UNIVERSIDADE FEDERAL DE PERNAMBUCO

DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

CAIO BEZERRA SOUTO MAIOR

REMAINING USEFUL LIFE PREDICTION VIA EMPIRICAL MODE DECOMPOSITION, WAVELETS AND SUPPORT VECTOR MACHINE

RECIFE, 2017

CAIO BEZERRA SOUTO MAIOR

REMAINING USEFUL LIFE PREDICTION VIA EMPIRICAL MODE DECOMPOSITION, WAVELETS AND SUPPORT VECTOR MACHINE

Master thesis presented to UFPE for the master's degree attainment as part of the requirements of the Programa de Pós-Graduação em Engenharia de Produção (concentration area: Operations Research).

Advisor: Prof.ª Isis Didier Lins, DSc.

Co-Advisor: Prof. Márcio José das Chagas Moura, DSc

RECIFE, /2017

Catalogação na fonte

Bibliotecária: Rosineide Mesquita Gonçalves Luz / CRB4-1361 (BCTG)

S728r Souto Maior, Caio Bezerra.
Remainig useful life prediction via empirical mode decomposition, wavelets and support vector machine/ Caio Bezerra Souto Maior. – Recife, 2017.
62f., il., figs., gráfs., tabs.

Orientadora: Profa. Isis Didier Lins, DSc.
Coorientador: Prof. Márcio José das Chagas Moura, DSc.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia de Produção, 2017.
Inclui Referências e Apêndice.

1. Engenharia de Produção. 2. Prognostic and health monitoring . 3. Empirical mode Decomposition. 4. Wavelets support vector machine. 5. Remaining useful life. 6. Reliability prediction. I. Lins, Isis Didier (Orientadora). II. Moura, Márcio José das Chagas (Coorientador). III. Título.

658.5 CDD (22.ed) UFPE/BCTG-2017/ 124

UNIVERSIDADE FEDERAL DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

PARECER DA COMISSÃO EXAMINADORA
DE DEFESA DE DISSERTAÇÃO DE
MESTRADO ACADÊMICO DE

CAIO BEZERRA SOUTO MAIOR

"REMAINING USEFUL LIFE PREDICTION VIA EMPIRICAL MODE DECOMPOSITION, WAVELETS AND SUPPORT VECTOR MACHINE"

ÁREA DE CONCENTRAÇÃO: PESQUISA OPERACIONAL

A comissão examinadora composta pelos professores abaixo, sob a presidência do(a) primeiro(a), considera o candidato **CAIO BEZERRA SOUTO MAIOR, APROVADO.**

Recife, 21 de fevereiro de 2017.

____

Profª. ISIS DIDIER LINS, Doutora (UFPE)

____

Profª. CRISTIANO ALEXANDRE VIRGÍNIO CAVALCANTE, Doutor (UFPE)

____

Prof. ABRAÃO DAVID COSTA DO NASCIMENTO, Doutor (UFPE)

To my parents, Leonardo and Lúcia.

To my sister, Thalita.

ACKNOWLEDGEMENTS

Gostaria de demonstrar meus profundos e sinceros agradecimentos a:

- Meus pais, Leonardo e Lúcia, por se esforçarem ao máximo para me garantir uma boa formação e me encorajar, tanto nos estudos como nos demais aspectos. Agradeço a eles o apoio incondicional em toda a minha vida. Minha irmã, Thalita, que sempre esteve ao meu lado e buscou me ajudar da maneira que ela podia.
- Minha vó Maria, por seu amparo em qualquer questão relacionada não só à educação, mas também à vida. Minha vó Rosa, por sua vibração e alegria com as conquistas realizadas nesses anos.
- Todos os meus amigos, os mais próximos e os mais distantes, que me apoiaram em bons e maus momentos durante a universidade, encorajando-me a seguir em frente.
- CEERMA por me fornecer oportunidades de estar em ambiente voltado ao pleno desenvolvimento acadêmico.
- Professora Isis Lins, minha professora orientadora, pela disponibilidade e pela ajuda imprescindível e indispensável nas várias dúvidas ao longo desse tempo.
- Professor Márcio Moura, meu professor co-orientador, pelo suporte e guia neste trabalho, bem como sugestões importantíssimas de melhorias.
- PPGEP e à CAPES por terem disponibilizado o conhecimento e o apoio financeiro durante meu mestrado acadêmico.

"One step ahead and you are not in the same place"

Chico Science

ABSTRACT

The useful life time of equipment is an important variable related to reliability and maintenance. The knowledge about the useful remaining life of operation system by means of a prognostic and health monitoring could lead to competitive advantage to the corporations. There are numbers of models trying to predict the reliability's variable behavior, such as the remaining useful life, from different types of signal (e.g. vibration signal), however several could not be realistic due to the imposed simplifications. An alternative to those models are the learning methods, used when exist many observations about the variable. A well-known method is Support Vector Machine (SVM), with the advantage that is not necessary previous knowledge about neither the function's behavior nor the relation between input and output. In order to achieve the best SVM's parameters, a Particle Swarm Optimization (PSO) algorithm is coupled to enhance the solution. Empirical Mode Decomposition (EMD) and Wavelets rise as two preprocessing methods seeking to improve the input data analysis. In this paper, EMD and wavelets are used coupled with PSO+SVM to predict the rolling bearing Remaining Useful Life (RUL) from a vibration signal and compare with the prediction without any preprocessing technique. As conclusion, EMD models presented accurate predictions and outperformed the other models tested.

Key-words: Prognostic and Health Monitoring. Empirical Mode Decomposition. Wavelets. Support Vector Machine. Remaining Useful Life. Reliability Prediction.

RESUMO

O tempo de vida útil de um equipamento é uma importante variável relacionada à confiabilidade e à manutenção, e o conhecimento sobre o tempo útil remanescente de um sistema em operação, por meio de um monitoramento do prognóstico de saúde, pode gerar vantagens competitivas para as corporações. Existem diversos modelos utilizados na tentativa de prever o comportamento de variáveis de confiabilidade, tal como a vida útil remanescente, a partir de diferentes tipos de sinais (e.g. sinal de vibração), porém alguns podem não ser realistas, devido às simplificações impostas. Uma alternativa a esses modelos são os métodos de aprendizado, utilizados quando se dispõe de diversas observações da variável. Um conhecido método de aprendizado supervisionado é o *Support Vector Machine* (SVM), que gera um mapeamento de funções de entrada-saída a partir de um conjunto de treinamento. Para encontrar os melhores parâmetros do SVM, o algoritmo de *Particle Swarm Optimization* (PSO) é acoplado para melhorar a solução. *Empirical Mode Decomposition* (EMD) e *Wavelets* são usados como métodos pré-processamento que buscam melhorar a qualidade dos dados de entrada para PSO+SVM. Neste trabalho, EMD e Wavelets foram usadas juntamente com PSO+SVM para estimar o tempo de vida útil remanescente de rolamentos a partir de sinais de vibração. Os resultados obtidos com e sem as técnicas de pré-processamento foram comparados. Ao final, é mostrado que modelos baseados em EMD apresentaram boa acurácia e superaram o desempenho dos outros modelos testados.

Palavras-chave: *Prognostic and Health Monitoring. Empirical Mode Decomposition. Wavelets. Support Vector Machine.* Tempo de vida útil residual. Previsão em Confiabilidade.

FIGURE LIST

LIST OF CONTENT

1 INTRODUCTION

1.1 Opening Remarks

The manufacturing and industrial sectors are facing an increasing demand to produce larger amounts of goods with better quality, which normally lead to keep the operating process at maximum requirement. Therefore, unscheduled downtimes bring severe problems to the production system and represent unplanned costs. Nevertheless, the impacts due to failure could be pictured in just inconvenience and setbacks but also can damage the system, injure people and, in extremes cases, cause death.

Reliability can be understood as the probability of a system to properly perform the tasks for it was designed, under certain conditions, during a predefined time length (RAUSAND & HOYLAND, 2004) and represents an essential role in the system performance. Reliability could be associated with maintenance, defining an optimal point to execute it. Monitoring and controlling are crucial, enhancing the productive availability and efficiency, avoiding failures, and reducing costs.

In this context, condition-based maintenance (CBM), or predictive maintenance, is a decision-making strategy using condition monitoring information to optimize the availability of operating plants (KAN, TAN & MATHEW, 2015). CBM enables the early detection of faults or failures in order to reduce downtime and operating costs, facilitate proactive responses, and improve the productivity as well as reliability, availability, maintainability and safety (RAMS) of equipment.

According to Si *et al.* (2011), the Remaining Useful Life (RUL) is the useful life left on an asset at a particular time of operation and is typically random and unknown. In fact, RUL is related with several factors, such as the currently degradation status, the operation environment and the system function and it must be estimated from available sources of information such as condition and health monitoring.

Hence, the development of a prognostic plan to properly monitor and manage the system becomes necessary to accurately estimate RUL. Pecht & Jaai (2010) comment about the prognostics and systems health management (PHM) as the most promising discipline of technologies and methods with the potential of solving RAMS problems. PHM not only analyzes the current system state but also aims to infer the degradation behavior in the future.

The system to be analyzed could be a specific component, equipment or also the entire system.

CBM/PHM technologies are developed and applied to a large variety of machines, systems, and processes in the transportation, industrial, and manufacturing sectors. Rotating equipment has received special attention due to its critical operating regimes, frequent failure modes and availability of measurements (vibration, temperature, etc.) intended to allow detection and isolation of incipient failures (VACHTSEVANOS *et al.*, 2007). Bearings are one of the most important components in rotating machine and their failure is one of the most frequent reasons for machine breakdown. The purpose of a ball bearing is to reduce rotational friction and support radial and axial loads and they are a very critical component in various machines.

Many researchers have recently considered on-line continuous monitoring of vulnerable motor components (e.g., bearings) based on different measurements. According to Vachtsevanos *et al*. (2007), the sensor measurements that affect fault diagnosis may be viewed from two different perspectives: static or process-related measurements such as temperature, speed, position, pressure, and flow rate and those that are characterized by their high bandwidth (i.e., they contain high frequency components, such as ultrasonic and vibration measurements, acoustic recordings, and alternating current or voltage).

Typical defects of roller bearings are cracks, spalling, flaking and indentation, mainly due to possible fatigue, wear, overloading and misalignment. The vibration signals generated by faults are often used for damage detection, since they often carry some significant dynamic information. The analysis of bearing's behavior from the vibration signal represents an important role in RUL: in general, when RUL decreases there is an increase in the vibration signal and it oscillates in higher amplitudes. Hence a longer or shorter RUL can be inferred from the vibration signal related to the bearings.

Based on vibration signal, many learning models (and hybrid forms) have been successfully applied to fault detection and prediction, such as in Mitoma, Wang, & Chen (2008), Lei, He, & Zi (2008), Bin *et al.* (2012) and Mahamad, Saon, & Hiyama (2010). Learning models require a training data set for an input variable to create an implicit knowledge about the behavior of the output variable. Support Vector Machines (SVM) (Cortes & Vapnik, 1995) and Artificial Neural Networks (ANN) (Haykin, 1998) are two well-known learning models.

In particular, SVM has been successfully applied to different fields, e.g. financial, environmental, reliability, power systems (KIM, 2003; NAJAFI *et al*., 2016; DROGUETT *et al.,* 2014; LINS *et al.,* 2013) and is particularly useful when the process or function that maps inputs into output is unknown. SVM is a data-driven kernel-based method that learns from an available data set formed by observed pairs of inputs and output. Its training step entails the resolution of a convex and quadratic optimization problem for which the Karush-Kuhn-Tucker (KKT) first order conditions are necessary and sufficient for global optimality.

So, compared with ANN, SVM do not have the drawback of be stuck in local optima. The basis of SVM relies on the structural risk minimization (SRM) principle that aims at minimizing the upper bound of the generalization error. Thus, the objective function has one part related with the model ability in predicting unseen data and the other one concerns training errors. The main idea of the SRM principle is to find a model with adequate capacity to describe the given training data set, creating a trade-off between model's capacity and training accuracy.

The learning model accuracy strongly depends on the quality of the input data. Some previous studies directly used the original series as the input variables in the construction of a forecasting model, which may lead to missing some features or to the consideration of irrelevant information (e.g. noise), generating poor predictions. Hence, some techniques can be used as preprocessing tools in order to improve data input quality and, consequently, to obtain superior predictions from the learning method.

Among those preprocessing techniques, the Empirical Mode Decomposition (EMD), proposed by Huang *et al.* (1998), converts the data in a more suitable form by decomposing the original series into a sum of simplest ones. According to Huang *et al.* (2014), EMD is adaptive, empirical, direct and intuitive. Each series is decomposed into a set of components called intrinsic mode functions (IMFs).

Wavelets are other well known preprocessing techiniques based in time-frequency analysis, originally proposed by Morlet *et al.* (1982) who introduce filter banks called wavelets function and scaling function. The idea behind Wavalets Transform is the same for the Short-Time Fourier Transform (ALLEN, 1977), but the former present the best frequency/time resolution trade-off, given that windows of various length are applied.

In machine learning, a considerable challenge is to provide the best parameters to be used in training step, since in principle, those are defined *a priori*. Normally, trial and error is not a consistent method to define those parameters due to low efficiency and long delay. Still, the best set of parameters often depends of the specific problem studied and its data set. Therefore, optimizations metaheuristics, such as Ant Colony System (ACS), Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), lead to satisfactory parameters' values avoiding the problem of set it in an erroneous way. In particular, PSO coupled with SVM have been successfully applied in reliability problems and fault machinery prediction (LINS *et al.*, 2013; DROGUETT *et al.*, 2014; GARCIA NIETO *et al.*, 2015).

Thus, a hybrid model involving a data preprocessing method, an optimization algorithm for parameter tuning and a machine learning structure is expected to present high predictive ability. The proposed methodology combines a preprocessing technique, EMD or Wavelets, with PSO+SVM to create models to estimate the RUL of rolling bearings from vibration signals. The performance of both models is compared with the model without any preprocessing technique.

1.2 Justification

The main objective of the maintenance is to guarantee the maximum availability and continuity to the operational function and its different maintenance schedules should be mainly based on system reliability. The exponential growth in system complexity has been followed by an increased demand for higher reliability. Estimation of RUL is helpful to manage life cycles of machines and to reduce maintenance costs.

Different metrics can be obtained from an operational system in order to track its degradation process and to build an accurate relationship between the current health condition state and RUL. Metrics such as vibration, acoustic emission, temperature, corrosion, among others, can represent the evolution of degradation, and their analyses are as necessary as arduous.

SVM is a promising algorithm for RUL estimation because it can deal with small training sets and multi-dimensional data (LIU *et al.,* 2016). Many SVM-based methods have been proposed to predict RUL of some key components and hybrid methodologies usually improve RUL estimation accuracy and overcome limitations of the individual methods (MAIOR *et al.,* 2016).

A number of hybrid approaches have been successfully applied, many of those as mixing of a preprocessing technique and an optimization algorithm, such as the one proposed in this work. However, recent researches utilize the combination of techniques but do not analyze the real gain of the hybrid methodology when compared to the learning method isolated. In this context, the present work compares EMD+PSO+SVM models and Wavelets+PSO+SVM models with a model without preprocessing techniques and investigates whether the hybrid methodology actually provides significant gain. The preprocessing technique usually increases the computational effort, once both EMD and Wavelets decompose the original series into many others, and the trade-off of this approach has to be taken into account. This result could support the solution of reliability problems as well as the preparation of suitable maintenance policies.

1.3 Objective

1.3.1 General Objective

This dissertation aims to compare two different preprocessed predictive models for RUL in industrial bearing with a model without any preprocessing technique. A hybrid method using EMD+PSO+SVM and Wavelets+SVM+PSO are proposed and compared with a PSO+SVM model, once all should track the correct behavior of the bearing's vibration in many stages of the degradation. The objective is to define the real gain with the use of EMD and Wavelets as data preprocessing techniques.

1.3.2 Specific Objectives

In order to achieve the general objective, some specific targets are defined:

- Understanding about the vibration signal and how it can be analyzed, in the time domain or in the frequency one seeking out to correctly use the methods applied.
- Get access to real data base for bearing vibration or experimental results in order to provide an authentic example.
- Apply EMD+PSO+SVM and Wavelets+PSO+SVM methodologies for reliability prediction problems and compare the results with the PSO+SVM methodology to confirm the efficiency of the methodologies.

1.4 Dissertation Layout

Besides this introduction chapter, this master thesis has four additional chapters briefly described as follows:

- **Chapter 2**: Presents the theoretical background required to the dissertation accomplishment as well as a short review about the related recent works. It clarifies points about PHM, Rolling Bearing and Vibration Signal, EMD, Wavelets, SVM and PSO.
- **Chapter 3**: Explains and the methodology and the model to be employed in this work.
- **Chapter 4**: Presents the real data and the procedures for a RUL prediction of rolling bearings from a vibration signal.
- **Chapter 5**: Contains a summary of the main aspects and results presented in this master thesis along with some limitations and propositions for future works.

2 THEORETICAL BACKGROUND AND LITERATURE REVIEW

This chapter presents definitions and explanations about the key topics of this work: Prognostic and Health Monitoring, Rolling Bearing and Vibration Signal, Empirical Mode Decomposition, Wavelets, Support Vector Machine and Particle Swarm Optimization. Also, a short review of the present works in the area is presented as a literature review.

2.1 Prognostic and Health Monitoring

Widodo & Yang (2011) explain that in the least years, studies about prognostic systems have been highlighted in different fields of science, like in maintenance and reliability research. The benefits of prognostics systems to engineering asset management are evident. Many advantages can be obtained from such as reducing the production downtime, spare-parts inventory, maintenance cost, and safety hazards.

A machine prognostics system is addressed to estimate the RUL of machine components based on various methods such as based on traditional distribution, data-driven methods, physic-based models and probability-based methods. Among them, data-driven and physic-based methods are more popular than the probability-based methods.

Adopting a health monitoring and prognostic technique system requires a continuous monitoring of the performance of the key parameters and especial attention should be given to detect any kind of disturbance in those parameters. Prognostics and health monitoring is a technology used to monitor degradation in engineering systems, understand when failure may occur and provide a cost-effective strategy for scheduled maintenance. Health monitoring and prognostics of engineering systems or products has become very important as failures may cause severe damage to the system, environment and users, and may result in significant costly repairs (SUTHARSSAN *et al.*, 2012).

Survival analysis is the name for a collection of statistical techniques used to describe and quantify time to event data. In survival analysis, we use the term ‘failure’ to define the occurrence of the event of interest and the term ‘survival time’ to specify the length of time taken for failure to occur. Situations where survival analysis have been used include prognostics of life time machine components, time from diagnosis to death in clinical trial, duration of industrial dispute, time from infection to disease onset, among others (WIDODO & YANG, 2011).

Over time, different definitions of prognostics was proposed in literature, as seen in Engel *et al.* (2000), Wu, Hu, & Zhang (2007) and Heng et al. (2009), for example. Commonly, prognostics could be explained from four characteristics: (1) it is, or should be, performed at the component or sub-component level; (2) involves predicting the time progression of a specific failure mode from its incipience to the time of component failure; (3) an appreciation of future component operation is required; and (4) prognostics is related to, but not the same as, diagnostics.

The most all-encompassing description of prognostics is presented by ISO13381-1; it defines prognostics as 'an estimation of time to failure and risk for one or more existing and future failure modes'. This implies that the field of prognostics is not only interested in predicting the effects of known failure modes on asset life, but also how these may initiate other failure modes (SIKORSKA *et al.*, 2011). Even though not a formal definition, it is common to characterize the prognostics output in two components: an estimated time until failure, normally called RUL and associated confidence limits (SIKORSKA *et al.*, 2011).

In order to understand the idea of prognostics and diagnostics, it is important to verify the many steps related with RUL and its confidence limits. The same component could have many failure modes, triggered reasons and a particular deterioration behavior, even when exposed to the same operational conditions. Anomalous events, such as changes in operating conditions, maintenance actions or other failures, may also occur and may accelerate one or more particular failure modes progressions (e.g. a bearing fault causes high vibration that induces and accelerates mechanical seal degradation).

Therefore, some information is necessary to determinate the RUL, such as: what is the current degradation state of the component; which failure mode has initiated the degradation and how severe is the degradation, that means, 'where' the component is on the particular degradation curve. These questions are associated with diagnostics. Other questions, which are in turn related with prognostics, are about how quickly is degradation expected to progress from its current state to functional failure; what novel events will change this expected degradation behavior and how other factors could affect RUL estimate.

Diagnostics is related with the current state of a component, in other words, what has happened in the past until now. Prognostics is related with what will happen with the component, that means, what is expected for the future. Continuous monitoring of the component is necessary to detect any new event, provide the diagnosis, estimate the new prognosis and then, update the system accordingly.

In order to analyze the process and achieve a realistic prognostic estimation, it is necessary to choose a suitable mathematical model to be used. There is no model type strictly better than other, since each problem should be dealt in a particular way. Complex models, which normally implies in higher reliability, could provide worse estimations compared with the simplest ones, due to the poor variables interpretation, due to the limited accuracy or even due to the lack of human knowledge to apply the model.

A number of different types of models could be seen in the literature and it is hard to analyze each one separately. Hence, it is necessary to group in classes some similar models, although there is no general consensus. Sikorska *et al*. (2011) propose a modified classification approach specifically designed for RUL prediction, with the models in four main groups and a varying number of subgroups, namelly:

(a) Knowledge-based models: These assess the similarity between an observed situation and a database of previously defined failures and deduce the life expectancy from previous events. Sub-categories include the following:

a. Expert systems.

b. Fuzzy systems.

(b) Life expectancy models: These determine the life expectancy of individual machine components with respect to the expected risk of deterioration under known operating conditions. Sub-categories are separated into statistical and stochastic models and include the following:

a. Stochastic models:

i. Aggregate reliability functions.

ii. Conditional probability methods, including the RUL probability density function and Bayesian networks.

b. Statistical models:

i. Trend extrapolation.

ii. Auto-regressive Moving Average (ARMA) models and variants.

iii. Proportional Hazards Modeling.

(c) Artificial Neural Networks: These compute an estimated output for the remaining useful life of a component/machine, directly or indirectly, from a mathematical representation of the component/system that has been derived from observation data rather than a physical understanding of the failure processes. They are further grouped into models used for:

a. Direct RUL forecasting;
b. Parametric estimation for other models.

(d) Physical models: These compute an estimated output for the remaining useful life of a component/machine from a mathematical representation of the physical behavior of the degradation processes. Types of physical models tend to be application (failure mode) specific and are therefore not classified further.

Also, Liao & Köttig (2014) categorized models into three different classes:

- Experience-based models
- Date-driven models
- Physics-based models.

They explain that due to system complexity, data availability, and application constraints, there is no universally accepted best model to estimate RUL and that data-driven models in hybrid approaches have been used for anomaly detection to trigger RUL prediction process.

The advantages and disadvantages of all methods should be careful understood and analyzed before the model choice. As previously mentioned, there is not one best model, but surely there is a suitable model for any situation. In this master thesis, a well-known method called SVM is used, which in the first classification present from Sikorska *et al.* (2011) would be in the third category along with ANN, given that it is primarily based in the learning from data and is classified in the data-driven category in the classification from Liao & Köttig (2014).

2.2 Rolling Bearings and Vibration Signal

Rolling bearings are an essential and critical component of rotating machines with its use and study widespread inside industrial applications. Figure 1 shows the principal components of a rolling bearing, highlighting the outer race, the inner race, the ball and the cage. Fault diagnosis of the rolling bearings has been the subject of extensive research. This process includes the acquisition of information, feature extraction and condition recognition (NIKOLAOU & ANTONIADIS, 2002).

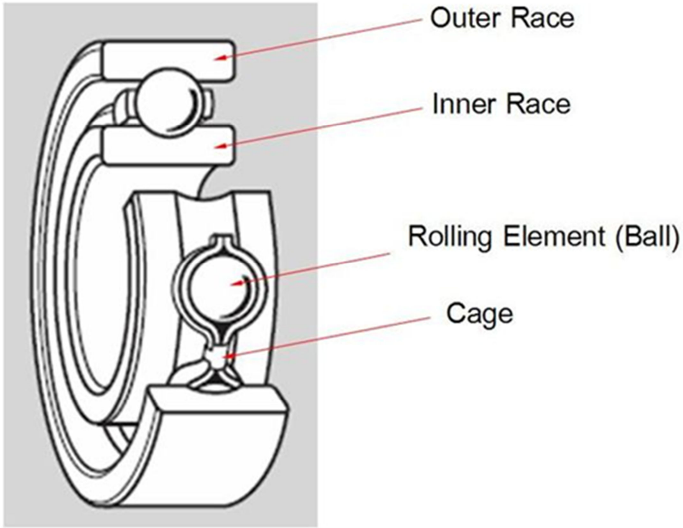

Figure 1: Elements of a Rolling Element Bearing

Different methods are used for the acquisition of information and they may be broadly classified depending on the type of measurements: vibration and acoustic, temperature and wear debris analysis (TANDON & CHOUDHURY, 1999). Among these, vibration measurements are commonly used in the condition monitoring and diagnosis of the rotating machinery mainly due to the easy-to-measure signals and plausible analysis.

The vibration measurement of the rolling bearing can be made using some accelerate sensors that are placed on the bearing house. When faults occur in the roller bearing, the vibration signal of the roller bearing would be different from the signal under the normal state. Localized faults in rolling element bearings produce a series of broadband impulse responses in the acceleration signal as the bearing components repeatedly strike the fault. The precise location of the fault determines the nature of the impulse response series, and Figure 2 shows the typical cases (RANDALL & ANTONI, 2011). Here, each element of the rolling bearing has its own rotation frequency (i.e. BPFO as BallPass Frequency Outer race, BPFI as BallPass Frequency Inner race, FTF as Fundamental Train Frequency – cage – and BSF as Ball Spin Frequency), which leads to composed complex signal.

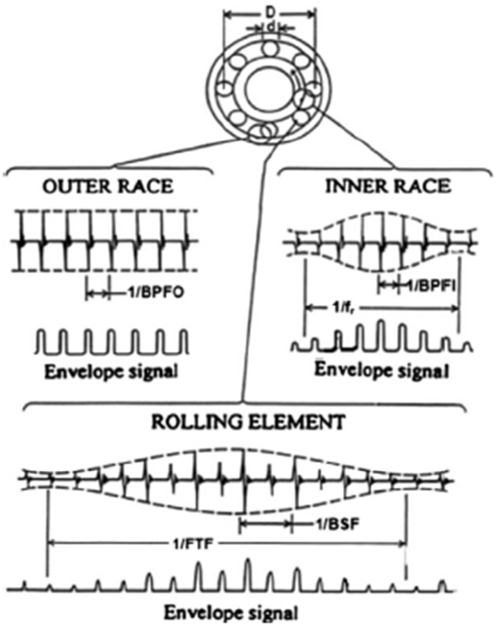

Figure 2: Signals and envelope signals from local faults in rolling element bearings

The quality of prognostics is directly impacted by the quality of the diagnosis values. There are many standard vibration-based metrics that are traditionally used for machinery diagnostics, including entropy, root mean square, signal amplitude, variance, kurtosis, as well as higher order statistics. While many of these metrics generally trend up in value as a spall grows, prognostics requires an estimate of absolute fault size. This more stringent requirement is difficult for most existing algorithms to meet, especially when operating conditions are variable. Some of the most applied metrics used in vibration analysis can be explained as follows:

- **Entropy** measures the uncertainty/complexity of the signal. Specifically, Shannon entropy is commonly referred to as a measure of information loss. Shannon entropy of a normalized signal measures the amount of randomness and sparseness of data. Thus, a signal with minimal amount of Shannon entropy can be treated as the most periodic and easier to predict. When roller elements pass over a defect on bearing components, different series impulses will be generated, causing significant change in the value of entropy (HEMMATI, ORFALI & GADALA, 2016). Shannon (1948) defines entropy as follows:

$$H(x) = -\sum_{i=1}^{N} p(x_i)\log\big(p(x_i)\big), \qquad \sum_{i=1}^{N} p(x_i) = 1 \tag{2.1}$$

where $p(x_i)$ is the probability of observing the ith possible value of a random variable $X = (x_1, x_2, \dots, x_n)$. For further details and applications with Shannon Entropy, see Qiu *et al.* (2006).

- **Root mean square (RMS)**, also known as the quadratic mean, provides an indication of the overall signal energy, i.e. the power content in the vibration signal. When a bearing is faulted, the vibration signature is changed by the impacts produced by rolling elements passing over the fault, resulting in increased energy. The RMS value for a time series x_i with length N is calculated as:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{2.2}$$

- **Variance** measures the statistical dispersion of a signal. The impacts in a spalled bearing should increase the variability in the signal and it is calculated based on the follow equation for a time series x_i with length N:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{2.3}$$

Where the population mean is

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2.4}$$

- **Kurtosis**, the fourth statistical moment of the data related with the peakedness or flatness of the distribution, is used as measure of the “spikiness” of the data. Intuitively, abrupt impacts between a ball and the spall should cause spikes in the signal, increasing kurtosis. It is calculated as follows:

$$k = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{N * (\sigma^2)^2} \tag{2.5}$$

- **Crest Factor** is the ratio of the peak vibration level to the RMS, and is often used to detect changes in signal patterns due to impulse vibration sources, normally not captured in the RMS analysis alone. For normal conditions, its value is between 2 and 6. It is calculated as follows:

$$Crest\ Factor = \frac{Peak\ Level}{RMS} \tag{2.6}$$

- **Peak signal amplitude** is related to the crest factor. An increase in signal energy should cause an increase in peak signal amplitude. As with RMS, the variability of the peak signal amplitude increases with spall length.
- Higher order statistics, such as the sixth statistical moment **M6**, are often useful for the purposes of diagnosis. Normally, **M6** and kurtosis present similar results.

A useful study comparing some of the most common vibration-based diagnostics metrics for bearings, gears and other machinery can be found in Lybeck, Marble, & Morton (2007). The paper compares the correlation between the metric value and the known spall length and concludes that Root Mean Square, Variance and Peak Signal Amplitude are the more reliable, among the previously mentioned.

However, numerous of previous researches in literature have shown that each feature is only effective for a certain defect at a certain stage. Thus, which feature to be selected for machine fault prognostics, in particular, with machine degradation, is still a challenge that needs more investigation.

2.3 Empirical Mode Decomposition

Starting with the work of Huang *et al.* (1998), a remarkable method to analyze non-linear and non-stationary data series was developed and have been used in many types of applications. Rojas *et al.* (2013) comment that EMD is an empirical, intuitive, direct and adaptive method and it does not require any previous assumption. The main idea is that any data series could be decomposed in a small number of simplest oscillation series, called Intrinsic Mode Functions (IMFs). The goal is to obtain IMFs regarding data characteristics in the time scale (HUANG & WU, 2008).

Generally, any complex signal can be possibly separated into a small number of IMFs, represented by $c_i(t)$, and a trend $r(t)$. For a number N of IMFs generated, the original series $x(t)$ is expressed as follows:

$$x(t) = \sum_{i}^{N} c_i\,(t) + r(t) \tag{2.7}$$

Huang *et al.* (1998) define an IMF as a function that satisfies two conditions: (1) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; and (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. Ideally, the requirement should be 'the local mean of the data being zero'. For non-stationary data, the 'local mean' involves a 'local time scale' to compute the mean, which is impossible to define. As a surrogate, it is used the local mean of the envelopes defined by the local maxima and the local minima to force the local symmetry instead.

Yang *et al.* (2007) mention that each IMF represents a frequency-amplitude modulated narrow band, normally associated with a specific physical process. The goal of EMD is to, empirically, identify the IMFs from the data series features and decompose it according to its unique characteristics. Figure 2 represents the generation of IMFs in the EMD process.

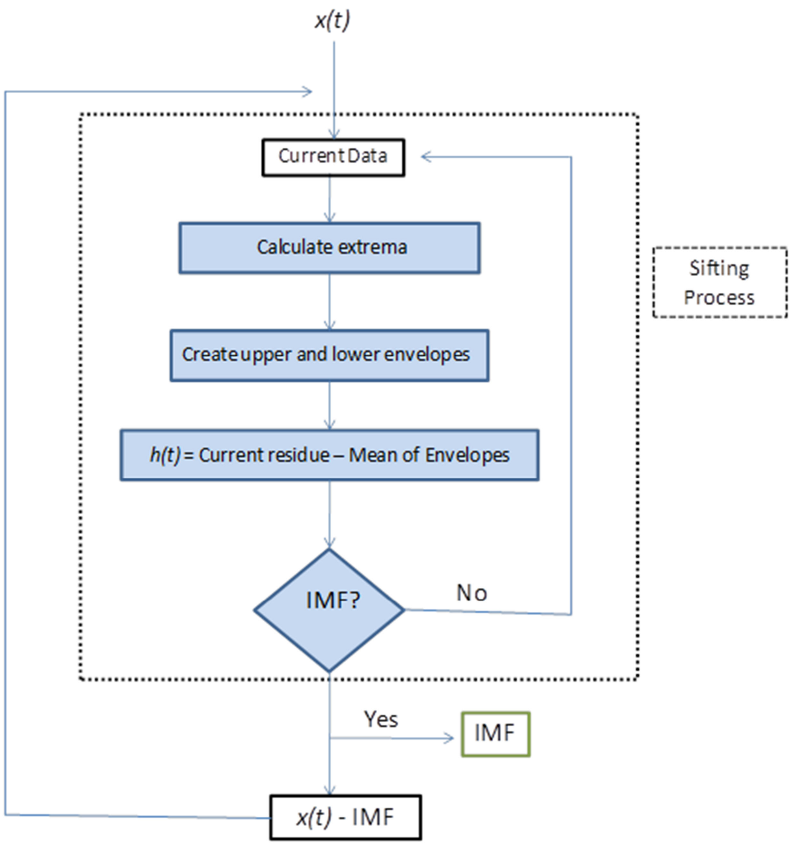

Figure 3: Descriptive flowchart of the EMD process

A process called sifting is performed in order to uncover each IMF. The decomposition is based in the following assumptions: (1) the signal has at least two extrema – one maximum and one minimum; (2) the characteristics time scale is defined by the time lapse between the extrema; and (3) if the data were totally devoid of extrema but contained only inflection points, then it can be differentiated once or more times to reveal the extrema (HUANG *et al*., 1998).

In a signal, if there is a local minimum greater than zero between two successive local maxima, or if there exists a local maximum less than zero between two successive local minima, the segment between these two successive local maxima (or local minima) is called a riding wave (YANG *et al.,* 2007) and is shown in Figure 4. The sifting goal is to remove the riding waves, so as to make the wave profile more symmetric. The sifting process can be described in the following steps:

1. Identify all local extrema (maximum and minimum) of the series $x(t)$;
2. Connect all the local extrema with a cubic spline line to create the upper and lower envelops, e_u, e_l, respectively;
3. Calculate the envelope mean $m(t) = (e_u + e_l)/2$;
4. Obtain $h(t) = x(t) - m(t)$, candidate to be an IMF;
5. Verify if $h(t)$ satisfies both conditions that define an IMF. If it satisfies, an IMF was generated with the residue $m(t) = x(t) - h(t)$ replacing the initial series $x(t)$. Otherwise, $h(t)$ would be the new series $x(t)$ and return to step 1.
6. Once the step 5 is achieved and an IMF is generated, save $c_i(t) = h(t)$ as the i-th IMF. Then, a series residue $r(t) = x(t) - c_i(t)$ becomes the new series $x(t)$ and a new loop starts in step 1.

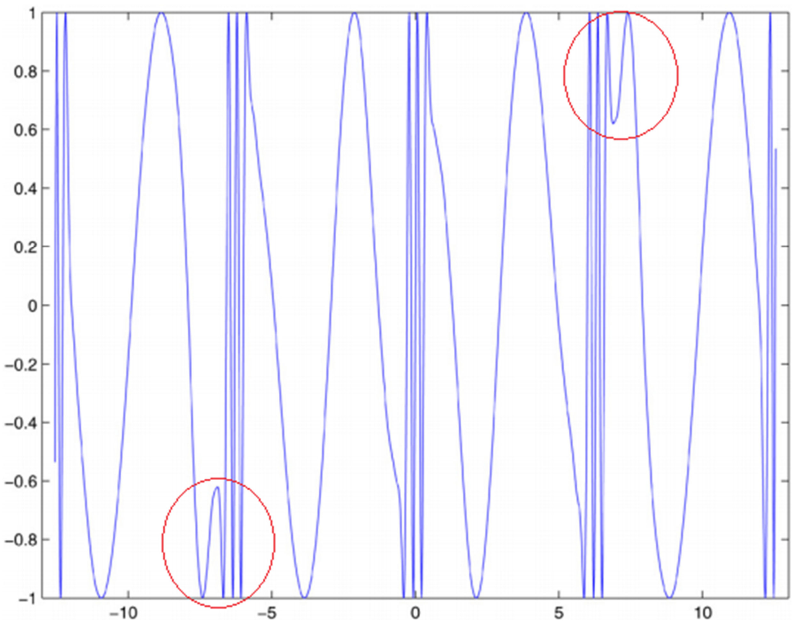

Figure 4: Example of riding waves.
Adapted from Yang *et al.* (2007)

Eftekhar, Toumazou, & Drakakis (2013) affirm that in its core form, the literature has done little to modify this process. Modifications come in how some of the steps are carried out, including the maxima/minima detection, envelope formation (interpolation), the way an IMF is identified and how one stops the sifting process.

At the end of the sifting process, a number of IMFs are generated plus a final residue $r(t)$. The sifting has two effects: (1) riding waves are eliminated; and (b) uneven amplitudes are smoothed. While the first condition is absolutely necessary for the instantaneous frequency to be meaningful once riding waves normally represents the mixing between more than one oscillatory mode, the second condition is also necessary in case the neighboring wave amplitudes have a large disparity. Unfortunately, the second effect, when carried to the extreme, could obliterate the physically meaningful amplitude fluctuations (HUANG *et al.*, 1998).

Therefore, the sifting process should be applied with care, for carrying the process to an extreme could make the resulting IMF a pure frequency modulated signal of constant amplitude. To guarantee that the IMFs retain enough physical sense of both amplitude and frequency modulations, a stop criterion for the sifting process have to be determined. This can be accomplished by limiting the size of the standard deviation, SD, computed from two consecutive sifting results as:

$$SD_k = \sum_{t=0}^{T} \left[\frac{|(h_{1(k-1)}(t) - h_{1k}(t))|^2}{h_{1(k-1)}(t)^2} \right] \tag{2.8}$$

With k representing the number of iterations until the process achieves an IMF and t stand for the data. The sifting process stops when SD_k is smaller than a preset value. After a great number of tests, Huang *et al.* (1998) explain that a typical values for SD should be set between 0.2 and 0.3 to guarantee physical meaning.

Overall, $c_1(t)$ should contain the finest scale or the shortest period component of the signal. We can separate $c_1(t)$ from the rest of the data by

$$x(t) - c_1(t) = r_1(t) \tag{2.9}$$

Since the residue r_1 still contains information of longer period components (small frequencies), it is treated as the new data and it is subjected to the same sifting process as described above. This procedure can be repeated on all the subsequent r_j's, and the result is

$$r_1 - c_2 = r_2, \cdots, r_{N-1} - c_N = r_N \tag{2.10}$$

The sifting process also can be stopped by any of the following predetermined criteria: either when the component, c_N, or the residue, r_N, becomes so small that it is less than the predetermined value of substantial consequence, or when the residue, r_N, becomes a function from which no more IMF can be extracted. Even for data with zero mean, the final residue can still be different from zero; for data with a trend, then the final residue should be that trend. Finally, the series could be presented as:

$$x(t) = \sum_{i=1}^{N} c_i + r_N \tag{2.11}$$

Thus, original series $x(t)$ is decomposed in N-intrinsic modes and a residue r_N, which can be either the mean trend or a constant.

2.4 Wavelets

A widespread technique applied in the field of signal analysis, the wavelet transform was first proposed by the geophysicist Jean Morlet in Morlet *et al.* (1982) and was later formalized by Grossmann & Morlet (1984) and Goupillaud & Morlet (1984) , particularly for the processing of seismic measurements. The word wavelet means ‘small wave’ and according to Daubechies (1990), Morlet’s original name for the wavelets was ‘wavelets of constant shape’. It was supposedly chosen to contrast them with the analyzing functions in the short-time Fourier transform, which do not have a constant shape.

Wavelet transforms offer a description of a process through decomposition of a signal onto a set of basis functions, called wavelets and these wavelets are obtained from a single prototype wavelet, called a mother wavelet. The representation of the process occurs by an infinite series expansion of dilated/contracted and translated versions of a mother wavelet, each multiplied by an appropriate coefficient.

Two important properties of wavelets are the admissibility and the regularity conditions. Kumar & Foufoula-Georgiou (1997) remark that a wavelet transform is chosen so that it has:

- Sufficiently fast decay, to obtain localization in space (called regularity condition);
- Zero mean (called admissibility condition).

The first condition ensures that the wavelet is not a sustaining wave, while the second condition ensures that it has a wiggle (i.e. is wave-like). In practical applications, it is possible to use different well known wavelet transforms for distinct purposes and its choice depends on the specific signal characteristics and the analysis to be done.

Wavelet transform can be considered as a mathematical tool that converts a signal in time domain into a different form, i.e. a series of wavelet coefficients, in time-scale domain (YAN, GAO & CHEN, 2014). The Continuous Wavelet Transform (CWT) of a function $x(t)$ is defined as the integral transform:

$$W_x(\lambda, t, \psi) = \frac{1}{\sqrt{\lambda}} \int_{-\infty}^{\infty} x(u) \bar{\psi}\left(\frac{u-t}{\lambda}\right) du \qquad (2.12)$$

In the expression, λ is a scale parameter, t is a location parameter and $\bar{\psi}_{\lambda,t}(u)$ represents the complex conjugate of $\psi_{\lambda,t}(u)$, a family of wavelet functions. There are different examples of wavelets defined for the continuous wavelet transforms, such as the Mexican Hat wavelet, Meyer wavelet and Shannon Wavelet. Morlet’s wavelet is one of the most popular wavelets and is defined as:

$$\psi(t) = \frac{1}{\sqrt{2\pi}} e^{[-t^2/_2 + j2\pi b_0 t]} \tag{2.13}$$

Since Morlet's wavelet is a complex waveform, it may be decomposed into its real and imaginary parts, respectively:

$$\psi_r(t) = \frac{1}{\sqrt{2\pi}} e^{[-t^2/_2 \cos(2\pi b_0 t)]} \tag{2.14}$$

$$\psi_i(t) = \frac{1}{\sqrt{2\pi}} e^{[-t^2/_2 \sin(2\pi b_0 t)]} \tag{2.15}$$

where b_0 is a constant and $b_0 > 0$ to satisfy the admissibility condition. The Morlet's wavelet is depicted in Figure 5.

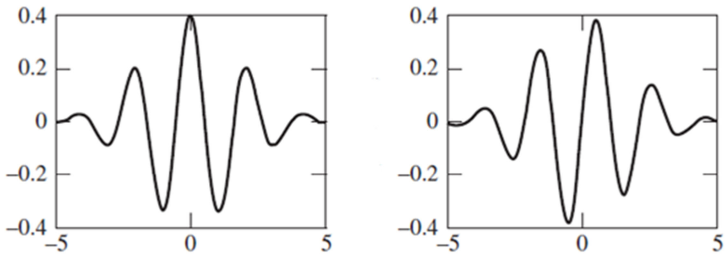

Figura 5: Morlet's wavelet: real (left) and imaginary (right) parts

Source: Sanei & Chambers (2007)

Guohua *et al.* (2006) comment that wavelet analysis decomposes a signal into two parts, called approximations and details. Approximations consist of the high scale low frequency components, while details consist of the low scale high frequency components. While approximations offer general information of a signal, details offer detailed information of a hidden pattern in the signal. The authors also remind that wavelet functions are obtained from recursive relationships, where the first wavelet is the mother wavelet function.

Peng & Chu (2004) remind that Meyer and Mallat developed the idea of multi-resolution analysis (MRA) that made it very easy to construct other orthogonal wavelet bases. A more important event was that the MRA led to the famous fast wavelet transform – a simple and recursive filtering algorithm to compute the wavelet decomposition of the signal to its finest scale approximation.

Ingrid Daubechies was another important researcher in wavelets applications. She popularized wavelets with her work in Daubechies (1993), constructing orthogonal wavelet bases compactly supported, allowing more liberty in the choice of the basis wavelet functions at a little expense of some redundancy. Daubechies, along with Mallat, is therefore credited with the development of the wavelet from continuous to discrete signal analysis. In the discrete wavelet formalism (DWT), the scale λ and the time t are discretized as following:

$$\lambda = \lambda_0^m, \qquad t = n\lambda_0^m t_0 \tag{2.16}$$

where m and n are integers. So the continuous wavelet function $\psi_{\lambda,t}(u)$ in Eq. (2.12) become the discrete wavelets given by

$$\psi_{\lambda,t}(u) = \lambda_0^{-\frac{m}{2}} \psi(\lambda_0^{-m} u - nt_0) \tag{2.17}$$

The discretization of the scale parameter and time parameter leads to the discrete wavelet transform, defined as:

$$W_x(m, n, \psi) = \frac{1}{\sqrt{\lambda_0^m}} \int_{-\infty}^{\infty} x(u)\bar{\psi}(\lambda_0^{-m} u - nt_0) du \tag{2.18}$$

Also, there are many discrete functions that can be used as mother wavelets, such as Haar Wevelet, Legendre Wavelet and Symlet. A popular wavelet used in DWT is the Daubechies Wavelet. In general the Daubechies wavelets are chosen to have the highest number A of vanishing moments, for given support width $2A - 1$ (Daubechies, 1993). The Daubechies wavelets are not defined in terms of the resulting scaling and wavelet functions; in fact, they cannot be written in closed form. However, due to its successful and widespread applications, Daubechies wavelets were used in this work.

2.5 Support Vector Machine

Wang (2005) defines SVM as a supervised learning method aiming to map input-output from a dataset called training data $D = \{(x_1, y_1), \ldots (x_m, y_m)\}$. The objective is to find the function $f(x)$ with the smallest penalization with respect to the deviation from the real data and, at the same time, as flat as possible. Specifically, the learning problem is defined according its output y as:

a. Classification problems, with y assuming discrete values that represent categories. If only two categories are considered (e.g. y = -1 or y = +1), the problem at hand is of binary classification. Otherwise, if three or more categories are taken into account, it is the case of a multi-classification problem.

b. Regression problems: y is real-valued and its relation with the input vector x is given by a function.

SVM was firstly proposed by Vladimir Vapnik based on the principle of the Structural Risk Minimization (SRM) and has its concepts based on the Statistical Learning Theory (VAPNIK, 2000). Regardless of whether classification or regression, the problem could be seen as follows: there exist a mapping function $y = f(x)$, unknown, of real values and, possibly, non-linear between an input vector x and an output scalar y and the only available information is the data D, used in the learning process, with $f(x)$ as solution. This means to solve a convex and quadratic optimization problem with the Karush-Kuhn-Tucker (KKT) conditions as necessary and sufficient to guarantee a global optimum. When SVM is applied to regression problems, it is called Support Vector Regression (SVR).

The first order KKT conditions are necessary to ensure the solution from a non linear programming problem to be optimum, given that the regularity conditions are satisfied. In a convex problem, these conditions are sufficient to a global optimum (ZHAO & DIMIROVSKI, 2004).

The goal is to find the hyperplan that best represents the input-output mapping from the data D. It is important to emphasize that the goal is not to look for the perfect alignment between the function $f(x)$ and D, but the best representation for the mapping. Furthermore, it is not desirable the strict alignment, once a trade-off have to be made between the data fitness and the generalization ability to predict new data. The equation of the regression hyperplane is:

$$f\,(\boldsymbol{x}) = \boldsymbol{w^T x} + b \tag{2.19}$$

with $\boldsymbol{x}$ expressing the input data and $\boldsymbol{w^T}$ and b the coefficients to be determined. They are estimated from the follow regularized risk function:

$$R(C) = C\frac{1}{m}\sum_{i=0}^{m} \psi_\varepsilon(y_i, f_i) + \frac{1}{2}\boldsymbol{w^T w} \tag{2.20}$$

and

$$\psi_\varepsilon(y_i, f_i) = \begin{cases} |y_i - f_i| - \varepsilon & if \quad |y_i - f_i| \geq \varepsilon \\ 0 & \text{othewise} \end{cases} \tag{2.21}$$

where y_i is the variable real value, that is , the original data and f_i is the estimated value to the same variable on the same time. Equation (2.20) is known as the Vapnik's ε -insensitive loss function that implies the non-penalization when the points are inside de tube with radius ε. For calculus convenience, ξ_i is defined when the data is above the tube and ξ_i^* when the data is below the tube, and represent the slack variables.

Hence, ε measures the performance in the training process and is related to the first term of Equation (2.20). The second term of the same equation is used as a smoothness function, once SVM also aim to get $f(x)$ as flat as possible and is also necessary to minimize the term related with the machine's capacity represented by $\boldsymbol{w^T w}$, which is the squared norm of $\boldsymbol{w}$. Therefore, C is a tradeoff between the empirical risk and the model's smoothness, with its value defined *a priori*, in the same way as the parameter ε. The primal problem is defined:

$$\min_{\mathbf{w},b,\xi,\xi^*} \frac{1}{2} \boldsymbol{w}^T\boldsymbol{w} + C \cdot \sum_{i=o}^{m} (\xi_i + \xi_i^*) \tag{2.22}$$

where m is the number of training points and subject to:

$$y_i - \boldsymbol{w}^T\boldsymbol{x}_i - b \leq \varepsilon + \xi_i \tag{2.23}$$

$$\boldsymbol{w}^T\boldsymbol{x}_i + b - y_i \leq \varepsilon + \xi_i^* \tag{2.24}$$

$$\xi_i \geq 0, \qquad i = 1,2,\dots,m \tag{2.25}$$

$$\xi_i^* \geq 0, \qquad i = 1,2,\dots,m \tag{2.26}$$

From that, the corresponding primal Lagrangian function is as follows:

$$\mathcal{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\xi}^*, \boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}, \boldsymbol{\beta}^*) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=0}^{l}(\xi_i + \xi_i^*) - \sum_{i=0}^{l}(\beta_i\xi_i + \xi_i^*\beta_i^*) - \sum_{i=0}^{l}\alpha_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b - y_i + \varepsilon + \xi_i\right) - \sum_{i=0}^{l}\alpha_i^*(y_i, \boldsymbol{w}^T\boldsymbol{x}_i - b + \varepsilon + \xi_i^*) \tag{2.27}$$

Where $\boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$ are the l-dimensional vectors of Lagrange multipliers associated to constrains (2.23), (2.24), (2.25) and (2.26), respectively. It may be observed that α_i and α_i^* both cannot be strictly positive since there is no point satisfying (2.25) and (2.26) simultaneously.

Equation (2.27) needs to be minimized with respect to the primal variables $\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\xi}_i^*$, and maximized with respect to the dual variables $\boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}, \boldsymbol{\beta}^*$. KKT conditions state that the partial derivatives of $\mathcal{L}$ with respect to primal variables must vanish when evaluated at the optimal point. By introducing some equalities provided by the KKT complementarity conditions back into $\mathcal{L}$, the latter becomes function only of the dual variables α_i, α_i^*, and the dual Lagrangian is obtained:

$$\max L_D(\alpha, \alpha^*) = -\frac{1}{2}\sum_{i,j}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\boldsymbol{x}_i^T\boldsymbol{x}_j - \sum_i(\varepsilon - y_i)\alpha_i - \sum_i(\varepsilon + y_i)\alpha_j^* \tag{2.28}$$

subject to:

$$\sum_i(\alpha_i - \alpha_i^*) = 0 \tag{2.29}$$

$$0 \le \alpha_i \le C \tag{2.30}$$

$$0 \le \alpha_i^* \le C \tag{2.31}$$

The solution of the dual training problem provides the optimal estimated regression function:

$$f(\boldsymbol{x}) = \boldsymbol{w}_0^T\boldsymbol{x} + b = \sum_{i=1}^{l}(\alpha_i + \alpha_i^*)\,\boldsymbol{x}_i^T\boldsymbol{x} + b \tag{2.32}$$

To solve the linear regression in SVR, it is necessary to calculate the dot products, $\boldsymbol{x}_i^T\boldsymbol{x}_j$ and $\boldsymbol{x}_i^T\boldsymbol{x}$, as presented in the Equations (2.28) and (2.32). To deal with non-linearity, mapping functions $\Phi(x)$ are applied and the dot product, $\Phi^T(x)\Phi(x)$, is solved in higher dimension. However, in addition to the selection of a proper mapping function, the explicit calculation of the dot products with mapped input vectors can be computationally expensive. In this way, dot products are replaced by kernel functions $K(x_i, x_j)$ defined in the input space that implicitly maps x into a higher dimension. One can see an interesting tutorial about kernel methods in Jäkel, Schölkopf & Wichmann (2007). The regression problem could be also solved from the Kernel functions and the regression function is expressed as:

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(x_o, x_i) + b \tag{2.33}$$

In this master thesis, the kernel function adopted is Gaussian Radial Basis Function (RBF), expressed by $K(x_l, x_o) = \exp(-\gamma \|x_l - x_o\|^2)$, where γ is also a model parameter. For more discussion about Kernel Methods applied to SVM, see Shawe-Taylor & Sun (2014).

After the choice of the function parameters, it is necessary to evaluate the precision of the estimated function. Hence, errors are calculated comparing real values y_i and predicted values $\hat{y}_i$. One of the most common performance error measures is MSE (Mean Square Error), described to a series with *m* data points as follows:

$$\text{MSE} = \left(\frac{1}{m}\right) \sum_{i=0}^{m} (y_i - \hat{y}_i)^2 \tag{2.34}$$

The SVR performance strongly depends on the choice of parameters C, ε e γ, defined *a priori*. In order to determinate the best set of parameters, optimization algorithms are frequently applied.

2.6 Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a probabilistic optimization heuristic inspired by the social behavior of biological organisms (e.g., birds and fishes), specifically on the ability of animal groups to work as a whole in order to find some desirable position. This seeking behavior is artificially modeled by PSO, which has been mainly used in the quest for solutions of non-linear optimization problems in a real-valued search space (BRATTON & KENNEDY, 2007).

The basic element of PSO is a particle, which can fly throughout the search space toward an optimum by using its own information as well as that provided by other particles within its neighborhood. For a problem with n-variables, each possible solution can be considered as a particle with a position vector of dimension n and the population of particles is defined as swarm (SAMANTA & NATARAJ, 2009)

The performance of a particle is determined by its fitness that is assessed by calculating the objective function of the problem to be solved. Mathematically, a particle i is characterized by the following three vectors:

- Its current position in the *n*-dimensional search space $\mathbf{s}_i = (s_{i1}, s_{i2}, \cdots, s_{in})$.;

- The best individual position it has thus far occupied during motion $\mathbf{p}_i = (p_{i1}, p_{i2}, \cdots, p_{in})$;
- Its velocity $\mathbf{v}_i = (v_{i1}, v_{i2}, \cdots, v_{in})$.

Traditionally, the current positions $\mathbf{s}_i$ and velocities $\mathbf{v}_i$ are initialized by sampling from uniform distributions over the intervals of definition of the decisions variables and by setting a maximum velocity value $\mathrm{v}_{\max}$, respectively. Then, the particles move throughout the search space in successive iterations driven by the following set of update equations:

$$\mathbf{v}_{ij}(t+1) = v_{ij}(t) + c_1u_1\left[p_{ij}(t) - s_{ij}(t)\right] + c_2u_2[p_{gj}(t) - s_{gj}(t)] \quad (2.35)$$

$$\mathbf{s}_{ij}(t+1) = s_{ij}(t) + v_{ij}(t+1),\ \ j = 1,2,\dots,n \quad (2.36)$$

where c_1 and c_2 are constants, u_1 and u_2 are independent uniform random numbers generated at every update along each individual direction $j = 1,2,\dots,n$ and $\boldsymbol{p}_g(t)$ is the n-dimensional vector of the best position encountered by any neighbor of particle i. Note that velocities and positions at time $t+1$ are influenced by the distances of the particle's current position from its own best experience $\boldsymbol{p}_i(t)$ and the neighborhood's best experience $\boldsymbol{p}_g(t)$, in a cooperative fashion (KENNEDY & EBARHAR, 1995).

The PSO search stops when some criteria are reached. Three criteria were adopted to stop the algorithm: (i) maximum number of iterations; (ii) the global best particle is the same for 10% of the maximum number of iterations and (iii) the global best fitness value in consecutive iterations have difference smaller 10^{-9}, based on the objective function. In this work, the objective function used was the Mean Squared Error.

The number of works involving preprocessing techniques, coupled or not with optimizations algorithms and machine learning are growing apace over the years and some of them are commented in the next session.

2.7 Previous works

Zio (2009) provides an interesting paper covering many aspects of maintenance and reliability problems, the contributions to system safety and on its role within system risk analysis. Specifically on PHM, diagnostics and prognostics are important aspects in the context of the development of monitoring programs. According to Dong & He (2007) diagnostics procedures consist in detection, isolation and identification of a failure, while prognostics procedures are related with the prediction of the next failure moment (RUL) given the system's current condition, normally influenced by external facts, typically classified as non time invariable, non-linear and dependent one to another. As result, RUL prediction could demand sophisticated probabilistic models in order to realistically capture the complexities of the dynamic behavior of the system's reliability, which could lead to complex mathematics formulations and expensive numeric methods, even if they do not lead to a desired accuracy (MOURA & DROGUETT, 2009).

Fan & Tang (2013) also comment that many methods based in different models were applied in reliability problems and RUL prediction, under assumptions that could, or not, be true in reality, as techniques based on Markov models (e.g. Zhao *et al.*, 2014). This leads to an important alternative to the practical applications in non-linear reliability predictions: the data-driven learning methods.

Statistical data-driven approaches rely on the availability of data and the nature of the data. In this context, many studies compare some data driven models (e.g. Multilayer Perceptron, Nearest Neighbour, Decision Tree and SVM) in different applications (Tan *et al.*, 2005; Geng, Zhan, & Zhou 2005; Acharya *et al.*, 2016; Raghavendra *et al.*, 2016; Cuss, McConnell, & Guéguen, 2016). Also, Khelif, Chebel-Morello, & Zerhouni (2015) applied Unsupervised Kernel Regression to predict RUL for Li-ion batteries with interesting practical results.

Here, we highlight SVM as supervised learning method that generates a mapping function of a labeled training dataset in a classification problem or a regression problem. Moura *et al.*(2011), Lins *et al.* (2013) and Droguett *et al.* (2014) successfully applied SVM to reliability problems in time series data. Hu et *al.* (2014) adopt an SVM model to estimate the remaining useful life and the performance reliability based on the predefined threshold for failure in manufacturing industry.

Benkedjouh *et al.* (2013) reported the importance of PHM in rotation machines, as it increases the reliability and decreases machinery downtimes. They estimate the RUL in rolling bearings using SVM. Hu *et al.* (2010) analyze the monitoring of RUL for axial piston pump, applying an EMD+SVM methodology to predict RUL and confirming the better results provided from hybrid methods.

Peng & Chu (2004) and, more recently, Yan, Gao, & Che (2014) present an interesting review of wavelets applications focused on rotary machine fault diagnosis. Konar &

Chattopadhyay (2011) applied CWT coupled with SVM for fault detection of induction motors. Liu *et al.* (2013) used a Wavelet+PSO+SVM methodology to analyze vibration signals from rolling bearing elements, working with diagnosis, not with prognostics.

EMD+SVM algorithm applied in vibration signal to estimate failure in rolling bearing could be seen in Yang, Yu, & Cheng (2007) and prognostic is performed by Liu *et al.* (2014) with the EMD+SVM approach to predict the RUL of subsystems in heavy rolling-mills.

Rai & Upadhyay (2016) provide an interesting review on signal processing techniques utilized in the fault analysis of rolling element bearings. The current work in failure analysis from vibration signal used different types of hybrid methods as seen in Wang, Jia & Li (2014), Su *et al.* (2015) and Saxena, Bannett, & Sharma (2016).

3 METHODOLOGY AND DATA SET CONTENT

3.1 Applied Methodology

The methodology applied in this master thesis is presented in Figure 6. Vibration data from rolling bearings are used as input for the methodology. The data sets contain a large amount of observed values and, due to the computational cost, the learning model (PSO+SVM) cannot handle such an extensive data, being necessary to reduce the actual amount of data used. The reduction was done in two steps: the first step concerned a feature extraction to summarize 2560 points, which represent a discrete recording of the vibration signal, into just one point; the second reduction was sampling these data in specific frequencies depending on which degradation state the bearing was. Further details about the data set are exposed in the next section.

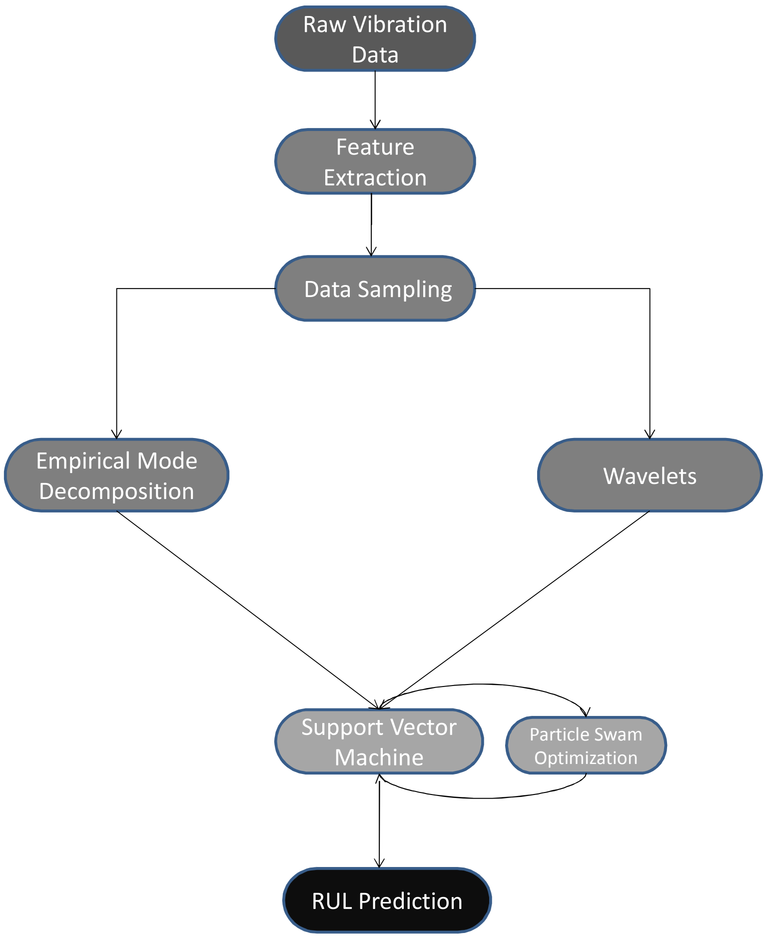

Figure 6: Methodology applied for RUL prediction

After the sampling, EMD or Wavelets was performed as shown in Figure 7 and 8, respectively. In each case, two distinct inputs for regression models were created differing in which regressors it used: for EMD, one model containing all IMFs plus the final residue and the other model containing just the final residue; for Wavelets, one model consisted of wavelet functions of each level plus the last scaling function and other model consisted of just the last scaling function. Once SVM highly depends of the data input, the idea behind using the final residue for EMD and the last scaling function for Wavelets was to provide an input enough smooth but still caring important characteristics of the signal.

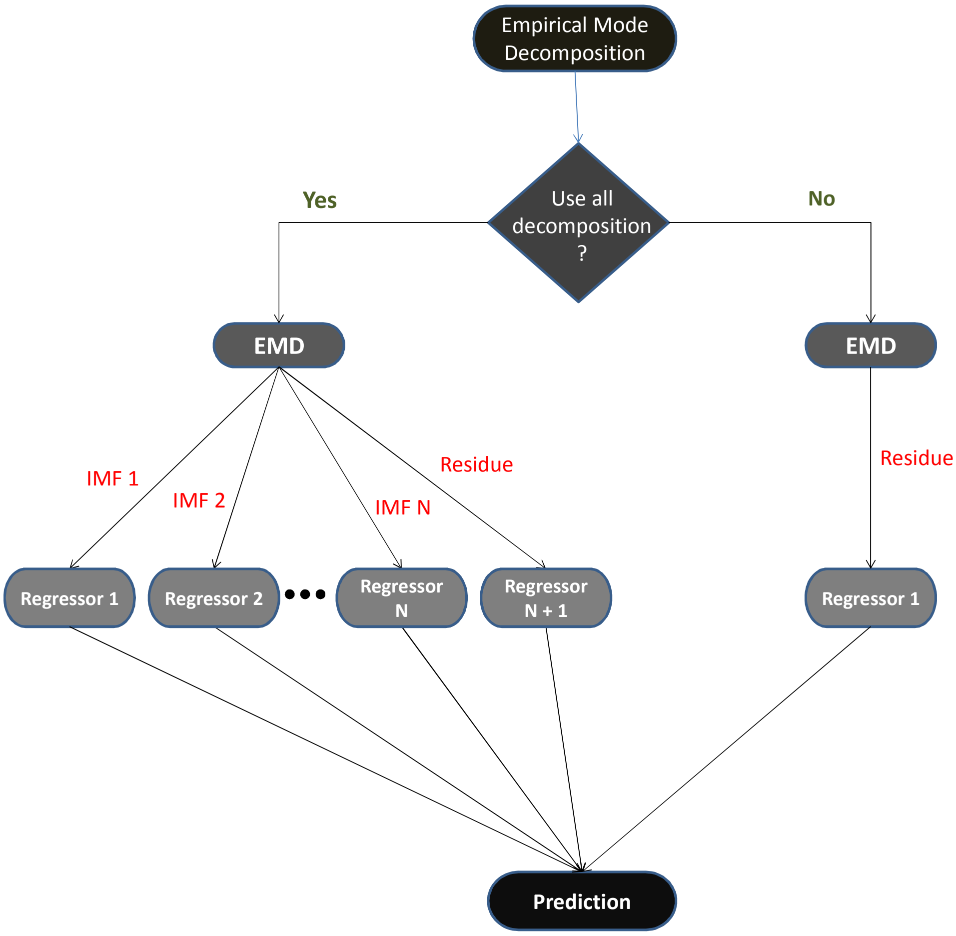

Figure 7: Methodology applied when EMD is used

The next step was to provide the PSO+SVM model with this preprocessed data. Then, the algorithm creates a regression function for each model (1 – IMFs + Residue; 2 – Residue; 3 – Wavelets + Scaling; 4 – Scaling; 5 – No preprocessing) and the results could be compared based on the performance of RUL prediction.

This methodology was applied in two different cases based on the same data set. The first application was performed when the whole vibration run-to-failure data set was provided, i.e., there was data until failure happens. There, a regression model was created and estimation about the RUL was done for each point until the failure based on the vibration data. In the end, the line trend of all points was the predicted RUL. The second application was more challenging, once only part of the test set was provided, that means that there was only the vibration signal until some point far from the failure time. The goal was to correctly estimate the RUL based on the developing behavior of the vibration signal and on the regression functions estimated over the run-to-failure data sets.

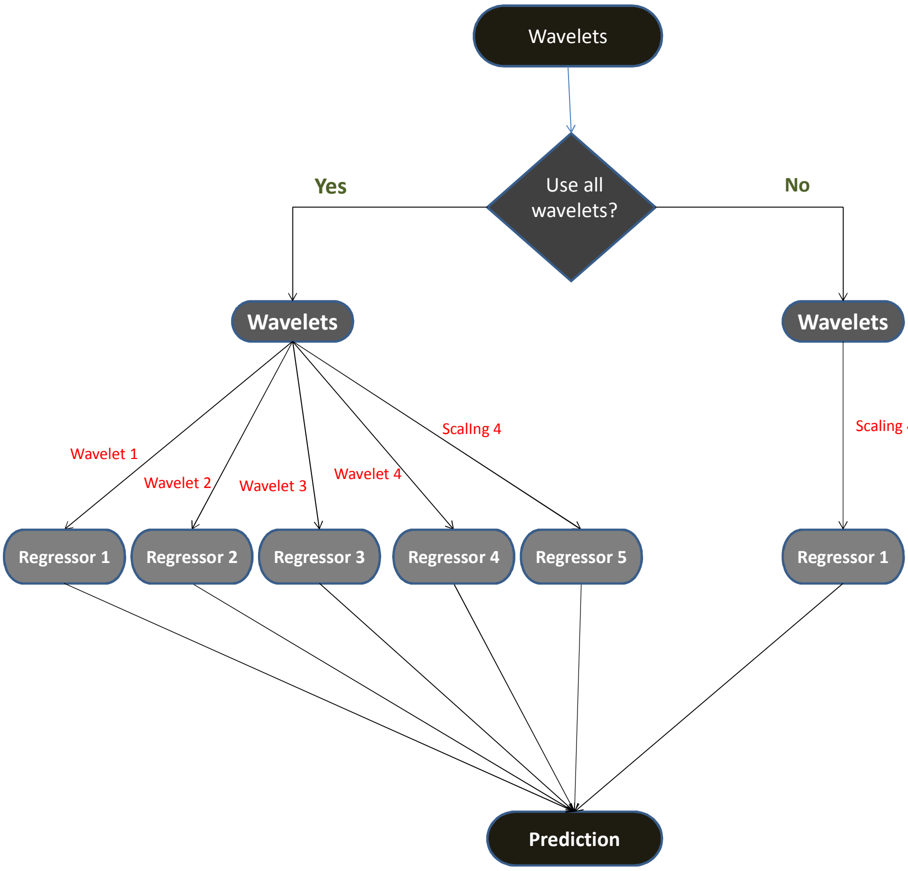

Figure 8: Methodology applied when Wavelets is used

EMD process and Wavelets process were performed in R® by the packages "EMD" (D. Kim & Oh, 2009) and "wavelets", respectively. In the Wavelets process, the Daubechies function was used as mother wavelet and 4 decomposition levels were applied. PSO+SVM were performed in MATLAB® with integrated software "libsvm" (CHANG & LIN, 2011; LINS *et al.*, 2012). For all the applications, the used softwares were R® 3.1.2 (64 bit) and MATLAB® 7.14, processed in a 2.3 GHz, 3.9 GB RAM'S computer and Windows 7 Professional® operational system.

3.2 Overview of the data set

The presented methodology was applied to a real data set provided by FEMTO-ST Institute (Nectoux *et al.*, 2012). The data was used in the IEEE PHM 2012 Data Challenge, focused on the estimation of the RUL of bearings. Experiments were carried out on a laboratory experimental platform (PRONOSTIA), Figure 9, that enables accelerated degradation of bearings under constant and/or variable operating conditions, while gathering online health monitoring data (rotating speed, load force, temperature, vibration).

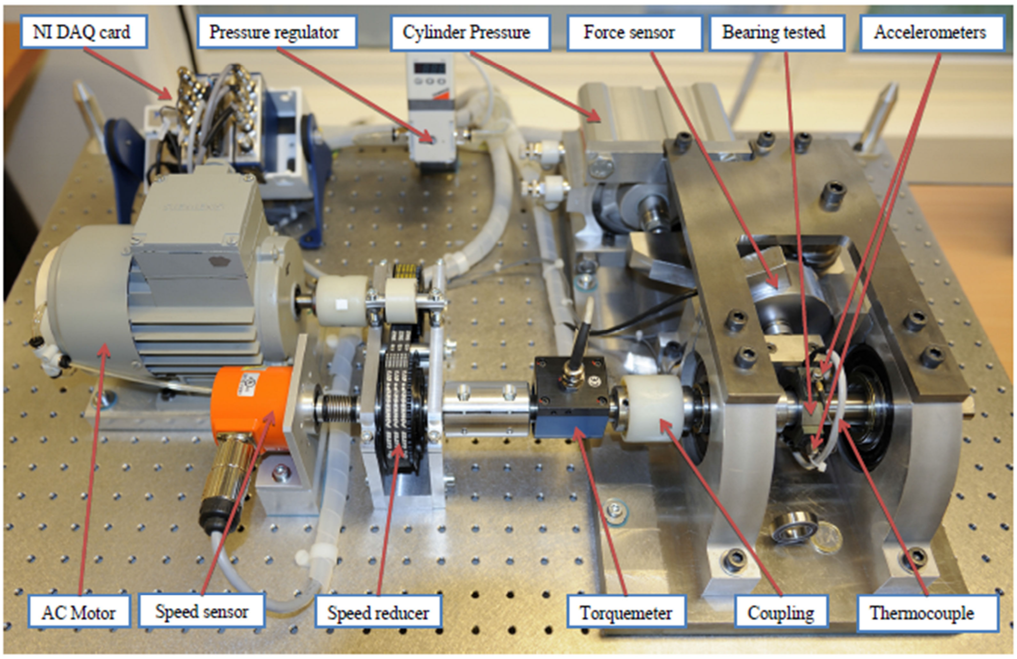

Figure 9: Overview of PRONOSTIA
From Nectoux *et al.* (2012)

Regarding the PHM Challenge, three different categories of data were provided, each of them with a different load and rotational speed applied. In each category, they were constant during all the experiment. Six run-to-failure data sets were provided in order to build some prognostics models, and participants were asked to estimate accurately the RUL for 11 remaining bearings. Monitoring data of the 11 test bearings were truncated so that participants were supposed to predict the RUL. Also, no assumption about the type of failure that had occurred was given, i.e., nothing was known about the nature and the origin of the degradation: balls, inner or outer races, cage. The challenge data sets were characterized by a small amount of training data and a high variability in experiment durations (from 1h to 7h) and thereby, achieving good results was quite difficult (NECTOUX *et al.*, 2012).

Furthermore, even if the data provided for the challenge concerns constant operating conditions for each realized experiment, the current design of PRONOSTIA enables to provide data related to bearings degraded under variable operating conditions. The main objective of PRONOSTIA is to provide experimental data that characterize the degradation of ball bearings along their whole operational life (until their total failure). This experimental platform allows conducting bearings' degradations in only few hours. Also, compared to other bearing test beds proposed in literature, the data provided by the PRONOSTIA platform are different in the sense that they correspond to 'normally' degraded bearings. This means that the defects are not initiated on the bearings and that each degraded bearing contains almost all the types of defects (balls, rings and cage).

3.3 Bearings degradation: run-to-failure experiments

PRONOSTIA platform enables to perform run-to-failure experiments. In order to avoid propagation of damages to the whole test bed (and for security reasons), tests were stopped when the amplitude of the vibration signal measured by accelerometers overpassed 20g measure. Figure 10 depicts an example of what one can observe on the ball bearing components before and after an experiment, as well as a vibration raw signal gathered during a whole experiment. Note that the degradation of bearings depicts very different behaviors leading to very different experiment duration (until the failure).

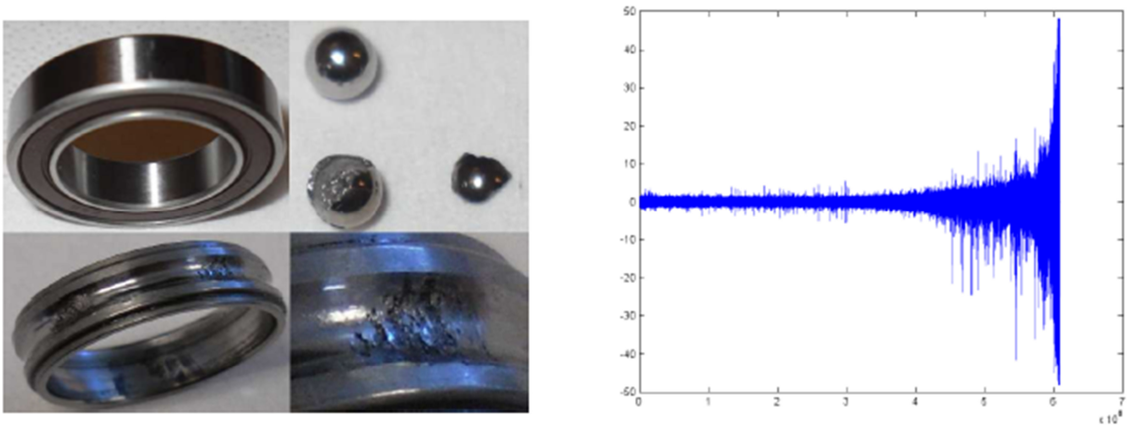

Figure 10: Normal and degraded bearings and vibration raw signal from an experiment
Adapted from Nectoux *et al.* (2012)

The experiments from which the data set was derived involved three different loading conditions. Condition 1 had seven ball bearings operated at 1800 rpm with 4000 N radial load. Complete run-to-failure data for algorithm training were provided for two of the bearings, and truncated data for algorithm testing were provided for the other five bearings. Condition 2 featured seven ball bearings operated at 1650 rpm with 4200 N radial load. Of the seven bearings, complete run to-failure data for training were provided for two, and data for testing were provided for five of the bearings. Condition 3 featured three bearings operated at 1500 rpm with 5000 N radial load. Data from two of the bearings were provided for training and data from the other bearing was provided for testing.

Two accelerometers were mounted on the bearing housing to measure vibration in the vertical and horizontal directions. Data sampling was conducted at 10 seconds intervals at a 25.6 kHz sampling rate and 0.1 seconds duration; hence, each observation contained 2560 points.

3.4 Data Processing

As previously mentioned, data sampling was monitored at a 10 seconds interval and each observation contained 2560 points. For example, the experiment of Bearing 1, which was performed in the first condition, lasted 7 hours and 47 minutes, which means that 2803 observations and 7,175,680 points were recorded. This huge data represents the measure of just one bearing at one condition in one accelerometer. It is easy to see the necessity of data processing to achieve a set of representative points so as to decrease the required computational effort.

Therefore, the feature extraction was done in order to represent the 2560 recorded points in just one point for each observation, once vibration variation within the 0.1 seconds duration interval is negligible when compared with the whole experiments duration. Hence, the whole data set provided in competition had, on average, about 1300 data points for each bearing, in each accelerometer (horizontal and vertical).

Three metrics were considered: peak amplitude, kurtosis and entropy. Each of them was calculated over the observed values on the horizontal axis, on the vertical axis and on the vector sum of the horizontal and vertical points, thereby resulting in a total of nine metrics. The absolute amplitude vibration considered was the average of the five highest absolute peak acceleration values measured in each observation. Averaging was done to reduce the effect of data noise.

In general, the data was very disturbed and not a particular behavior was presented for any metric. Further, anomalous features such as a considerable number of high peaks at the beginning of the experiment followed by small peaks after a while was observed in many bearings. It seems to indicate an improvement on this experimental test, which is not correct. This type of behavior indicates the difficulty in data processing, as well as to decide which summary metric should be used for further analysis.

4 APPLICATION EXAMPLES

As previously mentioned, the data set presented an unusual behavior for every metric analyzed and the metric chosen was the one that presented the most ‘normal’ expected degradation behavior. For the application of the master thesis methodology, a training set is necessary, from which the machine will learn about the bearing degradation behavior and a test set, where the same machine will try to correctly predict another bearing behavior. A comparison among models with EMD, with Wavelets and without any preprocessing technique is made in order to identify what is the gain, if present, of those preprocessing techniques in the data set.

After several investigations, the absolute frequency amplitude was considered the most suitable among all the metrics and was the work’s focus. Furthermore, the horizontal absolute amplitude vibration presents a better behavior and was the chosen one. Comparisons between the bearings in their respective conditions were made and the condition 1 was selected to be analyzed, by applying the methodology to the Bearing 1 (learning set) and the Bearing 3 (test set).

Bearing 1 had 2803 observations in a run-to-failure test. SVM is a supervised learning method, which requires the necessity of both y, the response variable, and $\boldsymbol{x}$, the regressor/input variables. In all cases, the response variable was the RUL and the regression variables were the amplitude vibrations. In the EMD case, one model where each IMF and the Residue were considered as a regressor variable and other model where just the Residue was the regressor variable; in the Wavelet case, one model contains each Wavelet and the last Scaling function as regressors and the other model considered just the last Scaling function. The last model, without EMD or Wavelets as preprocessing techniques, had the signal resulted from the vibration feature extraction as regressor.

The direct point prediction is not expected to be reliable, due to the high variability and disorder of the data, but the trend of all predictions should express the realistic RUL estimation, as seen in Figure 11:

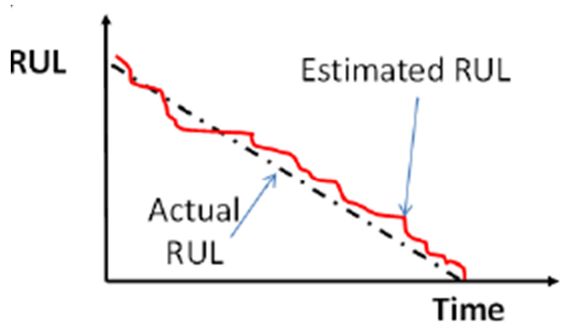

Figure 11: Expected estimated RUL behavior
From Sutrisno *et al.* (2012)

The goal was training the machine with the learning set (Bearing 1) and test the machine with the test set from Bearing 3. The predicted RUL would be compared with the real RUL, also provided by the FEMTO-ST Institute in Nectoux *et al.* (2012). This comparison is only possible because the FEMTO-ST Institute provided the complete data set after the end of the challenge in 2012. Bearing 3 was a test bearing and had 1802 truncated observations and 2374 complete observations.

Even so, after the feature extraction, the processing of 2803 data by PSO+SVM is still computational expensive, once that PSO is a probabilistic algorithm that needs to explore through the search space for the best solution. Therefore, the data were divided in four different regions, each one representing one degradation phase of the bearing. These four regions are shown in Figure 12.

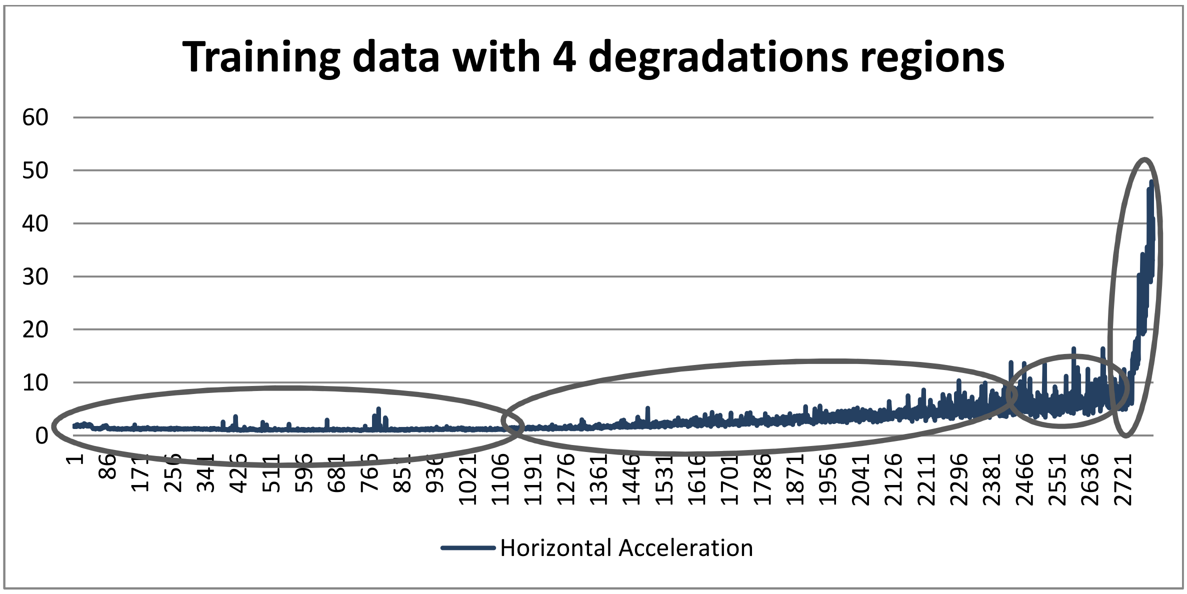

Figure 12: The four different regions of degradation

The four regions represent different degradation states of Bearing 1. The first one represents a healthy state, where the bearing is still considered new and had duration of 12000 seconds (1200 samples). The second region represents an initial degraded state, where there is an increase in the vibration mean, but still considered stable and this phase lasts 13000 seconds (1300 samples). In the third, the mean keep growing and also the variance starts to increase, considered as a high degradation state, lasting 2400 seconds (240 samples). The fourth is almost a failed state, with the mean strongly increasing in each observation and with the variance oscillating heavily, during 510 seconds (51 samples).

In order to reduce the data quantity, the data sampling in every region had a different frequency, once the more stable the bearing is, the less necessary is the monitoring every ten seconds. The suitable values used in the data sampling was 400 seconds for the first region (30 points), 200 seconds to the second region (65 points), 100 seconds for the third region (25 points) and 30 seconds to the last one (17 points). In this way, the data set was reduced to 137 points, which are easily handled by the PSO+SVM method. Similar procedure was made with data of Bearing 3, reducing the 2374 data points to just 120, using the same sampling frequency in each of the degradation regions applied to Bearing 1.

After applying all the methodology and predicting the RUL, it is necessary to get back to real space between observations, i.e. before the sampling was performed, and compare the results with the 2374 points from the original data. It was done adding the respective sampling time lapse in all four degradation regions, so as to the RUL predicted vector has “length” 2374 (i.e. it is created a “sparse” vector with non-zero values only in the sampling time).

4.1 Complete Test Set Case

In the first application case, all test data of Bearing 3 were provided and estimations about the RUL were made for all data, i.e. every test point until the failure has an estimated RUL. As previously mentioned, it is not expected a good point prediction, but the trend should correctly express the degradation behavior. Figure 13 depicts an example (e.g. PSO+SVM model with no preprocessing technique) of the applied procedure. Each one of the 120 points has a real RUL, represented by darks diamonds, and its own prediction, represented by the light grey squares. A linear trend is calculated from the predictions and is it presented as a solid black line, thereby creating a good representation of the real RUL. This procedure was performed for the five models under analysis: IMFs + Residue; Residue; Wavelets + Scaling; Scaling; no preprocessing.

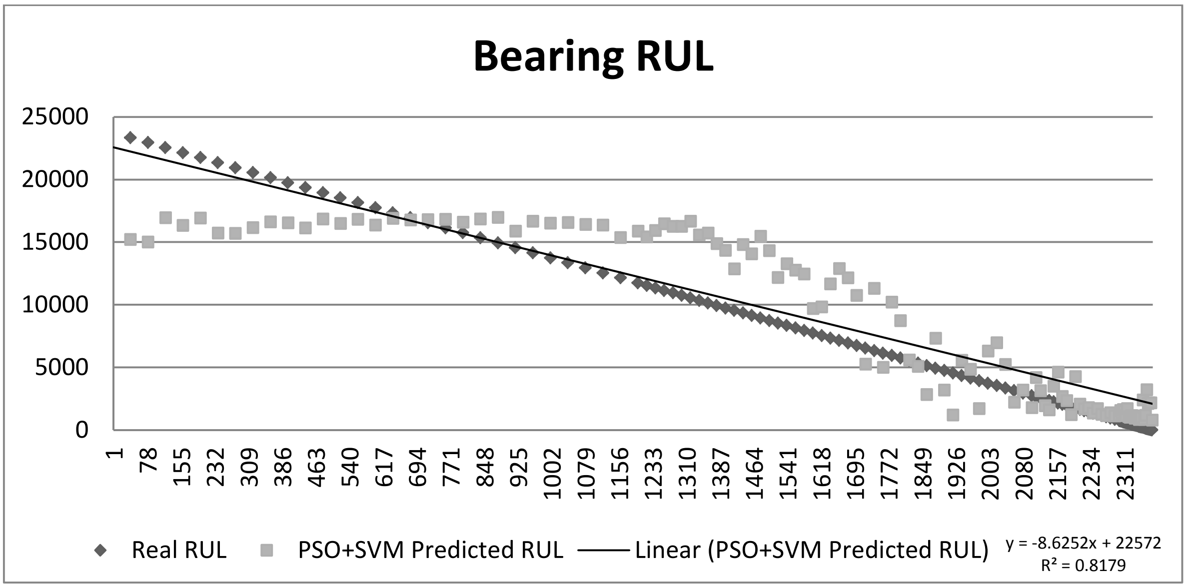

Figure 13: RUL estimation for complete data case

In order to correctly estimate the RUL, two metrics were applied: in the first, it was considered a simple average between the y axis cross value and 10 times the x cross value, once both should represent the predicted RUL for the whole experiment. The multiplication by 10 of x axis was done just to treat both axis as time axis (multiplication of the number of samples by time lapse between each sample). Then we subtract the real RUL and divide this value for the real RUL to get a Percentage Error (PE). This is expressed in the equation below:

$$PE = \frac{\frac{(y\ cross\ value + 10\ (x\ cross\ value))}{2} - Real\ RUL}{Real\ RUL} \% \qquad (2.377)$$

The second metric was an Absolute Percentage Error (APE), where we first calculate the absolute difference value between y axis cross value and the real RUL; then we calculate the absolute difference value between 10 times the x cross value and the real RUL. Ideally, both should have the same value, due to the fact the y axis represents the real RUL as well as the x axis times 10 (seconds), and the experiment was a run-to-failure (i.e., the last RUL is zero). This absolute value quantifies the distance error. To compare the models, we calculate the percentage error, dividing the RUL prediction for both models to the real RUL, in this case, 23740 seconds.

$$APE = \frac{\frac{(|Real\ RUL - \ y\ cross\ value| + |Real\ RUL - \ 10\ (x\ cross\ value)|)}{2}}{Real\ RUL} \% \quad (2.388)$$

Both errors are presented in Table 1 for each tested model. Model using as regressor only the Residue provided by EMD preprocessing technique present the smallest PE as well as APE, representing the best model in this application. Still, all PE and almost all APE presented errors lesser than 10%, even if the singularity of the data. Still, EMD+PSO+SVM model presents almost the same PE and APE, which mean that RUL correctly decrease between samples, errors representing just a time translations, i.e., every RUL prediction differs from the same time lapse from the real RUL. Parameters from PSO of each model can be found in Appendix A.

Table 1: Table of errors for all tested models in first application

Model	Regressors	PE	APE
EMD+PSO+SVM	IMFs + Residue	2.54%	2.54%
Residue+PSO+SVM	Residue	0.34%	1.45%
Wavelets+PSO+SVM	Wavelets + V4	0.48%	8.70%
V4+PSO+SVM	V4	1.00%	15.00%
PSO+SVM	Direct Vibration Data	2.66%	7.58%

4.2 Truncated Test Set Case

The second application case was to perform exactly the IEEE PHM 2012 Data Challenge provided by the FEMTO-ST Institute. In this case, only truncated data was provided for the test bearing and the challenge was to calculate the RUL from it. Once more, all procedure applied in the first case was done: feature extraction first, then sampling to reduce the amount of data, training with Bearing 1 data and test with Bearing 3 truncated data (i.e. there is no data until failure). Again, point estimations were not expected to be good predictions, but the series trend should represent the bearing's behavior.

Bearing 3 truncated test set consisted of 1802 points that are depicted Figure 14. As well as Bearing 1 training set, it is expected to Bearing 3 to pass through all four degradations regions, even if the truncated data still does not present all of them. By analyzing test data, Bearing 3 only reached first and second degradation regions and inference has to be done to further degradation zones.

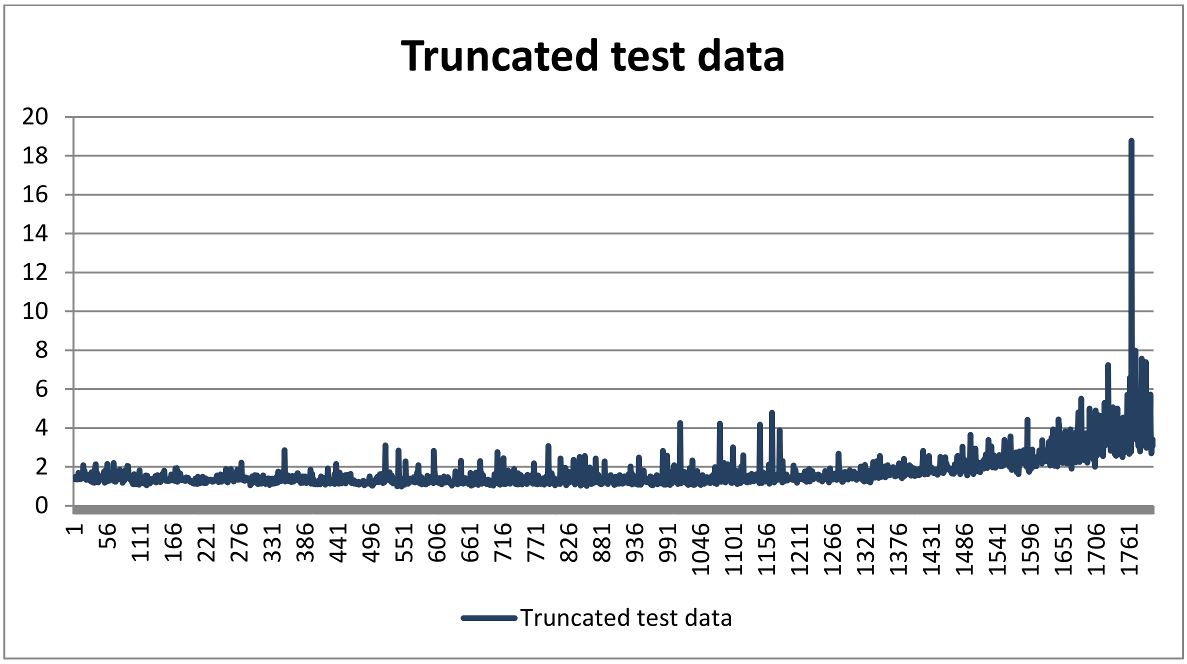

Figure 14: Truncated test set used in second application

From Figure 12, note that vibration in first degradation region (i.e. health stage) is almost stationary, with negligible fluctuations compared with the whole vibration data set. Analogously, data from first region of the test bearing were considered until point 1200 and the remaining were from second region. Note that there is only one outgoing peak near the end of truncated data (even if its values is still far from a failure value), but this not correspond to a trend change and it was probably due to a noise in test. Thus, it is still considered second region. Once truncated test data only provide with information from regions 1 and 2, use data from region 1 will not represent any gain about bearing degradation. Even more, using data from region 1 will only deviate the overall trend.

Therefore, in this application, none of data from the first region was used. Also, sampling of the second region was changed to contain more information, i.e. it was used the sampling rate of the fourth region (every 30 seconds/3 samples). Thus, data test was reduced to 200 test points used for estimation. Again, it is applied the same idea as in the first example and Figure 15 illustrate one example model (PSO+SVM without preprocessing techniques) of how conducted was carried predictions. Dark gray line represents the real RUL and light grey light represents estimations. Thin black line represent the trend overall predictions.

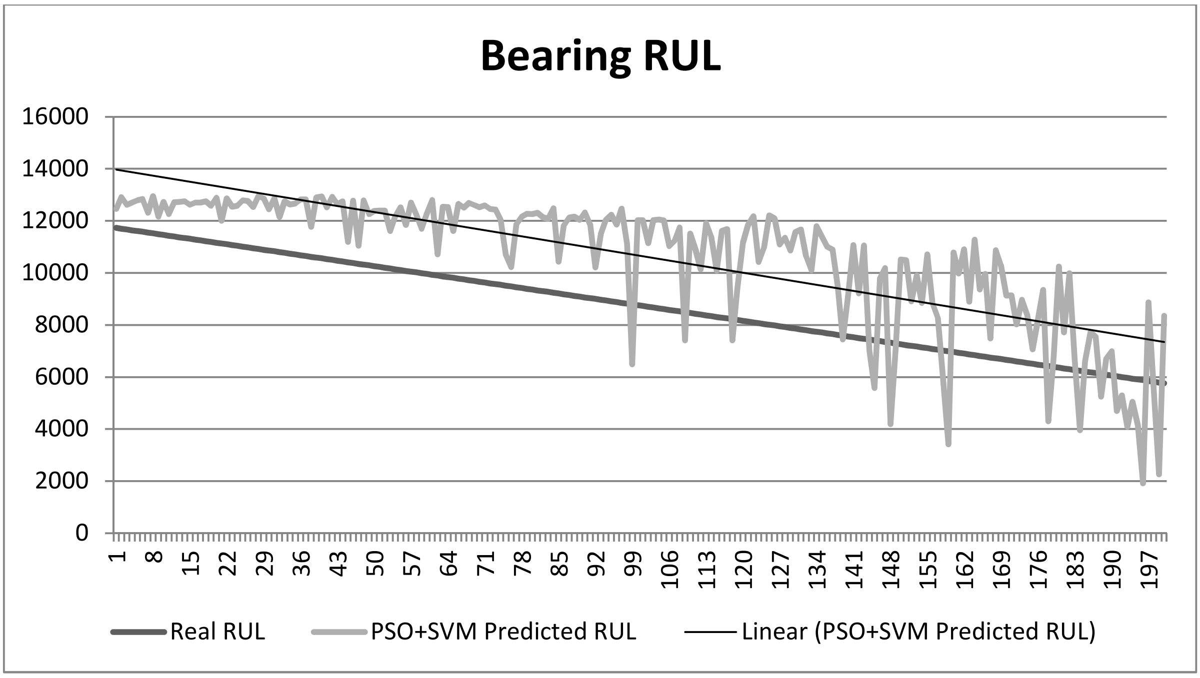

Figure 15: Predictions for RUL in the IEEE PHM Data Challenge

Here, note that predictions do not occur until y axis crossing, i.e., until failure. So, to define RUL estimation for the challenge, thin black line (trend) was extrapolated until it crossed y axis and this value was used as the desired estimation. RUL for each model was compared with the true RUL value of 5370 (noticed from the last time of truncated data to the failure time of the bearing) and models with good performance should predict values similar to this. For errors measurement, PE was used once there is only one estimated RUL for each model. Results are shown in Table 2. Errors for models with Wavelets application are not displayed, once they predicted negative values for RUL, which does not occur in reality. Parameters from PSO of each model can be found in Appendix A.

Table 2: Table of errors for tested models in second application

Model	Regressors	Percentage Errors
EMD+PSO+SVM	IMFs + Residue	15.39%
Residue+PSO+SVM	Residue	24.90%

PSO+SVM	Direct Vibration Data	58.53%

In this application, EMD+PSO+SVM model presented the best performance. Again, it is important to highlight the difficulty of this challenge, which is evidenced by the consideration of errors of this order of magnitude as acceptable. Indeed, in comparison, Sutrisno *et al.* (2012) presented the prediction errors for the winner of the challenge and the error for this bearing was 37%, worse than two of our estimations.

5 CONCLUDING REMARKS

5.1 Conclusions

The concern about systems efficiency has been the focus in most of the companies and the technological progress subordinated this efficiency to machinery and equipments. Enhance the availability and utility of system becomes necessary and condition monitoring programs have presented vital importance.

This work proposed a comparison between the ability of models using EMD or Wavelets as preprocessing technique to a PSO+SVM learning algorithm to correctly predict the RUL of rolling bearings from a vibration signal. Moreover, the comparison was applied to a real dataset provided for a PHM Challenge competition. Due to the duty of a challenge, the data was difficult to analyze and it provided some odd features.

Two cases of applications were considered: (i) when the whole test data set was provided and (ii) when a truncated test data set was provided. In the first case, almost all models presented good estimations for RUL, especially the ones with EMD+PSO+SVM models, the one with just the Residue as regressor and the one using IMFs and the Residue as regressors. To measure model's performance, two types of errors (PE and APE) were applied. The second case was more challenging, given that fewer amounts of data were provided, with the bearing in the early stages of degradation. In this case, only some models could provide acceptable RUL, but the ones that did it provided good estimates. Even more, two of three models provided better RUL predictions than the winner of the PHM Challenge competition.

Even if, in general, performing PSO+SVM learning algorithm already represent a reasonable estimation, apply preprocessing techniques to treat the data could represent a gain in terms of performance prediction on estimating the RUL for rolling bearings. Moreover, in the exemplified examples, EMD overcomes Wavelets providing better results.

Therefore, more reliable predictions could be performed improving information related to operational conditions. Decision-making about RAMS problems are support with these information, aiming to reduce costs and risk, and maximize efficiency and production capacity.

5.2 Limitations and future works

Despite usually producing satisfactory predictions, PSO+SVM only could handle relatively small samples due to the probabilistic PSO algorithm. Thus, the dataset had to be reduced in more than 10 times to be processed and this could lead to loss of information and misinterpretations. Nevertheless, vibration signals are often too large, given that the sampling frequency is very high, thus data set reduction is almost a mandatory step.

Besides, the data set was considerable confusing and the application of the methodology in a more 'normal' data set could verify, or not, the same conclusions of this master thesis. IEEE PHM Challenge also provides information of other bearings that could be used for more test analysis. Furthermore, the data set also provides information about temperature, not used in this analysis. This could point to other ways to interpreting the degradation signal and possible forms to predict the RUL.

For future research, a comparison between the EEMD (Ensemble Empirical Mode Decomposition) preprocessing could be done, to verify if this approach leads to a better prediction. EEMD tries to solve some problems of EMD, such as the high variability in the series extremes and mode mixing.

Another future research could be the elaboration of a complete maintenance plan based in this prediction in order to maximize the availability and minimize some kind of cost. Even more, an online monitoring using this approach that is updated *in loco* should bring great results to industrial company machinery availability.

REFERENCES

Acharya, U. R., Fujita, H., Sudarshan, V. K., Mookiah, M. R. K., Koh, J. E., Tan, J. H., … Ng, K. H. (2016). An integrated index for identification of fatty liver disease using radon transform and discrete cosine transform features in ultrasound images. *Information Fusion*, *31*, 43–53.

Allen, J. (1977). Short term spectral analysis, synthesis, and modification by discrete Fourier transform. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *25*(3), 235–238.

Benkedjouh, T., Medjaher, K., Zerhouni, N., & Rechak, S. (2013). Remaining useful life estimation based on nonlinear feature reduction and support vector regression. *Engineering Applications of Artificial Intelligence*, *26*(7), 1751–1760.

Bezerra Souto Maior, C., das Chagas Moura, M., Didier Lins, I., Lopez Droguett, E., & Henrique Lima Diniz, H. (2016). Remaining Useful Life Estimation by Empirical Mode Decomposition and Support Vector Machine. *IEEE Latin America Transactions*, *14*(11), 4603–4610.

Bin, G. F., Gao, J. J., Li, X. J., & Dhillon, B. S. (2012). Early Fault Diagnosis of Rotating Machinery based on Wavelet Packets—Empirical Mode Decomposition Feature Extraction and Neural network. *Mechanical Systems and Signal Processing*, *27*, 696–711.

Bratton, D., & Kennedy, J. (2007). Defining a Standard for Particle Swarm Optimization. In *2007 IEEE Swarm Intelligence Symposium* (pp. 120–127).

Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 27.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*(3), 273–297.

Cuss, C. W., McConnell, S. M., & Guéguen, C. (2016). Combining parallel factor analysis and machine learning for the classification of dissolved organic matter according to source using fluorescence signatures. *Chemosphere*, *155*, 283–291.

Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, *36*(5), 961–1005. https://doi.org/10.1109/18.57199

Daubechies, I. (1993). Ten Lectures on Wavelets. *SIAM Review*.

Dong, M., & He, D. (2007). A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. *Mechanical Systems and Signal Processing*, *21*(5), 2248–2266.

Droguett, E. L., Lins, I. D., Moura, M. C., Zio, E., & Jacinto, C. M. (2014). Variable selection and uncertainty analysis of scale growth rate under pre-salt oil wells conditions using support vector regression. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *229*(4), 319–326.

Eftekhar, A., Toumazou, C., & Drakakis, E. M. (2013). Empirical Mode Decomposition: Real-Time Implementation and Applications. *Journal of Signal Processing Systems*, *73*(1), 43–58.

Engel, S. J., Gilmartin, B. J., Bongort, K., & Hess, A. (2000). Prognostics, the real issues involved with predicting life remaining. In *2000 IEEE Aerospace Conference. Proceedings (Cat. No.00TH8484)* (Vol. 6, pp. 457–469).

Fan, J., & Tang, Y. (2013). An EMD-SVR method for non-stationary time series prediction. *2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE)*, 1765–1770.

García Nieto, P. J., García-Gonzalo, E., Sánchez Lasheras, F., & De Cos Juez, F. J. (2015). Hybrid PSO-SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability. *Reliability Engineering and System Safety*, *138*, 219–231.

Geng, X., Zhan, D. C., & Zhou, Z. H. (2005). Supervised nonlinear dimensionality reduction for visualization and classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *35*(6), 1098–1107.

Goupillaud, P., Grossmann, A., & Morlet, J. (1984). Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, *23*(1), 85–102.

Grossmann, A., & Morlet, J. (1984). Decomposition of Hardy Function into Square Integrable Wavelets of Constant Shape. *SIAM J. Math. Anal.*, *15*(4), 723–736.

Guohua, G., Yu, Z., Guanghuang, D., & Yongzhong, Z. (2006). Intelligent Fault Identification Based On Wavelet Packet Energy Analysis and SVM. *2006 9th International Conference on Control, Automation, Robotics and Vision*, *1*(3), 1–5.

Haykin, S. (1998). Neural Networks. *Convergence*, 1–16.

Hemmati, F., Orfali, W., & Gadala, M. S. (2016). Roller bearing acoustic signature extraction by wavelet packet transform, applications in fault detection and size estimation. *Applied Acoustics*, *104*, 101–118.

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

Hu, W., Wang, S., He, Z., & Zhao, S. (2010). Prognostic analysis based on updated grey model for axial piston pump. In *Proceedings of the 2010 5th IEEE Conference on Industrial Electronics and Applications, ICIEA 2010* (pp. 571–575).

Hu, Y., Liu, S., Lu, H., & Zhang, H. (2014). Remaining useful life assessment and its application in the decision for remanufacturing. *Procedia CIRP*, *15*, 212–217.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., … Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*.

Huang, N. E., & Wu, Z. (2008). A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, *46*(2), 1–23.

Huang, S., Chang, J., Huang, Q., & Chen, Y. (2014). Monthly streamflow prediction using modified EMD-based support vector machine. *Journal of Hydrology*, *511*, 764–775.

Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2007). A tutorial on kernel methods for categorization. *Journal of Mathematical Psychology*.

Kan, M. S., Tan, A. C. C., & Mathew, J. (2015). A review on prognostic techniques for non-stationary and non-linear rotating systems. *Mechanical Systems and Signal Processing*, *62–63*, 1–20.

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Neural Networks, 1995.*

Proceedings., IEEE International Conference on, *4*, 1942–1948 vol.4.

Khelif, R., Chebel-Morello, B., & Zerhouni, N. (2015). Experience based approach for Li-ion batteries RUL prediction. In *IFAC Proceedings Volumes (IFAC-PapersOnline)* (Vol. 48, pp. 761–766).

Kim, D., & Oh, H. (2009). EMD: A Package for Empirical Mode Decomposition and Hilbert Spectrum. *The R Journal*, *1*(1), 40–46.

Kim, K. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1–2), 307–319. https://doi.org/10.1016/S0925-2312(03)00372-2

Konar, P., & Chattopadhyay, P. (2011). Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs). *Applied Soft Computing*, *11*(6), 4203–4211.

Kumar, P., & Foufoula-Georgiou, E. (1997). Wavelet analysis for geophysical applications. *Reviews of Geophysics*, *35*(4), 385.

Lei, Y., He, Z., & Zi, Y. (2008). A new approach to intelligent fault diagnosis of rotating machinery. *Expert Systems with Applications*, *35*(4), 1593–1600.

Liao, L., & Köttig, F. (2014). Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Transactions on Reliability*, *63*(1), 191–207.

Lins, I. D., Araujo, M., Das, M., Moura, C., Silva, M. A., & Droguett, E. L. (2013). Computational Statistics and Data Analysis Prediction of sea surface temperature in the tropical Atlantic by support vector machines. *Computational Statistics and Data Analysis*, *61*, 187–198.

Lins, I. D., Moura, M. D. C., Zio, E., & Droguett, E. L. (2012). A particle swarm-optimized support vector machine for reliability prediction. *Quality and Reliability Engineering International*, *28*(2), 141–158.

Lins, I., Moura, M., & Droguett, E. (2013). Failure prediction of oil wells by support vector regression with variable selection, hyperparameter tuning and uncertainty analysis. *Chemical Engineering Transactions*, *33*, 817–822.

Liu, JF *et al*. Remote Fault Diagnosis System Based on EMD and SVM for Heavy Rolling-mills; *Engineering Solutions for Manufacturing Processes IV*, PTS 1 and 2,Vol. 889-890; p. 681-686; 2014.

Liu, Z., Cao, H., Chen, X., He, Z., & Shen, Z. (2013). Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings. *Neurocomputing*, *99*, 399–410.

Liu, Z., Wang, L., Zhang, Y., & Chen, C. L. P. (2016). A SVM controller for the stable walking of biped robots based on small sample sizes. *Applied Soft Computing*, *38*, 738–753.

Lybeck, N., Marble, S., & Morton, B. (2007). Validating Prognostic Algorithms: A Case Study Using Comprehensive Bearing Fault Data. In *Aerospace Conference, 2007 IEEE* (pp. 1–9). https://doi.org/10.1109/AERO.2007.352842

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

Mitoma, T., Wang, H., & Chen, P. (2008). Fault diagnosis and condition surveillance for plant rotating machinery using partially-linearized neural network. *Computers and Industrial*

Engineering, *55*(4), 783–794.

Morlet, J., Arens, G., Fourgeau, E., & Giard, D. (1982). Wave propagation and sampling theory—Part II: Sampling theory and complex waves. *Geophysics*, *47*(2), 222–236.

Moura, M. C., & Droguett, E. L. (2009). Mathematical formulation and numerical treatment based on transition frequency densities and quadrature methods for non-homogeneous semi-Markov processes. *Reliability Engineering and System Safety*, *94*(2), 342–349.

Moura, M. D. C., Zio, E., Lins, I. D., & Droguett, E. (2011). Failure and reliability prediction by support vector machines regression of time series data. *Reliability Engineering & System Safety*, *96*(11), 1527–1534.

Najafi, G., Ghobadian, B., Moosavian, A., Yusaf, T., Mamat, R., Kettner, M., & Azmi, W. H. (2016). SVM and ANFIS for prediction of performance and exhaust emissions of a SI engine with gasoline-ethanol blended fuels. *Applied Thermal Engineering*, *95*, 186–203.

Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., & Varnier, C. (2012). PRONOSTIA : An experimental platform for bearings accelerated degradation tests. In *IEEE International Conference on Prognostics and Health Management, PHM'12* (pp. 1–8).

Nikolaou, N. G., & Antoniadis, I. A. (2002). Rolling element bearing fault diagnosis using wavelet packets. *NDT & E International*, *35*(3), 197–205.

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

Peng, Z. K., & Chu, F. L. (2004). Application of the Wavelet Transform in Machine Condition Monitoring and Fault Diagnostics: A Review with Bibliography. *Mechanical Systems and Signal Processing*, *18*(2), 199–221.

Qiu, H., Lee, J., Lin, J., & Yu, G. (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of Sound and Vibration*, *289*(4–5), 1066–1090.

Raghavendra, U., Rajendra Acharya, U., Fujita, H., Gudigar, A., Tan, J. H., & Chokkadi, S. (2016). Application of Gabor wavelet and Locality Sensitive Discriminant Analysis for automated identification of breast cancer using digitized mammogram images. *Applied Soft Computing Journal*, *46*, 151–161.

Rai, A., & Upadhyay, S. H. (2016). A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribology International*.

Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics-A tutorial. *Mechanical Systems and Signal Processing*, *25*(2), 485–520.

Rausand, M., & Hsyland, A. (2004). *System Reliability Theory: Models, Statistical Methods, and Applications*. *Wescon/96*.

Rojas, A., Górriz, J. M., Ramírez, J., Illán, I. A., Martínez-Murcia, F. J., Ortiz, A., … Moreno-Caballero, M. (2013). Application of Empirical Mode Decomposition (EMD) on DaTSCAN SPECT images to explore Parkinson Disease. *Expert Systems with Applications*, *40*, 2756–2766.

Samanta, B., & Nataraj, C. (2009). Use of particle swarm optimization for machinery fault detection. *Engineering Applications of Artificial Intelligence*, *22*(2), 308–316.

Sanei, S., & Chambers, J. a. (2007). *EEG Signal Processing*. *Chemistry & biodiversity* (V. 1).

Saxena, M., Bannett, O. O., & Sharma, V. (2016). Bearing Fault Evaluation for Structural Health Monitoring, Fault Detection, Failure Prevention and Prognosis. In *Procedia Engineering* (Vol. 144, pp. 208–214).

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(July 1928), 379–423.

Shawe-Taylor, J., & Sun, S. (2014). *Academic Press Library in Signal Processing: Volume 1 - Signal Processing Theory and Machine Learning*. *Academic Press Library in Signal Processing* (Vol. 1).

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

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

Su, Z., Tang, B., Liu, Z., & Qin, Y. (2015). Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine. *Neurocomputing*, *157*, 208–222.

Sutharssan, T., Stoyanov, S., Bailey, C., & Rosunally, Y. (2012). Prognostics and health monitoring of high power LED. *Micromachines*, *3*, 78–100.

Sutrisno, E., Oh, H., Vasan, A. S. S., & Pecht, M. (2012). Estimation of remaining useful life of ball bearings using data driven methodologies. *2012 IEEE Conference on Prognostics and Health Management*, *2*, 1–7.

Tan, A. C., Naiman, D. Q., Xu, L., Winslow, R. L., & Geman, D. (2005). Simple decision rules for classifying human cancers from gene expression profiles. *Bioinformatics*, *21*(20), 3896–3904.

Tandon, N., & Choudhury, A. (1999). A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology International*, *32*(8), 469–480.

Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A., & Wu, B. (2007). *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*.

Vapnik, V. (2000). *The Nature of Statistical Learning Theory*. *Springer* (Vol. 8).

Wang, C, Jia, LM & Li, XF . A study of fault diagnosis method for the train axle box based on EMD and PSO-LSSVM. *2014 Third International Conference on instrumentation & Measurement, Computer, Communication and Control (IMCCC).*Pag. 520-525.

Wang, L. (2005). Support Vector Machines : Theory and Applications.

Widodo, A., & Yang, B. S. (2011). Machine health prognostics using survival probability and support vector machine. *Expert Systems with Applications*, *38*(7), 8430–8437.

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

Yan, R., Gao, R. X., & Chen, X. (2014). Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Processing*, *96*(PART A), 1–15.

Yang, Y., Yu, D., & Cheng, J. (2007). A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. *Measurement: Journal of the International Measurement Confederation*, *40*(9–10), 943–950.

Zhao, J., & Dimirovski, G. M. (2004). Quadratic stability of a class of switched nonlinear systems. *IEEE Transactions on Automatic Control*, *49*(4), 574–578.

Zhao, Z., Quan, Q., & Cai, K.-Y. (2014). A Profust Reliability Based Approach to Prognostics and Health Management. *IEEE Transactions on Reliability*, *63*(1), 26–41.

Zio, E. (2009). Reliability engineering: Old problems and new challenges. *Reliability Engineering and System Safety*. https://doi.org/10.1016/j.ress.2008.06.002

APPENDIX A

Complete Test Set Case

Model	PSO Parameters		
	C	ε	γ
EMD+PSO+SVM	1499.99999	277.7990	0.194966
Residue+PSO+SVM	1499.99996	33.8814	0.177431
Wavelets+PSO+SVM	1499.98568	9.6023	0.290502
V4+PSO+SVM	1499.99968	30.4670	0.100154
PSO+SVM	1492.50063	760.5810	0.106303

Truncated Test Set Case

Model	PSO Parameters		
	C	ε	γ
EMD+PSO+SVM	1499.99997	603.2615	0.106401
Residue+PSO+SVM	1499.99996	607.9639	0.100001
PSO+SVM	1499.99988	20.1639	0.100001