SPIM

Thèse de Doctorat

école doctorale sciences pour l'ingénieur et microtechniques

UNIVERSITÉ DE FRANCHE-COMTÉ

A Robust & Reliable Data-driven Prognostics Approach Based on Extreme Learning Machine and Fuzzy Clustering

Kamran JAVED

SPIM

Thèse de Doctorat

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

UNIVERSITÉ DE FRANCHE-COMTÉ

THÈSE présentée par

Kamran JAVED

pour obtenir le

Grade de Docteur de
l'Université de Franche-Comté

Spécialité : **Automatique**

A Robust & Reliable Data-driven Prognostics Approach Based on Extreme Learning Machine and Fuzzy Clustering

Unité de Recherche : FEMTO-ST, UMR CNRS 6174

Soutenue le 9 Avril, 2014 devant le Jury :

Louise **TRAVÉ-MASSUYÈS**	Présidente	Directeur de Recherche, LAAS CNRS, Toulouse (Fr)
Said **RECHAK**	Rapporteur	Prof., Ecole Nationale Polytechnique d'Alger, Algérie
Enrico **ZIO**	Rapporteur	Prof., Politecnico di Milano, Italie
Antoine **GRALL**	Examinateur	Prof., Université de Technologie de Troyes (Fr)
Janan **ZAYTOON**	Examinateur	Prof., Université de Reims (Fr)
Rafael **GOURIVEAU**	Co-Directeur de thèse	Maitre de Conférences, ENSMM, Besançon (Fr)
Noureddine **ZERHOUNI**	Co-Directeur de thèse	Prof., ENSMM, Besançon (Fr)

To my loving parents Javed and Shaheen, for everything I am now.
To my respectable uncle Qazi Mohammad Farooq for his support, who has been a constant source of inspiration.

Abstract

Prognostics and Health Management (PHM) aims at extending the life cycle of engineering assets, while reducing exploitation and maintenance costs. For this reason, prognostics is considered as a key process with future capabilities. Indeed, accurate estimates of the Remaining Useful Life (RUL) of an equipment enable defining further plan of actions to increase safety, minimize downtime, ensure mission completion and efficient production.
Recent advances show that data-driven approaches (mainly based on machine learning methods) are increasingly applied for fault prognostics. They can be seen as black-box models that learn system behavior directly from Condition Monitoring (CM) data, use that knowledge to infer its current state and predict future progression of failure. However, approximating the behavior of critical machinery is a challenging task that can result in poor prognostics. As for understanding, some issues of data-driven prognostics modeling are highlighted as follows. 1) How to effectively process raw monitoring data to obtain suitable features that clearly reflect evolution of degradation? 2) How to discriminate degradation states and define failure criteria (that can vary from case to case)? 3) How to be sure that learned-models will be robust enough to show steady performance over uncertain inputs that deviate from learned experiences, and to be reliable enough to encounter unknown data (i.e., operating conditions, engineering variations, etc.)? 4) How to achieve ease of application under industrial constraints and requirements? Such issues constitute the problems addressed in this thesis and have led to develop a novel approach beyond conventional methods of data-driven prognostics. The main contributions are as follows.

- The data-processing step is improved by introducing a new approach for features extraction using trigonometric and cumulative functions, where features selection is based on three characteristics, i.e., monotonicity, trendability and predictability. The main idea of this development is to transform raw data into features that improve accuracy of long-term predictions.

- To account for robustness, reliability and applicability issues, a new prediction algorithm is proposed: the Summation Wavelet-Extreme Learning Machine (SW-ELM). SW-ELM ensures good prediction performances while reducing the learning

time. An ensemble of SW-ELM is also proposed to quantify uncertainty and improve accuracy of estimates.

- Prognostics performances are also enhanced thanks to the proposition of a new health assessment algorithm: the Subtractive-Maximum Entropy Fuzzy Clustering (S-MEFC). S-MEFC is an unsupervised classification approach which uses maximum entropy inference to represent uncertainty of unlabeled multidimensional data and can automatically determine the number of states (clusters), i.e., without human assumption.

- The final prognostics model is achieved by integrating SW-ELM and S-MEFC to show evolution of machine degradation with simultaneous predictions and discrete state estimation. This scheme also enables to dynamically set failure thresholds and estimate RUL of monitored machinery.

Developments are validated on real data from three experimental platforms: PRONOSTIA FEMTO-ST (bearings test-bed), CNC SIMTech (machining cutters), C-MAPSS NASA (turbofan engines) and other benchmark data. Due to realistic nature of the proposed RUL estimation strategy, quite promising results are achieved. However, reliability of the prognostics model still needs to be improved which is the main perspective of this work.

Keywords: Prognostics, Data-driven, Extreme learning Machine, Fuzzy Clustering, RUL

Acknowledgements

Special thanks to Almighty for the wisdom & perseverance, that he bestowed upon me throughout my Ph.D.

I am thankful to FEMTO-ST Institute & Université de Franche-Comté Besançon (France) for the Ph.D. position & a unique research environment.

I would like to express my respect & gratitude to my supervisors Assoc. Prof. Rafael Gouriveau & Prof. Noureddine Zerhouni, for their continuous guidance, & valuable inputs to my research.

I would like to thank my thesis reporters: Prof. Said Rechak, Prof. Enrico Zio, & the examiners: Prof. Louise Travé-Massuyès, Prof. Antoine Grall, Prof. Janan Zaytoon for taking time to review my thesis.

I would like to acknowledge for the helpful suggestions from Assoc. Prof. Emmanuel Ramasso, Assoc. Prof. Kamal Medjaher & Dr. Tarak Benkedjouh. I also want to thank my other colleagues, without particular order: Syed Zill-e-Hussnain, Bilal Komati, Baptiste Veron, Nathalie Herr, Ahmed Mosallam, Haithem Skima & many others for the friendly environment & fruitful discussions.

Many thanks to my wife Momna for her support, & my little angel Musa who never let me go bored of my work.

I am grateful to my parents, my uncle, my brothers & other family members who supported me over the years.

Finally, I won't forget my friend Salik, who encouraged me to pursue my research career with this opportunity.

Kamran Javed

Contents

List of Figures

List of Tables

List of Algorithms

Acronyms & Notations

ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Networks
CBM	Condition Based Maintenance
CI	Confidence Intervals
CM	Condition Monitoring
CNC	Computer Numerical Control machine
$CVRMSE$	Coefficient of Variation of the Root Mean Square Error
CWT	Continuous Wavelet Transform
DD	Data-driven approach
DWT	Discrete Wavelet Transform
D_{SE}	Standardized Euclidean Distance
ED	Euclidean Distance
ELM	Extreme Learning Machine
EOL	End Of Life
FCM	Fuzzy C-Means
FFNN	Feed Forward Neural Network
FT	Failure Threshold
Hyb	Hybrid approach
M	Models of prediction
$MAPE$	Mean Average Percent Error
MEFC	Maximum Entropy based Fuzzy Clustering
MEI	Maximum entropy inference
MFE	Mean Forecast Error
MRA	Multi-Resolution Analysis
msp	Multi-steps ahead predictions
NFS	Neuro-Fuzzy System
NW	Nguyen Widrow procedure
PHM	Prognostics and Health Management
POF	Physics of failure

R	Correlation coefficient
RMS	Root Mean Square
RUL	Remaining Useful Life
SC	Subtractive Clustering
SLFN	Single Layer Feed Forward Neural Network
SW-ELM	Summation Wavelet-Extreme Learning Machine
SW-ELME	Summation Wavelet-Extreme Learning Machine Ensemble
S-MEFC	Subtractive-Maximum Entropy Fuzzy Clustering
WT	Wavelet transform

b	Hidden node bias (i.e., SLFN)
c	Number of clusters
C	Constant
$C\acute{F}$	Cumulative Feature
d	Dilate (or scale) daughter wavelet
f	Activation function for hidden nodes
$\bar{f}$	Average output from two different activation functions
F or $\acute{F}$	Feature
H	Hidden layer output matrix
$H^{\dagger}$	Moore-Penrose generalized inverse
H_{avg}	Average of two hidden layer output matrix
ℓ_D	Learning data for prognostics model
$\ddot{M}$	Monotonicity
$\tilde{N}$	Number of hidden nodes
$\hat{o}_j^m$	Predicted output of m^{th} model against the j^{th} input sample
o	Prediction model output (i.e., SLFN)
$\overline{O}$	Averaged output from multiple prediction models
$R2$	Coefficient of determination
s	Translate (or shifts) daughter wavelet
t	Outputs samples
tc	Current time
tD	Time of degradation
tf	Time at which prediction passes the FT
U	Fuzzy partition matrix
v	Cluster center
V	Cluster centers matrix
w	Input-hidden neuron weights
x	Inputs samples
β	Output weight matrix
Ψ	Mother wavelet
μ_{ij}	Membership of the i^{th} data point to the j^{th} centroid
$\sigma^2_{\bar{o}_j}$	Variance
σ	Standard deviation

General introduction

Any machine during its service life is prone to degrade with the passage of time. However, the failure of engineering assets or their critical machinery can not only be a loss in the manufacturing process or timely services to the customers, but can also have major negative outcomes. In some cases, machinery malfunctioning can even result in irreversible damages to the environment or in safety problems, for example, jet crash due to engine failure, rail accident due to bearing failure, etc. This highlights the need to maintain critical machinery before a failure could happen.

Generally, maintenance can be defined as, "the combination of all technical and associated administrative actions intended to retain a system in a state at which it can perform the required function" [66]. With recent advances in modern technology, industrials and researchers are progressing toward enhanced maintenance support systems that aim at improving reliability and availability of critical engineering assets while reducing overall costs. Therefore, the role of maintenance has changed from a "necessary evil" to a "profit contributor" [194]. Also, the trends have grown rapidly and shifted from fail-fix maintenance to predict-prevent maintenance. In this context, one should speak about the evolution in maintenance policies [66], that began from unplanned corrective maintenance practices for non-critical machinery, and then shifted to a planned preventive maintenance (PM). More precisely, the PM is performed by either pre-defined schedule or by performing Condition-Based Maintenance (CBM) on the basis of current state of the machinery. Further, the CBM evolved to predictive maintenance strategy, that is based on forecasts of future evolution of machine degradation. Therefore, upon detection of failure precursors, prognostics [1] becomes a necessary step to anticipate (and predict) the failure of degrading equipment at future times to estimate the Remaining Useful Life (RUL) [226]. Indeed, it is assumed that adequate actions (either maintenance tasks, either load profile changes) must be performed in a timely manner, such that, critical failures that could lead to major breakdowns or huge wastes can be avoided, which enables optimizing machinery service delivery potential during its lifetime. To fulfill such time critical needs, an enhanced application of CBM is through prognostics, and therefore, CBM has evolved into the concept of Prognostics and Health Management.

[1] Note: In literature keywords prognosis, prognostic or prognostics are used, but for the thesis the word "prognostics" is considered.

For modern industry, PHM appears to be a key enabler for improving availability of critical engineering assets while reducing inopportune spending and security risks. Within the framework of PHM, prognostics is considered as a core process for deploying an effective predictive maintenance scheme. Prognostics is also called as the “prediction of a system’s lifetime”, as its primary objective is to intelligently use the monitoring information of an in-service machinery, and to predict its RUL before a failure occurs, given the current machine condition and its past operation profile [105].
Machinery operates in a dynamic environment, where its behavior is often non-linear due to different factors like environmental and operating conditions, temperature, pressure, noise, engineering variance, etc. As a result, the condition monitoring data gathered from machinery are subject to high levels of uncertainty and unpredictability. Also, with lack of knowledge and understanding about complex deterioration processes, predicting behavior of an in-service machinery can be a complicated challenge. Although in the past decade there are several efforts around the world, real prognostics systems to meet industrial challenges are still scarce. This can be due to a highly complex and non-linear operational environment of industrial systems, which makes it hard to establish effective prognostics approaches that are: robust enough to tolerate uncertainty, reliable enough to show acceptable performance under diverse conditions, and applicable enough to fit industrial constraints. As a result, the need for an enhanced prognostics approach can be pointed out.
In this thesis, the developments are achieved following a thorough literature review on three different approaches for prognostics, i.e., physics based, data-driven, and hybrid approaches. This enables identifying the key challenges related to implementation of a prognostics model, i.e., robustness, reliability and applicability. To account for such challenges, a novel approach for prognostics is introduced by applying advanced techniques from data-driven category of prognostics approaches. The overall performances of data-driven prognostics framework are enhanced by focusing on data-processing, health assessment, and behavior prediction steps. Most importantly, as compared to conventional approaches of data-driven prognostics, RUL estimation is achieved by integrating two newly developed rapid machine learning algorithms. The proposed developments also give a new direction to perform prognostics with a data-driven approach.

The structure of the thesis manuscript is organized as follows.

Chapter 1 gives an overview of PHM, and the role of prognostics. Following that a thorough survey on the classification of prognostics approaches is presented including their pros and cons. Different RUL estimation strategies with data-driven approaches are also reviewed, and the challenges of prognostics modeling are defined.

Chapter 2 discusses the importance of data-processing, its impact on prognostics modeling, and on the accuracy of RUL estimates. Therefore, developments are focused on features extraction and selection steps, and aim at obtaining features that clearly reflect machine degradation, and can be easily predicted. In order, to validate the propositions, two PHM case studies are considered: real data of turbofan engines from PHM challenge

2008, and real data of bearings from PHM challenge 2012.

Chapter 3 presents a new rapid learning algorithm, the Summation Wavelet-Extreme Learning Machine (SW-ELM) that enables performing "long-term predictions". For issues related to time series (i.e., approximation, one step-ahead prediction, multi-step ahead prediction) and challenges of prognostics modeling, the performances of SW-ELM are benchmarked with different approaches to show its improvements. An ensemble of SW-ELM is also proposed to quantify uncertainty of the data / modeling phase, and to improve accuracy of estimates. In order, to further validate the proposed prediction algorithm, two PHM case studies are considered: real data of a CNC machine, and real data of bearings from PHM challenge 2012.

Chapter 4 is dedicated to the complete implementation of prognostics model, and its validation. Firstly, a new unsupervised classification algorithm is proposed namely, the Subtractive-Maximum Entropy Fuzzy Clustering (S-MEFC), that enables estimating the health state of the system. Secondly, a novel strategy for RUL estimation is presented by integrating SW-ELM and S-MEFC as a prognostics model. The proposed approach allows tracking the evolution of machine degradation, with simultaneous predictions and discrete state estimation and can dynamically set the failure thresholds. Lastly, to investigate the efficiency of the proposed prognostics approach, a case study on the real data of turbofan engines from PHM challenge 2008 is presented, and the comparison with results from recent publications is also given.

Chapter 5 concludes this research work by summarizing the developments, improvements and limitations. A discussion on the future perspectives is also laid out.

Chapter 1

Enhancing data-driven prognostics

This chapter presents a thorough survey on Prognostics and Health Management literature, importance of prognostics, its issues and uncertainty. Also, a detailed classification of prognostics approaches is presented. All categories of prognostics approaches are assessed upon different criteria to point out their importance. Thereby, recent strategies of RUL estimation are reviewed and the challenges of prognostics modeling are defined. Thanks to that, the problem statement and objectives of this thesis are finally given at the end of the chapter.

1.1 Prognostics and Health Management (PHM)

1.1.1 PHM as an enabling discipline

With aging, machinery or its components are more vulnerable to failures. Availability and maintainability of such machinery are of great concern to ensure smooth functioning and to avoid unwanted situations. Also, the optimization of service and the minimization of life cycle costs / risks require continuous monitoring of deterioration process, and reliable prediction of life time at which machinery will be unable to perform desired functionality. According to [87], for such requirements, the barriers of traditional Condition-Based Maintenance (CBM) for widespread application, identified during a 2002 workshop organized by National Institute of Standards and Technology (USA), are as follows:

1. inability to continually monitor a machine;

2. inability to accurately and reliably predict the Remaining Useful Life (RUL);

3. inability of maintenance systems to learn and identify impending failures and recommend what action should be taken.

These barriers can be further redefined as deficiencies in sensing, prognostics and reasoning. Also, over the last decade, CBM has evolved into the concept of Prognostics and Health Management (PHM) [87], due to its broader scope. Basically, PHM is an emerging engineering discipline which links studies of failure mechanisms (corrosion, fatigue, overload, etc.) and life cycle management [191]. It aims at extending service cycle of an engineering asset, while reducing exploitation and maintenance costs. Mainly, acronym PHM consists of two elements [85, 87, 226].

1. Prognostics refers to a prediction / forecasting / extrapolation process by modeling fault progression, based on current state assessment and future operating conditions;

2. Health Management refers to a decision making capability to intelligently perform maintenance and logistics activities on the basis of diagnostics / prognostics information.

PHM has by and large been accepted by the engineering systems community in general, and the aerospace industry in particular, as the future direction [172]. Among different strategies of maintenance (Appendix A.1), PHM is contemporary maintenance strategy that can facilitate equipment vendors, integrators or operators to dynamically maintain their critical engineering assets. According to [191], in U.S military two significant weapon platforms were designed with a prognostics capability: Joint Strike Fighter Program [85], and the Future Combat Systems Program [22]. As technology is maturing, PHM is also and active research at NASA for their launch vehicles and spacecrafts [149]. In other words, PHM is a key enabler to facilitate different industries to meet their desired goals e.g. process industry, power energy, manufacturing, aviation, automotive, defence, etc. Some of the key benefits of PHM can be highlighted as follows:

- increase availability and reduce operational costs to optimized maintenance;
- improve system safety (predict to prevent negative outcomes);
- improve decision making to prolong life time of a machinery.

PHM makes use of past, present and future information of an equipment in order to assess its degradation, diagnose faults, predict and manage its failures [226]. Considering such activities, PHM is usually described as the combination of 7 layers adapted from Open System Architecture for CBM [119, 140], that all together enable linking failure mechanisms with life management (Fig. 1.1). We can divided these layers into three main phases 1) observe, 2) analyze and 3) act. Details are as summarized in Table 1.1.

TABLE 1.1: 7 layers of PHM cycle

Observe	1. *Data acquisition*: gather useful condition monitoring (CM) data records using digitized sensors. 2. *Data processing*: perform data cleaning, denoising, relevant features extraction and selection.
Analyze	3. *Condition assessment*: assess current condition of monitored machinery, and degradation level. 4. *Daignostics*: perform diagnostics to detect, isolate and identify faults (see Appendix A.2). 5. *Prognostics*: perform prognostics to project current health of degrading machinery onto future to estimate RUL and associate a confidence interval.
Act	6. *Decision support*: (off-line) recommend actions for maintenance / logistic (e.g. service quality), and (on-line) system configuration (safety actions). 7. *Human Machine Interface*: interact with different layers, e.g. prognostics, decision support and display warnings etc.

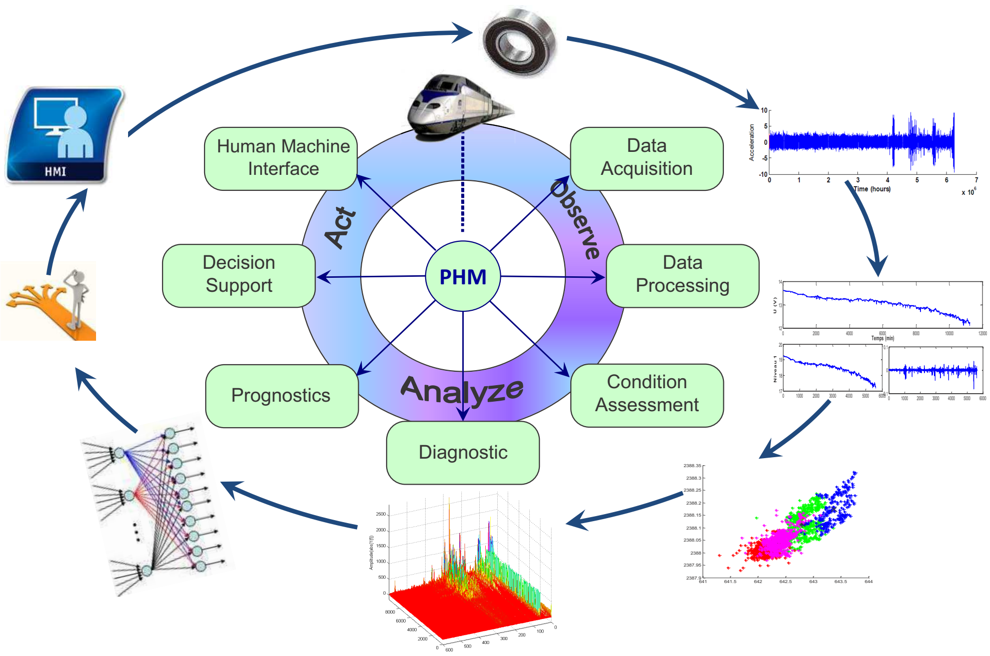

FIGURE 1.1: PHM cycle (adapted from [119])

Within analysis phase, prognostics is considered as a key task with future capabilities, which should be performed efficiently for successful decision support to recommend actions for maintenance [13], or system configuration [14]. Therefore, prognostics is a core process in PHM cycle, to decide plan of actions, increase safety, minimize downtime, ensure mission completion and efficient production [49].

1.1.2 Prognostics and the Remaining Useful Life

The concept of prognostics was initially introduced in the medical domain. Medical prognostics is defined as "the prediction of the future course and the outcome of disease process" [8]. In engineering, prognostics is generally understood as the process of monitoring health of an engineering asset, and of predicting the life time at which it will not perform the required function. According to literature, there are several definitions of prognostics [178], but, few interesting ones are given as follows:

1. 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 [67];

2. predictive diagnostics, which includes determining the remaining life or time span of proper operation of a component [86];

3. the prediction of future health states based on current health assessment, historical trends and projected usage loads on the equipment and / or process [207];

4. failure prognostics involves forecasting of system degradation based on observed system condition [133];

5. the forecast of an asset's remaining operational life, future condition, or risk to completion [83].

A common acceptance can be dressed from above definitions: prognostics can be seen as the "prediction of a system's lifetime" as it is a process whose objective is to predict the Remaining Useful Life before a failure occurs. Also, let retain the definition proposed by the International Organization for Standardization [99]:
"prognostics is the estimation of time to failure and risk for one or more existing and future failure modes".
Note that, the implementation of prognostics for a particular application can be made at different levels of abstraction like: critical components, sub-system or entire system. Also, RUL is expressed by considering units corresponding to the primary measurement of use for overall system. For example, in case of commercial aircrafts the unit is cycles (i.e., related to number of take-offs), for aircraft engines it is hours of operation, for automobiles it is kilometers (or miles). Having said that, whatever the level of abstraction and the unit to define that, prognostics facilitate decision makers with information, that enables them to change operating conditions (e.g. load). Consequently, it may prolong service life of the machinery. In addition, it also benefits the planners to manage upcoming maintenance and to initiate a logistics process, that enables a smooth transition from faulty machinery to fully functional [175].
The main capability of prognostics is to help maintenance staff with insight of future health states of a monitored system. This task is mainly composed of two different steps. The first step is related to the assessment of current health state (i.e., severity or degradation detection), which can also be considered under detection and diagnostics.

Different pattern recognition techniques can be applied to this phase. The second step aims at predicting (or forecasting) degradation trends to estimate the RUL, where time series techniques can be applied [63].
As for illustration of RUL estimation task, consider left part of Fig. 1.2, where for sake of simplicity the degradation is considered as a one-dimensional signal. In such case, the RUL can be computed between current time **tc** after degradation has been detected **tD**, and the time at which predicted signal passes the failure threshold (FT) (assumed or precisely set by an expert), i.e., time **tf**, with some confidence to the prediction. The FT does not necessarily indicate complete failure of the machinery, but a faulty state beyond which there is a risk of functionality loss [171], and end of life (EOL). Therefore, RUL can be defined by Eq. (1.1):

$$RUL = tf - tc \tag{1.1}$$

where **tf** is a random variable of time to failure, and **tc** is the current time.

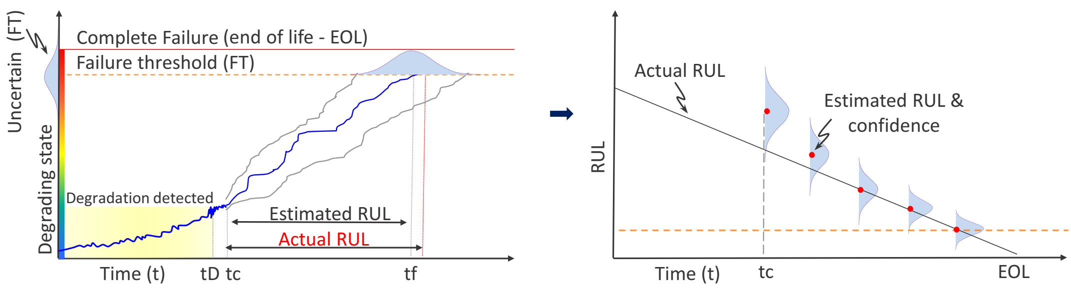

FIGURE 1.2: Illustration of prognostics and RUL estimates

Providing a confidence to predictions is essential for prognostics due to inherent uncertainties of deterioration phenomena, unclear future operating conditions and errors associated to the prognostics model. Therefore, decision making should be based on the bounds of RUL confidence interval rather than single value [178]. Also, narrow confidence intervals represent high precision / accuracy and are preferred over wide confidence intervals that indicate large uncertainty, thus risky decisions should be avoided. Therefore, the combined uncertainty of RUL estimate is not only due to prediction but due to the FT as well (which should be precise).
The right part of Fig. 1.2 shows a situation, where estimated RUL value is updated when new data arrives at each time interval. In this manner, different RULs are estimated with some confidence according to availability of data from monitored machinery. Obviously, accuracy of RUL estimates should increase with time, as more data are available.

1.1.3 Prognostics and uncertainty

Aging of machinery in a dynamic environment is a complex phenomena, and it is often non-linear function of usage, time, and environmental conditions [191]. Following that, real monitoring data from machinery condition (vibration, temperature, humidity, pressure, etc.,) are usually noisy and open to high variability due to direct or indirect impact of usage and environmental conditions related to the degradation of failing machinery. In other words, real-life machinery prognostics is subject to high levels of uncertainty, either due to gathered data or either due to degradation mechanisms. As for example, consider a process, where a component degrades from a healthy state to failure state (see Fig. 1.3). The same component can have different degrading curves due to different failure modes (e.g. in case of bearings inner race, outer race, cage crack), even exposed to same operating conditions. Although, failure modes can be for a same bearing, but still each mode can result different states of degradation that result different RULs. In such situations, RUL estimation becomes a complicated challenge for the prognostics model, that requires timely predicting the future unknown and intelligently assessing the faulty state.

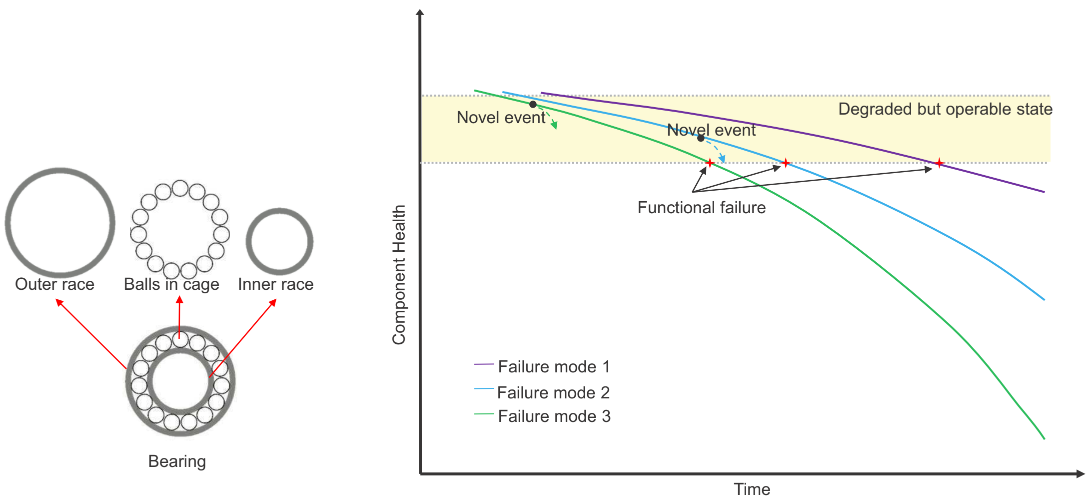

FIGURE 1.3: Component health evolution curves (adapted from [178])

From the above discussions, some key issues of prognostics can be pointed out.

- How to tackle inherent uncertainties of data?
- How to represent uncertainty of machine deterioration process (i.e., from good to faulty state)?
- How to avoid uncertainty of FTs, to enable accurate RUL estimates?

Such issues can be related to different sources of uncertainty [40].

1. Input uncertainties that can be due to initial damage, material properties, manufacturing variability, etc.

2. Model uncertainties that can be due to modeling error related to inaccurate predictions (for data-driven prognostic approaches it can be related to incomplete coverage of data model training [71]). However, such uncertainties can be reduced by improved modeling methods.

3. Measurement uncertainties that are related to data collection, data processing, etc., and can be managed to a better level. For example sensor noise, loss of information due to data processing, etc.

4. Operating environment uncertainties like unforeseen future loads or environments.

Whatever the type of uncertainties in prognostics, they can impact the accuracy of RUL estimates which prevents to recommend timely actions for decision making process or system configuration. In brief, for prognostics system development, three different processes are essential to handle uncertainty [40].

- To represent uncertainty of data, which include common theories like fuzzy set theory, probability theory, etc.;
- To quantify uncertainty related to different sources as correctly as possible;
- To manage the uncertainty by processing the data in an effective manner.

For example consider situation in Fig. 1.4, that illustrates the uncertainty of the RUL estimation by a prognostics model and updates its predictions when new data arrives. Initially the equipment is new and accumulated damage is minor, therefore uncertainty regarding the unknown future can be high. When the damage grows and the failure point is closer, the prognostics model can have much less uncertainty to estimate the exact time to failure. In addition, it is necessary to account for different sources of uncertainties of prognostics, that are associated with "long-term predictions" of machinery health. Without such information a prognostics model has limited use and cannot be integrated

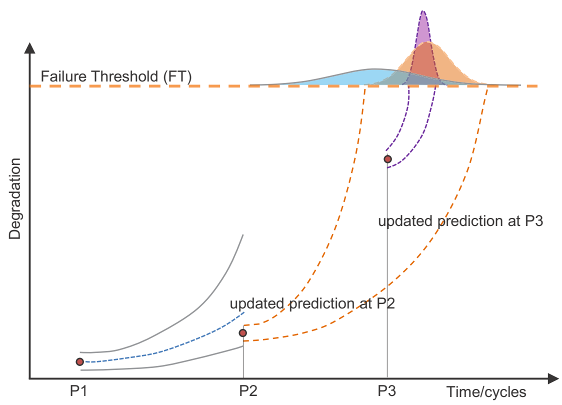

FIGURE 1.4: Uncertainty bands associated with RUL estimations (adapted from [191])

for critical applications. Therefore, in this thesis, uncertainty of prognostics is dealt by: quantifying uncertainty due to data and modeling phase, representing uncertainty using fuzzy set theory, and managing uncertainty by processing data.

1.2 Prognostics approaches

The core process of prognostics is to estimate the RUL of machinery by predicting the future evolution at an early stages of degradation. An accurate RUL estimation enables to run the equipment safely as long as it is healthy, which benefits in terms of additional time to opportunely plan and prepare maintenance interventions for most convenient and inexpensive times [105]. Due to the significance of such aspects, study on PHM has grown rapidly in recent years, where different prognostics approaches are being developed. Several review papers on the classification of prognostics approaches have been published [60, 74, 105, 115, 153, 156, 177, 178, 192, 193]. In spite of divergence in literature, we bring discussions on common grounds, where prognostics approaches are classified into three types: 1) physics based prognostics, 2) data-driven prognostics and 3) hybrid prognostics. But still, these classes are not well addressed in literature, which requires a detailed survey.

1.2.1 Physics based prognostics

1.2.1.1 Overview

The physics based or model based approaches for prognostics use explicit mathematical representation (or white-box model) to formalize physical understanding of a degrading system [155]. RUL estimates with such approaches are achieved on the basis of acquired knowledge of the process that affects normal machine operation and can cause failure. They are based on the principle that failure occurs from fundamental processes: mechanical, electrical, chemical, thermal, radiation [154]. Common approaches of physics based modeling include material level models like spall progression models, crack-growth models or gas path models for turbine engines [84, 178, 191]. To identify potential failure mechanisms, such methods utilize knowledge like loading conditions, geometry, and material properties of a system [153]. To predict the behavior of the system, such methods require detailed knowledge and through understanding of the process and mechanisms that cause failure. In other words, failure criteria are created by using physics of failure (POF) analysis and historic data information about failed equipment [88]. Implementation of physics based approach has to go through number of steps that include, failure modes and effects analysis (FMEA), feature extraction, and RUL estimation [153].
It should be noted that in literature, different works categorize physics based (or model-based) prognostics as POF and system modeling approach [155, 158]. However, they should be limited to POF [192, 197], because system modeling approaches are dependent on data-driven methods to tune parameters of physics based model and should be classified as hybrid approach for prognostics (see section 1.2.3).

1.2.1.2 Application point of view

In general, physics based approaches are application specific. They assume that system behavior can be described analytically and accurately. Physics based methods are suitable for a situation where accuracy outweighs other factors, such as the case of air vehicles [168]. POF models are usually applied at component or material level [197]. However, for most industrial applications physics based methods might not be a proper choice, because fault types can vary from one component to another and are difficult to identify without interrupting operating machinery [84]. In addition, system specific knowledge like material composition, geometry may not be always available [155]. Besides that, future loading conditions also affect fault propagation. Therefore in a dynamic operating environment, the model may not be accurate due to assumptions, errors and uncertainty in the application [197]. In such cases POF models are combined with data-driven methods to update model parameters in an on-line manner, which turns into a hybrid approach (and is discussed in section 1.2.3).

1.2.2 Data-driven prognostics

Data-driven (DD) prognostics approaches can be seen as black box models that learn systems behavior directly from collected condition monitoring (CM) data (e.g. vibration, acoustic signal, force, pressure, temperature, current, voltage, etc). They rely on the assumption that the statistical characteristics of system data are relatively unchanged unless a malfunctioning occurs. Such methods transform raw monitoring data into relevant information and behavioral models (including the degradation) of the system. Therefore, data-driven methods can be low cost models with an advantage of better applicability, as they only require data instead of *prior* knowledge or human experts [110, 111, 155].

According to literature, several studies are performed to categorize data-driven prognostics approaches. [63, 220] classified data-driven methods into machine learning and statistical approaches. [60, 156] classified data-driven approaches as artificial intelligence (AI) techniques and Statistical techniques. A survey on AI approaches for prognostics was presented by [176], where data-driven approaches were categorized as conventional numerical methods and machine learning methods. [158] classified data-driven prognostics methods as, evolutionary, machine learning / AI and state estimation techniques. However, we classify data-driven approaches for prognostics into two categories. 1) machine learning approaches and 2) statistical approaches.

1.2.2.1 Machine learning approaches

Machine learning approaches are branch of AI that attempt to learn by examples and are capable to capture complex relationships among collected data that are hard to describe. Obviously, such methods are suitable for situations where physics based modeling are not favorable to replicate behavior model [111]. Depending on the type of available data, learning can be performed in different ways. Supervised learning can be applied to labeled data, i.e., data are composed of input and the desired output is known. Unsupervised learning is applied to unlabeled data, i.e., learning data are only composed of input and desired output is unknown. Semi-supervised learning that involves both labeled (few data points) and unlabeled data. A partially supervised learning is performed when data have imprecise and / or uncertain soft labels, (i.e., learning data are composed of input and desired outputs are known with soft labels or belief mass [50]). Machine learning is a rapidly growing field in PHM domain, and vast numbers of algorithms are being developed. In brief, machine learning approaches for prognostics can be categorized with some examples as follows.

Connectionist methods - Flexible methods that use examples to learn and infer complex relations among data e.g.:

- Artificial neural networks (ANN) [9, 107, 134];
- Combination of ANN and Fuzzy rules, e.g. Neuro-Fuzzy systems [59, 107].

Bayesian methods - Probabilistic graphical methods mostly used in presence of uncertainty, particularly dynamic Bayesian approaches e.g.:

- Markov Models and variants, e.g. Hidden Markov Models (HMM) [138, 165];
- State estimation approaches, e.g. Kalman Filter, Particle Filter and variants [20, 174, 178].

Instance Based Learning methods (IBL) - Obtain knowledge from stored examples that were presented for learning and utilize this knowledge to find similarity between learned examples and new objects:

- K-nearest neighbor algorithm [142];
- Case-based reasoning for advanced IBL [197].

Combination methods - Can be an effective combination of supervised, unsupervised methods, semi-supervised and partially supervised methods or other possible combinations to overcome drawbacks of an individual data-driven approach, for e.g.:

- Connectionist approach and state estimation techniques [18];
- Connectionist approach and clustering methods [108, 164];
- Ensemble of different approaches to quantify uncertainty and to achieve robust models [17, 20, 89].

Note that, some of the above mentioned categories can also include supervised, unsupervised, semi-supervised learning or partially supervised approaches, however we avoid any strict classification.

1.2.2.2 Statistical approaches

They estimate the RUL by fitting the probabilistic model to the collected data and extrapolate the fitted curve to failure criteria. Statistical approaches are simple to conduct. Like machine learning approaches, statistical methods also require sufficient condition monitoring (CM) data to learn behavior of degrading machinery. However, they can have large errors in case of incomplete data, and the nature of data has therefore its own importance in this category.
[177] presented a state-of-the-art review of statistical approaches, where the taxonomy was mainly based on nature of CM data. From this systematic review, some commonly known prognostics approaches can be: regression based methods, stochastic filtering or state estimation methods like Kalman Filters Particle Filters and variants, Hidden Markov models and variants etc. Further details about this taxonomy are described in [177]. It should be noted that, Bayesian techniques cited just above can also be addressed as machine learning approaches. Other methods in this category can be classical time series prediction methods like Auto-Regressive Moving Average and variants or proportional hazards models [178].

1.2.2.3 Application point of view

In general, the strength of data-driven approaches is their ability to transform high-dimensional noisy data into low-dimensional information for prognostics decisions [60]. However, data-driven methods encounter common criticism that they require more data as compared to physics based approach, which is not surprising. Obviously sufficient quantities of run-to-failure data are necessary for data-driven models to learn and capture complex relations among data. In this context, sufficient quantity means that data has been observed for all fault-modes of interest [191]. However, some industrial systems can not be allowed to run until failure due to their consequences.
Beside that, quantity and quality of data are also important. Indeed, real machinery operates in highly non-linear environment and monitored data could be of high variability, sensor noise, which can impact on performance of data-driven methods. Therefore, it is essential to properly process acquired data in order to obtain good features to reduce modeling complexity, and increase accuracy of RUL estimates.
Machine learning approaches have the advantage that, they can be deployed rapidly and with low implementation cost as compared to physics based methods. In addition, they can provide system-wide scope, where scope of physics based approaches can be limited. Machine learning approach for prognostics could be performed with a connectionist feed forward neural network [107] to predict continuous state of degradation recursively, until it reaches FT. However, such methods provide point predictions and do not furnish any confidence [111]. In comparison, Bayesian methods can be applied to manage uncertainty of prognostics [40], but, again RUL estimates rely on FT. Instance based learning methods do not need FT and can directly estimate the RUL by matching similarity among stored examples and new test instances [142]. They can also be called as experience based approaches [197]. Combination of different machine learning approaches

can be a suitable choice to overcome drawbacks of an individual method [164]. But, whatever the approach considered for building a prognostics model, it is important to involve operating conditions and actual usage environment.
Lastly, statistical approaches for prognostics also require large amount of failure data to implement models. These approaches are economical and require fewer computations as compared to machine learning approaches. For e.g. methods like classical regression are simpler to build. However, they do not consider operating conditions, actual usage environments and failure mechanism [220].

1.2.3 Hybrid prognostics

A hybrid (Hyb) approach is an integration of physics based and data-driven prognostics approaches, that attempts to leverage the strengths from both categories. The main idea is to benefit from both approaches to achieve finely tuned prognostics models that have better capability to manage uncertainty, and can result in more accurate RUL estimates. According to literature, hybrid modeling can be performed in two ways [157]: 1) series approach, and 2) parallel approach.

1.2.3.1 Series approach

In PHM discipline, series approach is also known as system modeling approach that combines physics based approach having *prior* knowledge about the process being modeled, and a data-driven approach that serves as a state estimator of unmeasured process parameters which are hard to model by first principles [160]. Several works in recent literature address series approach as model based prognostics [19, 155, 158, 174]. However it cannot be regarded as model based, because, physics based model is dependant on a data-driven approach to tune its parameters (see Fig. 1.5).

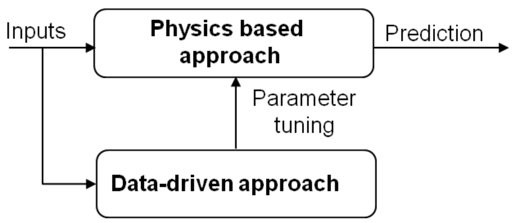

FIGURE 1.5: Series approach for hybrid prognostics model (adapted from [71])

In brief, the representation (or modeling) of an engineering asset is made with mathematical functions or mappings, like differential equations. Statistical estimation methods based on residuals and parity relations (i.e., difference of predictions from a model and system observations) are applied to detect, isolate and predict degradation to estimate the RUL [130, 192]. Practically, even if the model of process is known, RUL estimates might be hard to achieve, where the state of the degrading machinery may not be observable directly or measurements may be affected by noise [71]. In this case, a mathematical model is integrated with on-line parameter estimation methods to infer degrading state and furnish reliable quantification of uncertainty. State estimation techniques can be Bayesian methods like Kalman filter, Particle filter and variant [178], that update the prediction upon collection of new data [57, 58] (see Fig. 1.6).

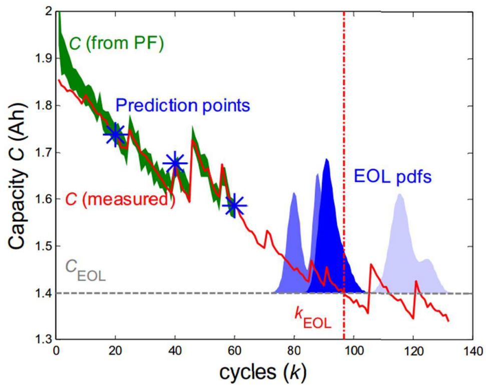

FIGURE 1.6: Battery capacity predictions and new pdf of RUL [169]

As for example from recent literature, [16] developed a physics based model relying on particle filtering to predict the RUL of turbine blades. An approach to RUL estimation of power MOSFETs (metal oxide field effect transistor) was presented by [41], which used an extended Kalman filter and a particle filter to accomplish prognostics models. An unscented Kalman filter based approach was applied for prognostics of PEMFC (polymer electrolyte membrane fuel cell) [221]. Recently, another interesting application on prognostics PEMFC was presented by [112], using particle filter that enabled to include non-observable states into physical models. [33] proposed a particle filter based approach to track spall propagation rate and to update predictions. [10, 11] presented a Matlab based tutorial that combines physical model for crack growth and particle filter, which uses the observed data to identify model parameters. [17] proposed a series approach concerning the prediction of the RUL of a creeping turbine blade.

1.2.3.2 Parallel approach

Physics based approaches make use of system specific knowledge, while data-driven approaches utilize in situ monitoring data for prognostics. Both approaches can have their own limitations and advantages. A parallel combination can benefit from advantages of physics based and data-driven approach, such that the output from resulting hybrid model is more accurate (see Fig. 1.7). According to literature, with parallel approach, the data-driven models are trained to predict the residuals not explained by the first principle model [71, 185].
In PHM discipline still different terminologies are being used for such modeling. [21]

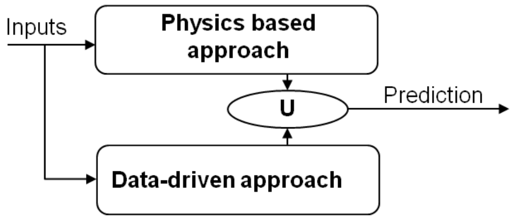

FIGURE 1.7: Parallel approach for hybrid prognostics model (adapted from [71])

called it as parallel hybrid approach, to build a model by combining a data-driven ensemble to POF model for an application of choke valve. In some works, such a combination of physics based and data-driven approaches is also called as fusion prognostics, that also requires an accurate mathematical model to represent a system for POF, and data for the data-driven approach [182]. As for some examples, [44] presented a fusion approach for prognostics of multilayer ceramic capacitors. A fusion methodology for electronics products was proposed by [117]. A case study was performed on computer by considering environmental and operational loads that a system is exposed to throughout its life cycle. [155] presented a road map for information and electronics-rich systems, where the proposed fusion approach was illustrated on an application of printed circuit card assembly. A hybrid approach to fuse outputs from model-based and data-driven approaches was proposed by [82].

1.2.3.3 Application point of view

Series approach for hybrid prognostics requires detailed knowledge of degrading process. However, for the complex systems in a dynamic industrial environment, it's hard to achieve accurate mathematical models. Also, it is important to precisely have FTs to estimate the RUL.
The need for implementation of parallel hybrid prognostics model lies in the limitation of building a prognostics model with an individual approach i.e., data-driven or model-based approach. Therefore, accuracy of parallel approach should be higher. However, implementation of such models include several steps, which can limit their applicability in real-time for some cases, due to computational complexity factor [182]. For example, the main steps to achieve RUL estimates by a parallel hybrid approach can be, parameter identification and monitoring, feature extraction and healthy baseline creation, anomaly detection, parameter isolation, POF models, failure definition, parameter trending and RUL estimation [44]. Finally, parallel hybrid prognostics approach has higher complexity than series hybrid approach.

1.2.4 Synthesis and discussions

1.2.4.1 Proposed classification of prognostics approaches

According to above discussions, prognostics approaches can be broadly categorized as physics based, data-driven, and hybrid approaches (see Fig. 1.8).

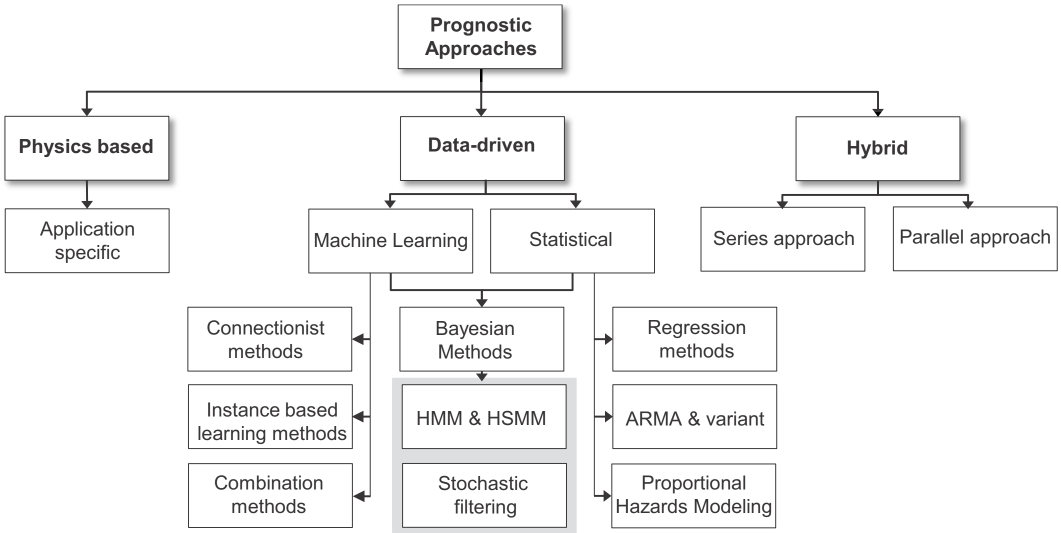

FIGURE 1.8: Classification of prognostics approaches

Among these categories, physics based approaches require modeling POF progression, and can produce precise results. As compared to data-driven methods, they need less data, however, they are component / defect specific [63]. Building an analytical model can be hard or even impossible for higher system level.
On the other hand, data-driven approaches are considered as model free, because they do not require any mathematical formulation of the process. They solely depend on sufficient run-to-failure data. The main strength of data-driven approaches is their ability to transform high dimensional noisy data into lower dimensional information for diagnostic / prognostics decision [60]. They can be a good choice when it is hard to build POF model of complex non-linear system. But, gathering enough CM data is not always possible, and the applicability of data-driven approaches is limited. They are also known as black-box approaches, and are not suitable for applications that require transparency (e.g. credit card, earthquake, etc.) [131].
A hybrid of physics based and data-driven approaches could be a suitable choice to benefit from both categories. This area is also growing rapidly and several works have been published in recent years. Although with hybrid approach, reliability and accuracy of prognostics model is gained significantly [226], but in parallel such methods can have higher computational cost which makes them difficult for some applications.
In spite of several efforts, real prognostics systems are still scarce, because whatever the

prognostics approach either physics based, data-driven or hybrid, they are subject to particular assumptions [178]. In addition, each approach has its own advantages and disadvantages, which limits their applicability. Thereby, for a particular application (either for system level or for component level) a prognostics approach should be selected by considering two important factors: 1) performance and 2) applicability.

1.2.4.2 Usefulness evaluation - criteria

In general, prognostics domain lacks in standardized concepts, and is still evolving to attain certain level of maturity for real industrial applications. To approve a prognostics model for a critical machinery, it is required to evaluate its performances *a priori*, against certain issues that are inherent to uncertainty from various sources. However, there are no universally accepted methods to quantify the benefit of a prognostics method [191], where, the desired set of prognostics metrics is not explicit and less understood. Methods to evaluate the performances of prognostics have acquired significant attention in recent years. From a survey, [171, 173] provided a functional classification of performance metrics, and categorize them into four classes.

1. Algorithm performance: metrics in this category evaluate prognostics model performance by errors obtained from actual and estimated RUL. Selection among competitive models is performed by considering different accuracy and precision criteria, e.g. Mean Absolute Percentage Error (MAPE), standard deviations, etc.

2. Computational performance: metrics in this category highlight the importance of computational performance of prognostics models, especially in case of critical systems that require less computational time for decision making. Therefore, for a particular prognostics approach computational performance can be easily measured by CPU time or elapsed time (or wall-clock time).

3. Cost Benefit Risk: metrics in this category are related to cost benefits that are influenced by accuracy of RUL estimates. Obviously, operational costs can be reduced if RUL estimates are accurate. Because, this will result in replacement of fewer components before the need and also potentially fewer costly failures [171]. For example, metrics in this class can be the ratio of mean time between failure (MTBF) and mean time between unit replacement (MTBUR), return on investment (ROI), etc.

4. Ease of algorithm Certification: metrics in this category are related to the assurance of an algorithm for a particular application (see [171] for details).

From the above classification, Cost Benefit Risk metrics have a broad scope, and obviously it is hard to quantify probable risks that are to be avoided. The Ease of algorithm Certification metrics are related to algorithm performance class, because if the prognostics model error / confidence is not mastered, it can not be certified.
In addition to above classification, a list of off-line metrics is also proposed by [171, 173] to assess prognostics models before they are applied to a real situation. For example,

this includes metrics of: accuracy and precision, prognostics horizon, prediction spread, horizon / prediction ratio, which are again related to algorithm performance class. Therefore, finally this thesis focuses only on metrics from algorithm and computational performance classes for prognostics model evaluation (i.e., class 1 and 2).

As far as applicability of the prognostics approach is concerned, thanks to the discussions on the application point of view for each prognostics approach (in sections 1.2.1.2, 1.2.2.3, 1.2.3.3) one can point out important criteria for applicability assessment:

- requirement of degradation process model;
- failure thresholds;
- generality or scope of the approach;
- learning experience, i.e., run-to-failure data;
- transparency or openness.
- modeling complexity and computation time.

1.2.4.3 Prognostics approach selection

According to above discussions, in order to select a prognostics approach, performance and applicability factors are assessed upon different criteria (based on previous survey). An interval $I = [1, 5]$ is considered to assign weights for each approach according to the given criteria, where 1 represents *min* weight and 5 represents *max* weight. For example in Table 1.2, consider the first entry "without degradation process model" (applicability factor). For this criteria, data-driven approach has *max* weight (i.e., 5), because it does not require any explicit model of degradation process for prognostics. Similarly, physics based approach has been assigned *min* weight (i.e., 1), as it is dependent on mathematical model of degradation process. In this manner, weights for each criteria are carefully given for both factors (i.e., applicability and performance) in Table 1.2, and are further averaged to finally select a particular prognostics approach. For more clarity, a plot of averaged weights in terms of applicability vs. performance is also shown in Fig. 1.9. The assessment clearly shows that data-driven methods have higher applicability as compared to other approaches, but performances need further improvement. Considering the importance of such broader aspects, following topics focus on data-driven approaches, particularly combination methods from machine learning category (section 1.2.2.1). Data-driven prognostics with combination of machine learning methods is a less explored area, but apparently growing rapidly due to its good potential to overcome drawbacks of conventional data-driven approaches. Like other machine learning methods, combination methods can be deployed quickly and cheaply, and can provide system wide coverage. Finally, combination approach can also be suitable to meet most of the criteria related to performance and applicability factors, which are considered in this thesis (i.e., Table 1.2).

TABLE 1.2: Prognostics approach selection

Applicability

Criteria	**POF**	**DD**	**Hyb**
Without degradation process model	1	5	3
Failure threshold	1	4	3
Generality	1	5	2
Learning experiences	4	1	3
Transparency	5	1	4
Complexity / Time	3	4	2
Average weights	2.5	**3.6**	2.8

Performance

Criteria	**POF**	**DD**	**Hyb**
High Accuracy	4	2	5
High Precision	4	1	5
Average weights	4	2	**5**

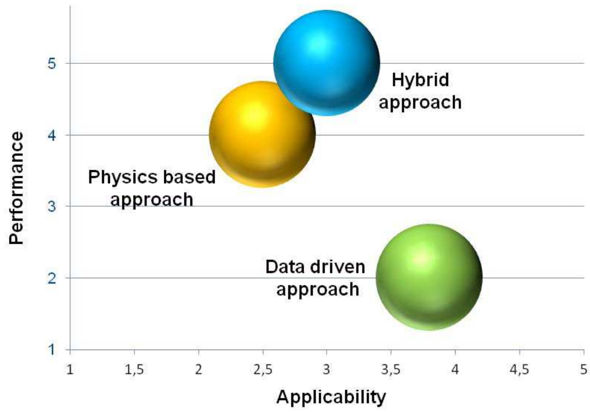

FIGURE 1.9: Prognostics approaches: applicability vs. performance

1.3 Frame of data-driven prognostics

Prognostics cannot be seen as a single maintenance task, the whole aspect of failure analysis and prediction must be viewed as a set of activities that are necessary to be performed. Therefore, when focusing on prognostics process, one can underline a flow that goes from multidimensional data through the RUL of equipment. Generally frame of data-driven prognostics is based on the following necessary steps: 1) data acquisition, 2) data processing (for feature extraction and selection), and 3) prognostics modeling (i.e., to estimate the RUL). Main aspects of each step are discussed as follows.

1.3.1 Data acquisition

Data acquisition is the process of collecting and storing useful data from the system, which can be manipulated by a computer. This is a necessary step to implement PHM program for machinery diagnostic and prognostic. The data collected for PHM can be classified into two types: 1) event data and 2) condition monitoring data [105].

1. Event data: event data records are combination of information on what actually happened (e.g. breakdown, installation, overhaul, etc) and what were the causes and / or and what was done (e.g. preventive maintenance, minor repairs, etc.) for the targeted engineering asset.

2. Condition monitoring (CM) data: CM data are collected thorough a procedure of monitoring parameters of health condition / state of the machinery, in order to clearly identify the changes that can develop faults or can even lead to failures. Such parameters can be vibration, force, temperature, voltage, acoustic, humidity, etc., where various sensors can be applied to collect data for such parameters like accelerometers to measure vibration, dynamometers to measure force, etc.

1.3.2 Data pre-processing

Raw data acquired from the machinery are normally redundant and noisy and cannot be used directly by a prognostics model. In other words, relevant information related to degrading machinery is often hidden in raw data, and should be processed to extract / select features. Also, the effectiveness of a prognostics model is closely related to the quality of features, that can impact uncertainty of prognostics. Besides that, it is also important to identify features that are sensitive to machine condition and clearly reflect failure progression to serve need of modeling task [38, 211]. Feature selection or variable selection is an important process to obtain a subset of useful features that can be used for prognostics model construction.

1.3.3 Prognostics modeling strategies for RUL estimation

The primary objective of prognostics is to build an effective model that is capable of predicting the evolution of degrading features (or variables) under anticipated operating conditions, and can provide RUL of the machinery. However, estimation of RUL is usually hard to perform. In recent years, research to build accurate prognostics models has gained popularity, and there are rapid advances especially with data-driven approaches. According to the discussions, we classify the RUL estimation strategies with a data-driven approach into three groups, 1) univariate degradation based modeling, 2) direct RUL prediction and 3) multivariate degradation based modeling.

1.3.3.1 Univariate degradation based modeling

The univariate degradation based prognostics modeling rely on the prediction of continuous degrading state followed by a failure criteria, where RUL estimate is obtained when degrading signal intersects a pre-defined FT (see Fig. 1.10). In this case, the data-driven approaches learn the model of degradation from time series observations through regression, trending or stochastic process modeling [36, 219]. The main difficulty with this method is to define FTs, which is often very hard to achieve and can restrict the applicability of univariate degradation based approach. Due to this problem, some works in literature assume FTs to estimate the RUL [28, 186]. However, such assumptions are less realistic for prognostics application and should be avoided, because, the uncertainty of FT can lead to poor prognostics and negative outcomes.

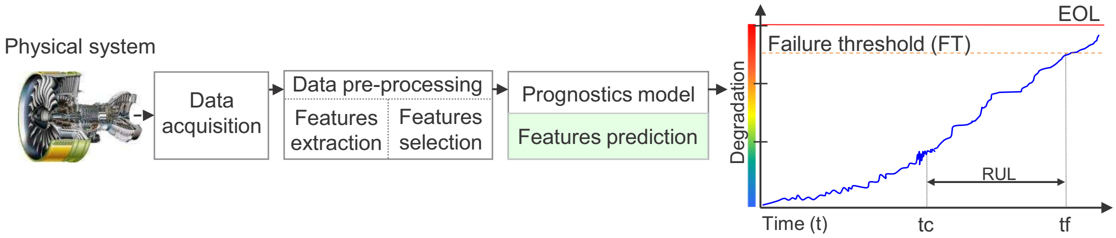

FIGURE 1.10: Univariate degradation based modeling strategy

1.3.3.2 Direct RUL prediction (or similarity based prognostics)

Direct RUL prediction approach learns from the data directly, the relation between observed trends and equipment EOL time to obtain RUL. Therefore, the RUL is derived from data-driven model by applying a pattern matching process (or finding similarity) between the current observation and the knowledge of equipment RUL [71, 199, 227] (see Fig. 1.11). This method does not require definition of failure criteria, but rely on smooth and monotonic trends to facilitate pattern matching [142]. However, even a small deviation from matched history case, either due to uncertain operating conditions or non-linearity due to noise phenomena can lead to large RUL errors. In addition, it is also necessary to have sufficient knowledge on RULs available in training dataset. Lastly, the similarity search procedure can be costly as well [197].

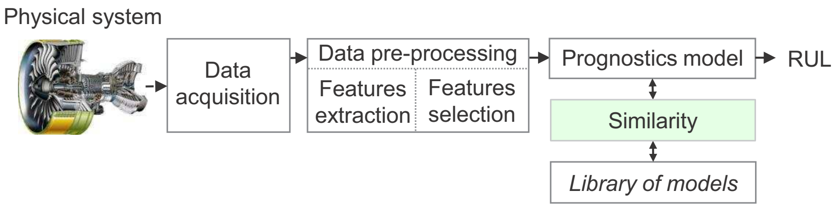

FIGURE 1.11: Direct RUL prediction strategy

1.3.3.3 Multivariate degradation based modeling

With univariate degradation based approach, failure definitions could be a difficult task to achieve. To resolve that issue, a new idea for RUL estimation was proposed by [61]. According to this proposition, prognostics modeling is performed by using a prediction model and a health assessment model to estimate the RUL. The complete illustration of this method was given in [164]. Later, this approach was further developed and named as joint approach for RUL estimation [166]. Mainly, the aim of multivariate degradation

based modeling is to integrate a prediction model and a classification model to precisely set FT. Therefore, a RUL can be achieved by making continuous state predictions of multidimensional features data and simultaneous estimation of discrete states to determine the modes of degradation [109, 166] (see Fig. 1.12).
This approach is new in the domain of data-driven prognostics, and relatively more realistic as compared to former methods and closely aligned with engineering reasoning for prognostics, i.e., with degradation phenomena, fault modes or severity of defect, failure definition, etc. However, the main limitation of this approach is the assumption about the number of degrading states. In fact, each machine could behave differently, even under same condition, which means they can have different states of degradation from a healthy state to a faulty state.
Thus, the number of states cannot be fixed (or same) for all machines in the fleet. As a result FTs can vary and should be assigned dynamically, rather than static FT due to fixed number of states as presented in [166]. Nevertheless, multivariate degradation based modeling shows a strong potential and appears to be more realistic as compared to direct RUL prediction and univariate degradation based modeling approaches. Therefore, in this thesis, multivariate degradation based modeling is considered as a building block for prognostics modeling step, and its further enhanced in the succeeding chapters.

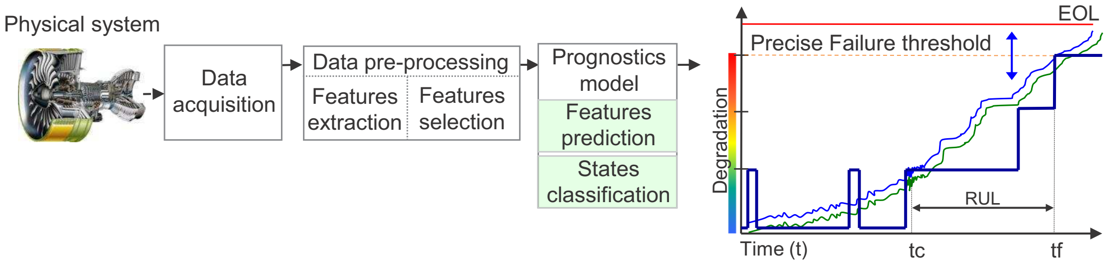

FIGURE 1.12: Multivariate degradation based modeling strategy

1.4 Open issues & problem statement for the thesis

1.4.1 Defining challenges of prognostics modeling

According to literature, various approaches for prognostics exist, i.e., physics based, data-driven and hybrid approaches. However, real prognostics systems to meet industrial challenges are still scarce. This can be due to highly complex and non-linear operational environment of industrial machinery, which makes it hard to establish efficient prognostics approaches, that are robust enough to tolerate uncertainty, and reliable enough to show acceptable performance under diverse conditions. In addition, the applicability of prognostics approaches is also necessary to meet industrial constraints and requirements. Finally, prognostics approaches should be enhanced by handling simultaneously all three challenges, robustness, reliability and applicability, which are still open areas. However, practitioners still encounter difficulties to identify their relationships and to define them. We propose to make it as follows.

1.4.1.1 From data to robustness

Real industrial systems are intrinsically not perfect and the usefulness of gathered data are highly dependent on the variability of phenomena, sensor nonlinearity, etc. Also, the degradation of an engineering asset cannot always be directly measured, so that indirect observations must be imagined. This complicates understanding (and modeling) of complex and uncertain behavior of real systems. Following that, it is obviously difficult to provide a prognostics model that is insensitive to uncertainty, and is capable of capturing dynamics of degrading asset in an accurate manner. Robustness of prognostics models appears to be an important aspect [128], and still remains a critical issue [37].

- Robustness: It can be defined as the "ability of a prognostics approach to be insensitive to inherent variations of input data". It means that, whatever the subset from the entire learning frame is used, the performances of a robust prognostics model should not impair. In other words, a robust model should have steady performance when it is exposed to variations in learning data sets having same context, i.e., operating conditions, geometrical scale, material, coating etc. An illustration is given in Fig. 1.13.

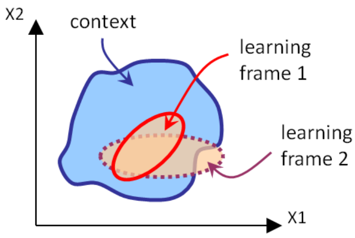

FIGURE 1.13: Illustration of challenge: robustness

1.4.1.2 From robustness to reliability

Even if the prognostics approach appear to be robust to tolerate uncertainty, it should also be reliable enough to be used for the context that is different from the one considered during the modeling phase [126, 136]. In other words, the prognostics should cope with the variations related to the context, such as, multiple operating conditions, geometric scale or materials differences of components, etc. Robustness and reliability [1] of a

[1]Note: classical definition of reliability "the ability of an item to perform a required function under given conditions for a given time interval" [66] is not retained here. Actually, the acception used in this thesis is according to application machine learning approaches in PHM, that do not consider reliability as dependability measure [34].

prognostics approach appear to be closely related [156], and both should be considered as important to ensure the accuracy of RUL estimates.

- Reliability: It can be defined as the "ability of a prognostics approach to be consistent in situations when new / unknown data are presented". It means that, the prognostics model should show consistent performance (in terms of accuracy and precision), in situations when data with different context are presented to the model. In other words, a reliable prognostics model can adapt variations related to context, either in case it is exposed to new data with small deviation from learned cases (i.e., context is partially known), or when totally unknown data with large deviations are presented (i.e., unknown context). An illustration is given in Fig. 1.14.

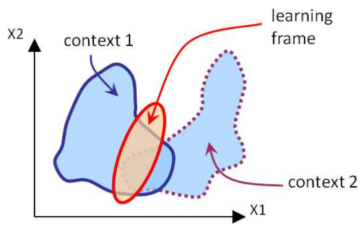

FIGURE 1.14: Illustration of challenge: reliability

1.4.1.3 From reliability to applicability

Besides robustness and reliability criteria, a prognostics model has to be chosen according to the implementation requirements and constraints that restrict the applicability of the approach. Mainly, these constraints can be related to the quality and quantity of data, the generalization capability that is expected, the complexity and computational time required by the model, the human assumptions that clearly impact accuracy of results, etc., [65, 178]. The applicability problem still remains a technical challenge as well.

- Applicability: It can be defined as the "ability of a prognostics approach to be practically applied under industrial constraints". An applicability of a prognostics model can be referred to its suitability or ease of implementation for a particular application, i.e., requirements like failure definition, human intervention, model complexity, computation time, theoretical limits of the approach or any hypothesis.

1.4.2 Toward enhanced data-driven prognostics

1.4.2.1 Issues to be addressed

Although in recent years, a vast number of prognostics methods have been proposed for different applications areas, the progress to build an effective and efficient prognostics approach is still limited. Usually, lack of understating about complex and non-linear behavior of degrading machinery under dynamic operating environment prevents practitioners to develop precise mathematical models for prognostics. In such situations data-driven prognostics approaches can be a good alternate to learn behavior of degrading machinery directly from the data, without any physical understanding about the degradation process. However, performances of prognostics model can suffer due different factors like: inherent uncertainties associated to deterioration process, lack of sufficient quantities of run-to-failure data, sensor noise, unknown environmental and operating conditions, and engineering variations. Obviously in such situations it could be quite hard to precisely infer the exact state of degrading machinery, and to further predict the evolution of degradation from the collected data. In this context, inherent uncertainty of data (e.g. due to variability of phenomena, sensor non-linearity, etc,) can affect robustness of the prognostics model in the learning frame, and also small or large deviations from learned experiences (e.g. due to different context) can affect reliability of the prognostics model in the testing frame. Moreover, the applicability of the prognostics model is equally important from engineering point of view, to meet real-time constraints and requirements for a particular industrial application. According to these discussions, key issues of data-driven prognostics can be summarized as follows.

- How to obtain right information in the form of features that clearly reflect deterioration phenomena?
- How to improve robustness and reliability of prognostics model to handle uncertain inputs and deviations from learned experiences?
- How to improve the applicability of the prognostics model to meet constraints and requirements of a real application?
- How to validate the performances of a prognostics model?

The issues highlighted above confirm that, the need to enhanced data-driven prognostics is inevitable. Therefore, the main assumptions, objective and contributions of this thesis are provided in the following sections.

1.4.2.2 Assumptions

To address identified problems, following assumptions have been made for the proposed prognostics approach.

1. There is at least one way to extract features that are representative of machinery conditions.
2. Collected data should consist of historical run-to-failure observation sequences.
3. It is assumed that degradation has been already detected (see Fig. 1.2).
4. Degrading machinery can not go under self healing.
5. It is assumed that prognostics is achieved at component level.

1.4.2.3 Objective and contributions

According to issues and assumptions, the objective of this thesis is to develop an enhanced data-driven prognostics approach that can estimate the RUL in a realistic manner, with an acceptable level of accuracy under modeling challenges of robustness, reliability and applicability (Fig. 1.15).

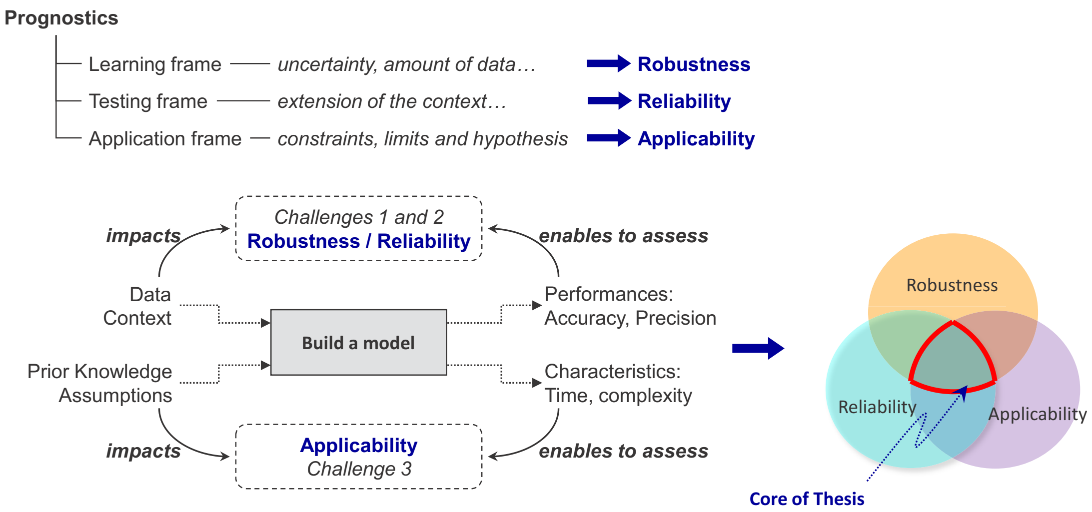

FIGURE 1.15: Thesis key points scheme - toward robust, reliable & applicable prognostics

More precisely, the core of the thesis is to build an efficient prognostics model dealing with all three challenges simultaneously, i.e., robustness, reliability and applicability. Leaving aside data acquisition step, data-driven prognostics framework is enhanced by focusing on data-processing and prognostics modeling steps. According to that, the main contributions of this thesis are as follows.

1. State prognostics modeling challenges (CHAPTER 1).
2. Propose a new data-processing approach to extract / select suitable features (CHAPTER 2).
3. Develop a new algorithm for long-term predictions (CHAPTER 3).
4. Develop a new algorithm to dynamically assess the health states of a degrading machinery. Also, to develop an innovative strategy for RUL estimation that performs simultaneous predictions, discrete state estimation and dynamically assign FTs (CHAPTER 4).
5. Validate all developments on three real data applications (CHAPTER 2, CHAPTER 3 and CHAPTER 4).

1.5 Summary

This chapter presents a thorough survey on PHM literature and its growing importance as a modern strategy to maintain and manage life cycle of critical machinery. Also prognostics process is pointed out as a key enabler to ensure mission achievement of systems while reducing costs and risks. According to that, a detailed classification of prognostics approaches is presented. Data-driven prognostics appear to have better applicability as compared to other approaches, but still lacks in accuracy performances. Following that, open challenges of robustness, reliability and applicability are defined, and the main issues / requirements for data-driven approach are discussed. This enabled to highlight the objective of thesis. Following chapters are the building blocks toward enhanced data-driven prognostics, and focus on "data-processing" step to extract / select suitable features, and on "prognostics modeling" step to improve multivariate degradation based modeling strategy for RUL estimation.

Chapter 2

From raw data to suitable features

This chapter presents data-processing approaches to obtain suitable features, that reduce the uncertainty of prognostics and result in accurate RUL estimates. Following a survey on features extraction and selection, the data pre-treatment approach is presented to extract features having monotonic trends, that clearly reflect evolution of machine degradation. To perform feature selection, recent metrics for feature fitness are considered from the literature. Moreover, a data post-treatment approach is introduced, which emphasize to further reduce multidimensional features data by assessing predictability of features. The underlying idea is that, prognostics should take into account the ability of a practitioner (or its models) to perform "long-term predictions". Thus, proposed method emphasizes on the post selection phase, and aims at showing that it should be performed according to the predictability of features: as there is no interest in retaining features that are hard to be predicted. The proposed developments are validated on real data of turbofan engines from PHM data challenge 2008 and bearings from PHM data challenge 2012.

2.1 Problem addressed

2.1.1 Importance of features for prognostics

According to literature, prognostics may encounter various situations regarding collected information and data from past, present or future behavior. Among different approaches for prognostics (physics based, data-driven, hybrid approaches [84, 105, 178]), data-driven techniques are easier to deploy when its hard to understand first principles of a complex machinery to build a diagnostics or prognostics model [109, 213]. They are black-box models that learn system behavior directly from CM data and use that knowledge to infer its current state and to predict failure progression to estimate RUL.

Generally, modeling of data-driven prognostics (with machine learning methods) has to go through necessary steps of learning and testing, which is dependent on features extracted / selected from raw CM data. In brief, firstly, raw data are collected from machinery and pre-processed to extract useful features to learn the degradation model. Secondly, in the test phase, the learned model is used to predict future behavior and to validate model performance [107]. For example, consider multivariate degradation based modeling strategy from data-driven category of prognostics approaches (Fig. 2.1), which performs prognostics by integrating a prediction model and a classification model to achieve RUL estimate.

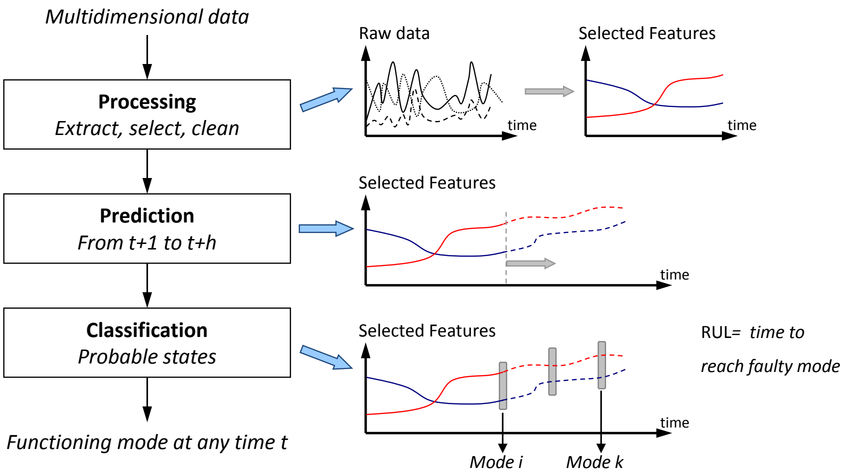

FIGURE 2.1: From data to RUL

The raw sensory data are normally redundant and noisy and cannot be used directly for prognostics. Transforming raw data into features serve need of degradation modeling in prognostics. The extracted features can be seen as multidimensional time series data. However, real industrial systems are intrinsically not "perfect" and usefulness of gathered data are highly dependent on variability phenomena, sensor non-linearity, etc. The form of extracted features can be non-linear, noisy or smooth, etc. Most importantly, the quality of extracted features has a direct affect on the learned prognostics model. In addition, features that properly reflect failure progression or have clear physical meaning can be easily predicted by the prognostics model. Therefore, three main problems can be highlighted.

1. Even if most of data-driven approaches are able to cater non-linearity of degrading signals, features with monotonic behavior are likely to lead to better RUL estimates. Therefore, how to obtain features that clearly reflect failure progression?

2. Some of the classical extracted features like Root Mean Square (RMS) or Kurtosis etc., do not show any variation until a few time before failure [110, 129]. Therefore, how to obtain features that are trendable and correlate to failure progression?

3. There is no way to ensure that the most relevant features (that contain the main information among all features) are those that can be easily predicted, and can result in better RUL estimates. Therefore, how to evaluate that features are predictable over long-term horizon?

Consequently, such situations prevent performing RUL estimates in a timely manner to optimize maintenance actions.

2.1.2 Toward monotonic, trendable and predictable features

The efficacy of a prognostics model is closely related to the form and trends of features, that could affect RUL estimates and result large uncertainty. Although, it could be very challenging to obtain features that explicitly reflect failure progression, nevertheless, it is desired to determine the quality of features prior to prognostics modeling phase. In this context, features for prognostics should have essential characteristics like monotonicity and trendability [48]. For understanding, consider Fig. 2.2, where effects of features with different characteristics are presented.

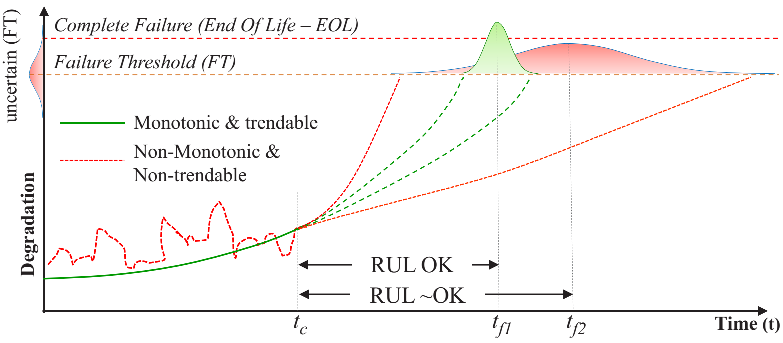

FIGURE 2.2: Effects of features on prognostics

Obviously, feature that clearly reflect failure progression (i.e., monotonic and trendable) up to current time **tc** has high predictability, and can lead to accurate RUL estimate with less uncertainty. Because, predicted fault is expected to grow in a same fashion till it reaches FT at time tf_1. On the other hand, when a non-monotonic and non-trendable feature is presented to a prognostics model, its performances may impair, or rather, it would be impossible for the model to predict future unknown from current time **tc** to tf_2, leading to large uncertainties that risk decision making. Therefore, it is strictly desired to extract features that can not only simplify prognostics modeling, but also lead to accurate RUL estimates. This is the aim of the propositions presented in the following sections.

- Firstly, a survey on data-processing is presented;
- Secondly, data pre-treatment approach is presented to extract / select monotonic and trendable features;

- Lastly, data post-treatment approach is introduced to further reduce multidimensional features data by assessing their predictability.

2.2 Data processing

2.2.1 Feature extraction approaches

Feature extraction is of great importance for machinery diagnostics / prognostics. In literature, a large number of signal processing techniques (or feature extraction methods) have been proposed. *Prior* to any selection among different possibilities, it is required to investigate an appropriate method for a specific application [105, 211]. However, there is huge literature on this topic, which is beyond the scope of the thesis. Let's highlight the three main categories of signal processing techniques which are shown in Fig. 2.3.

2.2.1.1 Time domain

Time-domain feature extraction is directly performed on the sensed waveforms (e.g. acoustic emissions, vibration signal) to identify signatures. In a classical manner time-domain approach extracts characteristic features using statistics like mean, variance, root mean square (RMS), standard deviation, etc. These are suitable in case of stationary signals. Otherwise extracted features may show sensitivity to variation in data and inherit non-linearity which can complicate prognostics modeling and may prevent RUL prediction in a timely manner. Apart from statistical approaches, time-domain analysis can be further categorized into model-based methods, like autoregressive moving average, or signal processing methods like synchronous averaging or correlation [211].

2.2.1.2 Frequency domain

Features extracted from time domain techniques are considered to be suitable for fault detection. However, frequency-domain techniques are considered more effective for fault diagnostic, because, they have good ability to identify and isolate frequency components. The most widely applied technique in this category is Fast Fourier Transform (FFT). Other methods that belong to this category are cepstrum, spectral analysis, higher-order spectra and envelop analysis [105, 211, 214]. The main limitation of frequency-domain techniques is their inability to deal with non-stationary signals, unfortunately which is the case in degrading machinery.

2.2.1.3 Time-frequency

Time-frequency techniques aim at investigating signals in both time and frequency domains. They are considered to be powerful tools to analyze non-stationary signals. Some of the popular time-frequency techniques proposed in literature are: Short Time Fourier Transform (STFT), Wigner-Ville Distribution [32], Wavelet Transform (WT) [47], Empirical Mode Decomposition (EMD) [97], Hilbert-Huang Transform based on EMD and Hilbert spectral analysis [96], spectral kurtosis [12] and cyclostationary analysis [73]. A

brief description of each approach can be found in [105, 211].
According to literature [43], EMD and WT are the two outstanding examples among signal processing techniques since the last two decades. In brief, EMD is a self-adaptive signal processing technique that is suitable for non-linear and non-stationary processes [180]. However, the main weakness of EMD its high sensitivity to noise, and it also runs into the problem of mixing modes [122, 203]. In addition to that, EMD is also reported to have characteristics like wavelet [70], which encourages to use WT as a substitute in studying the behavior of the time-frequency signals [148]. Besides that, EMD is popular in demodulation applications, whereas WT is commonly used in vibration content characterization [43].
On the other hand, WT have also gained attention among researchers / industrials, and is considered as effective approach for analyzing non-stationary signals, especially when they come from rotating machinery like bearings [25, 26, 187, 216]. WT also addresses, the limitations of STFT with fixed window size in time-frequency plane, by adaptively partitioning the time-frequency planes for a range of window sizes to analyze the signal.

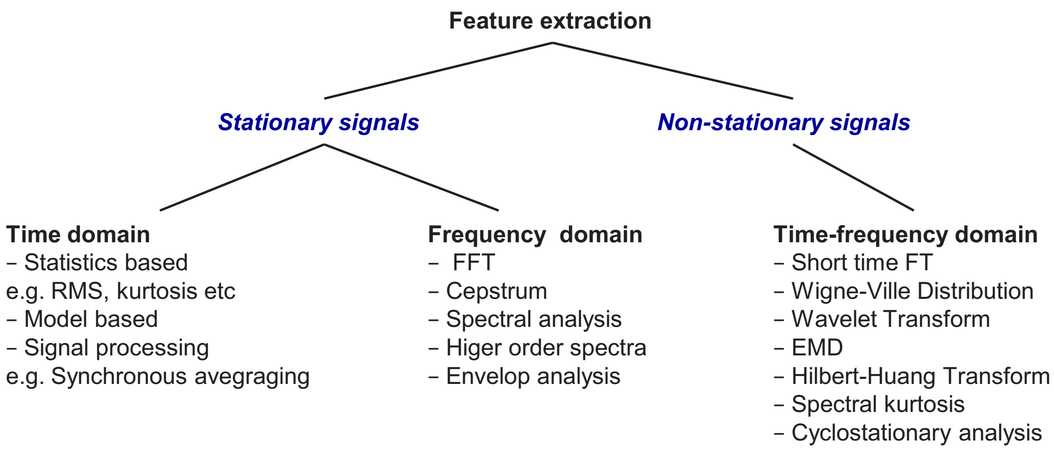

FIGURE 2.3: Feature extraction approaches (adapted from [211])

In this thesis, WT is considered for the analysis of non-stationary vibration / acoustic signals. However, as far as our knowledge, there is no clear rule for vibration signals, but obtaining reliable features is the main requirement [161]. Even the application of WT cannot guarantee ideal features for prognostics, and its performances can vary from case to case, or may be dependent on sensor non-linearity that affect observations of phenomena, manual settings, etc., which needs to be further addressed to extract features, that clearly reflect degradation process.

2.2.2 Feature selection approaches

The main aim of feature selection methodology is to avoid irrelevant or redundant features that do not provide any useful information and are hard to be predicted. Features dimensionality reduction can be performed in two ways:

1. by drawing features in a new space by applying methods like Principal Component Analysis or variants, Singular Value Decomposition, Self-Organizing Map, or clustering methods [28, 141]. However, in this case the resulting features obtained after transformation are different from actual features;

2. by selecting a feature subset based on highest information content, that can significantly improve prognostics. According to recent literature on prognostics, different metrics to characterize suitability of features are proposed for feature selection on the basis of essential characteristics like monotonicity, trendability and prognosability [48].

Leaving aside conventional approaches of feature transformation, recent works confirm that features subset selection by the second case is also vital to prognostics for its efficacy toward accurate RUL estimates [38, 49, 110, 129, 141, 198]. Therefore metrics for feature selection are used in this thesis.

2.3 Proposition of a new data pre-treatment procedure

2.3.1 Outline

PHM approaches derive useful features directly from routinely monitored operational data from machinery. As with change of time, wear in machinery increases and phenomena like excessive temperature, vibration or acoustic emission are observed [116, 132]. Prognostics approaches assume that features are constant until a fault occurs in a component or system. However, real systems operate in dynamic environment, where rapid changes in machinery can take place. Although, feature extraction process transforms redundant CM data into features to show behavior of machinery, one cannot be sure that extracted features properly reflect failure progression and can serve the need for prognostics modeling due to high variability phenomena.
As for example, consider the case of bearings. They are the most common components in rotating machinery, and normally constitute a large percentage of failures in such machines. According to the research survey of Electric Power Research Institute, bearing faults account for the 40% of motor faults [118]. RUL estimation of such components is a challenging task for prognostics approaches, that are based on vibration analysis [198]. However, vibration signals are normally noisy and also subject to variations from a component or a system to another [26, 198]. In such situations, it is quite challenging for signal processing approaches like time domain, frequency domain or time-frequency (wavelet) analysis to furnish vibration based features that have monotonic trends.
The global objective of the proposed approach is to achieve ease in implementing a

prognostics model with improved accuracy. Mainly, in this section, developments on data-preprocessing step are focused on extracting novel features that clearly reflect failure progression in a meaningful way. The methodology is demonstrated on vibration data of bearings, a kind of signal widely used for prognostics, even if RUL estimates are hard to be performed [26, 198]. Nevertheless, the proposed approach is not limited to bearings, and may be applied to other applications that require vibration /acoustic signal processing.

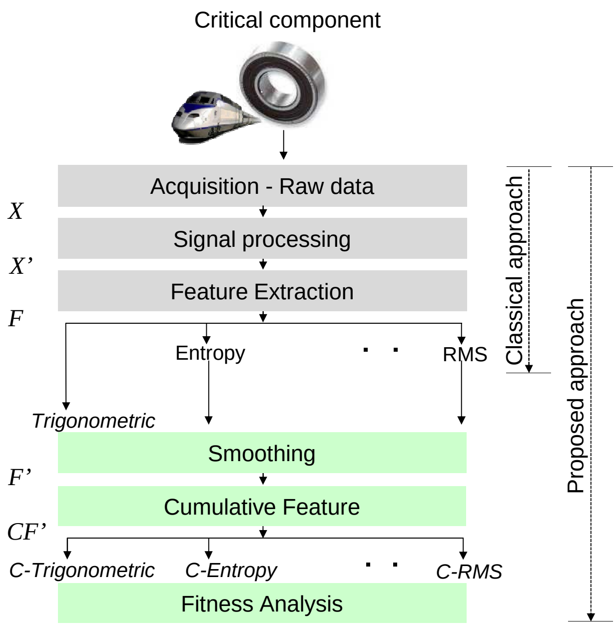

FIGURE 2.4: Proposed approach to obtain monotonic and trendable features

The complete procedure is depicted in Fig. 2.4.

- In the first phase, time-frequency analysis is performed by applying Discrete Wavelet Transform due to its good ability to deal with raw vibration data from rotating machinery. Following that, features are extracted / selected in a new manner: trigonometric functions are firstly applied to extract features, and then smoothed to remove unwanted noisy part.
- The extracted features are further transformed by performing a running total (or point-wise addition) and simultaneous scaling to build a cumulative feature having monotonic behavior and early trend. Note that, strategy of transformation can also be performed on classical features like RMS, skewness, kurtosis, energy, etc.
- Finally, multidimensional data (of extracted features) are analyzed for the fitness.

2.3.2 Steps for feature extraction / selection

As stated before, the approach is applied to vibration signals from monitoring of bearings. The details of each step are described in following topics.

2.3.2.1 Discrete Wavelet Transform and Trigonometric Features

In time-frequency domain, wavelet transform is considered as an effective tool to handle non-stationary signals, as it interprets the signal in time as well as frequency. WT is of two types: Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). CWT has a limit of impracticality with digital computers, where DWT is used for practical reasons, and is mainly achieved by discretization of CWT. An important implementation of DWT is known as Multi-Resolution Analysis (MRA) [135], to analyze the signal at different frequencies with different resolutions (Fig. 2.5a). MRA is accomplished by two functions: scaling and wavelet that are associated with a low pass filter (L_{PF}) and a high-pass filter (H_{PF}) respectively [26]. In brief, consider a discrete signal $x[n]$ that has 512 samples, and frequency span of $(0 - \pi)$ rad/s. At the first level of decomposition, the signal is passed through a L_{PF} that gives approximation coefficients (A), and a H_{PF} that gives detail coefficients (D), followed by a down sampling by 2 (see Fig. 2.5b for an illustration). At this level for $A1$, the entire signal is characterized by only half number of samples (as the output of L_{PF} has 256 points), whereas the frequency resolution is doubled since the spanned frequency band is also halved $(0 - \frac{\pi}{2})$. The frequency band can be further broken down into lower resolutions by recursive decomposition of approximation part at current level [25, 26, 135, 205].

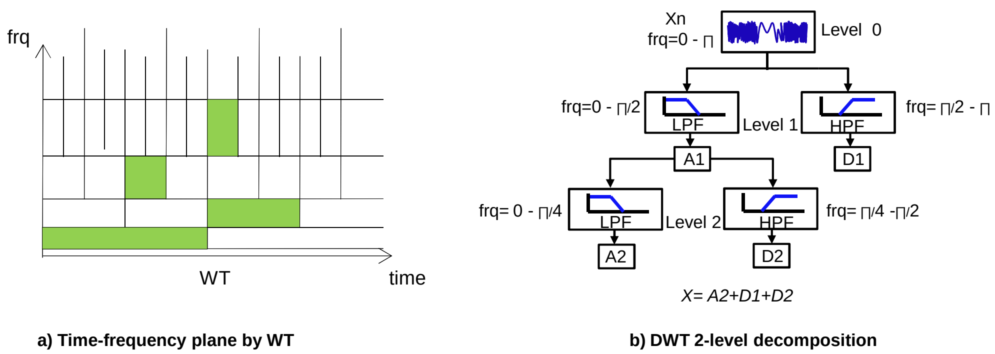

FIGURE 2.5: WT and DWT

Mainly, for applications like vibration signal, MRA requires three types of information:

1. the type of mother wavelet;
2. the level of decomposition;
3. the type of features to be extracted.

In literature, it is suggested that Daubichies wavelet of 4th order (D4) and 3-5 levels of decomposition are suitable for analysis of vibration signals from bearings [26, 56, 187]. However, D4 and 4-level of decomposition are considered for the proposed approach.

Classical features - The classical approach for feature extraction with DWT is performed at a required level of decomposition (and using approximation coefficients) to obtain different features. The extracted features can be entropy [56], Root Mean Square (RMS), variance [42] or other statistical features e.g. skewness, kurtosis, etc. However, each feature can be sensitive to different faults or the severity of degradation. For example, vibration based features RMS and kurtosis from the degraded bearings show variation only few time before failure (i.e., rate of change increases significantly Fig. 2.6), which can limit their use such features for prognostics. Therefore, a new set of features is introduced in the next topic.

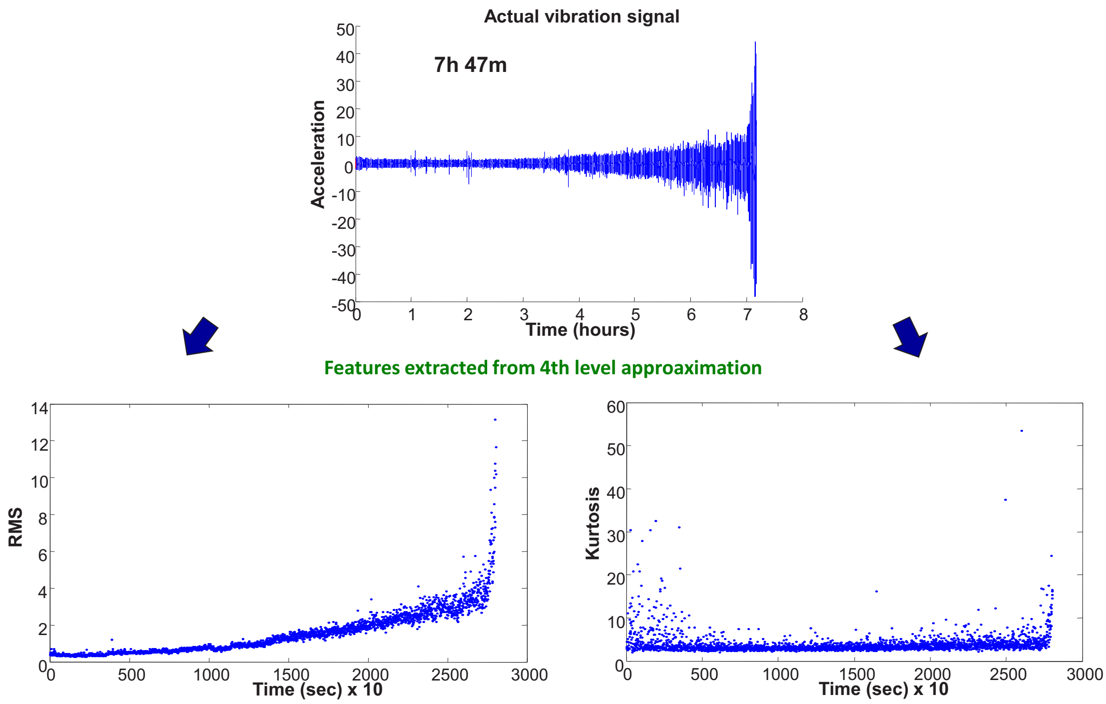

FIGURE 2.6: Classical features extracted from a degraded bearing

Trigonometric features - In this case, at a required level of decomposition (of vibration data), features extraction is performed by using a combination of statistics and trigonometric functions. Mainly, the trigonometric functions can be either monotonically increasing or decreasing, e.g. inverse hyperbolic cosine, inverse hyperbolic sine, inverse hyperbolic secant, etc. In this context, they can be grouped in two classes:

- functions that have domain $(-\infty, \infty)$, e.g. inverse hyperbolic sine (asinh), inverse tangent (atan), etc;
- functions that have different domain value but not $(-\infty, \infty)$, e.g. inverse hyperbolic cosine (acosh), inverse hyperbolic secant (asech), hyperbolic tangent (tanh).

For the second class, input values outside the domain are transformed to complex outputs (i.e., complex numbers), which can be further explored. However, the first class appears to be more relevant to a real data, as it allows domain values from $(-\infty, \infty)$. Therefore, we limit the study to the first class only.
The main benefit of using trigonometric functions is that: such function transform the input data to different scale such that, resulting features have better trends and low scale as compared to classical features. To achieve that, a trigonometric function operates on array (X) element-wise $(x_j,\ j = 1,2,..,n)$ to scale, and a standard deviation (SD) applied to the scaled array for extracting feature value: 1) standard deviation of inverse hyperbolic sin Eq. 2.1, 2) standard deviation of inverse tangent Eq. 2.2.

$$\sigma \left(log \left[x_j + (x_j^2 + 1)^{1/2} \right] \right) \tag{2.1}$$

$$\sigma \left(\frac{i}{2} log \left(\frac{i + x_j}{i - x_j} \right) \right) \tag{2.2}$$

For illustration, an example of features extracted with classical approach from a degraded bearing are presented in Fig. 2.6, and features extracted using Eq. 2.1 and Eq. 2.2 are show in Fig. 2.7. The results clearly depict that, trigonometric features have earlier trends and monotonic behavior, thanks to scaling by trigonometric functions. Finally, different features extracted from 4th level approximation (i.e., terminal nodes, Fig. 2.5b) of the decomposed vibration signal are listed in Table 2.1.

TABLE 2.1: Features extracted from 4th level Approximation

Trigonometric Features (proposed)	
Standard Deviation of inverse hyperbolic sine	$\sigma \left(log \left[x_j + (x_j^2 + 1)^{1/2} \right] \right)$
Standard Deviation of inverse tangent	$\sigma \left(\frac{i}{2} log \left(\frac{i+x_j}{i-x_j} \right) \right)$
Classical Features	
Entropy (threshold)	$E(x) = \sum_j E(x_j)$
Energy	$e = \sum_{j=0}^{n} x_j^2$
Root mean square	$RMS = \sqrt{\frac{1}{n}(x_1^2 + ... + x_n^2)}$
Skewness	$\frac{\sum_{j=1}^{n}(x_j - \bar{X})^3}{(n-1)\sigma^3}$
Kurtosis	$\frac{\sum_{j=1}^{n}(x_j - \bar{X})^4}{(n-1)\sigma^4}$
Upper bound	$max(x) + \frac{1}{2}\frac{max(X) - min(X)}{n-1}$

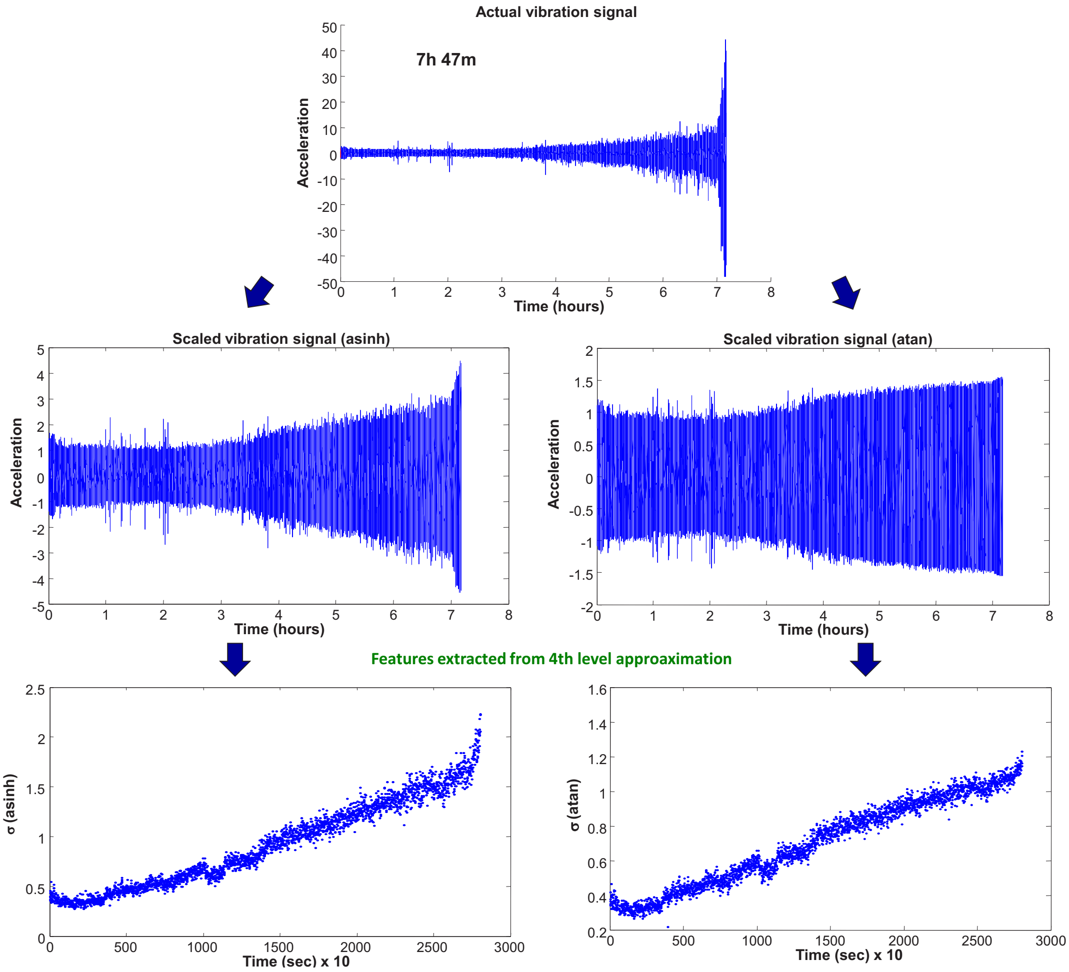

FIGURE 2.7: Trigonometric features extracted from a degraded bearing

2.3.2.2 Feature Smoothing

To reduce variability of extracted features and to filter unwanted noise, smoothing process is performed on each feature F from multidimensional time series to capture important trends. This step is met by applying LOESS filter with span value 0.3 (i.e., 30 %) [137]. In brief, LOESS is a popular smoothing method based on locally weighted regression function and a second degree polynomial. The smoothing operation is considered to be local because, each smoothed value is obtained by data points within a defined neighborhood. Therefore, given some scattered data, LOESS filter can locally perform a weighted fit with the n-nearest points. Further details on LOESS filter can be found in [100].

2.3.2.3 Cumulative features

Any machine is bound to degrade as time grows. In this context, the main idea of this step is obtain features that can clearly reflect failure progression and satisfy requirements mentioned in section 2.1.2. In fact, features with non-linear behavior don't clearly indicate state of the machinery under operation i.e., they don't distinguish among good, degrading or failure states. Note that such non-linearity can also represent self-healing as well. Here we assume that it is not possible to undergo self-recovery, which is the case for components like bearings, gears, or cutting tools. However, this assumption does not hold for batteries that may undergo self-recovery during non-use [49].
Therefore, we propose a straightforward but effective strategy to build features, which aims at transforming an extracted feature into its cumulative form to have monotonicity and trendability characteristics. The transformation task is achieved by applying a cumulative function on a time series, to perform a (point-wise) running total and a scaling operation simultaneously. This procedure results features that can clearly distinguish among different states of degrading machinery.

$$C\acute{F_v} = \frac{\sum_{i=1}^{N} F_v^{'}(i)}{\left| \left(\sum_{i=1}^{N} F_v^{'}(i) \right) \right|^{1/2}}, v = 1,2....,k \tag{2.3}$$

where, $\sum_{i=1}^{N} \acute{F_v}(i)$ represents running total of a "smoothed vth" feature $\acute{F_v}(i)$ up to "N" points, and $C\acute{F_v}$ represents transformed cumulative feature. It should be noted that, the cumulation of a feature can be sensitive to noise, therefore, features smoothing must be performed *a priori* (section 2.3.2.2).

2.3.2.4 Analyzing fitness of features

This step aims at identifying subset of features that have better fitness for further prognostics modeling. The central assumption is that, monotonic and trendable features can lead to more accurate RUL estimates, as compared to features that have contrary behavior. Therefore two simple metrics are considered to assess suitability (or fitness) of a feature: 1) Monotonicity (from literature) 2) Trendability (proposed).

- "Monotonicity" characterizes underlying increasing or decreasing trend [48]:

$$\ddot{M} = \left| \frac{no.\, of\ d/dx > 0}{n-1} - \frac{no.\, of\ d/dx < 0}{n-1} \right| \tag{2.4}$$

  where "n" is for number of observations in a particular feature. Monotonicity $\ddot{M}$ is measured by absolute difference of "positive" and "negative" derivatives for each feature. The value of $\ddot{M}$ can be from 0 to 1: highly monotonic features will have $\ddot{M}$=1 and for non-monotonic features $\ddot{M}$=0.

- "Trendability" is related to the functional form of an extracted feature and its correlation to time. That is, how the state of an engineering asset varies with time. Surely, a constant function (of feature) will have zero correlation with time, where a linear function will have strong correlation. In similar way, correlation can vary with increase in non-linearity (i.e., a non-linear feature will result low correlation). To measure trendability a straight forward formula is given in Eq. (2.5).

$$R = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}} \tag{2.5}$$

where R is the correlation coefficient between two variables x and y. Correlation can be positive, negative, perfect or no correlation. Thus, $R \in [-1;1]$ is the correlation coefficient between feature x and time index y.

2.4 First PHM case study on real data

2.4.1 Bearings datasets of IEEE PHM Challenge 2012

To demonstrate the effectiveness of our contributions, developments are applied to the dataset from IEEE PHM 2012 Data Challenge. Data are mainly composed of run-to-failure vibration signals related to ball bearings from an experimental platform PRONOSTIA [5, 144]. In brief, PRONOSTIA is dedicated to test and validate fault detection, diagnosis and prognostics methods on ball bearings (Fig. 2.8).

FIGURE 2.8: PRONOSTIA testbed - FEMTO-ST Institute, AS2M department

This platform has been designed and realized at AS2M department of FEMTO-ST Institute. It allows to perform accelerated degradations of bearings (until failure) by constant and/or variable operating-conditions, while gathering on-line CM data (load force, rotating speed, vibration - horizontal and vertical - and temperature). Run-to-failure data can be obtained in few hours (see Fig. 2.9), where a degraded bearing could have almost all types of defects: inner race, outer race, rolling element, cage. Note that, the experiments are performed on normally degraded bearings and, assuming no *prior* knowledge about the type of defects. Therefore, data-driven prognostics techniques should be applied.

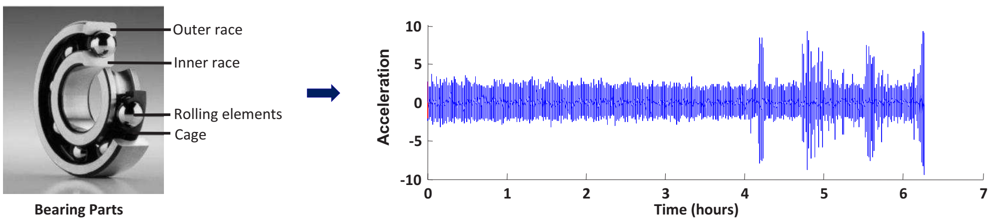

FIGURE 2.9: Bearing degradation: run-to-failure vibration signal

For PHM challenge, data were recorded with specific sampling frequency to capture complete frequency spectrum of degrading bearings. The sampling frequency of acceleration and temperature measures is set to 25.6 kHz and 0.1 Hz respectively. Three loads condition were considered and experiments were stopped when amplitude of the vibration signal overpasses 20g limit. To learn a prognostics model, authors of the experiments provided run-to-failure data from 6 bearings, whereas, for test, incomplete data were given from 11 bearings (see Table 2.2). Further details can be found in [5, 144]. For all the experiments, vibration data from horizontal accelerometer are considered.

TABLE 2.2: Datasets of IEEE PHM 2012 Prognostics challenge

Datasets	**Condition 1** (1800rpm & 4000N)	**Condition 2** (1650rpm & 4200N)	**Condition 3** (1500rpm & 5000N)
Learning set	$Bearing1-1$	$Bearing2-1$	$Bearing3-1$
	$Bearing1-2$	$Bearing2-2$	$Bearing3-2$
Test set	$Bearing1-3$	$Bearing2-3$	$Bearing3-3$
	$Bearing1-4$	$Bearing2-4$	
	$Bearing1-5$	$Bearing2-5$	
	$Bearing1-6$	$Bearing2-6$	
	$Bearing1-7$	$Bearing2-7$	

2.4.2 Feature extraction and selection results

The proposed feature extraction approach is applied on data from all 17 bearings to extract features listed in Table 2.1. For simplicity, bearings are labeled as Ber_{i-j}, where i represents the load condition and j represents the bearing number.

2.4.2.1 Trigonometric features vs. classical features

Here, we compare performances of proposed trigonometric features with classical ones which are listed in Table 2.1, on first bearing Ber_{1-1} (see Fig. 2.10). Vibration data appear to be noisy and have low trendability: the vibration signal is almost constant until 4^{th} hour, but grows suddenly till the end of the life. Results comparison of Fig. 2.10a and 2.10b shows that trigonometric features clearly depict failure progression with high monotonicity and trendability, where classical features depict high level of noise, low monotonicity and trendability, and higher scales.

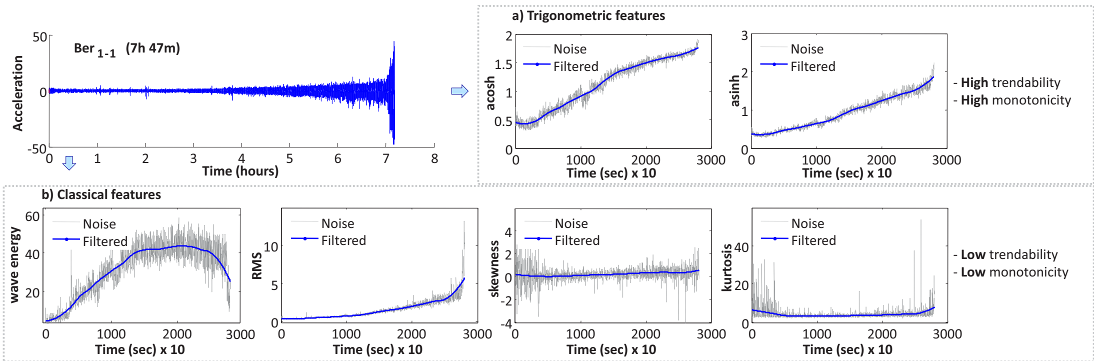

FIGURE 2.10: Trigonometric features and classical features

Back to the accuracy of prognostics problem, one can point out that classical features (RMS, Kurtosis, Skewness, etc.) are not well adapted to catch machine conditions (health states). In addition, they can have large scales, which require normalization before feeding a prognostics model. Therefore, they should be avoided for real application, which strengthens the interest of trigonometric features.

2.4.2.2 Classical features vs. cumulative features

Trigonometric features appear to be suitable as compared to classical features listed in Table 2.1, but such results are achieved for only few bearings like Ber_{1-1}, as depicted in Fig. 2.11. It could be quite challenging to obtain monotonic and trendable features in cases where raw data do not show clear progression of failure. As for an example, consider the bad case of Ber_{1-6} shown in Fig. 2.12, for which it is difficult to obtain good results by all features including trigonometric functions. In this case one can neither perform correct health assessment of bearing condition, nor can accurately predict the RUL, which enables pointing out the interest of cumulative features.
Therefore, consider now the proposition of cumulative transformation, which is an effective and straightforward approach to deal with this problem. In brief, following a smoothing task (by LOESS filter), cumulative features are obtained in simple manner by performing running total on a particular feature and simultaneous scaling operation. For example in Fig. 2.11 or Fig. 2.12, the RMS features extracted from the vibration

signal (from 4th level of decomposition) is transformed to its respective cumulative features C-RMS. Similarly, for both bearing (Ber_{1-1} and Ber_{1-6}) all the extracted features (Table 2.1) are transformed to build respective cumulative features (using Eq. (2.3)).

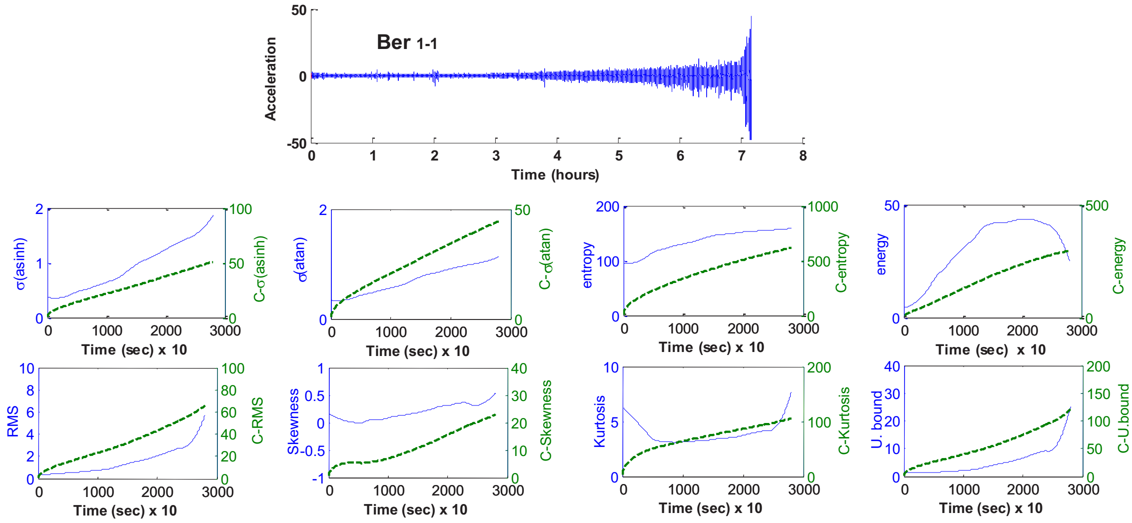

FIGURE 2.11: Extracted features and their respective cumulative features (good case)

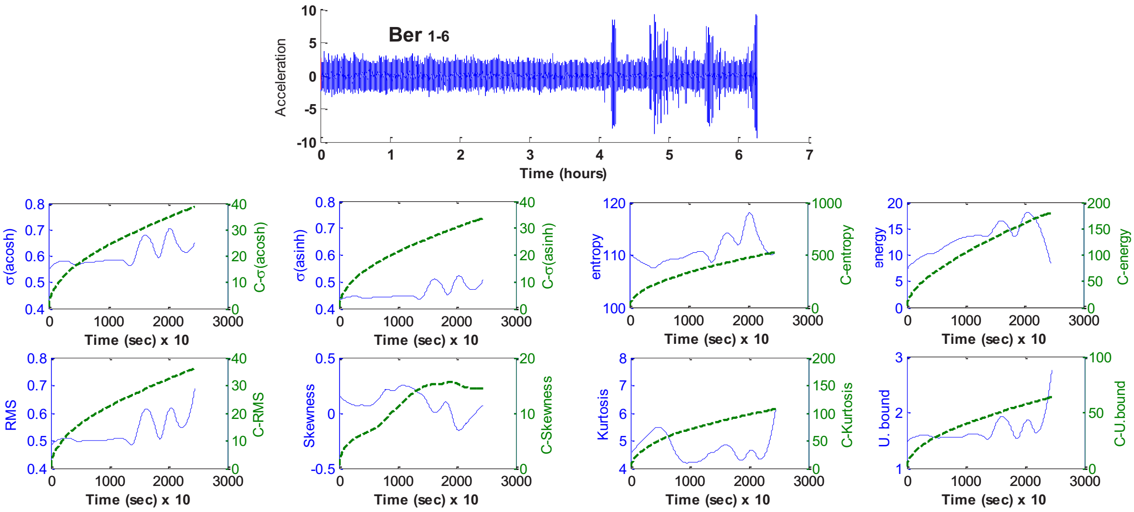

FIGURE 2.12: Extracted features and their respective cumulative features (bad case)

The results from both cases i.e., Ber_{1-1} and Ber_{1-6}, clearly show that the transformed feature have higher monotonicity and trendability than the classical ones. It is also confirmed by analyzing fitness measures $\ddot{M}$ and R (Eq. (2.4) and (2.5)), which are summarized in Table 2.3. The fitness plots from these results are presented in Fig. 2.13 to

TABLE 2.3: Features fitness analysis

Ber_{1-1} Good case

Feature	R	$\ddot{M}$	**Cumulative Feature**	R	$\ddot{M}$
σ(asinh)	0.99	0.90	C-σ(asinh)	0.99	1
σ(atan)	0.99	0.90	C-σ(atan)	0.99	1
Entropy	0.95	0.93	C-Entropy	0.98	1
Energy	0.81	0.43	C-Energy	0.99	1
RMS	0.91	0.90	C-RMS	0.99	1
Skewness	0.92	0.50	C-Skewness	0.97	0.89
Kurtosis	0.13	0.34	C-Kurtosis	0.97	1
Upper bound	0.85	0.79	C-Upper bound	0.99	1

Ber_{1-6} Bad case

Feature	R	$\ddot{M}$	**Cumulative Feature**	R	$\ddot{M}$
σ(asinh)	0.74	0.21	C-σ(asinh)	0.98	1
σ(atan)	0.73	0.19	C-σ(atan)	0.98	1
Entropy	0.70	0.08	C-Entropy	0.98	1
Energy	0.75	0.41	C-Energy	0.99	1
RMS	0.73	0.21	C-RMS	0.98	1
Skewness	-0.51	0.06	C-Skewness	0.92	0,66
Kurtosis	-0.26	0.24	C-Kurtosis	0.97	1
Upper bound	0.64	0.10	C-Upper bound	0.98	1

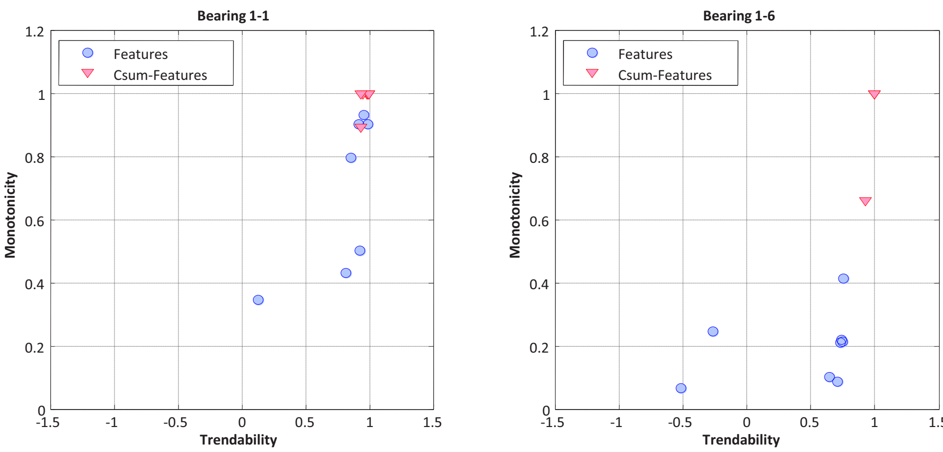

FIGURE 2.13: Fitness plots

further highlight the benefit of cumulative transformation.
Finally, fitness analysis is performed to compare classical features extraction procedure and the proposed approach (Fig. 2.4). Both methods are thoroughly examined on all 17 bearings by assessing trendability R and monotonicity $\ddot{M}$ characteristics of extracted

features. Mean performances by each approach are summarized in Table 2.4. According to results, one can clearly notice that, cumulative features have higher fitness as compared to features obtained from classical procedure. Also, cumulative features based on trigonometric functions (C-σ(atan) and C-σ(asinh) in Table 2.4) appear to be the more monotonic and trendable ones. Same conclusion can be drawn qualitatively by considering Fig. 2.14 that compares the form of features, extracted from classical and proposed approaches.

TABLE 2.4: Comparing features fitness (mean performances on 17 bearings)

Feature	R	$\ddot{M}$	**Cumulative Feature**	R	$\ddot{M}$
σ(asinh)	0.47	0.28	C-σ(asinh)	**0.984**	**1**
σ(atan)	**0.47**	**0,31**	C-σ(atan)	**0.984**	**1**
Entropy	0.47	0.31	C-Entropy	0.983	1
Energy	0.36	0.29	C-Energy	0.982	1
RMS	0.45	0.28	C-RMS	0.978	0.99
Skewness	-0.06	0.11	C-Skewness	-0.5	0.61
Kurtosis	-0.04	0.15	C-Kurtosis	0.972	1
Upper bound	0.39	0.21	C-Upper bound	0.973	0.99

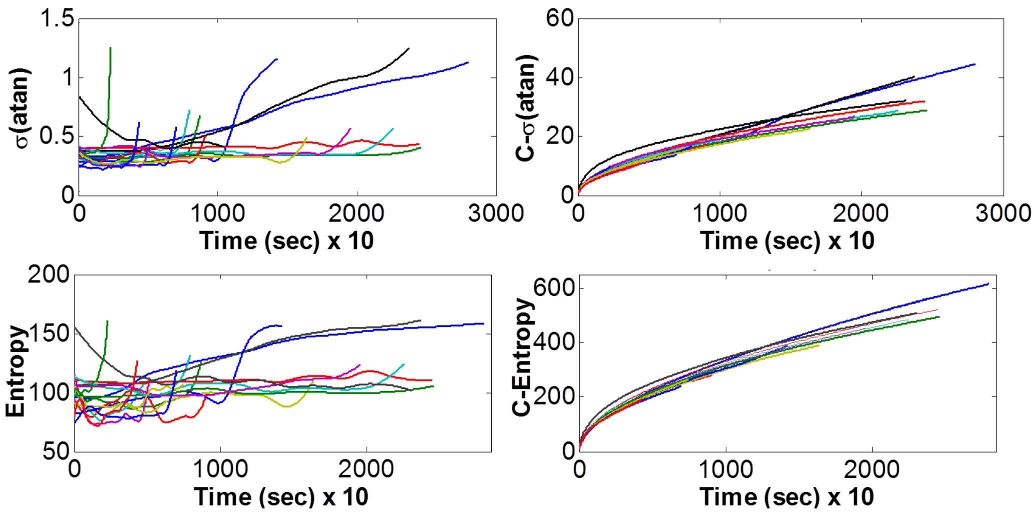

FIGURE 2.14: Comparison of classical and cumulative features on 17 bearings

As a synthesis, whatever the case is (and thereby, whatever load conditions), cumulative transformation lead to highly monotonic and trendable wear patterns. On the opposite, classical procedure result highly non-linear and complicated patterns that do not clearly show machine condition, which can impair performance of the prognostics model to a very low accuracy, and should result in large uncertainty of RUL estimates.

2.5 Predictability based prognostics: underlying concepts

2.5.1 Proposition of a new data post-treatment procedure

As discussed earlier, the standard of machine learning approaches for prognostics is divided into two phases, i.e., learning phase and testing phase section. In classical way prognostics model is learned by set of features that are acquired from a sensor signal. The model must be retrained upon these features, until significant performance level is achieved. This approach can be time expensive because some of the features can be very hard to be predicted. In other words, there is no use in retaining such features that are not predictable. According to existing procedure of data-driven prognostics, critical phase of prediction should be met in appropriate manner for further classification and RUL estimation (see Fig. 2.1), where learning phase of prognostics modeling should consider the important steps of "feature selection" and "prediction modeling" in a simultaneous manner in order to retain or reject features on the basis of predictability. From aforesaid procedure, two issues can be pointed out. Firstly, there is no unique way to select most relevant features that are predictable and contribute for better RUL estimation. Secondly, the predictability should be assessed according to a prediction model as well as the horizon of prediction. The main purpose of this method is to reconsider the learning phase of data-driven approaches by considering both steps "feature selection" and "prognostics modeling" as complementary and closely related.

2.5.1.1 Time series models and accuracy over horizon

In fact, predictability is not a well defined concept. Generally, predictability attributes to the capability in making predictions of future occurrence on the basis of past information. It should depict a goodness of predictions, so that undesirable events can be avoided by making accurate forecasts in a timely manner. Predictions can be incorrect, therefore, it is important to understand their quality in a framework that is dependent on the considered time series. Also, predictability can be affected by different factors that vary from event to event and make modeling process more complicated [201].
According to literature, predictability states for the degree of correctness in forecasting or shows the usefulness of a forecasting method [7]. In this context, it must be considered that up to what extent accurate predictions of a given time series can be provided by an applied modeling approach. Therefore, metrics are required in order to show significance of accurate prediction modeling. Remarkably, there are few works that focus on the predictability aspect by considering modeling accuracy. [201] used seasonally adjusted coefficient of efficiency to evaluate predictability of univariate stream flow process. However, in this study need of a suitable forecasting approach as well as model performance measure is highlighted for a particular domain. [113] presented a quantitative metric to measure time series predictability using genetic programming. [62] provided an improvement of those developments. [183] defined metrics to determine suitable predictors for genomic sequence: quantitative metrics that depict the ability of time series to be predicted by a particular approach. However, they were useful for single step-ahead forecasting methods.

Accuracy of prediction is greatly affected by horizon of prediction. A time series can be well predicted over a short horizon, but difficult to be predicted for a long term horizon. As error grows with increasing horizon, prediction accuracy is reduced, and this denotes low predictability of a time series. In this context, [53] proposed a general measure of predictability to assess relative accuracy over different horizons for macroeconomic application. [7] presented new metrics for predictability that were applied to multi-step ahead predictions of surrogated time series. However, no consensual point of view appears in existing contributions.

2.5.1.2 Defining predictability

As a synthesis, either considering correctness or horizon of prediction, literature points out that predictability is closely related to accuracy of predictions that are evaluated against certain error tolerance. In others words, assessing a prognostics model requires the user to define the limit of prediction he would like to obtain, as well as the performance of prediction which follows from that. Discussions enables to explicitly state that predictability is not only closely related to the type of prognostics model one mean to use, but also to the horizon of prediction that is evaluated as useful (short-term, mid-term, long-term). Also it depends on a limit of accuracy one wants to achieve (see Fig. 2.15). Finally, we define predictability as follows.

- Predictability is the ability of a given time series TS to be predicted with an appropriate modeling tool M that facilitates future outcomes over a specific horizon H and with a desired performance limit L.

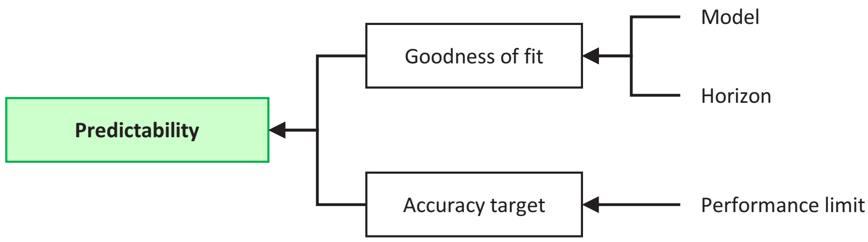

FIGURE 2.15: Compounds of predictability concept

Formally, we propose to formulate it as:

$$Pred\,(TS/M,H,L) = exp^{-\left|ln\left(\frac{1}{2}\right).\frac{MFE^{H}_{TS/M}}{L}\right|} \tag{2.6}$$

where, H states for the horizon of prediction, L is the limit of accuracy that is fixed, and MFE is the mean forecast error in between the actual values of TS and the predicted ones (thanks to M):

$$MFE^{H}_{TS/M} = \frac{1}{H}.\sum_{i=1}^{H} e_{i}= \frac{1}{H}.\sum_{i=1}^{H}\left(M^{i} - TS^{i}\right) \tag{2.7}$$

Perfect value for MFE is 0. $MFE > 0$ indicates under forecast and $MFE < 0$ over forecast. Predictability has an exponential form (Fig. 2.16) and is as higher (maximum=1) as the MFE is lower. A TS can be considered as predictable if its predictability coefficient is in between 0.5 and 1, i.e., if the MFE is in between 0 and the limit value L chosen by the user.

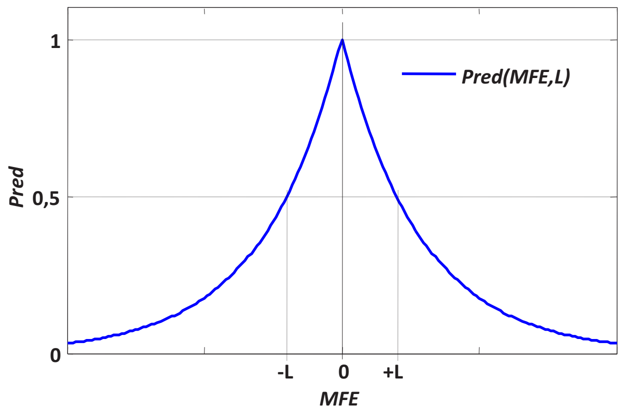

FIGURE 2.16: Illustration of predictability measure

2.5.2 Predictability based feature selection procedure

As stated before, (and illustrated in Fig. 2.1) the prognostics model uses a set of selected features to provide RUL estimation. The learning phase of the model has to be reiterated until suitable prognostics performances are obtained ("try and error approach"). However, this can be a waste of time, because some features can be very hard (even impossible) to be predicted, i.e., since there is no certainty that an accurate prognostics model can be provided. In other words, there is no interest in retaining features that cannot be forecasted in time. Therefore, learning phase of a prognostics model should be extended to the selection of features: not only the user aims to build the model for prognostic, but he also has to define the appropriate set of features that can be more accurately predicted over different horizons. Following that, the "features selection" phase should be performed while building a prognostics model. On this basis, features set obtained from classical data-mining techniques can be further reduced to final set of predictable features in accordance to learned prediction models. Consider Fig. 2.17 for understanding such methodology, where the depicted procedure aims at defining which features are predictable (according to a model and a horizon of prediction). This enables either to retain or reject each potential couple of "feature-model" to be used for prognostics.

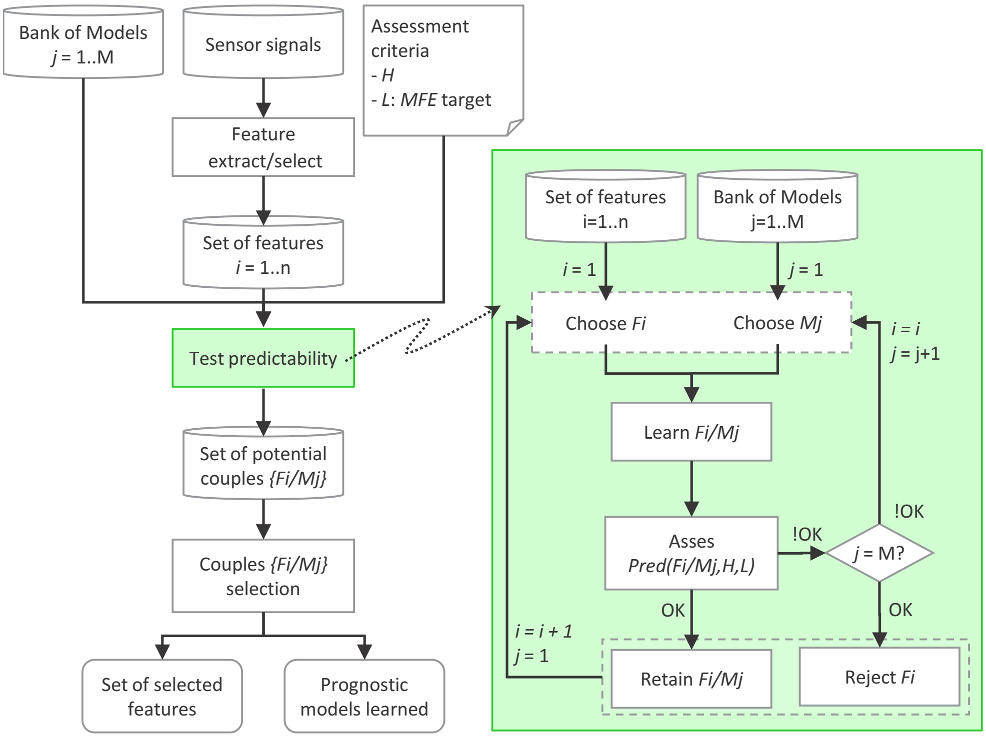

FIGURE 2.17: Selection procedure based on predictability

2.6 Second PHM case study on real data

2.6.1 Outline

2.6.1.1 Turbofan Engines datasets of IEEE PHM Challenge 2008

The proposed developments are applied to the challenge dataset of diagnostics and prognostics of machine faults from first international conference of prognostics and health management (2008). The data were collected from C-MAPSS (Commercial Modular Aero-Propulsion System Simulation), which is a tool for simulating a large commercial turbofan engine (see Fig. 2.18). This challenge dataset consists of multidimensional time series signals (26 variables or features) from different degrading instances and contaminated with measurement noise. Each set of time series comes from a different engine of a same fleet. Each engine starts from different initial conditions and manufacturing conditions are not known to the user. Each engine begins from a normal state but, due to some fault occurrence, starts to degrade. Thus, the fault magnitude increases with time until failure state takes place (see [175] for details). The RUL estimates are in units of time, i.e., hours or cycles. From the dataset among 26 available features 8 were pre-selected, as suggested in previous works by using information theory and Choquet Integral [164]. The selected features from different sensors listed in Table 2.5 (i.e., F1 to F8), are filtered with "moving average" filtering technique to remove noise content (Fig. 2.19), and are used as a starting point to show the impact of proposed "predictability-based" features selection on prognostics (section 2.5.2).

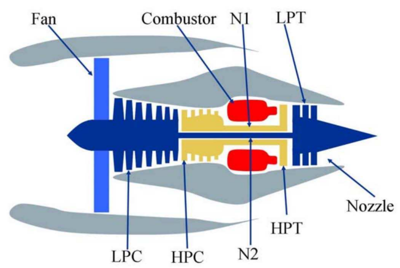

FIGURE 2.18: Engine diagram [72]

Sensor	Description	Unit
2	Total temperature at LPC outlet	$^\circ R$
3	Total temperature at HPC outlet	$^\circ R$
4	Total temperature at LPT outlet	$^\circ R$
8	Physical fan speed	rpm
11	Static pressure at HPC outlet	psia
13	Corrected fan speed	rpm
15	Bypass ratio	–
17	Bleed Enthalpy	–

TABLE 2.5: C-MPASS output features

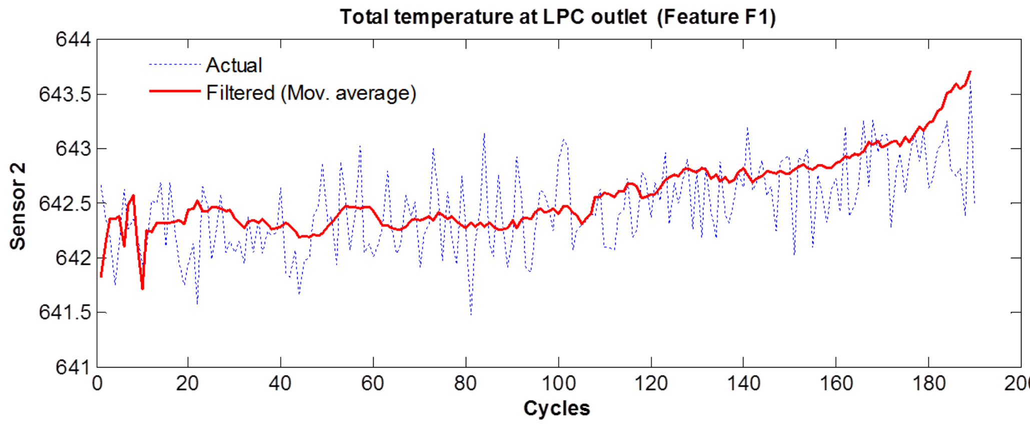

FIGURE 2.19: Filtered feature from engine dataset

2.6.1.2 Simulation settings

Among different RUL estimation methods from data-driven prognostics category, the multivariate degradation based modeling strategy is considered. As discussed in section 1.3.3, this strategy is more realistic for prognostics modeling, which estimates the RUL by simultaneous predictions and health state estimation. Therefore, building a prognostics model requires a predictor to perform long-term predictions of continuous states and an unsupervised classifier to handle unlabeled data (i.e., in the absence of ground truth for FTs) to estimate discrete states. Note that, the aspect of FTs is discussed in Chapter 4. For illustration, the prediction phase is met by two different connectionist approaches, Artificial Neural Network (ANN) and Neuro Fuzzy System (NFS). The unsupervised classification phase is performed by Fuzzy C-Means clustering algorithm (FCM), for details see [29]. Finally, to validate the proposed methodology (i.e., predictability based prognostics), for each test RUL is estimated with whole set of pre-selected features {F1 - F8} and with predictable features. Performances for both cases are compared to show the impact of predictably on RUL estimates. The complete procedure is shown in Fig. 2.20, where the details are provided in the following topics.

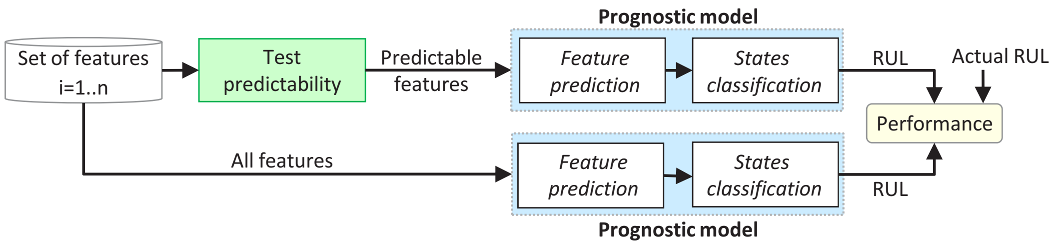

FIGURE 2.20: Procedure to analyze the impact of predictability

2.6.2 Building a prognostics model

2.6.2.1 Prediction phase: choice of feature / model couples

In brief, among different prognostics approaches, data-driven methods have great potential due to their ability to learn from examples and to model non-linear phenomena. Also, when input-output data set are main source of information to develop better understanding of systems current health state [98]. Due to such properties, machine learning methods are of great interest for prognostics. Among these techniques, adaptive networks like ANN and NFS are increasingly used for prediction problems [45, 52, 200]. These connectionist approaches are capable to capture complex relationship among inputs and outputs and have good approximation capability for non-linear modeling of real systems. Moreover, they have shown good performances in prognostics applications as well [64, 196, 210].
The ANN is to be tuned (learning phase) by Levenberg-Marquardt algorithm which is considered to be faster and more effective learning technique as compared to other algorithms [81]. The adaptive neuro-fuzzy inference system (ANFIS) proposed by [104] is also considered, due to its good potential for forecasting in maintenance applications [64]. Each approach has its own benefits as well as limitations, which are not deeply presented here, but interested reader can refer to [80, 123] for more theoretical details.

General formalization - Connectionist systems like ANNs or NFS aim at approximating an input-output function. This kind of systems must be tuned to fit to the studied problem thanks to a learning phase of parameters. Let $[X]$ be the input data set, $[Y]$ the output data set. The approximation function can be formalized as:

$$[\hat{Y}] = f_t([X],[\theta]) \tag{2.8}$$

where $[\hat{Y}]$ states for the estimated output set $[Y]$, and $[\theta]$ for the set of parameters that have to be tuned during the learning phase.
In a similar manner, let's formalize the problem of connectionist-based multi-steps

ahead prediction of a univariate time series (like a feature for prognostic). A univariate time series TS_t is a chronological sequence of values describing a physical observation made at equidistant intervals: $TS_t = \{x_1,x_2,\ldots,x_t\}$ (where t states for the temporal index variable). With these notations, the multi-steps ahead prediction problem consists in estimating a set of future values of the time series: $[\hat{X}_{t+1\to t+H}] = [\hat{x}_{t+1}\ ,\ \hat{x}_{t+2}\ ,\ \hat{x}_{t+3}\ ,\ \ldots\ ,\ \hat{x}_{t+H}]$ where H states for the final prediction horizon. This approximation can be expressed as:

$$[\hat{X}_{t+1\to t+H}] = msp([X_t]) \tag{2.9}$$

where, "msp" states for "multi-steps ahead prediction", and $[X_t] \in TS_t$ is know as the set of regressors used (for example $[X_t] = [x_t\ ,\ x_{t-1}\ ,\ x_{t-2}]$).

Multi-steps ahead predictions with an iterative approach - The msp can be obtained in different ways and by using different connectionist approaches (structure + learning algorithm). [76] dress an overview of those approaches and suggested that iterative approach is a more suitable choice for prognostics.
Mainly, msp are achieved by using a single tool (an ANN or a NFS) that is tuned to perform a one-step ahead prediction $\hat{x}_{t+1}$. This estimated value is used as one of the regressors of the model to estimate the following ones and the operation is repeated until the estimation of $\hat{x}_{t+H}$. The procedure is illustrated in Fig 2.21. Formally:

$$\hat{x}_{t+h} = \begin{cases} \text{if } h=1, f^1\left(x_t,\ldots,x_{t+1-p},[\theta^1]\right) \\ \text{elseif } h\in\{2,\ldots,p\}, \\ f^1\left(\hat{x}_{t+h-1},\ldots,\hat{x}_{t+1},x_t,\ldots,x_{t+h-p},[\theta^1]\right) \\ \text{elseif } h\in\{p+1,\ldots,H\}, \\ f^1\left(\hat{x}_{t+h-1},\ldots,\hat{x}_{t+h-p},[\theta^1]\right) \end{cases} \tag{2.10}$$

where $\left\{f^1,[\theta^1]\right\}$ states for the one-step ahead prediction model (ANN or NFS) with its parameters set calculated during the learning phase, p the number of regressors used, i.e., the number of past discrete values used for prediction. Note that, from the time $h > p$, predictions are made only on evaluated data and not on observed data.

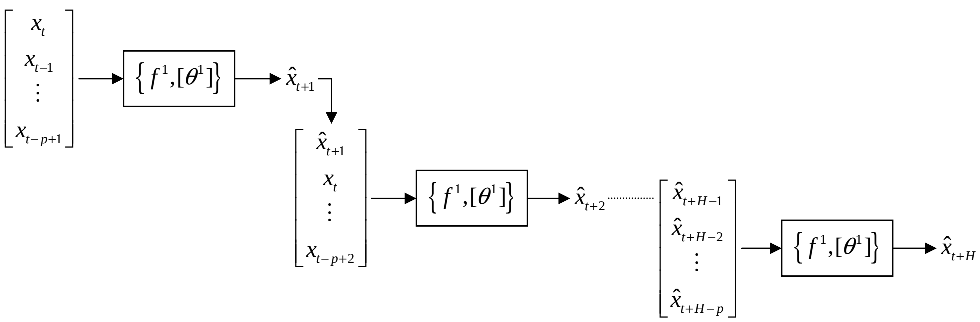

FIGURE 2.21: Iterative model for multi-steps predictions [76]

Application - The *msp* technique has been used to forecast features in time to illustrate the procedure of feature selection based on predictability concept. Experiments are performed by using an ANN and a NFS as potential models for feature predictions, where each model is tuned according to parameters shown in Table 2.6. The training of each model is met by multidimensional time series data of 40 turbofan engines from the challenge data record file "$train - FD001.txt$" (downloaded from [143]), where 5 engines data from the same file are used to perform tests. It should be noted that for both off-line training and on-line testing tasks, the challenge data record file "$train - FD001.txt$" is considered. The *msp* are performed from time $tc = 50$ till the defined prediction horizon ($tc + 134$), using an iterative approach. According to the dimension of feature space, "M" univariate prediction models (either ANN or NFS) are learned in the off-line phase, i.e., feature-model couple. In the on-line phase, the learned models are tested on data from 5 different engines. Prediction results obtained from each test are thoroughly investigated on predictability criteria to retain or reject each couple of potential "feature-model" to be used for prognostics, where limit of predictability L is set to 0.5 (see Fig. 2.17).

TABLE 2.6: Prediction models - Settings

ANN-Parameters	**Settings**
Input / Hidden / Output layer neurons	3 / 5 / 1
Hidden / Output layer Activation function	sigmoidal / linear
Training Algorithm	Levenberg-Marquardt

ANFIS-Parameters	**Settings**
Input / Output layer neurons	3 / 1
Number / type of input membership function	3 / Pi-shaped
Rules / Fuzzy Inference System	27 / First order Sugeno
Defuzzification method	Weighted Average
Training Algorithm	Hybrid Method

2.6.2.2 Predictability results from prediction phase

As stated before, in order to exclude unpredictable features, predictability analysis is performed on each test. For illustration, simulation results from a single test are reported in Table 2.7 over different indexes of prediction horizon for a better understanding. The obtained results show that features F2 and F3 do not satisfy predictability criteria, neither by ANN nor by ANFIS, as it is clearly indicated by their lower values of predictability (i.e., $Pred < 0.5$). Similar findings about F2 and F3 are obtained from other test cases as well, by applying both connectionist tools.

This behavior is shown in Fig. 2.22, that presents a global picture of features predictability over prediction horizon ($tc + 134$) steps ahead. Note that, ANFIS showed better performance with higher predictability values as compared to ANN for most of the simulations. Also results show that predictability is highly dependent on the horizon of prediction and can vary from one prediction model to another.

TABLE 2.7: Predictability results on a single test

Features	Approach	H=tc+50	H=tc+120	H=tc+134
F1	ANFIS	0.934	0.606	0.504
	ANN	0.770	0.762	0.6173
F2	ANFIS	*0.005*	*0.0002*	*4.8e-05*
	ANN	*0.017*	*9.0e-06*	*4.6e-07*
F3	ANFIS	*0.0025*	*0.0025*	*5.2e-05*
	ANN	*0.0023*	*2.6e-14*	*3.09e-17*
F4	ANFIS	0.965	0.870	0.841
	ANN	0.982	0.876	0.840
F5	ANFIS	0.915	0.8925	0.925
	ANN	0.904	0.592	0.507
F6	ANFIS	0.943	0.9908	0.957
	ANN	0.947	0.995	0.963
F7	ANFIS	0.993	0.927	0.904
	ANN	0.966	0.907	0.888
F8	ANFIS	0.187	0.540	0.888
	ANN	0.970	0.637	0.360

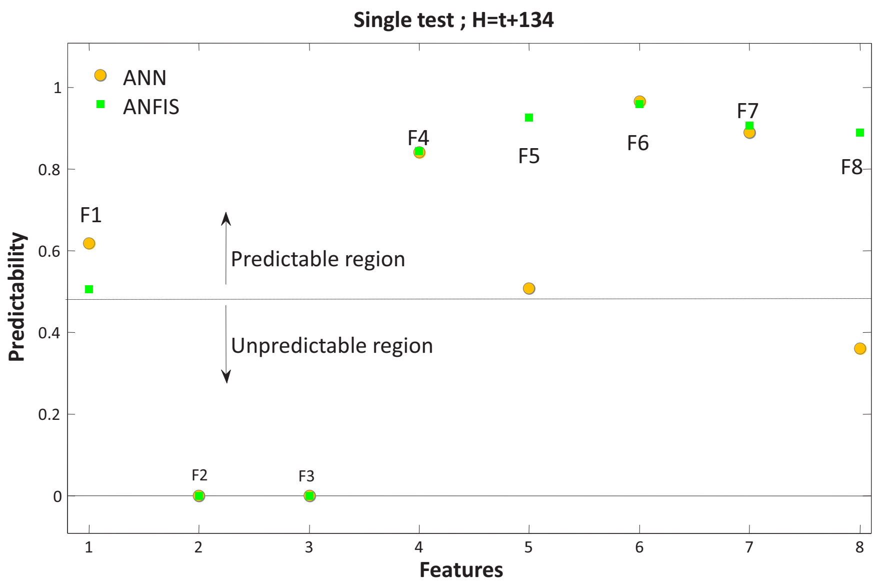

FIGURE 2.22: Predictability of features for $H = tc + 134$

As for illustration, the upper part of Fig. 2.23 depicts prediction results on feature F5 with both models ANN and ANFIS. The lower part presents the corresponding predictability measures over the horizon of prediction. As expected, ANFIS shows better prediction and higher predictability as compared to ANN with changing horizon. In other words, predictability measure not only shows the significance of a tool but also gives confidence in making predictions over the increasing horizon. In addition, higher predictability values can provide better confidence in predictions. Therefore, the prediction results from ANFIS are only used for further classification.

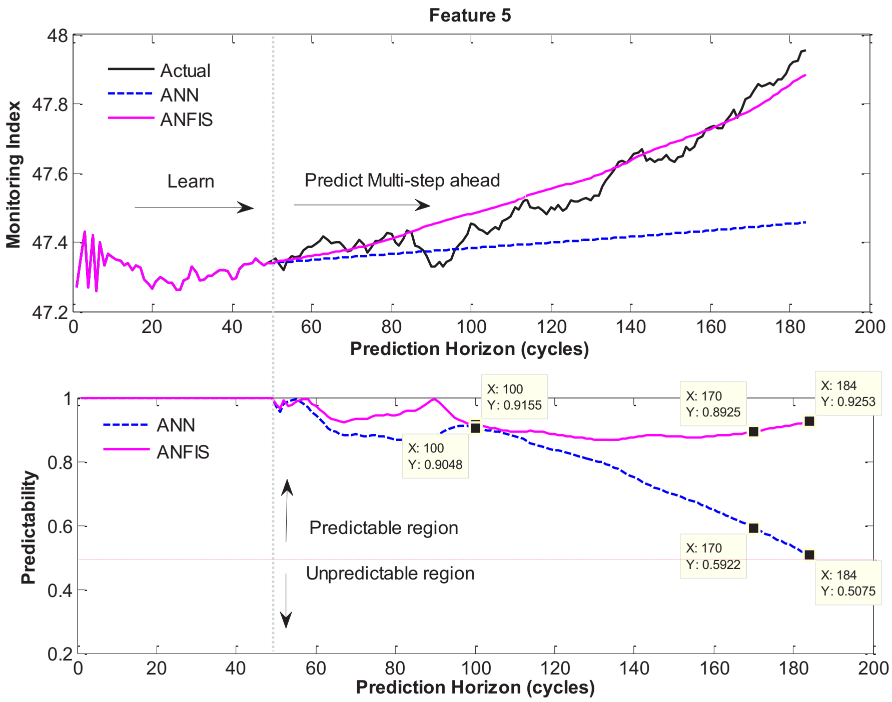

FIGURE 2.23: Example of degrading feature prediction

2.6.2.3 Unsupervised classification phase for health assessment

The main aim of the classification phase is to determine most probable states of the degrading equipment, and thus to provide a snapshot of time from projected degradations. In this step, the temporal predictions made by the predictor are analyzed by a classifier to estimate discrete states. Due to the absence of ground-truth information the unsupervised classification phase is met by well known FCM approach to handle unlabeled data [29]. Therefore, given the number of clusters and initial centers values, FCM can partition the data into groups of similar objects (or clusters) and can optimize center positions.

In brief, consider a dataset $\ell_D = \{X_i\}_{i=1}^N$ containing N unlabeled samples of n-dimension. FCM is used to obtain clusters c among multidimensional data, and centers belonging to each cluster $V = \{v_j\}_{j=1}^{c}$. Formally, the FCM clustering is attained by assigning membership grade (μ_{ij}) between [0, 1] to every data point that corresponds to each cluster center (or in other words belong to various groups), based on the measured distance between a data point and center of the cluster. If a data point is closer to a particular cluster center, a greater membership value is assigned. The summation of membership grades from all data points correspond to a membership equal to '1'. FCM aims to operate in an iterative manner, assuming the number of clusters is known it reduces following objective function:

$$J = \sum_{i=1}^{N} \sum_{j=1}^{c} (\mu_{ij})^m . \|x_i - v_j\|^2 \tag{2.11}$$

where, $\|x_i - v_j\|^2$ represents the Euclidean Distance (ED) (to measure similarity between) between the i^{th} data point and the j^{th} cluster center v (Eq. 2.12), fuzzy partition matrix can be represented as $U = [\mu_{ij}]_{c \times N}$, where μ_{ij} describes the membership of the i^{th} data point to the j^{th} centroid, and $m > 1$ is a weighting exponent. The key steps of FCM algorithm are summarize in Appendix A.3.

$$ED = \sqrt{\sum_{i=1}^{n} (x_i - v_j)^2} \tag{2.12}$$

2.6.2.4 Health assessment

Briefly, in the learning phase, a classifier is built to identify states of degrading equipment. For this purpose, different functioning modes have to be considered. In this case, four states of degradation are assumed, namely: steady state, degrading state, transition state and critical state. FCM is used to assign temporal predictions to different classes based on fuzzy partitioning (to cluster) multidimensional data. The number of clusters is preset to 4 for the FCM algorithm. Accordingly, the multidimensional training data of 40 engines are partitioned into four groups, where each group represents a particular state of degradation. It should be noted that, each engine has different life span or number of cycles. For illustration purpose, 2D visualizations from clustering phase are shown in Fig. 2.24.

In the on-line test phase, discrete states are assigned to continuous state predictions (or *msp*) by measuring the ED between predictions (of multidimensional data) and clusters centers (Eq. 2.12). Following that, class labels can be given by looking at the closest match. For more clarity, measured distances for the case of all features and predictable features are shown in Fig. 2.25. Finally, RUL estimate is obtained, when transition from degrading state (d) at time **tc** to faulty state (f) at time **tf** occurs (see Eq. (2.13)).

$$transition\ d \overset{State}{\rightarrow} f \Rightarrow RUL = S_{d \rightarrow f} - tc \tag{2.13}$$

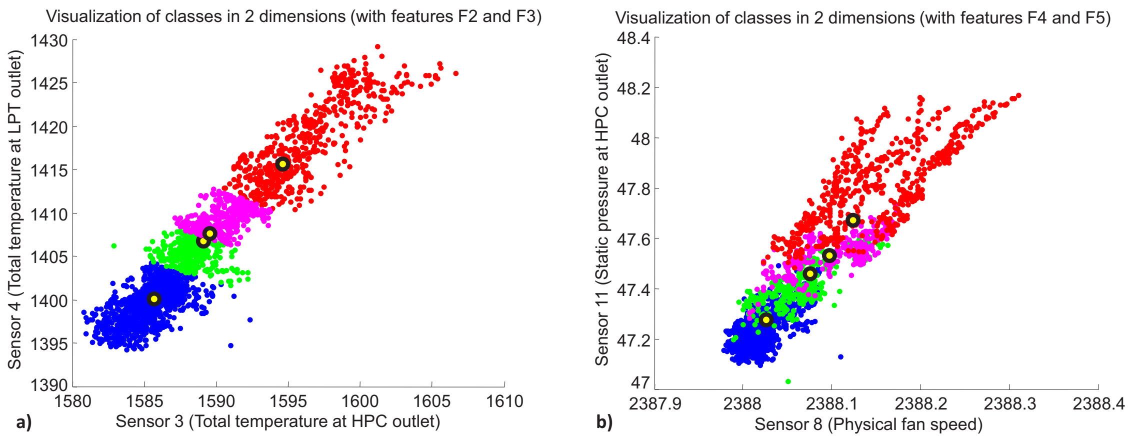

FIGURE 2.24: Visualization of classes from multidimensional data of 40 engines

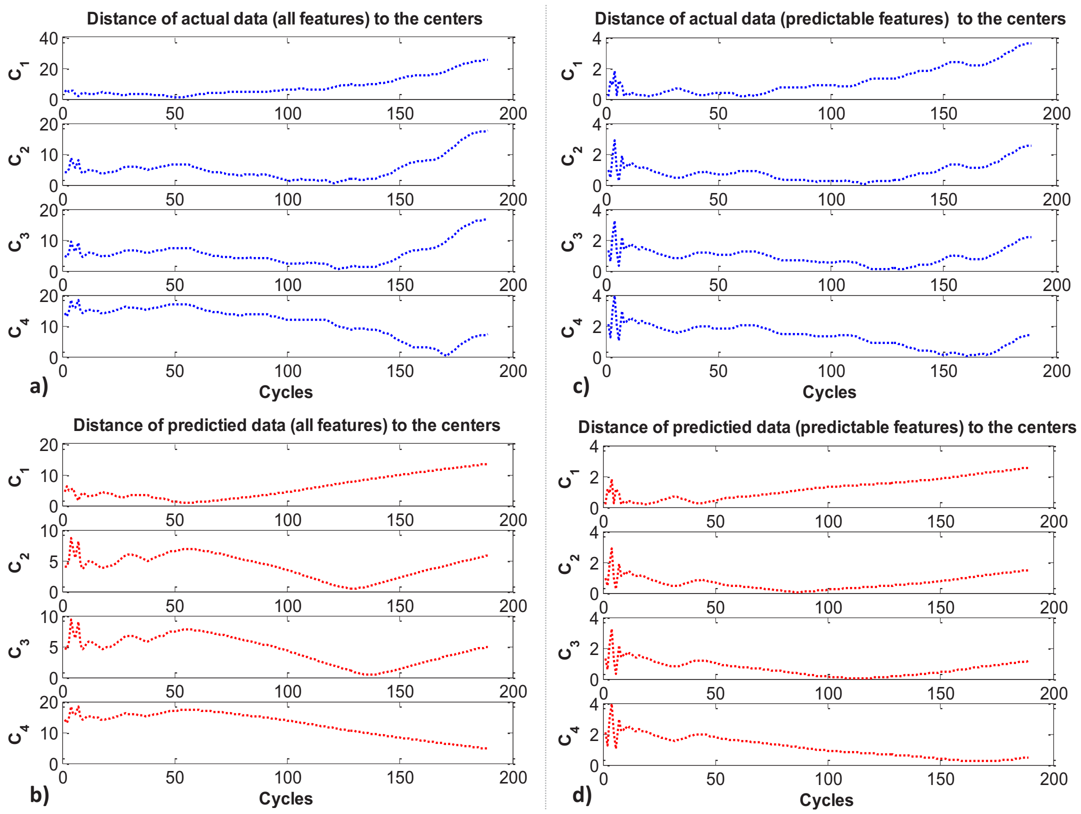

FIGURE 2.25: ED results $(test-1)$: $a)$,$b)$ All features and $c)$,$d)$ Predictable features

2.6.3 Prognostics results and impact of predictability

According to features prediction results using ANN and ANFIS predictors in Table 2.7. Prognostics should be performed by using only predictable features, that leads to better

RUL estimation i.e., with features {F1 ; F4 - F8}. Let's validate this assumption by estimating the RUL for each multidimensional time series predictions from ANFIS, using on one side, the whole set of pre-selected features {F1 - F8}, and on the other side, the final "predictability-based" selected features for a comparison. To show that, the temporal predictions from $test-1$ are classified to estimate the RUL. The results are organized according to two different cases (Fig. 2.26 and 2.27).
In the first case, classification is performed with "all features" {F1 - F8}, whereas in the second case the classification is performed with "predictable features" only i.e., excluding {F2 and F3}. It can be clearly seen from the results that, the RUL estimated from second case of classifications is closer to the actual RUL, thus, validating better prognostics accuracy and improvements achieved from proposed methodology.

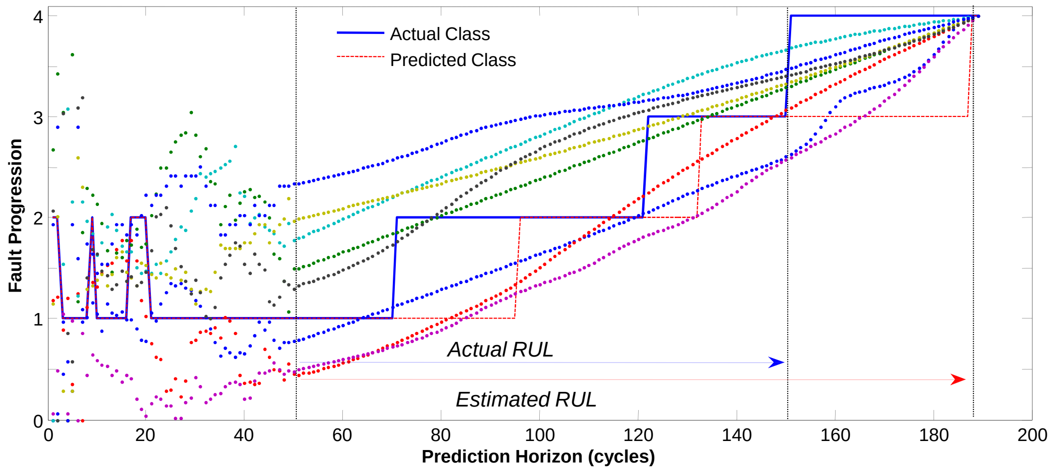

FIGURE 2.26: Classification with all features

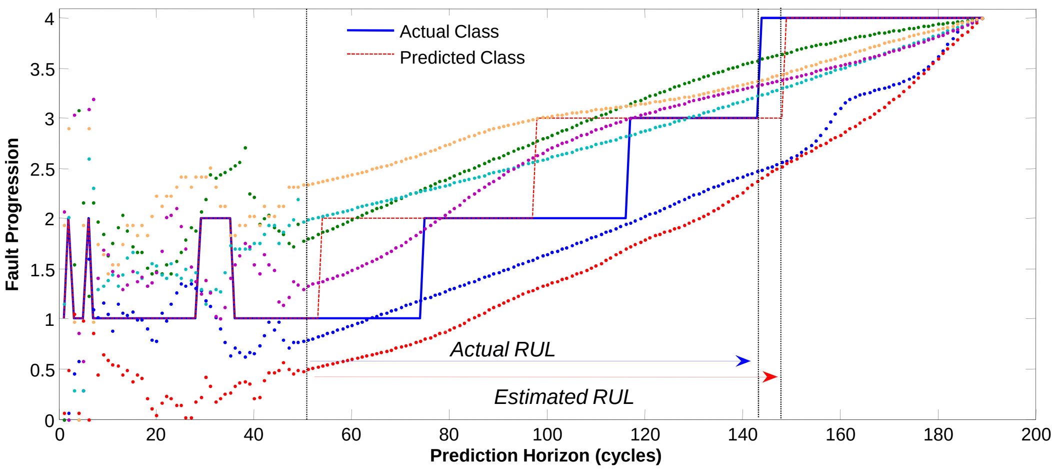

FIGURE 2.27: Classification with predictable features

The RUL percentage errors obtained from simulations are summarized in Table 2.8.

TABLE 2.8: RUL percentage error with ANFIS

Test	**All features**	**Selected Features**
1	7.096 %	0.636 %
2	11.83 %	1.898 %
3	24.34 %	1.265 %
4	15.95 %	0.6211 %
5	1.324 %	0.632 %
Mean % error	**12,10 %**	**1,01 %**

Results arrangement show that for each test, percentage error of RUL by considering all features {F1 - F8} is much higher as compared to the case of selected features, i.e., excluding F2 and F3. Also, results seem to be more stable among all tests when using the selected set of predictable features. Thus, features selection procedure based on predictability enhances significantly prognostics results.

2.6.4 Observations

According to predictability driven prognostics results with multivariate degradation based modeling strategy, the key observations are summarized as follows.

- Performances of RUL estimates are obviously closely related to accuracy of the *msp* model. As for example, ANFIS appears to be more suitable as compared to classical ANN for long-term predictions. This aspect has to be addressed to encounter the challenges of robustness and reliability (CHAPTER 3).

- Health assessment is performed by FCM algorithm, which is sensitive to parameter initialization and has inconsistent performance. Also, FCM clustering results are based on number of clusters set by the user, e.g. in this application even the life span of each engine is different, the multidimensional data are clustered into 4 groups. Therefore, it is required to further enhance the performances (i.e., robustness, reliability) of health assessment model and also its applicability (CHAPTER 4).

2.7 Summary

This chapter discusses the importance of data-processing to extract and select features with essential characteristic like monotonicity, trendability and predictability. A new approach for feature extraction and selection is proposed by performing pre-treatment of raw CM data to obtain cumulative features that show high monotonicity and trendability. The proposed developments are validated on real data of bearings from PHM challenge 2012. Surely, such features can lead to highly accurate long term predictions and ease in implementation of prognostics model. This aspect is validated in Chapter 3. Secondly, another concept of predictability driven prognostics is presented to perform post-treatment of features subset for further reduction. Indeed, there is no use of features that are hard to be predicted. The proposed developments are validated on real data of turbofan engines from PHM challenge 2008. Prognostics modeling is achieved by multivariate degradation based modeling strategy. For the sake of illustration ANN and ANFIS are used as prediction models and FCM algorithm is used for unsupervised classification phase. Besides that, some weaknesses of multivariate degradation based modeling are also highlighted in section 2.6.4. To further enhance this strategy, and according to the objective of this thesis (section 1.4.2.3). Chapter 3 presents a new approach for "long-term predictions", and Chapter 4 presents a new approach for unsupervised classification.

Chapter 3

From features to predictions

This chapter presents a new connectionist approach to perform "long-term predictions" for prognostics. More precisely, a Summation Wavelet-Extreme Learning Machine algorithm is proposed to address issues related to prognostics modeling challenges, i.e., robustness, reliability, applicability. Moreover, an ensemble of SW-ELM models is proposed to quantify the uncertainty due to data and modeling phase.
Firstly, for issues related to time-series, i.e., approximation, one step-ahead prediction, and multi-steps ahead prediction, SW-ELM is benchmarked with Extreme Learning Machine, Levenberg Marquardt and ELMAN algorithms on six industrial data sets. Secondly, to validate the efficiency of SW-ELM with respect to challenges, a real case of Computer Numerical Control (CNC) machine is considered for tool wear monitoring task. Thorough comparisons with ELM algorithm are given to show enhanced robustness and reliability of SW-ELM on different data sets.
*Finally, to validate the concept of cumulative features (*Chapter *2), SW-ELM is used as a potential model to perform "long-term predictions" with cumulative features. For this case study, SW-ELM is applied to real data of bearings from PHM challenge 2012.*

3.1 Long-term predictions for prognostics

According to multivariate degradation based modeling strategy (Chapter 1), the prognostics phase is composed of two complementary modules: a prediction engine that forecasts observations in time, and a classifier which provides the most probable states of the machinery. The RUL is obtained from the estimated time to reach the faulty state (see Fig. 1.12). Obviously, for prognostics part, the prediction phase is critical and must be dealt in an accurate manner for recommending timely maintenance actions / system configurations (e.g. load profile changes). Therefore, while building a prediction model it is important tackle following issues.

- How to handle complex multidimensional data?
- Is the model capable enough for "long-term" predictions?
- How to associate confidence to predictions for decision making?

Generally, the dynamic behavior of a machinery can be represented by time series measurements (or features) collected over long periods of time. Usually the time series data are inherently non-stationary, and properties of underlying system are not known. Also such data can be high dimensional, and are source of behaviors (states) that the machine may have undergone in the past [55]. For example, consider a dataset $\ell_D = \{X_i\}_{i=1}^N$ containing N unlabeled samples of n time series Eq. (3.1).

$$\ell_D = \begin{bmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \dots & \vdots \\ x_{N1} & \dots & x_{Nn} \end{bmatrix} \tag{3.1}$$

Given such data, machine learning methods from data-driven prognostics category can be applied to predict the behavior of the degrading machinery. More precisely, connectionist approaches like Artificial Neural Networks (ANN) and Neuro Fuzzy systems (NFs) benefit from a growing interest [76]. According to PHM literature, several works confirm strong approximation capabilities of such connectionist approaches [9, 18, 45, 64, 98, 134, 136, 164, 170, 196, 202, 217].
To recall, the aim of a connectionist approach is, to "learn" the relationship between input data (i.e., training inputs) and the output data (or training targets). Suppose that, actual relations is $[\hat{Y}] = f_t([X],[\theta])$ (Eq. 2.8). The learning problem is to estimate the unknown function f_t using training data, which is know as supervised learning. Further, for prognostics, the learned connectionist approach is used for performing multi-steps ahead prediction (*msp*) over a long prediction horizon.
In literature, different methods can be applied to achieve *msp* with connectionist approaches (iterative approach, direct approach, DirRec approach, MISMO approach, parallel approach, see [76] for details). However, for prognostics the future is unknow, except "iterative approach" all other *msp* approaches require *prior* knowledge of final horizon step, which is preset by the practitioner. As for illustration, consider Fig. 3.1, where the iterative approach is presented as the only one to achieve RUL estimates, for whatever the actual horizon [76]. This advantage can also be seen especially, while performing simultaneous discrete state estimation (or health assessment by a classifier).
Besides that, while building a prediction model, it is also important to consider robustness, reliability and applicability challenges, which affect prognostics performances. Therefore, following topics aim at introducing a new connectionist approach for performing long-term predictions namely, Extreme Learning Machine (ELM).

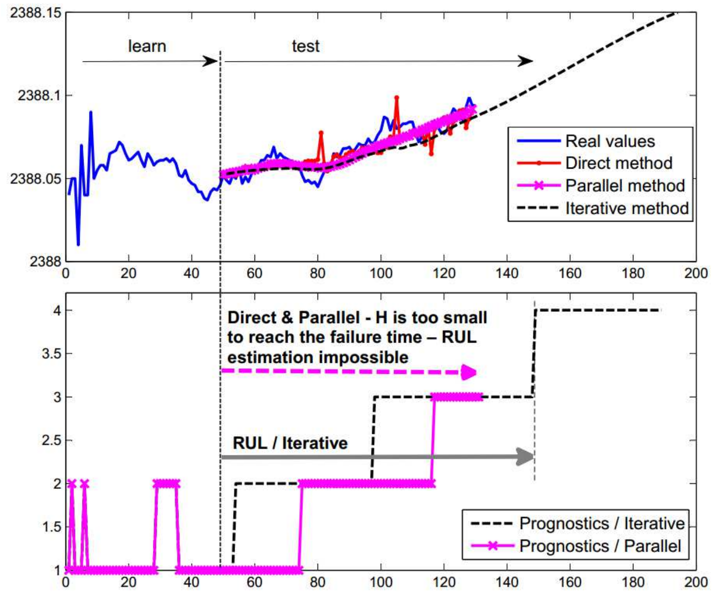

FIGURE 3.1: Discrete state estmation and RUL estimation [76]

3.2 ELM as a potential prediction tool

3.2.1 From ANN to ELM

ANN is a mathematical model which replicates the functionality of biological neurons in the human brain [103]. In simple words, an ANN can be addressed as a parallel structure of multi-input multi-output (MIMO) data processing nodes, that appears to the user as a black box. ANNs are extensively applied to different domains such as engineering, physics, medicine to solve problems like pattern classification, clustering, optimization, function approximation or prediction [103, 162, 179]. In PHM domain, ANNs are a special case of adaptive networks that are most commonly used among machine learning methods [84, 217]. Indeed, they have the ability to learn from examples and can capture complex relationships among CM data (even if noisy). They do not require *a priori* assumptions on the properties of given data set, neither they assume any *a priori* model [184].

In brief, ANN can have two types of architectures: feed-forward networks (FFNNs) and recurrent neural networks (RNNs). FFNN has connected strengths in forward direction, and RNN has cyclic connections Fig. 3.2. Such models must be tuned to learn parameters like weights and bias, in order to fit the studied problem. In other words, all the parameters of FFNNs and RNNs need to be adjusted, and there exist dependence among parameters of different layers [69]. According to Jaeger's estimation, around 95% of literature is on FFNNs [102].

Constructing a good neural network model is non-trivial task and practitioners still have

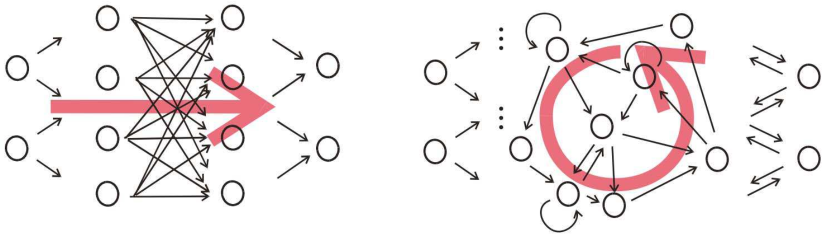

Figure 3.2: Structure of FFNN and RNN [102]

to encounter several issues that may affect the performance of ANNs and limit their applicability (ease of application) for real industrial case [179]. As for examples, such problems involve: parameter initialization, complexity of hidden layer, type of activation functions, slow iterative tuning, presence of local minima, over-fitting, generalization ability [111]. To overcome such drawbacks, comparatively two new learning schemes recently proposed for FFNN and RNN, namely, Extreme Learning Machine [94], and Echo State Network (ESN) [101]. Both learning schemes are based on random projection, because input-hidden layer parameters are randomly initialized, and learning is only achieved by solving the inverse of a least square problem. A comparative study between ELM and ESN on non-linear prediction benchmark problems shows that, ESN requires more training time as compared to ELM for acquiring the dynamic performance [31]. Both are sensitive to the number of hidden neurons. Moreover, for small data ESN was robust and show good generalization, however, with sufficient training samples, in most cases ELM show better generalization performances [31]. Most importantly, to the best of our knowledge, ELM has been proved for its universal approximation capability [90, 91, 92, 95], whereas for ESN its not the case. In addition, recent survey show the advantages of ELM over conventional methods to train ANNs [93]. Such findings highlight the importance of ELM as a suitable candidate for prognostics case, to achieve "long-term predictions".

3.2.2 ELM for SLFN: learning principle and mathematical perspective

Unlike conventional learning algorithms for FFNNs, ELM is a new learning scheme for SLFN that has been recently proposed by Huang et al. [94]. Almost all learning algorithms for SLFNs require adjustment of parameters that results dependence between different layers of parameters like, weights and biases. Therefore, many iterative tuning steps are required by traditional learning algorithms [95]. In opposite to that, ELM algorithm avoids slow iterative learning procedure and only requires a one-pass operation to learn SLFN. This is mainly due to the fact that there is no bias for the output layer neuron (see Fig. 3.3).

In brief, to initiate one-pass learning operation, the hidden node network parameters (weights and biases) are randomly generated without any *prior* knowledge or training procedure [39]. Consequently, the ELM turns into a system of linear equations and the

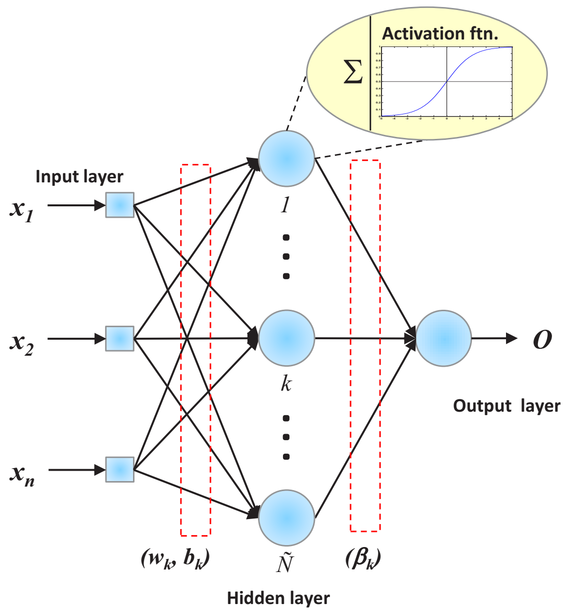

FIGURE 3.3: Single layer feed forward neural network

unknown weights between the hidden and output layer nodes can be obtained analytically by only applying Moore-Penrose generalized inverse procedure [124, 167].
Let note n and m the numbers of inputs and outputs (i.e., targets), N the number of learning data samples (x_i,t_i), where $i \in [1 \ldots N]$, $x_i = [x_{i1},x_{i2},...,x_{in}]^T \in \Re^n$ and $t_i = [t_{i1},t_{i2},...,t_{im}]^T \in \Re^m$, and $\tilde{N}$ the number of hidden nodes, each one with an activation function $f(x)$. For each sample j, the output o_j is mathematically expressed as:

$$\sum_{k=1}^{\tilde{N}} \beta_k . f(w_k . x_j + b_k) = o_j \ , j = 1,2,...., N \tag{3.2}$$

where $w_k = [w_{k1},w_{k2},...,w_{kn}]^T \in \Re^n$, is an input weight vector connecting the k^{th} hidden neuron to the input layer neurons, $(w_k \cdot x_j)$ is the inner product of weights and inputs, and $b_k \in \Re$ is the bias of k^{th} neuron of hidden layer. Also, $\beta_k = [\beta_{k1},\beta_{k2},...,\beta_{km}]^T \in \Re^m$, is the weight vector to connect the k^{th} neuron of hidden layer and output neurons.
Therefore, in order to minimize the difference between network output o_j and given target t_j, $\sum_{j=1}^{\tilde{N}} \|o_j - t_j\| = 0$, there exist β_k, w_k and b_k such that:

$$\sum_{k=1}^{\tilde{N}} \beta_k . f(w_k . x_j + b_k) = t_j \ , \ j = 1,2,...., N \tag{3.3}$$

which can be expressed in matrix form as,

$$H\beta = T \tag{3.4}$$

where,

$$H\left(w_1,\ldots,w_{\tilde{N}},x_1,\ldots,x_{\tilde{N}},b_1,\ldots,b_{\tilde{N}}\right) = \begin{bmatrix} f(w_1.x_1+b_1) & \ldots & f(w_{\tilde{N}}.x_1+b_{\tilde{N}}) \\ \vdots & \ldots & \vdots \\ f(w_1.x_N+b_1) & \ldots & f(w_{\tilde{N}}.x_N+b_{\tilde{N}}) \end{bmatrix}_{N\times\tilde{N}} \tag{3.5}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N}\times m} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N\times m} \tag{3.6}$$

The least square solution of the linear system defined in Eq. (3.4), with minimum norm (magnitude) of output weights β is:

$$\hat{\beta} = H^{\dagger}T = \left(H^T H\right)^{-1} H^T T \tag{3.7}$$

where $H^{\dagger}$ represents the Moore-Penrose generalized inverse solution for the hidden layer output matrix H [167]. Finally, learning of ELM can be summarized in Algorithm 1.

Algorithm 1 Learning scheme of an ELM

Assume

- n inputs, m outputs, $\tilde{N}$ hidden nodes ($k = 1 \ldots \tilde{N}$)

Require

- N learning data samples
- An activation function $f(x)$ (e.g sigmoid)

1: Randomly assign parameters of hidden nodes i.e., weights and bias (w_k,b_k).
2: Obtain the hidden layer output matrix H.
3: Find the output weight matrix β: $\beta = H^{\dagger}T$, where $H^{\dagger}$ represents the Moore-Penrose generalized inverse solution for the hidden layer output matrix H [167].

3.2.3 Discussions: ELM for prognostics

3.2.3.1 Better applicability of ELM

ELM approach shows rapid learning and good generalization ability for SLFNs [95]. It requires less human intervention, because, it doesn't have any control parameters to be manually tuned, except the number of hidden layer neurons, which makes it appropriate for real applications [30]. In addition, an ELM can also show good performance in situations where it is hard to obtain enough data to train SLFN and recent study confirms the advantages of ELM over earlier approaches for ANN [93]. This shows fitness and improved applicability of an ELM for real applications like prognostics, that encounter difficulties like data scarcity, time complexity, human involvement etc. However, as far as authors know, there is no explicit application of ELM for machine failure prognostics.

3.2.3.2 Issues and requirements

ELM method has two parameters that are set by the user, number of hidden layer neurons and variance of the hidden layer weights [151]. The interaction between improper initialization of weights and neurons of the hidden layer may impair performance of the ELM model [30]. In view of expected performances of a prognostics model highlighted in section 1.4, practical considerations related to model accuracy and implementation issues should be addressed for real applications.

- Due to **random initialization** of parameters (weights and biases), ELM model may require high complexity of hidden layer [162]. This may cause ill-condition, and reduce robustness of ELM to encounter variations in the input data, and the expected output of the model may not be close to the real output [77]. Also the randomly initialized weights can affect model generalization ability which should also be considered.

- It is required to carefully choose hidden neuron **activation functions** that can participate in better convergence of the algorithm, show good ability to transform non-linear inputs, and also result in a compact structure of network (i.e, hidden neurons) for a suitable level of accuracy [91, 103, 111].

- Finally, ELM can perform estimates after learning phase, however, it does not quantify uncertainty of modeling phase like any ANN. Therefore, it is necessary to bracket unknown future with **confidence intervals** to show reliability of estimates and also to facilitate decision making step for prognostics applications.

Therefore, to address these issues and requirements, the next section presents a new learning scheme for SLFN, which is an improved variant of ELM, namely, the Summation Wavelet- Extreme Learning Machine (SW-ELM).

3.3 SW-ELM and Ensemble models for prognostics

Combining neural networks and wavelet theory as an approximation or prediction models appears to be an effective solution in many applicative areas. However, when building such systems, one has to face parsimony problem, i.e., to look for a compromise between the complexity of the learning phase and accuracy performances. Following topics focus on building a new structure of connectionist network, the SW-ELM that enables good accuracy and generalization performances, while limiting the learning time and reducing the impact of random initialization procedure.

3.3.1 Wavelet neural network

3.3.1.1 Concept and structure

Wavelet theory is an outcome of multidisciplinary struggles, that brought together engineers, mathematicians and physicists [159]. The term wavelet means a “little wave”.

Mainly, a Wavelet Transform (WT) of continuous form behaves as a flexible time-frequency window, that shrinks when analyzing high frequency and spreads when low frequency behavior is observed [15]. The WT can be divided into two types, Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). CWT can be considered as the inner product of a signal $x(t)$ with a basis function $\Psi^*_{s,d}$, which can be formulated as follows (Eq.3.8).

$$CWT(d,s) = \frac{1}{\sqrt{d}} \int_{-\infty}^{+\infty} x\,(t)\,\Psi^* \left(\frac{t-s}{d}\right) dt \tag{3.8}$$

where Ψ is a mother wavelet function, $*$ denotes complex conjugation, and d and s are scale (or dilate) and translate (or shift) factors for the mother wavelet. It should be noted that CWT has the drawback of impracticality with digital computers, where, DWT is used in practice. Thus, the scale and translation factors are evaluated on a discrete grid of time scale to generate scaled and shifted daughter wavelets from a given mother wavelet Ψ, for e.g. see Fig. 3.4. Therefore, discretized wavelet (DW) for the sequence of samples x_i from $x(t)$ can be formulated as follows (Eq. 3.9). Further details can be found in [15].

$$DW = \sum_i x_i\, d_i^{-1/2} \Psi \left(\frac{t-s_i}{d_i}\right) \tag{3.9}$$

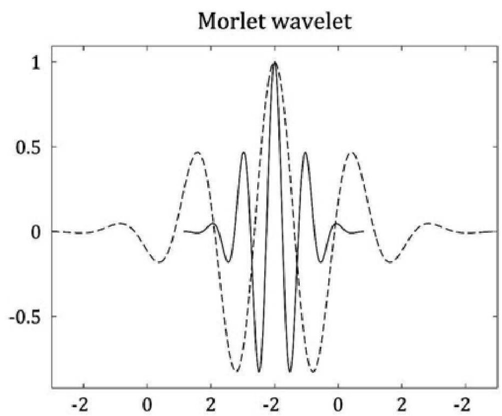

FIGURE 3.4: Mother and daughter wavelet from Morlet function

An analogy can be found between the expression of the DWT and ANN output. Indeed, Eq. (3.9) can be seen as the output expression of a SLFN that would have an activation function Ψ for a hidden node, with a linear neuron in the output layer. This lead to a combination of WT theory with basics from FFNNs, to produce a new category of neural networks known as WNN. Generally, such combination can be classified into two types. In the first case, the wavelet part is decoupled from network learning. In this manner, a signal is decomposed on some wavelets and its wavelet coefficients are furnished to a FFNN. In the second category, wavelet theory and FFNN are combined into a hybrid

structure. The scope of following topics cover the latter category.
Let note n the number of inputs of a WNN, and $\tilde{N}$ the number of hidden nodes with wavelet functions (eg. Morlet, see Fig. 3.3). According to this, the output of a WNN can be formulated as:

$$y = \sum_{k=1}^{\tilde{N}} v_k \Psi_k \left(\sum_{j=1}^{n} w_{kj} x_j \right) \tag{3.10}$$

$$\Psi_k \left(x \right) = \left| d_k \right|^{-1/2} \Psi \left(\frac{x - s_k}{d_k} \right) \tag{3.11}$$

where $x = x_1$,x_2,...,x_n depicts the input values, Ψ_k represents the family of daughter wavelets that are scaled and translated from a single mother wavelet function Ψ, d_k and s_k are the corresponding scale (dilate) and translation factors. Lastly, $w_k = [w_{k1}, w_{k2}, \ldots, w_{kn}]^T \in \Re^n$ is an input weight vector connecting the k^{th} hidden neuron to the input layer neurons, and v_k is the weight to connect the k^{th} neuron of hidden layer and the output.

3.3.1.2 Issues and requirements

According to literature, the initialization of dilation and translation factors (d_k and s_k in Eq. (3.11)) is considered as a critical phase. Indeed, it is necessary to properly initialize these parameters for fast convergence of algorithm [39, 150]. As wavelets functions vanish rapidly, improper initialization may lead to the following issues:

- a wavelet can be too local for a very small dilation;
- improper translation may be out of interest domain.

Therefore, the random tuning of dilation and translation factors of wavelets is inadvisable, and parameters initialization should be based on the input domain. This aspect is taken into account in the proposed model (section 3.3.2).

3.3.2 Summation Wavelet-Extreme Learning Machine for SLFN

3.3.2.1 Structure and mathematical perspective

As mentioned in previous section, ELM is quiet efficient as compared to traditional methods to learn SLFNs. However, issues like parameters initialization, model complexity and choice of activation functions have to be carefully addressed for improved performance. Since structure and proper choice of activation functions enhances the capability of the SLFN network to encounter low and high frequency signals simultaneously. The number of neurons required for hidden layer decreases and a compact structure is achieved [15, 103, 124]. Therefore, we propose the Summation Wavelet-Extreme Learning (SW-ELM) algorithm. The proposed structure of SW-ELM takes advantages of WT and SLFN (see Fig. 3.5), for which two significant modifications are performed in the hidden layer.

- **Structure** - Non-linear transformations are dealt in a better manner by using a conjunction of two distinct activation functions (f_1 and f_2) in each hidden node rather than a single activation function. The output from a hidden node is the average value after performing transformation from dual activations ($\bar{f} = (f_1 + f_2)/2$).

- **Activation functions** - To improve convergence of algorithm, an inverse hyperbolic sine ([124], Eq. (3.12)) and a Morlet wavelet ([39, 159], Eq. (3.13)) are used as dual activation functions.

$$f_1 = \theta(X) = log\left[x + (x^2+1)^{1/2}\right] \tag{3.12}$$

$$f_2 = \psi(X) = cos(5x)\, e^{\left(-0.5x^2\right)} \tag{3.13}$$

Such a combination of activation functions makes the network more adequate to deal with dynamic systems.
According to modifications mentioned above, for each sample j Eq. 3.2 can be adapted and the output o_j is mathematically expressed as:

$$\sum_{k=1}^{\tilde{N}} \beta_k.\bar{f}\left[(\theta,\psi)(w_k.x_j + b_k)\right] = o_j \;, j = 1,2,...,N \tag{3.14}$$

where $w_k = [w_{k1},w_{k2},...,w_{kn}]^T \in \Re^n$, is an input weight vector connecting the k^{th} hidden neuron to the input layer neurons, $(w_k.x_j)$ is the inner product of weights and inputs, and $b_k \in \Re$ is the bias of k^{th} neuron of hidden layer. Also, $\beta_k = [\beta_{k1},\beta_{k2},...,\beta_{km}]^T \in \Re^m$, is the weight vector to connect the k^{th} neuron of hidden layer and output neurons. Finally, $\bar{f}$ shows the average output from two different activation functions i.e., an inverse hyperbolic sine activation function θ and a Morlet wavelet activation function ψ. In order to minimize the difference between network output o_j and given target t_j, $\sum_{j=1}^{\tilde{N}} \|o_j - t_j\| = 0$, there exist β_k, w_k and b_k such that:

$$\sum_{k=1}^{\tilde{N}} \beta_k.\bar{f}\left[(\theta,\psi)(w_k.x_j + b_k)\right] = t_j \;,\; j = 1,2,...,N \tag{3.15}$$

which can be expressed in matrix form as,

$$H_{avg}\beta = T \tag{3.16}$$

where H_{avg} is a $\left[N \times \tilde{N}\right]$ matrix expressed as,

$$H_{avg}\left(w_1,\ldots,w_{\tilde{N}},x_1,\ldots,x_{\tilde{N}},b_1,\ldots,b_{\tilde{N}}\right) =$$

$$\bar{f}(\theta,\psi)\begin{bmatrix} (w_1.x_1 + b_1) & \ldots & \left(w_{\tilde{N}}.x_1 + b_{\tilde{N}}\right) \\ \vdots & \ldots & \vdots \\ (w_1.x_N + b_1) & \ldots & \left(w_{\tilde{N}}.x_N + b_{\tilde{N}}\right) \end{bmatrix} \tag{3.17}$$

and,

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \tag{3.18}$$

Finally, the least square solution of the linear system defined in Eq. (3.16), with minimum norm (magnitude) of output weights β is:

$$\hat{\beta} = H_{avg}^{\dagger} T = \left(H_{avg}^T H_{avg}\right)^{-1} H_{avg}^T T \tag{3.19}$$

where $H_{avg}^{\dagger}$ represents the Moore-Penrose generalized inverse solution for the hidden layer output matrix H_{avg} [167].

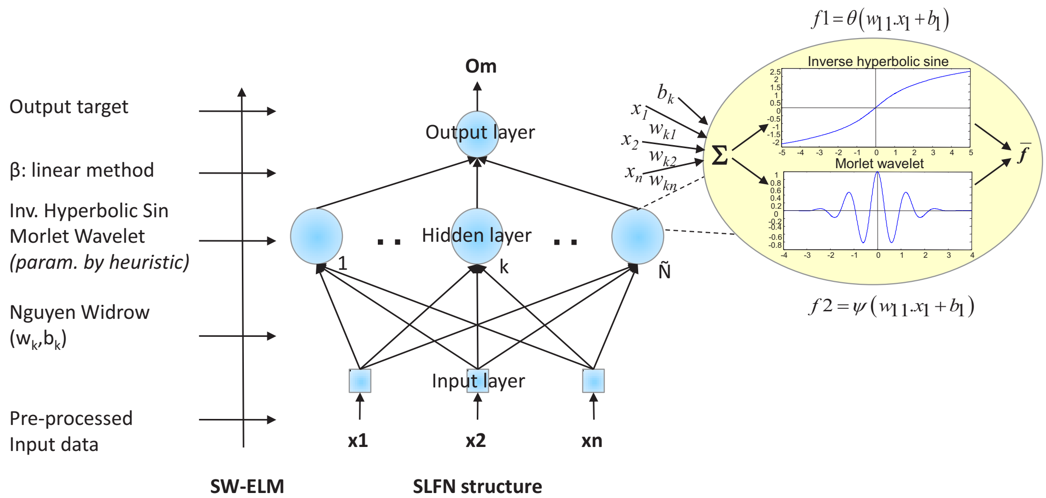

FIGURE 3.5: Structure and learning view of proposed SW-ELM

3.3.2.2 Learning scheme of SW-ELM

Main learning phase derives from Eq. (3.17) and (3.19). However, according to issues and requirements related to ELM and WNN (sections 3.3.1.2 and 3.2.3.2), it is required to take care of parameters initialization task and to provide a better starting point to algorithm. We address those aspects by considering to types of parameters: those from the wavelets (dilation and translation factors) and those from the SLFN (weights and bias for input to hidden layer nodes). Details of the learning scheme are given in Algorithm 2, and an outline of the main steps is synthesized hereafter.

- **Initializing wavelet parameters** - To initialize wavelet dilation and translation parameters (d_k and s_k in Eq. (3.11)) before the learning phase, a heuristic approach is applied to generate daughter wavelets from a mother wavelet function (in our case, a Morlet wavelet). Dilation and translation values are adapted by considering the domain of the input space where wavelet functions are not equal to zero [150]. This procedure is detailed in Algorithm 2. Besides that, wavelet function with a small dilation value are low frequency filters, whereas with increasing dilation factors the wavelet behave as high frequency filter [15]. Finally, the effects of random initialization of wavelet parameters are avoided, and the initialization guarantees that wavelet function stretches over the whole input domain [150].

- **Initializing weights and bias** - The hard random parameters initialization step is substituted by well known Nguyen Widrow (NW) procedure to initialize weights and bias [145]. The intent is to provide a better starting point for learning. NW method is a simple alteration of hard random initialization that aims at adjusting input weights intervals and hidden bias according to the input-hidden layer topology. It has been argued that NW method has shown improved performances over others methods for random parameters initialization of ANNs [152].

Algorithm 2 Learning scheme of the SW-ELM

Require - N learning data samples (x_i, t_i), n inputs $(j = 1 \dots n)$
- $\tilde{N}$ hidden nodes $(k = 1 \dots \tilde{N})$
- An inverse hyperbolic sine and a Morlet wavelet activation functions

Ensure - Initialize weights and bias from SLFN, initialize Morlet parameters
- Find hidden to output weights matrix β

SW-ELM learning procedure

Initialization of wavelet parameters [150]
- Define the input space domain intervals
    - Compute $[x_{jmin}\ ;\ x_{jmax}]$: {domain containing the input x_j in all observed samp.}
- Define dilation and translation parameters per domain
    - Compute $D_{kj} = 0{,}2 \times [x_{jmax} - x_{jmin}]$: {temporal dilatation parameter for input x_j}
    - Compute $M_{kj} = [x_{jmin} + x_{jmax}]/2$: {temporal translation parameter for input x_j}
- Initialize Morlet parameters (d_k and s_k)
    - Compute $d_k = mean(D_{kj})_{j=1\dots n}$: {dilatation factor}
    - Compute $s_k = mean(M_{kj})_{j=1\dots n}$: {translation factor}

Initialization of weights and bias parameters by Nguyen Widrow approach [145]
- Initialize small (random) input weights $w_{k(old)}$ in $[-0.5\ ;\ +0.5]$: {weights from input nodes to hidden nodes}
- Adjust weights parameters by applying NW approach
    - Compute $\beta_{factor} = C \times \tilde{N}^{\frac{1}{n}}$: {$C$ is a constant ≤ 0.7}
    - Compute $w_{k(new)} = \beta_{factor} \times \frac{w_{k(old)}}{\|w_{k(old)}\|}$: {normalized weights}
- Initialize bias values b_k
    - b_k = random number between $-\beta_{factor}$ and $+\beta_{factor}$

Adjust linear parameters: those from the hidden to the output layers
- Obtain hidden layer output matrix H_{avg} using Eq. (3.17)
- Find the output weight matrix $\hat{\beta}$ in Eq. (3.19) by applying Moore-Penrose generalized inverse procedure

3.3.3 SW-ELM Ensemble for uncertainty estimation

Despite the fact that ELM based algorithms have several advantages over traditional methods to learn SLFN, one of the main shortcoming can be, that their solution vary for each run due to random initialization of learning parameters (weights and biases). Thus unsatisfactory or low performances can occur. This is also a common existing issue of classical ANNs, that parameters are randomly initialized without *prior* information of final parameters (from hidden to output layer) [139, 181]. Also, such models do not furnish any indication about the quality of outcomes in order to facilitate practitioner with decision making. That is, considering the uncertainties which arise either due to model misspecification or either due to variations of input data by probabilistic events [114]. Although there is no single algorithm or model for prognostics that works for all sorts of situations, the integration (ensemble) of multiple models appears to be less likely to be in error than an individual model [89, 114]. Due to such issues, in literature it is preferred to apply an ensemble strategy of multiple models to improve robustness and to show reliability of estimates [89].
Mainly, the ensemble strategy is achieved by integrating several SW-ELM models, where each individual model is initialized with different parameters (Fig. 3.6). This strategy enables building confidence intervals (CI) that indicate the presence of uncertainty and can facilitate in decision making (e.g. to perform further maintenance actions). Following that, the desired output $\overline{O}$ can be obtained by averaging outputs from multiple SW-ELM models.

$$\overline{O}_j = \frac{1}{M}\sum_{m=1}^{M} \hat{o}_j^m \tag{3.20}$$

where $\hat{o}_j^m$ is the predicted output of m^{th} model against the j^{th} input sample.
Assuming that SW-ELM models are unbiased, variance $\sigma^2_{\overline{o}_j}$ due to random initialization of parameters and input variations can be estimated via variance of M models outcomes [114].

$$Var(\overline{O}_j) = \sigma^2_{\overline{o}_j} = \frac{1}{M-1}\sum_{m=1}^{M}\left(\hat{o}_j^m - \overline{O}_j\right)^2 \tag{3.21}$$

Thanks to variance $\sigma^2_{\overline{o}_j}$, the final estimated output can be constructed according to a confidence limit interval (CI):

$$\mathrm{CI} = \left[\overline{O}_j \pm z^* \times \sqrt{\sigma^2_{\overline{o}_j}}\right] \tag{3.22}$$

where z^* is the critical value for CI up to desired level, i.e. the $100(1-\alpha)\%$ quantile of a normal distribution. In our case, confidence level is **95**%, where $z^* = 1.96$ according to standard Gaussian distribution table.

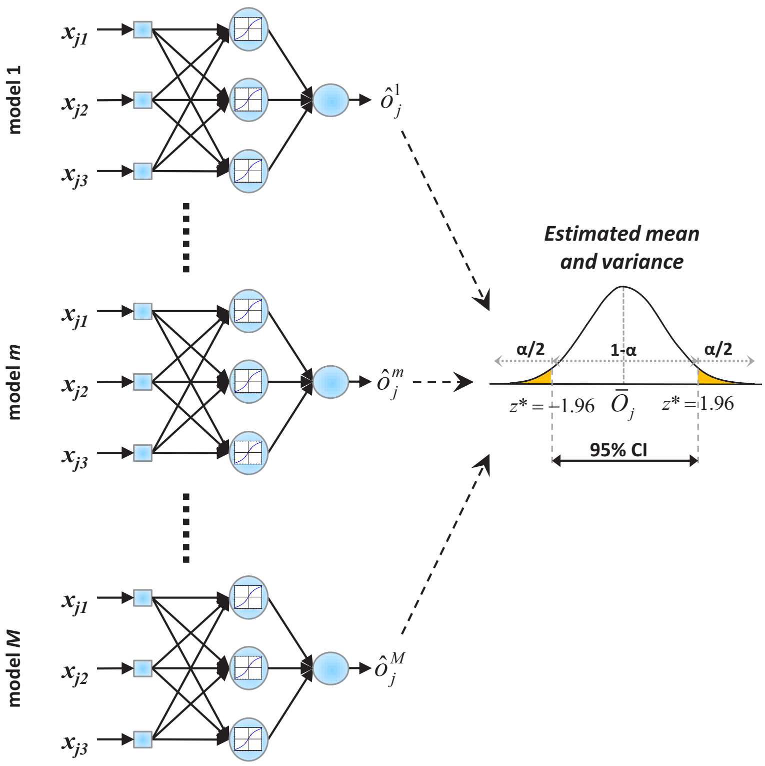

Figure 3.6: SW-ELME and confidence limit interval (CI)

3.4 Benchmarking SW-ELM on time series issues

3.4.1 Outline: aim of tests and performance evaluation

The aim of this part is to demonstrate enhanced performances of the proposed SW-ELM, that is benchmarked with the ELM, Levenberg-Marquardt algorithm for SLFN and ELMAN network. For simulation purpose, a sigmoid function is used as hidden node activation function for ELM, LM-SLFN and ELMAN network, whereas for SW-ELM, dual activations are used: an inverse hyperbolic sine function and a Morlet wavelet (Eq. (3.12) and (3.13)). Note that, number of hidden neurons for each model are assigned using trial and error approach, which obviously could not guarantee optimal structure. Simulations are carried on six real datasets from industry as shown in Table 3.1, where information about model inputs / outputs and training / testing samples are also mentioned (see section 2.4, 3.5.1 and Appendix A.5 for further details on the datasets). Three kind of issues related to time series application are considered: two "input-output" approximation problems, two one-step ahead prediction problems, and two *msp* problems. To compare performances, three criteria are selected.

1. Computation time to learn the dataset for a single trial ($Time$).

2. Model fitting accuracy is evaluated by the coefficient of determination ($R2$) that should be close to 1 and coefficient of variation of the Root Mean Squared Error ($CVRMSE$) that should be as low as possible (close to 0 - it is often expressed in percentage) (see Appendix A.4).

3. Network complexity is reflected by the number of hidden neurons ($Hid - nodes$).

With each approach (i.e., ELM, SWELM, LM-SLFN, ELMAN), *50* simulations are performed on a particular dataset and the best results are summarized.

TABLE 3.1: Datasets to benchmark performances for time series application

Data	Description	Input	Output	Train	Test
Approximation problems					
Pump [170] Appendix A.5	Condition monitoring	Root mean square, Variance	Fault code	73 (samp.)	19 (samp.)
CNC [222] section 3.5.1	Condition monitoring	Max force, Cutting amp. Amp. Ratio, Avg. force	Tool wear	C33, C09, C18 (450 samp.)	C18 (165 samp.)
One step-ahead prediction problems					
Ind. Dryer [1] Appendix A.5	Predict temperature	Fuel flow, Fan speed Flow raw material Bulb temp. y_t	Bulb temp. y_{t+1}	500 (samp.)	367 (samp.)
Hair dryer [2] Appendix A.5	Predict temperature	Voltage of device x_t Air temp. y_t	Air temp. y_{t+1}	500 (samp.)	500 (samp.)
Multi-step ahead prediction problems					
NN3 [6] Appendix A.5	Time series forecast	Time series (4 reg.) $(x_t, x_{t-1}, x_{t-2}, x_{t-3})$	Same series $x_{t+1 \to t+18}$	51, 54, 56, 58, 60, 61, 92, 106	All series (18 samp.)
Turbofan [175] section 2.6.1.1	Predict degradation	Degradation series 3 reg. (x_t, x_{t-1}, x_{t-2})	Same series $x_{t+1 \to t+H}$	90 engines	5 engines $H \in [103, 283]$

3.4.2 First issue: approximation problems

In case of approximation or estimation problems, real datasets from two different degrading machineries are used for condition monitoring task. The first dataset is from a *Carnallite surge tank pump*, which is used to approximate fault codes, where the second dataset is from a CNC milling machine to estimate wear of degrading cutting tools (Table 3.1). Results in Table 3.2 show that SW-ELM for both test datasets performs best approximations, with a compact structure, and it requires less learning time.

1. ELM has clear superiority of fast learning times ($5.8e^{-004}$ and $5.0e^{-004}$ sec) for both datasets. In addition, SW-ELM and ELM can learn much rapidly as compared to LM-SLFN or ELMAN network. This is mainly due to the advantages of single-step learning phase of ELM based models, whereas LM-SLFN and ELMAN network required additional 50 epochs for the *Pump* dataset and 10 epochs for *CNC* dataset to achieve better accuracy performances.

2. The accuracy of SW-ELM show a best fitting among all methods for both datasets as measured by $R2$ i.e., ($R2$=0.96 and $R2$=0.92). Comparative plots representing approximation of fault codes and tool wear estimation are shown in Fig. 3.7 and Fig. 3.8, where model fitting errors are also depicted for a better understanding of results. Note that, for more clarity, plots of only two methods with higher accuracies are compared.

3. Lastly, when considering model complexity factor, the number of hidden neurons required by LM-SLFN and ELMAN network appears to be twice (i.e., 30 hidden nodes) as compared to ELM based models when applied to Pump dataset. However, same network complexity of 4 hidden nodes is sufficient for all methods with CNC dataset.

Table 3.2: Approximation problems: results comparison

	Pump				**CNC**			
Method	Hid-node	Time (s)	Epoch	R2	Hid-node	Time (s)	Epoch	R2
SW-ELM	**15**	6.5e-004	–	**0.96**	**4**	7.7e-004	–	**0.92**
ELM	15	**5.8e-004**	–	0.94	4	**5.0e-004**	–	0.77
LM-SLFN	30	1.02	50	0.79	4	0.22	10	0.80
ELMAN	30	8.88	50	0.81	4	0.21	10	0.77

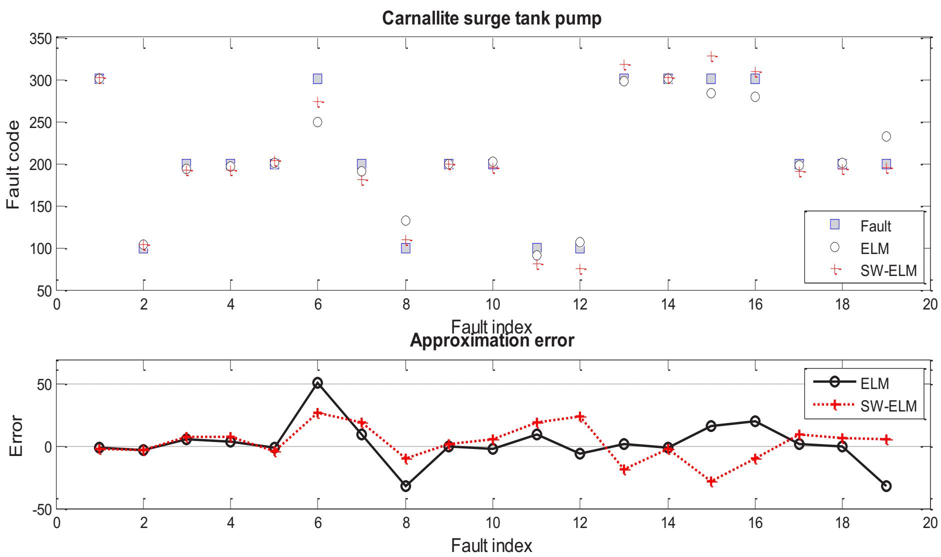

Figure 3.7: Fault code approximations and corresponding errors

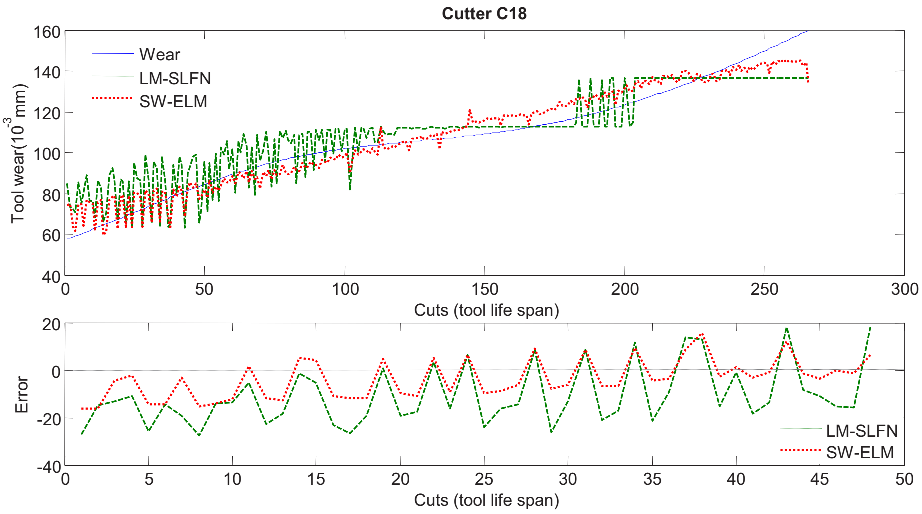

FIGURE 3.8: Tool wear estimation and corresponding errors

3.4.3 Second issue: one-step ahead prediction problems

In case of one-step ahead prediction issue, the datasets used are from an *industrial dryer* and a *mechanical hair dryer*. In both cases, the aim of tests is to predict temperature variations (Table 3.1). In order to evaluate model performances, same criteria as for approximation problems are used. Comparative performances from all models are summarized in Table 3.3. Again, results indicate that SW-ELM shows improved performances over ELM, LM-SLFN and ELMAN networks.

1. ELM can still perform faster learning (0.0012 and $3.4e^{-004}$ sec) with both datasets, even if the learning data size is increased (500 samples). Like in previous issue, the learning times of SW-ELM is still close to ELM.

2. The accuracy indicator of SW-ELM shows a higher prediction performance i.e., $R2 = 0.85$ with dataset of an *industrial dryer*. However the accuracy of SW-ELM and ELM are same ($R2 = 0.944$) in case of second data of a *mechanical hair dryer*. It should be noted that, in the second test all methods show good prediction performances. As for illustration, comparative plots of predictions with only two methods are shown in Fig. 3.9 and Fig. 3.10 by considering model accuracy, where predictions errors are also depicted. In Fig. 3.9, one can note that SW-ELM catches better non-linearity from the signal since the error is quite constant among all predictions.

3. ELM based models show a more compact structure (20 hidden nodes) as compared to LM-SLFN and ELMAN network (30 hidden nodes) when applied to *industrial dryer* data. However, again small network complexity of 4 hidden nodes is sufficient for all methods with the second data of *mechanical hair dryer*.

TABLE 3.3: One step-ahead prediction problems: results comparison

	Industrial Dryer				Mechanical hair dryer			
Method	Hid-node	Time (s)	Epoch	R2	Hid-node	Time (s)	Epoch	R2
SW-ELM	**20**	0.0024	–	**0.85**	**4**	6.1e-004	–	**0.944**
ELM	20	**0.0012**	–	0.66	4	**3.4e-004**	–	**0.944**
LM-SLFN	30	1.03	50	0.81	4	0.21	10	0.9434
ELMAN	30	8.9	50	0.80	4	0.20	10	0.9434

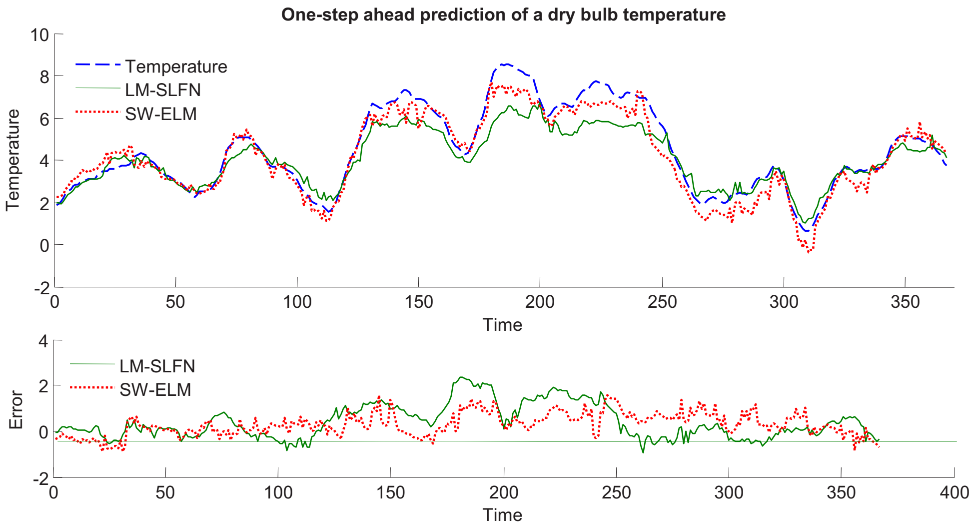

FIGURE 3.9: One-step ahead prediction of bulb temperature and corresponding errors

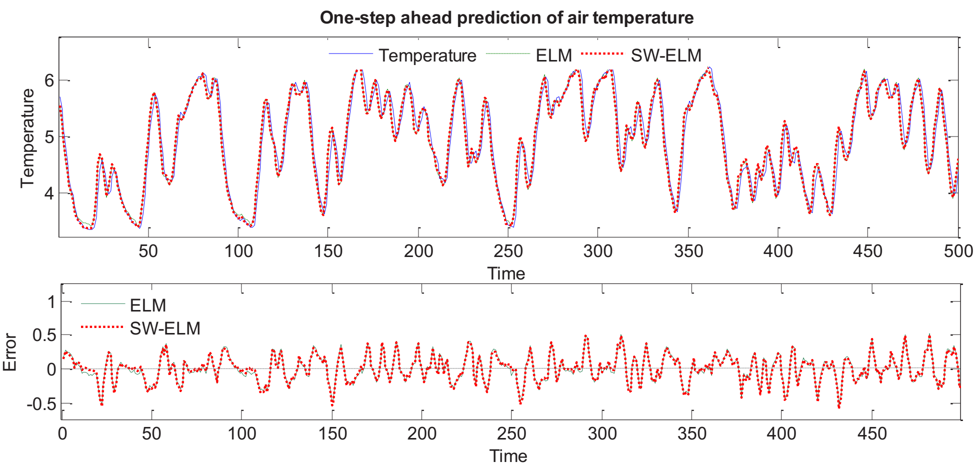

FIGURE 3.10: One-step ahead prediction of Air temperature and corresponding errors

3.4.4 Third issue: multi-steps ahead prediction (*msp*) problems

In case of *msp* problems, again two data sets are used to test models performances using iterative approach (section 2.6.2.1). The first test data, are from *NN3 competition*, where horizon of prediction is defined as last 18 values of each time series for test. The second dataset are from NASA data repository, and is composed of run to failure sensor measurements from degrading *Turbofan engines*, where the horizon of prediction is from a current time **tc** up to end of degradation (see Table 3.1 data record $train - FD001.txt$ is used to train and test the models). Similarly, like in previous issues, 50 trials are performed and best results are presented. The *msp* accuracy of each model is assessed by computing $CVRMSE$, that enables evaluating relative closeness of predictions, which should be as low as possible. Whereas, computational time (for a single trial) and network complexity are evaluated similarly as in previous cases. Models performances are summarized in Table 3.4.

A comparative plot is shown in Fig. 3.11 with two models having higher accuracies of *msp* with *NN3 competition* data for time series 61. Averaged performances of prediction models for all randomly selected time series (51, 54, 56, 58, 60, 61, 92, 106) are given in Table 3.4.

In case of *Turbofan engines* dataset, the comparative plot of two models on test-engine 5 is shown in Fig. 3.12, where SW-ELM explicitly shows good prediction capability over long horizon. Averaged performances of prediction models for all tests are given in Table 3.4.

1. Among all methods, ELM requires less computation time for learning with both datasets (i.e., $5.5e^{-004}$ and 0.004 sec), where SW-ELM also shows rapid learning behavior closer to ELM. As explained before, this is mainly due to the one-pass learning phase.

2. Most importantly, accuracy indicator ($CVRMSE$) of SW-ELM is highest in comparison to others models for both datasets (i.e., *NN3 competition* and *Turbofan engines*). A synthetic but explicit view of that is given in Fig. 3.13. Note that such performances of SW-ELM are obtained with a same structure like other methods (see point 3).

3. The network complexity of all models is same with both datasets (i.e., 30 hidden nodes and 3 hidden nodes). Note that for all test on 5 *Turbofan engines* even with more complex models of ELM, LM-SLFN and ELMAN network poor prediction performances where obtained, therefore the only complexity of 3 hidden nodes is compatible with the data.

Finally, again SW-ELM enables a good compromise among model accuracy, learning time and network complexity.

TABLE 3.4: Multi-setp ahead prediction problems: results comparison

	NN3 (avg.)				**Turbofan** (avg.)			
Method	Hid-node	Time (s)	Epoch	CVRMSE	Hid-node	Time (s)	Epoch	CVRMSE
SW-ELM	**30**	0,0014	–	**10.83%**	**3**	0.006	–	**0.042%**
ELM	30	**5.5e-004**	–	11.06%	3	**0.004**	–	0.0578%
LM-SLFN	30	0.20	10	11.51%	3	0.72	10	0.0570%
ELMAN	30	0.45	10	**10.83%**	3	0.75	10	0.0570%

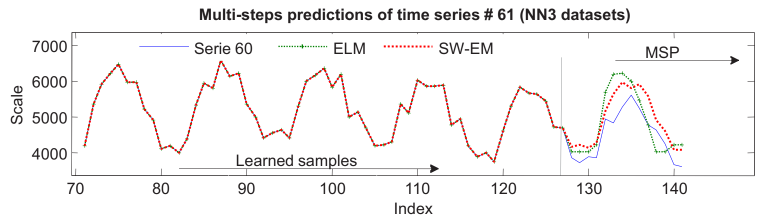

FIGURE 3.11: *msp* of time series 61 (NN3)

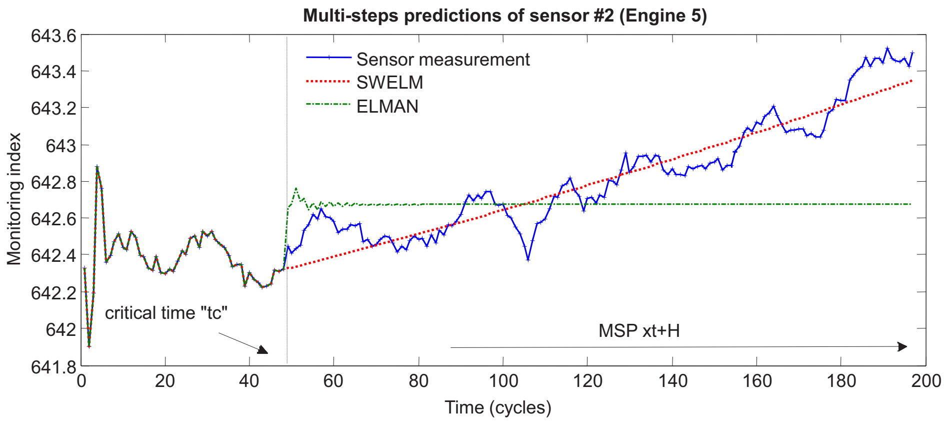

FIGURE 3.12: *msp* of sensor 2 (Engine 5)

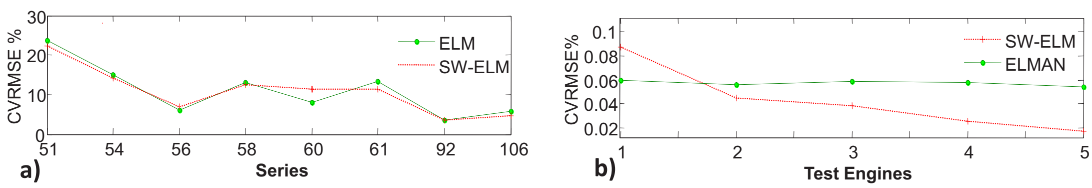

FIGURE 3.13: Accuracy of: a) *msp* for 8 NN3 series, b) *msp* for 5 Turbofan engines

3.5 PHM case studies on real data

3.5.1 First case study: cutters datasets & simulation settings

3.5.1.1 Industrial significance

To investigate suitability of the proposed approach (i.e., SW-ELM and Ensemble), a case study is carried on a high speed CNC machine for tool wear monitoring task. The high speed machining process has become the most important and cost-effective means in manufacturing industry to produce products with high surface quality [218]. This process is performed in a dynamic environment and under diverse conditions, where a cutting tool acts as a backbone of machining process. Failure of cutting tool can affect the product quality and cause machine down-time [224]. Mainly, during cutting treatment of the metal work-piece, the cutter wear can increase due to irregular varying loads on its flutes that are always engaged and disengaged with the surface of work-piece [51]. This may result in increased imperfections in the work-piece surface i.e., dimensional accuracy of finished parts. Due to complex nature of cutting tool (flute) wear phenomena, the uncertainties of machining conditions and the variability in cutting tool geometry make the modeling of tool wear a complicated task [188]. Most CNC machines are unable to detect tool wear on-line. Therefore, the estimate of cutting performance is usually performed through indirect method of tool condition monitoring (without shutting down the operating machine) by acquiring monitoring data that can be related to suitable wear models [224]. As for some examples, [127] used acoustic emission signal, [121, 218] used force signals, [79] used vibration signals. Among all those approaches, cutting force signals are preferred for modeling: force measurements are easy to manipulate, have been found as suitable indicators in association with tool wear (due to high sensitivity), and are considered as the most useful to predict cutting performance (due to good measurement accuracy) [218, 223]. As for an example, consider the cutting force signals in three axes (Fx, Fy, Fz) are depicted in Fig. 3.14.

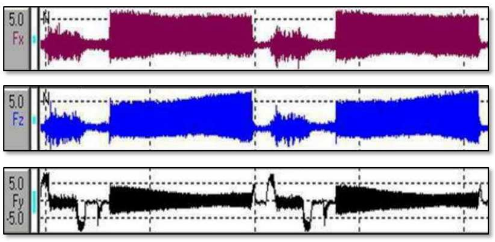

FIGURE 3.14: Cutting force signals in three axes (Fx, Fy, Fz) [223]

3.5.1.2 Experimental arrangements

The real data were provided by SIMTECH Institute in Singapore, where a high-speed CNC milling machine (Roders Tech RFM 760) was used as a test-bed. In the machining treatment, the spindle speed was set to 10360 rpm. The material of work-piece used was Inconel 718. Also, 3 tungsten carbide cutters with 6 mm ball-nose / 3-flutes were used in the experiments. To achieve high quality surface finish, the metal work-piece was made via face milling to remove its original skin layer of hard particles. During milling, the feed rate was 1.555 mm / min, the Y depth of generated cuts was 0.125 mm and the Z depth of cuts was 0.2 mm [136]. Authors of experiments recorded CM data in terms of tool wear represented by force, acoustic and vibration properties of cutting operation. In our case, the cutting force signal has been used instead, because of its high sensitivity to tool wear and good measurement accuracy [222]. A dynamometer was positioned between the metal work-piece and the machining table in order to acquire cutting forces in the form of charges, that are converted to voltage (voltage signal was sampled by PCI 1200 board at 50 kHz). The acquired data were composed of 315 cuts made by 3 different cutters that are utilized in experiments, namely C18, C33 and C09. During experiments, the cutting operation was stopped after each cut and tool wear was measured via Olympic microscope. The cutting force is mainly affected by cutting conditions, cutter geometry, coating and properties of work-piece. Also, considering complications of tool wear modeling, it is important to highlight the characteristics of all cutters that are used. That is, cutting tools C33 and C18 had the same geometry but different coatings, while cutting tool C09 has its own geometry and coating.

Table 3.5: Type of cutting tools used during experiments

Cutters	**Geometry**	**Coating**
C18	Geom1	Coat1
C33		Coat2
C09	Geom2	Coat3

3.5.1.3 Tool wear estimation - methodology

The methodology described below is depicted in Fig. 3.15.

Data acquisition and pre-processing - As explained before, force signals are utilized to build the behavior models of degrading cutters. Mainly, 16 features were derived from force signals and a subset of 4 features are selected to learn behavior models (see Fig. 3.16) [136]. The detail of feature extraction / selection can be found in [126, 222].

Tool wear modeling - All the collected data were stored so that the complete database represents records from 3 cutters, whereas each record contained collected feature subset and its measured wear for 315 cuts. The database is used to establish the model which enables to generate a relation between input features and measured wear output. For that purpose, the SW-ELM is used to predict the tool wear, and it is benchmarked with ELM model. Note that, due to inherit complexity of tool degradation

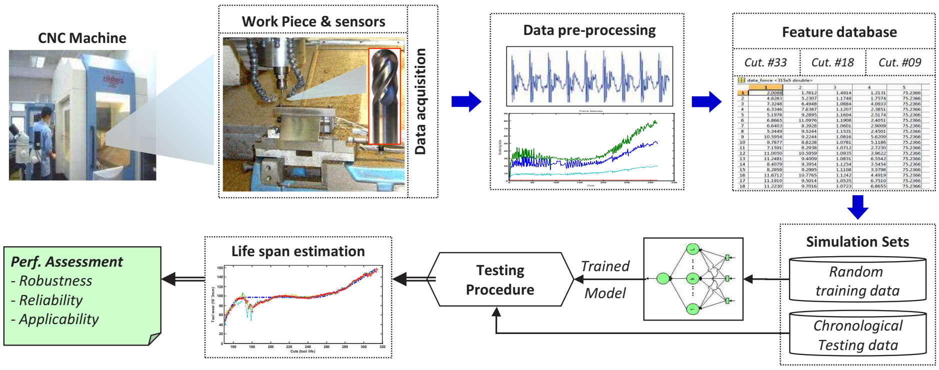

FIGURE 3.15: Cutting tool wear estimation - methodology

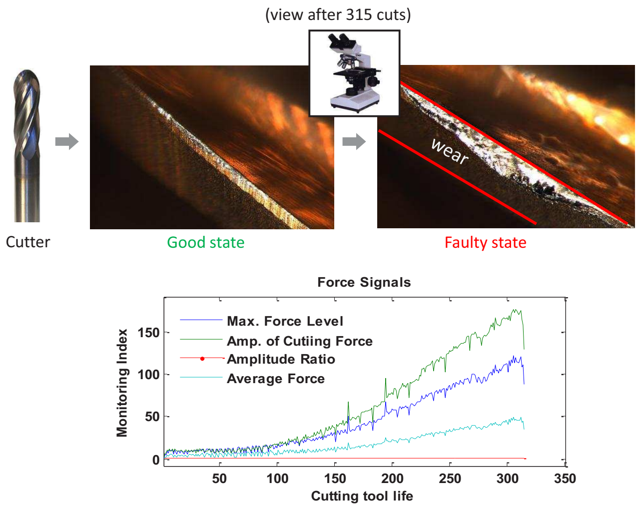

FIGURE 3.16: Force features extracted from 315 cuts

phenomena, it is hard to deduce complex relationships between features and measured wear.

Simulations - The simulations presented in next sections aim at estimating the life span of degrading cutters (cuts that can be made safely). Learning and testing datasets have been defined accordingly: for each model (SW-ELM, ELM) same datasets from the database are used to learn the association between input features and machinery behavior. Besides that, to improve the generalization performance of ELM as well, and considering the discussion in section 3.2.3.2, small random weights of input-hidden nodes are initialized for ELM model i.e., $[-0.1, +0.1]$, rather than default values $[-1, +1]$. Provided a learning data set for a model (ELM or SW-ELM), 100 trials are performed and model performances are averaged for a specific network complexity (for 4 to 20 hidden nodes). It should be noted that model performance for a single trial is equal to an accumulated performance from 10 different simulations: for which data subsets (of cutting tools) are learned after random permutation (to introduce variation in data), while test are performed on remaining data in chronological order (see Fig. 3.15). For clarity, detailed results are furnished with different model complexity (hidden neuron), data variability / model steadiness and related tool wear estimation.

Performance criteria - To discuss the robustness, reliability, and applicability of the models (and according to the analysis scheme proposed in Fig. 1.15), performances of the models are assessed in terms of accuracy and network complexity. More precisely, 3 criteria are computed: the Mean Average Percent Error of prediction ($MAPE$) that should be as low as possible, the coefficient of determination ($R2$) that should be closer to 1 (see Appendix A.4), and the learning time ($Time$). The obtained results are divided into two different cases for better understanding (section 3.5.1.4 and 3.5.1.5).

3.5.1.4 Results discussion: robustness and applicability

This case aims at evaluating the prediction accuracy of each model by assessing its robust performance with a single cutting tool (with data of same attributes) to establish a model i.e., C18, C33 and C09. For learning, data set from a particular cutting tool is created by randomly selecting 150 data samples, while rest of the data (165 data samples) are presented in chronological order to test the accuracy of the learned model (Fig. 3.17). This procedure is repeated 10 times for each model-cutting tool couple and is considered as a "single trial". Comparative analysis on robust tool wear prediction performances from all model-cutting tool couples are shown in Table 3.6. Among all model-cutting tool couples with varying learning data, SW-ELM model has more robust performance on tests as compared to ELM, even when small learning data are presented to both models. However, the average learning time of ELM is slightly faster than SW-ELM. The detailed simulations from the obtained results are presented in Fig. 3.18. In brief, Fig. 3.18a, shows an average accuracy ($R2$) performance for 5 different network complexities, where best results are achieved by SW-ELM for tests on cutter C09 and C18 (with no. hidden nodes 16 and 12). Considering these results, Fig. 3.18b compares the steadiness of both models (SW-ELM and ELM) for 100 trials. One can see that SW-ELM is more robust to input variations as its test accuracy ($R2$) is steady for 100 trials with tests on both cutters i.e., C09 and C18. In addition, Fig. 3.18c depicts a comparison on average results of tool wear estimation on cutter C09.

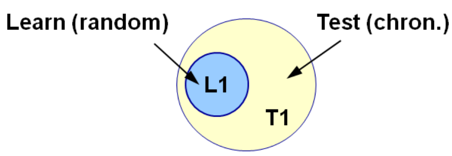

FIGURE 3.17: Robustness: learning and testing with one cutter

TABLE 3.6: Robustness and applicability for a single cutter model

Cutter 09	SW-ELM	ELM
Hidden nodes	**16**	16
Activation function	asinh & Morlet	sig
Training time (s)	0.0009	**0.0005**
$R2$	**0.824**	0.796

Cutter 18	SW-ELM	ELM
Hidden nodes	**12**	12
Activation function	asinh & Morlet	sig
Training time (s)	0.0007	**0.0004**
$R2$	**0.955**	0.946

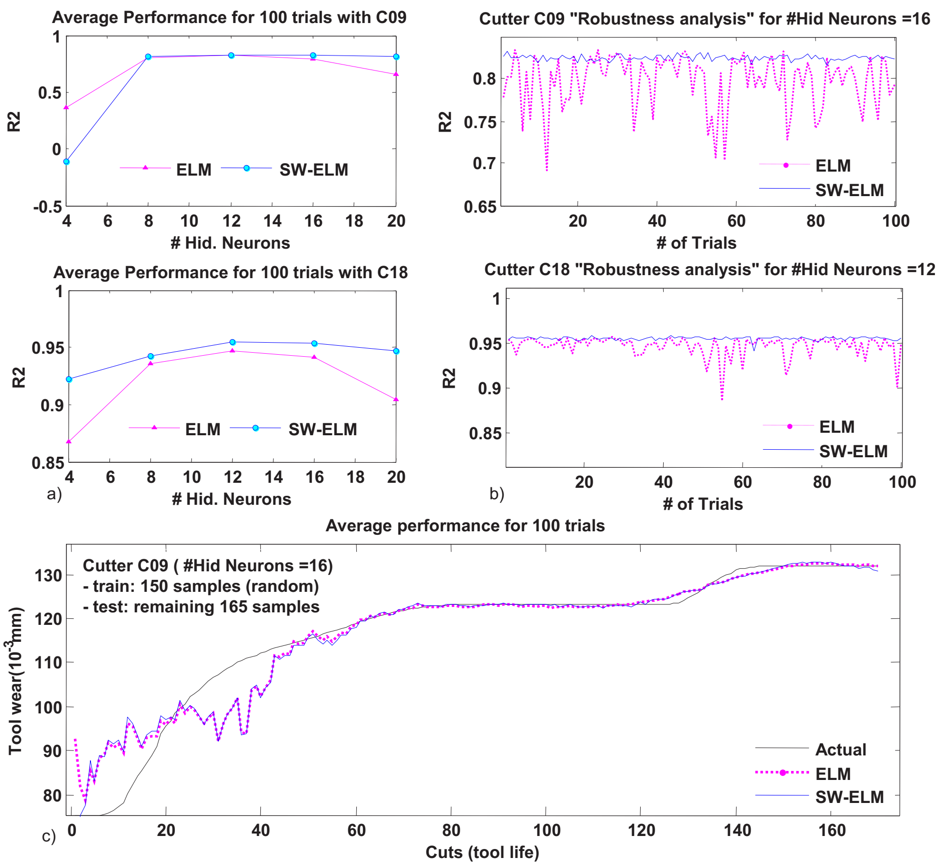

FIGURE 3.18: Robustness analysis: partially unknown data for a "single cutter" model

3.5.1.5 Results discussion: reliability and applicability

Reliability / partially known data - This case aims at evaluating the reliability performances of each model (ELM and SW-ELM) when data from multiple cutters with different attributes (geometrical scale and coating - see Table 3.5) are utilized to establish a model. In order to build a "multi-cutters" model, a partial data set of 450 samples from all cutters are presented in random order for learning and 165 data samples from any of these cutters are presented for test (Fig. 3.19). This procedure is repeated 10 times for each multi-cutters model and is considered as "single trial" like in the previous case. A comparative analysis on reliable tool wear prediction performances (on data of different attributes) from all model-cutting tool couples is shown in Table 3.7.

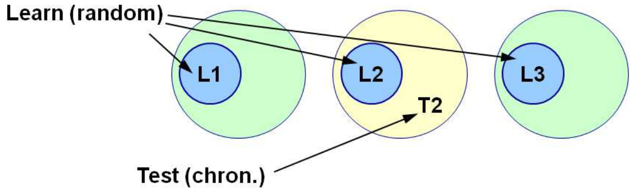

FIGURE 3.19: Reliability: learning and testing with three cutters

TABLE 3.7: Reliability and applicability for 3 cutters models

Train: C33, C18, C09 Test: C18	**SW-ELM**	**ELM**
Hidden nodes	**20**	20
Activation function	asinh & Morlet	sig
Training time (s)	0.002	**0.001**
$R2$	**0.837**	0.836

Train: C33, C18, C09 Test: C33	**SW-ELM**	**ELM**
Hidden nodes	**16**	16
Activation function	asinh & Morlet	sig
Training time (s)	0.002	**0.0009**
$R2$	**0.847**	0.80

It can be observed from results that, even if the learning data are increased, both methods are still fast. However, in this case, when looking at average learning times for both tests, again ELM requires slightly less time for learning for same complexity of models. As far as accuracy ($R2$) performance is concerned, SW-ELM shows superiority and consistency on data of different attributes as compared to ELM, thanks to improvements made in Algorithm 2. Like the previous case, detailed simulations from the results are presented in Fig. 3.20. In brief, Fig. 3.20a, shows an average accuracy ($R2$) performance for 5 different network complexities, where best results are achieved by SW-EM for tests on cutter C18 and C33 (with no. hidden nodes 20 and 16). Considering these results, Fig. 3.20b compares the steadiness of both models (SW-ELM and ELM) for 100 trials. One can see that SW-ELM is more stable to input variations, as its accuracy ($R2$) is consistent for 100 trials with tests on both cutters i.e., C18 and C33. Finally, Fig. 3.20c shows a comparison on the average results of tool wear estimation on cutter C33.

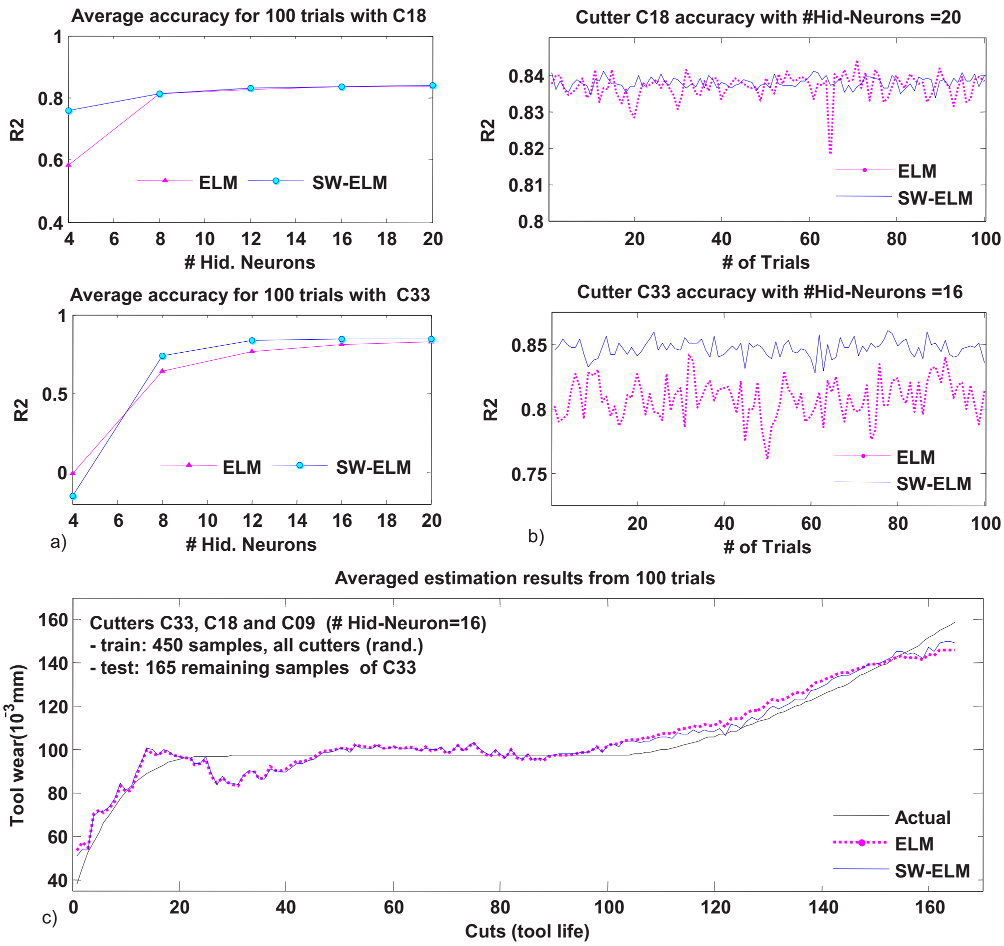

FIGURE 3.20: Reliability analysis: partially unknown data for "multi-cutter" model

Reliability / totally unknown data - This case aim at evaluating the reliability performances of a model (ELM or SW-ELM) when unknown data are presented for tests. For this purpose, test performances of each model are assessed by establishing a reference model from learning complete life span of two different cutting tools and testing its wear prediction capability on data from another cutter that is totally unknown to the model (and with different geometric and coating characteristics). Therefore learning data of 630 samples from two different cutters are presented randomly to train the model, whereas 315 data samples from a third cutter are presented in a chronological order for testing purpose Fig. 3.21, (i.e., performed only once for a trial, which is different from previous cases). The procedure is repeated for 100 trails for a given model complexity (no. of hidden neuron 4 to 20). Averaged performances from multi-tools model learned with cutting tool C33 and C09 and tested on cutter C18 are shown in Table 3.8.

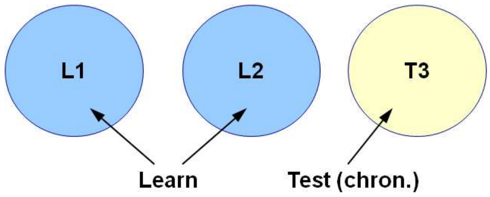

FIGURE 3.21: Reliability: learning and testing with unknown data

TABLE 3.8: Reliability and applicability for unknown data

Train: C33 & C09 Test: C18	**SW-ELM**	**ELM**
Hidden nodes	**4**	4
Activation function	asinh & Morlet	sig
Training time (s)	0.0009	**0.0004**
$R2$	**0.701**	0.44

It can be observed from the results that SW-ELM shows better reliability performance as compared to ELM, but requires slightly more time to train on data. However, it should be noted that, when cutters C33 and C09 are presented as unknown data to the prediction models, the prediction accuracy of each approach decreased to a poor level. Indeed, results of SW-ELM for cutter C18 are better because the scale of its wear pattern exists between cutters C33 and C09 as wear increases with respect to number of cuts (Fig. 3.24). In addition, as stated before, cutters C33 and C18 have same geometry while cutter C09 has its own geometry. Finally, detailed simulations for performance comparison on model stability (on 100 trials) and tool wear estimation are shown in Fig. 3.22 and Fig. 3.23, which indicates that the reliability still needs to be improved when unknown data of different attributes (context) are presented to the model.

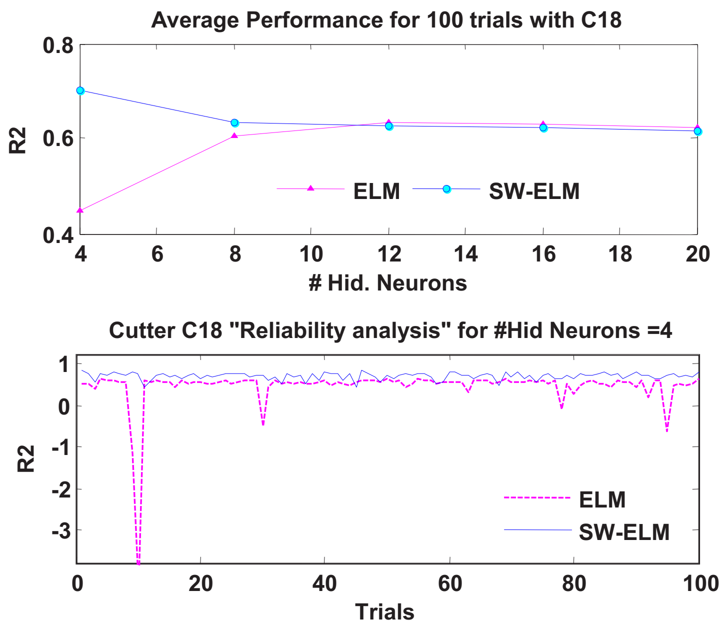

FIGURE 3.22: Reliability analysis with totally unknown data

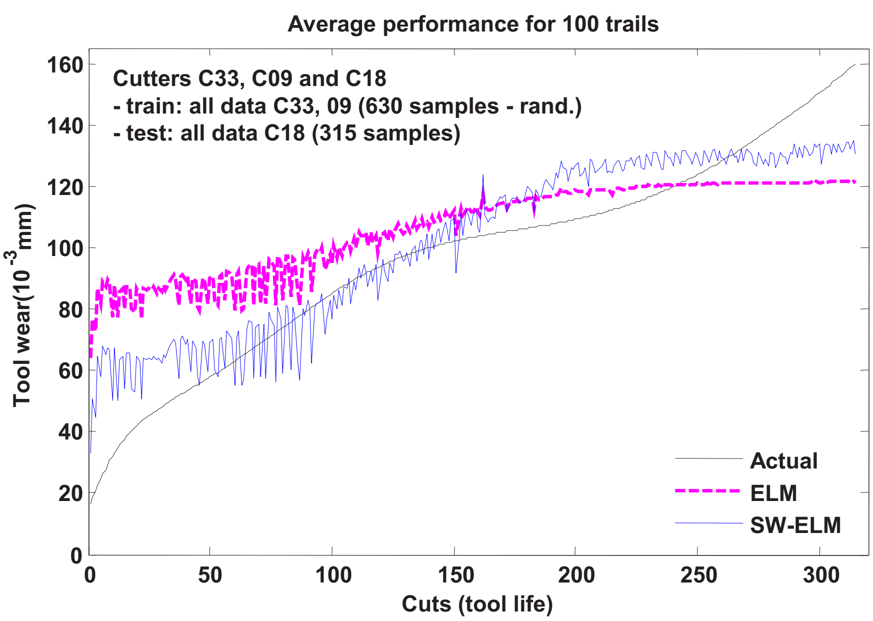

FIGURE 3.23: Estimation with multi-cutter model on unknown data

3.5.1.6 SW-ELME and its reliability performance

Simulation settings - Considering the superiority of SW-ELM over ELM, the ensemble strategy is only performed with SW-ELM approach. In this strategy, the initial step is to determine structure of a single SW-ELM model. That is, a structure of model which results satisfactory performances with minimum possible complexity. Following that, multiple SW-ELM models of same complexity are integrated to produce averaged output and to improve accuracy of estimates (Table 3.9). However, to be more explicit on that, each individual model is learned with same dataset, but initialized with different parameters, i.e., weights and biases.
Tests are performed on three different cutters i.e., C18, C33 and C09, where each individual SW-ELM model is learned with 680 samples and tested on unknown data of 265 samples. For example, learning set includes cutter data C18, C33 and C09 (up to 50 samples) and testing set is composed of remaining 265 samples from C09 (starting from current time **tc=50** till the last cut **315**) that are totally unknown to the model.

TABLE 3.9: Tool wear estimation with SW-ELME - Settings

SW-ELME	**Settings**
No. of models in ensemble	100
Hidden neuron per model	10
Actv. functions	asinh & Morlet
Current time **tc**	50

Wear limit - Setting a precise wear limit is vital to any prognostics approach. Any assumptions should be avoided for practical reasons. Therefore, it should be precisely assigned for a particular application. For CNC application, tool wear evolution is classified in 5 wear states, "initial wear, slight wear, moderate wear, severe wear and worn-out" [68]. As for all cutters tool life span are same, we use a simple partition based clustering approach, namely, K-means algorithm to cluster their wear patterns (see [208] for clustering details). The benefit of this step is to identify worn-out state for all cutters. Therefore, threshold can be set when tool degrades from severe wear to worn-out state (see Fig. 3.24 where max wear is common limit for tests i.e., wear limit is set to 130 $10^{-3}mm$). However, the aspect of wear limits or thresholds is addressed in Chapter 4 with details.

Results and discussion - Results from all cases i.e., C18, C33 and C09 with SW-ELME model are summarized in Table 3.10. The cutting tool life span or remaining life is estimated by setting a common wear limit among all cutters. That is, the number of cuts that can be successfully made until defined wear limit. Finally, 95% CI intervals indicating lower and upper confidence bounds of estimates with an SW-ELME model are furnished for each test case (see Fig. 3.25).

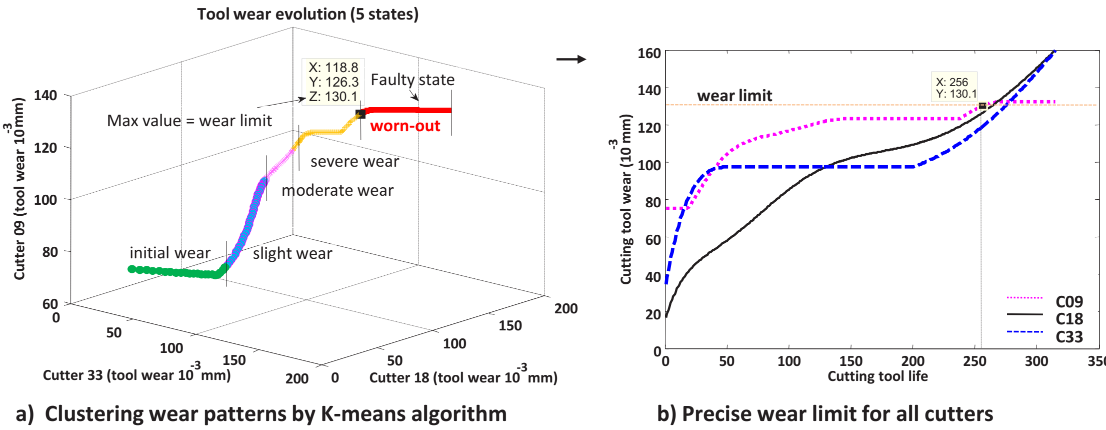

FIGURE 3.24: Tool wear evolution and precise wear limit

TABLE 3.10: Reliability and applicability for unknown data with SW-ELME

Cutter	**C18**	**C33**	**C09**
Time (s)	0.154	0.187	0.187
$MAPE$	9.475%	9.77%	14.20%
$R2$	0.768	0.525	-6.99
Actual cuts	265	278	257
Estimated cuts	287	261	280
Lower 95% CI	257	238	219
Upper 95% CI	291	286	292

It can be observed from the results in Table 3.10, that, due to ensemble strategy and increased data size, the total learning time SW-ELME model is M times greater than an individual SW-ELM, but still remains very short. As far as other performances are concerned, the accuracy of SW-ELME model is surely superior than a single SW-ELM model as indicated in Table 3.10. For example, if a comparison is made between results on cutter C18 in Tables 3.8 and 3.10, the $R2$ is improved from 0.701 to 0.768, where the complexity of each SW-ELM in an ensemble is set to 10 hidden neurons (Table 3.9). Moreover, the overall performances of SW-ELME are still satisfactory on C18 and C33 (with $R2 = 0.768$ and $R2 = 0.525$). However, poor results are obtained for test C09. Obviously, performance of model decreases when the unknown test data has large deviation from learned data or when data are not within the bounds of data used during learning. For clarity, see Fig. 3.24b which enables to see that wear pattern of cutter C18 is between C33 and C09. Besides that, when looking at the CIs, true target value is within the confidence level as indicated by lower and upper bounds in Table 3.10. However, in case of cutter C09 unsatisfactory results are achieved with a wide CI.

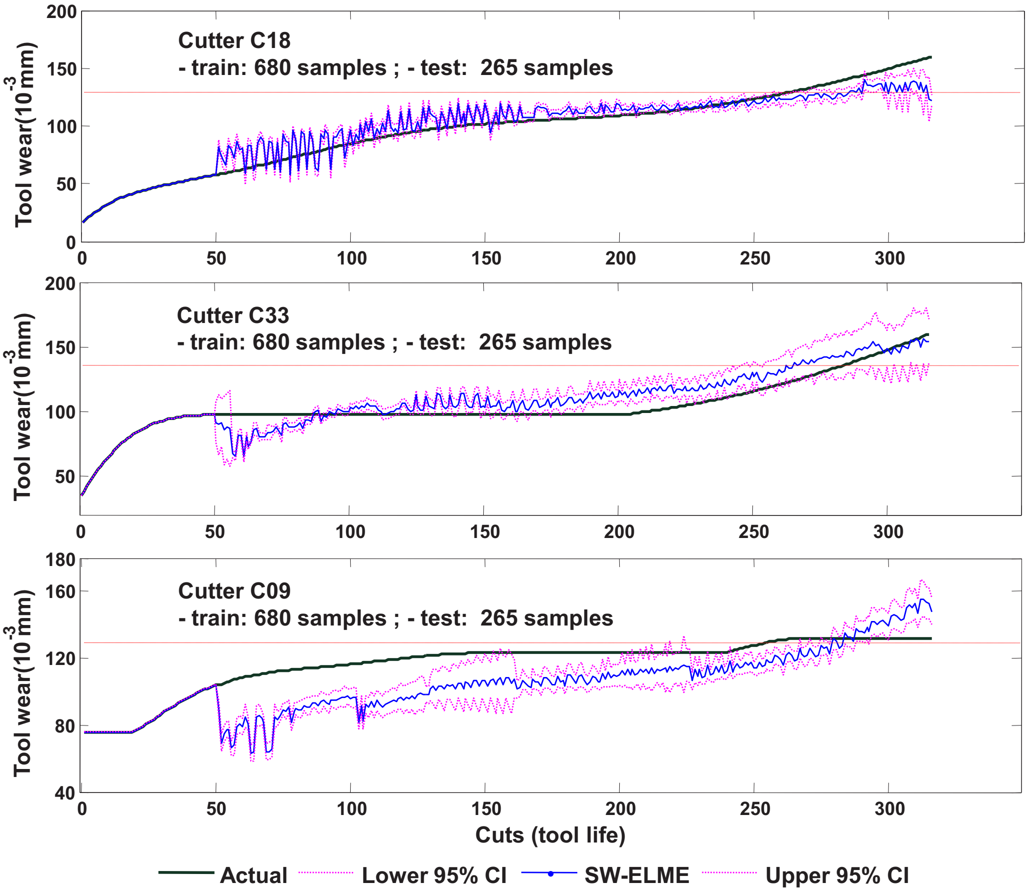

FIGURE 3.25: Cutting tools wear estimation and CI construction

3.5.2 Second case study: bearings datasets of PHM Challenge 2012

Predicting continuous state of a degrading machinery acts as prerequisite of a prognostics model to achieve RUL estimates. In other words, accuracy of prognostics is related to its ability to predict future behavior of equipment. However, for a real application, performances of a prognostics model are closely related to critical issues that limit its applicability, like the horizon of prediction, the expected accuracy, the acceptable uncertainty level, human intervention, or convergence of algorithm. Such aspects should be considered while building a prediction model. Therefore, a case study is performed on IEEE PHM 2012 challenge bearing datasets (section 2.4). The aim of these experiments is not only to show long-term prediction capability SW-ELM, but also to highlight the significance of the cumulative features extraction procedure proposed in section 2.3.1. The performances of SW-ELM are compared for *msp* with cumulative features and classical features (section 2.1.2) using iterative approach as described in section 2.6.2.1.

3.5.2.1 Learning and testing schemes

To perform *msp*, a compact structure SW-ELM with topology 3 inputs, 4 hidden and 1 output nodes is used. Note that, 3 inputs represent regressors from a particular feature (x_t, x_{t-1}, x_{t-2}), where output represents *msp* ($x_{t+1 \to t+H}$), H$\in$ *msp* horizon. To minimize the impact of random parameters on learning and testing tasks, small norms of input-hidden weights and bias are initialized by setting $C = 0.001$ (i.e., for beta value in NW procedure see Algorithm 2). Learning of SW-ELM model is performed on run-to failure data sets from bearings Ber_{1-1} and Ber_{1-2} under operating conditions 1800 *rpm* and 4000 N [5, 144]. In test phase, incomplete data sets from 5 bearings under same operating conditions (i.e., Ber_{1-3} to Ber_{1-7}) are used to predict the RUL (see details about data in section 2.4). In other words, predictions are performed from a current time (of test data) to the end of the bearings life. Assuming that a single model cannot guarantee accuracy of predictions for all tests, 100 SW-ELM models (with different parameters) are learned, and the best model with minimum learning error is selected for testing phase. Note that, the learning of a single model is achieved in a rapid manner in just 0.0063 sec, due to one-step learning phase (section 3.3.2).

3.5.2.2 Results and discussion

For illustration, predictions results with classical and proposed feature extraction approaches are compared for test bearings Ber_{1-3} and Ber_{1-7} in Fig. 3.26. One can qualitatively note the accuracy of predictions with proposed methodology: the figure depicts very minor difference between the actual and predicted trends, where predictions with classical feature show poor results. This conclusion can be extended for all tested bearings (see Table 3.11 for all results): predictions with cumulative features are achieved with high accuracy and generalization even for long prediction horizons like $Ber1-7$ ("757 steps") with $MAPE = 0.46\,\%$. Such results are of good omen for prognostics: since predictions follow trending behavior of features in an accurate manner (as illustrated in Fig. 2.2). The complete approach (cumulative features with SW-ELM) should lead to low uncertainty of RUL estimates.

TABLE 3.11: Long term prediction results for 5 test bearings

Bearing	**Predicted steps**	**MAPE (classical)**	**MAPE (proposed)**
$1-3$	573	2.09 %	1.32 %
$1-4$	289	32.14 %	1.89 %
$1-5$	161	0.38 %	0.186 %
$1-6$	146	0.68 %	0.184 %
$1-7$	757	16.92 %	0.46 %

MAPE: Mean Absolute Percent Error - $[0,\infty[$, Perfect score = 0

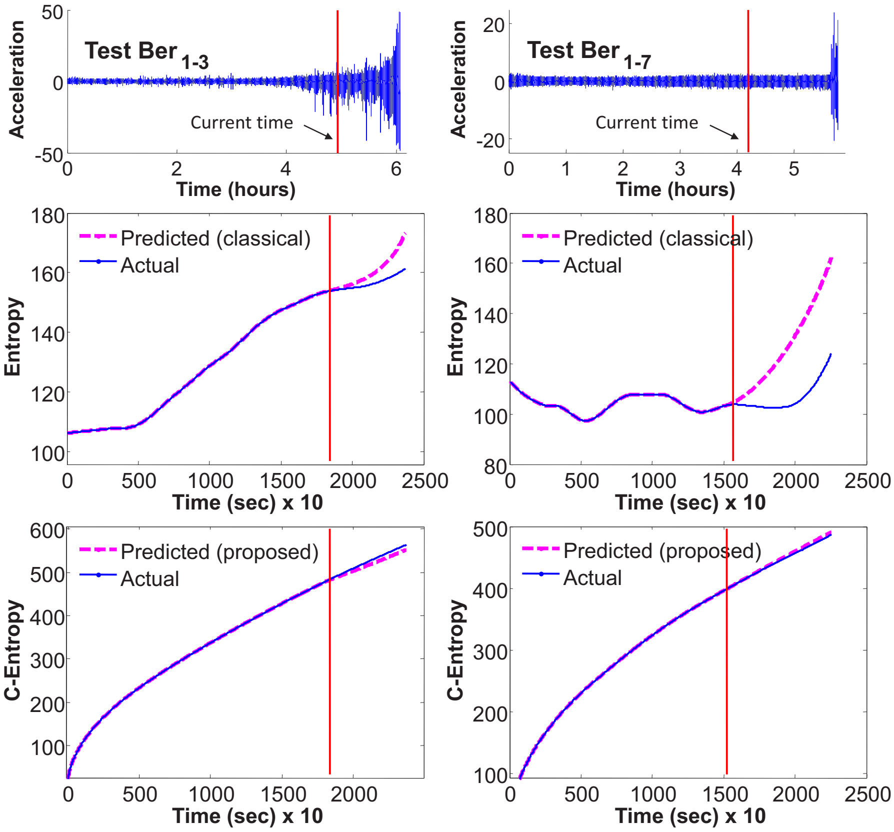

FIGURE 3.26: Examples of multi-step ahead predictions (Classical vs. Proposed)

3.6 Summary

This chapter addresses issues related to time series application and challenges of prognostics modeling, to develop a new prediction model. Therefore, SW-ELM algorithm is proposed as a potential approach for performing "long-term predictions" for a prognostics model (i.e., multivariate degradation based modeling strategy). In addition, to account for uncertainty from data and modeling phase, an ensemble of SW-ELM models is proposed (SW-ELME) to associate confidence to estimates. These, developments are thoroughly investigated on several data sets, and benchmarked with other methods to show improvements. Moreover, two PHM case studies are also considered: firstly, real data of a CNC machine are used to analyze robustness and reliability of SW-ELM as compared to existing ELM. Secondly, bearings datasets of IEEE PHM Challenge 2012 are used to perform long-term prediction with SW-ELM (using cumulative and classical features). CHAPTER 4 focuses on development of an unsupervised classification model, which is integrated with SW-ELM to enhance "multivariate degradation based modeling strategy".

Chapter 4

From predictions to RUL estimates

This chapter presents an enhanced multivariate degradation based modeling strategy for data-driven prognostics. To estimate the RUL of a machinery, prognostics modeling phase requires precise knowledge about failure thresholds (FTs). However, prognostics can have large uncertainty, when there is absence of prior knowledge about actual states of degrading machinery (i.e., unlabeled data).
To address this issue, and also to further improve prognostics performances (i.e., robustness, reliability, applicability), firstly, a new clustering algorithm is proposed to handle unlabeled data and to represent the uncertainty of degradation process namely, Subtractive-Maximum Entropy Fuzzy Clustering (S-MEFC). Following that, a novel approach for prognostics is presented to enhance "multivariate degradation based modeling strategy" to avoid uncertain FTs. In brief, the proposed prognostics model integrates two new algorithms, namely, the SW-ELM (introduced in CHAPTER *3) and the S-MEFC, in order to predict the behavior of degrading machinery, to dynamically assign the FTs, and to automatically determine different states of degradation. Indeed, for practical reasons there is no interest in assuming FT for RUL estimation. The effectiveness of the approach is evaluated by applying it to real dataset of turbofan engines from PHM challenge 2008.*

4.1 Problem addressed: dynamic threshold assignment

4.1.1 Failure threshold: univariate approach and limits

Forecasting future of complex machinery is a complicated task, which is encountered in many PHM applications for industry [166]. Mainly, for PHM applications, prognostics models are applied to perform "long-term predictions" of continuous observations followed by a failure thresholds (FT) to detect faulty state, and to estimate the RUL of the degrading machinery. The FT does not necessarily indicate complete failure of the machinery, but beyond which risk of functionality loss [171]. As discussed in section 1.1.2, in a classical manner the RUL estimate is based on the study of one-dimensional signal: RUL is computed between a current time to initiate prediction and the time at which degrading signal passes the FT or a limit on damage level. The assessment of system's health requires the threshold to be tuned precisely which can be a practical limitation especially in situations when the signal does not represent any physical meaning [166]. Therefore, uncertainty of RUL estimates is not only due to prediction, but to FTs as well. The uncertainty of FT can lead to large RUL errors. According to [166], to ensure reliability of RUL estimates, the use of multidimensional degradation signals is preferred rather than one-dimension signal. Practically, this restricts the use of classical thresholding techniques, where specification of FT is a critical issue and there is lack of standard approaches for this purpose [116, 155]. In other words, the use of multidimensional degrading signals makes thresholding techniques more complicated.

4.1.2 Failure threshold: multivariate approach and limits

With multivariate degradation based modeling strategy for RUL estimates, machine degradation can be represented by continuous and discrete states (section 1.3.3.3), where continuous state shows value of degrading signal, while discrete states depict fault mode. Therefore, to perform prognostics with multivariate degradation based modeling, it is required to build an efficient tool to predict continuous states of signals (that represent features) and a classifier to estimate discrete states reflecting the probable states of degradation. Most importantly with this approach, it is not only the question of performing good predictions, but also to stop prediction process at right time, which is the task of a classifier that sets FTs. Describing the problem of prognostics by means of continuous states and discrete states, is new in PHM community and it's only addressed in few publications [108, 164, 166]. The use of discrete states not only avoids classical way of assuming thresholds, but can also benefit in reducing uncertainty of RUL estimates. However, works presented in [108, 164, 166] have a common drawback that the number of discrete states are pre-assumed for continuous observations and the RUL estimates are achieved accordingly. This assumptions is not realistic in case of degrading machinery, because, the behavior of each machine can differ according to operational environment. Even if two identical machines are exposed to similar operating conditions, most probably, their degrading behavior would be different, and each could represent different level of degradation. In such situations, it can be quite challenging or even impossible for previous approaches (that pre-assume fault modes or discrete states) to assign FTs, which limits their applicability.

4.1.3 Toward dynamic failure threshold

During operation each machine can have different states or degradation levels toward failure or faulty condition [147, 44]. In other words, different failure mechanisms have different trending parameter which requires FT to be assigned dynamically according to a particular case. However, prognostics can be quite challenging when few knowledge or previous experiences on degradation process are available [65], or when there is absence of ground truth. In other words, prognostics with data-driven models can be difficult, as one has to handle unlabeled data with no *prior* knowledge about actual states or transitions between these states (see Fig. 4.1). According to above discussions, two key issues can be pointed out.

- How to manage unlabeled multidimensional data to determine states of degrading machinery?
- How to dynamically assign FTs without any *prior* knowledge?

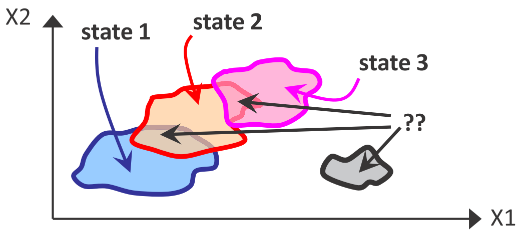

FIGURE 4.1: Transition from healthy to degrading state [75]

To enhance multivariate degradation based modeling strategy and to handle unlabeled data, an unsupervised classification procedure is necessary to assess the health of a degrading machinery. Therefore, in order to dynamically set the FTs, the classifier should have the capability to find hidden structure in an unlabeled data that represent degradation from healthy to faulty state, and secondly, the data classification should clearly show the transition between states. The important aspects of unsupervised classification are discussed in detail as follows.

4.2 Clustering for health assessment: outline & problems

4.2.1 Categories of classification methods

Basically for PHM, the main aim of classification phase is to assess the health state of the machinery. This task requires multidimensional data (variables or features) in order to determine the health of monitored machinery. Briefly, classification methods are of four types: supervised, unsupervised, semi-supervised and partially supervised (Fig. 4.2). In context to PHM, the case of supervised classification is considered when the data are accompanied with a ground truth (or labels), which indicates the actual

structure of data classes. Therefore, data are already labeled and the health states are known. An unsupervised classification is applied when there is absence of ground truth (or labels): unsupervised classification (or data clustering) to identify hidden data structure is only based on data from sensors. The third category, of semi-supervised classification is applied when only few data points have their identity represented by labels for a specific data class [75]. Lastly, the fourth category of partially supervised classification is applied, when the class memberships of learning data are partially known and are expressed as imprecise and or / uncertain soft labels (or belief mass) [50], (for example see recently published article [163]).
Among these categories, supervised approaches appear to be irrelevant for real PHM applications in the absence of ground truth. Even the semi-supervised and partially supervised approaches can have limited applicability as they assume that ground truth can be known with limited uncertainty. Therefore, an underlying "data clustering" (or an unsupervised classification) step is often required to assess the health for a PHM application. However, unsupervised techniques still have some limits, that are highlighted in the following topic.

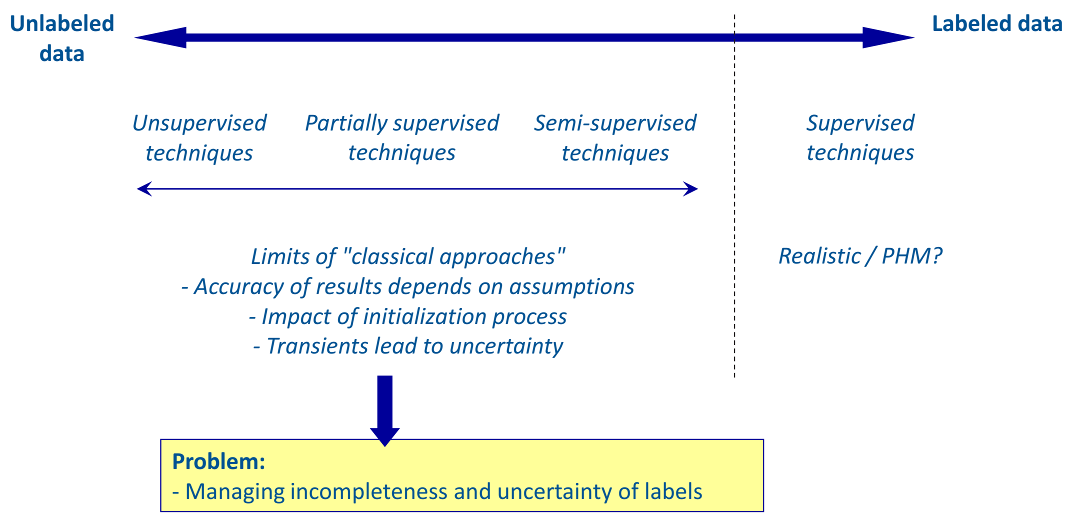

FIGURE 4.2: Usefulness and limits of classification methods in PHM (adapted from [75])

4.2.2 Clustering and discrete state estimation

Mainly, the goal of clustering is to organize data into homogeneous groups such that, intra group similarity is minimized to have compact clusters and inter group dissimilarity is maximized to increase separation among clusters. However, time series data for clustering can be static or dynamic. In case of static data, the time series features do not vary with time or the change is negligible. In case of dynamic data, the time series features vary with time. CM data for prognostics falls in the later category. In addition, the time series data can be discrete-valued or real valued, non-uniformly sampled

or uniformly sampled, univariate or multivariate and finally can have equal or unequal length [204]. Given unlabeled multidimensional data, it is required to determine groups of time series with similar properties to assess the health of monitored machinery.
For prognostics modeling, the discrete state estimation phase can be performed in two steps. Firstly, an off-line step is required, where data are clustered to determine the hidden structure representing states of degradation, i.e., good state to faulty state. This step can also be called as learning phase to build an unsupervised classifier. The second step corresponds to labeling new data in test phase. Therefore, in the on-line phase, discrete states (or labels) are assigned to new data by looking at distance similarity between multidimensional time series and clusters obtained in the off-line phase. To illustrate this procedure with respect to multivariate degradation based modeling strategy, consider 2D visualization in Fig. 4.3. The continuous state time series data are firstly assessed by the unsupervised classifier to determine clusters and transition from good state to faulty state, where in test phase new data from prediction are matched with obtained clusters. On the basis of distance similarity, data closer to particular cluster centers are assigned labels that represent discrete states (S), and the RUL can be estimated when the transition from degrading (d) state to faulty (f) occurs, indicating a FT to stop prediction process (see Eq. (2.13)).

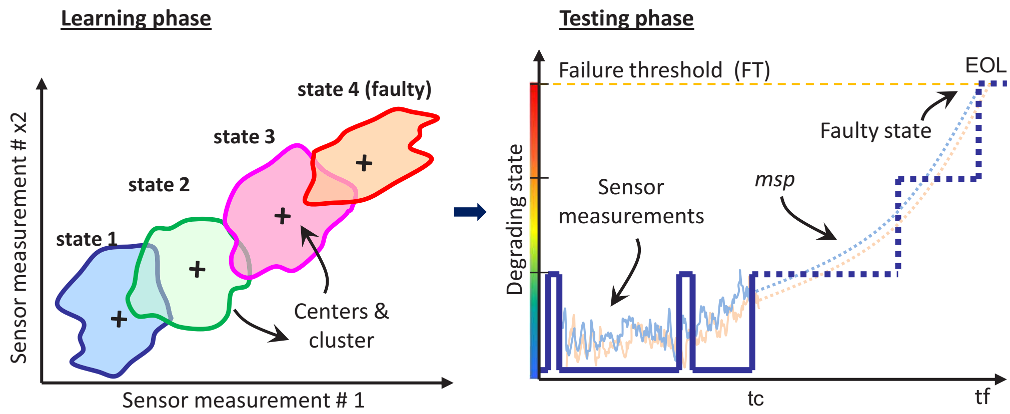

FIGURE 4.3: Discrete state estimation from multidimensional data

4.2.3 Issues and requirement

According to literature, clustering approaches can be classified into five categories: density based methods, graph based methods, hierarchical methods, partition methods and model based methods [204]. As clustering methods are totally data-driven, each approach can have its own limitations, however some key issues can be pointed out according to prognostics modeling challenges defined in section 1.4.

- Clustering algorithms can be sensitive to noise and outliers, which can influence cluster centers and distort clusters. As a result, the robustness of a prognostics model is affected.

- Clustering algorithms can have inconsistent performance, such that each run gives different results, which can affect robustness and reliability of a prognostics model.

- Clustering algorithms can be time consuming, dependent on human involvement to initialize parameters (i.e., number of clusters and initial centers values), limited to specific data types (e.g. time series with equal lengths). Such issues reduce the applicability of the clustering approach for a prognostics model.

As for real prognostics application where machinery operates in a dynamic environment and data are affected by inherent non-linearities, with the absence of *prior* knowledge of ground truth, clustering of an unlabeled data could be quite hard. In this case, it is also important to represent uncertainty of unlabeled multidimensional data like, transitions among different states, which can be performed by theories like fuzzy sets, probability, etc. Therefore, to achieve the goal of dynamic FTs, it is required to propose a new algorithm that accounts for issues mentioned above and can improve the efficiency of the prognostics model.

4.3 Discrete state estimation: the S-MEFC Algorithm

4.3.1 Background

4.3.1.1 Maximum Entropy Fuzzy Clustering (MEFC)

The clustering algorithms attempt to find an inherent structure for a given set of unlabeled data points by partitioning data into subgroups or clusters (representing similar objects). Typically, the partition based clustering algorithm (either hard or fuzzy) run as follows: assuming the initial positions of cluster center (or centroids) a membership function generates partition, which is measured by a cost function for fitness analysis. An iterative approach is implemented until the membership function create a partition that reduces the cost function [27]. Most importantly, whatever the clustering algorithm, it should provide compact clusters by minimizing the distance between similar objects (or data points) and separate dissimilar objects objects by maximizing their distances. In other words, intra-cluster distances should be minimized and inter-cluster distances should be maximized to obtain boundaries of clusters.
According to literature, to handle the issue of vague (imprecise) boundaries of clusters, fuzzy logic is considered as an effective tool [23]. Also, there are several approaches for fuzzy clustering, but, such techniques may induce biases in the choice of membership function [27]. Thereby, [125] proposed an algorithm namely, the maximum entropy based fuzzy clustering algorithm (MEFC), that ensures a maximum fairness to handle imprecise data and minimize such choice of bias of membership function via maximum entropy principle. In addition, MEFC algorithm is considered to be robust as compared

to FCM which is recognized to be sensitive to noise and outliers [215]. As compared to other fuzzy clustering approaches, the maximum entropy function also give a clear physical meaning for clustering data. This means, that data points closer to cluster centers will have higher memberships (representing higher entropy values), as compared to data points that are far from cluster centers [215] (see Fig. 4.4).

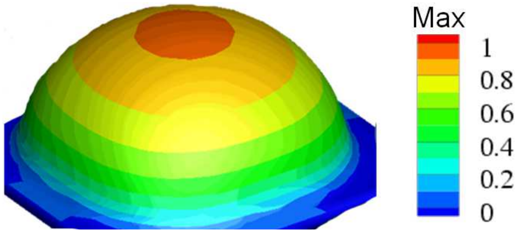

FIGURE 4.4: Entropy contour plot

Cluster estimation - Consider a dataset $\ell_D = \{X_i\}_{i=1}^N$ as already given by Eq. (3.1). MEFC is used to estimate clusters c among multidimensional data, and their centers $V = \{v_j\}_{j=1}^c$. To represent uncertainty of unlabeled multidimensional data, the maximum entropy inference (MEI) problem aims at assigning membership grade (μ_{ij}) between [0, 1] to every data point which avoids bias [125]. Formally, the entropy of cluster centers (H) with respect to data, i.e., the clustering entropy is given as:

$$H = -\sum_{i=1}^{N}\sum_{j=1}^{c}\mu_{ij}log\left(\mu_{ij}\right) \tag{4.1}$$

the entropy of data points with respect to the cluster centers to improve partitioning (i.e., clustered entropy), the loss function is given as:

$$L = \sum_{i=1}^{N}\sum_{j=1}^{c}\mu_{ij}D_{ij}^2 \tag{4.2}$$

where D_{ij} is the euclidean distance (ED) (Eq. 2.12) between the i^{th} data point and the j^{th} cluster. Fuzzy partition matrix can be represented as $U = [\mu_{ij}]_{c\times N}$, where μ_{ij} describes the membership of the i^{th} data point to the j^{th} cluster and satisfies following conditions.

$$0 \leq \mu_{ij} \leq 1 \qquad \forall\, i,j \tag{4.3}$$

$$0 \leq \sum_{j=1}^{N}\mu_{ij} < N \quad \forall\, i \tag{4.4}$$

$$\sum_{i=1}^{c}\mu_{ij} = 1 \qquad \forall\, j \tag{4.5}$$

A fuzzy clustering based on MEI aims at obtaining data points that minimize Eq. (4.2) and the membership assignments that meet Eq. (4.3-4.5). Further, to maximize Eq. (4.1) subject to Eq. (4.2) and constraints by Eq. (4.3-4.5), a Langrangian multiplier approach is used which gives following solutions [125]:

$$\mu_{ij} = \frac{e^{-D_{ij}^2/2\sigma^2}}{\sum_{k=1}^{c} e^{-D_{ik}^2/2\sigma^2}} \quad \forall\, i,j \tag{4.6}$$

where σ represents the fuzziness parameter, and $2\sigma^2$ is "temperature" in statistical physics. When $\sigma = 0$, the clustering becomes hard, when $\sigma \to \infty$ each pattern is equally assigned to all clusters [125]. Therefore, σ has similar properties like fuzziness parameter "m" in FCM clustering, however it has more clear physical meaning. In should be noted that the resulting Eq. (4.6) satisfies (4.2) only when centers v_j are fixed. Similarly, when μ_{ij} are fixed, the centers are updated with a following solution achieved from a proof similar to [29] for FCM:

$$v_j = \frac{\sum_{i=1}^{N} \mu_{ij} \cdot x_i}{\sum_{i=1}^{N} \mu_{ij}} \quad \forall\, j \tag{4.7}$$

which means that v_j is the center of j^{th} cluster.

Issues and requirements - According to [215], the performance of MEFC algorithm is dependent on the choice of the number of clusters and their initial centers. Therefore, it is required to properly initialize such parameters, to achieve an optimized solutions and avoid issues like inconsistency due to parameter initialization, which gives different results for different parameter inputs.

4.3.1.2 Subtractive Clustering

According to literature, to overcome issues of parameter initialization and to achieve the global optimum solution, methods like genetic algorithm and particle swarm optimization were applied [225]. However, such methods are computationally expensive which limits their applicability, for real prognostics applications. A simple and effective algorithm has been proposed by Yager and Filev [209], to estimate number of clusters and their initial centers, namely mountain clustering. But again, the algorithm was computationally expensive, because the computations grow exponentially according to the dimensions of the problem [46]. Finally, an extension of Mountain Clustering was proposed by Chiu [46], called as Subtractive Clustering (SC) which has an additional advantage to handle high dimensional data.

Briefly, SC is a one pass approach to estimate cluster centers among numeric data (based on the density function), that does not require any iterative optimization. Therefore, this method can be used to provide number of clusters and initial centers for algorithms like FCM or its other variants [125]. According to [54], SC is robust clustering algorithm could (a) remove outliers (robust to outliers that can impact clusters centers), (b) resolve conflicting data and (c) discard unnecessary patterns. In addition, it does not have inconsistency issue like FCM that gives different results for different simulations [23], which shows reliability of the method.

Cluster estimation - Consider a dataset $\ell_D = \{X_i\}_{i=1}^N$ as already given by Eq. (3.1). SC is used to estimate clusters c among multidimensional data and their centers $V = \{v_j\}_{j=1}^c$. Without loss of generality, data are normalized in each dimension such that all are bounded by a hypercube [46]. Each data point is considered as a potential cluster center. A density measure at each data point is computed, such that a data point having more neighboring data (or density of surrounding points) has higher chance to be selected as a cluster center, as compared to data point with fewer surrounding data points. Obviously, this approach can surely remove the effects of noisy data and outliers. The density or potential of a data point x_i, is given as:

$$D_i = \sum_{j=1}^{N} exp\left(-\frac{\|x_i - x_j\|^2}{(r_a/2)^2}\right) \tag{4.8}$$

where r_a is a positive constant that defines radius of neighborhood. A data point outside the range have less influence on the potential. Thus, the density measure for a data point is a function of its distances to neighboring data points (see Fig. 4.5).

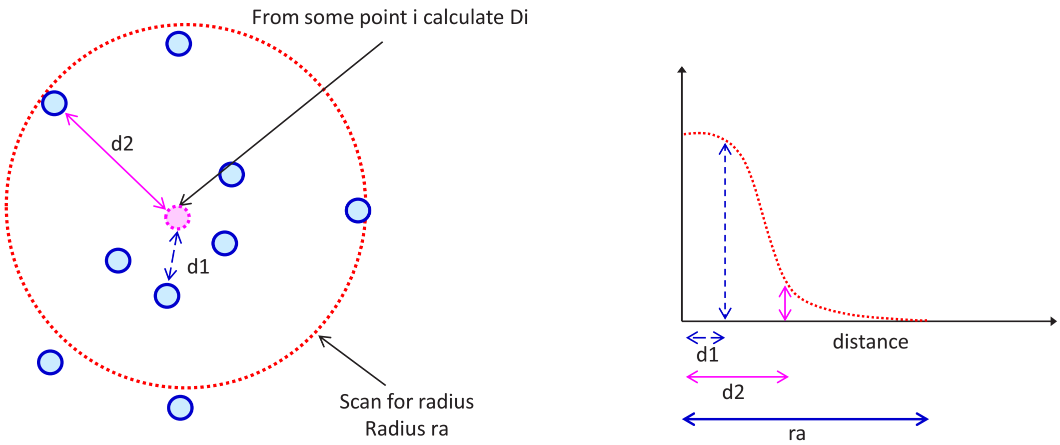

FIGURE 4.5: Density measurement for each data point [4]

Following the computation of density of each data point, the first cluster center x_{c1} is selected representing highest density value D_{c1}. Then again, the density measure of each data point is calculated as follows:

$$D_i = D_i - D_{c1}\, exp\left(-\frac{\|x_i - x_{c1}\|^2}{(r_b/2)^2}\right) \tag{4.9}$$

where r_b is a positive constant (i.e., $r_b = 1.5\, r_a$ to avoid closely spaced clusters). According to Eq. (4.9), after subtraction data points closer to the first cluster center

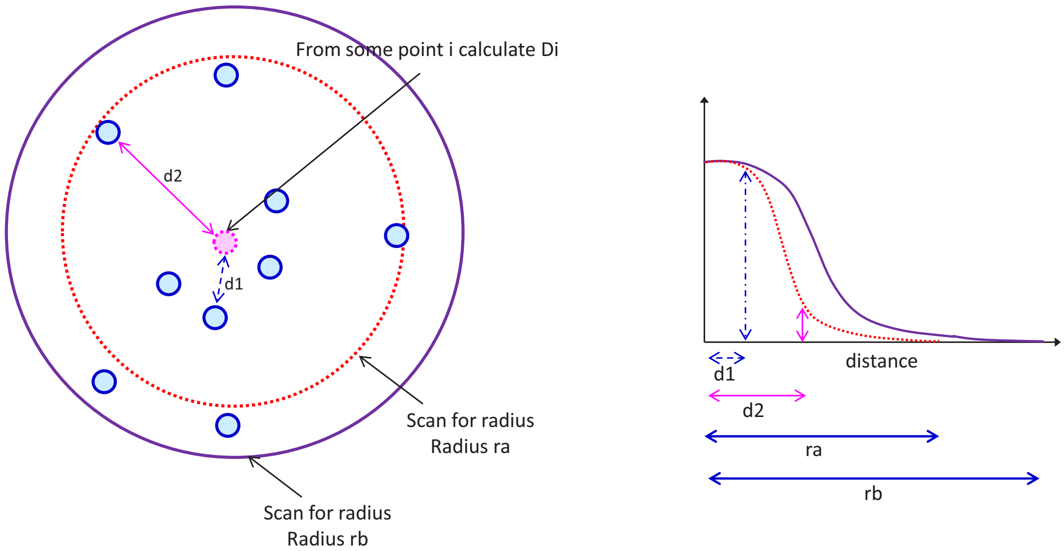

FIGURE 4.6: Revised density measuring procedure [4]

will have lower density as compared to data points that are far (Fig. 4.6). Therefore, the next cluster center x_{c2} is selected representing the highest density D_{c2}. This procedure is revised until sufficient number of clusters are achieved (see [46] for details). Lastly, according to cluster estimation procedure, SC appears to be a computationally efficient algorithm, that only requires one parameter to be set by the user i.e., radius of neighborhood, which shows its better applicability.

Issues and requirements - Although SC algorithm is a consistent algorithm with improved computational efficiency, robustness and applicability advantages. However, the main problem with this algorithms is the absence of uncertainty representation for the clustered data. Besides that, the centers are not optimized and require further adjustment [23]. Therefore, SC can be combined with iterative optimization (e.g. fuzzy clustering) methods to fine tune the centroids. This idea is presented in the following topic.

4.3.2 S-MEFC for discrete state estimation

We propose a new unsupervised classification algorithm namely, the Subtractive-Maximum Entropy Fuzzy Clustering (S-MEFC), that enables estimating the health state of the system. Mainly, S-MEFC algorithm takes benefits of density based SC algorithm [46] and of a MEFC approach by means of MEI [125]. S-MEFC algorithm is summarized in Algorithm 3. In brief, consider a training dataset $\ell_D = \{X_i\}_{i=1}^{N}$ as already given by Eq. (3.1) containing N unlabeled samples of n time series. SC approach is used to automatically determine clusters c in multidimensional data, and their centers $V = \{v_j\}_{j=1}^{c}$ (see [46]). SC is a one-pass approach that requires only one parameter to be

set by the user, i.e., the radius of neighborhood ra, but, the centers positions are not optimized and need to be adjusted. For this task we propose to use fuzzy clustering approach by means of MEI. The obtained centers, V from SC serve the need of MEFC clustering algorithm, that avoids any random initialization of centers. To optimize centers positions and to assign membership to each data point in a particular cluster, given σ a fuzziness parameter (set by user), the algorithm runs in an iterative manner until termination criteria (ϵ) is met. The MEI based fuzzy partition matrix can be represented as $U = [\mu_{ij}]_{c\times N}$, where μ_{ij} represents the membership degree of i^{th} object in j^{th} cluster.
It should be noted that, the key component in clustering is to measure similarity between two data being compared [204]. This can strongly influence shape of clustering, because different distance similarity measures can lead to different partition of data. In our case we use Standardized Euclidean Distance D_{SE} distance metric, while updating cluster partitions matrix U and centers matrix V with MEFC algorithm. It is very similar to ED except that every dimension is divided by its standard deviation. This leads to better clustering than would be achieved with ED, when each dimension has different scales. As for illustration, consider Fig. 4.7, where clustering is performed with D_{SE} (left part) for data with actual scales, in comparison to clustering with ED metric (right part). Better clustering is achieved by D_{SE} due to equalization of variances on each axis. Note that, it is also recommended in literature [195, 206], to use D_{SE} rather than ED.

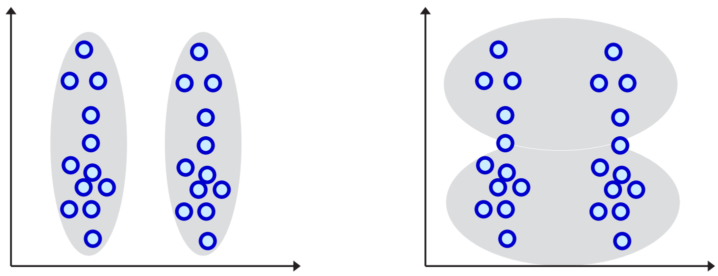

FIGURE 4.7: Example of clustering with D_{SE} (left) and ED (right) metrics [3]

Let x, v each be a n-dimensional vector and SD the standard deviation. The D_{SE} distance between data points and the centers is computed as follows:

$$D_{SE}(x,v) = \sqrt{\sum_{k=1}^{n} \left(1/SD_k{}^2\right) (x_k - v_k)^2} \tag{4.10}$$

Algorithm 3 S-MEFC

Require - Learning dataset Eq. (3.1)
- Fix ra, ϵ, $\sigma > 0$

Ensure - Cluster centers V and fuzzy partitioning U

learning procedure

1: Obtain initial cluster centers v^{old} using SC (section 4.3.1.2)
2: Compute fuzzy partition matrix U using MEI (section 4.3.1.1)

$$\mu_{ij} = \frac{e^{-D_{ij}^2/2\sigma^2}}{\sum_{k=1}^{c} e^{-D_{ik}^2/2\sigma^2}} \quad \forall\, i,j \tag{4.11}$$

3: Adjust cluster centers v^{new}

$$v_j^{new} = \frac{\sum_{i=1}^{N} \mu_{ij}.x_i}{\sum_{i=1}^{N} \mu_{ij}} \quad \forall\, j \tag{4.12}$$

4: Repeat step 2 and 3 untill termination criteria is met

$$\left\| v^{new} - v^{old} \right\| < \epsilon \tag{4.13}$$

4.4 Enhanced multivariate degradation based modeling

4.4.1 Outline: dynamic FTs assignment procedure

According to discussions in Chapter 1 (section 1.3.3.3), we propose to perform RUL estimates by enhancing classical multivariate degradation based modeling strategy. The classical approach is mainly based on assumption, that all the machines can have same states of degradation e.g. 3 or 4 states. According to that, FTs are assigned by partitioning unlabeled multidimensional data into specific number of groups set by the user [164, 166]. However, in reality this is not true, because each machine can have different levels of degradation, even when similar operating conditions are applied. For practical reasons such assumptions are avoided to enhance multivariate degradation based modeling, by proposing S-MEFC algorithm that enables to improve health assessment phase of prognostics model.

The proposed prognostics model integrates two algorithms namely, the SW-ELM (section 3.3.2) and the S-MEFC (section 4.3.2) to estimate RUL in a realistic manner. In other words, the main task of SW-ELM is to perform long-term multi-step ahead predictions (msp), where the task of S-MEFC is to stop the prediction process by estimating discrete states, which also enables to set FTs when transition from degrading state to faulty state occurs (see Eq. (2.13)). In comparison, the classical approach does not ac-

count for these important issues.
The flow diagram of prognostics using "enhanced multivariate degradation based modeling strategy " with dynamic threshold assignment procedure is shown in Fig. 4.8. Here two main phases are highlighted: 1) off-line phase to learn the predictor and the classifier, and 2) the on-line phase to simultaneously perform continuous state predictions and discrete state estimation, which also enables to dynamically assign FTs. The detailed descriptions of each phase are given in the following sections.

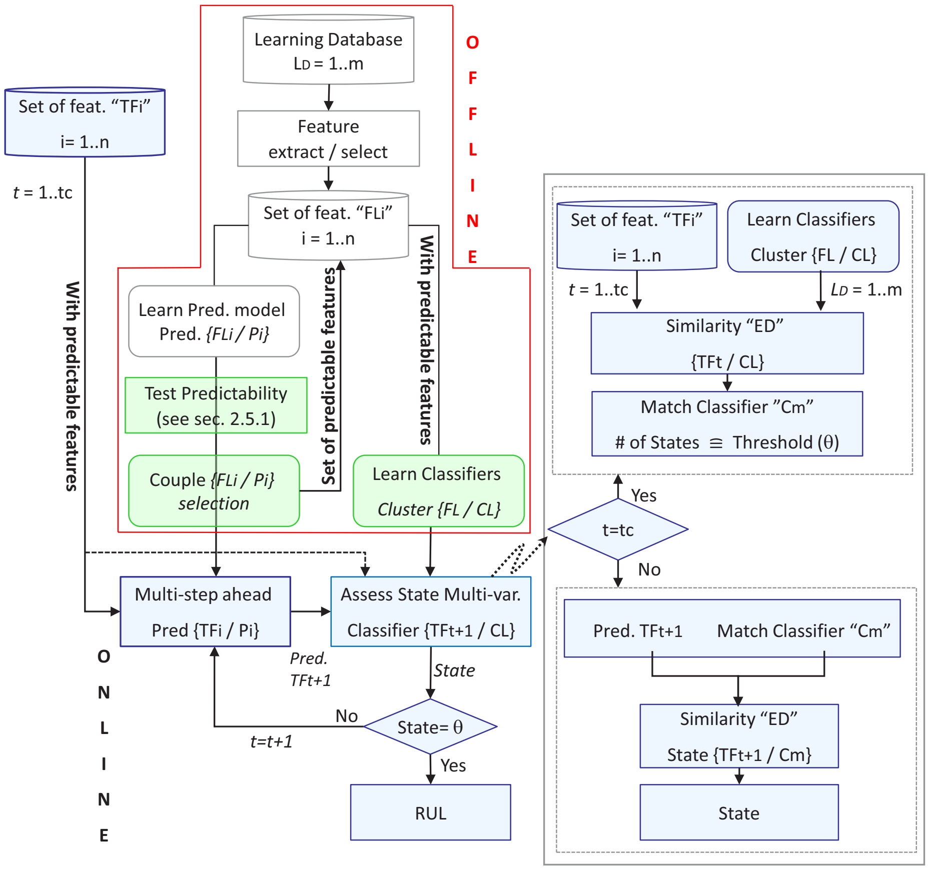

FIGURE 4.8: Enhanced multivariate degradation based modeling strategy

4.4.2 Off-line phase: learn predictors and classifiers

Consider training dataset containing multidimensional signal (sensor measurement or feature) F_{Li}, which can be from multiple learning instances **L**. According to dimension of feature space in the data, **n** univariate predictors P_i are build using SW-ELM and trained with data from **L** instances (Fig. 4.9, left part). The features set is further reduced by post-treatment (section 2.5.1) to final set of predictable features according to learned prediction models. This enables either to retain or reject each potential couple of "feature-model" to be used for prognostics. Following that, an unsupervised classifier is build using S-MEFC for each learning instance considering multidimensional signals F_i that are predictable (Fig. 4.9, right part). Each classifier C_L can represent different cluster or states depending on particular learning instance, that enables to dynamically set FTs in the on-line phase.

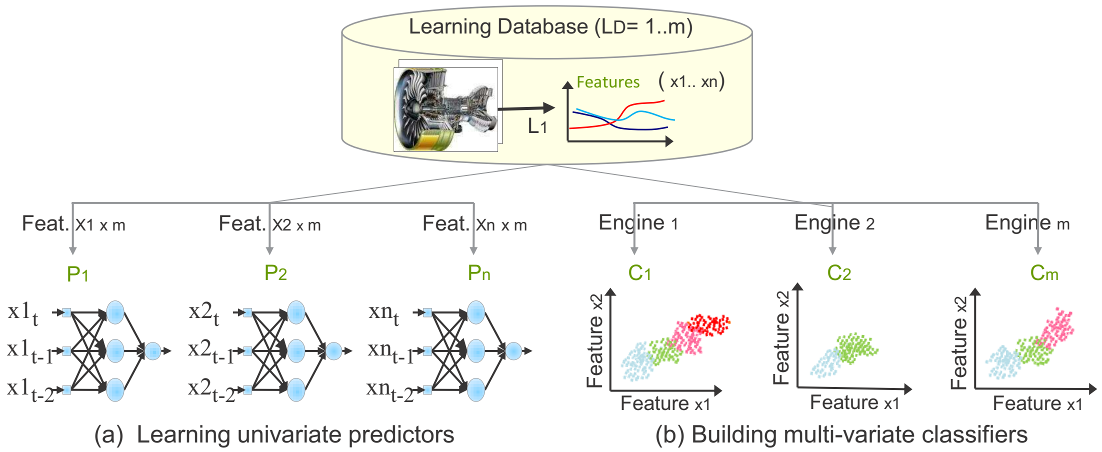

FIGURE 4.9: off-line phase

4.4.3 On-line phase: simultaneous predictions & discrete estimation

In the off-line phase, **n** univariate predictors and C_L classifiers are build with predictable features. In the on-line phase (Fig. 4.10), consider a new test instance TF_i containing partial data from multidimensional signals (i.e., predictable ones) up to some current time **tc** (or piece of trajectories up to **tc**, from where prognostics should be initiated). *Prior* to RUL estimation task, threshold is dynamically assigned by looking at distance similarity (ED) among learned classifiers and indexes of test data from instance TF_i (e.g. multidimensional data from index= 1:**tc**). Note that each index of TF can have similarity to a particular C_L. This step can be simply met by applying "mode operation" to identify most frequently matched classifier with indexes of test data. Suppose that, most of the indexes TF match classifier C_m that is build in the learning phase. Now consider that the total states (cluster) in C_m are **4**, which means 4^{th} state is faulty, and threshold for the test instance TF_i is set according to that. When the matched C_m is identified, the severity of degradation at the current state **tc** is also determined by distance similarity between **tc** and states of C_m. After assigning threshold and assessment of severity of degradation at current state, RUL estimation task is initiated. Finally, **n** predictors are used to perform *msp* in an iterative manner [76], and simultaneously discrete states are estimated by distance similarity (ED), due to best match C_m (see Fig. 4.3). The *msp* continues until discrete state reaches threshold point i.e., transition from degrading state to faulty (see Eq. (2.13)).

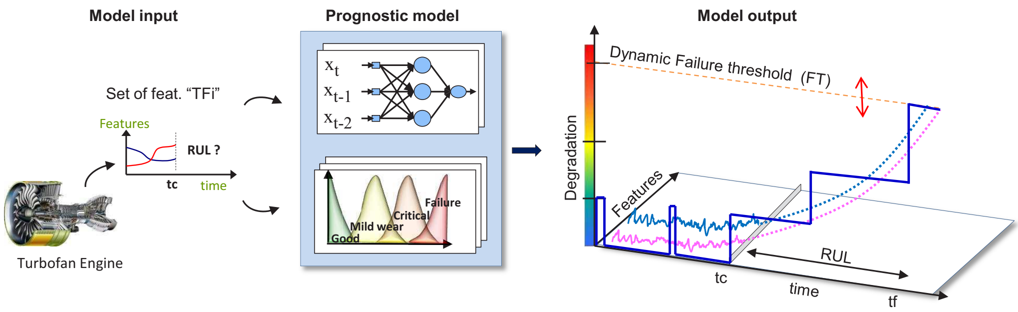

FIGURE 4.10: Simultaneous predictions and state estimation

4.5 PHM case study on real data

4.5.1 Turbofan Engines datasets of IEEE PHM Challenge 2008

The aim of this case study is to validate the performances of the proposed "enhanced multivariate degradation based modeling strategy". For this purpose, the real data of turbofan engines are considered, which are available at NASA data repository [143]. In this case study, complete data of text files "$train - FD001.txt$" and "$test - FD001.txt$" are used, (whereas, in our earlier experiments in section 2.6.1.1, data of only 40 engines from file "$train - FD001.txt$" were considered).
To recall, it should be noted that, each file record is composed of 100 instances (i.e., turbofan engines), and each instance either from the train or test file has 26 multidimensional time series signals contaminated with measurement noise. However, only 8 features were used from each instance for the experiments presented in CHAPTER 2 (see Table 2.5). As an example, consider the left part of Fig. 4.11, where run-to-failures data of "sensor measurement 2" are shown from 100 engines in the training data set. One can clearly notice that, the data are highly noisy and should be filtered before feeding the prognostics model. On the other hand, the right part of Fig. 4.11 shows the life spans distribution of 100 engines in the training data. The variability in learning histories can affect the training and testing performances.

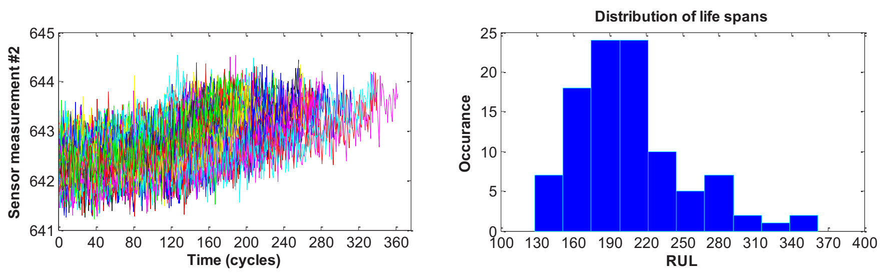

FIGURE 4.11: Sensor measurement from 100 engines (training data) and life spans

The tests file "$test - FD001.txt$" is composed of pieces of trajectories up to current time **tc** (for all of 100 instances), and the remaining life span or RUL is unknown. Note that, RUL estimates are in units of time, i.e., number of cycles.

4.5.2 Performance evaluation and simulation setting

For results evaluation, estimated RULs are compared with actual RULs provided in the file "$rul - FD001.txt$". Most importantly, for a given test instance **TF** (up to current time **tc**), an interval $I = [-10,13]$ is considered to assess RUL estimates as on-time, early or late (Fig. 4.12). This interval is further used to compute final score for comparison (see [175]), which should be as low as possible. In PHM context, it is generally desirable to have early RUL estimates rather than late RULs, since the main aspect is to avoid failures. Therefore, for engines degradation scenario, an early RUL estimate is preferred over late RUL.
In brief, the scoring function for the challenge data is asymmetric around the true failure time to heavily penalize late RUL estimates as compared to early RULs [175]. However, with the increase in error the scoring penalty grows exponentially. The scoring function (s) is given by Eq. (4.14) as follows:

$$s = \begin{cases} \sum_{i=1}^{n} e^{-\left(\frac{d}{a1}\right)} - 1 \; for \; d < 0 \\ \\ \sum_{i=1}^{n} e^{\left(\frac{d}{a2}\right)} - 1 \; for \; d \geq 0 \end{cases} \tag{4.14}$$

where, $a1$ and $a2$ are parameters to control the asymmetric preference, with values $a1 = 10$ and $a2 = 13$ (i.e., $I = [-10,13]$). The RUL error is denoted by d (i.e., estimated RUL - Actual RUL), and n represents the number of units under test (or instances).

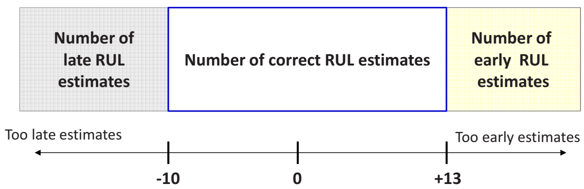

FIGURE 4.12: Prediction interval [166]

Moreover, prognostics performances are further assessed by following criteria.

1. Total computation *time* to learn and test entire dataset (i.e., fleet of 200 engines).
2. Accuracy of prognostics model evaluated by the coefficient of determination ($R2$), that should be close to 1.

According to the proposed scheme in this chapter, SW-ELM is used to build univariate predictors. The network topology of each SW-ELM predictor is set as follows: 3 regressors input node, 5 hidden nodes, 1 output node and constant $C = 0.1$ (i.e., for beta value in NW procedure, see Algorithm 2). As for state estimation, the S-MEFC classifier is used with a neighborhood radius $ra = 0.4$, and the fuzziness parameter $\sigma = 0.38$.

4.5.3 Prognostics results on complete test data

4.5.3.1 Results comparison: with all features & with predictable features

According to previous discussions in section 2.6.2.2 prognostics should be performed by using only predictable features, that leads to better RUL estimates. To avoid any confusion, the RUL estimates for turbofan data are expressed in units time i.e., number of cycles. From the experiments on training data, it was found that, among multidimensional data 8 features, only 5 are predictable, i.e., features {F1 ; F4 - F7} or sensor (2, 8, 11, 13, 15) as shown in Table 2.5. Let us again validate this assumption by estimating the RUL for each multidimensional time series test data set, by using, on one side, the whole set of pre-selected features {F1 - F8}, and on the other side, these final "predictability-based" selected features.
To estimate the RULs for both cases, SWELM and S-MEFC model are learned with 100 instances and tests are performed on 100 as well. It should be noted that for the learning and testing tasks, datasets are used with actual scales, without any normalization. As for example, the RUL estimation result with simultaneous prediction (by SWELM) and state estimation (by S-MEFC) involving multidimensional signals {F1 - F8} from the first test instance is shown in Fig. 4.13.

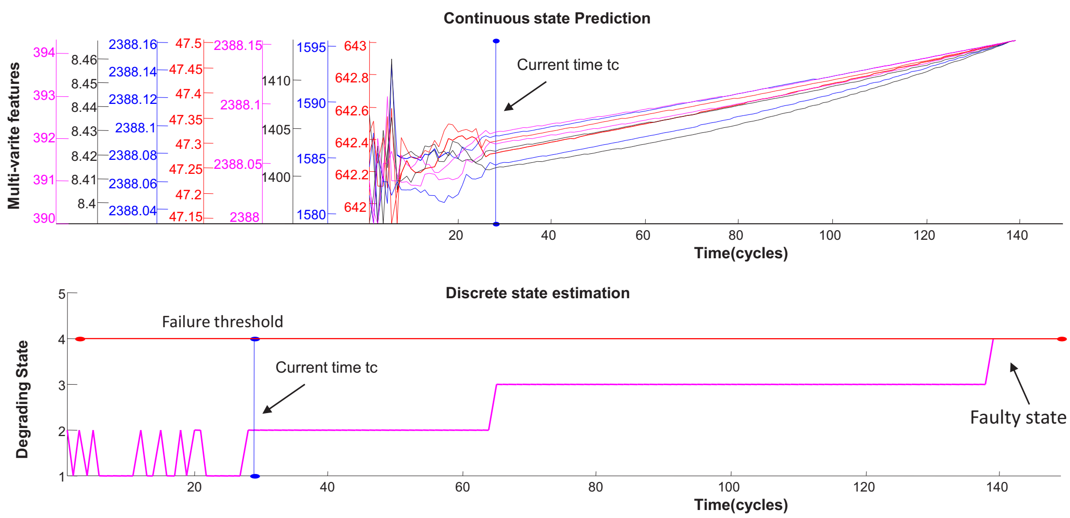

FIGURE 4.13: RUL estimation - Test 1

As mentioned in section 4.1, the RUL estimate is achieved by predicting multidimensional features and simultaneously estimating discrete state to stop the prediction task by dynamic FT. The dynamic FTs assignment is possible due to S-MEFC algorithm, that can automatically determine different clusters (or states) among multidimensional data. Obviously, this is a realistic approach, because each machine can have different level of degradation from good state to faulty state, and any assumption abouts states should be avoided. For e.g. consider Fig. 4.14, where clustering results on run-to-failure data on two engines from file "$train - FD001.txt$" are provided for different sensor measurements.

The right part of the figure represents a color code sequence which corresponds to the membership of data points in the clusters shown in the left part of Fig. 4.14 with the same color. In case of turbofan engine 1, data are automatically partitioned into 4 groups, whereas for turbofan engine 100 data are automatically partitioned into 6 groups using S-MEFC algorithm. Similarly, each classifier C_L (for a particular engine in the training data) can represent different states of degradation.

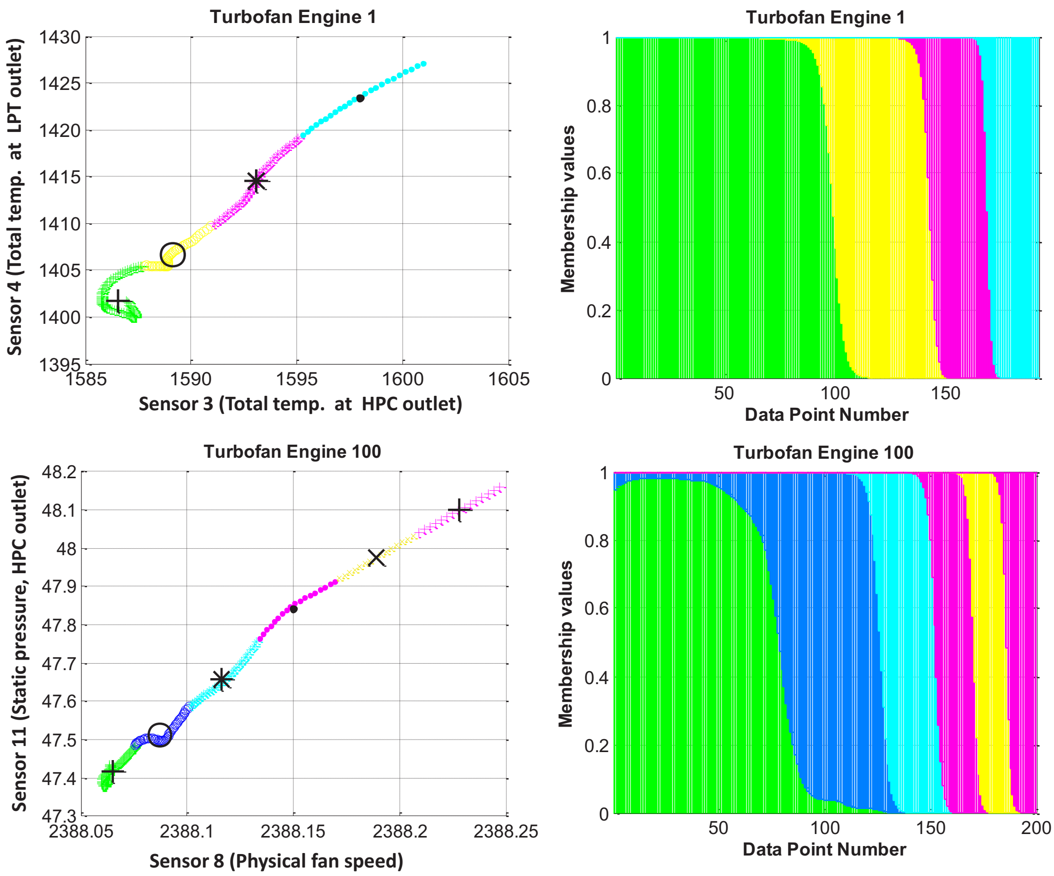

FIGURE 4.14: Visualization of classes and data point membership functions

The dynamic thresholds are assigned by looking at distance similarity in multidimensional data of test instance and learned classifiers C_L. Fig. 4.15, shows variations of FTs with all features {F1 - F8} and with predictable features {F1 ; F4 - F7} for complete test data of 100 engines. Mainly, the difference of the results is due to the number of features used to build a classifier C_L (because unpredictable features can lead to poor prognostics results).

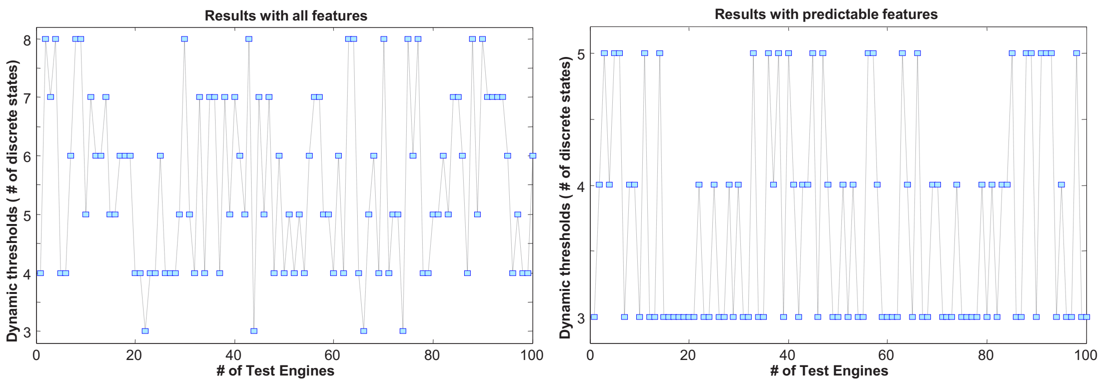

FIGURE 4.15: Dynamic threshold assignment results (for discrete states)

Now, lets validate our proposed methodology to get correct RUL estimates falling in an interval $I = [-10,13]$, i.e., enhanced multivariate degradation based modeling strategy with predictable features. Table 4.1 summarizes performances for all tests initiated at current time tc, and compares prognostics results with all features {F1 - F8} and with predictable features {F1 ; F4 - F7}.

TABLE 4.1: Prognostics model results comparison for 100 test engines

Criteria	**With all features**	**With predictable features**	By [166]
RUL error distribution span	$I = [-85{,}74]$	**I=[-39,60]**	$I = [-85{,}120]$
On time estimates	32	48	53
Early estimates	34	40	36
Late estimates	34	12	11
$R2$	0.55	**0.614**	N/A
Time	5m 33sec	**3m 54sec**	N/A
Score	4463	**1046**	N/A

Among 100 tests instances, when using predictable features RUL estimates of **48** cases fall in interval $I = [-10,13]$, i.e., on-time estimates. The amount of early estimates is **40**, and amount of late estimates is only **12**. It should be noted that early estimates are preferable than late estimates that can cause catastrophic situations [175].
The estimated RULs with the proposed approach are also compared with actual RULs provided in the file "$rul - FD001.txt$", where the overall accuracy achieved is **R2=0.614** with a global score of **1046** (whereas score is 4463 when using all features). Most importantly, the RUL error distribution with the proposed approach (using predictable features) has lowest span **I=[-39,60]** as compared to other methods (Fig. 4.17). This shows robust performance of the prognostics model when exposed to test data without any knowledge about the RULs and their variability. As shown in Fig. 4.16, the estimated RULs are closely mapped to actual RULs.

Besides that, it is also important for a prognostics model to meet real time constraints. In this context, the total computation time with the proposed approach to learn and test the data from fleet of **200** turbofan engines is just **3m 54sec**.

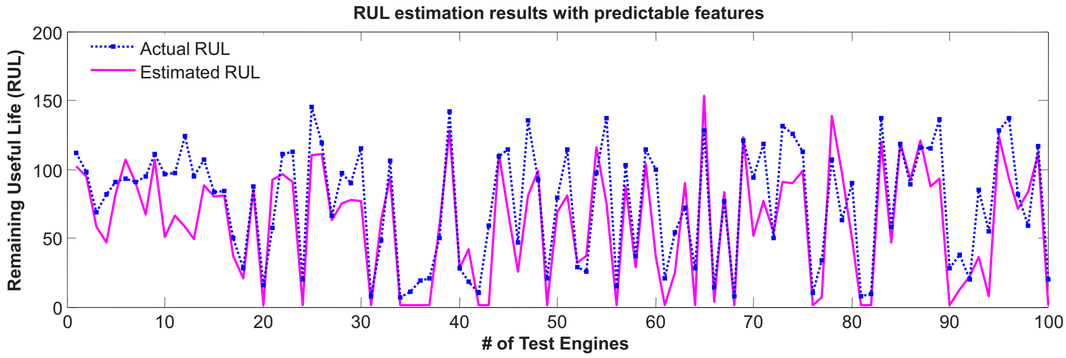

FIGURE 4.16: Actual RUL vs. estimated RUL (100 Tests)

4.5.3.2 Results comparison: to validate performances with other approaches

As expected, predictability based prognostics model has much better results as compared to prognostics with all features (including unpredictable ones). Obviously, features that are hard to be predicted can lead to poor prognostics results. But still, it is required to further compare the performances of the "enhanced multivariate degradation based modeling strategy", to other approaches in the literature. Unfortunately the authors who use data "$train - FD001.txt$" and "$test - FD001.txt$" do not clearly mention errors of RUL estimates and scoring (for each test) which prevents benchmarking approaches. In literature only one publication was made with same data of 200 turbofan engines (and also the same Features {F1 ; F4 - F7}) by applying "classical multivariate degradation based modeling" strategy [166], but again limited information on results were given. Nevertheless, it will help to make a final comparison with RUL estimates achieved with our method.
The RUL error distribution obtained with proposed approach is compared with the recent results published by [166], (see Fig. 4.17). According to these results, the proposed approach has better precision of RUL estimates as indicated by the compact distribution of RUL errors (i.e. from **I=[-39, 60]**). In comparison, the results provided by [166] are less precise with a wider distribution of RUL errors (i.e., approximately from I=[-85, 120]). However, [166] achieved higher number of on-time estimates as compared to our method, i.e., **53** vs. 48. But, the RUL error distribution has a wide spread on interval I=[-85, 120], which shows very early/ late estimates and will result large score according to criteria of data challenge Eq. (4.14). Therefore, when the proposed method is evaluated on the basis of scoring criteria, the achieved total score is **1046**, (see Table 4.1). The score curve for all 100 tests is given in Fig. 4.18, where only **5%** RUL estimates are very early, which affects the scoring. However, the score plot clearly reflects that,

the proposed method has preference for early RUL estimates rather than late RULs. Further, it is not possible to compare our score with other approaches in the literature, due to the absence of scoring results with this data (i.e., $train - FD001.txt$).

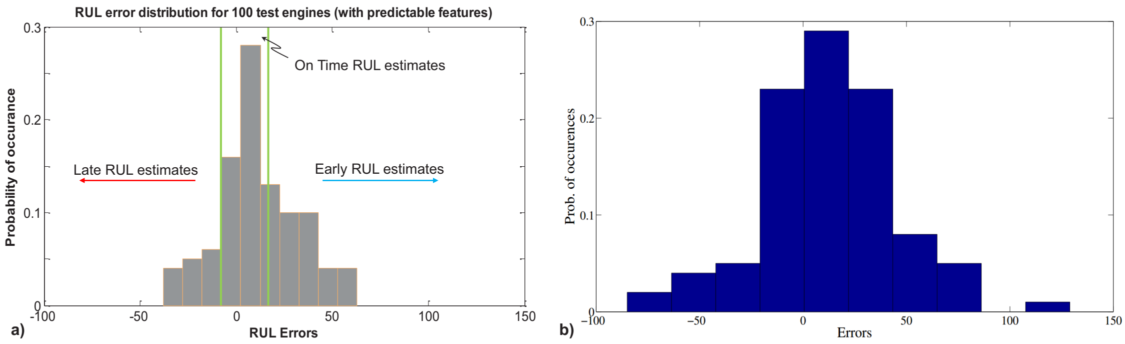

FIGURE 4.17: RUL error distribution a) proposed approach and b) by [166]

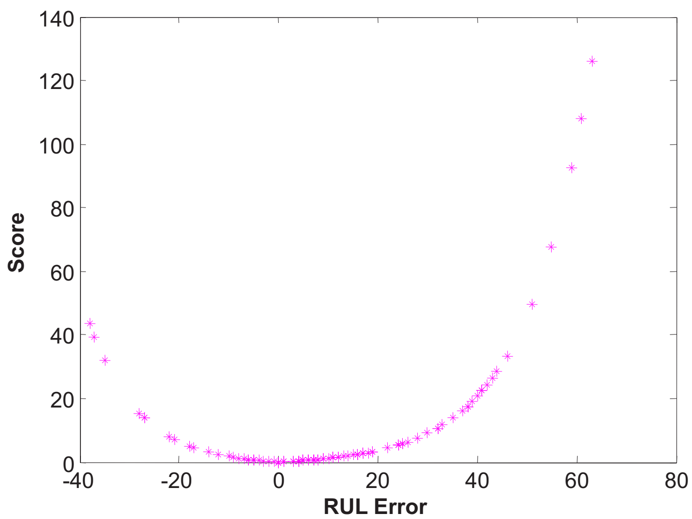

FIGURE 4.18: Scores for 100 tests (enhanced multivariate degradation based modeling)

Lastly, in case of the total computation time comparison, the proposed approach is much faster (i.e., **3m 54sec**) than "trajectory similarity based prediction approach" by [197], which could take few hours even with only data of **100** test instances (i.e., engines).

4.5.4 Benefits of the proposed approach

As stated before, it is not possible to further benchmark our approach with previous works. Nevertheless, leaving aside performances criteria, one can point out the following aspects that highlight the benefits of our approach from application point of view.

- Does not require any normalization of data and work on actual scales;
- Manages unlabeled data of different temporal lengths;
- Requires no assumption about the data and can dynamically estimates states;
- Failure thresholds are dynamically assigned without any *prior* knowledge;
- Fast approach and requires only four parameters to be set by the user.

4.6 Summary

This chapter addresses the challenges of prognostics modeling while building a classification model, and also discusses the key importance of FTs for the reliability of RUL estimates. Firstly, the S-MEFC algorithm is proposed as a potential approach to handle unlabeled multidimensional data, and to dynamically assess the health of degrading machinery. Secondly, an enhanced strategy for RUL estimates is proposed, that integrates SW-ELM and S-MEFC for performing simultaneously predictions and discrete state estimation and to dynamically assign FTs. Further, to investigate the efficiency of proposed prognostics model, a complete prognostics application is considered for real data of turbofan engines from PHM challenge 2008. RUL estimation results of the proposed approach are also compared with recently published articles to show its significance.

Chapter 5

Conclusion and future works

This chapter concludes the research work by summarizing issues addressed in the thesis, developments and contributions. A discussion on limits of the approach is also laid out to address future perspectives.

Conclusion

To optimize machinery service delivery potential during its lifetime, the main goal of prognostics is to intelligently use condition monitoring data from an in-service machinery, and estimate the Remaining Useful Life (RUL) before a failure. Therefore, it is a core process of Prognostics and Health Management (PHM) discipline that links the studies of failure mechanism and life cycle management of critical engineering assets. However, in a dynamic operational environment, the deterioration process of a machinery can be affected by different factors like engineering variances, failure modes, environmental and operating conditions. The data acquired from such machinery are usually noisy and subject to high level of uncertainty / unpredictability, and estimation of RUL is usually a hard task. To perform an accurate prognostics, one has to address the challenges of modeling phase i.e., robustness to encounter uncertain inputs, reliability to encounter unknown data operating conditions, engineering variations, etc., and applicability to meet industrial constraints and requirements. According to this, data-driven prognostics approaches appear to be a valuable alternative to deal with uncertainty of available data, in applications where a deep understanding of system physics is not available.
In this thesis, a novel prognostics approach is developed that aims at enhancing data-driven prognostics in a realistic manner, with an acceptable level of accuracy under modeling challenges of robustness, reliability and applicability. For this purpose, mainly two aspects are considered: data processing and prognostics modeling.
In Chapter 1, the literature on PHM and classification of prognostics approaches is extensively reviewed. Prognostics approaches, (i.e., physics based, data-driven, and hybrid) are assessed upon different criteria that are related to two important factors: 1)

applicability and 2) performances. The assessment shows that data-driven methods have higher applicability as compared to physics based and hybrid approaches, but their performances require further improvement that can be linked with robustness and reliability of the approach. Different strategies for data-driven prognostics are further reviewed which enables us to point out that the "multivariate degradation based modeling strategy" for RUL estimation has a strong potential, and is more realistic as compared to direct RUL prediction and to univariate degradation based modeling. Therefore, it is considered as a building block for prognostics modeling step. Besides that, for open challenges of prognostics modeling, i.e., robustness, reliability and applicability, new definitions are also proposed, which are not clearly defined in the literature.

Whatever the prognostics approach is, it requires suitable time series features or variables that are sensitive to machine degradation, reflect failure progression, and can be predicted easily. In Chapter 2 we propose the data pre-treatment and data post-treatment methods for feature extraction and selection. In brief, for data pre-treatment firstly, raw data are analyzed by applying discrete wavelet transform. Secondly feature extraction is performed in a new manner by using trigonometric functions and cumulative transformation. Finally, feature selection is performed by assessing their fitness in terms of monotonicity and trendability characteristics. Experiments on real data bearings from PHM challenge 2012 show significant improvements by the proposed data pre-treatment procedure, as compared to classical way of feature extraction and selection. The data post-treatment phase emphasize to further reduce multidimensional features data by assessing their predictability, as there is no interest in retaining features that are hard to be predicted. To show the impact of data post-treatment procedure, the multivariate degradation based modeling strategy is applied to estimate the RUL by simultaneous predictions and discrete state estimation. Experiments on real data of turbofan engines from PHM data challenge 2008 show that, the use of multidimensional data with predictable features is beneficial in improving accuracy of RUL estimates. In addition, some weaknesses of multivariate degradation based modeling strategy are also identified, to be addressed in succeeding chapters.

To account for issues related to time series prediction and the challenges of prognostics modeling (i.e., robustness, reliability and applicability), a new prediction algorithm is proposed in Chapter 3: the Summation Wavelet-Extreme Learning Machine (SW-ELM). SW-ELM is a supervised approach that does not require any iterative learning like classical neural network and has only two parameters to be set by user. For time series prediction issues, the performances of SW-ELM are benchmarked with existing Extreme Learning Machine (ELM), Levenberg-Marquardt algorithm for SLFN and ELMAN network on six different data sets. Results show improved performances achieved by SW-ELM in terms of computation time, model complexity and prediction accuracy. To validate the efficiency of SW-ELM with respect to robustness and reliability challenges, a real case of Computer Numerical Control (CNC) machine is considered for tool wear monitoring task. Moreover, it is observed that, lower magnitude (or norm) of weights and bias of SW-ELM during learning phase can play a significant role in improving testing accuracy. Although, performances of SW-ELM are of great interest but, such rapid learning methods should be used as ensemble of multiple models to

achieve more robust and accurate performances rather than a single model. Therefore, an ensemble of SW-ELM is also proposed to quantify uncertainty and improve accuracy of estimates. Note that, during experiments the reliability of SW-ELM is assessed on unknown data of cutting tools with engineering variations like geometry and coating. Moreover, SW-ELM is further used to perform "long-term predictions" with real data of degraded bearings from PHM challenge 2012, with constant operating conditions. The multi-step ahead prediction (*msp*) results with cumulative features are also compared by *msp* using classical features.
In Chapter 4 prognostics performances are further enhanced by a new health assessment algorithm: the Subtractive-Maximum Entropy Fuzzy Clustering (S-MEFC). S-MEFC is an unsupervised classification approach that can automatically determine the number of states (clusters) from multidimensional data, i.e., without human assumption and has consistent performance for each run. S-MEFC uses Maximum Entropy Inference to represent uncertainty of data rather than traditional fuzzy memberships. Moreover, a novel strategy for RUL estimation is also presented by integrating SW-ELM and S-MEFC as prognostics model. Both models are run simultaneously in order to perform "long-term predictions" and discrete state estimation, and to dynamically set failure thresholds (FTs). Lastly, the efficiency of the proposed "enhanced multivariate degradation based modeling strategy", is investigated on a case study of turbofan engines from PHM challenge 2008, and the results are thoroughly compared with recent publications. From results analysis we find that, the proposed RUL estimation strategy is more realistic as compared to existing methods in terms of applicability criteria like computational time, model complexity and parameters, human assumptions and domain knowledge. In addition, the accuracy of RUL estimates show better robustness of the proposed approach over fleet of test engines with different life spans. Besides that, the reliability of the prognostics approach is also improved by avoiding uncertain FTs based on assumptions and are assigned dynamically in precise manner for each test case.

Obviously no model is perfect, however, the proposed developments on data-processing and prognostics modeling show good potential and improved performances as compared to traditional methods. In addition, they also fulfill four main conditions of data-driven prognostics, as mentioned in [24], 1) ability to extract features related to damage, 2) to set FTs, 3) fault propagation model to project fault growth up to the FT and 4) associate confidence. The main limits of proposed data-driven approach are as follows.

1. How to automate the initialization of two important parameters of the prognostics model, which are to be set by the user, i.e., number of neurons in the hidden layer of SW-ELM, and the radius of neighborhood of a cluster in S-MEFC?

2. How to further manage the uncertainty related to data and prognostics modeling?

3. How to integrate variable operating conditions to enhance learning / testing performances of the prognostics model?

Nevertheless, this research work is a step ahead in PHM domain toward maturity of prognostics.

Future works

In general, inherent uncertainty in the machinery is a major hurdle while developing prognostics models, and indeed one of the factors that add to uncertainty is variable operating conditions. Such issue can greatly affect reliability of the prognostics model. Indeed, the variations of operating conditions can affect the behavior of degrading machinery, for e.g. increase in load can speedup the deterioration process and reduce RUL of machinery, while decrease in load results opposite behavior. Thus, operating conditions should be used as inputs of the prognostics model, and this topic needs to be further explored. Also, another problem could be the absence of future operating conditions, which demands to intelligently manage the learning phase of data-driven prognostics model, to avoid large uncertainty of predictions phase.
Among different limits highlighted above, the key areas for future focus are uncertainty management and integration of variable operating conditions. Therefore, current developments of this thesis will be further extended as follows.

1. To manage and quantify uncertainty related to data and prognostics modeling;
2. To investigate the effects of variable operating conditions on the proposition of cumulative features;
3. To integrate variable operating conditions as inputs of the prognostics model;
4. To extend the development to higher level rather than a component level.

Bibliography

[1] Industrial dryer dataset
ftp://ftp.esat.kuleuven.ac.be/sista/data/process_industry.

[2] Mechanical hair dryer dataset
ftp://ftp.esat.kuleuven.ac.be/sista/data/mechanical.

[3] Microarray cluster analysis and applications
http://www.science.co.il/enuka/essays/microarray-review.pdf.

[4] Subtractive clustering density
http://www.spikingneurons.com/.

[5] IEEE PHM 2012 prognostic challenge
www.femto-st.fr/f/d/ieeephm2012-challenge-details.pdf.

[6] NN3 forecasting competition
http://www.neural-forecasting-competition.com/nn3/index.htm.

[7] Abbas, S. R., and Arif, M. New time series predictability metrics for nearest neighbor based forecasting. In *Multitopic Conference, 2006. INMIC'06. IEEE* (2006), IEEE, pp. 100–105.

[8] Abu-Hanna, A., and Lucas, P. Prognostic models in medicine. *Methods of Information in Medicine 40*, 1 (2001), 1–5.

[9] Aka, R., Lia, Y.-F., Vitellia, V., and Zio, E. A genetic algorithm and neural network technique for predicting wind power under uncertainty. *Chemical Engineering 33* (2013), 1–6.

[10] An, D., Choi, J.-H., and Kim, N. H. A tutorial for model-based prognostics algorithms based on matlab code. In *Annual Conference of the Prognostics and Health Management Society, Mineapolis, Sept. 23-27* (2012).

[11] An, D., Choi, J.-H., and Kim, N. H. Prognostics 101: A tutorial for particle filter-based prognostics algorithm using matlab. *Reliability Engineering & System Safety 115*, 0 (2013), 161–169.

[12] Antoni, J. The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing 20*, 2 (2006), 282–307.

[13] Asmai, S. A., Hussin, B., and Yusof, M. M. A framework of an intelligent maintenance prognosis tool. In *Computer Research and Development, 2010 Second International Conference on* (2010), IEEE, pp. 241–245.

[14] Balaban, E., and Alonso, J. J. An approach to prognostic decision making in the aerospace domain. In *Annual Conference of the Prognostics and Health Management Society, Mineapolis, Sept. 23-27* (2012).

[15] Banakar, A., and Azeem, M. F. Artificial wavelet neural network and its application in neuro-fuzzy models. *Applied Soft Computing 8*, 4 (2008), 1463–1485.

[16] Baraldi, P., Cadini, F., Mangili, F., and Zio, E. Model-based and data-driven prognostics under different available information. *Probabilistic Engineering Mechanics 32*, 0 (2013), 66–79.

[17] Baraldi, P., Cadinia, F., Mangilia, F., and Zioa, E. Prognostics under different available information. *Chemical Engineering 33* (2013), 163–168.

[18] Baraldi, P., Compare, M., Sauco, S., and Zio, E. Ensemble neural network-based particle filtering for prognostics. *Mechanical Systems and Signal Processing 41*, 1-2 (2013), 288–300.

[19] Baraldi, P., Mangili, F., and Zio, E. Investigation of uncertainty treatment capability of model-based and data-driven prognostic methods using simulated data. *Reliability Engineering & System Safety 112*, 0 (2013), 94–108.

[20] Baraldi, P., Mangili, F., Zio, E., et al. A kalman filter-based ensemble approach with application to turbine creep prognostics. *IEEE Transactions on Reliability 61*, 4 (2012), 966–977.

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

[22] Barton, P. H. Prognostics for combat systems of the future. In *Autotestcon, 2006 IEEE* (2006), IEEE, pp. 599–604.

[23] Bataineh, K., Naji, M., and Saqer, M. A comparison study between various fuzzy clustering algorithms. *Editorial Board 5*, 4 (2011), 335.

[24] Bechhoefer, E. *Data Driven Prognostics for Rotating Machinery.* Ed. Seifedine Kadry. Hershey: IGI Global, 2013.

[25] Bellini, A., Filippetti, F., Tassoni, C., and Capolino, G.-A. Advances in diagnostic techniques for induction machines. *Industrial Electronics, IEEE Transactions on 55*, 12 (2008), 4109–4126.

[26] Bendjama, H., Bouhouche, S., and Boucherit, M. S. Application of wavelet transform for fault diagnosis in rotating machinery. *International Journal of Machine Learning and Computing 2*, 1 (2012), 82–87.

[27] Beni, G., and Liu, X. A least biased fuzzy clustering method. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 16*, 9 (1994), 954–960.

[28] Benkedjouh, T., Medjaher, K., Zerhouni, N., and Rechak, S. Health assessment and life prediction of cutting tools based on support vector regression. *Journal of Intelligent Manufacturing* (2013), 1–11.

[29] Bezdek, J.-C. *Pattern Recognition with Fuzzy Objective Function Algorithm.* Plenum New York, 1981.

[30] Bhat, A. U., Merchant, S., and Bhagwat, S. S. Prediction of melting point of organic compounds using extreme learning machines. *Industrial and Engineering Chemistry Research 47*, 3 (2008), 920–925.

[31] Bin, L., Yibin, L., and Xuewen, R. Comparison of echo state network and extreme learning machine on nonlinear prediction. *Journal of Computational Information Systems 7*, 6 (2011), 1863–1870.

[32] Boashash, B., and Black, P. An efficient real-time implementation of the wigner-ville distribution. *Acoustics, Speech and Signal Processing, IEEE Transactions on 35*, 11 (1987), 1611–1618.

[33] Bolander, N., Qiu, H., Eklund, N., Hindle, E., and Taylor, R. Physics-based remaining useful life prediction for aircraft engine bearing prognosis. In *Annual Conference of the Prognostics and Health Management Society* (2009).

[34] Bosnić, Z., and Kononenko, I. An overview of advances in reliability estimation of individual predictions in machine learning. *Intelligent Data Analysis 13*, 2 (2009), 385–401.

[35] Butler, S. *Prognostic Algorithms for Condition Monitoring and Remaining Useful Life Estimation.* PhD thesis, National University of Ireland Maynooth, 2012.

[36] Caesarendra, W., Widodo, A., Pham, H., and Yang, B.-S. Machine degradation prognostic based on rvm and arma/garch model for bearing fault simulated data. In *Prognostics and Health Management Conference, PHM'10* (2010), IEEE, pp. 1–6.

[37] Camci, F., and Chinnam, R. B. Health-state estimation and prognostics in machining processes. *Automation Science and Engineering, IEEE Transactions on 7*, 3 (2010), 581–597.

[38] Camci, F., Medjaher, K., Zerhouni, N., and Nectoux, P. Feature evaluation for effective bearing prognostics. *Quality and Reliability Engineering International 29*, 4 (2013), 477–486.

[39] Cao, J., Lin, Z., and Huang, G.-B. Composite function wavelet neural networks with extreme learning machine. *Neurocomputing 73*, 7-9 (2010), 1405–1416.

[40] Celaya, J. R., Saxena, A., and Goebel, K. Uncertainty representation and interpretation in model-based prognostics algorithms based on kalman filter estimation. In *Annual Conference of the Prognostics and Health Management Society* (2012).

[41] Celaya, J. R., Saxena, A., Saha, S., and Goebel, K. F. Prognostics of power mosfets under thermal stress accelerated aging using data-driven and model-based methodologies. In *Annual Conference of the Prognostics and Health Management Society, Montreal* (2011).

[42] Chebil, J., Noel, G., Mesbah, M., and Deriche, M. Wavelet decomposition for the detection and diagnosis of faults in rolling element bearings. *Jordan Journal of Mechanical & Industrial Engineering 4*, 5 (2010), 260–266.

[43] Chen, B., Zhang, Z., Sun, C., Li, B., Zi, Y., and He, Z. Fault feature extraction of gearbox by using overcomplete rational dilation discrete wavelet transform on signals measured from vibration sensors. *Mechanical Systems and Signal Processing 33* (2012), 275–298.

[44] Cheng, S., and Pecht, M. A fusion prognostics method for remaining useful life prediction of electronic products. In *Automation Science and Engineering, CASE 2009. IEEE International Conference on* (2009), IEEE, pp. 102–107.

[45] Chinnam, R. B., and Baruah, P. A neuro-fuzzy approach for estimating mean residual life in condition-based maintenance systems. *International Journal of Materials and Product Technology 20*, 1 (2004), 166–179.

[46] Chiu, S. L. Fuzzy model identification based on cluster estimation. *Journal of intelligent and Fuzzy systems 2*, 3 (1994), 267–278.

[47] Chui, C. K. *An introduction to wavelets*, vol. 1. Academic press, 1992.

[48] Coble, J., and Hines, J. W. Identifying optimal prognostics parameters from data: A genetic algorithms approach. In *Annual Conference of the Prognostics and Health Management Society* (2009).

[49] Coble, J., and Hines, J. W. Applying the general path model to estimation of remaining useful life. *International Journal of Prognostics and Health Management 2*, 1 (2011), 74–84.

[50] Côme, E., Oukhellou, L., Denoeux, T., and Aknin, P. Learning from partially supervised data using mixture models and belief functions. *Pattern recognition 42*, 3 (2009), 334–348.

[51] Das, S., Hall, R., Herzog, S., Harrison, G., and Bodkin, M. Essential steps in prognostic health management. In *Prognostics and Health Management (PHM), 2011 IEEE Conference on* (2011), IEEE, pp. 1–9.

[52] De Freitas, J. G., MacLeod, I., and Maltz, J. Neural networks for pneumatic actuator fault detection. *Transaction-South African Institute of Electrical Engineers 90* (1999), 28–34.

[53] Diebold, F. X., and Kilian, L. Measuring predictability: theory and macroeconomic applications. *Journal of Applied Econometrics 16*, 6 (2001), 657–669.

[54] Doan, C., Liong, S., and Karunasinghe, D. Derivation of effective and efficient data set with subtractive clustering method and genetic algorithm. *Journal of Hydroinformatics 7* (2005), 219–233.

[55] Dong, M. A tutorial on nonlinear time-series data mining in engineering asset health and reliability prediction: concepts, models, and algorithms. *Mathematical Problems in Engineering* (2010).

[56] Dong, S., Tang, B., and Chen, R. Bearing running state recognition based on non-extensive wavelet feature scale entropy and support vector machine. *Measurement 46*, 10 (2013), 4189–4199.

[57] Doucet, A., De Freitas, N., Gordon, N., et al. *Sequential Monte Carlo methods in practice., ISBN: 978-1-4757-3437-9*, vol. 1. Springer, 2001.

[58] Doucet, A., Godsill, S., and Andrieu, C. *On sequential simulation-based methods for Bayesian filtering.* Department of Engineering, University of Cambridge UK, 1998.

[59] Dragomir, O. E., Dragomir, F., Gouriveau, R., and Minca, E. Medium term load forecasting using anfis predictor. In *Control & Automation (MED), 2010 18th Mediterranean Conference on* (2010), IEEE, pp. 551–556.

[60] Dragomir, O. E., Gouriveau, R., Dragomir, F., Minca, E., Zerhouni, N., et al. Review of prognostic problem in condition-based maintenance. In *European Control Conference, ECC'09* (2009), pp. 1585–1592.

[61] Dragomir, O. E., Gouriveau, R., Zerhouni, N., Dragomir, F., et al. Framework for a distributed and hybrid prognostic system. In *4th IFAC Conference on Management and Control of Production and Logistics, MCPL'2007* (2007), vol. 3, pp. 431–436.

[62] Duan, M. *Time Series Predictability.* Graduate School Milwaukee: Marquette University, 2002.

[63] Eker, O., Camci, F., and Jennions, I. Major challenges in prognostics: Study on benchmarking prognostics datasets. *PHM Europe, 6th–8th July* (2012).

[64] El-Koujok, M., Gouriveau, R., and Zerhouni, N. Towards a neuro-fuzzy system for time series forecasting in maintenance applications. In *IFAC World Congress, Korea* (2008).

[65] El-Koujok, M., Gouriveau, R., and Zerhouni, N. Reducing arbitrary choices in model building for prognostics: An approach by applying parsimony principle on an evolving neuro-fuzzy system. *Microelectronics Reliability 51* (2011), 310–320.

[66] EN, N. 13306. *Terminologie de la maintenance* (2001).

[67] Engel, S. J., Gilmartin, B. J., Bongort, K., and Hess, A. Prognostics, the real issues involved with predicting life remaining. In *Aerospace Conference Proceedings, 2000 IEEE* (2000), vol. 6, IEEE, pp. 457–469.

[68] Ertunc, H. M., and Oysu, C. Drill wear monitoring using cutting force signals. *Mechatronics 14*, 5 (2004), 533–548.

[69] Feng, G., Huang, G.-B., Lin, Q., and Gay, R. Error minimized extreme learning machine with growth of hidden nodes and incremental learning. *Neural Networks, IEEE Transactions on 20*, 8 (2009), 1352–1357.

[70] Flandrin, P., Rilling, G., and Goncalves, P. Empirical mode decomposition as a filter bank. *Signal Processing Letters, IEEE 11*, 2 (2004), 112–114.

[71] Francesca, M. *Development of advanced computational methods for Prognostics and Health Management in energy components and systems.* PhD thesis, Politecnico di Milano, 2013.

[72] Frederick, D. K., DeCastro, J. A., and Litt, J. S. User's guide for the commercial modular aero-propulsion system simulation (c-mapss).

[73] Gardner, W. A., Napolitano, A., and Paura, L. Cyclostationarity: Half a century of research. *Signal processing 86*, 4 (2006), 639–697.

[74] Gorjian, N., Ma, L., Mittinty, M., Yarlagadda, P., and Sun, Y. A review on reliability models with covariates. In *Engineering Asset Lifecycle Management.* Springer, 2010, pp. 385–397.

[75] Gouriveau, R., Ramasso, E., Zerhouni, N., et al. Strategies to face imbalanced and unlabelled data in phm applications. *Chemical Engineering Transactions 33* (2013), 115–120.

[76] Gouriveau, R., and Zerhouni, N. Connexionist-systems-based long term prediction approaches for prognostics. *Reliability, IEEE Transactions on 61*, 4 (2012), 909–920.

[77] Guopeng, Z., Zhiqi, S., Chunyan, M., and Zhihong, M. On improving the conditioning of extreme learning machine: a linear case. In *Information, Communications and Signal Processing. ICICS 09. 7th International Conference on* (Piscataway, NJ, USA, 2009).

[78] Gupta, J. S., Trinquier, C., Lorton, A., and Feuillard, V. Characterization of prognosis methods: an industrial approach. In *European Conference of the Prognostics and Health Management Society* (2012).

[79] Haddadi, E., M. R Shabghard, and M. M Ettefagh. Effect of different tool edge conditions on wear detection by vibration spectrum analysis in turning operation. *Journal of Applied Sciences 8*, 21 (2008), 3879–3886.

[80] Hagan, M.-T., Demuth, H.-B., and Beale, M.-H. *Neural Network Design.* Boston, MA: PWS Publishing, 1996.

[81] Hagan, M. T., and Menhaj, M. B. Training feedforward networks with the marquardt algorithm. *Neural Networks, IEEE Transactions on 5*, 6 (1994), 989–993.

[82] Hansen, R. J., Hall, D. L., and Kurtz, S. K. A new approach to the challenge of machinery prognostics. *Journal of engineering for gas turbines and power 117*, 2 (1995), 320–325.

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

[84] Heng, A., Zhang, S., Tan, A. C., and Mathew, J. Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing 23*, 3 (2009), 724–739.

[85] Hess, A., Calvello, G., and Dabney, T. Phm a key enabler for the jsf autonomic logistics support concept. In *Aerospace Conference, 2004. Proceedings. 2004 IEEE* (2004), vol. 6, IEEE, pp. 3543–3550.

[86] Hess, A., Calvello, G., and Frith, P. Challenges, issues, and lessons learned chasing the. In *Aerospace Conference, 2005 IEEE* (2005), IEEE, pp. 3610–3619.

[87] Hess, A., Stecki, J. S., and Rudov-Clark, S. D. The maintenance aware design environment: Development of an aerospace phm software tool. In *Proc. PHM08* (2008).

[88] Hong, S., Zhou, Z., and Lv, C. Storage lifetime prognosis of an intermediate frequency (if) amplifier based on physics of failure method. *Chemical Engineering 33* (2013).

[89] Hu, C., Youn, B. D., Wang, P., and Taek Yoon, J. Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliability Engineering & System Safety 103* (2012), 120–135.

[90] Huang, G.-B., and Chen, L. Convex incremental extreme learning machine. *Neurocomputing 70*, 16 (2007), 3056–3062.

[91] Huang, G.-B., and Chen, L. Enhanced random search based incremental extreme learning machine. *Neurocomputing 71*, 16 (2008), 3460–3468.

[92] Huang, G.-B., Chen, L., and Siew, C.-K. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *Neural Networks, IEEE Transactions on 17*, 4 (2006), 879–892.

[93] Huang, G.-B., Wang, D. H., and Lan, Y. Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics 2*, 2 (2011), 107–122.

[94] Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. Extreme learning machine: a new learning scheme of feedforward neural networks. In *Neural Networks, Proceedings. 2004 IEEE International Joint Conference on* (2004), vol. 2, IEEE, pp. 985–990.

[95] Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. Extreme learning machine: theory and applications. *Neurocomputing 70*, 1 (2006), 489–501.

[96] Huang, N. E., and Shen, S. S. *Hilbert-Huang transform and its applications*, vol. 5. World Scientific, 2005.

[97] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. 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 454*, 1971 (1998), 903–995.

[98] Huang, R., Xi, L., Li, X., Richard Liu, C., Qiu, H., and Lee, J. 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), 193–207.

[99] ISO13381-1. *Condition monitoring and diagnostics of machines prognostics Part1: General guidelines.* International Standard, ISO, 2004.

[100] Iwaniec, J., Lisowski, W., and Uhl, T. Nonparametric approach to improvement of quality of modal parameters estimation. *Journal of Theoretical and Applied Mechanics 43* (2005), 327–344.

[101] Jaeger, H. The echo state approach to analyzing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report 148* (2001).

[102] Jaeger, H. *Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the echo state network approach.* GMD-Forschungszentrum Informationstechnik, 2002.

[103] Jalab, H. A., and Ibrahim, R. W. New activation functions for complex-valued neural network. *International Journal of the Physical Sciences 6*, 7 (2011), 1766–1772.

[104] Jang, J.-S., and Sun, C.-T. Neuro-fuzzy modeling and control. *Proceedings of the IEEE 83*, 3 (1995), 378–406.

[105] Jardine, A. K., Lin, D., and Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing 20*, 7 (2006), 1483–1510.

[106] Jarrell, D. B., SISK, D. R., and Bond, L. J. Prognostics and condition-based maintenance: A new approach to precursive metrics. *Nuclear technology 145*, 3 (2004), 275–286.

[107] Javed, K., Gouriveau, R., Zemouri, R., Zerhouni, N., et al. Improving data-driven prognostics by assessing predictability of features. *Prognostics and Health Management Society 2011* (2011), 555–560.

[108] Javed, K., Gouriveau, R., Zemouri, R., Zerhouni, N., et al. Features selection procedure for prognostics: An approach based on predictability. *8th IFAC Int. Symp. On Fault Dectection, Supervision and Safety of Technical Processes* (2012), 25–30.

[109] Javed, K., Gouriveau, R., and Zerhouni, N. Novel failure prognostics approach with dynamic thresholds for machine degradation. In *IECON 2013-39th Annual Conference on IEEE Industrial Electronics Society, Vienna, Austria*, IEEE, pp. 4402–4407.

[110] Javed, K., Gouriveau, R., Zerhouni, N., and Nectoux, P. A feature extraction procedure based on trigonometric functions and cumulative descriptors to enhance prognostics modeling. In *Prognostics and Health Management (PHM), 2013 IEEE Conference on* (2013), IEEE, pp. 1–7.

[111] Javed, K., Gouriveau, R., Zerhouni, N., Zemouri, R., and Li, X. Robust, reliable and applicable tool wear monitoring and prognostic: approach based on

an improved-extreme learning machine. In *IEEE Conf. on Prognostics and Health Management* (Denver, CO, USA, 2012).

[112] Jouin, M., Gouriveau, R., Hissel, D., Pra, M.-C., and Zerhouni, N. Prognostics of {PEM} fuel cell in a particle filtering framework. *International Journal of Hydrogen Energy 39*, 1 (2014), 481–494.

[113] Kaboudan, M. A measure of time series predictability using genetic programming applied to stock returns. *Journal of Forecasting 18*, 5 (1999), 345–357.

[114] Khosravi, A., Nahavandi, S., Creighton, D., and Atiya, A. F. Comprehensive review of neural network-based prediction intervals and new advances. *Neural Networks, IEEE Transactions on 22*, 9 (2011), 1341–1356.

[115] Kothamasu, R., Huang, S. H., and VerDuin, W. H. System health monitoring and prognostics-a review of current paradigms and practices. *The International Journal of Advanced Manufacturing Technology 28*, 9-10 (2006), 1012–1024.

[116] Kumar, S., and Pecht, M. Modeling approaches for prognostics and health management of electronics. *International Journal of Performability Engineering 6*, 5 (2010), 467–476.

[117] Kumar, S., Torres, M., Chan, Y., and Pecht, M. A hybrid prognostics methodology for electronic products. In *Neural Networks. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on* (2008), IEEE, pp. 3479–3485.

[118] Lau, E. C., and Ngan, H. Detection of motor bearing outer raceway defect by wavelet packet transformed motor current signature analysis. *Instrumentation and Measurement, IEEE Transactions on 59*, 10 (2010), 2683–2690.

[119] Lebold, M., and Thurston, M. Open standards for condition-based maintenance and prognostic systems. In *5th Annual Maintenance and Reliability Conference* (2001).

[120] Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., and Siegel, D. Prognostics and health management design for rotary machinery systemsreviews, methodology and applications. *Mechanical Systems and Signal Processing 42*, 1 (2014), 314–334.

[121] Leem, C. S., Dornfeld, D., and Dreyfus, S. A customized neural network for sensor fusion in on-line monitoring of cutting tool wear. *Journal of engineering for industry 117*, 2 (1995), 152–159.

[122] Lei, Y., Lin, J., He, Z., and Zuo, M. J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing 35*, 1 (2013), 108–126.

[123] Li, C., and Cheng, K.-H. Recurrent neuro-fuzzy hybrid-learning approach to accurate system modeling. *Fuzzy Sets and Systems 158*, 2 (2007), 194–212.

[124] Li, M.-B., Huang, G.-B., Saratchandran, P., and Sundararajan, N. Fully complex extreme learning machine. *Neurocomputing 68* (2005), 306–314.

[125] Li, R.-P., and Mukaidono, M. A maximum-entropy approach to fuzzy clustering. In *Fuzzy Systems, 1995. International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium., Proceedings of 1995 IEEE International Conference on* (1995), vol. 4, IEEE, pp. 2227–2232.

[126] Li, X., Lim, B., Zhou, J., Huang, S., Phua, S., Shaw, K., and Er, M. Fuzzy neural network modelling for tool wear estimation in dry milling operation. In *Annual Conference of the Prognostics and Health Management Society* (2009).

[127] Li, X., and Yao, X. Multi-scale statistical process monitoring in machining. *Industrial Electronics, IEEE Transactions on 52*, 3 (2005), 924–927.

[128] Liao, L. *An adaptive modeling for robust prognostics on a reconfigurable platform.* PhD thesis, University of Cincinnati, 2010.

[129] Liao, L. Discovering prognostic features using genetic programming in remaining useful life prediction. *Industrial Electronics, IEEE Transactions on 61*, 5 (2014), 2464–2472.

[130] Liu, J., Djurdjanovic, D., Marko, K. A., and Ni, J. A divide and conquer approach to anomaly detection, localization and diagnosis. *Mechanical Systems and Signal Processing 23*, 8 (2009), 2488–2499.

[131] Liu, J., Wang, W., Ma, F., Yang, Y., and Yang, C. A data-model-fusion prognostic framework for dynamic system state forecasting. *Engineering Applications of Artificial Intelligence 25*, 4 (2012), 814–823.

[132] Loutas, T., and Kostopoulos, V. *Utilising the Wavelet Transform in Condition-Based Maintenance: A Review with Applications.* 2012.

[133] Luo, J., Namburu, M., Pattipati, K., Qiao, L., Kawamoto, M., and Chigusa, S. Model-based prognostic techniques [maintenance applications]. In *AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference. Proceedings* (2003), IEEE, pp. 330–340.

[134] Mahamad, A. K., Saon, S., and Hiyama, T. Predicting remaining useful life of rotating machinery based artificial neural network. *Computers & Mathematics with Applications 60*, 4 (2010), 1078–1087.

[135] Mallat, S. G. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 11*, 7 (1989), 674–693.

[136] Massol, O., Li, X., Gouriveau, R., Zhou, J., and Gan, O. P. An exts based neuro-fuzzy algorithm for prognostics and tool condition monitoring. In *Control Automation Robotics & Vision (ICARCV), 2010 11th International Conference on* (2010), IEEE, pp. 1329–1334.

[137] MATLAB. *curve fitting toolbox.* The MathWorks Inc., Natick, Massachusetts, 2010.

[138] Medjaher, K., Tobon-Mejia, D. A., and Zerhouni, N. Remaining useful life estimation of critical components with application to bearings. *Reliability, IEEE Transactions on 61*, 2 (2012), 292–302.

[139] Meireles, M. R., Almeida, P. E., and Simões, M. G. A comprehensive review for industrial applicability of artificial neural networks. *Industrial Electronics, IEEE Transactions on 50*, 3 (2003), 585–601.

[140] MIMOSA-CBM. Condition-based maintenance http://www.mimosa.org/?q=wiki/mimosa-osa-cbm.

[141] Mosallam, A., Medjaher, K., and Zerhouni, N. Nonparametric time series modelling for industrial prognostics and health management. *The International Journal of Advanced Manufacturing Technology 69*, 5-8 (2013), 1685–1699.

[142] Mosallam, A., Medjaher, K., Zerhouni, N., et al. Bayesian approach for remaining useful life prediction. *Chemical Engineering Transactions 33* (2013), 139–144.

[143] NASA. Prognostic data repository http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/.

[144] Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Morello, B., Zerhouni, N., and Varnier, C. Pronostia: An experimental platform for bearings accelerated life test. In *IEEE International Conference on Prognostics and Health Management, Denver, CO, USA* (2012).

[145] Nguyen, D., and Widrow, B. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In *Neural Networks, 1990., 1990 IJCNN International Joint Conference on* (1990), IEEE, pp. 21–26.

[146] Niu, G., Yang, B.-S., and Pecht, M. Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance. *Reliability Engineering & System Safety 95*, 7 (2010), 786–796.

[147] Nystad, B. H., Gola, G., and Hulsund, J. E. Lifetime models for remaining useful life estimation with randomly distributed failure thresholds. In *First european conference of the prognostics and health management society* (2012), vol. 3.

[148] Olhede, S., and Walden, A. The hilbert spectrum via wavelet projections. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 460*, 2044 (2004), 955–975.

[149] Osipov, V., Luchinsky, D., Smelyanskiy, V., and et al. In-flight failure decision and prognostics for the solid rocket booster. In *AIAA 43rd AIAA/ASME/SAE/ ASEE Joint Propulsion Conference and Exhibit, Cincinnati, OH, July 8-11* (2007).

[150] Oussar, Y., and Dreyfus, G. Initialization by selection for wavelet network training. *Neurocomputing 34*, 1 (2000), 131–143.

[151] Parviainen, E., Riihimäki, J., Miche, Y., and Lendasse, A. Interpreting extreme learning machine as an approximation to an infinite neural network. In *In KDIR* (2010), pp. 65–73.

[152] Pavelka, A., and Prochazka, A. Algorithms for initialization of neural network weights. In *In Proceedings of the 12th Annual Conference, MATLAB 2004* (2004), vol. 2, pp. 453–459.

[153] Pecht, M. *Prognostics and health management of electronics.* Wiley Online Library, 2008.

[154] Pecht, M., and Gu, J. Physics-of-failure-based prognostics for electronic products. *Transactions of the Institute of Measurement and Control 31*, 3-4 (2009), 309–322.

[155] Pecht, M., and Jaai, R. A prognostics and health management roadmap for information and electronics-rich systems. *Microelectronics Reliability 50*, 3 (2010), 317–323.

[156] Peng, Y., Dong, M., and Zuo, M. J. Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology 50*, 1-4 (2010), 297–313.

[157] Penha, R.L., t. H., and B.R. Upadhyaya. Application of hybrid modeling for monitoring heat exchangers. In *3rd Meeting of the Americas - America's Nuclear Energy Symposium,* (Miami, FL, October 16-18, 2002).

[158] Peysson, F., Ouladsine, M., Noura, H., Leger, J.-B., and Allemand, C. New approach to prognostic systems failures. In *Proceedings of the 17th IFAC World Congress* (2007).

[159] Pourtaghi, A., and Lotfollahi-Yaghin, M. Wavenet ability assessment in comparison to ann for predicting the maximum surface settlement caused by tunneling. *Tunnelling and Underground Space Technology 28* (2012), 257–271.

[160] Psichogios, D. C., and Ungar, L. H. A hybrid neural network-first principles approach to process modeling. *AIChE Journal 38*, 10 (1992), 1499–1511.

[161] Rafiee, J., Rafiee, M., and Tse, P. Application of mother wavelet functions for automatic gear and bearing fault diagnosis. *Expert Systems with Applications 37*, 6 (2010), 4568–4579.

[162] Rajesh, R., and Prakash, J. S. Extreme learning machines-a review and state-of-the-art. *International journal of wisdom based computing 1* (2011), 35–49.

[163] Ramasso, E., and Denoeux, T. Making use of partial knowledge about hidden states in hmms: an approach based on belief functions. *IEEE Transactions on Fuzzy Systems (submitted)* (2013).

[164] Ramasso, E., and Gouriveau, R. Prognostics in switching systems: Evidential markovian classification of real-time neuro-fuzzy predictions. In *Prognostics and Health Management Conference, 2010. PHM'10* (2010), IEEE, pp. 1–10.

[165] Ramasso, E., Placet, V., Gouriveau, R., Boubakar, L., Zerhouni, N., et al. Health assessment of composite structures in unconstrained environments using partially supervised pattern recognition tools. *Prognostics and Health Management Society* (2012), 1–11.

[166] Ramasso, E., Rombaut, M., and Zerhouni, N. Joint prediction of continuous and discrete states in time-series based on belief functions. *Cybernetics, IEEE Tran. on 43*, 1 (2013), 37–50.

[167] Rao, C. R., and Mitra, S. K. *Generalized Inverse of Matrices and its Applications*. John Wiley and Sons, New York, 1971.

[168] Roemer, M. J., Nwadiogbu, E., and Bloor, G. Development of diagnostic and prognostic technologies for aerospace health management applications. In *Aerospace Conference, 2001, IEEE Proceedings* (2001), vol. 6, IEEE, pp. 3139–3147.

[169] Saha, B., and Goebel, K. Modeling li-ion battery capacity depletion in a particle filtering framework. In *Proceedings of the annual conference of the prognostics and health management society* (2009).

[170] Samhouri, M., Al-Ghandoor, A., Ali, S. A., Hinti, I., and Massad, W. An intelligent machine condition monitoring system using time-based analysis: neuro-fuzzy versus neural network. *Jordan Journal of Mechanical and Industrial Engineering 3*, 4 (2009), 294–305.

[171] Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., and Schwabacher, M. Metrics for evaluating performance of prognostic techniques. In *Prognostics and Health Management, 2008. PHM 2008. International Conference on* (2008), IEEE, pp. 1–17.

[172] Saxena, A., Celaya, J., Saha, B., Saha, S., and Goebel, K. Evaluating prognostics performance for algorithms incorporating uncertainty estimates. In *Aerospace Conference, 2010* (2010), IEEE, pp. 1–11.

[173] Saxena, A., Celaya, J., Saha, B., Saha, S., and Goebel, K. Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and Health Management 1*, 1 (2010), 20.

[174] Saxena, A., Celaya, J. R., Roychoudhury, I., Saha, S., Saha, B., and Goebel, K. Designing data-driven battery prognostic approaches for variable loading profiles: Some lessons learned. In *European Conference of Prognostics and Health Management Society* (2012).

[175] Saxena, A., Goebel, K., Simon, D., and Eklund, N. Damage propagation modeling for aircraft engine run-to-failure simulation. In *Prognostics and Health Management, 2008. PHM 2008. International Conference on* (2008), IEEE, pp. 1–9.

[176] Schwabacher, M., and Goebel, K. A survey of artificial intelligence for prognostics. In *AAAI Fall Symposium* (2007), pp. 107–114.

[177] Si, X.-S., Wang, W., Hu, C.-H., and Zhou, D.-H. Remaining useful life estimation–a review on the statistical data driven approaches. *European Journal of Operational Research 213*, 1 (2011), 1–14.

[178] Sikorska, J., Hodkiewicz, M., and Ma, L. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing 25*, 5 (2011), 1803–1836.

[179] Singh, R., and Balasundaram, S. Application of extreme learning machine method for time series analysis. *International Journal of Intelligent Technology 2*, 4 (2007), 256–262.

[180] Sun, H., He, Z., Zi, Y., Yuan, J., Wang, X., Chen, J., and He, S. Multiwavelet transform and its applications in mechanical fault diagnosis a review. *Mechanical Systems and Signal Processing 43*, 1-2 (2014), 1–24.

[181] Sun, Z.-L., Choi, T.-M., Au, K.-F., and Yu, Y. Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems 46*, 1 (2008), 411–419.

[182] Sutharssan, T., Stoyanov, S., Bailey, C., and Rosunally, Y. Prognostics and health monitoring of high power led. *Micromachines 3*, 1 (2012), 78–100.

[183] Teodorescu, H.-N., and Fira, L.-I. Analysis of the predictability of time series obtained from genomic sequences by using several predictors. *Journal of Intelligent and Fuzzy Systems 19*, 1 (2008), 51–63.

[184] C.M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, Walton Street, Oxford OX2 6DP, 1995.

[185] Thompson, M. L., and Kramer, M. A. Modeling chemical processes using prior knowledge and neural networks. *AIChE Journal 40*, 8 (1994), 1328–1340.

[186] Tobon-Mejia, D., Medjaher, K., and Zerhouni, N. Cnc machine tool's wear diagnostic and prognostic by using dynamic bayesian networks. *Mechanical Systems and Signal Processing 28* (2012), 167–182.

[187] Tobon-Mejia, D. A., Medjaher, K., Zerhouni, N., and Tripot, G. A data-driven failure prognostics method based on mixture of gaussians hidden markov models. *Reliability, IEEE Transactions on 61*, 2 (2012), 491–503.

[188] Torabi, A. J., Joo, E. M., Xiang, L., Siong, L. B., Lianyin, Z., Jie, P. S., Junhong, Z., Lin, S., Sheng, H., and Tijo, J. T. T. A survey on artificial intelligence technologies in modeling of high speed end-milling processes. In *Advanced Intelligent Mechatronics, 2009. AIM 2009. IEEE/ASME International Conference on* (2009), IEEE, pp. 320–325.

[189] Tran, V. T., and Yang, B.-S. Machine fault diagnosis and prognosis: The state of the art. *The International Journal of Fluid Machinery and Systems (IJFMS) 2*, 1 (2009), 61–71.

[190] Tran, V. T., and Yang, B.-S. An intelligent condition-based maintenance platform for rotating machinery. *Expert Systems with Applications 39*, 3 (2012), 2977–2988.

[191] Uckun, S., Goebel, K., and Lucas, P. Standardizing research methods for prognostics. In *PHM Int. Conf. on* (2008), pp. 1–10.

[192] Vachtsevanos, G., Lewis, F.-L., Roemer, M., and Hess, A. *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. New Jersey, Hoboken: Wiley & Sons, 2006.

[193] Van Noortwijk, J. A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety 94*, 1 (2009), 2–21.

[194] Waeyenbergh, G., and Pintelon, L. A framework for maintenance concept development. *International Journal of Production Economics 77*, 3 (2002), 299–313.

[195] Wang, J.-c., Hu, J., Huang, X.-x., and Xu, S.-c. Assessment of different genetic distances in constructing cotton core subset by genotypic values. *Journal of Zhejiang University SCIENCE B 9*, 5 (2008), 356–362.

[196] Wang, P., and Vachtsevanos, G. Fault prognostic using dynamic wavelet neural networks. *Artificial Intelligence for Engineering Design Analysis and Manufacturing 15*, 04 (2001), 349–365.

[197] Wang, T. *Trajectory similarity based prediction for remaining useful life estimation.* PhD thesis, University of Cincinnati, 2010.

[198] Wang, T. Bearing life prediction based on vibration signals: A case study and lessons learned. In *Prognostics and Health Management, IEEE Conference on* (USA, 2012).

[199] Wang, T., Yu, J., Siegel, D., and Lee, J. A similarity-based prognostics approach for remaining useful life estimation of engineered systems. In *International Conference on Prognostics and Health Management* (2008), IEEE, pp. 1–6.

[200] Wang, W. An adaptive predictor for dynamic system forecasting. *Mechanical Systems and Signal Processing 21*, 2 (2007), 809–823.

[201] Wang, W., Gelder, P. V., and J.K Vrijling, title = Measuring predictability of Daily Streamflow Processes Based on Univariate Time Series Model, b. . i. y. . . v. . . p. . .

[202] Wang, W. Q., Golnaraghi, M. F., and Ismail, F. Prognosis of machine health condition using neuro-fuzzy systems. *Mechanical Systems and Signal Processing 18*, 4 (2004), 813–831.

[203] Wang, Y., He, Z., and Zi, Y. A comparative study on the local mean decomposition and empirical mode decomposition and their applications to rotating machinery health diagnosis. *Journal of Vibration and Acoustics 132*, 2 (2010), 021010.

[204] Warren Liao, T. Clustering of time series dataa survey. *Pattern Recognition 38*, 11 (2005), 1857–1874.

[205] Wave-Tutorial. Tutorial on wavelets.

[206] Wu, T.-J., Hsieh, Y.-C., and Li, L.-A. Statistical measures of dna sequence dissimilarity under markov chain models of base composition. *Biometrics 57*, 2 (2001), 441–448.

[207] Wu, W., Hu, J., and Zhang, J. Prognostics of machine health condition using an improved arima-based prediction method. In *Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on* (2007), IEEE, pp. 1062–1067.

[208] Xu, R., Wunsch, D., et al. Survey of clustering algorithms. *Neural Networks, IEEE Transactions on 16*, 3 (2005), 645–678.

[209] Yager, R. R., and Filev, D. P. Approximate clustering via the mountain method. *Systems, Man and Cybernetics, IEEE Transactions on 24*, 8 (1994), 1279–1284.

[210] Yam, R., Tse, P., Li, L., and Tu, P. Intelligent predictive decision support system for condition-based maintenance. *The International Journal of Advanced Manufacturing Technology 17*, 5 (2001), 383–391.

[211] Yan, W., Qiu, H., and Iyer, N. Feature extraction for bearing prognostics and health management (phm)-a survey (preprint). Tech. rep., DTIC Document, 2008.

[212] Yang, Q. *Model-based and data driven fault diagnosis methods with applications to process monitoring*. PhD thesis, Case Western Reserve University, 2004.

[213] Yin, S., Luo, H., and Ding, S. Real-time implementation of fault-tolerant control systems with performance optimization. *Industrial Electronics, IEEE Transactions on 61*, 5 (2014), 2402–2411.

[214] Yu, J. A hybrid feature selection scheme and self-organizing map model for machine health assessment. *Applied Soft Computing 11*, 5 (2011), 4041–4054.

[215] Zaidi, H., Diaz-Gomez, M., Boudraa, A., and Slosman, D. Fuzzy clustering-based segmented attenuation correction in whole-body pet imaging. *Physics in Medicine and Biology 47*, 7 (2002), 1143–1160.

[216] Zarei, J., and Poshtan, J. Bearing fault detection using wavelet packet transform of induction motor stator current. *Tribology International 40*, 5 (2007), 763–769.

[217] Zemouri, R., Gouriveau, R., and Zerhouni, N. Improving the prediction accuracy of recurrent neural network by a pid controller. *International Journal of Systems Applications, Engineering & Development 4*, 2 (2010), 19–34.

[218] Zhai, L., Er, M., Li, X., Gan, O., Phua, S., Huang, S., Zhou, J., San, L., and Torabi, A. Intelligent monitoring of surface integrity and cutter degradation in high-speed milling processes. In *Annual Conference of the Prognostics and Health Management Society* (2010).

[219] Zhang, G. P. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing 50* (2003), 159–175.

[220] Zhang, H., Kang, R., and Pecht, M. A hybrid prognostics and health management approach for condition-based maintenance. In *Industrial Engineering and Engineering Management, 2009. IEEM 2009. IEEE International Conference on* (2009), IEEE, pp. 1165–1169.

[221] Zhang, X., and Pisu, P. An unscented kalman filter based approach for the health-monitoring and prognostics of a polymer electrolyte membrane fuel cell. In *Annual Conference of the Prognostics and Health Management Society* (2012).

[222] Zhou, J., Li, X., Gan, O. P., Han, S., and Ng, W. K. Genetic algorithms for feature subset selection in equipment fault diagnosis. In *Engineering Asset Management*. Springer, 2006, pp. 1104–1113.

[223] Zhou, J.-H., Pang, C. K., Lewis, F. L., and Zhong, Z.-W. Intelligent diagnosis and prognosis of tool wear using dominant feature identification. *Industrial Informatics, IEEE Transactions on 5*, 4 (2009), 454–464.

[224] Zhou, J.-H., Pang, C. K., Zhong, Z.-W., and Lewis, F. L. Tool wear monitoring using acoustic emissions by dominant-feature identification. *Instrumentation and Measurement, IEEE Transactions on 60*, 2 (2011), 547–559.

[225] Zhu, J.-H., and Li, H.-G. Subtractive fuzzy c-means clustering approach with applications to fuzzy predictive control. *WSEAS Transactions on Systems and Control 6*, 2 (2011), 33–43.

[226] Zio, E. Prognostics and health management of industrial equipment. *Diagnostics and Prognostics of Engineering Systems: Methods and Techniques* (2012), 333–356.

[227] Zio, E., and Di Maio, F. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. *Reliability Engineering & System Safety 95*, 1 (2010), 49–57.

Appendix A

Appendix A

A.1 Taxonomy of maintenance policies

According to literature, maintenance policies can be broadly classified into two categories namely: Corrective maintenance and Preventive maintenance, (see Fig. A.1), [78, 84, 146, 158, 189, 190, 220].

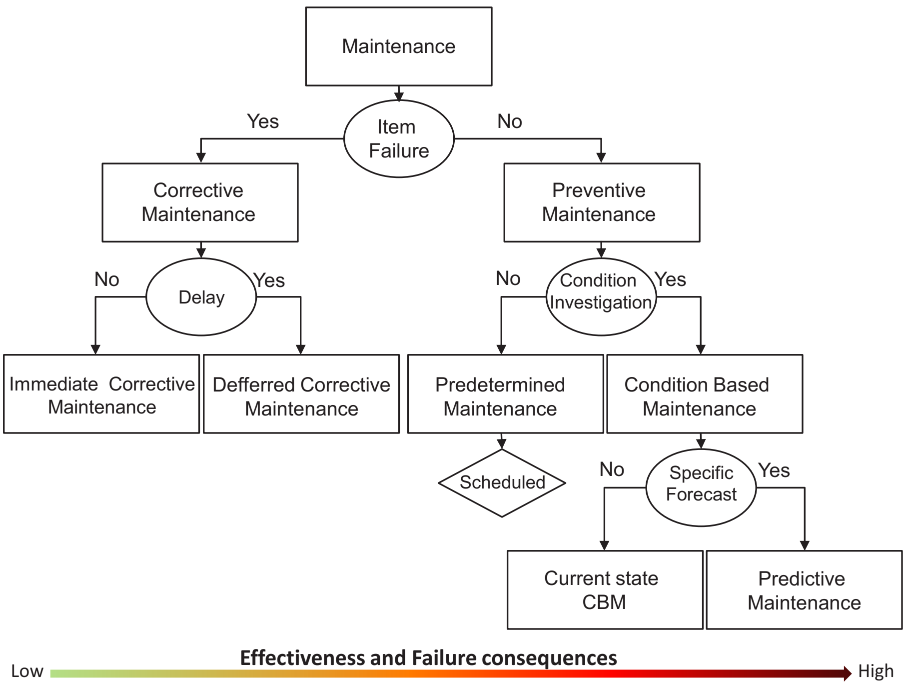

FIGURE A.1: Classification of maintenance policies in industry

A.1.1 Corrective maintenance

Corrective maintenance is an earliest form of maintenance that is unplanned (or reactive) and is undertaken after machinery breakdown. A failed item is repaired or replaced with a new one to store functional capabilities. This type of maintenance can be further categorized as immediate corrective maintenance or deferred corrective maintenance [78]. However, such maintenance is suitable for non-critical engineering assets, where failure consequences are slight and no immediate safety risks are involved. In other words, it should be used when there are no serious consequences of unscheduled downtime.

A.1.2 Preventive maintenance

Preventive maintenance is a planned (or proactive) strategy for maintaining machinery, that was introduced in 1950s [189]. The main objective was to increase availability of machinery as compared to corrective maintenance. Preventive maintenance can be further classified as predetermined maintenance and Condition Based Maintenance (CBM).

A.1.2.1 Predetermined maintenance

With predetermined maintenance repairs are performed on the basis of pre-defined schedule (or fixed intervals). Maintenance tasks are carried periodically (like lubrication, calibration, refurbishing, checking and inspecting machinery) after fixed time intervals to decrease the deterioration phenomena [190]. However, the maintenance procedures do not involve evaluation of current state in particular. Although predetermined maintenance can be a simple approach to maintain equipment, it can be costly due to unnecessary stoppages or replacements of operating machinery. Another problem of this approach is that failures are assumed to occur in well defined intervals [115].

A.1.2.2 Condition-Based Maintenance (CBM)

CBM strategy is a step forward from predetermined maintenance, as it utilizes real-time condition monitoring data to maintain machinery, and repair only when need arises [63, 158]. Therefore, unnecessary maintenance procedures are avoided and maintenance costs are reduced significantly. A CBM program has three main stages that are discussed in [105]: 1) data-acquisition, 2) data pre-processing and 3) decision making. The first stage is a building block for CBM procedure, where monitored information is gathered and stored. In the second stage feature extraction and selection tasks are performed. Lastly, the third stage of decision making recommends maintenance actions through diagnostics (for fault detection, isolation and identification) and prognostics of degrading equipment. Having knowledge of the current state of the equipment, and predicting its future behavior are the basis of CBM [106]. Therefore, in industrial context CBM approach can be of two types.

Current state CBM: in this case CBM is based only on the current-state of the equipment and do not include prediction / forecasting of future states of the degrading equipment. Therefore, maintenance decisions are based on the estimation of current-state only and equipment is run as long as it is healthy [78, 158].

Predictive maintenance: it is an innovative maintenance approach that is founded on assessment of current health state of equipment and on the prediction of future evolution (i.e., prognostics) using specific future conditions. It means, the same item could have different service life, when different future conditions are applied [18, 78, 105]. Identification of problems at early stages and estimation of remaining operating life enables the machinery to run till healthy state, and provide time to opportunely plan / scheduled necessary maintenance actions prior to break down.

A.2 Relation between diagnostics and prognostics

Like prognostics, diagnostic is also an important element of PHM cycle (section 1.1.1). In literature, there is a little disagreement, that prognostics is related to and highly dependent upon diagnostics. Mainly, diagnostics involves fault / failure detection, isolation and identification of damage that has occurred, while prognostics is concerned with RUL estimation and associated confidence interval (see Fig. A.2).
Therefore, in terms of relation between diagnostics and prognostics, diagnostics process detects and identifies a fault / failure that has caused degradation, where prognostics process generates a rational estimation of RUL until complete failure [178]. However, in general, diagnostics is a reactive process that is performed after the occurrence of fault / failure and cannot stop machine downtime in, where a prognostics is a predict-to-prevent behavior [120]. It should be noted that there is a difference between "fault diagnostics" and "failure diagnosis". The terminology of fault diagnostics implies that, the machinery under observation is still operational, however, cannot continue operating indefinitely without any maintenance intervention. Where, failure diagnostics is performed on machinery that has ceased to perform its functionality [35].
As compared to prognostic, diagnostics domain is well developed and has several applications in industry. The main approaches for diagnostics can be model based or data-driven, however these topics are not covered in the thesis (see [212]).

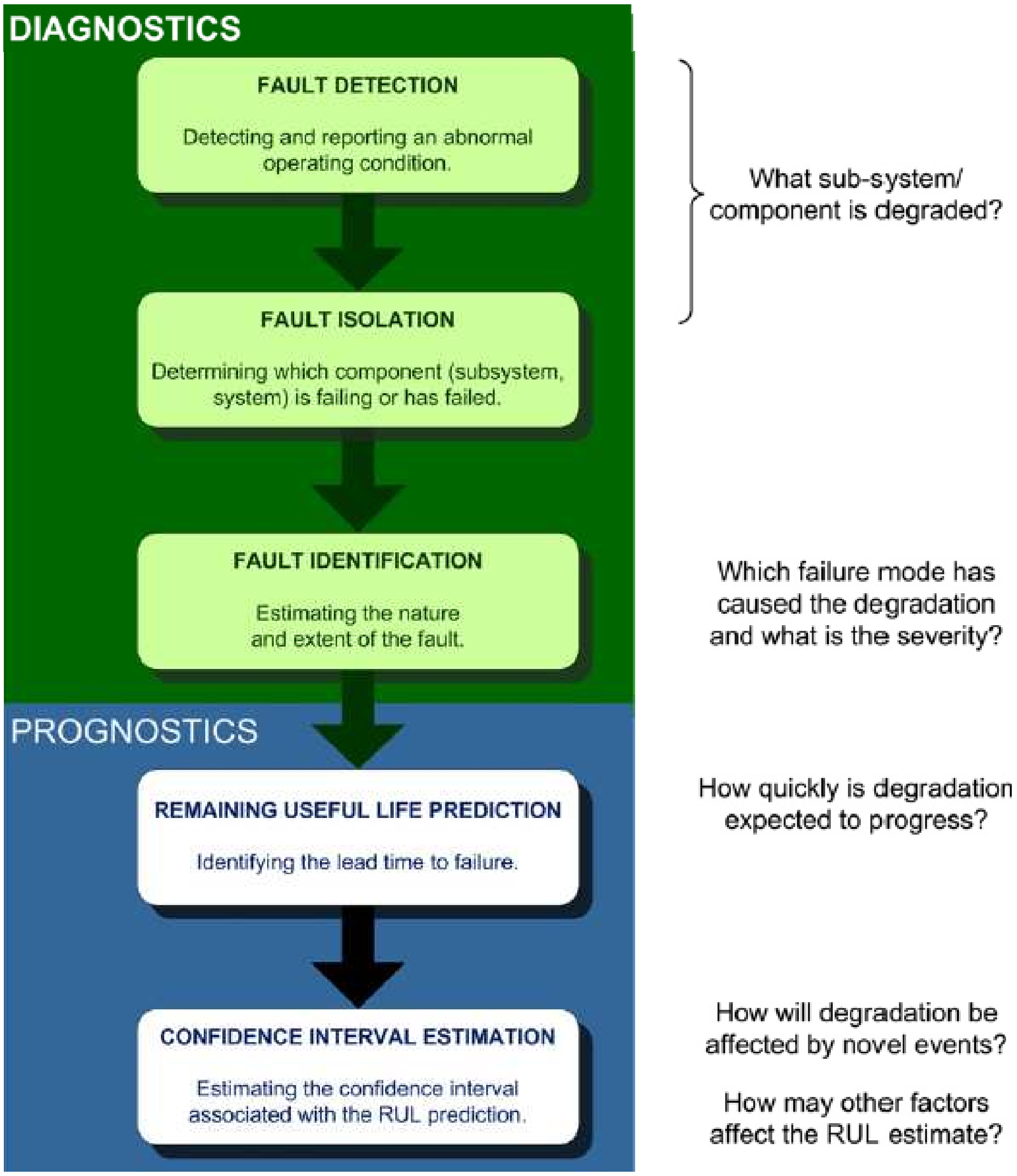

FIGURE A.2: Steps to obtain prognostics results & relationship to diagnostics [178]

A.3 FCM algorithm

According to [29], the main steps of FCM algorithms are summarized as follows.

Algorithm 4 Key steps of FCM algorithm

Require - Learning dataset Eq. (3.1)
- No. of clusters c, initial Cluster centers V, Fuzziness parameter m

Ensure - Cluster centers V and fuzzy partitioning U

learning procedure

1: Compute fuzzy partition matrix U

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{\frac{2}{m-1}}} \; \forall \, i,j \tag{A.1}$$

2: Adjust cluster centers v^{new}

$$v_j^{new} = \frac{\sum_{i=1}^{N} \mu_{ij}^{m} . x_i}{\sum_{i=1}^{N} \mu_{ij}^{m}} \; \forall \, j \tag{A.2}$$

3: Repeat step 1 and 2 untill termination criteria is met

$$\left\| v^{new} - v^{old} \right\| < \epsilon \tag{A.3}$$

A.4 Metrics for model accuracy

- MAPE: the mean absolute percentage error, or known as mean absolute percentage deviation (MAPD). MAPE usually expresses accuracy as a percentage and can be formulated as follows,

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{A.4}$$

  where y_i indicates the actual value and $\hat{y}_i$ indicates the predicted value.

- R2: evaluates the variability in the actual values explained by the model and can be computed by dividing explained variation by total variation:

$$R2 = 1 - \frac{SSE}{SS_{yy}} \tag{A.5}$$

where SSE is the measure of deviation of observations from the predicted values, i.e., $SSE = \sum_i (y_i - \hat{y}_i)^2$, and SS_{yy} is the measure of observation from the mean value, i.e., $SS_{yy} = \sum_i (y_i - \bar{y})^2$.

- CVRMSE: Coefficient of Variation of the Root Mean Squared Error evaluates the relative closeness of the predictions to the actual values and can be formulated as,

$$CVRMSE = 100 \times \frac{RMSE}{\bar{y}} \tag{A.6}$$

where $RMSE$ is the root mean square error i.e.,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{A.7}$$

It should be noted that, all metrics are unit less.

A.5 Benchmark datasets

A.5.1 Carnallite surge tank dataset

The dataset was provided by Arab Potash company [170]. Mainly, the data were acquired from a critical component called as Carnallite surge tank pump to perform condition monitoring task in an intelligent manner. The main objective was to replace visual inspection (human) for condition monitoring with an automated system to detect faults and avoid breakdowns.
Data were collected from component item 13, and were further divided into three bins, each containing two different time signal features, namely the root mean square (RMS) and the variance (σ^2) Fig. A.3, and their corresponding fault code Fig. A.4. The tests aim at approximating the output "fault codes" by feeding neural networks with two inputs (RMS and σ^2). The whole dataset consists of 92 samples, from which 73 samples were used to learn the models, whereas the remaining 19 were used for testing purpose.

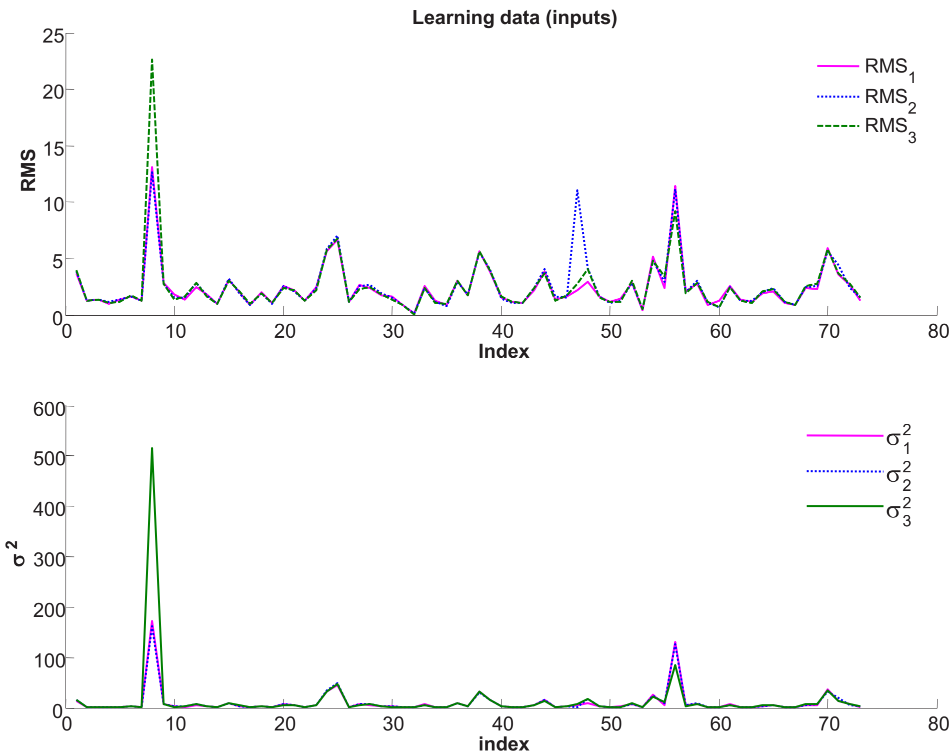

FIGURE A.3: Input data: RMS and variance

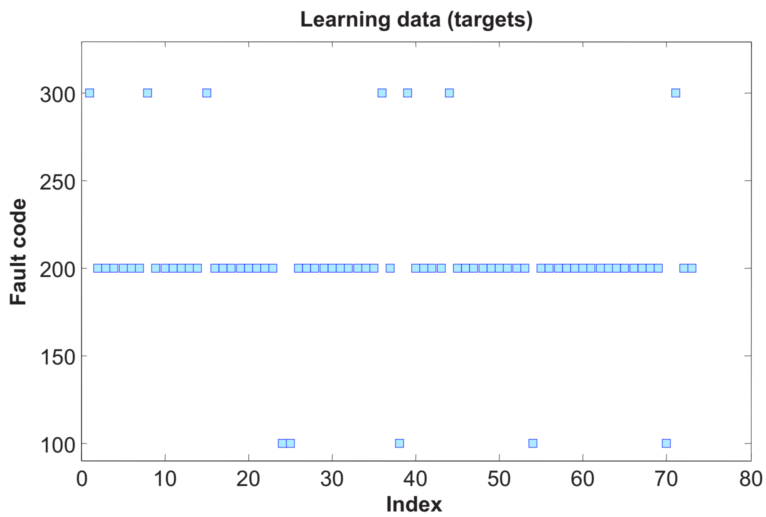

FIGURE A.4: Output targets

A.5.2 Industrial dryer dataset

The industrial dryer dataset used for one-step ahead prediction were contributed by Jan Maciejowski from Cambridge University [1]. Data consists of three exogenous variables, namely the fuel flow rate (x_1), the hot gas exhaust fan speed (x_2), and the rate of flow of raw material (x_3) Fig. A.5, and a corresponding endogenous variable, the dry bulb temperature (y). The tests aim at predicting the temperature of the dry bulb at next step $\hat{y}(t+1)$ by using current input values $x_1(t)$, $x_2(t)$, $x_3(t)$. An additional input of one regressors from the dry bulb temperature is also utilized ($y(t)$) Fig. A.6. The complete dataset consists of 867 samples, from which 500 samples were used to learn the models, whereas the remaining 367 were used for testing purpose.

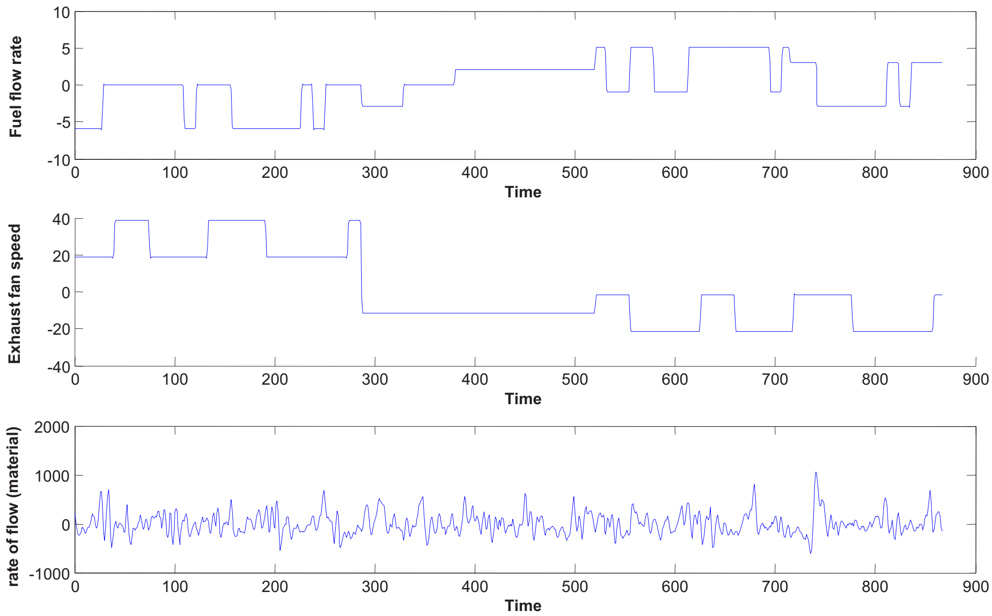

FIGURE A.5: Input data

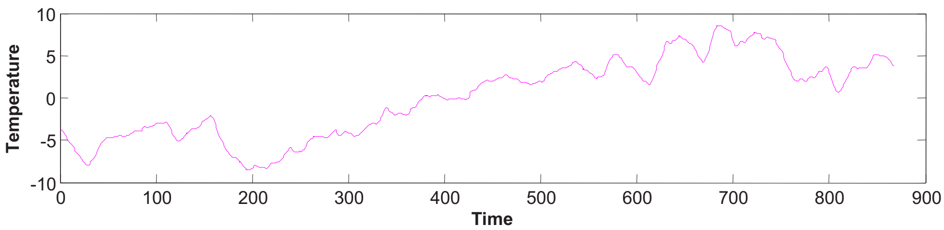

FIGURE A.6: Output: dry bulb temperature

A.5.3 Mechanical hair dryer dataset

The hair dryer dataset used for one-step ahead prediction were contributed by W. Favoreel from the Kuleuven Unversity [2]. The air temperature of the mechanical hair dryer is linked to voltage of the heating device. The data have one exogenous variable x_1 for the voltage of the heating device, and one endogenous variable y for the air temperature Fig. A.7.

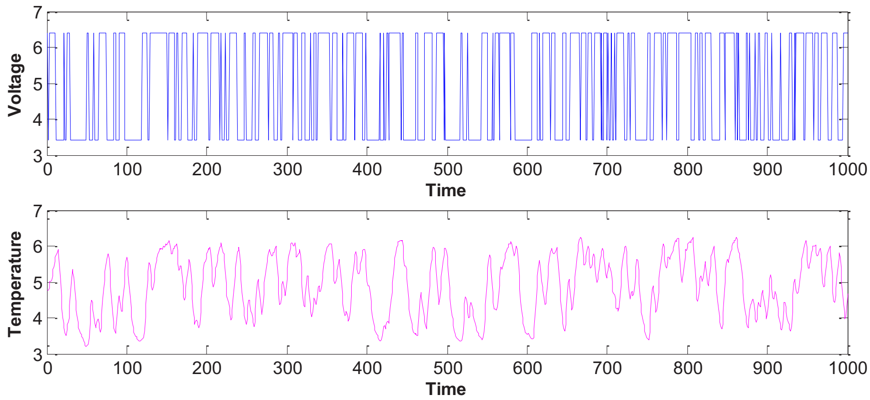

FIGURE A.7: Voltage of heading device and air temperature

A.5.4 NN3 forecasting data

To evaluate multi-steps ahead prediction performances, simulations were made on data sets from the NN3 competition, that were originally provided to test the accuracy of computational intelligence methods (notably neural networks) in time series forecasting [6]. Mainly, the advantage of the data resides in diversity and quantity of time series: the complete data are composed of 111 monthly time series derived from homogeneous population of empirical business time.
Among all data, tests were carried out on time series numbers 51, 54, 56, 58, 60, 61, 92, 106 (randomly chosen), without data processing. A plot of those time series is given in Fig. A.8 which reflects the diversity of behaviors to be predicted. As for the horizon of prediction (and according to NN3 competition), the last 18 values of each time series were used for test, whereas the remaining data (the previous ones) were used to learn the models.

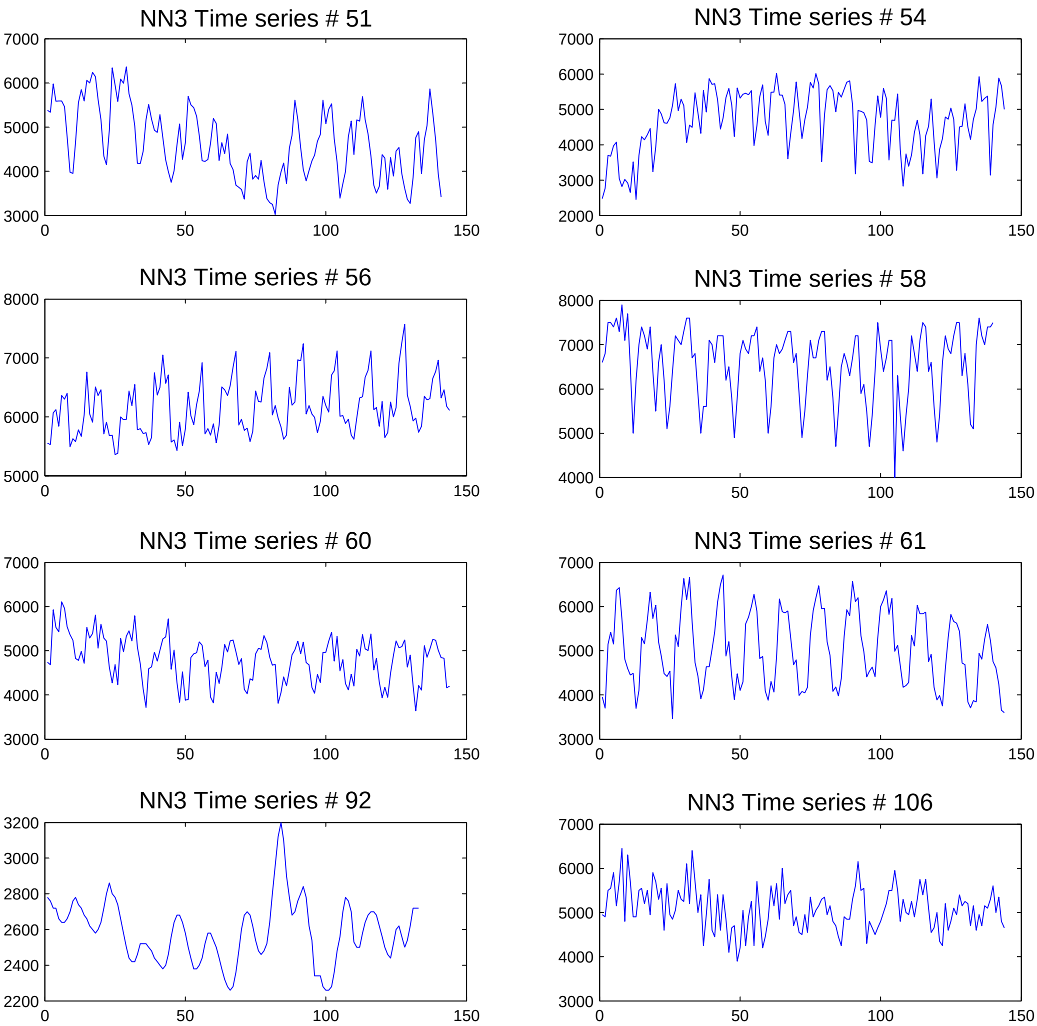

FIGURE A.8: Time series from NN3 datasets for multi-steps predictions

Publications *(February 2011 - January 2014)*

Journal Articles

- Javed K., Gouriveau R., Zerhouni N, SW-ELM: A summation wavelet extreme learning machine algorithm with *a priori* parameter initialization, *Neurocomputing*, Vol. 123, pp. 299-307, January 2014.
- Javed K., Gouriveau R., Zerhouni N, Enabling health monitoring approach based on vibration data for accurate prognostics, *IEEE Transactions on Industrial Electronics*, Accepted.
- Javed K., Gouriveau R., Zerhouni N, , Li X., Improving Robustness and Reliability of Tool Wear Monitoring and Prognostics: A Summation Wavelet-Extreme Learning Machine Approach, *Engineering Applications of Artificial Intelligence*, Revised Manuscript Under Review.

In progress

- Javed K., Gouriveau R., Zerhouni N, A new multivariate approach for prognostics based on Extreme Learning Machine & Fuzzy clustering.
- Javed K., Gouriveau R., Zerhouni N, A survey and taxonomy of prognostics approaches.

Conferences

- Javed K., Gouriveau R., Zerhouni N, Novel failure prognostics approach with dynamic thresholds for machine degradation, In: Proceedings of the 39th Annual Conference of the IEEE Industrial Electronics Society, IECON 2013, Vienna, Austria, 10-13 November 2013.
- Javed K., Gouriveau R., Zerhouni N, A feature extraction procedure based on trigonometric functions and cumulative descriptors to enhance prognostics modeling and estimates, In: Proceedings of the IEEE International Conference on Prognostics and Health Management, IEEE PHM 2013, Gaithersburg, MD, USA, 24-27 June 2013.
- Javed K., Gouriveau R., Zemouri R., Zerhouni N, Features Selection Procedure for Prognostics: An Approach Based on Predictability, In: Proceedings of the 8th IFAC International Symposium on Fault Detection, Supervision and Safety for Technical Processes, SAFEPROCESS 2012, Mexico, 29-31 August 2012.

- Javed K., Gouriveau R., Zerhouni N., Zemouri R., Li X., Robust, reliable and applicable tool wear monitoring and prognostic: approach based on an Improved-Extreme Learning Machine, In: Proceedings of the IEEE International Conference on Prognostics and Health Management, IEEE PHM 2012, Denver, CO, USA, 18-21 June 2012.
- Javed K., Gouriveau R., Zemouri R., Zerhouni N, Improving data-driven prognostics by assessing predictability of features, In: Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM 2011, Montreal, QC, Canada, 25-29 September 2011.

Résumé :

Le pronostic industriel vise à étendre le cycle de vie d'un dispositif physique, tout en réduisant les coûts d'exploitation et de maintenance. Pour cette raison, le pronostic est considéré comme un processus clé avec des capacités de prédiction. En effet, des estimations précises de la durée de vie avant défaillance d'un équipement, Remaining Useful Life (RUL), permettent de mieux définir un plan d'action visant à accroître la sécurité, réduire les temps d'arrêt, assurer l'achèvement de la mission et l'efficacité de la production.
Des études récentes montrent que les approches guidées par les données sont de plus en plus appliquées pour le pronostic de défaillance. Elles peuvent être considérées comme des modèles de type boite noire pour l'étude du comportement du système directement à partir des données de surveillance d'état, pour définir l'état actuel du systèm et prédire la progression future de défauts. Cependant, l'approximation du comportement des machines critiques est une tâche difficile qui peut entraîner des mauvais pronostic. Pour la compréhension de la modélisation du pronostic guidé par les données, on considère les points suivants. 1) Comment traiter les données brutes de surveillance pour obtenir des caractéristiques appropriées reflétant l'évolution de la dégradation? 2) Comment distinguer les états de dégradation et définir des critères de défaillance (qui peuvent varier d'un cas à un autre)? 3) Comment être sûr que les modèles définis seront assez robustes pour montrer une performance stable avec des entrées incertaines s'écartant des expériences acquises, et seront suffisamment fiables pour intégrer des données inconnues (c'est à dire les conditions de fonctionnement, les variations de l'ingénierie, etc.)? 4) Comment réaliser facilement une intégration sous des contraintes et des exigences industrielles? Ces questions sont des problèmes abordés dans cette thèse. Elles ont conduit à développer une nouvelle approche allant au-delà des limites des méthodes classiques de pronostic guidé par les données. Les principales contributions sont les suivantes.

– L'étape de traitement des données est améliorée par l'introduction d'une nouvelle approche d'extraction des caractéristiques à l'aide de fonctions trigonométriques et cumulatives qui sont basées sur trois caractéristiques : la monotonie, la "trendability" et la prévisibilité. L'idée principale de ce développement est de transformer les données brutes en indicateurs qui améliorent la précision des prévisions à long terme.
– Pour tenir compte de la robustesse, de la fiabilité et de l'applicabilité, un nouvel algorithme de prédiction est proposé: Summation Wavelet-Extreme Learning Machine (SW-ELM). Le SW-ELM assure de bonnes performances de prédiction, tout en réduisant le temps d'apprentissage. Un ensemble de SW-ELM est également proposé pour quantifier l'incertitude et améliorer la précision des estimations.
– Les performances du pronostic sont également renforcées grâce à la proposition d'un nouvel algorithme d'évaluation de la santé du systèm: Subtractive-Maximum Entropy Fuzzy Clustering (S-MEFC). S-MEFC est une approche de classification non supervisée qui utilise l'inférence de l'entropie maximale pour représenter l'incertitude de données multidimensionnelles. Elle peut automatiquement déterminer le nombre d'états, sans intervention humaine.
– Le modèle de pronostic final est obtenu en intégrant le SW-ELM et le S-MEFC pour montrer l'évolution de la dégradation de la machine avec des prédictions simultanées et l'estimation d'états discrets. Ce programme permet également de définir dynamiquement les seuils de défaillance et d'estimer le RUL des machines surveillées.

Les développements sont validés sur des données réelles à partir de trois plates-formes expérimentales: PRONOSTIA FEMTO-ST (banc d'essai des roulements), CNC SIMTech (fraises d'usinage), C-MAPSS NASA (turboréacteurs) et d'autres données de référence. Toutefois, la perspective principale de ce travail est d'améliorer la fiabilité du modèle de pronostic.
Mots-clés : Prognostics, Data-driven, Extreme learning Machine, Fuzzy Clustering, RUL

SPIM

■ 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