Artigo Completo nº 20170615102340

Prognosis Comparison Between Empirical Mode Decomposition And Wavelets On Support Vector Machine Based Models

Caio B. Souto Maior, Isis D. Lins, Márcio C. Moura.
Center for Risk Analysis and Environmental Modeling, Department of Production Engineering, Federal University of Pernambuco, Recife, Brazil

Enrique L. Droguett
Mechanical Engineering Department, University of Chile, Santiago, Chile

1. Introduction

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, 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.

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 [1]. 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).

According to [2] , Remaining Useful Life (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, 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. CBM 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 [3] .

Different metrics can be obtained in order to track the degradation of a system and build an accurate relationship between the current health condition state and RUL. Many metrics (e.g. vibration, acoustic emission, temperature, corrosion) can represent the evolution of degradation, and their analyses are as necessary as arduous [4, 5].

SVM is a promising algorithm for RUL estimation because it can deal with small training sets and multi-dimensional data [6]. In particular, SVM has been successfully applied to different fields, e.g. financial, environmental, reliability, power systems [7, 8; 9; 10] and is particularly useful when the process or function that maps inputs into output is unknown. 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 [11].

The learning model accuracy strongly depends on the quality of the input data. The direct use of the original series as the input variables in forecasting model could 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 [12], converts the data in a more suitable form by decomposing the original series into a sum of Intrinsic Mode Functions (IMFs). According to [13], EMD is adaptive, empirical, direct and intuitive.

Wavelets Transform are other well known preprocessing techniques based on time-frequency analysis, originally proposed by [14] who introduce filter banks called wavelets function and scaling function. The idea behind Wavalets Transform is the same for the Short-Time Fourier Transform [15], but the former presents the best frequency/time resolution trade-off, given that windows of various lengths are applied.

Due to the challenge to find suitable parameters for SVM, Particle Swarm Optimization (PSO) was used coupled with the learning machine to enhance its predictive capacity.

2. General Objectives

The present work compares EMD-based models and Wavelets-based models coupled with Support Vector Machine and Particle Swarm optimization against a model without any preprocessing techniques and investigates whether the hybrid methodology actually provides significant gain. It is expected that pre-processing techniques could enhance the

3. Theoretical Background

3.1 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. Normally, the main component considered in a rolling bearing are the outer race, the inner race, the ball and the cage. Fault diagnosis of the rolling bearings has been the subject of extensive research [16]. This process includes the acquisition of information, feature extraction and condition recognition [17].

Different methods are used for the acquisition of information and may be broadly classified depending on the type of measurements: vibration and acoustic, temperature and wear debris analysis [18]. 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.

When faults occur in the roller bearing, the vibration signal 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 Fig. 1 shows the typical cases [19]. 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.

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 [20].

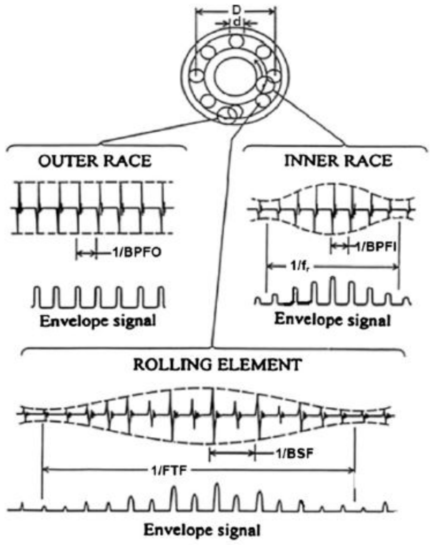

Fig. 1. Signals and envelopes from local faults in rolling element bearings [19]

3.2 Empirical Mode Decomposition

Starting with the work of [12], a remarkable method to analyze non-linear and non-stationary data series was developed and have been used in many types of applications. The main idea is that any data series could be decomposed into a small number of simpler oscillation series, called IMFs. The goal is to obtain IMFs regarding data characteristics in the time scale [21].

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:

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

Ref. [12] defines 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. Then, EMD aims to identify empirically the IMFs from the data series features and decompose it according to its unique characteristics in a process called sifting.

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 [12]. The sifting goal is to remove 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 envelopes, 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.

At the end of the sifting process, a number of IMFs are generated plus a final residue $r(t)$. Those number of IMFs generated may vary once it depends on the intrinsic characteristics of the series $x(t)$. The sifting process should be applied cautiously, once carrying the process to an extreme could cause the IMF to have no physical 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 siftings. In practice, the number of IMFs created is less than 10.

Generally, $c_1(t)$ 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

$$x(t) - c_1(t) = r_1(t) \tag{2}$$

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

$$r_1 - c_2 = r_2, \cdots, r_{N-1} - c_N = r_N \tag{3}$$

Finally, the series could be presented as:

$$x(t) = \sum_{i=1}^{N} c_i + r_N \tag{4}$$

Thus, original series $x(t)$ is decomposed into a number N of IMFs and a residue r_N.

3.3 Wavelets Transform

A widespread technique applied in the field of signal analysis, the wavelet transform was first proposed by [14]. 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. 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.

Wavelet transform can be considered as a mathematical tool that converts a signal in time domain into a different form [22]. The Continuous Wavelet Transform (CWT) of a function $x(t)$ is defined as the integral transform:

$$W_x(\lambda, t, \psi) = \frac{1}{\sqrt{\lambda}} \int_{-\infty}^{\infty} x(u) \bar{\psi}\left(\frac{u-t}{\lambda}\right) du \tag{5}$$

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.

Reference [23] argues 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.

Ingrid Daubechies popularized wavelets with her work in [24], 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. In the discrete wavelet formalism (DWT), the scale λ and the time t are discretized as following:

$$\lambda = \lambda_0^m, \qquad t = n\lambda_0^m t_0 \tag{6}$$

where m and n are integers. The continuous wavelet function $\psi_{\lambda,t}(u)$ in Eq. (5) becomes the discrete wavelets given by:

$$\psi_{\lambda,t}(u) = \lambda_0^{-\frac{m}{2}} \psi(\lambda_0^{-m} u - n t_0) \tag{7}$$

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

$$W_x(m, n, \psi) = \frac{1}{\sqrt{\lambda_0^m}} \int_{-\infty}^{\infty} x(u) \bar{\psi}(\lambda_0^{-m} u - n t_0) du \tag{8}$$

Many discrete functions can be used as mother wavelets, such as Haar Wavelet, Legendre Wavelet and Symlet. A popular wavelet used in DWT is the Daubechies. 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.

3.4 Support Vector Machine and Particle SwarmOptimization

Support Vector Machine (SVM) is a supervised learning method aiming to map input-output from a dataset called training data $D = \{(x_1, y_1), \dots (x_m, y_m)\}$ [25]. 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.

SVM was firstly proposed by Vladimir Vapnik based on the principle of the Structural Risk Minimization and has its concepts built on the Statistical Learning Theory [26]. 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 $\boldsymbol{x}$ 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. 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:

$$f(\boldsymbol{x}) = \boldsymbol{w^T x} + b \tag{9}$$

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:

$$R(C) = C\frac{1}{m}\sum_{i=0}^{m} \psi_{\varepsilon}(y_i, f_i) + \frac{1}{2}\boldsymbol{w^T w} \tag{10}$$

and

$$\psi_{\varepsilon}(y_i, f_i) = \begin{cases} |y_i - f_i| - \varepsilon & if \quad |y_i - f_i| \geq \varepsilon \\ 0 & \text{othewise} \end{cases} \tag{11}$$

where y_i is the variable real value (i. e. the original data) and f_i is the estimated value to the same variable on the same time. Equation (11) 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 they represent slack variables.

Hence, ε measures the performance in the training process and is related to the first term of Equation (11). 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 $\boldsymbol{w}$. Ergo, C is a trade-off between the empirical risk and the model's smoothness, with its value defined *a priori*, as well as the parameter ε. The primal problem is defined as:

$$\min_{\mathbf{w},b,\xi,\xi^*} \frac{1}{2}\, \boldsymbol{w}^T\boldsymbol{w} + C \cdot \sum_{i=o}^{m} (\xi_i + \xi_i^*) \tag{12}$$

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

$$y_i - \boldsymbol{w}^T\boldsymbol{x}_i - b \leq \varepsilon + \xi_i \tag{13}$$

$$\boldsymbol{w}^T\boldsymbol{x}_i + b - y_i \leq \varepsilon + \xi_i^* \tag{14}$$

$$\xi_i, \xi_i^* \geq 0, \qquad i = 1,2, \dots, m \tag{15}$$

Hence, the corresponding primal Lagrangian function could be determined from the Lagrange multipliers. Also, the dual problem can be formulated and, from the KKT conditions, find the global solution, which is also the primal solution. The function $f(x)$ is obtained as follows:

$$f(\boldsymbol{x}) = \boldsymbol{w}_0^T\boldsymbol{x} + b = \sum_{i=1}^{l} (\alpha_i + \alpha_i^*)\, \boldsymbol{x}_i^T\boldsymbol{x} + b \tag{16}$$

where α_i and α_i^* are the Lagrange multipliers. To solve the linear regression, it is necessary to calculate the dot products, $\boldsymbol{x}_i^T\boldsymbol{x}_j$ and $\boldsymbol{x}_i^T\boldsymbol{x}$, and the kernel functions, $K(x_i, x_j)$, could be applied. Hence, it is produced:

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x_o, x_i) + b \tag{17}$$

In this paper, the kernel function adopted is the 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.

In machine learning, a considerable challenge is to provide the best set of parameters to be used in training step, since in principle, those are defined *a priori*. Therefore, optimizations metaheuristics, such as Ant Colony System, Genetic Algorithm and 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 in [27]; [9]; [28].

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 [29].

The fundamental element of PSO is a particle, which can fly all over the search space toward an optimum by using its own information as well as that produced by 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 [30]. The performance of a particle is determined by its fitness.

Mathematically, a particle i is characterized by the following three vectors: (1) its current position in the n-dimensional search space, (2) the best individual position it has held during motion and (3) its velocity. The particles move all over the search space in successive iterations driven by update equations. The PSO search stops when some criteria are reached.

4 METHODOLOGY

The methodology formulated is presented in Fig. 2 and was applied to a real data set provided by FEMTO-ST Institute [31]. The data was used in the IEEE PHM 2012 Data Challenge, focused on the estimation of the RUL of bearings and vibration data from rolling bearings were used as input.

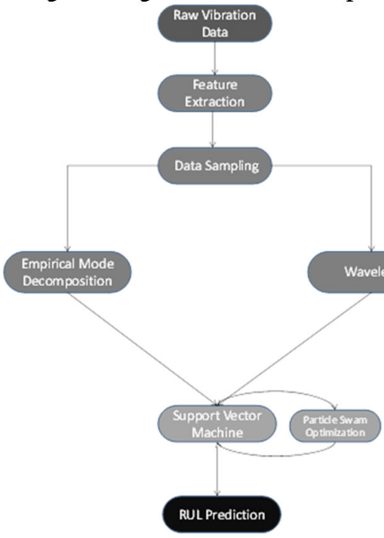

Fig. 2. Methodology applied for RUL prediction

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 (e. g. mean, kurtosis, highest absolute value); 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.

5 APPLICATION EXAMPLE

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 (rotating speed, load force, temperature, vibration). Even if the data provided for the challenge concerns constant operating conditions for each realized experiment, PRONOSTIA enables to provide data related to bearings degraded under varied operating conditions. The main objective is to provide experimental data that characterize the degradation of ball bearings along their whole operational life (until their total failure). For further information, see [31].

For the application, 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 predict correctly the behavior of another bearing that is submitted to the same operating conditions. The IEEE PHM Data Challenge provided a set of vibration data from a training bearing, called in this paper as Bearing 1. Based on its behavior, estimations for the RUL of a test bearing, called Bearing 3, were performed. A comparison among models with EMD, Wavelets Transforms and without any preprocessing technique is made to identify what is the benefit of these methods.

Bearing 1 had 2803 observations in a run-to-failure test. SVM supervised learning method requires 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 vibrations amplitude. In the EMD case, two models were created: 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, two other models were also created: 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.

It is not expected that the direct point prediction to be reliable due to the high variability of the data, but the trend of all predictions should express the realistic RUL estimation, as seen in Fig. 3 [32]:

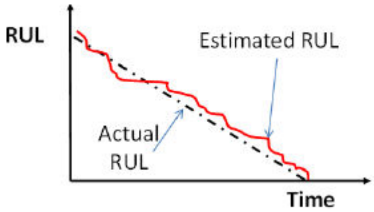

Fig. 3. Expected estimated RUL behavior

The data provided has 2803 discrete recordings and, after the feature extraction, process the data with 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 into four different regions, each one representing one degradation phase of the bearing. In order to reduce the data quantity, a data sampling was performed in every region with a different frequency, once the more stable the bearing is, the less necessary is monitoring. The four regions are shown in Fig. 4.

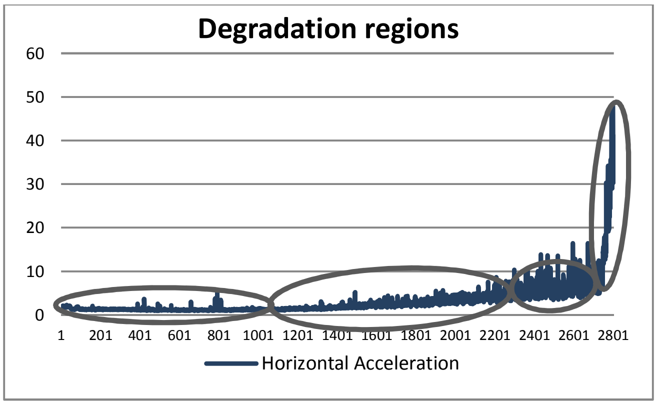

Fig. 4. The four different regions of degradation

Estimations of RUL of test Bearing 3 were performed 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. Fig. 5 depicts an example (e.g. PSO+SVM model with no preprocessing technique) of the applied procedure. Each point has a real RUL and its own prediction. A linear trend is calculated from predictions and it is 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.

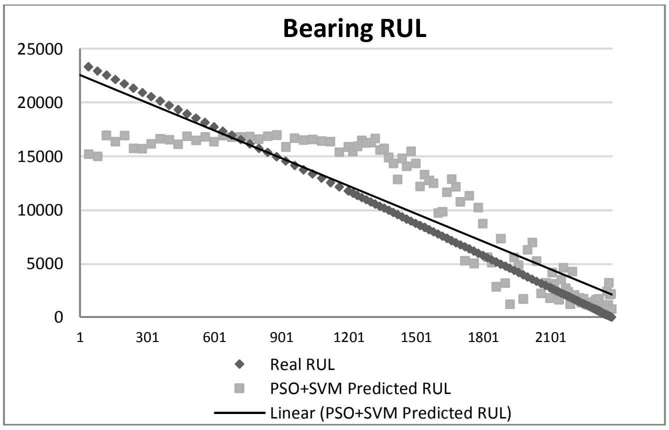

Fig. 5. RUL estimation for complete data case

In order to measure the quality of the estimate RUL, we calculate the Absolute Percentage Error (APE), which quantifies the distance error from the real RUL and the estimated one. To compare the models, we compute the percentage error, dividing the RUL prediction for both models to the real RUL, which, in this case, was 23740 seconds. Both errors are presented in Table 1 for each tested model. As it can be seen from Table 1, even the worst EMD-based model is at least three times better than the others.

TABLE I. ERROS FOR ALL TESTED MODELS

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

6 CONCLUSION AND FUTURE WORKS

This work proposed a comparison between the ability of models using EMD or WT as preprocessing technique to a PSO+SVM learning algorithm to predict the RUL of rolling bearings from a vibration signal. Moreover, the comparison was applied to a real data set 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.

EMD+PSO+SVM based models (which also includes Residue+PSO+SVM model) presented best performance compared with others. Even if, in general, performing PSO+SVM learning algorithm already represents 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, for the presented examples, EMD overcomes Wavelets providing better results.

For future research, a comparison between Ensemble Empirical Mode Decomposition (EEMD) [33] 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.

REFERENCES

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

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

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

[4] El-Thalji I., Jantunen E. "A summary of fault modelling and predictive health monitoring of rolling element bearings". *Mechanical Systems and Signal Processing*, vol. 60, 252–272. (2015)

[5] Ambhore N., Kamble D., Chinchanikar S., Wayal V. "Tool condition monitoring system: A review". *Materials Today: Proceedings*, vol. 2, n. 4–5, 3419–3428. (2015)

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

[7] Kim K. "Financial time series forecasting using support vector machines". *Neurocomputing*, vol. 55, n. 1–2, 307–319. (2003)

[8] Najafi G., Ghobadian B., Moosavian A., Yusaf T., Mamat R., Kettner M., Azmi W.H. "SVM and ANFIS for prediction of performance and exhaust emissions of a SI engine with gasoline-ethanol blended fuels".

Applied Thermal Engineering, vol. 95, 186–203. (2016)

[9] Droguett E.L., Lins I.D., Moura M.C., Zio E., Jacinto C.M. "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*, vol. 229, n. 4, 319–326. (2014)

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

[11] Souto Maior C.B., Moura M. das C., Lins I.D., Droguett E.L., Diniz H.H.L. "Remaining Useful Life Estimation by Empirical Mode Decomposition and Support Vector Machine". *IEEE Latin America Transactions*, vol. 14, n. 11, 4603–4610. (2016)

[12] Huang N.E., Shen Z., Long S.R., Wu M.C., Shih H.H., Zheng Q., Yen N.-C., Tung C.C., Liu H.H. "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)

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

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

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

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

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

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

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

[20] Lybeck N., Marble S., Morton B. "Validating Prognostic Algorithms: A Case Study Using Comprehensive Bearing Fault Data". In Aerospace Conference, 2007 IEEE (pp. 1–9). (2007)

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

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

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

[24] Daubechies I. "Ten Lectures on Wavelets". *SIAM Review*. (1993)

[25] Wang L. "Support Vector Machines : Theory and Applications". (2005)

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

[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*, vol. 33, 817–822. (2013)

[28] García Nieto P.J., García-Gonzalo E., Sánchez Lasheras F., De Cos Juez F.J. "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*, vol. 138, 219–231. (2015)

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

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

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

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

[33] Wu Z., Huang N.E. "Ensemble Empirical Mode Decomposition: a Noise-Assisted Data Analysis Method". *Advances in Adaptive Data Anal*, vol. 2009, n. 1, 1–41. (2009)