Article citation info:
MAIOR CBS, CHAGAS MOURA M, LINS ID. Particle swarm-optimized support vector machines and pre-processing techniques for remaining useful life estimation of bearings. Eksploatacja i Niezawodnosc – Maintenance and Reliability 2019; 21 (4): 610–619, http://dx.doi.org/10.17531/ein.2019.4.10.

Caio Bezerra Souto MAIOR
Márcio das CHAGAS MOURA
Isis Didier LINS

PARTICLE SWARM-OPTIMIZED SUPPORT VECTOR MACHINES AND PRE-PROCESSING TECHNIQUES FOR REMAINING USEFUL LIFE ESTIMATION OF BEARINGS

ZASTOSOWANIE MASZYN WEKTORÓW NOŚNYCH ZOPTYMALIZOWANYCH METODĄ ROJU CZĄSTEK ORAZ TECHNIK PRZETWARZANIA WSTĘPNEGO DO OCENY POZOSTAŁEGO OKRESU UŻYTKOWANIA ŁOŻYSK

The useful life time of equipment is an important variable related to system prognosis, and its accurate estimation leads to several competitive advantage in industry. In this paper, Remaining Useful Lifetime (RUL) prediction is estimated by Particle Swarm optimized Support Vector Machines (PSO+SVM) considering two possible pre-processing techniques to improve input quality: Empirical Mode Decomposition (EMD) and Wavelet Transforms (WT). Here, EMD and WT coupled with SVM are used to predict RUL of bearing from the IEEE PHM Challenge 2012 big dataset. Specifically, two cases were analyzed: considering the complete vibration dataset and considering truncated vibration dataset. Finally, predictions provided from models applying both pre-processing techniques are compared against results obtained from PSO+SVM without any pre-processing approach. As conclusion, EMD+SVM presented more accurate predictions and outperformed the other models.

Keywords: *big data, vibration signal, bearings, remaining useful life, empirical mode decomposition, wavelets transform, support vector machine, particle swarm optimization*

Okres użytkowania sprzętu jest ważną zmienną związaną z prognozowaniem pracy systemu, a możliwość jego dokładnej oceny daje zakładom przemysłowym znaczną przewagę konkurencyjną. W tym artykule pozostały czas pracy (Remaining Useful Life, RUL) szacowano za pomocą maszyn wektorów nośnych zoptymalizowanych rojem cząstek (SVM+PSO) z uwzględnieniem dwóch technik przetwarzania wstępnego pozwalających na poprawę jakości danych wejściowych: empirycznej dekompozycji sygnału (Empirical Mode Decomposition, EMD) oraz transformat falkowych (Wavelet Transforms, WT). W niniejszej pracy, EMD i falki w połączeniu z SVM wykorzystano do prognozowania RUL łożyska ze zbioru danych IEEE PHM Challenge 2012 Big Dataset. W szczególności, przeanalizowano dwa przypadki: uwzględniający kompletny zestaw danych o drganiach oraz drugi, biorący pod uwagę okrojoną wersję tego zbioru. Prognozy otrzymane na podstawie modeli, w których zastosowano obie techniki przetwarzania wstępnego porównano z wynikami uzyskanymi za pomocą PSO + SVM bez wstępnego przetwarzania danych. Wyniki pokazały, że model EMD + SVM generował dokładniejsze prognozy i tym samym przewyższał pozostałe badane modele.

Słowa kluczowe: *duże dane, sygnał drgań, łożyska, pozostały okres użytkowania, empiryczna dekompozycja sygnału, transformata falkowa, maszyna wektorów nośnych, optymalizacja rojem cząstek*

1. Introduction

System prognosis is a key factor within the condition-based maintenance (CBM) strategy and has been highlighted in different fields of science (Widodo and Yang [54]; Sutharssan *et al.* [47]). In this context, Remaining Useful Life (RUL) is a rather common measure used to characterize equipment performance (Sikorska, Hodkiewicz and Ma [46]). According to Si *et al.* [42], RUL is the useful life left at a particular time of operation, and is typically random and unknown. In fact, RUL is related with several factors (e.g. current degradation state, operating environment, system function) and should be estimated from available sources of information such as condition and health monitoring sensors. Even though there is no universally accepted best model to estimate RUL (Liao and Köttig [25]), current promising statistical methods have dealt with real-time big data (Bousdekis *et al.* [5]).

In fact, different signals can be collected in order to track the degradation of a system, and then build an accurate relationship between the current health condition state and RUL. Many signals (e.g. vibration, acoustic emission, temperature) can represent the evolution of degradation, and their analyses are as necessary as arduous (Chang *et al.* [7]; El-Thalji and Jantunen [13]; Ambhore *et al.* [2]). In this context, rotating equipment has received special attention due to its critical operating regimes, frequent failure modes and availability of measurements (e.g. vibration), allowing detection and isolation of incipient failures (Vachtsevanos *et al.* [54]).

Support Vector Machines (SVM) have been a successful technique for RUL estimation once it can deal with relative multi-dimensional datasets (Liu *et al.* [28]). Several SVM-based methods (Soualhi, Medjaher and Zerhouni [44]; Saha, Goebel and Christophersen [41]; Patil *et al.* [34]) have been proposed to predict RUL, taking into account that hybrid methodologies usually improve estimation accuracy and overcome limitations of individual methods (Souto Maior *et al.* [48]).

However, SVM learning performance strongly depends on the quality of the input data. In fact, the direct use of the original series as input variables could consider irrelevant information (e.g. noise) and/or miss important features, which may generate imprecise predictions. Hence, specific techniques can be used as pre-processing tools in order to improve data input quality, and then obtain superior predictions from the learning method.

A notable pre-processing technique is Empirical Mode Decomposition (EMD), which decomposes the original series into a sum of Intrinsic Mode Functions (IMFs). According to Huang *et al.* [21], EMD is adaptive, empirical, direct and intuitive. Other specific pre-processing approach based on time-frequency analysis is Wavelet Transforms (WT). The idea behind WT is the same for the short-time fourier transform (Allen [1]), concentrating analysis on frequency filters. However, WT presents the best frequency/time resolution trade-off once it applies windows (filters) of various lengths.

Hence, this work proposes analyzing the ability of EMD-based models and WT-based models to correctly predict RUL when coupled with optimized-SVM. We compared and evaluated the prediction performance when applying both pre-processing techniques as well as predictions obtained without them. The big database considered was provided by FEMTO-ST Institute for the IEEE PHM 2012 Data Challenge focused on the estimation of the RUL for bearings (Nectoux *et al.* [35]) from vibration data.

The remainder of this article is organized as follows: Section 2 presents concepts and a theoretical background about rolling bearing and vibration signals, EMD, WT and SVM. Section 3 describes the methodology and steps adopted on the creation of models to estimate RUL of bearings. Section 4 presents the vibration big database and two cases in which the methodology was applied as well as the results of this application. Section 5 concludes remarks.

2. Theoretical background

2.1. Rolling Bearings and Vibration Signal

Rolling bearings are critical components of rotating machines and its fault diagnosis has subject of extensive research (Rai and Upadhyay [37]; Nikolaou and Antoniadis [33]). Generally, the main component considered on the analysis of localized defects in rolling bearings are the outer race, inner race, ball and cage (Prabhakar, Mohanty and Sekhar [38]).

Regarding to monitoring information (i.e. data), signals are broadly classified depending on the its specific type: vibration and acoustic, temperature and wear debris analysis (Tandon and Choudhury [52]). Particularly, vibration signals are a remarkable indicators for determining failure modes because they are easy-to-measure and provides adequate information, being commonly used in the condition monitoring and diagnosis of the rotating machinery (Chang *et al.* [7]; McKee *et al.* [30]). In this context, several standard vibration-based measures are commonly used for diagnosis purposes, including entropy, root mean square, signal amplitude, variance, kurtosis, as well as higher order statistics (Lybeck, Marble and Morton [29]).

In a fault state, vibration signals presents different pattern from healthy state, which allows failure identification (Chang *et al.* [7]). Indeed, localized faults in rolling bearing components produce a series of broadband impulse responses in the acceleration signals. Each component of the bearing rolling (e.g. outer and inner race; ball) has its own rotation frequency and wave behavior, which leads to a composed and complex signal (Randall and Antoni [41]) as depicted in Figure 1. Pre-processing techniques (e.g. EMD and WT) represents an alternative to deal with complex series creating a more manageable data, yet still carrying the important information.

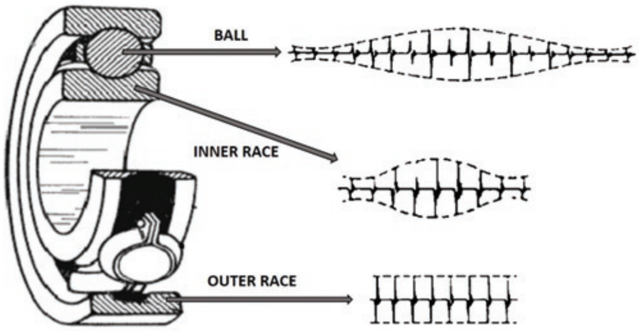

Fig. 1. Signals from local faults in rolling element bearings. Adapted from Randall and Antoni [38]

2.2. Empirical Mode Decomposition

A robust method to analyze non-linear and non-stationary data, Empirical Mode Decomposition (EMD) was developed by Huang *et al.* [19] and have been used in many types of applications. Its main idea is that a data series could be decomposed into a small number of simpler oscillation functions, called Intrinsic Mode Functions (IMFs). Then, the objective is to obtain IMFs regarding data characteristics in time scale (Huang and Wu [21]). Figure 2 depicts a general example of EMD decomposition (6 IMFs and a residue).

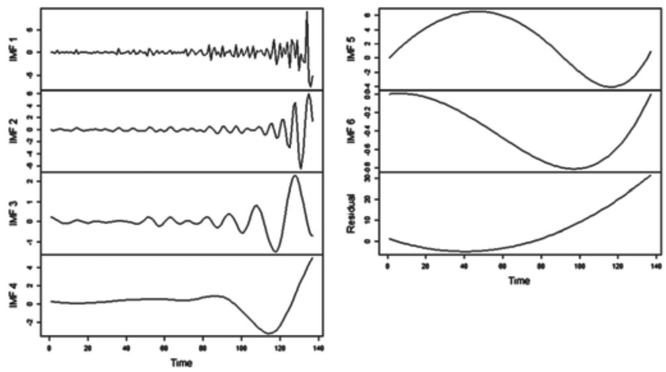

Fig. 2. General decomposition presented by EMD

Generally, any complex signal can be possibly separated into a small number of IMFs and a trend (or residue) r, indexed on $t \in T$, where T is the time interval (set of moments) considered. For a number N of IMFs, the original series $x(t)$ is expressed as follows:

$$x(t) = \sum_{i=1}^{N} \mathrm{IMF}_i(t) + r(t) \tag{1}$$

Huang *et al.* [19] defines IMF as a function that satisfies two conditions: (1) in the whole data set, the number of extrema and zero crossings must either equal or differ at most by one; and (2) at any point, the mean value of the envelop defined by the local maxima and the envelope defined by the local minima is zero. Then, EMD empirically identifies the IMFs through a process called sifting, which is based on three 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 data has not extrema, but only contains inflection points, then it can be differentiated once or more times to reveal the extrema. The sifting goal is to remove riding waves to make the wave profile more symmetric. The sifting process can be described in the following steps (see. Figure 3):

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 envelopes, $e_u(t), e_l(t)$, respectively;
3. Calculate the envelope mean $m(t) = \frac{e_u(t) + e_l(t)}{2}$
4. Obtain $h(t) = x(t) - m(t)$, which is candidate to be IMF;
5. Verify if $h(t)$ satisfies conditions defining an IMF. If it satisfies, an IMF was generated and the new series $x(t) - h(t)$ replaces the initial series $x(t)$. Otherwise, $h(t)$ would be processed again in step 1.

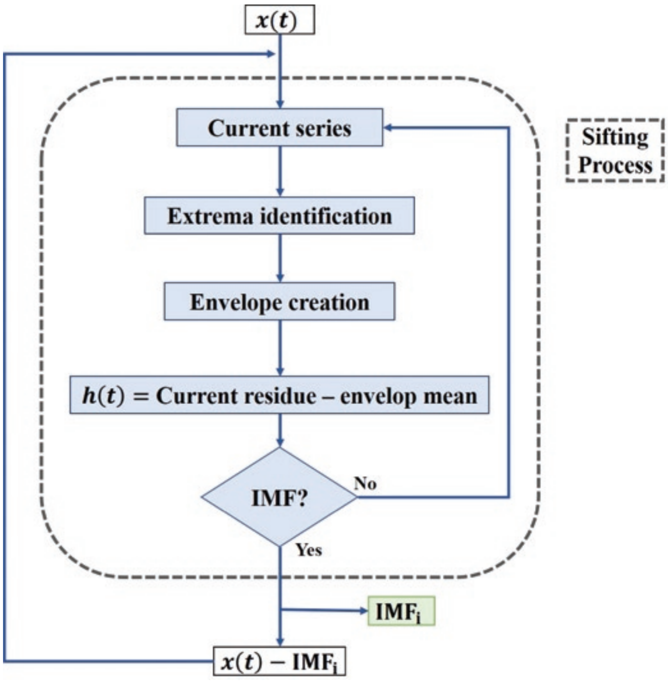

Fig. 3. Sifting process in EMD. Adapted from Souto Maior et al. [45]

At the end of the sifting process, a number of IMFs are generated as well as a final residue $r(t)$. The number of IMFs may vary depending on the intrinsic characteristics of $x(t)$. If the sifting process is carried to an extreme, the candidate IMF could have no physical meaning in sense of both amplitude and frequency modulations. Thus, a stop criterion for the sifting process has to be determined, which can be accomplished by limiting the standard deviation value computed from two consecutive sifting and/or the number of sifting iterations, as originally proposed by Huang *et al.* [19] and still in use (Eftekhar, Toumazou and Drakakis [13]). In practice, the number of IMFs created is lower than 10.

Generally, $\text{IMF}_1(t)$ should contain the finest scale or the shortest period component of the signal. Since the reminder signal $r_1(t)$, i.e. $x(t) - \text{IMF}_1(t)$, still contains information of longer periods (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 an all the subsequent iterations (Equation 2):

$$r_1(t) - IMF_2(t) = r_2(t), \cdots, r_{N-1}(t) - IMF_n(t) = r_N(t) \tag{2}$$

Finally, the original series $x(t)$ is represented as a sum of a number N of IMFs(t) and a residue $r_N(t)$, as presented in Equation (1).

2.3. Wavelet transform

Wavelet Transforms (WT) was first proposed by Morlet *et al.* [31] and has been a widespread technique applied in the field of signal analysis. WT is a mathematical tool that converts a signal of time domain using a wavelet basis function (i.e. a series of wavelet coefficients in time-scale domain) into a different form (Mallat [28]; Yan, Gao and Chen [57]). Kumar and Foufoula-Georgiou [22] remarks that a WT is chosen so that it has two important properties: admissibility (i.e. zero mean) and regularity (i.e. sufficient fast decay, to obtain localization) conditions.

The representation of the transform process occurs by an infinite series expansion of dilated/contracted and translated versions of a mother wavelet, each multiplied by an appropriate coefficient. Hence, the same signal could be represented in different forms, allowing multiple analysis. In practical applications, it is possible to use different well-known WT for distinct purposes and its choice depends on the specific signal characteristics. Figure 4 depicts a general signal processed by WT.

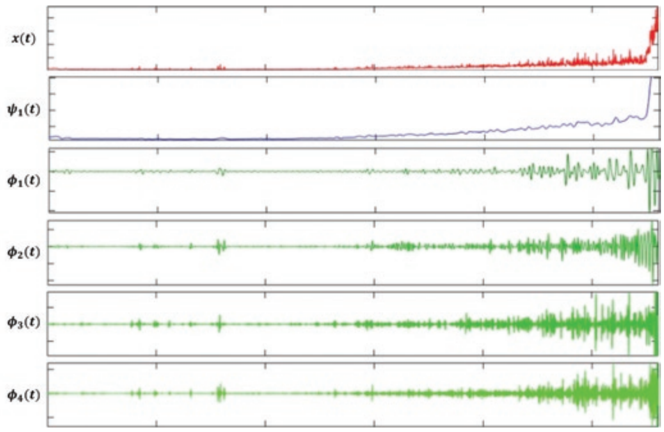

Fig. 4. General decomposition presented by WT

In time domain, a general wavelet dictionary $\{\psi_{u,s}\}$ can be defined as the dilated with the parameter $s > 0$, and translated by $u \in R$ of the mother wavelet ψ as follows(Chen *et al.* [8]):

$$\psi_{s,u}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) \tag{3}$$

Hence, the WT of a function $x(t)$ is calculated by:

$$W(u,s) = \int_{-\infty}^{\infty} x(t) \psi_{s,u}(t) dt \tag{4}$$

Guohua *et al.* [17] argues that wavelet analysis decomposes a signal into two parts, called approximations and details, in which the former consists of high scale low frequency components and offers general information, while the latter corresponds to the low scale high frequency portions and provides detailed hidden patterns.

Daubechies [10], along with Mallat [30], popularized WT, allowing more liberty in the choice of the basis wavelet functions at a little expense of some redundancy, and is credited with the development of the wavelet from continuous to discrete signal analysis. Considering Equation (3), if s represents a continuous variable, then $W(u,s)$ is the continuous WT of $x(t)$ while if $s = a^i$, a is the scale parameter, then $W(u,s)$ is the discrete WT of $x(t)$ (Chen *et al.* [8]). Daubechies wavelets basis relies on the scaling function $\phi(t)$, with set of (filter) coefficients $\{a_k\}_{k\in Z}$, and wavelets function $\psi(t)$, with set of (filter) coefficients $\{b_k\}_{k\in Z}$, satisfying the following refinement (Bakhoday-Paskyabi, Valinejad and Azodi [3]):

$$\phi_k(t) = \sqrt{2}\sum_k a_k \phi(2t-k) \tag{5}$$

$$\psi_k(t) = \sqrt{2}\sum_k b_k \phi(2t-k) \tag{6}$$

Hence, in WT decomposition, the discrete series $x(t)$ of M points is decomposed in distinct levels (e.g. j layer) of $\phi_k(t)$ and $\psi_{j,k}(t)$, each one related with a specific time-frequency characteristic, was follows (Chun-Lin [10]):

$$x(t) = \frac{1}{\sqrt{M}}\sum_k W(j_0,k)\phi_k(t) + \frac{1}{\sqrt{M}}\sum_j\sum_k W(j,k)\psi_{j,k}(t) \tag{7}$$

For discrete WT, Daubechies wavelets were used in this work due to its successful and acknowledged applications (Rafiee, Rafiee and Tse [36]; Genovese *et al.* [16]).

2.4. Support vector machine and particle swarm optimization

Support Vector Machine (SVM) is a supervised learning method which aims at create an mapping function between an input vector $\boldsymbol{x}$ and an output scalar y based on the training data set $D = \{(\boldsymbol{x}_1, y_1), \ldots (\boldsymbol{x_m}, y_m)\}$ (Wang [56]). The objective is to find the function $f(\boldsymbol{x})$ with the smallest penalization with respect to the deviation from the real data and, at the same time, as flat as possible. Depending on the nature of output y (i.e. whether binary/categorical or real numbers), SVM assess different learning problems: (i) classification problem, when dealing with categorical classes (e.g. heath state of a machinery); and (ii) regression problem, when dealing with quantitative and real-valued parameters (e.g. RUL estimation) (Lins *et al.* [26]).

SVM is based on the principle of the Structural Risk Minimization and its concepts are built on the Statistical Learning Theory (Vapnik [55]). This means to solve a convex and quadratic optimization problem in which the Karush-Kuhn-Tucker (KKT) condition are necessary and sufficient conditions to guarantee a global optimum. The goal is not to look for the perfect alignment between the function $f(\boldsymbol{x})$ and D, but the best representation for the mapping (i.e. a trade-off between the data fitness and the generalization ability to predict new data). The regression hyperplane equation is represented by:

$$f(\boldsymbol{x}) = \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + b \tag{8}$$

with $\boldsymbol{x}$ expressing the input data, and $\boldsymbol{w}^{\boldsymbol{T}}$ and b the coefficients to be estimated minimizing the following regularized risk function:

$$\min_{\omega,b} C\frac{1}{m}\sum_{i=0}^{m}\psi_\varepsilon(y_i, f_i) + \frac{1}{2}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w} \tag{9}$$

in which:

$$\psi_\varepsilon(y_i, f_i) = \begin{cases} |y_i - f_i| - \varepsilon & if \quad |y_i - f_i| \geq \varepsilon \\ 0 & otherwise \end{cases} \tag{10}$$

where y_i is the i-th real output (i.e. the original data) while f_i is the i-th estimated value. Equation (10) is known as the Vapnik's ε-insensitive loss function, which implies a non-penalization when the points are inside a tube with radius ε. Hence, ε measures the performance in the training process related to the first term of Equation (9). The second term of the same equation is used as a smoothness function of $f(\boldsymbol{x})$ and is related to the machine's capacity of generalization represented by $\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}$. Yet, C is a trade-off for penalization between the empirical risk and the model's smoothness.

In addition, the problem could be formulated using the primal-dual relation, which states that the solution from dual problem is also solution for the primal one. In practice, the dual problem is the one actually solved and, from the KKT conditions, a global solution is achieved. For more information related with the primal-dual problem, see Wright [55]. Hence, $f(\boldsymbol{x})$ is obtained in terms of the dual problem from Equation (8) as follows:

$$f(\boldsymbol{x}, \alpha, \alpha^*) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\boldsymbol{x}_i^T\boldsymbol{x} + b \tag{11}$$

where α_i and α_i^* are the dual Lagrange multipliers. To solve the linear regression, it is necessary to calculate the dot products, $\boldsymbol{x}_i^T\boldsymbol{x}$, $i = 1,2,\ldots,l$. The generalization for non-linear regression is possible by using kernel functions, $K(\boldsymbol{x}_i, \boldsymbol{x})$, $i = 1,2,\ldots,l$. Hence, Equation (11) becomes Equation (12):

$$f(\boldsymbol{x}, \alpha, \alpha^*) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(\boldsymbol{x}_i, \boldsymbol{x}) + b \tag{12}$$

We here adopted the gaussian Radial Basis Function (RBF) as the kernel function, which is expressed by $K(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right)$, where γ is also a model parameter. One of several advantages of RBF over others kernel functions is to provide great flexibility requiring just one parameter (Lins *et al.* [26]).

A considerable challenge is to provide the best set of parameters to be used in training step. Therefore, metaheuristics, such as Particle Swarm Optimization (PSO), may lead to satisfactory parameters' values. PSO is a probabilistic optimization heuristic inspired by the social behavior of biological organisms (e.g., birds and fishes) and on the ability of animal groups to work as a whole in order to find some desirable position. This seeking behavior artificially modeled by PSO provides useful results in the quest for solutions of non-linear optimization problems in a real-valued search space (Bratton and Kennedy [6]). PSO-optimized SVM have been successfully applied in reliability problems (Droguett *et al.* [11]; Lins, Moura and Droguett [25]; García Nieto *et al.* [15]) and thus were here adopted for choosing to enhance models performance.

3. Methodology

The methodology proposed in this paper is presented in Figure 5 and was applied to an public bigdata set provided by FEMTO-ST Institute (Nectoux *et al.* [35]). The data was generated in the IEEE PHM 2012 Data Challenge focused on the estimation of the RUL for bearings based on vibration signals. Further details about the data set are exposed in next section.

The big datasets contain large quantity of information and, due to the computational cost and hardware restrictions, the learning model

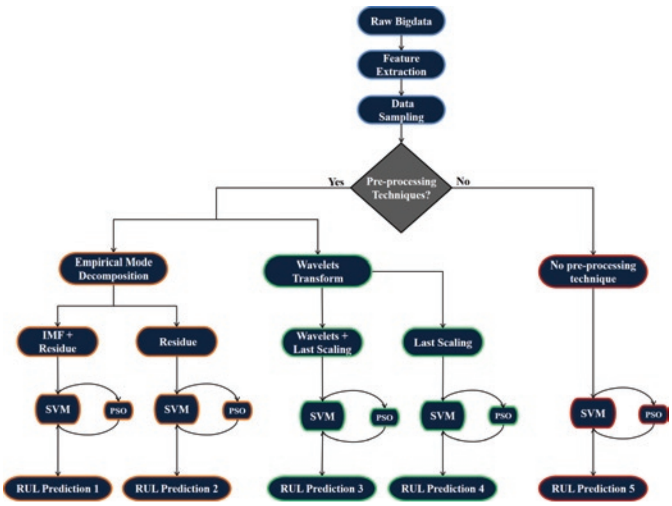

Fig. 5. Methodology applied for RUL prediction

cannot directly handle such an extensive data. Hence, to cover the massive data, two previous steps were performed for data dimension reduction: (i) feature extraction and (ii) data sampling. The former aims to reduce 2,560 vibration signal points into a representative measure (e.g. mean, kurtosis, or the highest absolute value), while the latter consists in sampling from original data (e.g. with frequency rate depending on the degradation state). Indeed, for healthier states of bearing, lower sampling frequency is necessary, while for more degraded states, higher sampling frequency is required. The procedures described above intended to extract only substantial information to be handled by the pre-processing techniques (EMD/WT).

After sampling, EMD or WT was performed. In each case, two distinct regression models were created. For EMD, one model contained all IMFs and the residue, while the other model contained just the final residue. For WT, one model consisted in wavelet functions of each level and the last scaling function, while the other model consisted in just the last scaling function. Given that SVM highly depends on the data input, the idea of using just the final residue (EMD-case) and the last scaling function (WT-case) was to provide an input possibly smooth enough yet still carries valuable aspects of the signal. For EMD, the number of IMFs generated directly depends on the characteristics of signal; a maximum tolerance of 20 sifting was used. In WT, Daubechies function was used as mother wavelet and 4 decomposition levels were applied.

The next step was to input PSO+SVM model with the previous processed data. Then, we evaluated each model (1 – IMFs + Residue; 2 – Residue; 3 – Wavelets + Scaling; 4 – Scaling; 5 – No preprocessing) based on the performance of RUL prediction. This methodology was applied in two different cases. The first application was performed with the complete data set provided. In this case, we considered data until failure (red dotted line in Figure 6), and a regression model was created to estimate the RUL considering each data point. The second application was more challenging once only part of the test set is provided, i.e., there was vibration signals just until some point far from failure time (yellow dotted line in Figure 6). In this case, the goal is to estimate correctly the RUL based on the current behavior of the vibration signals. Further details about both cases are presented in next session.

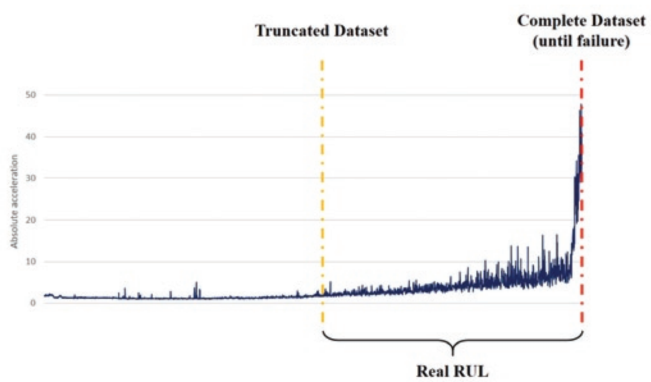

Fig. 6. Different cases considered on the test phase of application example

4. Application example

The presented methodology was applied to a real bigdata set provided by FEMTO-ST Institute (Nectoux *et al.* [35]). Experiments were carried out on a laboratory experimental platform (PRONOSTIA), that enables accelerated degradation of bearings under constant and/or variable operating conditions, while gathering online health monitoring data (e.g. vibration). The main objective is to provide experimental data that characterize the degradation of ball bearings along their complete operational life (until their total failure). Yet, considering the nature of a PHM challenge, data was complex and tricky, which really jeopardize the prediction capacity of the proposed models. The database have become popular during recent years, however many applications only use the complete dataset (Ren *et al.* [39]; Fumeo, Oneto and Anguita [14]) or do not reproduce the design and metrics of the Challenge (Boškoski *et al.* [4]; Mao *et al.* [29]) which is done in this work. For further information, see Nectoux *et al.* [32].

In our applications, we divided the dataset into training and test groups, where the former is necessary to teach SVM about the bearing degradation behavior, while the latter tries to predict correctly the behavior of an unseen bearing. The IEEE PHM Data Challenge provided one set of vibration data from a bearing to be used in the training phase, which is here called as 'Training Bearing', and had 2,803 observations in a run-to-failure experiment. Based on its behavior, estimations for RUL should be performed for another bearing (i.e. the test phase), here named as 'Test Bearing'. Finally, comparison between predictions obtained from models with EMD, WT and without a pre-processing technique are made in order to identify the most suitable approach.

SVM supervised learning method requires both y (i.e. the response variable) and $\boldsymbol{x}$ (i.e. the regressor/input) variables. In all cases, the response variable was the RUL and the regression variables were the vibrations signal. As previously mentioned, in EMD case, two models were created: one considering each IMF and the residue as regressors, and the other considering only the residue as the regressor. In WT case, two other models were also created: one containing each Wavelet and the last Scaling function as regressors, and the other considering only the last scaling function. The last model, without EMD or WT as pre-processing techniques, had the direct signal used as regressor.

In both application cases, it is not expected the direct point prediction to be enough precise due to the high variability of the data. However, the trend of all predictions should express the realistic RUL

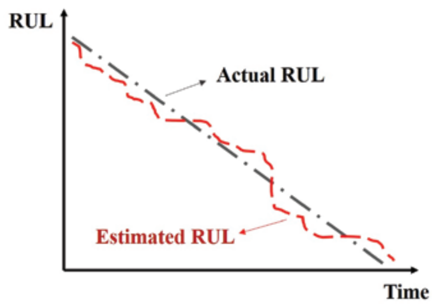

Fig. 7. Expected estimated RUL behavior. Adapted from Sutrisno et al. [48]

estimation, as seen in Figure 7 (Sutrisno *et al*. [51]). Thus, we expect the final residue for EMD and/or the last scaling function for WT to provide interesting results since both have intrinsic attributes related with flatness and physical meaning of the signal trend.

The bigdata set provided for 'Training Bearing' had 2,803 recordings, each one containing 2,560 points for horizontal vibration and other 2,560 points for vertical vibration. Hence, more than 14 million points were provided only for training purposes. In our analysis, a third signal composed by the vectorial sum of the horizontal vibration and vertical vibration was also computed. For each of the three vibration signals (i.e. vertical, horizontal and vectorial sum), three metrics were calculated in the feature extraction step: absolute peak amplitude, kurtosis and entropy. Figure 8 summarizes the data provided and the feature extraction process for the large amount of information.

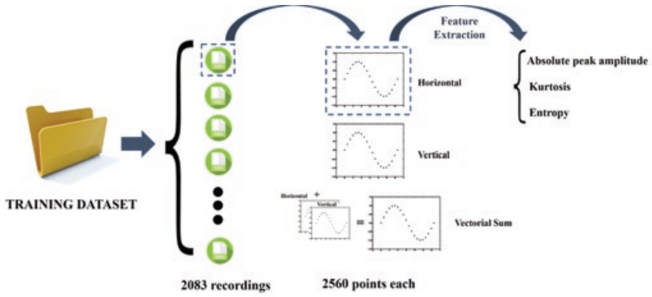

Fig. 8. Feature extraction process considering the dataset

By conducting several investigations and tests, the absolute peak amplitude proved to be the most suitable feature to be used in the next steps, similarly as in other papers (e.g. Rohlmann *et al.* [40]; Chen *et al.* [9]). Specifically, the horizontal vibration signal presented better initial results and was the chosen one to be analyzed. Moreover, dealing with the absolute amplitude, we considered the average of the five highest absolute peak acceleration values measured in each observation (archive). Averaging was done in order to alleviate the effect of data noise(Lee and Yun [24]).

After feature extraction, the data were categorized in four different regions in similar procedure to ISO 10816 (Standardization [49]) that deals with condition monitoring based on vibration. Therefore, each region represents a degradation phase of the bearing. We considered that a change of region takes place when the vibration trend line for the current region suffers a sudden increase of inclination (e.g. a new crack appears). In order to reduce the amount of information, a data sampling was performed in every region with distinct sampling frequency (i.e. the more unstable the bearing is, the more necessary is to monitoring). Table 1 depicts the sampling frequency, the total duration for each degradation region and the number of points actually used after sampling. To illustrate, the third region was sampled every 100 seconds during the 2500 seconds in which the bearing stayed in this region, providing a total of 25 points. The four regions are shown in Figure 9.

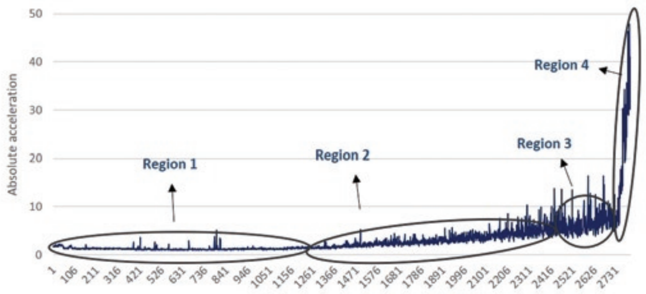

Fig. 9. The four different regions of degradation

Table 1. Sampling frequency and duration for each degradation region

Degradation Region	Sampling Frequency (in seconds)	Total time (in seconds)	Number of points considered
1	400	12000	30
2	200	13000	65
3	100	2500	25
4	0	510	17

4.1. Complete Dataset (until Failure)

In the first case, estimations of RUL for 'Test Bearing' were performed for all data, i.e. every test point until failure has an estimated RUL. As previously mentioned, it is not expected a good punctual prediction, but the trend should correctly express the degradation behavior. Hence, each model provides its own point prediction and a linear trend is created based on these points (i.e. take every predicted value from specific model and calculate the linear regression model representing them). This procedure was performed for the five models under analysis: 1) IMFs + residue; 2) residue; 3) Wavelets + last scaling; 4) Last scaling; and 5) no pre-processing.

In order to measure the quality of the estimated RUL, the Absolute Percentage Error (APE) was calculated, quantifying the distance error from the real RUL to the estimated one. Table 2 presents the APE as well as the number of support vectors related to each model. As it can be seen, EMD-based models (Model 1 and Model 2) presented superior performance compared with WT-based models and the model with no pre-processing technique. Moreover, even the worst EMD-based model was almost three times better than the others. As one might expected, Model 2, which use only the Residue as regressor, clearly presented the best performance.

Table 2. Errors for all tested models

Model	Regressors	Number of support vectors	APE
1	IMFs + Residue	62	2.54%
2	Residue	120	1.45%
3	Wavelets + Last Scaling	126	8.70%
4	Last Scaling	117	15.00%
5	Direct Vibration Data	100	7.58%

4.2. Truncated Dataset (IEEE PHM Data Challenge)

The second case aims to replicate exactly the IEEE PHM 2012 Data Challenge, which presents unclear end-of-life signature and unbalanced dataset (Huang *et al.* [18]). In this case, only truncated data was provided for 'Test Bearing' and the challenge was to estimate the actual RUL based on its initial degradation behavior (see Figure 6). All procedures applied in the first case was done: feature extraction and sampling to reduce the amount of data, training with 'Training Bearing' and test with 'Test Bearing' truncated data (i.e. not data until failure).

Vibration data provided for 'Test Bearing' consisted of 1,802 records and is depict on Figure 10. It is expected the bearing to pass through all four degradations regions, even if the truncated data does not present all of them. By analyzing, 'Test Bearing' does not present several abrupt changes in the signal, seemingly representing only first and second degradation regions. Therefore, inference had to be done to further degradation zones.

Analogously to session 4.1, the first region of 'Test Bearing' lasted 12,000 seconds. Figure 10 shows that vibration from the healthier stage (i.e. first degradation region) is almost stationary, with negligible fluctuations, even though it represents most of the data. In addition, for 'Test Bearing', all remaining points belonged to the second region. Note that there is one outlier near the end of truncated data (even if its value is far from the failure vibration value presented 'Training Bearing' – around 47 units), but this does not correspond to a trend change and it was probably due to a noise in test. Thus, this was still considered belonging to second region. Given that, using data from region 1 will not represent any gain about bearing degradation. Moreover, using data from region 1 will only deviate the overall trend.

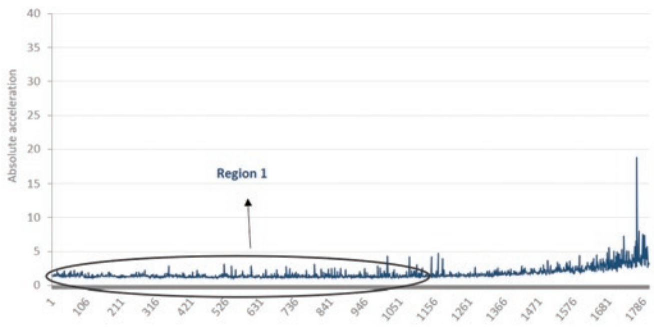

Fig. 10. Truncated test set used in second application

Therefore, in this case, none data from the first region was used. Also, sampling rate of the second region was adapted to provide more information, i.e. frequency of the fourth region was used (every 30 seconds). Thus, data test was reduced to nearly 200 test points used for estimation. Again, the same concept of first case was applied, with the predicted RUL being based on the overall trend of predictions. To define RUL estimation for the challenge, the trend line was extrapolated until it crosses the y-axis (i.e. trend estimate RUL equals to 0). Table 3 presents the models' performance based on the APE of the RUL estimation and real value of RUL. Errors for WT-models are not displayed, once they predicted negative values for RUL, which does not occur in reality.

Table 3. Errors for all tested models

Model	Regressors	APE
1	IMFs + Residue	15.39%
2	Residue	24.90%
5	Direct Vibration Data	58.53%

EMD-based models (Model 1 and 2), which also presented good performance in the first case, actually predicted best results. However, it is important to highlight the difficulty of this challenge, which is evident by considering the magnitude of errors in Table 3. Indeed, we compared our models with the winner of the challenge, which, based on the same evaluation metric, presented prediction errors of 37% (Sutrisno *et al.* [51]). The winner's prediction is worse than estimations provided by two of our models. Moreover, our best EMD-model (Model 1) reduced the error in more than 58%, which confirms the advantage in use the pre-processing method proposed.

5. Concluding remarks

This work compares the use of pre-processing techniques (i.e. EMD and WT) in order to increase prediction performance of RUL in PSO+SVM-based models. The comparison was applied to a real big data set of vibration signals of rolling bearings provided by an IEEE PHM Challenge competition. Two cases were performed: 1) considering the complete dataset (until failure); and 2) considering truncated (replicating the IEEE challenge). Even though performing alone PSO+SVM learning algorithm already provides reasonable estimations, applying pre-processing techniques yields gain in terms of prediction performance

Specifically, EMD based models (Model 1 and Model 2) presented the best performance compared with other approaches for both cases. Moreover, for the second case, two models provided better RUL predictions than the winner of the PHM Challenge competition. For future research, investigations about variations on the Wavelets approach (e.g. number of layers, type of mother wavelet) should be analyzed to improve its poor performance. Moreover, comparison with variants of EMD techniques (e.g. Ensemble Empirical Mode Decomposition (EEMD) (Wu and Huang [59]), Complete Ensemble Empirical Mode Decomposition (CEEMD) (Torres *et al.* [53]) could be done to verify if an even better prediction is achieved.

6. Acknowledgments

The authors thank the Brazilian research funding agencies 'Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)' and 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)' - Finance Code 001 - for the financial support through research grants.

References

1. Allen J. Short term spectral analysis, synthesis, and modification by discrete Fourier transform. IEEE Transactions on Acoustics, Speech and Signal Processing 1977; 25(3): 235-238, https://doi.org/10.1109/TASSP.1977.1162950.
2. Ambhore N, Kamble D, Chinchanikar S, Wayal. V. Tool condition monitoring system: A review. Materials Today: Proceedings 2015; 2(4-5): 3419-3428, https://doi.org/10.1016/j.matpr.2015.07.317.
3. Bakhoday-Paskyabi M, Valinejad A, Azodi H. D. Numerical solution of regularised long ocean waves using periodised scaling functions. Pramana 2019; 92(5): 71, https://doi.org/10.1007/s12043-019-1726-2.
4. Boškoski P, Gasperin M, Petelin D, Juricic D. Bearing fault prognostics using Rényi entropy based features and Gaussian process models, Mechanical Systems and Signal Processing 2015; 52-53: 327-337, https://doi.org/10.1016/j.ymssp.2014.07.011.
5. Bousdekis A, Magoutas B, Apostolou D. Mentzas G.Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance. Journal of Intelligent Manufacturing 2015; 29(6) 1303-1316, https://doi.org/10.1007/s10845-015-1179-5.
6. Bratton D. Kennedy J. Defining a Standard for Particle Swarm Optimization. 2007 IEEE Swarm Intelligence Symposium 2007; 120-127, https://doi.org/10.1109/SIS.2007.368035.
7. Chang L, Chung Y, Lin C, Chen J, Kuo C, Chen S. Mechanical Vibration Fault Detection for Turbine Generator Using Frequency Spectral Data and Machine Learning Model : Feasibility Study of Big Data Analysis. Sensors and Materials 2018; 30(4): 821-832, https://doi.org/10.18494/SAM.2018.1783.

8. Chen J, Li Z, Pan J, Chen G, Zi Y, Yuan J, Chen B, He Z. Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing 2016; 70-71: 1-35, https://doi.org/10.1016/j.ymssp.2015.08.023.
9. Chen X, Ding M, Wang T, Ding M, Wang J, Chen J, Yan J. Analysis and prediction on the cutting process of constrained damping boring bars based on PSO-BP neural network model. Journal of Vibroengineering 2017; 19(2): 878-893, https://doi.org/10.21595/jve.2017.18068.
10. Chun-Lin L. A Tutorial of the Wavelet Transform. Taipei: National Taiwan University, 2010.
11. Daubechies I. Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics 1993; 666-669, https://doi.org/10.1137/1.9781611970104.
12. Droguett E, Lins I, Moura M, Zio E, Jacinto C. 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 2014; 229(4): 319-326, https://doi.org/10.1177/1748006X14533105.
13. Eftekhar A, Toumazou C, Drakakis E. M. Empirical Mode Decomposition: Real-Time Implementation and Applications. Journal of Signal Processing Systems 2013; 73(1): 43-58, https://doi.org/10.1007/s11265-012-0726-y.
14. El-Thalji I, Jantunen E. A summary of fault modelling and predictive health monitoring of rolling element bearings. Mechanical Systems and Signal Processing 2015; 60: 252-272, https://doi.org/10.1016/j.ymssp.2015.02.008.
15. Fumeo E, Oneto L, Anguita D. Condition based maintenance in railway transportation systems based on big data streaming analysis. Procedia Computer Science 2015; 53: 437-446, https://doi.org/10.1016/j.procs.2015.07.321.
16. García Nieto P. J, García-Gonzalo E, Sánchez Lasheras F, Juezc de Cos. 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 2015; 138: 219-231, https://doi.org/10.1016/j.ress.2015.02.001.
17. Genovese L. Videau V, Ospici M, Deutsch T, Goedecker S, Méhaut J. Daubechies wavelets for high performance electronic structure calculations: The BigDFT project. Comptes Rendus Mécanique 2011; 339: 149-164, https://doi.org/10.1016/j.crme.2010.12.003.
18. Guohua G, Yu Z, Guanghuang D, Yongzhong Z. Intelligent Fault Identification Based On Wavelet Packet Energy Analysis and SVM. International Conference on Control, Automation, Robotics and Vision 2006; 1(3): 1-5, https://doi.org/10.1109/ICARCV.2006.345306.
19. Huang B, Jin C, Di Y, Lee J. Review of Data-Driven Prognostics and Health Management Techniques: Lessions Learned From Phm Data Challenge Competitions. Machine Failure Prevention Technology 2017.
20. Huang N. E, Shen Z, Long S, Wu M, Shih H, Zheng Q, Yen N, Tung C, Liu H. 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 1998; 903-995, https://doi.org/10.1098/rspa.1998.0193.
21. Huang N. E, Wu Z. A review on Hilbert-Huang transform: Method and its applications to geophysical studies. Reviews of Geophysics 2008; 46(2): 1-23, https://doi.org/10.1029/2007RG000228.
22. Huang S, Chang J, Huang Q, Chen Y. Monthly streamflow prediction using modified EMD-based support vector machine. Journal of Hydrology 2014; 511: 764-775, https://doi.org/10.1016/j.jhydrol.2014.01.062.
23. Kumar P, Foufoula-Georgiou E. Wavelet analysis for geophysical applications. Reviews of Geophysics 1997; 35(4), https://doi.org/10.1029/97RG00427.
24. Lee J. J, Yun C. B. Damage diagnosis of steel girder bridges using ambient vibration data. Engineering Structures 2006, https://doi.org/10.1016/j.engstruct.2005.10.017.
25. Liao L, Köttig F. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. IEEE Transactions on Reliability 2014, https://doi.org/10.1109/TR.2014.2299152.
26. Lins I, Araujo M, Moura M, Silva M, Droguett E. Prediction of sea surface temperature in the tropical Atlantic by support vector machines. Computational Statistics and Data Analysis 2013; 61: 187-198, https://doi.org/10.1016/j.csda.2012.12.003.
27. Lins I, Moura M, Droguett E. Failure prediction of oil wells by support vector regression with variable selection, hyperparameter tuning and uncertainty analysis. Chemical Engineering Transactions 2013; 33: 817-822.
28. Liu Z, Wang L, Zhang Y, Chen C. A SVM controller for the stable walking of biped robots based on small sample sizes. Applied Soft Computing 2016; 38: 738-753, https://doi.org/10.1016/j.asoc.2015.10.029.
29. Lybeck N, Marble S, Morton B. Validating Prognostic Algorithms: A Case Study Using Comprehensive Bearing Fault Data, Aerospace Conference 2007; 1-9, https://doi.org/10.1109/AERO.2007.352842.
30. Mallat S. A Wavelet Tour of Signal Processing. A Wavelet Tour of Signal Processing 2009.
31. Mallat S. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 1989, https://doi.org/10.1109/34.192463.
32. Mao W, He J, Tang J, Li Y. et al. Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network. Advances in Mechanical Engineering 2018; 10(12), https://doi.org/10.1177/1687814018817184.
33. McKee K. K, Forbes G, Mazhar I, Entwistle R, Hodkiewicz M, Howard I. A vibration cavitation sensitivity parameter based on spectral and statistical methods. Expert Systems with Applications 2015; 42(1): 67-78, https://doi.org/10.1016/j.eswa.2014.07.029.
34. Morlet J, Arens G, Fourgeau E, Giardet D. Wave propagation and sampling theory-Part II: Sampling theory and complex waves. Geophysics 1982; 47(2): 222-236, https://doi.org/10.1190/1.1441329.
35. Nectoux P, Gouriveau R, Medjaher K, Ramasso E, Chebel-Morello B, Zerhouni N, Varnier C. PRONOSTIA : An experimental platform for bearings accelerated degradation tests. IEEE International Conference on Prognostics and Health Management 2012; 1-8.
36. Nikolaou N. G, Antoniadis I. A. Rolling element bearing fault diagnosis using wavelet packets NDT & E International 2002; 35(3): 197-205, https://doi.org/10.1016/S0963-8695(01)00044-5.
37. Patil M. A, Tagade P, Hariharan K, Kolake S, Song T, Yeo T, Doob S. A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation. Applied Energy 2015; 159: 285-297, https://doi.org/10.1016/j.apenergy.2015.08.119.
38. Prabhakar S, Mohanty A. R, Sekhar A. S. Application of discrete wavelet transform for detection of ball bearing race faults. Tribology International 2002, https://doi.org/10.1016/S0301-679X(02)00063-4.
39. Rafiee J, Rafiee M. A, Tse P. W. Application of mother wavelet functions for automatic gear and bearing fault diagnosis. Expert Systems with Applications 2010, https://doi.org/10.1016/j.eswa.2009.12.051.

40. Rai A, Upadhyay S. H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. Tribology International 2016; 289-306, https://doi.org/10.1016/j.triboint.2015.12.037.
41. Randall R. B, Antoni J. Rolling element bearing diagnostics-A tutorial. Mechanical Systems and Signal Processing 2011; 25(2): 485-520, https://doi.org/10.1016/j.ymssp.2010.07.017.
42. Ren L, Sun Y, Cui J, Zhang, L. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. Journal of Manufacturing Systems 2018; 48: 71-77, https://doi.org/10.1016/j.jmsy.2018.04.008.
43. Rohlmann A, Schmidt H, Gast U, Kutzner I, Damm P, Bergmann G. In vivo measurements of the effect of whole body vibration on spinal loads. European Spine Journal 2014, https://doi.org/10.1007/s00586-013-3087-8.
44. Saha B, Goebel K, Christophersen J. Comparison of prognostic algorithms for estimating remaining useful life of batteries. Transactions of the Institute of Measurement and Contro 2009; 31(3-4): 293-308, https://doi.org/10.1177/0142331208092030.
45. Si X. S, Wang W, Hu C, Zhou D. Remaining useful life estimation - A review on the statistical data driven approaches. European Journal of Operational Research 2011; 213(1): 1-14, https://doi.org/10.1016/j.ejor.2010.11.018.
46. Sikorska J. Z, Hodkiewicz M, Ma L. Prognostic modelling options for remaining useful life estimation by industry. Mechanical Systems and Signal Processing 2011; 25: 1803-1836, https://doi.org/10.1016/j.ymssp.2010.11.018.
47. Soualhi A, Medjaher K. Zerhouni N. Bearing health monitoring based on hilbert-huang transform, support vector machine, and regression. IEEE Transactions on Instrumentation and Measurement 2015; 64(1): 52-62, https://doi.org/10.1109/TIM.2014.2330494.
48. Souto Maior C. B, Moura M, Lins L. Droguett, Diniz H. E. Remaining Useful Life Estimation by Empirical Mode Decomposition and Support Vector Machine. IEEE Latin America Transactions 2016; 14(11): 4603-4610, https://doi.org/10.1109/TLA.2016.7795836.
49. Standardization. ISO 10816-7: Mechanical vibration - Evaluation of machine vibration by measurements on non-rotating parts. Part 7: Rotodynamic pumps for industrial applications, including measurements on rotating shafts. Switzerland: ISO. 2009.
50. Sutharssan T, Stoyanov S. Bailey C, Rosunally Y. Prognostics and health monitoring of high power LED. Micromachines 2012; 3: 78-100, https://doi.org/10.3390/mi3010078.
51. Sutrisno E, Oh H, Vasan A, Pecht M. Estimation of remaining useful life of ball bearings using data driven methodologies. 2012 IEEE Conference on Prognostics and Health Management 2012; 2: 1-7, https://doi.org/10.1109/ICPHM.2012.6299548.
52. Tandon N, Choudhury A. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. Tribology International 1999; 32(8): 469-480, https://doi.org/10.1016/S0301-679X(99)00077-8.
53. Torres M. E, Colominas M, Schlotthauer G, Flandrin P. A complete ensemble empirical mode decomposition with adaptive noise. IEEE International Conference on Acoustics, Speech and Signal Processing 2011, https://doi.org/10.1109/ICASSP.2011.5947265.
54. Vachtsevanos G, Lewis F, Roemer M, Hess A. Biqing Wu t al. Intelligent Fault Diagnosis and Prognosis for Engineering Systems. Intelligent Fault Diagnosis and Prognosis for Engineering Systems 2007, https://doi.org/10.1002/9780470117842.
55. Vapnik V. The Nature of Statistical Learning Theory. New York: Springer, 2000, https://doi.org/10.1007/978-1-4757-3264-1.
56. Wang L. Support Vector Machines : Theory and Applications. 2005, https://doi.org/10.1007/b95439.
57. Widodo A, Yang B. S. Machine health prognostics using survival probability and support vector machine. Expert Systems with Applications 2011; 38(7): 8430-8437, https://doi.org/10.1016/j.eswa.2011.01.038.
58. Wright S. J. Primal-Dual Interior-Point Methods. Primal-Dual Interior-Point Methods 2011.
59. Wu Z, Huang N. E. Ensemble Empirical Mode Decomposition: a Noise-Assisted Data Analysis Method. Advances in Adaptive Data Anal 2009; 1-41, https://doi.org/10.1142/S1793536909000047.
60. Yan R, Gao R, X. Chen X. Wavelets for fault diagnosis of rotary machines: A review with applications. Signal Processing 2014, 96(Part A): 1-15, https://doi.org/10.1016/j.sigpro.2013.04.015.

Caio Bezerra Souto MAIOR
Márcio das CHAGAS MOURA
Isis Didier LINS
Center for Risk Analysis and Environmental Modeling – CEERMA
Department of Production Engineering
Universidade Federal de Pernambuco – UFPE
Av. Prof. Moraes Rego
1235 – University City
Recife – PE – Brazil – 50670-901

Emails: caio.maior@ufpe.br, marcio@ceerma.org, isis.lins@ceerma.org