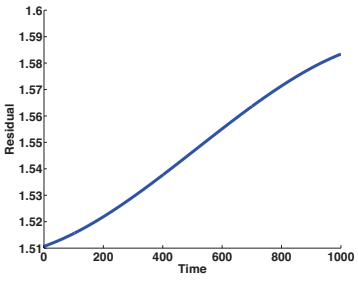
(a) Input time series. (b) Four levels WPD tree.

Figure 4.7: Example of how WPD can be used for feature extraction.

amplitudes can be used as different features that represent a particular input signal (Figure 4.8).



(a) Input time series. (b) Selected IMF signal.



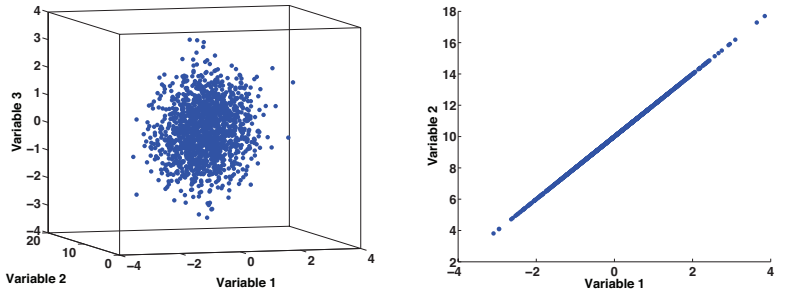
(c) Residual signal.

Figure 4.8: Example of how EMD can be used for feature extraction.

4.2.2 Feature selection approaches

Feature selection is the process of selecting smaller subset(s) from the data and neglect the irrelevant or weakly relevant subsets. The assumption is that not all features are im-

portant for modeling. Some features do not contribute to the model and removing them from the training set will increase the model efficiency. The main difference between feature extraction and feature selection is that the former generates new numerical representation of the input data whereas feature selection returns a subset of the original data. Also, selecting sub set of features can actually lead to better understanding of the process behind generating the data by keeping the sample points in a subspace of the original space. Figure 4.9a shows a plot of three variables where the variables seem to be random while figure 4.9b shows that if one variable is removed the relation between the remaining variables is perfectly linear. Thus, feature selection is important to reveal hidden structure with the data while keeping the original signals unchanged. Moreover,



(a) Hidden structure.

(b) Uncovered relation between variables.

Figure 4.9: Example of how feature selection can reveal hidden structures in the data.

dimension reduction of the data leads to performance improvement which can be applied either to the observations, variables or to both. The objective of feature selection can be summarized as follows:

1. Removing irrelevant features to increase model reliability.
2. Increasing comprehensibility of the data by finding hidden relations or structures in high dimensional data.
3. Reducing the execution time.

The selection of features can be performed in two different ways; features ranking and subset selection [156, 157]. *Features ranking algorithms* rank all available features according to some criterion and then select top k features where k can be determined by the user setting or analytically [158]. They can be used as a pre-processing methods because of their computational simplicity. *Subset selection algorithms* automatically assign a score to each possible subset of features based on specific criteria [159]. The selected subsets of features can be used for modeling or better understanding the hidden structures. In terms of the dependence of the learning models, feature selection methods can be classified into three main categories, namely filter, wrappers and embedded methods [160].

- Filter methods: they return a relevance index $R(A|B)$, which evaluates how relevant a given feature subset A is for the task Y given the data B. Relevance indexes can be implemented using correlation functions or information based function (Figure 4.10).

Figure 4.10: General scheme of filter method for feature selection.

- Wrappers methods: they search for a good subset by assessing subsets of features according to their usefulness to a given induction algorithm. The idea is that the induction algorithm runs on the data set with different sets of features. Features are removed from the data and the efficiency of the induction algorithm output is evaluated. The feature subset with the highest evaluation is selected for training (Figure 4.11).

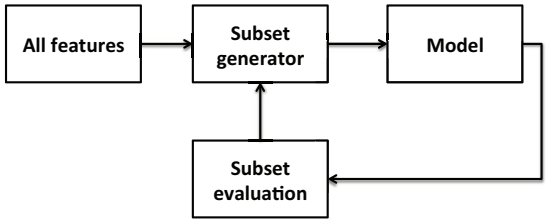


Figure 4.11: General scheme of wrapper method for feature selection.

- Embedded methods: these methods perform variable selection as part of the learning procedure and usually are built into induction algorithms. This approach has less computational complexity compared to wrapper algorithms (Figure 4.12).

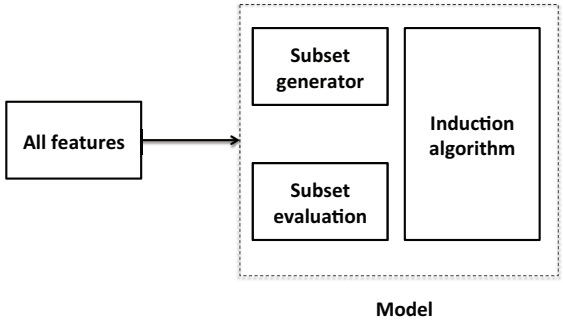


Figure 4.12: General scheme of embedded method for feature selection.

4.2.3 Feature reduction

It is the process of reducing the size of the input signal for better modeling and understanding of the input signal. The result of this step will be a compressed signal in a new domain which can reveal hidden structures in lower dimensions. One approach for data compression is the transform coding, which is usually applied for audio and image signals. It projects input vector x to a vector y using linear or nonlinear transformation T where:

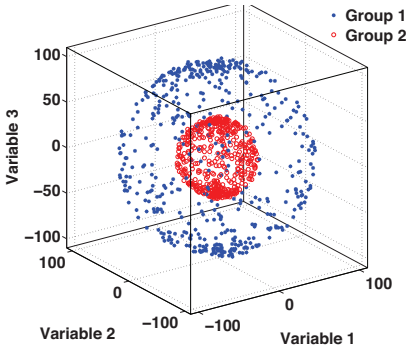
$$y = T(x) \tag{4.2}$$

The resulting vector y carries most of the information required for later processing, modeling or interpretation. The compression is done by choosing some y elements that contain the majority of the information and discard the rest. The result may not be identical to the original signal, but is expected to be close enough. The most used technique to perform transform coding is Karhunen-Loeve transform which is also known as principal component analysis (PCA) in machine learning community [161]. PCA can be defined as the orthogonal projection of the data onto a lower dimensional linear space known as principal subspace such that the variance of the projected data is maximized. The input to the PCA algorithm is a multidimensional data matrix. PCA projects the input data into a new multidimensional matrix, with the same size of the original data, and is called principal component scores. The elements of new projected matrix have the following characteristics:

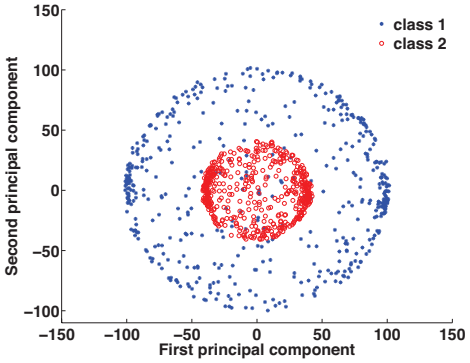
1. Linear combination of the original data.
2. The first component contains the maximum variance of the original data. The remaining components contain a less variance until the last component which contains the least amount of the variance.

The compression is done by choosing subset of the resulting components and neglect the rest. Another technique for doing transform coding is by using Kernel PCA (KPCA) [162]. KPCA is an extension of standard PCA which generalizes it to nonlinear dimensionality reduction using kernel methods. Figure 4.13 shows an example of the advantage of the KPCA over PCA. Figure 4.13a shows an example of three variables having two concentric sphere surfaces relationship. The total number of all data points can be divided into two main groups. The first two components of the standard PCA projection, depicted in Figure 4.13b, does not reveal the hidden structure. However, the first two components of the KPCA projection shows two separate groups in a lower dimension (Figure 4.13c). In this way, KPCA can reveal hidden structures in data that has nonlinear relationships which standard PCA could not reveal.

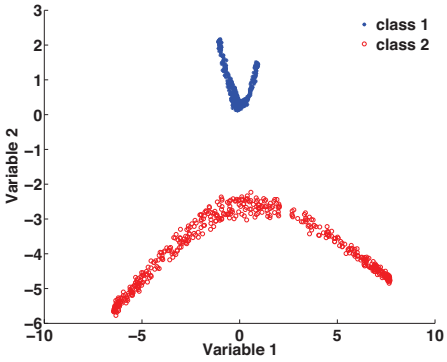
To sum up, choosing particular or group of method(s) for signal processing depends on both the problem of interest and the critical component to monitor. Before choosing specific data analysis method it is important to study the signal for getting more insight of how the degradation is progressing over time.



(a) Example of three variables have two concentric sphere surfaces relation.



(b) First two components of standard PCA.



(c) First two components of standard KPCA.

Figure 4.13: Example of how nonlinear projection can reveal hidden structures in the data.

4.3 The proposed method

The data acquired from the monitored component can not be used directly to build a model due to their redundancy and high dimensionality. Different data analysis techniques should be used to discover useful information about degradation process. Such information can be used to:

1. Present the sensor data in more comprehensible way.
2. Build more reliable models.

The information retrieved from the processed signal are know as features. *Features* can be defined as functions of the original measurement variables which represent relevant information for later modeling tasks. Features can be developed in the form of variables or observations (Figure 4.14). Such features are used to construct one or more Health Indicator (HI). A *HI* can be defined as a set of features extracted from monitored component which represent the component's degradation evolution as a function of time.

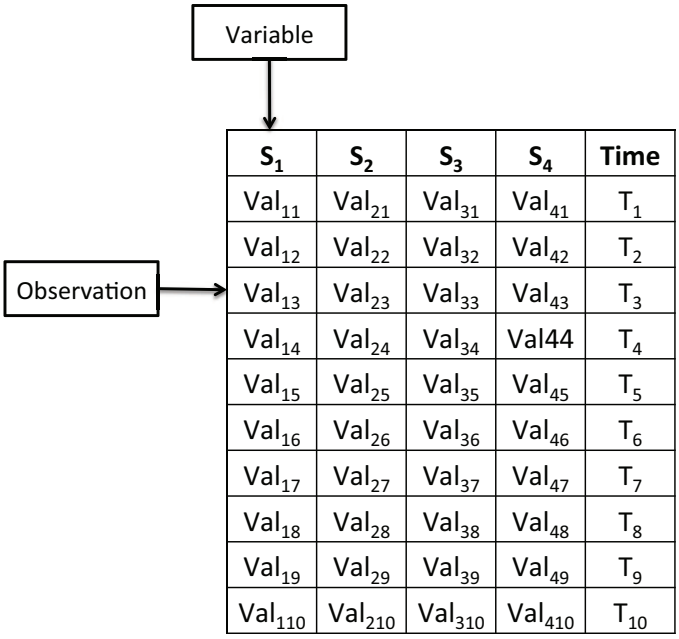


Figure 4.14: Features can be in the form of variable or observation.

However, construction of HIs from collected raw sensor data in a practical working environment is always a great challenge as sensory signals are usually multidimensional and obscured by noise.

In general, raw sensor data can be represented as time series signals, which are a collection of random variables ordered in time, such as:

$$X = X_t : t \in T, \tag{4.3}$$

where T is the index set. A time series is said to be stationary if its mean and variance are constant over time. Such signals are used to describe systems that can only evolve independently of time. In contrast, a non-stationary time series signal will have time variant mean and/or variance and is often used to represent the evolution of a system over time.

In this work, it is assumed that, the monitored critical component starts its life cycle in a healthy status and degrades through time and non-stationary time series signals are used to describe the acquired sensor signals. A non-stationary time series signal recorded from sensor S_i, until it reaches the EOL criteria, can be defined as random variable X_{ji} such that

$$X_{ji} = \begin{bmatrix} x_{ji} \\ : \\ x_{ni} \end{bmatrix} \tag{4.4}$$

where, $j = 1, ..., n$, is the time at which the sensor reading is observed and n is number of the last cycle at which the component reached the EOL criteria. Many sensors are used to record data from a critical component, therefore, the total sensor signals acquired from the same component can be defined as an $n \times m$ matrix D_{nm} such that

$$D_{nm} = \begin{bmatrix} x_{jk} & ... & x_{jm} \\ : & ... & : \\ x_{n1} & ... & x_{nm} \end{bmatrix} \tag{4.5}$$

where, $k = 1, .., m$, is the variable/sensor number and m is the total number of variables.

In this section, we describe the proposed signal analysis method (Figure 4.15). It builds on finding variables that contain information about the degradation behavior using unsupervised variable selection method. The relation between the selected variables and the End Of Life (EOL) time is mapped using two steps feature extraction methods, namely PCA and EMD. Finally, four different HIs are constructed from the mapped relation to represent the degradation as a function of time.

Figure 4.15: Scheme of the proposed health indicators construction approach.

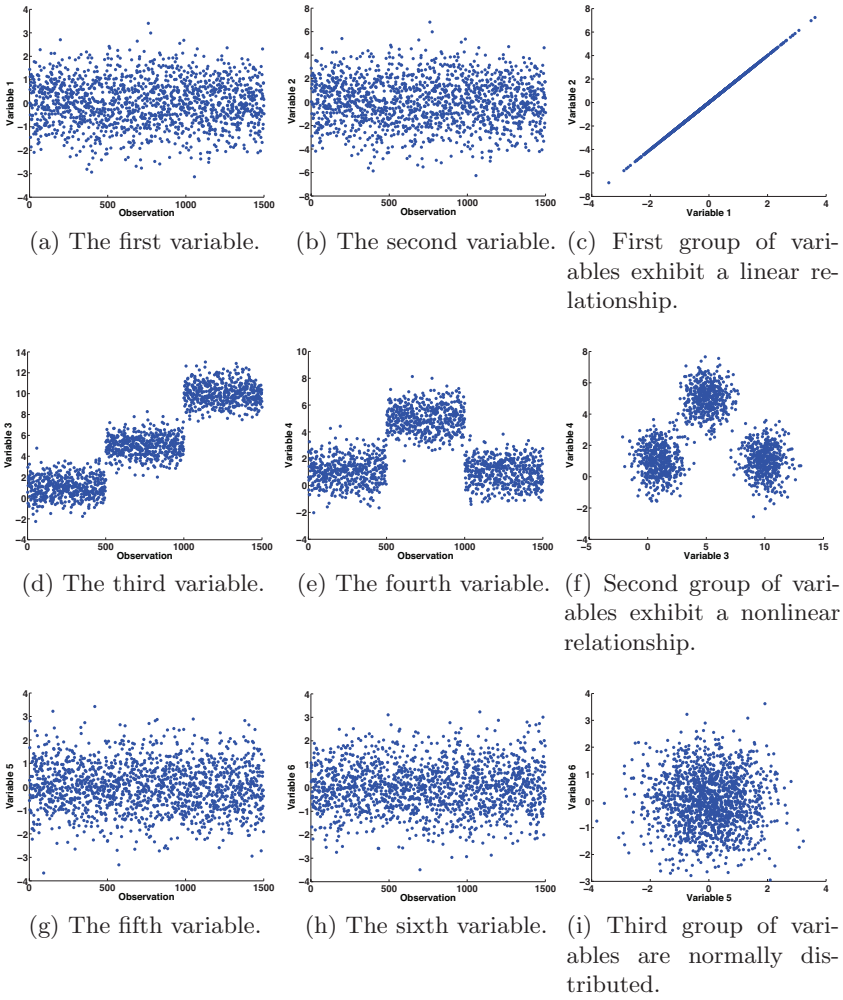
For illustration purpose, a "synthetic" data set is created. The data set contains six variables and 1500 observations. The first group of variables {1,2} exhibits a linear relationship and the second group of variables {3,4} exhibits a nonlinear relationship. The remaining group of variables {5,6} contains normally distributed pseudorandom numbers.

4.3.1 Variable selection

The high dimensional data gathered continuously from monitored component has created challenges to model such data in order to predict the RUL. The problem is that not all sensor data or variables hold valuable information about the health evolution of the system. Selecting only "interesting" variables will increase the RUL prediction efficiency while reducing the computational costs.

In this part, we propose and describe the proposed algorithm that explores a data set to look for non-random relationships among the input variables, and groups the variables without making any assumptions concerning the number of the variables in each group. This can be done in three main steps:

- Pair wise similarity measure using symmetrical uncertainty.



(a) The first variable. (b) The second variable. (c) First group of variables exhibit a linear relationship.

(d) The third variable. (e) The fourth variable. (f) Second group of variables exhibit a nonlinear relationship.

(g) The fifth variable. (h) The sixth variable. (i) Third group of variables are normally distributed.

Figure 4.16: The synthetic data set used to illustrate the proposed method.

- Grouping variables with hierarchical clustering.
- Assessing the quality of the variable groupings.

Computing the pair wise variable similarity

The pair wise similarity between two variables X and Y is measured with the symmetrical uncertainty (SU) [163], defined as:

$$SU(X,Y) = 2 \times \frac{I(X,Y)}{H(X) + H(Y)} \tag{4.6}$$

where $I(X,Y)$ is the mutual information

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{4.7}$$

and $H(X)$ is the Shannon entropy

$$H(X) = \sum_i P(x_i) log_b[P(x_i)] \tag{4.8}$$

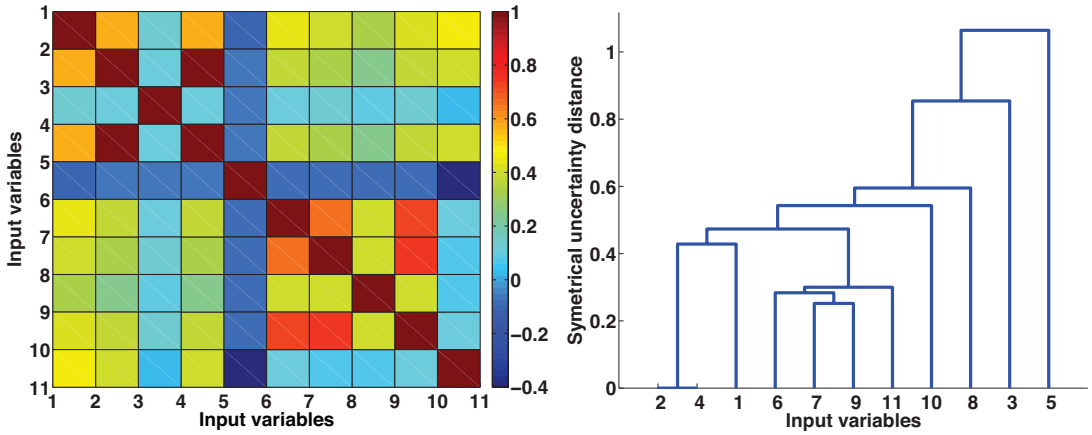
where b is the base of the logarithm used, $P(x_i)$ is the probability that $X = x_i$ and $H(X,Y)$ is the joint Shannon entropy of two variables X and Y. A variable Y is regarded more similar to variable X than to variable Z if $SU(X,Y) > SU(Z,Y)$. The SU measure treats two variables symmetrically and is normalized to the range $[0,1]$. The value 1 indicates that knowing one of the variables completely predicts the value of the other variable. On the other hand, the value 0 indicates independence between the two variables. The SU matrix is computed for all pairs of variables, using an adaptive binning method to estimate the entropies. It is then transformed into a distance matrix $D(i,j)$, where:

$$D(i,j) = 1 - SU(i,j) \tag{4.9}$$

Figure 4.17a shows an example of SU similarity measure for a data matrix contains eleven variables. The top row shows signal 1 plotted versus the remaining signals. Similarly, the second row from the top shows signal 2 plotted versus the rest of the signals. This continues in the same manner for the other rows. The diagonal shows SU measure plots for each variable and itself.

Grouping variables with hierarchical clustering

The distance metric is then used as input to an agglomerative hierarchical clustering (single linkage) algorithm [164]. The hierarchical clustering algorithm outputs a dendrogram, which is a tree representation of the relations between variables (Figure 4.17b). In order to get the actual clusters, an algorithm based on modified "L method", which



(a) SU similarity measure for a data matrix that contains 11 variables.

(b) Tree representation of variable relations.

Figure 4.17: Grouping input variables based on SU similarity measure.

was originally proposed by Salvador et al., is applied on the dendrogram [170]. The modification consists on calling the "L method" iteratively [157, 171]. At each iteration, the algorithm removes each detected knee and the points before it. The algorithm stops when it reaches the smallest distance in the dendrogram.

Assessing the quality of the variable groupings

The quality of the variable groupings is checked by doing a clustering on the observations (not on the variables) using a Self-Organizing Maps (SOM) [165]. The distortion should be low if there is a close relationship between the signals. The SOM is based on using K codebook vectors (cluster centers), w_k, with a defined topology among them. The parameters are fitted by minimizing the distortion measure:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} \wedge_k [x(n)] \|x(n) - w_k\|^2 \tag{4.10}$$

where N is the number of observations. The neighborhood function $\wedge_k$ defines the topology by assigning how much a data point $x(n)$ affects the codebook vector (cluster center) w_k. In the evaluation phase, after the clustering is done, the neighborhood function only assigns the closest codebook vector to a data point. The quality of variable groups is assessed in the following way. First, a SOM clustering is done on the suggested variable group and the resulting distortion is computed. Then, a new cluster is created which contains the same number of variables as the selected cluster. However, those generated variables are random, i.e. they are normally distributed. Similarly, the distortion of the new cluster is calculated. The ratio of these distortions is then used as a quality measure of the variable group:

$$Q = \frac{J_a}{J_b} \tag{4.11}$$

where J_a denotes the distortion value after SOM clustering on the selected variables and J_b denotes the distortion after SOM clustering on random variables that are equal in number to selected variables. If this ratio is much less than one then the clustered variables are much more closely related than the random variables and vice versa. A signal group is considered to be correctly identified if the algorithm finds it and also assigns to it a low Q value. The proposed algorithm worked excellently on the synthetic data set. It found three groups in the synthetic data set and assigned low Q values, 0.2545, for the first group and 0.4995 for the second, and high value, 0.9489, for the last group.

4.3.2 Feature extraction

Relationships change, between the selected variables, over time holds valuable information about the degradation process. The rate of the change can represent how close the monitored component is to the EOL time. Therefore, it is important to carefully choose informative features to be extracted from the resulting variables which can represent

such relation in a lower dimension. This can be done in two main steps, namely feature reduction and trend extraction.

Feature reduction

The dimension of the input data set consisting of the selected variables is reduced using PCA. The main idea of PCA is to reduce the dimension of a data set consisting of a large number of interrelated observations while retaining as much as possible of the variation in the input data set. This can be done by transforming the data to a new set of orthogonal variables. The new variables are ordered so that the first few contain the most of the variation present in all of the original input variables. Lets assume D_{nl} is a data set of the selected variables:

$$D_{nl} \subseteq D_{nm} \tag{4.12}$$

where, n is the number of the observations, l is the number of the selected variables and m is the total number of the variables before the selection. PCA starts by centering the mean of input data set by calculating the mean value of each variable:

$$\mu_l = \frac{1}{n}\sum_{i=1}^{n} D_{nl}, \forall_l \tag{4.13}$$

where μ_l is a vector of all variables' means. The centered data set, $D_{nl}^{centered}$, is calculated by subtracting μ_l from the input data set:

$$D_{nl}^{centered} = D_{nl} - \mu_l \tag{4.14}$$

so that

$$\mu_l^{centered} = \frac{1}{n}\sum_{i=1}^{n} D_{nl}^{centered} = 0, \forall_l \tag{4.15}$$

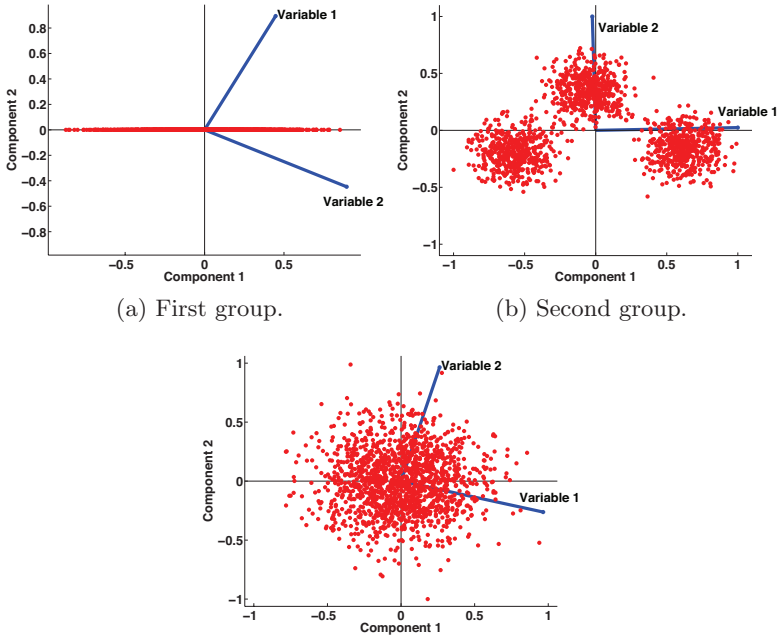
Then, pair wise covariance matrix C is computed for all pairs of variables:

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)} \tag{4.16}$$

where, X and $Y \in D_{nl}^{centered}$ and n is the number of the observations. Finally, the eigenvectors and eigenvalues are calculated for the C matrix:

$$Cv_l = \lambda_l v_l \tag{4.17}$$

where, λ_l are the eigenvalues, v_l are the eigenvectors for the covariance matrix C and l is the total number of resulting components which is equal to the number of the selected variables. The resulting components are ordered according to the variance that each component accommodates using the eignevalues. The resulting eignevalue for the first group is 100% for the first component and 0% for the second. For the second group, the greatest variance is in the direction of the first component and the eignevalue is



(a) First group.

(b) Second group.

(c) Third group.

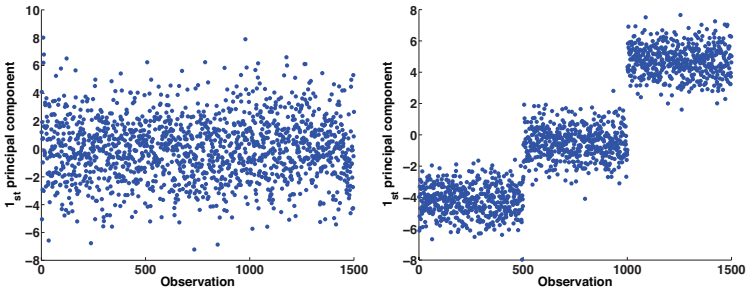
Figure 4.18: Biplot of the observations with respect to their components. The variables axes represent the original axes.

76.0020%. Finally, the third group's variances are 50.5469% and 49.4531% for the first and the second components respectively (Figure 4.18). The first component retains the maximum variance and therefore it will be used in the later steps (Figure 4.19).

As expected, the first component of the third group does not show any interesting information as the variables were randomly distributed (Figure 4.19c). The component extracted from the second group showed monotonic change (Figure 4.19b). The first group's component, depicted in Figure 4.19a, does not show any interesting features despite that the variables have linear relationship. The reason is that, the variance should be correlated with the time to be informative, i.e show progression over time. This is not the case with the first and of course the third groups. To illustrate that, figures 4.20a and 4.20c show that variables of the first and third groups are not correlated with the indexes/time and therefore their variances did not show any valuable information. On the other hand, figure 4.20b shows that variables of the second group are correlated with the time which leads to an informative first principal component.

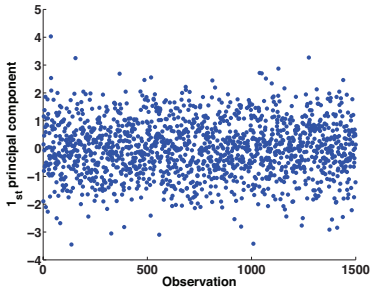
Trend extraction

The first principal component shows information about the change of the selected variables' relations over time and consequently EOL. The idea is to extract such information in a suitable form for HI construction. To do that, we propose to use the EMD algorithm. EMD is a method used to decompose a signal into a successive IMFs. The IMF



(a) First group.

(b) Second group.



(c) Third group.

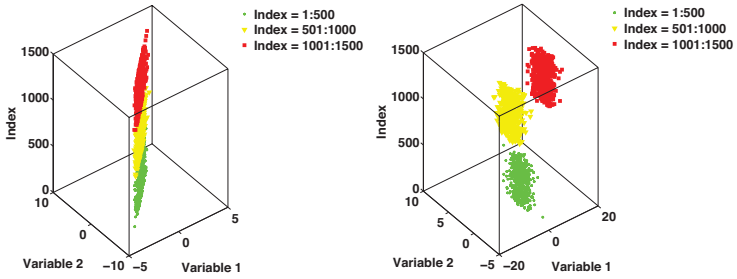
Figure 4.19: First components for the three groups of variables.

with the lowest frequency is the trend of the raw data. IMFs of higher orders reflect the dynamic characteristics of the data, such as depicted in Figure 4.21. A time series is considered as an IMF if it exhibits the following two properties:

- The number of local extrema of the time series and the number of it is zero-crossings must either be equal or differ by at most one.
- At any time, the mean value of the upper envelope determined by the local maxima and the lower envelope determined by the local minima is zero.

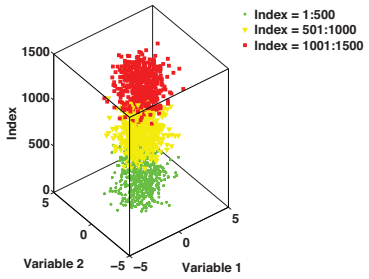
Given a non-stationary data series, $X(t)$, the EMD algorithm consists of the following steps [169]:

1. Find all the local maxima and minima of the input signal and compute the corresponding upper and lower envelopes using cubic spline respectively.
2. Subtract the mean value of the upper and lower envelopes from the original signal.
3. Repeat the previous steps until the signal remains nearly unchanged and obtain IMF_i.
4. Remove IMF_i from the signal and repeat the previous steps if the IMF_i is neither a constant nor a trend.



(a) First group resulting components.

(b) Second group resulting components.



(c) Third group resulting components.

Figure 4.20: Plot of the three groups of variables with respect to their indexes.

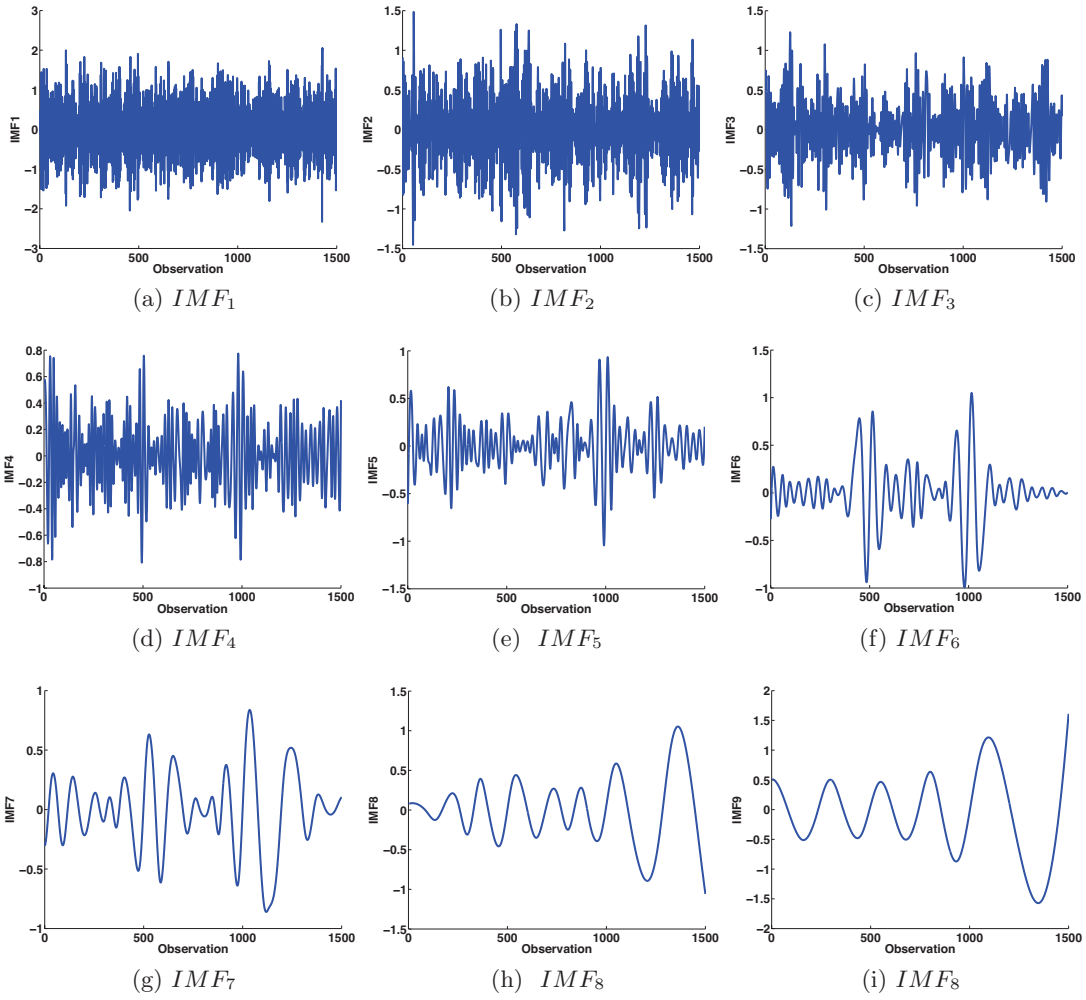
The remaining residual, should be constant or monotonic which can be represented as:

$$r_n(t) = X(t) - \sum_{i=1}^{n} imf_i(t) \tag{4.18}$$

where $X(t)$ is the original signal, i.e. first principal component of the selected signals, imf_i is the intrinsic mode function and $r(t)$ is the remaining residual. Figure 4.22 shows the resulting residuals for the three groups. The first and the last groups' residuals, shown in figure 4.22a and figure 4.22c respectively, are non monotonic which do not hold relevant information. Only the second group residual, depicted in figure 4.22b, is a monotonic signal.

The property of the generated residual can represent the relation between the generated trend and EOL time. To show how the generated residual can represent the relation between the generated trend and EOL time, an aging experiment is conducted on two bearings. In the first experiment, a degradation profile was applied on one of the bearings and the acceleration signals were acquired until the bearing completely worn out after 9 hours (Figure 4.23a). For the other bearing, no degradation profile was applied and also the sensor signals were recorded for 9 hours (Figure 4.23b).

EMD was applied to both of the two signals and the resulting residuals are shown in figure 4.24. The experiments show that the residual of the degraded component was a



(a) IMF_1 (b) IMF_2 (c) IMF_3

(d) IMF_4 (e) IMF_5 (f) IMF_6

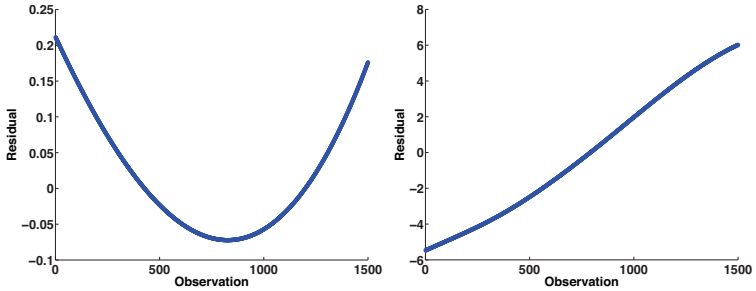
(g) IMF_7 (h) IMF_8 (i) IMF_8

Figure 4.21: Example of IMFs generated using EMD applied on the extracted component from the second group of variables.

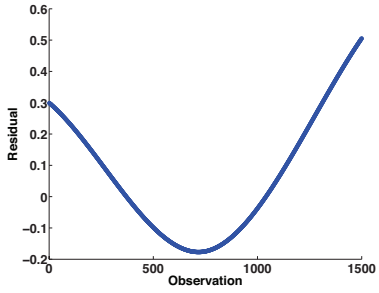
monotonic signal while the non degraded component generated almost a constant residual. Hence, the characteristic of the trend can represent the severity of the degradation. This property can be used to build HIs if appropriate measures are taken from the residuals. HIs can be used for visualizing the health status in a comprehensible form and can be used to build predictive models as shown in the next section.

4.3.3 Health indicators construction

Variations in the resulting EMD residuals can show evolution of the degradation over time. These variations can be described by using different statistical quantifiers to reduce the dimensionality and the computation and also to have a compact representation of

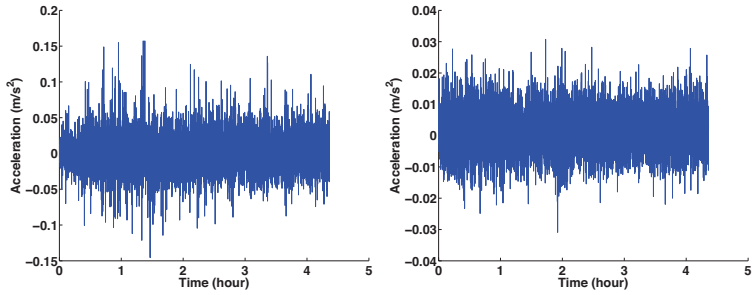


(a) First group resulting residual. (b) Second group resulting residual.



(c) Third group resulting residual.

Figure 4.22: Residual extracted from the three groups first components.



(a) Degraded bearing. (b) Non degraded bearing.

Figure 4.23: Vibration signals acquired from two bearings.

the degradation over time such as HIs. Many quantities can be calculated to represent the residuals, such as, mean, variance, multi-exponential function, curve fitting, discrete wavelet transform and discrete Fourier transform. However, selecting an appropriate approach is mainly problem specific. Recalling figure 4.24, the slope of the trend can be a discriminant characteristic of the trend. A trend with a higher RUL value tends to have smaller slope and vice versa. The intercept value can also change with each new observation and that could also be interesting to measure. Another suitable discriminant

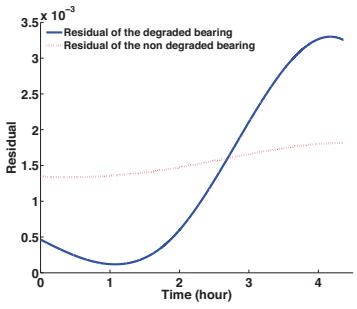


Figure 4.24: Differences in the resulting residual's slope according to the health status.

feature for this problem can be the arithmetic mean of the extracted trend. Every data value in the trend contributes to the mean value, and changing one of them will change the mean. Similarly, the variance of the extracted trend is a discriminant feature. It describes the spread of a trend and therefore can show a correlation with EOL. To measure these quantities, a first degree polynomial curve was fitted to each trend

$$Y = aX + b \tag{4.19}$$

where a is the slope, b is the y-intersect, Y is the residual value and X is the observation index vector. Both a and b are used as HIs values at a particular time (Figure 4.25a). Also, the mean and the variance were calculated for each trend, as shown in Eq. (4.20) and Eq. (4.21) respectively.

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \tag{4.20}$$

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2 \tag{4.21}$$

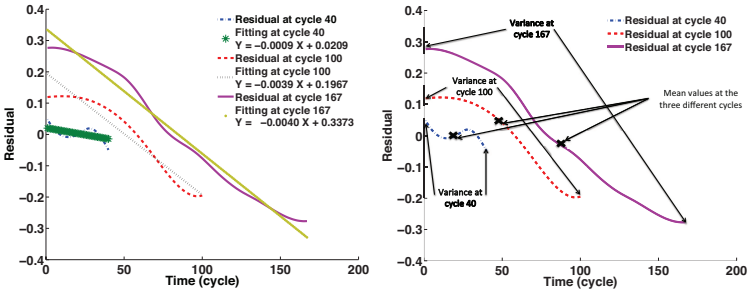
where x_j is the input observation and n is the length of the trend (Figure 4.25b). The set of HI values at a particular time can be described as

$$HI_t = [a, b, \bar{x}, s^2] \tag{4.22}$$

At each time/cycle all observations, starting from time t_0 till $t_{current}$, are processed to extract the proposed features. This process is repeated recursively till the end of the given training data set to construct four different HIs. The resulting HIs can be then defined as a matrix $HI_{n\times 5}$

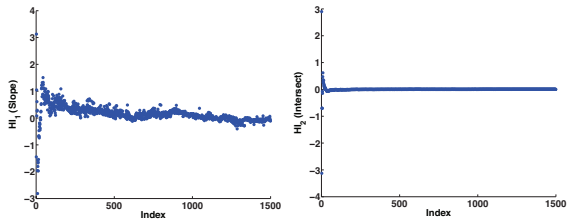
$$HI(t) = [a_t, b_t, \bar{x}_t, s_t^2, t] \tag{4.23}$$

where, t is the current time or cycle. Figure 4.26 shows plots of the HIs generated from the first group.

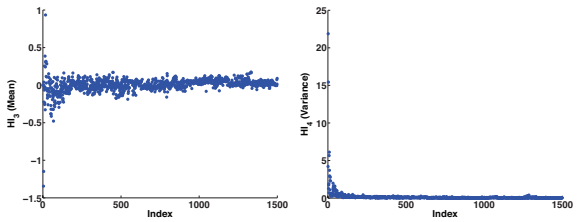


(a) HI using slope and intersect. (b) HI using mean and variance.

Figure 4.25: Constructing HIs at cycles 40, 100 and 167.



(a) Slope health indicator. (b) Intersect health indicator.



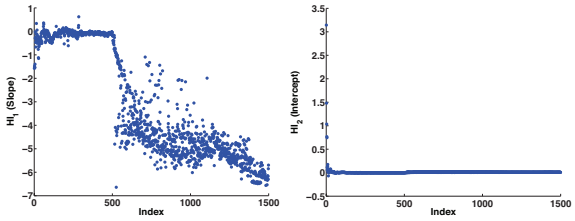
(c) Mean health indicator. (d) Variance health indicator.

Figure 4.26: The four HIs constructed for the first group of trends.

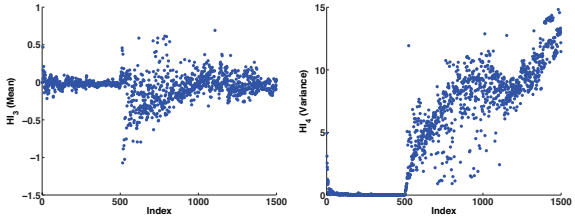
All HIs are not monotonic, which means that the selected variables do not show any variance over time. Figure 4.27 shows the resulting HIs for the second group.

It can be seen that only the slope and variance HIs show interesting behavior. The two HIs were constant in the first 500 cycles which is correct as the original variables are not changing. After 500 cycles, the two HIs show monotonic change, increasing for the slope and decreasing for the variance. Figure 4.28 shows plots of the third group, and similar to the first group, the HIs are not showing any monotonic change and that is because the variables are random.

The processes of feature extraction and construction are repeated on the selected variables for several times using data from similar components to create learning data D_l

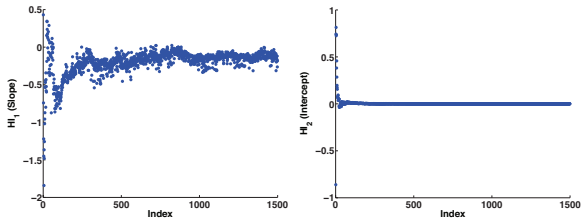


(a) Slope health indicator. (b) Intersect health indicator.

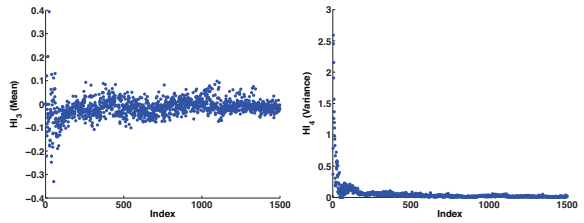


(c) Mean health indicator. (d) Variance health indicator.

Figure 4.27: The four HIs constructed for the second group of trends.



(a) Slope health indicator. (b) Intersect health indicator.



(c) Mean health indicator. (d) Variance health indicator.

Figure 4.28: The four HIs constructed for the third group of trends.

sets. The learning data is used to build reference model(s) of the monitored component. The overall HIs construction algorithm is summarized in Algorithm 4.1.

Algorithm 4.1: Health indicators construction

Data: $trainingData$
Result: D_l

for $\forall trainingData$ **do**
  $selectedVariables = FindBestGroup(D_{nm})$;
for $i = 1 : numberOf(trainingData)$ **do**
  $EOL = lengthOf(trainingData(i))$;
  for $j = 2 : n$ **do**
    $ip = selectedVariables(1 : j)$;
    $variance = GetFirstComponent(ip)$;
    $residual = GetEMDResidual(variance)$;
    $HI =$
    $append([HI_1(residual), HI_2(residual), HI_3(residual), HI_4(residual), i])$;
  $D_l = append([HI, EOL])$;

4.4 Conclusion

Probably, one of the most crucial problems in data-driven prognostics approaches research is the signal processing. The data acquired from monitored components is highly influenced by different sources of noise which can obscure the degradation evolution behavior. The data acquired is also multidimensional, which introduces another challenge such as the difficulty of selecting appropriate sensors to process and extract relevant information. The quality of data driven prediction algorithm depends on the information extracted from raw signals.

The goal of this chapter was to develop a method that can extract features representing the behavior of the monitored component and from these features extract smooth trends to represent the critical component's health evolution over time. The proposed method can be seen as a tool to build non parametric models from historical data. It did not build on many assumptions about the monitored component or the acquired signals. Instead, the method was used to let the data "speak" for itself.

The problem of selecting appropriate sensors data is tackled in this chapter. It has been shown that the method managed to select informative sensors and to neglect the sensors that contain no valuable information in unsupervised way. From the selected sensors, the method extracts sequential information from all historical data until the EOL time. These features are used to construct four different HIs. The resulting HIs showed good representation for the synthetic data evolution over time.

One of the remaining problems is how to use the generated offline models to predict the RUL of the new test data at a particular time. Another problem, given all the many sources of noise, is how to represent the uncertainty about the prediction result. These question will be answered in the next chapter.

5

Health assessment and remaining useful life estimation

"The farther backward you can look, the farther forward you are likely to see."
– Winston Churchill

5.1 Introduction

The constructed HIs from the training data sets represent different degradation behaviors for the monitored component over time. These HIs are saved in the offline data base as reference models. Furthermore, the behavior of a new component is expected to resemble the behavior of the most similar offline model. Thus, the problem of RUL estimation, for a new component, becomes a problem of finding the most similar offline model. Doing that can be very challenging due to the uncertainty that affects the constructed HIs such as the variability of the End Of Life (EOL) value for similar components working under same operating conditions (Figure 5.1). Other uncertainty sources can be unmodeled phenomenon, manufacturing variability, sensor noise, model approximations and unforeseen future environment changes.

In this chapter, we present a novel failure prognostic method for health assessment and RUL estimation of critical components. The method uses different machine learning algorithms, such as k-NN and Gaussian process regression to map the relation between sensor data and their EOL values. The method deduces the health status using discrete Bayesian filter applied on the online HI. Finally, the results of applying the proposed method on different real life applications, namely bearings, Lithium-ion batteries and turbofan are presented. The assumptions taken into account in this work can be summarized as follows:

1. The data acquired from the test component is taken at similar time-stamps to the offline signal.
2. The offline data set contains enough data to represent different degradation behaviors.

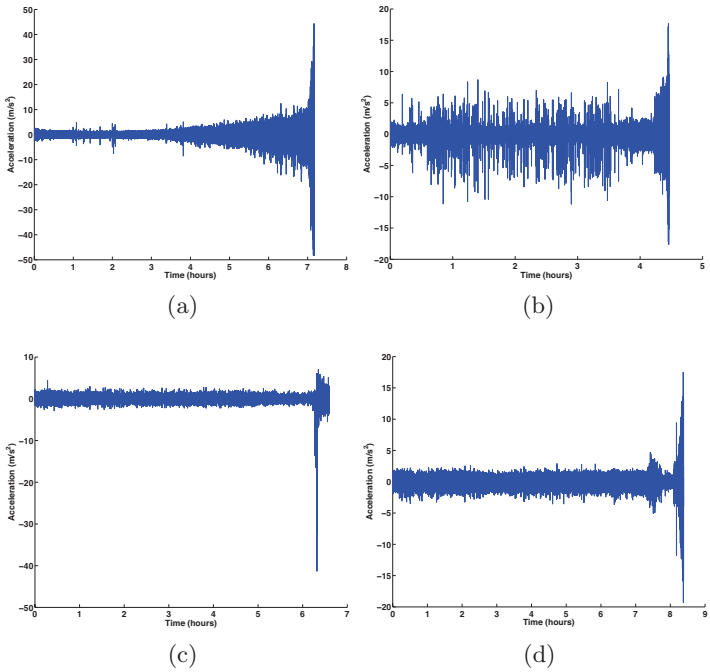


Figure 5.1: Example of raw vibration signals extracted from bearings.

5.2 The proposed method

In this section the online method for health assessment and RUL prediction is presented. The main goals of the proposed method are:

1. Performing health assessment for monitored component. This can be done first by estimating the health indicators values recursively from the monitored component sensor data using Bayesian filter. The Bayesian filter represents the uncertainty about the health status in a probabilistic form. This can be useful for decision making in later steps.

2. Predicting the RUL by finding the most similar offline model in the training database using two different approaches:

   - Classification: is the process of assigning a class for an input pattern based on a training set of data containing patterns whose class membership is known.
   - Regression: is the process of predicting a real value associated with an input pattern based on estimates of the relationship between training set of patterns whose associated values are known.

The sensor data of the monitored component is known as testing data set D_{tl}. D_{tl} can be defined as a matrix of the size $t \times l$, where t is the number of samples acquired from the test component until the current time and l is the number of sensor data selected by

the selection algorithm in the offline phase. This data set is acquired from a component that is not used to build the offline data sets. Similarly to the offline signals, four HIs are constructed from the test sensor signals (Figure 5.2). Then the method starts performing

Online HIs at time "t"

HI_{11}	HI_{21}	HI_{31}	HI_{41}	1
HI_{12}	HI_{22}	HI_{32}	HI_{42}	2
		:		
HI_{1t}	HI_{2t}	HI_{3t}	HI_{4t}	t

Figure 5.2: Constructed HIs from online data.

health assessment by estimating recursively the values of HIs given past history of the testing data set. The method looks in the offline data base for the most similar offline $HI(t)$ at the same time/cycle. The assumption is that, the future behavior of the new monitored component will be similar to the most similar offline model. Therefore, the EOL value of the new component equals to the EOL value of the selected offline model, $EOL_{selected}$. Finally, the RUL of the monitored component, RUL_t, can be calculated as follows:

$$RUL_t = EOL_{selected} - t, \forall EOL_{selected} > t \tag{5.1}$$

where t is the current time. The final output of the proposed method is the prediction of the RUL and estimates of the HIs at the predicted RUL (Figure 5.3).

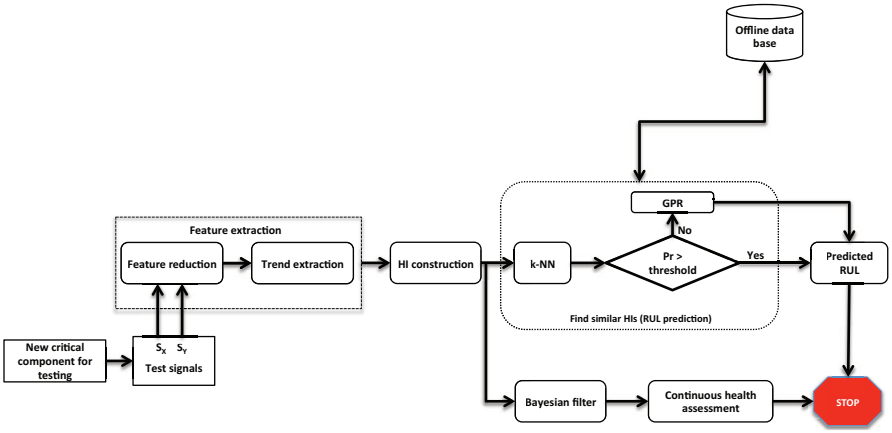


Figure 5.3: General scheme of the proposed method.

5.2.1 Health assessment

One way to do that is by applying recursive estimating algorithms. Such algorithms estimate the HIs from the online data until it reaches stopping criteria (Figure 5.4). For

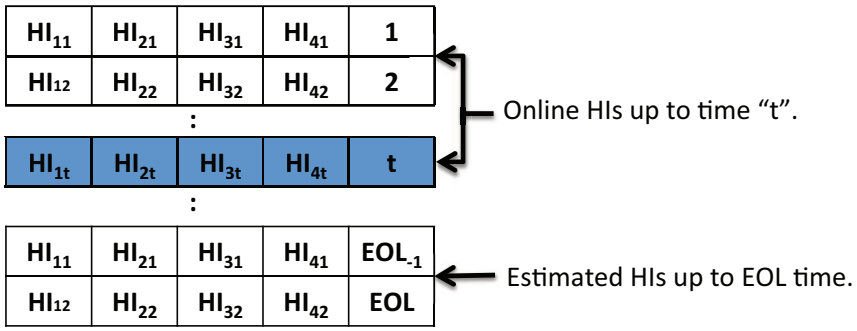


Figure 5.4: Recursive estimation of the online HIs up to the EOL time.

that purpose, a recursive Bayesian filter has been applied to the online trends [177]. The unobserved signal (hidden states) will be denoted as "x_t" at time "t". Assume also that x_t is complete, i.e., the knowledge of past states or measurements carry no additional information that would help to predict the future (Markov assumption). The measurements data acquired from the monitored component at time t will be denoted as "z_t", where:

$$z_{t_1:t_2} = z_{t_1}, z_{t_1+1}, z_{t_1+2}, ...z_{t_2} \tag{5.2}$$

In order to represent the uncertainty about the online trends in a probabilistic form, the system should be described by a dynamic stochastic model. For this purpose, two main quantities need to be observed: state transition and measurement transition probabilities. State transition probability specifies how the unobserved signal state evolves over the time and is defined as follows:

$$p(x_t|x_{0:t-1}, z_{1:t-1}) \tag{5.3}$$

where x_t is the current unobserved signal or hidden state, $x_{0:t-1}$ is the previous states and $z_{1:t-1}$ represents the sensory measurements. Using the complete state assumption, Eq. (5.3) can be expressed as follows:

$$p(x_t|x_{0:t-1}, z_{1:t-1}) = p(x_t|x_{t-1}) \tag{5.4}$$

The measurement probability specifies how a measurement "z" is generated from a hidden state "x", and is defined as follows:

$$p(z_t|x_{0:t}, z_{1:t-1}) \tag{5.5}$$

Given that the unobserved signal x is complete, Eq. (5.5) can be defined as:

$$p(z_t|x_{0:t}, z_{1:t-1}) = p(z_t|x_t) \tag{5.6}$$

The system's uncertainty over a state should be distinguished from the true state. To do so, uncertainty over state variable x_t will be denoted by $uncert(x_t)$, which is an abbreviation of posterior probability, as shown in the following equation:

$$uncert(x_t) = p(x_t|z_{1:t}) \tag{5.7}$$

Another useful probability distribution, which is calculated before incorporating the current measurement z_t to Eq. (5.7), is often referred to as prediction probability distribution and denoted as:

$$\overline{uncert}(x_t) = p(x_t|z_{1:t-1}) \tag{5.8}$$

A Bayesian filter is a recursive algorithm, that is, the $uncert(x_t)$ at time t is calculated from $uncert(x_{t-1})$ at time $(t-1)$. The input is the uncertainty at time $(t-1)$ along with the most recent measurement z_t. Algorithm 5.1 depicts the general formula for the Bayesian filter.

Algorithm 5.1: The general algorithm for Bayesian filter.

Input : $uncert(x_{t-1})$, z_t
Output: $uncert(x_t)$

forall the x_t **do**
  $\overline{uncert}(x_t) = \int p(x_t|x_{t-1}) uncert(x_{t-1})\, \mathrm{d}x$
  $uncert(x_t) = \eta p(z_t|x_t)\overline{uncert}(x_t)$

where $\eta = p(z_t|z_{1:t-1})$. Bayes filters are implemented in several different ways such as Kalman, extended Kalman, particle and histogram filters. Any implementation requires knowing three probability distributions: initial probability $p(x_0)$, measurement probability $p(z_t|x_t)$ and state transition probability $p(x_t|x_{t-1})$. Each technique depends on different assumptions regarding initial, measurement and state transition uncertainties. In this work, the histogram implementation for the Bayesian filter is proposed. The histogram filter decomposes the state space into finitely many regions and represents the cumulative posterior for each region by probability values. When applied to discrete spaces, such filters are called discrete Bayesian filters and when applied to continuous spaces, they are called histogram filters. In this work it is assumed that the space domain is discrete. Therefore, discrete Bayesian filter will be discussed in this section (see Algorithm 5.2).

Algorithm 5.2: Discrete Bayesian filter.

Input : $\{p_{k,t-1}\}$, z_t
Output: $\{p_{k,t}\}$

forall the k **do**
  $\bar{p}_{k,t} = \sum_i p(X_t = x_k|X_{t-1} = x_i) p_{i,t-1}$
  $p_{k,t} = \eta p(z_t|X_t = x_k)\bar{p}_{k,t}$

A discrete Bayesian filter is derived from the general Bayesian filter by replacing the integration with a finite sum. The input to the algorithm is a discrete probability distribution $\{p_{k,t}\}$ along with the recent measurement z_t. The first line of the Algorithm 5.2, $\bar{p}_{k,t} = \sum_i p(X_t = x_k|X_{t-1} = x_i) p_{i,t-1}$, calculates the prediction for the new state based on previous state uncertainty and state transition model. The prediction

is then updated in the second line, $p_{k,t} = \eta p(z_t|X_t = x_k)\bar{p}_{k,t}$, so as to incorporate the measurement.

Discrete Bayesian filters apply to problems with finite state space, where the random variable X_t can take finitely many values:

$$X_t = x_{1,t} \cup x_{2,t} \cup ...x_{k,t} \tag{5.9}$$

A straightforward decomposition of X_t is a multidimensional grid, where each $x_{k,t}$ is a bin or region. The size of each bin, dx, can be calculated as follows:

$$dx = (x_{max} - x_{min})/n; \tag{5.10}$$

where x_{max} is the maximum state value, x_{min} is the minimum state value and n is the number of bins. Each bin can then be represented as a Gaussian function with a mean value at each state and a common variance:

$$p(X_t|X_{t-1}) = \|dx \times \mathcal{N}(X_{k,t}, \sigma^2)\| \tag{5.11}$$

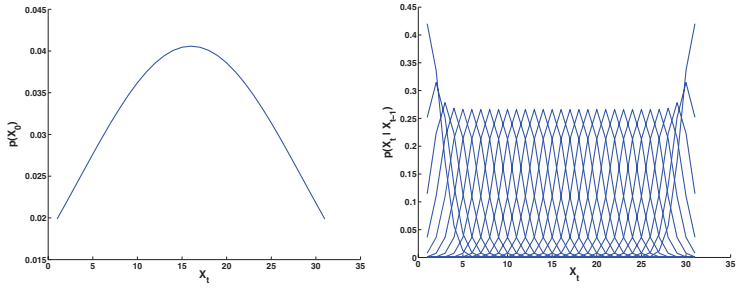
where $p(X_t = x_k|X_{t-1})$ is the state transition model, dx is the bin size and $\mathcal{N}(X_{k,t}, \sigma^2)$ is the normal distribution at state $X_{k,t}$. Moreover, equation (5.11) is normalised to turn this quantity into a probability distribution. Similarly, the measurement probability model can be calculated in the same manner as the transition model. Figure 5.5 depicts an example of different probability distributions for one estimation step. The prior discrete probability distribution is depicted in figure 5.5a. The state transition model, or measurement model, decomposed in n bins that vary between $[-1, 1]$, is shown in figure 5.5b. Finally the predicted and corrected probabilities are shown in figure 5.5c and figure 5.5d, respectively.

5.2.2 Remaining useful life estimation

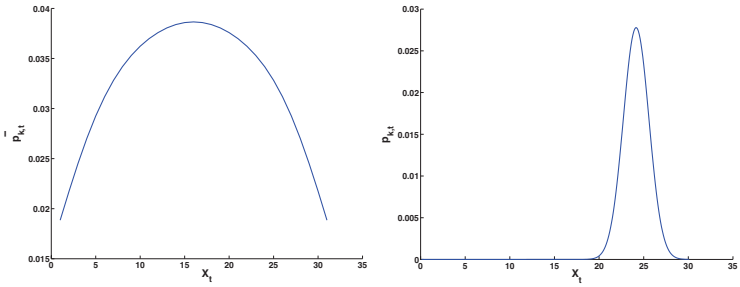
The second task of the online algorithm is to find the most similar offline model in the data base. EOL value of the selected offline model can be used as EOL of the new HIs. Then the RUL can be calculated from the EOL value and a probability value should be assigned to the predicted RUL. In this section, two different machine learning algorithms are used to predict the RUL given the offline data. The choice of a particular machine learning algorithm for selecting the most similar offline group of HIs from the offline data set is based on the ability of the algorithm to represent the decisions in probabilistic forms. One way to use these two machine learning models together is by constructing an ensemble model [172, 173]. Another way is to select the outcome of one model based on specific parameter. The former approach is performed in this work.

Classification

In order to select the most similar offline HI models, a k-NN classifier is applied in this work. k-NN is a non parametric learning algorithm which does not make any assumptions on the underlying data distribution [174]. A k-NN algorithm is considered one of



(a) Initial probability distribution (b) Transition or measurement model



(c) Predicted probability distribution (d) Posterior probability distribution

Figure 5.5: Example of different probability distributions.

the simplest of all machine learning algorithms and therefore it can be used for online applications. The final classification decision is based on largest posterior probability of the tested sample. Therefore, a probability value is assigned to the prediction output (Figure 5.6). Let each EOL value be represented by a particular class label C_k where

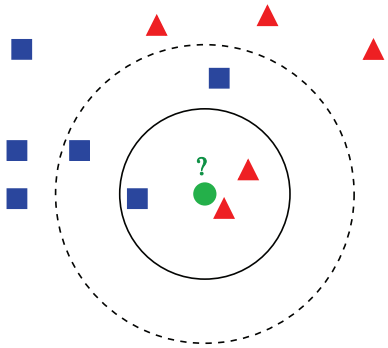


Figure 5.6: k-NN classifier.

two historical data that end at the same time belong to the same class and get the same label. Each group of HIs values, extracted from the test data at time "t", are considered

a four dimensional point defined as "α". Assume N_k points in class C_k with N points in total, where:

$$\sum_k N_k = N \tag{5.12}$$

In order to assign a new point α to a class C_k, a sphere centered on α and containing K points is drawn. Suppose this sphere has V volume and contains K_k points from class C_k. Then, the unconditional density associated with "α" is defined as:

$$p(\alpha) = \frac{K}{N \times V} \tag{5.13}$$

Similarly, estimate of the density associated with each class is given by:

$$p(\alpha|C_k) = \frac{K_k}{N_k \times V} \tag{5.14}$$

and the class priors are given by:

$$p(C_k) = \frac{N_k}{N} \tag{5.15}$$

Combining equations (5.12), (5.13) and (5.14) and using Bayes theorem, one can obtain the posterior probability of a class membership or the EOL value:

$$p(C_k|\alpha) = \frac{p(\alpha|C_k) \times p(C_k)}{p(\alpha)} = \frac{K_k}{K} \tag{5.16}$$

So, to minimize the miss classification error, the point α should be assigned to the class with largest posterior probability. The same procedure applies to classifying a new point α. Figure 5.7 depicts an example of how k-NN is utilized in the proposed method at time t. After constructing the HIs from the input trend signal, the HIs values are passed to the k-NN to classify the input and assign it to one of the offline classes. The HIs are compared to the offline data with the same time stamp t.

One drawback of the k-NN is that the RUL prediction depends on the offline data base. If the new online data is not similar to any of the offline models, k-NN will choose the closest offline model which leads to a huge prediction error. The problem is best explained by an example depicted in figure 5.8. Assume a new HI extracted from critical component and its RUL to be predicted at time t. The real EOL value for this component is 100 cycles. In the offline data base there is only offline data models and their EOL values are 148 and 55 cycles. Based on the k-NN, the RUL prediction output should be either too late or too early. It is then important to build a model of the offline data so that if the new HIs have no close offline model, the new model can predict its RUL.

Regression

Another way to predict the RUL is by using regression models [175]. The relation between the EOL values and the different offline HIs can be mapped using regression

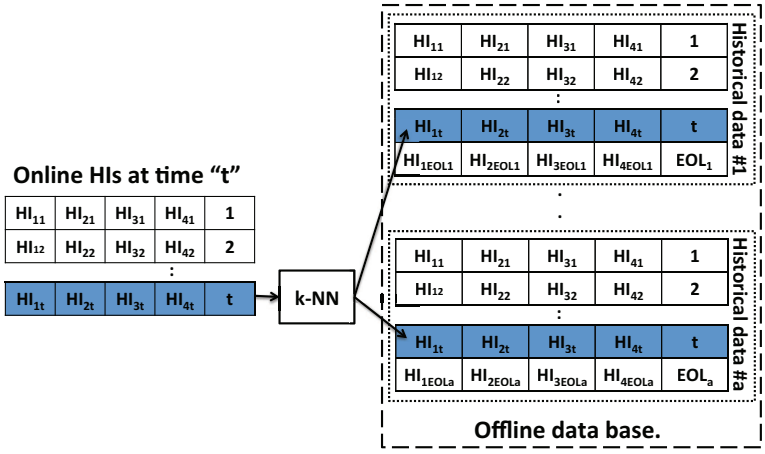


Figure 5.7: Finding the most similar offline HIs to the test HIs. The algorithm assigns the EOL of the selected HIs to the test HIs to calculate the RUL.

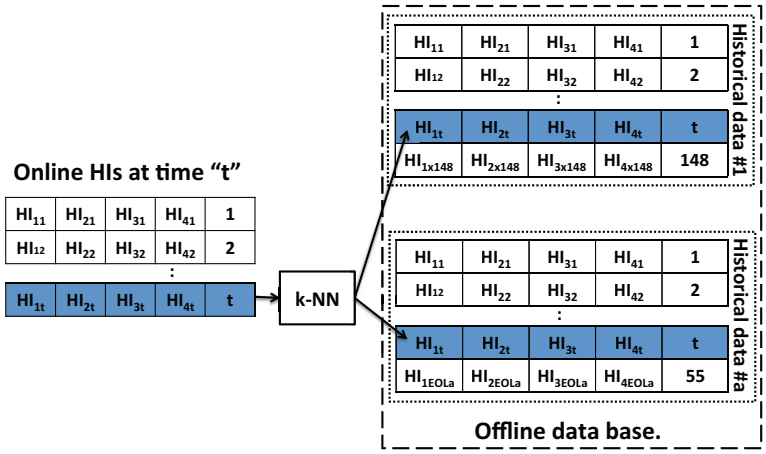


Figure 5.8: Classification can lead to a huge RUL estimation error.

models. In this way, if the new online HIs are never seen in the offline model, the error of the RUL prediction is decreased. It is also important to employ a regression algorithm that can predict the RUL and its associated confidence in a probabilistic form. For these reasons, Gaussian Process Regression (GPR) is applied. GPR is a flexible, powerful and probabilistic approach for Bayesian inference over functions [176]. Assume the offline HIs "x" to be the input at particular time "t" and the corresponding EOL values would be the output "y". The idea is to build a model to map the relation between the input and the output which is used to predict the value of the EOL for the new online HIs (Figure 5.9). In order to map from an input x to output $y = f(x)$, GPR defines the prior for output $f(x)$ in the form of distribution over functions specified by Gaussian

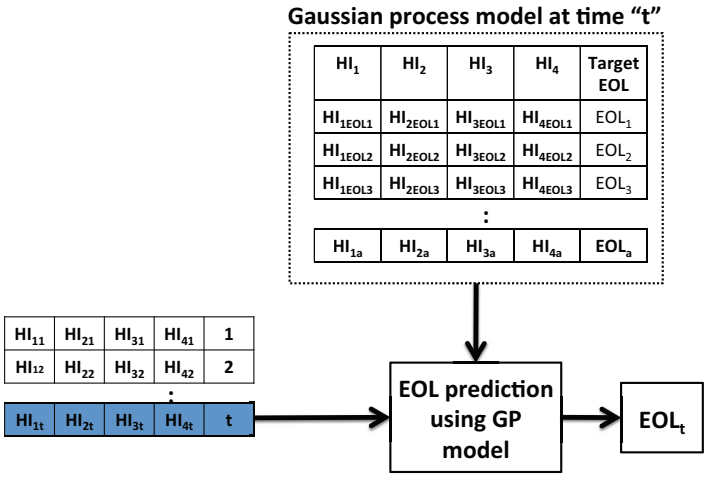


Figure 5.9: Mapping the relation between the input offline HIs and the output corresponding EOL values to predict the value of the EOL for the new data.

process (GP):

$$y = f(x) + \mathcal{N}(0, \sigma_n^2) \tag{5.17}$$

where, y is the dependent variable, x is the independent variable, $f(x)$ is the underlying GP and $\mathcal{N}(0, \sigma_n^2)$ is the Gaussian noise with zero mean and σ_n^2 variance. GP extends multivariate Gaussian distribution to infinite dimensions. Formally, it's a collection of random variables which have a joint multivariate Gaussian distribution. GP function $f(x)$ is specified by a mean function $m(x)$ and covariance function $k(x, x')$ collected for all possible pairs of the input vector x:

$$f(x) = \mathcal{GP}(m(x), k(x, x')) \tag{5.18}$$

where

$$m(x) = E[f(x)] \tag{5.19}$$

and

$$k(x, x')) = E[(f(x) - m(x))(f(x') - m(x'))] \tag{5.20}$$

The value of the covariance function expresses the relation between observations. A popular choice of the covariance function is the squared exponential function:

$$k(x, x')) = \sigma_f^2 \exp(\frac{-(x - x')^2}{2l^2}) + \sigma_n^2 \delta(x, x') \tag{5.21}$$

where, x and x' are two observations, l is a scaling parameter, σ_f^2 is the maximum allowable variance, σ_n^2 is the noise variance and $\delta(x, x')$ is the Kronecker delta function which is a function of two variables. The Kronecker function is 1 if the variables are equal, and 0 otherwise. Given a new observation y_*, the goal is to compute the posterior probability $p(y_*|y)$, i.e., the probability of the test data y_* given the training data y.

Since the assumption in GP is that data can be represented as samples from multivariate Gaussian distribution, the posterior probability distribution can be written as:

$$\begin{bmatrix} y \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix} \right)$$

where, $\mu = m(x)$ is the mean for training set, $\mu_* = m(x_*)$ is the mean for test set, K is the covariance for training set, K_* is the covariance for training-test, K_{**} is the covariance for test set and T is the matrix transpose. The posterior probability for y_* follows a Gaussian distribution:

$$y_*|y \sim \mathcal{N}(\mu_* + K_* K^{-1}(y - \mu), K_{**} - K_* K^{-1} K_*^T) \tag{5.22}$$

where, the best estimate for y_* is the mean of this distribution.

$$\overline{y}_* = \mu_* + K_* K^{-1}(y - \mu) \tag{5.23}$$

and the uncertainty in the estimate is represented in the variance.

$$var(y_*) = K_{**} - K_* K^{-1} K_*^T \tag{5.24}$$

To summarize, in the online phase the method does not apply variable selection on the input data. Instead, it only processes the same variables selected on the offline phase. Furthermore, the method performs feature reduction and trend extraction on the selected variables. Then, the method constructs four HIs from the extracted trends that represent the current health status. The method estimates the health status recursively using Bayesian filter. To estimate the RUL, the method looks for the most similar offline model using k-NN. If the posterior probability is less than a certain threshold, the GPR model is then used. The estimated value of the RUL can be used as a stopping criteria for the recursive Bayesian filter (Figure 5.10). Figure 5.11 shows the results of estimating the second group of four HIs at cycle number 1000. The estimation process gets as an input the prior probability distribution and the current observation. The estimator recursively estimates one step ahead the value of each HI. The stopping critiera for the estimation algorithm is the predicted EOL value using k-NN or GPR models. The

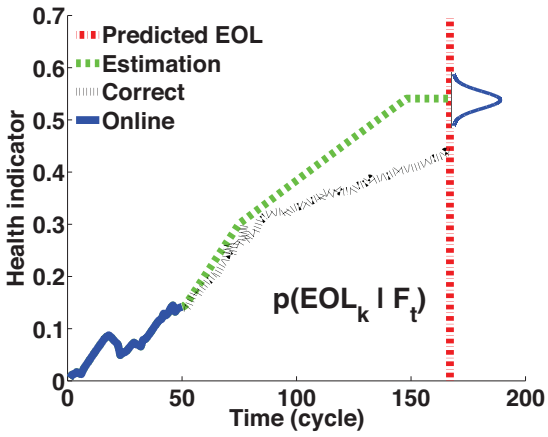
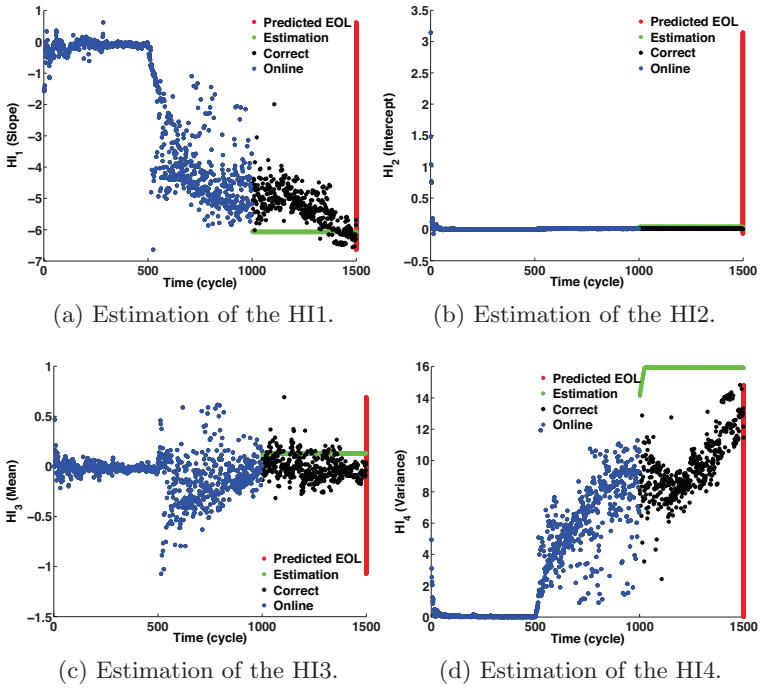


Figure 5.10: Final result of the online process for one HI.



(a) Estimation of the HI1.

(b) Estimation of the HI2.

(c) Estimation of the HI3.

(d) Estimation of the HI4.

Figure 5.11: Example of estimating the HIs at time 1000.

overall online algorithm is summarized in Algorithm 5.3.

Algorithm 5.3: The general algorithm of the online phase

Data: $\{trainingData, testData\}$
Result: $\{D_l, RUL\}$

Online phase
for $\forall testData$ **do**
    $selectedVariables = getSelectedVariables(testData)$
    $firstComponent = GetFirstComponent(selectedVariables)$;
    $residual = GetEMDResidual(firstComponent)$
    $testingFeatures = GetFeatures(residual)$;
    $EOL = kNN(testFeatures, D_l)$;
    $rulEstimation = discreteBayesianFilter(testFeatures, EOL)$;
    $RUL = (EOL, rulEstimation)$;

5.3 Applications and results

To verify the proposed method, three critical components that can be found in any avionic system are used in this work, namely bearings, batteries and turbofans. For the first component two data sets, PRONOSTIA and NASA bearings, are used only to verify the trend extraction phase. Then NASA batteries and NASA turbofans are used to verify the whole method. First, the trends have been extracted from the raw data then a selected trend has been modeled for RUL estimation. The details of the data-sets and the experiments are explained hereafter.

5.3.1 Bearings

The trend extraction approach has been verified using two different bearing data-sets, namely PRONOSTIA and NASA. The details of the data-sets and the experiments are explained in the following subsections.

PRONOSTIA

PRONOSTIA is an experimentation platform dedicated to test and validate bearings fault detection, diagnostic and prognostic approaches (Figure 5.12). It consists of four main parts: a rotating part, a degradation generation part (with a radial force applied on the tested bearings), a measurement part and test bearings. In this work, three data sets, acquired from this platform, have been used to validate the algorithm, where only the acceleration data have been used. For the first data set, the algorithm generates nine

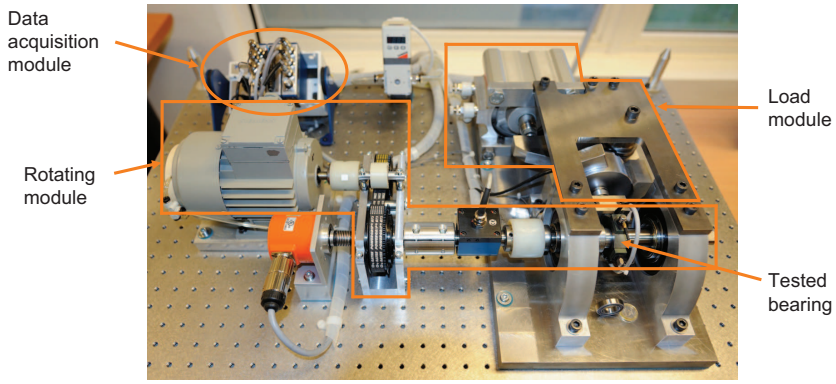
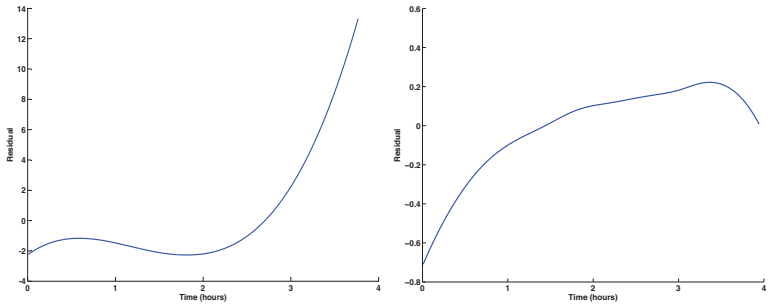


Figure 5.12: PRONOSTIA experimentation platform.

signals, each of that represents evolution of the bearing degradation over the time with different trajectories. Figure 5.13 shows two selected monotonic health indicators for the first experiment. As can be seen from the figure, the trends show a smooth increasing signals over the time, which can be used to deduce health status of the machinery. The plot on the left side is the result of fusing two auto-regression (AR) parameters for the

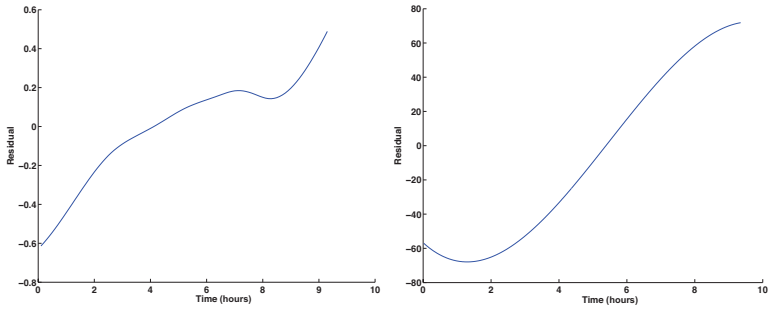
first sensor and one parameter for the second sensor. The plot on the right side shows another indicator which was a result of fusing also three features namely, skewness for the first sensor along with the second parameter of AR model and line integral feature extracted from the second sensor signal.



(a) (b)

Figure 5.13: PRONOSTIA first monotonic data set.

For the second data set, the algorithm generates 10 signals, each of that indicates the health degradation over the time. Figure 5.14 shows two selected monotonic health indicators for the first experiment. The plot on the left side is the result of fusing two features, maximum peak value and AR parameter, both have been extracted from the first sensor. The plot on the right side shows a smooth monotonic function. It is a result of fusing also two features, RMS and line integral acquired from the second sensor.

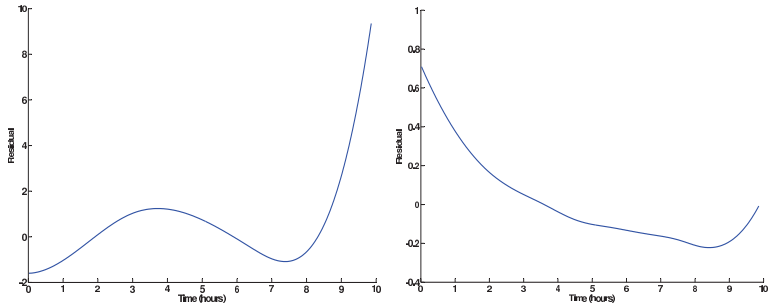


(a) (b)

Figure 5.14: PRONOSTIA second monotonic data set.

For the third data set, the algorithm generates 9 signals that indicate the health degradation over the time. Figure 5.15 shows two selected monotonic health indicators for the aforementioned experiment. The plot on the left side is the result of fusing three features, peak-to-peak, maximum peak value and AR parameter where all the features have been extracted from the first sensor. The plot on the right side shows a smooth decreasing monotonic function. It is a result of fusing three features, arithmetic mean

of Power Spectral Density (PSD) extracted from both sensors and kurtosis which has been extracted from the second sensor.

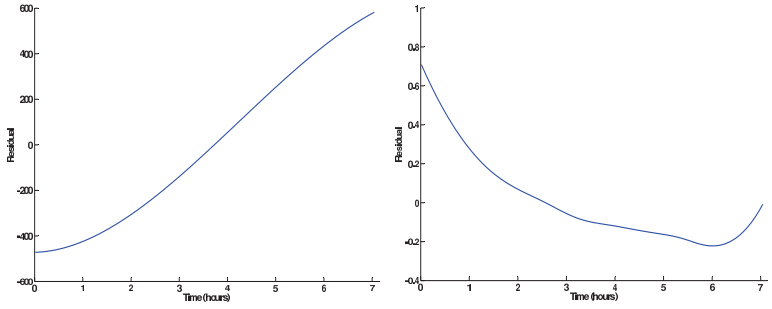


(a) (b)

Figure 5.15: PRONOSTIA third monotonic data set.

NASA data set

This data set is fully described in NASA's website (ti.arc.nasa.gov/tech/dash/pcoe/). In this work only the first two sensors readings have been used to validate the algorithm. For the first data set, the algorithm generates 10 trends from the raw data. Figure 5.16 shows two selected monotonic health indicators for the aforementioned experiment. The plot on the left side shows the result of fusing four features, namely peak-to-peak, maximum peak value and energy acquired from the second sensor and entropy from the first sensor. The plot on the right side shows a smooth decreasing monotonic function. It is a result of fusing two features, mutual information between both sensors and entropy which has been extracted from the second sensor. For the second data set the algorithm

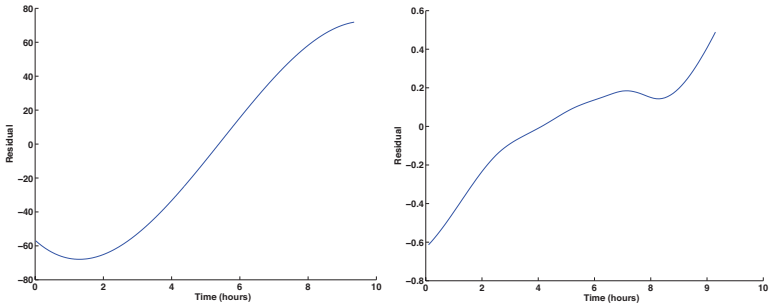


(a) (b)

Figure 5.16: NASA first monotonic data set.

generates 6 trends from the raw data. Figure 5.17 shows two selected monotonic health indicators for the aforementioned experiment. The plot on the left side shows the result of fusing two features; maximum peak value acquired from the second sensor and energy

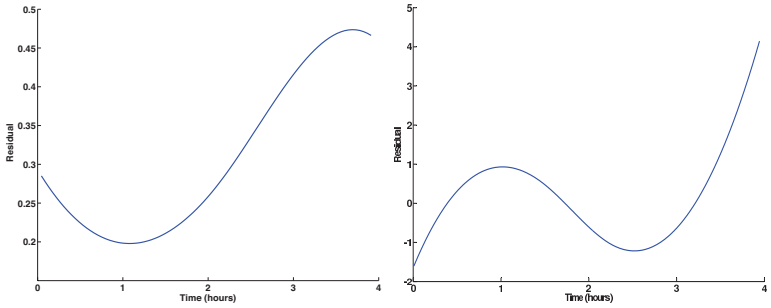
acquired from the first sensor. The plot on the right side shows a smooth decreasing monotonic function. It is a result of fusing two features, entropy of the first sensor and arithmetic mean of PSD acquired from the second sensor. For the final data set, the



(a) (b)

Figure 5.17: NASA second monotonic data set.

algorithm generates 10 trends from the raw data and non of them was monotonic. Figure 5.18 shows two selected non-monotonic health indicators for the same experiment. The plot on the left side is the result of fusing two features, peak-to-peak; acquired from the first sensor, and entropy of both sensors. The plot on the right side shows another indicator which was a result of fusing two features namely, root mean square and AR coefficient acquired from the second sensor.



(a) (b)

Figure 5.18: NASA third non-monotonic data set.

5.3.2 Turbofan engine data

The turbofan engine data sets are generated using Commercial Modular Aero-Propulsion System Simulation (C-MAPSS). They consist of four training files, four testing files and four RUL values files. The training files contain run to failure sensor records of a fleet of engines generated under different combinations of operational conditions and fault modes. Each engine is operating normally and it develops a fault at some point during

the operation until finally it reaches the system failure and the engine stops. The test files are generated in the same way; however, the sensor readings are omitted prior to system failure. The RUL files contain vector of true RUL values for the test data. Each training and test file contains 26 columns that represent different variables. The first two columns represent the engine number and the time in cycles, respectively. The next three columns represent the operational settings. The last 21 columns, or variables, represent different time series sensor data such as total temperature at fan inlet, pressure at fan inlet, physical fan speed, etc. Each row represents a data snapshot taken during a single cycle. In this work, the data file "train_FD001.txt" is used for offline training and "test_FD001.txt" is used for online testing. Each file contains data for 100 engines and the objective is to predict the number of remaining operational cycles before failure in the test set. The true RUL values for the test data are presented in the data file "RUL_FD001.txt".

Variable selection

One of the results of the selection algorithm is the pair of sensors number {8,13}, i.e. physical fan speed and corrected fan speed, respectively (Figure 5.19a). The selected group is interesting as the two variables are correlated and both are related to the fan speed. Then, the method starts constructing the monotonic trends iteratively from each pair at each time.

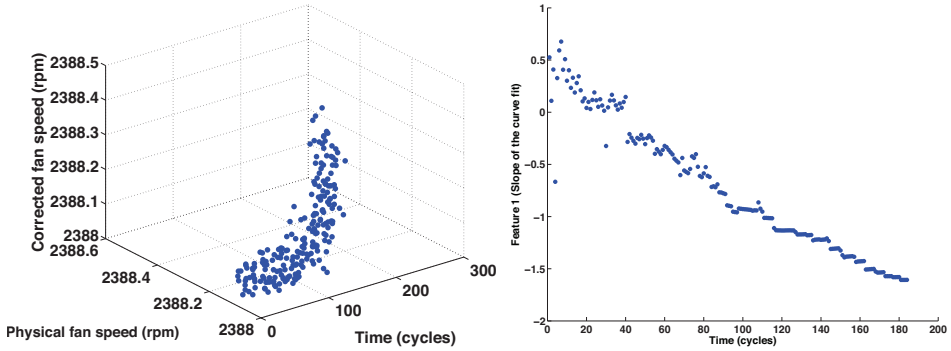
Health indicator

As mentioned before, four features are extracted from each trend at each time and labeled with end of life time to be saved in the offline database. The features represent the relation between the extracted trends and the engine's end of life. Figure 5.19b shows one of the four health indicators for the NASA training engine number 61. The indicator is monotonic and shows how the relation between the end of life and the extracted trend changes through the time. Each health indicator is then saved in offline database and labeled with the end of life time and will be used for predicting the RUL of the new test engine.

Prediction results

Figure 5.20 shows the predicted RUL for 4 engines at all cycles. It can be noticed that the accuracy of the predictions increases with the time. The prediction error at the last cycles is less than the errors at the beginning. To assess the performance of the proposed method, the Mean Absolute Percentage Error (MAPE) is calculated for all 100 online predictions:

$$MAPE(\%) = \frac{100\%}{n} \times \sum_{i=1}^{n} |\frac{RUL_i - RUL_i^*}{RUL_i}| \tag{5.25}$$



(a) The selected pair of sensors. (b) The slope health indicator.

Figure 5.19: Results of variable selection and health indicator construction for the NASA turbofan engine 61.

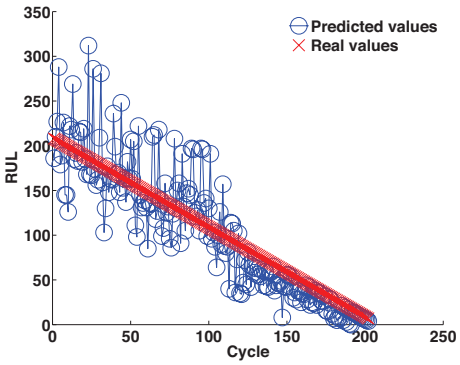
where RUL and RUL^* are the actual and predicted RUL values respectively and n is the number of total predictions. The error is calculated only for the last cycles of all 100 test signals. The MAPE over the 100 test data equals to 12.19%.

5.3.3 Lithium-ion battery data

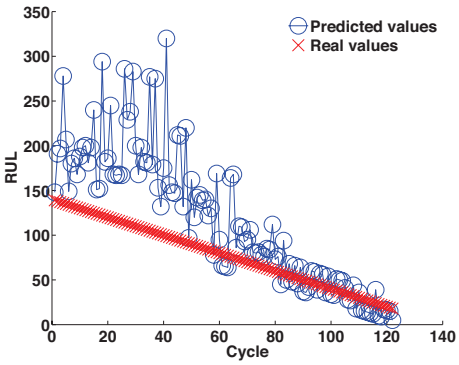
These data are collected on 34 lithium-ion batteries run through different operational profiles (e.g. charge, discharge and impedance) at different temperatures. In this work, only charge and discharge data are used. Each data set, corresponding to one experiment, consists of 11 variables such as charging voltage, charging current, temperature, discharging current, discharging voltage and capacity. The aging of the batteries was accelerated and the experiments continued until the batteries reached their end of life time. Each cycle is presented by the mean value to reduce the processing time. In order to validate the proposed method, a 3-fold cross-validation is applied, i.e. the available data sets are partitioned into three groups of equal size. Each group is then divided into training and testing data set as depicted in Table 5.1 and Table 5.2, respectively. Only 31 battery data sets are used in this experiment as three batteries, namely B0018, B0041 and B0053, do not have any similar data sets with the same end of life.

Variable selection

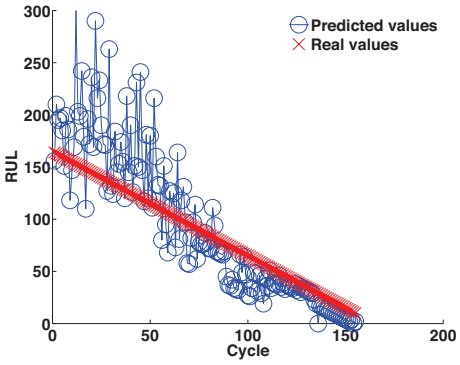
One of the results of the selection algorithm is the pair of variables {6, 11}, i.e. the voltage measured at discharge and the capacity of the battery (Figure 5.21). The selected pair is interesting because the two variables are correlated. Indeed, the capacity is related to the battery health as the decrease in the capacity indicates health degradation.



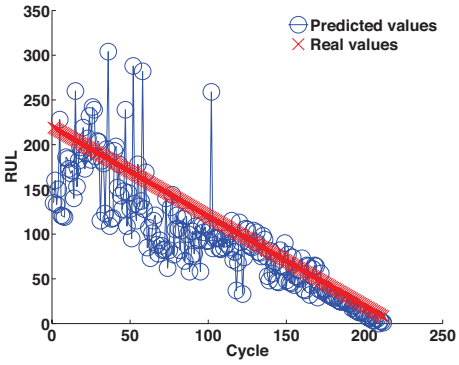
(a) RUL of engine 34.



(b) RUL of engine 41.



(c) RUL of engine 42.



(d) RUL of engine 81.

Figure 5.20: Results of predicting the RUL at all cycles for 4 engines.

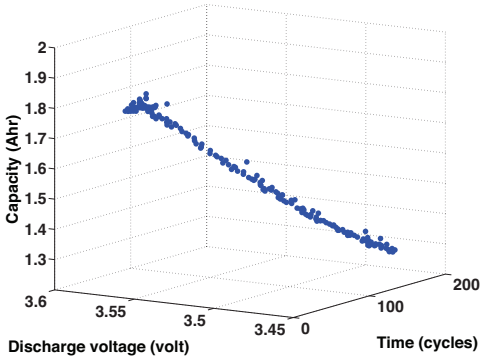


Figure 5.21: Selected pair of variables from the NASA battery B0005.

Health indicator

Four features are extracted from each trend at each time and labeled with end of life time to be saved in the offline database. Figure 5.22 shows two of the four health indicators

Table 5.1: Training data sets with three folds.

Fold #1	Fold #2	Fold #3	EOL
B0006	B0005	B0005	168
B0007	B0007	B0006	168
B0026	B0025	B0025	28
B0027	B0026	B0026	28
B0028	B0027	B0028	28
B0030	B0029	B0029	40
B0031	B0031	B0030	40
B0032	B0032	B0031	40
B0034	B0033	B0033	197
B0036	B0036	B0034	197
B0039	B0038	B0038	47
B0040	B0040	B0039	47
B0043	B0042	B0042	112
B0044	B0044	B0043	112
B0045	B0045	B0045	72
B0047	B0046	B0046	72
B0048	B0048	B0047	72
B0050	B0049	B0049	25
B0051	B0050	B0050	25
B0052	B0051	B0052	25
B0055	B0054	B0054	102
B0056	B0056	B0055	102

Table 5.2: Testing data sets with three folds.

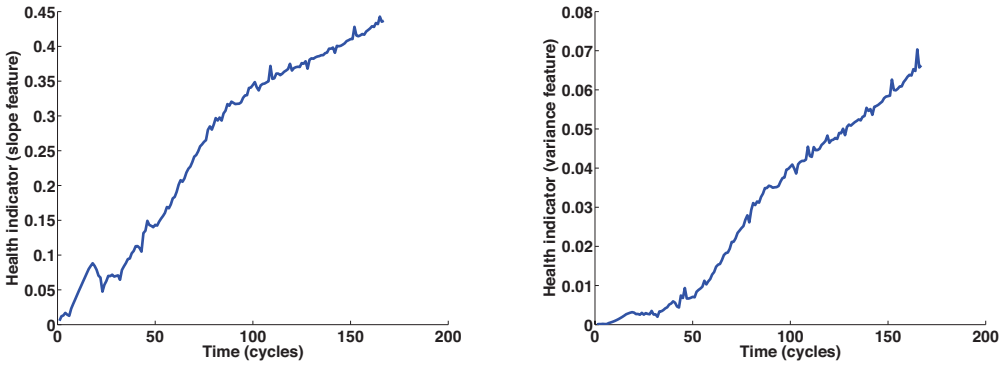
Fold #1	Fold #2	Fold #3	EOL
B0005	B0006	B0007	168
B0025	B0028	B0027	28
B0029	B0030	B0032	40
B0033	B0034	B0036	197
B0038	B0039	B0040	47
B0042	B0043	B0044	112
B0046	B0047	B0048	72
B0049	B0052	B0051	25
B0054	B0055	B0056	102

for the battery B0005. The indicators are monotonic and show how the relation between the end of life and the extracted trend changes through the time.

Prediction results

To assess the performance of the proposed method, a MAPE is calculated for all cycles of each battery (Figure 5.23). The average MAPE per fold is calculated as follows:

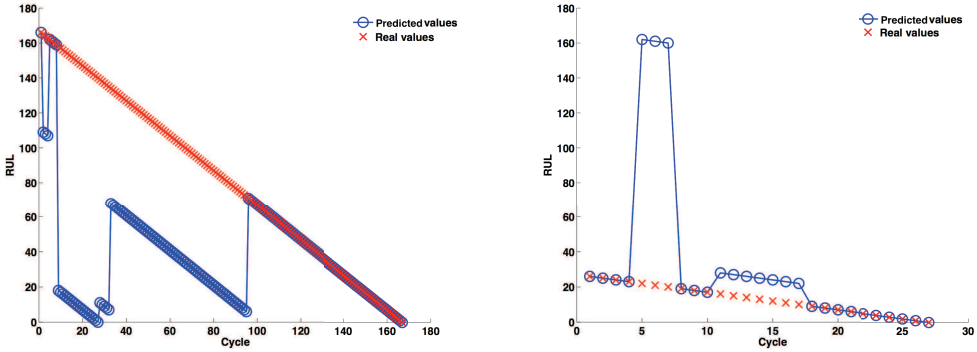
$$MAPE_f = \frac{1}{n} \times \sum_{i=1}^{n} MAPE_{i,f} \tag{5.26}$$



(a) Slope features over time. (b) Variance features over time.

Figure 5.22: Health indicators constructed from the NASA battery B0005.

where $MAPE_f$ is the average MAPE for a complete fold, $MAPE_{i,f}$ is the MAPE for test battery i in fold f. The final results are calculated and summarized in Table 5.3.



(a) RUL of battery B0005. (b) RUL of battery B0025.

Figure 5.23: Results of predicting the RUL at all cycles for 2 batteries.

Table 5.3: Mean absolute percentage error for the NASA battery data sets.

Fold #1	Fold #2	Fold #3	Average
28.0493%	26.3089%	28.3536%	27.5706%

Figure 5.23a shows plot of the predicted RUL for all cycles for battery B0005. It can be seen that the prediction accuracy increases with time, i.e, the longer the test trend is the higher the prediction accuracy. Figure 5.23b shows a plot of the RUL predicted

for the battery B0025. Only 10 cycles were considered as late prediction. However, the error was decreasing at the later cycles.

5.4 Conclusion

In this chapter the online method for health assessment and RUL prediction is presented. The method is based on two main parts, namely performing health assessment and estimating the RUL of the monitored component. The health assessment is done first by estimating the health indicators values recursively from the monitored component sensor data using Bayesian filter. The Bayesian filter represents the uncertainty about the health status in a probabilistic form. Furthermore, the RUL estimation is performed by finding the most similar offline model in the training database using two different Bayesian approaches such as classification and regression. The choice of Bayesian algorithm for selecting the most similar offline group of HIs from the offline data set and health state estimation is based on the ability of the algorithms to represent the decisions in probabilistic form. This can be useful for decision making in later steps.

A disadvantage of the proposed method is that it depends on the data. If the data set does not show all the degradation instances, the prediction error will increase. In order to tackle this problem, the proposed method utilized GPR to reduce such error. An advantage of the proposed method is that it can be applied to different components as long as the offline data base is available and the online HIs are represented in a similar representation like the offline models. furthermore, the proposed method utilizes machine learning algorithms which are based on Bayesian theory. In this way, uncertainty about the results or (RUL prediction) are represented in a probabilistic form. The probabilistic representation of the RUL prediction is important for two reasons. First, it can be useful for the decision making step. Second, it can be useful to enhance the prediction results by fusing different models according to the level of the uncertainty associated with each model. If the k-NN class posterior value is less than a certain threshold, the method uses GPR model to predict the RUL. Furthermore, the method was applied successfully on three different applications, namely bearings, batteries and turbofans and can be generalized to more applications given the same assumptions used in this work.

This chapter concludes the contributions presented in this thesis. The general conclusions and future work will be presented in the next chapter.

6

Conclusion and future work

Prognostics and Health Management (PHM) has become an important element to achieve efficient condition and predictive based maintenance. Different PHM tasks attracted significant research interest due to the need for failure prediction and decision making models. Such models are paramount for performing efficient predictive maintenance strategies for industrial systems. Prognostics models, in particular, are becoming important due to the need for health assessment and remaining useful life models. Building prognostics models for a whole industrial system, however, is challenging and still quite difficult in practice. Instead, component-oriented prognostics approaches are used to build models for specific critical subsystems or components. To do that, system experts have to first select the critical component(s) to be monitored and decide the type and the number of the monitoring sensors. Sensor data are then gathered, through the life cycle of the component, and analyzed to build prognostics models. Prognostics models can be realized using three main methods, namely model-based, data driven and hybrid methods. It can be very challenging to use physics based or hybrid models due to the increased complexity of the critical component. On the other hand, data driven methods are attracting a lot of research lately because such methods can build prognostics models faster with less costs compared to other approaches.

Data-driven methods employ different computational intelligence (CI) and machine learning (ML) models, such as artificial neural networks, fuzzy logic, Bayesian models and similarity based models, to learn the degradation mechanism from sensor data. This sensor data are recorded over the entire operating process of the desired critical components. If the recorded data contain the degradation behavior, data driven methods infer this knowledge without using the first principles. Unlike the other data driven models, Bayesian approaches have a natural way of representing the uncertainty in a probabilistic form. Furthermore, building Bayesian models does not require understanding the system behavior and it can be used to model multidimensional dynamic systems. There are two main approaches to build data driven models, namely cumulative degradation based prognostics approaches and direct RUL mapping approaches. Cumulative degradation based prognostics approaches use empirical models to map the degradation evolution of the desired system. These models are later used to estimate the new system health

status. After knowing the new system's current health status, the RUL can be predicted based on the expected future behavior. The drawback of such approaches is that some times it is difficult to identify the different degradation stages of the monitored component. On the other hand, direct RUL mapping approaches use empirical models to directly map the relation between sensor data and the corresponding EOL value. Direct RUL mapping approaches require only run to failure data with known EOL values. The advantage of such approaches is that they do not require prior knowledge about the degradation states of the monitored component.

In this thesis, we proposed an approach for direct remaining useful life prognostics of critical components based on Bayesian approaches. The approach is divided in two main phases, namely offline and online. In the offline phase, the approach starts by looking for "interesting" variables in the form of non-random relationships among measured sensory signals, or features, derived from signals. The assumption is that information about the degradation of a component can be extracted from the relationships between signals on-board the same component. The selected variables are then compressed using PCA into a compact form. Then, EMD algorithm is applied to extract monotonic trends that represent the degradation of the critical component through time. Next, statistical features are extracted from the trends to represent each trend in a compact form through the time and use them as health indicators (HI) of the monitored component. From the HIs, the original signals can be retrieved which can be used to interpret the true physical behavior of the monitored component. In the online phase, the new data acquired goes through the same steps as the offline phase. Then, histogram filter is used to recursively estimate the health status of the testing component. It also represents the uncertainty about the new data in a probabilistic form. The new acquired trends are then compared to the trends learned during the offline phase first by using k-NN classifier. If the posterior probability of the selected class is less than a certain threshold, then the method uses GP to approximate the closest correct group. In both cases, the method associates a probability value to the decision to represent the uncertainty about it. The performance of the proposed approach is evaluated using three data sets, namely, bearings, turbofan engines and lithium-ion batteries data downloaded from the NASA prognostic center of excellence website. The results show the efficiency of the method in finding important relationships in different applications without any prior knowledge. Furthermore, the prediction results show low MAPE error for two applications.

6.1 Contributions

While in Section 1.5 (page 6), a detailed list of the contributions of this work is presented, this section emphasis on particular aspects of PHM, where our work makes contributions.

1. Selecting informative variables that hold non random relationships, without prior knowledge about the component, to represent the relation between the input signals. In this way, the proposed method can discover new relations that were not known before. Such relations can help to understand the physical meaning of the failure mechanism.

2. Construct different health indicators that represent the degradation evolution of the monitored components regardless of the application. This health indicator is based on two main steps. The first step is to extract the EMD residual signal. This signal can represent the degradation behavior of the monitored component. If the component is degrading, the resulting trend becomes monotonic. On the other hand, if the component is not degrading, the resulting trend becomes constant. The second step is to extract time domain features from the resulting trends. In this way, the proposed method for health indicator construction reflects the behavior of the signal by just revealing the internal structure of the signal.

3. Besides the computational simplicity of the data-driven models used in this method, the choice of Bayesian based models was made to represent the uncertainty about the health status and the RUL estimation. Bayesian models have a natural way of representing the uncertainty in a probabilistic form, which can be useful for decision making.

4. Integrating different machine learning algorithms, such as k-NN, GP and histogram filter, to learn the degradation evolution, assess the health status of the components and estimate their RUL that increase the prediction efficiency.

5. Applying the proposed approach on different applications to show its generality as well as its efficiency.

6.2 Limitations

It is worth noting that the presented framework has limitations that have yet to be studied in order to be implemented as part of the future work. One limitation of the proposed framework is that the choosing the values of the different parameters, such as number of the neighbors for the k-NN, the threshold to choose between k-NN and GPR estimation and GPR parameters are based on try and error. Furthermore, the framework assesses the health and estimates the RUL at each cycle of the monitored component. Thus, the processing resources required by the method increase through time. This might not be feasible for applications with limited computational resources and long life span.

6.3 Future work

While significant advances have been achieved recently in the field of component oriented PHM research, there are still many open questions that need answers.

1. The use of event monitoring data is not common in the literature. Most of the work done utilizes condition monitoring data. Combining both data to realize PHM tasks might increase the efficiency of the proposed methods.

2. Although, many research works have proposed different criteria for selecting informative sensor data, yet most of these methods depends on visual inspection of the data sets. A unified way to select sensor data to build prognostics models might be interesting to study.

3. Data-driven methods are usually black-box approaches with no physical meaning of the developed models. In this work, the first step towards building the proposed models was to first select sensor data that have nonlinear relationship. The selected variables were later processed to derive prognostics models. It might be interesting to see the relation between the processed signals and their original signals, which might help to build physical models.

4. A possible research work to be addressed in the future research is health assessment and RUL estimation for data with variable operating conditions. Normally, the operating condition of an industrial system changes due to many reasons and therefore the proposed models should take account of such variations.

5. Another key aspect to be addressed is the effect of maintenance interventions in the industrial system. Replacing or fixing a faulty component has an effect on the monitoring signals. It is important to see how the prognostics models should deal with such changes.

6. To the author's knowledge, the effect of human operator on industrial systems has not been studied. Human operators might differ from each other in terms of skills and behaviors which might affect how they operate the industrial system. As an example, a non skilled bus driver might cause a faster deterioration of the bus compared to a skilled driver. This aspect should be considered in the future PHM research.

Bibliography

[1] Muller, Michael, Don Kasten, Kevin Comer, Michael Simek, Chris Tierney, and Gustavo Garcia. (1996) "Industrial Productivity Training Manual." In 1996 Annual IAC Director's Meeting, sponsored by DOE/OIT, The State University of New Jersey (Rutgers).

[2] Benjamin, K. Sovacool.(2008) "The Dirty Energy Dilemma: What's Blocking Clean Power in the United States." Praegar, Westport, CT.

[3] Konstantinos Tzanakakis, (2013) "The Railway Track and Its Long Term Behaviour: A Handbook for a Railway Track of High Quality." Springer, vol 2, DOI: 10.1007/978-3-642-36051-0.

[4] O'Kane, M.M., (2000) "Intelligent motors moving to the forefront of predictive maintenance", IEEE conference of Petroleum and Chemical Industry, Pages 217-223, DOI: 10.1109/PCICON.2000.882778.

[5] V. Deshpande, J. Modak. (2002) "Application of rcm for safety considerations in a steel plant". Reliability Engineering and System Safety, Volume 78, Issue 3, Pages 325-334. DOI: 10.1016/S0951-8320(02)00177-1

[6] C. Chu, J.m. Proth, P. Wolff, (1998) "Predictive maintenance: the one-unit replacement model", International Journal of Production economics, vol 54, pages 285-295.

[7] Zhen Zhao, Fu-li Wang, Ming-xing Jia, Shu Wang, Predictive maintenance policy based on process data, Chemometrics and Intelligent Laboratory Systems, Volume 103, Issue 2, 15 October 2010, Pages 137-143, DOI: http://dx.doi.org/10.1016/j.chemolab.2010.06.009

[8] Maintenance Terminology. European Standard. CEN (European Committee for Standardization), EN 13306:2001, Brussels, 2001.

[9] Tsang, A.H.C., Jardine, A. K.S., Kolodny, H. (1999) "Measuring maintenance performance: a holistic approach" International Journal of Operations and Production Management volume 19, Iss. 1, pp 691-715.

[10] Knezevic J. (1993) "Reliability, maintainability and support-ability engineering: a probabilistic approach". McGraw Hill.

[11] Riis, Jens O., James T. Luxhoj, and Uffe Thorsteinsson. (1997) "A situational maintenance model." International Journal of Quality & Reliability Management volume 14, issue 4, pages 349-366.

[12] Besnard, Francois, M. Patrikssont, AB. Strombergt, Adam Wojciechowski, and Lina Bertling. (2009) "An optimization framework for opportunistic maintenance of offshore wind power system." In IEEE Bucharest PowerTech. pages 1-7, DOI: 10.1109/PTC.2009.5281868. Bucharest, Romania.

[13] Yuanhang Wang, Chao Deng, Jun Wu, Yingchun Wang, Yao Xiong, (2014) "A corrective maintenance scheme for engineering equipment", Engineering Failure Analysis, Volume 36, Pages 269-283, DOI: http://dx.doi.org/10.1016/j.engfailanal.2013.10.006.

[14] Rosmaini Ahmad, Shahrul Kamaruddin, (2012) "An overview of time-based and condition-based maintenance in industrial application", Computers & Industrial Engineering, Volume 63, Issue 1, Pages 135-149,DOI: http://dx.doi.org/10.1016/j.cie.2012.02.002.

[15] Navarra, K.; Lawton, R.; Hearrell, N., (2007) "An Enterprise Strategy for Implementing Conditioned-Based Maintenance Plus (CBM+) Research in the USAF," IEEE Aerospace Conference, pp.1,7, DOI: 10.1109/AERO.2007.352916

[16] Jaw, L.C.; Merrill, W., (2008) "CBM+ Research Environment - Facilitating Technology Development, Experimentation, and Maturation," IEEE Aerospace Conference, pp.1,6, DOI: 10.1109/AERO.2008.4526641

[17] M Bevilacqua, M Braglia, (2000) "The analytic hierarchy process applied to maintenance strategy selection" Reliability Engineering & System Safety, volume 70, issue 1, pages 71-83.DOI: http://dx.doi.org/10.1016/S0951-8320(00)00047-8.

[18] Fausto Pedro Garcia Marquez, Andrew Mark Tobias, Jesus Maria Pinar Perez, Mayorkinos Papaelias "Condition monitoring of wind turbines: Techniques and methods", Renewable Energy, Volume 46, October 2012, Pages 169-178, DOI: http://dx.doi.org/10.1016/j.renene.2012.03.003.

[19] Andrawus, Jesse A., John Watson, Mohammed Kishk, and Allan Adam. (2006) "The selection of a suitable maintenance strategy for wind turbines." Journal of wind engineering, volume 30, issue 6, pages: 471-486.

[20] Almeida AT, Bohoris GA. (1995) "Decision theory in maintenance decision making". Journal of Quality in Maintenance Engineering.volume 1, issue: 1, pages: 39-45. DOI: 10.1108/13552519510083138.

[21] Triantaphyllou E, Kovalerchuk B, Mann L, Knapp GM. (1997) "Determining the most important criteria in maintenance decision making". Journal of Quality in Maintenance Engineering. volume: 3, issue: 1, pages: 16-24. DOI: 10.1108/13552519710161517.

[22] Marvin Rausand. (1998) “Reliability centered maintenance”. Reliability Engineering & System Safety, volume 60, issue 2, pages 121-132, DOI: http://dx.doi.org/10.1016/S0951-8320(98)83005-6.

[23] J Crocker, U.D Kumar. (2000) “Age-related maintenance versus reliability centered maintenance: a case study on aero-engines”. Reliability Engineering & System Safety, volume 67, issue 2 , pages 113-118, DOI: http://dx.doi.org/10.1016/S0951-8320(99)00052-6.

[24] Moubray, J., (1997) “Reliability-Centered Maintenance”. New York City, NY: Industrial Press, Inc.

[25] Anthony M. Smith. (1993) “Reliability Centered Maintenance”. New York: McGraw-Hill.

[26] Fausto Pedro Garcia Marquez and Diego J. Pedregal. (2007) “Applied RCM2 algorithms based on statistical methods”. International Journal of Automation and Computing. volume: 4, issue: 2, pages: 109-116. DOI:10.1007/s11633-007-0109-1.

[27] Ben-Daya MS, Duffuaa AR. (2009) “Handbook of maintenance management and engineering”. Springer verlag London Limited.

[28] Deshpande VS, Modak JP. “Application of RCM for safety considerations in a steel plant”. Reliability Engineering and System Safety 2002;3(78):32534.

[29] Garcia Marquez FP, Schmid F, Collado JC. “A reliability centered approach to remote condition monitoring. A railway points case study.” Reliability Engineering & System Safety 2003;80(1):3340.

[30] Fausto Pedro Garcia Marquez, Felix Schmid, Javier Conde Collado, (2003) “Wear assessment employing remote condition monitoring: a case study”, Wear, Volume 255, Issues 7-12, Pages 1209-1220, DOI: http://dx.doi.org/10.1016/S0043-1648(03)00214-X.

[31] Andrawus JA, Watson J, Kishk M, Adam A. The selection of suitable maintenance strategy for wind turbines. Wind Engineering 2009;30(6):47186.

[32] Byon E, Ding Y. Season-dependent condition-based maintenance for a wind turbine using a partially observed markov decision process. IEEE Transactions on Power Systems 2010;25(4):182334.

[33] McMillan D, Ault GW. Condition monitoring benefit for onshore wind turbines: sensitivity to operational parameters. IET Renewable Power Generation 2008;2(1):6072.

[34] Luna, Joel, Paul Kolodziejski, Nick Frankle, Donald C. Conroy, and Ron Shroder. (2009) “Strategies For Optimizing The Application Of Prognostic Health Management To Complex Systems.” In Machine Failure Prevention Technology Conference (MFPT).

[35] Haddad, G.; Sandborn, P.; Pecht, M., (2011) "Using real options to manage condition-based maintenance enabled by PHM," IEEE Conference on Prognostics and Health Management (PHM), pp.1,7, DOI: 10.1109/ICPHM.2011.6024318

[36] Bae, Sanghoon, Hanju Cha, and Youngsuk Suh (2014) "Study on Condition Based Maintenance Using On-Line Monitoring and Prognostics Suitable to a Research Reactor". European conference of the prognostics and health management society.

[37] Dragomir, Otilia Elena, Rafael Gouriveau, Florin Dragomir, Eugenia Minca, and Noureddine Zerhouni. "Review of prognostic problem in condition-based maintenance." In European Control Conference, ECC'09., pp. 1585-1592. 2009.

[38] Carl French, Data Processing and Information Technology, Cengage Learning Business Press, 1996, ISBN 1844801004.

[39] Stuart J. Russell and Peter Norvig. Artificial Intelligence: A Modern Approach. Pearson Education, 2003.

[40] L. H. Chiang, E. L. Russell, and R. D. Braatz, Fault Detection and Diagnosis in Industrial Systems, Springer, New York, NY, USA, 2001.

[41] Svensson M, Rognvaldsson T, Byttner S, West M, Andersson B (2011) Unsupervised deviation detection by GMM - a simulation study. IEEE international symposium on diagnostics for electric machines, power electronics and drives, pp 51-54.

[42] Isermann, R. (1984). Process fault detection based on modelling and estimation methods - a survey. Automatica, Vol. 20, No. 4, pp. 387-404.

[43] Gertler, J. (1998). Fault detection and diagnosis in engineering systems. Marcel Dekker, New York

[44] Frank, P.M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy. Automatica, Vol. 26, pp. 459-474.

[45] Chen, J. and R.J. Patton (1999). Robust model-based fault diagnosis for dynamic systems. Kluwer, Boston

[46] Patton, R.J., P.M. Frank and P.N. Clark (2000) Issues of fault diagnosis for dynamic systems. Springer, Berlin.

[47] Isermann, R. (1993). Fault diagnosis of machines via parameter estimation and knowledge processing. Automatica, Vol. 29, No. 4, pp. 815-835.

[48] Clark, R.N. (1978). A simplified instrument detection scheme. IEEE Trans. Aerospace Electron. Syst., Vol. 14, pp. 558- 563 and 456-465.

[49] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewen Yin, Surya N. Kavuri, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods", Computers & Chemical Engineering, Volume 27, Issue 3, 15 March 2003, Pages 293-311, DOI: http://dx.doi.org/10.1016/S0098-1354(02)00160-6.

[50] Markos Markou, Sameer Singh, "Novelty detection: a review-part 1: statistical approaches", Signal Processing, Volume 83, Issue 12, December 2003, Pages 2481-2497, DOI: http://dx.doi.org/10.1016/j.sigpro.2003.07.018.

[51] Junghui Chen, Chien-Mao Liao, "Dynamic process fault monitoring based on neural network and PCA", Journal of Process Control, Volume 12, Issue 2, February 2002, Pages 277-289, DOI: http://dx.doi.org/10.1016/S0959-1524(01)00027-0.

[52] S. Byttner, T. Rognvaldsson, M. Svensson, Consensus self-organized models for fault detection (COSMO), Engineering Applications of Artificial Intelligence, Volume 24, Issue 5, August 2011, Pages 833-839, DOI: http://dx.doi.org/10.1016/j.engappai.2011.03.002.

[53] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewen Yin, Surya N. Kavuri: A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. 293-311

[54] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Surya N. Kavuri: A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies. 313-326.

[55] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Surya N. Kavuri, Kewen Yin: A review of process fault detection and diagnosis: Part III: Process history based methods. 327-346

[56] P.M. Frank, Analytical and Qualitative Model-based Fault Diagnosis - A Survey and Some New Results, European Journal of Control, Volume 2, Issue 1, 1996, Pages 6-28, DOI: http://dx.doi.org/10.1016/S0947-3580(96)70024-9.

[57] Shen Yin, Steven X. Ding, Adel Haghani, Haiyang Hao, Ping Zhang, A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process, Journal of Process Control, Volume 22, Issue 9, October 2012, Pages 1567-1581, DOI: http://dx.doi.org/10.1016/j.jprocont.2012.06.009.

[58] K. Zhang. Fault Detection and Diagnosis for Multi-Actuator Pneumatic Systems. PhD thesis, Deparment of Mechanical Engineering, SUNY at Stony Brook, May 2011.

[59] Isermann, R. (1992). Estimation of physical parameters for dynamic processes with application to an industrial robot. Int. J. Control, Vol. 55, pp. 1287-1298.

[60] Willsky, A.S. (1976). A survey of design methods for failure detection systems. Automatica, Vol. 12, pp. 601-611.

[61] Himmelblau, D.M. (1978). Fault detection and diagnosis in chemical and petrochemical processes, pp. 343-393. Elsevier: Amsterdam.

[62] ISO 13381-1, Condition Monitoring and Diagnostics of Machines - Prognostics - Part 1: General Guidelines: International Standards Organization, 2004.

[63] Kai Goebel, Abhinav Saxena, Matt Daigle, Jose Celaya, Indranil Roychoudhury. (2012) "Introduction to Prognostics", European PHM conference, Online tutorial, last visited October 2014.

[64] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical Systems and Signal Processing 20 (7) (2006) 1483-1510.

[65] S.J. Engel, B.J. Gilmartin, K. Bongort, A. Hess, Prognostics, the real issues involved with predicting life remaining, Aerospace Conference Proceedings, vol. 6, IEEE, 2000, pp. 457-469.

[66] A. Hess, G. Calvello, P. Frith, Challenges, issues, and lessons learned chasing the "Big P", Real predictive prognostics, Part 1, IEEE Aerospace Conference , 2005, pp. 3610-3619.

[67] W. Wu, J. Hu, J. Zhang, Prognostics of machine health condition using an improved ARIMA-based prediction method, IEEE, Harbin, China, 2007, pp. 1062-1067.

[68] J. Luo, M. Namburu, K. Pattipati, L. Qiao, M. Kawamoto, S. Chigusa, Model-Based Prognostic Techniques, Anaheim, CA, United States: 2003, Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ, United States, 2003, pp. 330-340.

[69] T. Brotherton, "A testbed for data fusion for engine diagnostics and prognostics", IEEE Aerospace Conference, Big Sky MT, 2002, pp. 8-15.

[70] S. Katipamula, M.R. Brambley, "Methods for fault detection, diagnostics, and prognostics for building systems - a review", Part I, HVAC and R Research 11 (1) (2005) 3 - 25.

[71] S.A. Lewis, T.G. Edwards, "Smart sensors and system health management tools for avionics and mechanical systems", Digital Avionics Systems Conference, 1997.

[72] A.E. Smith, D.W. Coit, Y.C. Liang, "A neural network approach to condition based maintenance: case study of airport ground transportation vehicles", IMA Journal Management Mathematics on Maintenance, Reliability and Replacement April, 2003.

[73] P. Baruah, R.B. Chinnam, "HMMs for diagnostics and prognostics in machining processes", International Journal of Production Research 43 (6) (2005) 1275-1293.

[74] A. Heng, A.C.C. Tan, J. Mathew, N. Montgomery, D. Banjevic, A.K.S. Jardine, Intelligent condition-based prediction of machinery reliability, Mechanical Systems and Signal Processing 23 (5) (2009) 1600-1614.

[75] Wilcock, G. W., "Enhanced fault management for future IMA systems" IEEE Digital Avionics Systems Conference, vol 1, pp. 3.2 - 32-9. DOI: 10.1109/DASC.1997.635050

[76] Kalgren, P.W.; Byington, C.S.; Roemer, M.J., (2006) "Defining PHM, A Lexical Evolution of Maintenance and Logistics," IEEE Autotestcon conference, pp.353-358, DOI: 10.1109/AUTEST.2006.283685

[77] Filev, D.P., Tseng, F. Novelty detection based machine health prognostics. International Symposium on Evolving Fuzzy Systems pp. 193–199 (2006)

[78] Tavares, G., Zsigraiova, Z., Semiao, V., da Graca Carvalho, M.: Monitoring, fault detection and operation prediction of msw incinerators using multivariate statistical methods. Waste Management pp. 1635–1644 (2011)

[79] Miao, Q., Wang, D., , Pecht, M.: A probabilistic description scheme for rotating machinery health evaluation. Journal of Mechanical Science and Technology pp. 2421–2430 (2010)

[80] Cottrell, M., Gaubert, P., Eloy, C., Franscois, D., Hallaux, G., Lacaille, J., Verleysen, M.: Fault prediction in aircraft engines using self-organizing maps. Advances in Self-Organizing Maps pp. 37–44 (2009)

[81] Agogino, A., Tumer, K.: Entropy based anomaly detection applied to space shuttle main engines. IEEE Aerospace Conference (2006)

[82] Svensson, M., Rognvaldsson, T., Byttner, S., West, M., Andersson, B.: Unsupervised deviation detection by gmm; a simulation study. IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives pp. 51–54 (2011)

[83] Pontoppidan, N.H., Larsen, J.: Unsupervised condition change detection in large diesel engines. IEEE 13th Workshop on Neural Networks for Signal Processing pp. 565–574 (2003)

[84] CEMPEL, C.: Multidimensional condition monitoring of mechanical systems in operation. Mechanical Systems and Signal Processing pp. 1291–1303 (2003)

[85] CEMPEL, C.: Multifault condition monitoring of mechanical systems in operation. IMEKO XVII, Croatia pp. 1–4 (2003)

[86] Moshou, D., Kateris, D., Sawalhi, N., Loutridis, S., Gravalos, I.: Fault severity estimation in rotating mechanical systems using feature based fusion and self-organizing maps. Artificial Neural Networks pp. 410–413 (2010)

[87] Geramifard, O., Xu, J.X., Zhou, J.H., Li., X.: Continuous health assessment using a single hidden markov model. 11th International Conference on Control Automation Robotics and Vision pp. 1347–1352 (2010)

[88] Jie, S., Hong, G.S., Rahman, M., Wong, Y.S.: Feature extraction and selection in tool condition monitoring system. Advances in Artificial Intelligence pp. 487–497 (2002)

[89] Geramifard, O., Xu, J.X., Pang, C.K., Zhou, J.H., Li, X.: Data-driven approaches in health condition monitoring; a comparative study. 8th IEEE International Conference on Control and Automation pp. 1618–1622 (2010)

[90] Yang, Z., Zhong, J., Wong, S.F.: Machine learning method with compensation distance technique for gear fault detection. 9th World Congress on Intelligent Control and Automation pp. 632–637 (2011)

[91] Wald, R., Khoshgoftaar, T.M., Sloan, J.C.: Using feature selection to determine optimal depth for wavelet packet decomposition of vibration signals for ocean system reliability. IEEE 13th International Symposium on High-Assurance Systems Engineering pp. 236–243 (2011)

[92] Jiang, H., Penman, J.: Using kohonen feature maps to monitor the condition of synchronous generators. Workshop on Neural Network Applications and Tools pp. 89–94 (1993)

[93] Zhang, S., Hodkiewicz, M., Ma, L., Mathew, J., Kennedy, J., Tan, A., Anderson, D.: Machinery condition prognosis using multivariate analysis. Engineering Asset Management pp. 847–854 (2006)

[94] Zhang, X., Xu, R., Kwan, C., Liang, S., Xie, Q., Haynes, L.: An integrated approach to bearing fault diagnostics and prognostics. Proceedings of the 2005 American Control Conference pp. 2750–2755 (2005)

[95] Geramifard, O., Xu, J.X., Zhou, J.H., Li, X.: Continuous health condition monitoring: A single hidden semi-markov model approach. IEEE Conference on Prognostics and Health Management pp. 1–10 (2011)

[96] Zhou, J.H., Pang, C.K., Lewis, F.L., Zhong, Z.W.: Intelligent diagnosis and prognosis of tool wear using dominant feature identification. IEEE Transactions on Industrial Informatics pp. 454–464 (2009)

[97] Zhang, S.: Function estimation for multiple indices trend analysis using self-organizing mapping. IEEE Symposium on Emerging Technologies and Factory Automation pp. 160–165 (1994)

[98] Brown, D., Georgoulas, G., Chen, R., Ho, Y.H., Tannenbaum, G., Schroeder, J.: Particle filter based anomaly detection for aircraft actuator system. IEEE Aerospace conference, Monatan, USA (2009)

[99] Saha, B., Goebel, K., Poll, S., Christopherson, J.: A bayesian framework for remaining useful life estimation. Proc. Fall AAAI Symposium: AI for Prognostics, Arlington, VA (2007)

[100] Saha, B., Goebel, K., Poll, S., Christophersen, J.: Prognostics methods for battery health monitoring using a bayesian framework. IEEE Transactions on Instrumentation and Measurement pp. 291–296 (2009)

[101] Aiwina Heng, Sheng Zhang, Andy C.C. Tan, Joseph Mathew. "Rotating machinery prognostics: State of the art, challenges and opportunities." Mechanical Systems and Signal Processing, volume 23, issue 3, pages 724-739, April 2009. DOI: dx.doi.org/10.1016/j.ymssp.2008.06.009.

[102] R. Isermann, Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance. Heidelberg: Springer-Verlag, Heidelberg, 2006.

[103] G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, B. Wu, Intelligent Fault Diagnosis and Prognosis for Engineering Systems, John Wiley and Sons Inc., Hoboken, New Jersey, 2006.

[104] J. Luo, M. Namburu, K. Pattipati, L. Qiao, M. Kawamoto, S. Chigusa, Model-Based Prognostic Techniques, Anaheim, CA, United States: 2003, Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ, United States, 2003, pages 330-340.

[105] Fakher Chaari, Tahar Fakhfakh, Mohamed Haddar. "Analytical modelling of spur gear tooth crack and influence on gearmesh stiffness." European Journal of Mechanics - A/Solids, volume 28, issue 3, pages 461-468, June 2009. DOI: dx.doi.org/10.1016/j.euromechsol.2008.07.007.

[106] Zhenyou Zhang, Yi Wang, Kesheng Wang. "Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network". Journal of Intelligent Manufacturing, volume 24, issue 6, pages 1213-1227, December 2013. DOI: 10.1007/s10845-012-0657-2

[107] He, D., Li, R., Bechhoefer, E. (2012). "Stochastic modeling of damage physics for mechanical component prognostics using condition indicators." Journal of Intelligent Manufacturing, 23, 221226.

[108] M.A. Schwabacher, (2005) "A Survey of Data-Driven Prognostic," Aerospace Infotech, pages 26 - 29, Arlington, Virginia.

[109] G.E.P. Box, G.M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, 1976.

[110] F. Lewis, Applied Optimal Control and Estimation: Digital Design and Implementation, Prentice-Hall, Englewood Cliffs, NJ, 1992.

[111] R.S. Tsay, Time series and forecasting: brief history and future research, Journal of the American Statistical Association 95 (450) (2000), pages 638-643.

[112] W.Wu, J.Hu, J. Zhang, Prognostics of machine health condition using an improved ARIMA-based prediction method, IEEE, Harbin, China, 2007, pages 1062-1067.

[113] J. Yan, M. Koc, J. Lee, A prognostic algorithm for machine performance assessment and its application, Production Planning and Control 76 (2004), pages 796-801.

[114] J. Lee, J. Ni, D. Djurdjanovic, H. Qiu, H. Liao, Intelligent prognostics tools and e-maintenance, Computers in Industry 57 (6) (2006), pages 476-489.

[115] N. Gebraeel, M. Lawley, R. Liu, V. Parmeshwaran, Residual life predictions from vibration-based degradation signals: a neural network approach, IEEE Transactions on Industrial Electronics 51 (3) (2004), pages 694-700.

[116] B. Satish, N.D.R. Sarma, A fuzzy BP approach for diagnosis and prognosis of bearing faults in induction motors, IEEE Power Engineering Society General Meeting, IEEE, 2005, pages 2291-2294.

[117] R. Huang, L. Xi, X. Li, C. Richard Liu, H. Qiu, J. Lee, Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods, Mechanical Systems and Signal Processing 21 (1) (2007), pages 193-207.

[118] Z. Lei, L. Xingshan, Y. Jinsong, G. ZhanBao, A genetic training algorithm of wavelet neural networks for fault prognostics in condition based maintenance, in: Proceedings of the Eighth International Conference on Electronic Measurement and Instruments, IEEE, 2007, pages 584-589.

[119] A.P. Vassilopoulos, E.F. Georgopoulos, V. Dionysopoulos, Artificial neural networks in spectrum fatigue life prediction of composite materials, International Journal of Fatigue 29 (1) (2007), pages 20-29.

[120] Zhigang Tian. "An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring". Journal of Intelligent Manufacturing, volume 23, issue 2, pages 227-237, April 2012. DOI: 10.1007/s10845-009-0356-9

[121] Kamran Javed, Rafael Gouriveau, Noureddine Zerhouni." Novel failure prognostics approach with dynamic thresholds for machine degradation." 39th Annual Conference of the IEEE Industrial Electronics Society, (IECON), pages 4404-4409, 10-13 November 2013 DOI: 10.1109/IECON.2013.6699844

[122] Brezak, D., Majetic, D., Udiljak, T., Kasac, J. (2012). Tool wear estimation using an analytic fuzzy classifier and support vector machines. Journal of Intelligent Manufacturing, 23, 797809.

[123] Gajate, A., Haber, R., Del Toro, R., Vega, P., Bustillo, A. (2012). Tool wear monitoring using neuro-fuzzy techniques: A comparative study in a turning process. Journal of Intelligent Manufacturing, 23, 869882.

[124] Purushothaman, S. (2010). Tool wear monitoring using artificial neural network based on extended Kalman filter weight updation with transformed input patterns. Journal of Intelligent Manufacturing, 21, 717730.

[125] Yeo, S. H., Khoo, L. P., Neo, S. S. (2000). Tool condition monitoring using reflectance of chip surface and neural network. Journal of Intelligent Manufacturing, 11, 507514.

[126] G.P. Zhang. “Time series forecasting using a hybrid ARIMA and neural network model”, Neurocomputing 50 (2003), pages 159-175.

[127] N. Gorjian, L. Ma, M. Mittinty, P. Yarlagadda, and Y. Sun. Review on Degradation Models in Reliability Analysis, Proceedings of the 4th World Congress on Engineering Asset Management, 2009, 28-30 Sept, Athens, Greece.

[128] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak. “Health assessment and life prediction of cutting tools based on support vector regression”. Journal of Intelligent Manufacturing, article published online 19 April 2013, DOI: 10.1007/s10845-013-0774-6

[129] Tianyi Wang; Jianbo Yu; Siegel, D.; Lee, J. “A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems.” IEEE International Conference on Prognostics and Health Management, volume 1, number 6, pages 6-9, October 2008. DOI: 10.1109/PHM.2008.4711421.

[130] Emmanuel Ramasso, MichÃ¨le Rombaut, and Noureddine Zerhouni. “Joint Prediction of Continuous and Discrete States in Time-Series Based on Belief Functions.” IEEE Transactions on Cybernetics, voume 43, number 1, pages 37-50, Feb. 2013. DOI: 10.1109/TSMCB.2012.2198882.

[131] Penha, R.L., t. H., and B.R. Upadhyaya. (2002) “Application of hybrid modeling for monitoring heat exchangers”. Third meeting of the Americas - America’s Nuclear Energy Symposium, Miami, FL.

[132] Baraldi, P., Zio, E., Mangili, F., Gola, G., Nystad, B. H., et al. (2011) “An hybrid ensemble based approach for process parameter estimation in offshore oil platforms”. In Proceedings of EHPG Meeting, pp. 1-11.

[133] Jouin, M., Gouriveau, R., Hissel, D., Pra, M.-C., and Zerhouni, N. (2014) “Prognostics of PEM fuel cell in a particle filtering framework.” International Journal of Hydrogen Energy, Volume 39, Issue 1, Pages 481-494. DOI: 10.1016/j.ijhydene.2013.10.054

[134] Cheng, S., and Pecht, M. (2009) “A fusion prognostics method for remaining useful life prediction of electronic products”. EEE International Conference on Automation Science and Engineering, pp. 102-107.

[135] J.Z. Sikorska, M. Hodkiewicz, L. Ma. “Prognostic modelling options for remaining useful life estimation by industry.” Mechanical Systems and Signal Processing, volume 25, issue 5, pages 1803-1836, July 2011. DOI: dx.doi.org/10.1016/j.ymssp.2005.09.012.

[136] Nathan Crutchfield and James Roughton, Chapter 12 - Developing the Job Hazard Analysis, In Safety Culture, edited by Nathan Crutchfield and James Roughton, Butterworth-Heinemann, Oxford, 2014, Pages 235-248, ISBN 9780123964960, http://dx.doi.org/10.1016/B978-0-12-396496-0.00012-4.

[137] Popovic, V., Vasic, B., (2008) Review of Hazard Analysis Methods and Their Basic Characteristics, FME Transactions, 36(4), 181-187

[138] Harold R. Booher. (2003) "Handbook of Human Systems Integration", ISBN: 978-0-471-02053-0.

[139] Gupta, P., 1997. Application of DOW's Fire and Explosion Index Hazard Classification Guide to Process Plants in the Developing Countries, Journal of Loss Prevention in the Process Industries 10, p. 7.

[140] American Institute of Chemical Engineers, 1994. Dow's Fire and Explosion Index Hazards Classification Guide. 7th Edition, American Institute of Chemical Engineers, New York.

[141] Ulrich Hauptmanns, A decision-making framework for protecting process plants from flooding based on fault tree analysis, Reliability Engineering and System Safety, Volume 95, Issue 9, September 2010, Pages 970-980, DOI: http://dx.doi.org/10.1016/j.ress.2010.04.008.

[142] Simon French, Tim Bedford, Simon J.T. Pollard, Emma Soane, Human reliability analysis: A critique and review for managers, Safety Science, Volume 49, Issue 6, July 2011, Pages 753-763, DOI: http://dx.doi.org/10.1016/j.ssci.2011.02.008.

[143] James J. Rooney, Joe H. Turner, John S. Arendt, A preliminary hazards analysis of a fluid catalytic cracking unit complex, Journal of Loss Prevention in the Process Industries, Volume 1, Issue 2, April 1988, Pages 96-103, DOI: http://dx.doi.org/10.1016/0950-4230(88)80019-6.

[144] Jordi Dunjo, Vasilis Fthenakis, Juan A. Vilchez, Josep Arnaldos, Hazard and operability (HAZOP) analysis. A literature review, Journal of Hazardous Materials, Volume 173, Issues 1-3, 15 January 2010, Pages 19-32, DOI: http://dx.doi.org/10.1016/j.jhazmat.2009.08.076.

[145] Hu-Chen Liu, Long Liu, Nan Liu, Risk evaluation approaches in failure mode and effects analysis: A literature review, Expert Systems with Applications, Volume 40, Issue 2, 1 February 2013, Pages 828-838, DOI: http://dx.doi.org/10.1016/j.eswa.2012.08.010.

[146] Philip M. Myers, Layer of Protection Analysis - Quantifying human performance in initiating events and independent protection layers, Journal of Loss Prevention in the Process Industries, Volume 26, Issue 3, May 2013, Pages 534-546, DOI: http://dx.doi.org/10.1016/j.jlp.2012.07.003.

[147] Carol A. Wallace, Lynda Holyoak, Susan C. Powell, Fiona C. Dykes, HACCP - The difficulty with Hazard Analysis, Food Control, Volume 35, Issue 1, January 2014, Pages 233-240, DOI: http://dx.doi.org/10.1016/j.foodcont.2013.07.012.

[148] Cheng, Shunfeng, Michael H. Azarian, and Michael G. Pecht. (2010) "Sensor systems for prognostics and health management." Sensors vloume 10, Issue 6.

[149] Keith Worden, Wieslaw J. Staszewski, James J. Hensman, "Natural computing for mechanical systems research: A tutorial overview", Mechanical Systems and Signal Processing, Volume 25, Issue 1, January 2011, Pages 4-111, DOI: http://dx.doi.org/10.1016/j.ymssp.2010.07.013.

[150] Allison, Paul D. "Missing data". Vol. 136. Sage publications, 2001.

[151] Rubin, Donald B."Inference and missing data." Biometrika 63.3 (1976): 581-592.

[152] Alan L. Yuille, Peter W. Hallinan, David S. Cohen (1992) " Feature extraction from faces using deformable templates" International Journal of Computer Vision, Volume 8, Issue 2, pp 99-111, DOI: 10.1007/BF00127169.

[153] Oivind Due Trier, Anil K. Jain, Torfinn Taxt, (1996) "Feature extraction methods for character recognition - A survey" Pattern Recognition, Volume 29, Issue 4, Pages 641-662, DOI: http://dx.doi.org/10.1016/0031-3203(95)00118-2.

[154] JING LIN, LIANGSHENG QU, "FEATURE EXTRACTION BASED ON MORLET WAVELET AND ITS APPLICATION FOR MECHANICAL FAULT DIAGNOSIS" Journal of Sound and Vibration, Volume 234, Issue 1, 29 June 2000, Pages 135-148, DOI: http://dx.doi.org/10.1006/jsvi.2000.2864.

[155] Anderson, T. W. (2011) " The statistical analysis of time series" Vol. 19, John Wiley and Sons. Chicago

[156] Milewski, Robert, Pawet Malinowski, Anna Justyna Milewska, Piotr Ziniewicz, and Stawomir Wotczynski. (2010) "The usage of margin-based feature selection algorithm in IVF ICSI/ET data analysis." Studies in Logic, Grammar and Rhetoric 21, no. 34 pages: 35-46.

[157] Fikri Ali Mosallam, Ahmed. "Self-organized Selection of Features for Unsupervised On-board Fault Detection." Master's thesis, University of Örebro, 2010.

[158] Jianning Liang, Su Yang, Adam Winstanley, "Invariant optimal feature selection: A distance discriminant and feature ranking based solution", Pattern Recognition, Volume 41, Issue 5, May 2008, Pages 1429-1439, DOI: http://dx.doi.org/10.1016/j.patcog.2007.10.018.

[159] Isabelle Guyon and Andre Elisseeff. "An introduction to variable and feature selection". Journal of machine learning research, March 2003, Pages 1157-1182

[160] Isabelle Guyon, Steve Gunn, Masoud Nikravesh, Lofti A. Zadeh. "Feature Extraction: Foundations and Applications (Studies in Fuzziness and Soft Computing)". Springer, 1 edition, August 2006.

[161] Jolliffe, I. T. (1986). “Principal Component Analysis”, Springer Verlag, pa 487, DOI:10.1007/b98835.

[162] Scholkopf, B., Smola, A., Muller, K. R. (1997). “Kernel principal component analysis” In Artificial Neural Networks - ICANN’97, pa. 583-588. Springer Berlin Heidelberg.

[163] I. H. Witten and E. Frank (2005) “Data Mining: Practical Machine Learning Tools and Techniques”, Morgan Kaufmann, Amsterdam.

[164] Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2005). “The elements of statistical learning: data mining, inference and prediction”. The Mathematical Intelligencer, 27(2), 83-85.

[165] T. Kohonen, Self-Organizing Maps, Springer Series in Information Science, 1995.

[166] Kreyszig, Erwin. (2007) “Advanced engineering mathematics”. John Wiley & Sons.

[167] Allen, J.B., (1977) “Short term spectral analysis, synthesis, and modification by discrete Fourier transform,” IEEE Transactions on Acoustics, Speech and Signal Processing, vol.25, no.3, pp.235-238. DOI: 10.1109/TASSP.1977.1162950

[168] Liang, H.; Nartimo, I, (1998) “A feature extraction algorithm based on wavelet packet decomposition for heart sound signals,” International Symposium on Time-Frequency and Time-Scale Analysis. vol., no., pp. 93-96. DOI: 10.1109/TFSA.1998.721369.

[169] Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H. (1998) “The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis.” Proceedings of the Royal Society of London Series A. Mathematical, Physical and Engineering Sciences Pages 903-995.

[170] S. Salvador, P. Chan, (2004) “Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms,” IEEE International Conference on Tools with Artificial Intelligence, vol., no., pp. 576-584, DOI: 10.1109/ICTAI.2004.50.

[171] A. Mosallam, S. Byttner, M. Svensson, T. Rognvaldsson, (2011) “Nonlinear Relation Mining for Maintenance Prediction,” IEEE Aerospace Conference, pp.1-9, DOI: 10.1109/AERO.2011.5747581

[172] Zio, E., Baraldi, P., & Gola, G. (2008). “Feature-based classifier ensembles for diagnosing multiple faults in rotating machinery”. Applied Soft Computing, 8(4), 1365-1380.

[173] Baraldi, P., Razavi-Far, R., & Zio, E. (2011). “Bagged ensemble of Fuzzy C-Means classifiers for nuclear transient identification”. Annals of Nuclear Energy, 38(5), 1161-1171.

[174] Bishop C. Pattern recognition and machine learning. Springer: 2006.

[175] Piero Baraldi, Francesca Mangili, Enrico Zio, (2015) "A prognostics approach to nuclear component degradation modeling based on Gaussian Process Regression", Progress in Nuclear Energy, Volume 78, January 2015, Pages 141-154, IDOI: http://dx.doi.org/10.1016/j.pnucene.2014.08.006.

[176] Carl Edward Rasmussen and Chris Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.

[177] S. Thrun, W. Burgard, D. Fox, Probabilistic Robotics, MIT Press, Cambridge, MA, 2005.

Résumé :

La construction de modèles de pronostic nécessite la compréhension du processus de dégradation des composants critiques surveillés afin d'estimer correctement leurs durées de fonctionnement avant défaillance. Un processus de dégradation peut être modélisé en utilisant des modèles de connaissance issus des lois de la physique. Cependant, cette approche nécessite des compétences pluridisciplinaires et des moyens expérimentaux importants pour la validation des modèles générés, ce qui n'est pas toujours facile à mettre en place en pratique. Une des alternatives consiste à apprendre le modèle de dégradation à partir de données issues de capteurs installés sur le système. On parle alors d'approche guidée par des données.

Dans cette thèse, nous proposons une approche de pronostic guidée par des données. Elle vise à estimer à tout instant l'état de santé du composant physique et prédire sa durée de fonctionnement avant défaillance. Cette approche repose sur deux phases, une phase hors ligne et une phase en ligne. Dans la phase hors ligne, on cherche à sélectionner, parmi l'ensemble des des signaux fournis par les capteurs, ceux qui contiennent le plus d'information sur la dégradation. Cela est réalisé en utilisant un algorithme de sélection non supervisé développé dans la thèse. Ensuite, les signaux sélectionnés sont utilisés pour construire différents indicateurs de santé représentant les différents historiques de données (un historique par composant). Dans la phase en ligne, l'approche développée permet d'estimer l'état de santé du composant test en faisant appel au filtre Bayésien discret. Elle permet également de calculer la durée de fonctionnement avant défaillance du composant en utilisant le classifieur k-plus proches voisins (k-NN) et le processus de Gauss pour la régression. La durée de fonctionnement avant défaillance est alors obtenue en comparant l'indicateur de santé courant aux indicateurs de santé appris hors ligne.

L'approche développée a été vérifiée sur des données expérimentales issues de la plateforme PRONOSTIA sur les roulements ainsi que sur des données fournies par le Prognostic Center of Excellence de la NASA sur les batteries et les turboréacteurs.

Mots-clés : Pronostic de défaillances guidé par des données, Durée de fonctionnement avant défaillance, Traitement de données, Construction d'indicateurs de santé, Filtre bayésien discret, Processus de Gauss pour la régression.

Abstract:

Constructing prognostics models rely upon understanding the degradation process of the monitored critical components to correctly estimate the remaining useful life (RUL). Traditionally, a degradation process is represented in the form of physical or experts models. Such models require extensive experimentation and verification that are not always feasible in practice. Another approach that builds up knowledge about the system degradation over time from component sensor data is known as data driven. Data driven models require that sufficient historical data have been collected.

In this work, a two phases data driven method for RUL prediction is presented. In the offline phase, the proposed method builds on finding variables that contain information about the degradation behavior using unsupervised variable selection method. Different health indicators (HI) are constructed from the selected variables, which represent the degradation as a function of time, and saved in the offline database as reference models. In the online phase, the method estimates the degradation state using discrete Bayesian filter. The method finally finds the most similar offline health indicator, to the online one, using k-nearest neighbors (k-NN) classifier and Gaussian process regression (GPR) to use it as a RUL estimator. The method is verified using PRONOSTIA bearing as well as battery and turbofan engine degradation data acquired from NASA data repository. The results show the effectiveness of the method in predicting the RUL.

Keywords: Data-driven prognostics, Remaining Useful Life, Data processing, Health indicators construction, Discrete Bayes filter, Gaussian process regression

■ École doctorale SPIM 16 route de Gray F - 25030 Besançon cedex

■ tél. +33 (0)3 81 66 66 02 ■ **ed-spim@univ-fcomte.fr** ■ www.ed-spim.univ-fcomte.fr

UFC UNIVERSITÉ DE FRANCHE-COMTÉ