Bearing failure prognostics based on wavelets, support vector machines and bootstrap

Monalisa Cristina Moura dos Santos[a]
monalisamoura24@gmail.com
Hiago Henrique Gomes de Araujo[a]
hiago.araujo.00@gmail.com
Isis Didier Lins[a]
isis.lins@ceerma.org[a]
Caio Bezerra Souto Maior[a]
caiomaior@hotmail.com
Márcio das Chagas Moura[a]
marcio@ceerma.org
Enrique López Droguett[b]
elopezdroguett@ing.uchile.cl

[a]Centro de Estudos e Ensaios em Risco e Modelagem Ambiental – Departamento de Engenharia de Produção – Universidade Federal de Pernambuco
Av. da Arquitetura, s/n - Cidade Universitária, Recife - PE, 50740-550

[b]Departamento de Engenharia Mecânica – Universidade do Chile
Av. Beauchef 851, Santiago, Chile

ABSTRACT

The main objective of this paper is to provide point and interval estimates of the remaining useful life (RUL) of rolling bearings. The proposed methodology involves wavelets, support vector machines (SVM) and bootstrap methods and its input data are vibration measurements given by the IEEE 2012 Data Challenge. First of all, the data is processed using Wavelets so as to remove noise and then SVM and the Maximum Entropy Bootstrap (MEB) are used to provide point and interval predictions of the RUL. A comparison between Wavelets+SVM+MEB and SVM+MEB is performed and superior results are obtained using Wavelets+SVM+MEB.

1. INTRODUCTION

Over the last decades, many prognostics and diagnostics techniques were introduced and, with those strategies, were developed many philosophies, which can be applied to equipment maintenance. One of the crucial parts to develop and implement an effective diagnostic and prognostic program is based on the ability to anticipate a fault and to generate useful information. The prognostics is done through data collected on equipment, and, then, the RUL estimate gives significant information about the future conditions of the equipment.

The condition-based maintenance (CBM) uses operating equipment data to determine its condition and to plan repair and maintenance before its total failure. The prognostics and health management (PHM) incorporates the forecast of the future behavior of the equipment –e.g., the estimated RUL. The detection of faults or of an incipient fault is crucial to a system and the RUL forecast is necessary to various CBM/PHM studies [1].

Rolling bearings are widely used in diverse types of machine, since a simple fan or complex manufacturing equipment. Rolling bearings' failure are the most common cause of system breakdown.

However, development of diagnosis and prognosis methodologies are crucial to previously detect a failure at the rolling bearing, as well as equipment malfunction or degradation.

The behavior of a rolling bearing is mostly shown by data such as vibration signal, which is the one used in this paper. The vibration data come as a raw data and, to improve the model, it needs to be processed to eliminate noise. Many techniques of data processing are used to treat non-stationary times series, such as EMD (Empirical Mode Decomposition) [2] and wavelets [3]. The latter is well suited to handle non-stationary data and is used in this work to eliminate the unnecessary features of the data.

Given that the processed vibration signal remains non-stationary, the Maximum Entropy Bootstrap (MEB) is applied, this method was developed by Vinod (2004)[4]. There are other bootstrap methods devised to handle time series such as stationary bootstrap (POLITIS & ROMANO, 1994) [5] but they are suited to stationary data. MEB replicates the processed vibration data to generate a number of bootstrap samples. Each of these samples is used to estimate a regression model using Support Vector Machine (SVM). For each of these samples, SVM provides RUL point estimates. The actual RUL point estimate is extracted from a linear regression model based on the RUL predictions given by SVM. The various RUL point estimates are then used to construct RUL confidence intervals using the percentile method (EFRON e TIBISHINARI, 1993) [6].

Firstly, this paper describes the techniques used to processing the raw data, then it explains the RUL is forecasted under the PSO+SVM methodology. As a last step of the theoretical background, the Bootstrap and the maximum entropy bootstrap methods were shown, which the train set is replicated and, after, the confidence intervals are calculated and the RUL is punctual predicted. The next step of this paper is to explain the methodology that was chosen to predict the RUL and, then, show the results with the methodology validation and the conclusions of those predictions.

2. THEORETICAL BACKGROUND

2.1 WAVELET TRANSFORM

Wavelets has applicability on non-stationary data that requires a previous treatment to remove noise. They can be classified in two types: the continuous and the discrete. The processed signal can be divided in two groups: the approximation coefficients – components with high scale and low frequency (the time series used to apply at the prognostics methodology after the processing) - and the details coefficients – components with low scale and high frequency. Each coefficient is calculated from the pyramidal algorithm shown in Figure 1 [7]. These coefficients compose the transformation with the Wavelet Mother (Equation 3) and the Wavelet Father (ϕ) (Equation 10) [8].

$$f(t) = \sum_{k} c_{j_0,k}\phi_{j_0,k} + \sum_{j \geq j_0}\sum_{k} d_{j,k}\psi_{j,k} \quad (1)$$

$$\phi(x) = \sqrt{2}\sum_{k} l_k \phi(2t-k) \quad (2)$$

$$\psi(x) = \sqrt{2}\sum_{k} h_k \phi(2t-k) \quad (3)$$

j: Resolution level
k: Translation factor
n: Maximum
resolution level

t: Time

To build the Wavelets, it is necessary to choose a filter base to be used to analyze the data. Among the discrete Wavelet bases, the one chosen was the Daubechies, since it has successful applications in discrete data processing [3]. The Daubechies base, which is calculated in an iterative way, consists in a family of orthogonal Wavelets represented by filters of size N = 2p, in which p is the number of vanishing moments- vanishing moments are the zero-valued elements placed to start the algorithm in order to allow that the number of output coefficients is the desired one - and the output coefficients are located between [-p+1,p] [7].

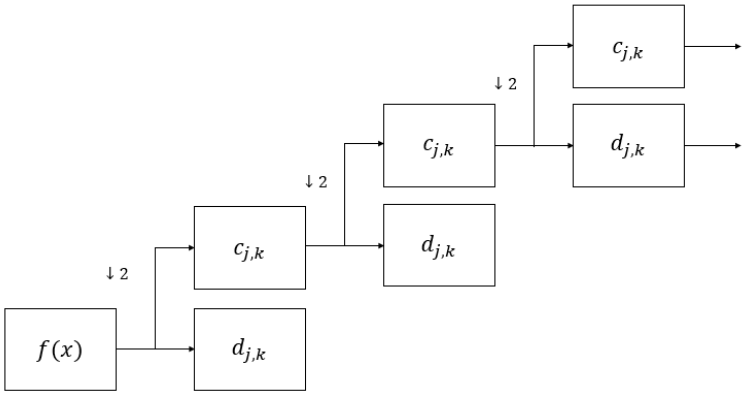

Figure 1 - Pyramidal Algorithm.

2.2 SUPPORT VECTOR MACHINES

SVM, also known as kernel machines, is an algorithm of statistical learning that can be used for classification or for the construction of regression models. SVM is based on the statistical learning theory and on the Structural Risk Minimization principle [9]. The non-parametric property is one of the main advantages of the SVM, that means it does not need to assume any hypothesis over the dataset distribution to apply this methodology.

2.2.1 SUPPORT VECTOR REGRESSION

The objective of a regression model is to find the functional dependence of a variable y over an independent vector of variables x. The input of Support Vector Regression (SVR) is a dataset (training) with the values of the dependent and independent variables. After processing these data, one can find an approach that represents the relation between y and x and calculate the y values based on new x. To associate y and x learning algorithm bases itself on the resolution of the following problem (Equation 4-7).

$$min_{w,b,\xi,\xi*} \quad \frac{1}{2}\ ||w^2|| + C.\sum_{l}(\xi_l + \xi_l^*), \qquad l = 1,2,\dots,L \tag{4}$$

$$s.a. \qquad y_l - w^T\phi(x_l) - b \leq\ \varepsilon + \xi_l, \qquad \forall l \tag{5}$$

$$-y_l + w^T\phi(x_l) + b \leq\ \varepsilon + \xi_l^*, \qquad \forall l \tag{6}$$

$$\xi_l, \xi_l^* > 0. \tag{7}$$

The objective function has two parts: the first, indicates the capacity of the SVM, the lower this value is, more generalist is the regression function encountered. The second part represents a cost caused by the errors between the real values and the predicted ones. These points are calculated for each point by and the parameter C represents a trade-off between these two parts of the algorithm.

The restriction represents the superior and inferior limits determined by a function with tube format with length 2 ε, which involves the values of the dependent variable. ξ_l and ξ_l^* are outputs of the optimization problem and represent slack variables. The right choice of a regression model can become really complex as the dimension of x increases; the scalar product may demand a big computational effort. This problem is related to the dimensionality curse phenomenon. The solution to this problem is solved by the introduction of a kernel function, $K(x_i, x_j) = \Phi^T(x_i)\Phi(x_j)$, whose principal advantage is to avoid the direct computation of dot products. The Gaussian Radial Basis Function (RBF) (Equation 8) is the most applied kernel function; one of its advantages is the necessity of only one parameter.

$$K(x_i, x_j) = exp[\frac{(x_i - x_j)^2}{2\sigma^2}]^d \tag{8}$$

For the resolution of the mathematical programming the dual version was utilized. It can be obtained by the Karush-Kuhn-Tucker conditions. The following model is encountered (9-12):

$$max_{\alpha,\alpha^*}\ L_D = -\frac{1}{2}\sum_i \quad \sum_j (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)x_j^T x_j \ - \sum_i (\varepsilon - y_i)\alpha_i - \sum_i (\varepsilon + y_i)\alpha_i^*, i, j = 1,2,\dots,l \tag{9}$$

$$s.a. \quad \sum_i (\alpha_i - \alpha_j^*) = 0 \tag{10}$$

$$0 \le \alpha_i \le C \tag{11}$$

$$0 \le \alpha_i^* \le C \tag{12}$$

α and α^* represent vectors of dimension l composed by the Lagrange multipliers for the problem. These values are encountered by the model resolution and are utilized on the regression model in which the index 0 indicates that the value is part of the optimal solution. The resolution of the dual model leads to the following regression model (equation 13):

$$f(x) = w_0^T \Phi(x) + b_0 = \sum_{i=1}^{l} (\alpha_{0i} - \alpha_{0i}^*) K(x_i, x) + b_0 \tag{13}$$

2.3 BOOTSTRAP

Bootstrap is a replication technique commonly used in statistics. It has big versatility and can be used in many applications, such as confidence interval construction and bias and standard deviation estimates (Figure 2). One big attractiveness of bootstrap methods is referred to its applicability without the necessity of assumptions over the data set distribution.

The resample process can be defined as the extraction of many samples that have similar nature of

the original sample.For a data set x, one will have several samples with a similar behavior to the original, $X^1, X^2, \ldots, X^B$. These samples are used to provide estimates such as bias, standard error and confidence intervals [4].

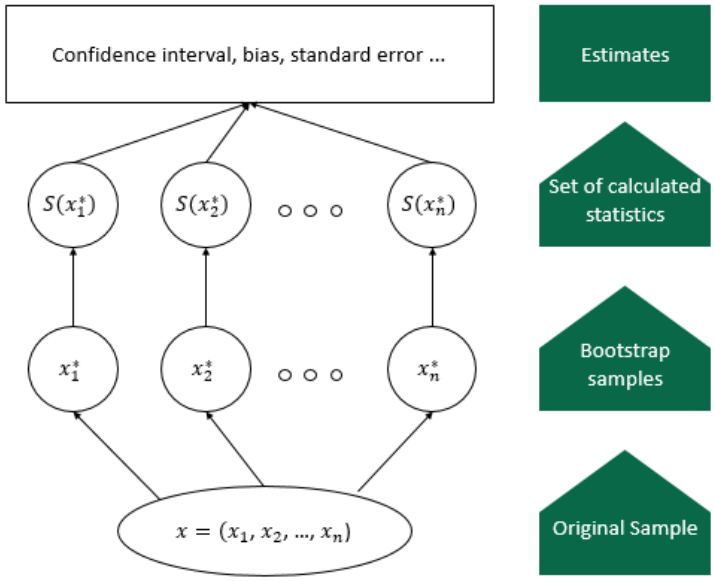

Figure 2 – Bootstrap Scheme.

The most basic bootstrap technique assumes that the original data set is independent and identically distributed (i.i.d) and may be described by the following step-by-step algorithm:

1. Let $X = \{x_1, x_2, \ldots, x_n\}$ be the original dataset which one we want to replicate using the bootstrap
2. Let $I = \{I_1, I_2, I_3, \ldots, I_n\}$ be a set of data distributed as a discrete uniform limited between 0 and n
3. A bootstrap sample is constructed by using I to resample with replacement X
4. The steps 2 to 4 are repeated until we obtain the chosen number of bootstrap samples

2.3.1 MAXIMUM ENTROPY BOOTSTRAP

The basic bootstrap is not an appropriate choice when the studied data set is not i.i.d. So this resample technique is not recommended to the study of time series, however there are many bootstrap methods developed for this purpose. Among these, the Maximum Entropy Bootstrap (ME Bootstrap), developed by Vinod, is applicable to non-stationary time series, therefore this technique is worth to be used on the study of a lot of real world situations that are modeled by non-stationary time series.

The biggest advantage of the ME Bootstrap is the possibility to do a more precise analysis than with hypothesis tests based on asymptotic theory, this technique is available for using with nonstationary time series, therefore it avoids the using of transformations that destroy the series format [4].

Let a set be composed by the population from which the analyzed time series was collected, the ME Bootstrap generates many replicates of this series as element of by a seven steps algorithm, which were developed in a way that the format and time dependence of the autocorrelation function (ACF) of the original time series are maintained at the replicated one [10].

The step-by-step algorithm is described as follow:

1. Sort the original data, obtaining a new time series $x_{(t)}$, and save the position of these values before the transformation

2. Compute the medians between the already sorted series values $z_t = \frac{(x_{(t)} - x_{(t-1)})}{2}, t = 2, 3, \ldots, T-1$;
3. Compute the trimmed mean of the errors $x_{(t)} - x_{(t-1)}$ and compute the values of z_1 and z_T using: $z_1 = x_{(1)} - m_{trm}$ e $z_T = x_{(T)} - m_{trm}$;
4. Construct the Maximum Entropy density function using the values of z as limits of the intervals. The distribution is a composition of uniform distributions with equal values of probabilities (Image 2). This distribution grants that the mean of any interval satisfice the mean preservation constraint, this constraint is developed to grant the hypothesis of the ergodic theorem
5. Simulate, by the inverse transform method, T values of the constructed time serie and sort them in ascending order
6. Sort again the original values using the saved position at the step 1. Therefore, the original format of the time series will be maintained, the values will always be positioned on the replicated series according to its dimensions. The biggest point at the original and replicated time series are positioned at the same location
7. Repeat the steps 2 to 6 so the number of replicates necessary for the analysis are collected

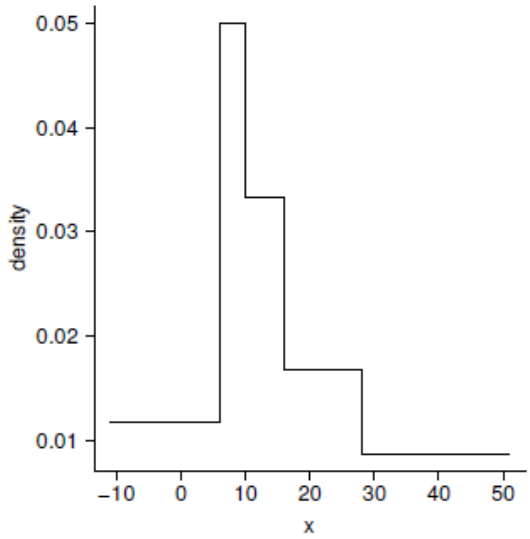

Figure 3 – Maximum entropy function (Source: Vinod e Lopes-de-Lacaqlle, 2009).

3. PROPOSED METHODOLOGY

The construction of confidence intervals of the SVM prediction over the RUL, the bootstrap was applied over the original time series, which was treated by Wavelets. The following step-by-step and the figure 4 describe the applied methodology:

1. Apply the wavelet at the original training time series;
2. Train the SVM algorithm for the original dataset;
3. Apply the encountered regression model to the test dataset;
4. Use linear regression to construct the point estimate of RUL;
5. Use the MEB to replicate the training time series;

6. Repeat the steps from 2 to 4 using the new test dataset

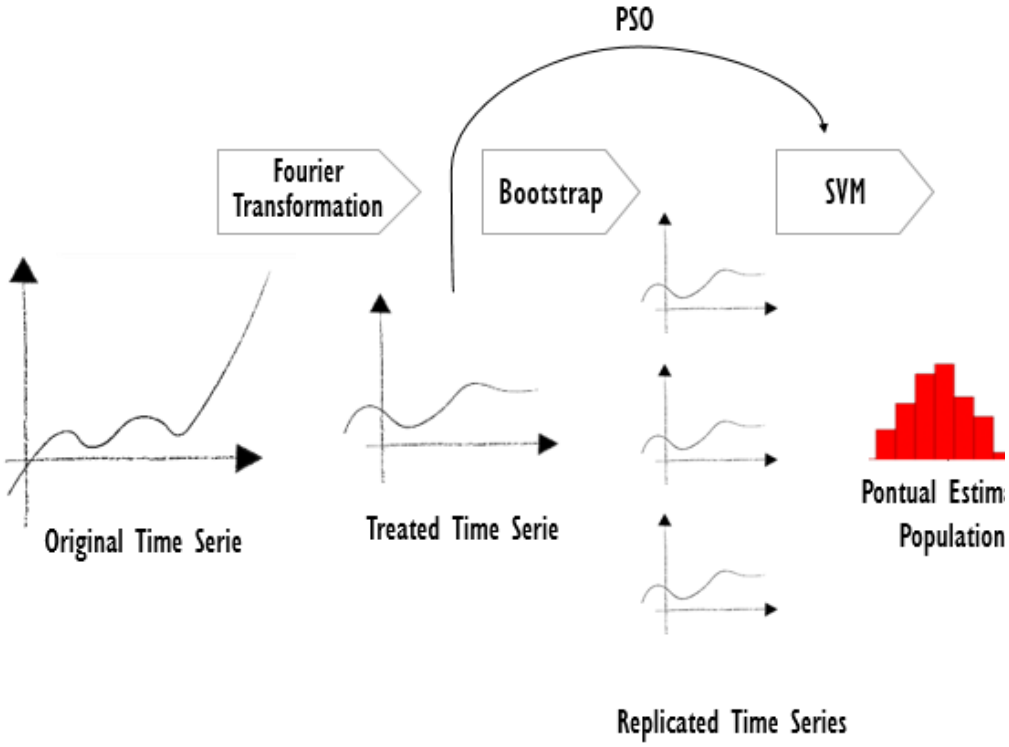

Figure 4 – Proposed methodology.

4. RESULTS

The data used in this paper is from a challenge focused on estimating the RUL of bearings and it was performed by IEEE PHM 2012 Data Challenge (IEEE Reliability Society and FEMTO-ST Institute). The experiment was performed with 1800 rpm and 4000N. The bearing's vibration data were analyzed on MATLAB, the training set was called "Bearing1" – which has the measures of the vibration signal until the complete fault – and the test set "Bearing3" –the one used to forecast the RUL. To process the data the Daubechies filter was chosen with four levels of decomposition for the two signals and then the original signal was compared with the processed one to see if the processing stage caused any improvement on the prediction.

Firstly, the training of the data from Bearing1 happens to compute the equation of regression that will be used on the train stage, those equations will be, at first, replicated to calculate the confidence interval under Maximal Entropy Bootstrap. Bearing1 data are composed by pairs (x_t, y_t) in which x_t is the vibration and the y_t is the corresponding RUL at a time t. To choose the coefficients for the SVR, a Particle Swarm Optimization (PSO) algorithm was used. To do the RUL forecasting through the test data set containing 1802 pairs (x_t, y_t) was used.

However, to obtain a precise estimate of the RUL, the test data is sampled, at first, cutting the initial stage of vibration since in that stage the bearing does not seem to have any changes on it vibration

behavior (because, at the beginning of the rotation, the vibration has less impact at the bearing). So, only the important stages of initial failure and in which the bearing is close to break are considered in the analysis. After, the test data set was sampled in a fixed time interval (300 seconds) until the truncated point. This sampling was chosen to adapt the RUL forecasted by de SVR to the linear regression model, so the point prediction could be done.

The RUL forecasts obtained by the SVR (Figure 5 and 6) were related to each acceleration and their corresponding input parameters shown by the Table 5. Nonetheless, it is necessary compute an RUL point estimate. Hence, a linear regression of the RUL vs. index (which is related to time) is done to predict future behaviors. Because of the linear regression, two point estimates of RUL were found: (i) the x cross value and (ii) the y cross value. They are both used to calculate the absolute percentage error given by the Equation 14:

$$APE = \frac{\frac{(|Real\ RUL - y\ cross\ value| + |Real\ RUL - x\ cross\ value|)}{2}}{Real\ RUL} \quad (14)$$

Table 5 - Values of the parameters used in SVR.

Parameters of the SVR		
Truncated data without processing		
γ	C	ε
84.2407	15.0837	0.0070
Truncated data with processing by wavelets		
γ	C	ε
81.3929	15.1844	0.04874

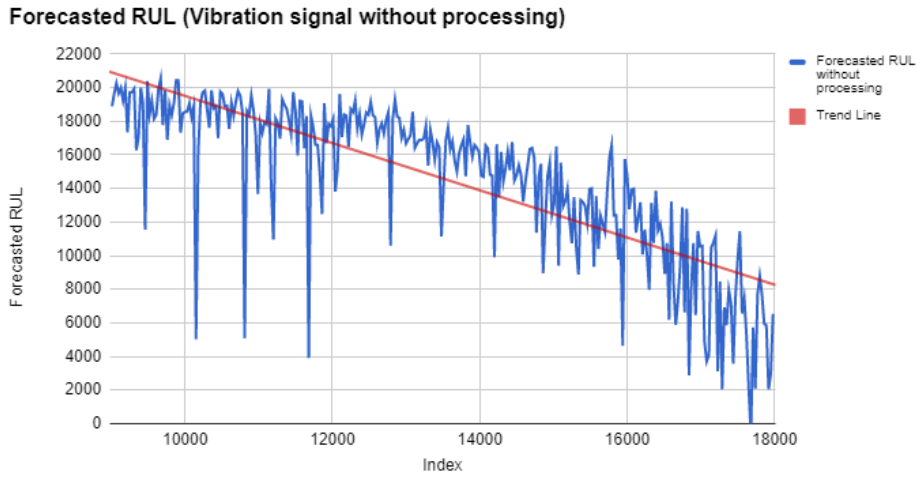

Figure 5 - Forecasted RUL performed under the Bearing3 truncated data without processing and its trend line.

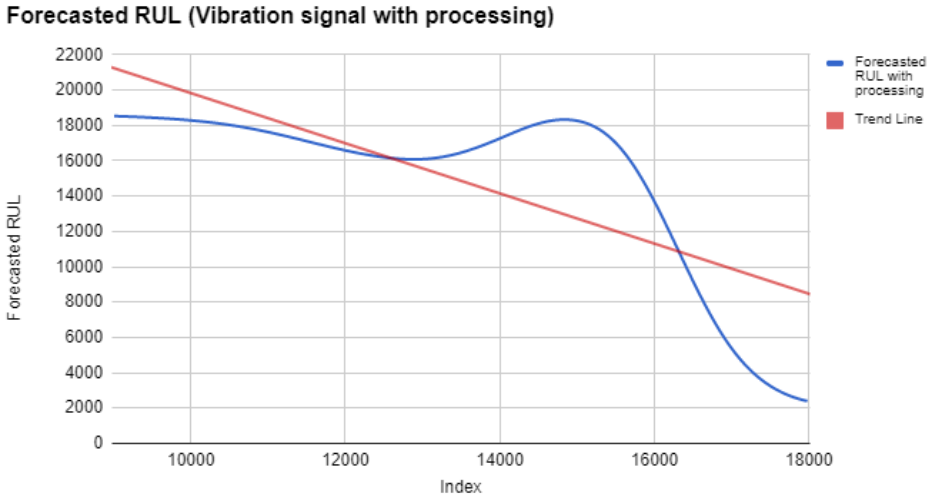

Figure 6- Forecasted RUL performed under the Bearing3 truncated data with processing and its trend line.

The point estimate of RUL, its absolute percentage error and linear model (Table 8):

Table 8 - Forecasted RUL' results.

Forecasted RUL for data without processing	**Forecasted RUL for data with processing**
$RUL_{estimated}$= 4119s	$RUL_{estimated}$= 4569s
Absolute Percentage Error =43,5%	Absolute Percentage Error= 40,7%
$y = -1{,}40x + 28909$	$y = -1{,}45x + 29616$

After the point prediction, the replication of the training dataset using the Maximum Entropy Bootstrap and calculating the point prediction for the same testing data set, the predicted interval was built by applying the percentile method at the set of predictions. The percentile method uses the quantiles of this set and has big popularity due to its simplicity. The number of replicates was 5000, which is greater than the minimum recommended in [11] for the method.

The intervals were made for the predicted RUL with and without wavelet processing (Table 9), and the histograms of the set of calculated point predictions (Figure 7 and 8) indicate information about the distribution of this statistic.

Table 9 – Results of RUL bootstrap confidence intervals

Training Dataset	**Inferior limit of confidence interval**	**Superior limit of confidence interval**	**Distance Between the limits**	**Point prediction**	**Original predicted value**
With Processing	3505.055	4407.769	902.714	4569	5730
Without Processing	2585.494	3631.323	1045.129	4119	5730

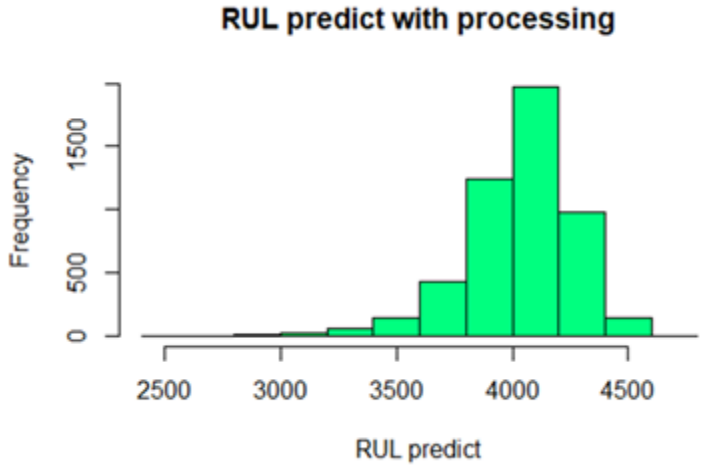

Figure 7 - Set of calculated RUL predicts with the replicated time series with processing.

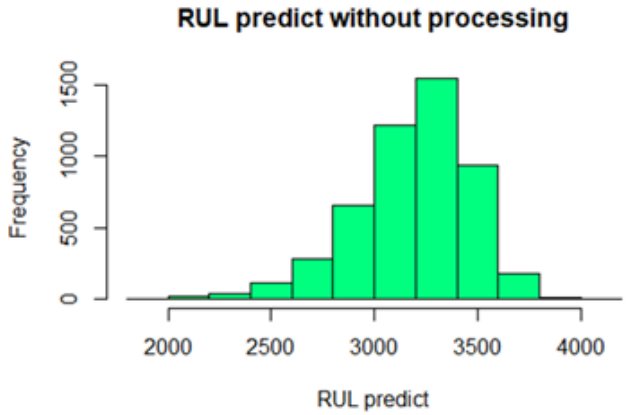

Figure 8 - Set of calculated RUL predicts with the replicated time series without processing.

5. CONCLUSION

Prognostics methodologies are a current topic in development at Condition-Based Maintenance strategies. In this paper, the main discussion was the point and interval prediction of the RUL rolling bearings. The data used is given by the IEEE PHM 2012 Data Challenge, a reliability challenge to foster research of new prognostics methodologies. Vibration signal of a rolling bearing was analyzed under the proposed methodology. As a first step, the raw data was processed using Wavelets – a transform that, in this paper, is applied to filter the noise of the vibration signal and to improve the forecast process. Confidence intervals of the RUL were alse estimated with the Maximal Entropy Bootstrap. This stage consists, basically, to replicate the time series of the training set and then to compute the RUL point prediction by SVM+PSO.

For the data without Wavelets processing, the estimated RUL was 4119s with an absolute percentage error of 43,5% and, for the data with Wavelets processing, the estimated RUL was 4569s with an absolute percentage error of 40,7%. The winner of the challenge estimated an RUL with an error of 37% [12] for the

forecasted RUL using the Bearing3 as the test set. So, as can be seen, the proposed methodology is a promising tool to forecast RUL based on vibration data.

After that, the interval estimates were built using MEB. The computed intervals included neither the original value nor the point prediction for the original data set**.** This fact needs to be further investigated and the procedure to obtain the RUL point estimate should be improved to be coupled with MEB so as to provide informative RUL interval estimates.

REFERENCES

[1] MOURA, M. C.; ZIO, E.; LINS, I. D.; DROGUETT, E. L., *Failure and reliability prediction by support vector machines regression of time series data. Reliability Engineering & System Safety*, 96: 1527-1534, (2011).

[2] HUANG, N. E., SHEN, Z., LONG, S. R., WU, M. C., SHIH, H. H., ZHENG, Q., … 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).

[3] DAUBECHIES, I. *Ten Lectures on Wavelets*. *SIAM Review* (1993).

[4] VINOD, H. D. *Ranking mutual funds using unconventional utility theory and stochastic dominance*. *Journal of Empirical Finance*, v. 11, n. 3, p. 353–377, (2004).

[5] POLITIS, DIMITRIS N.; ROMANO, JOSEPH P. *The stationary bootstrap. Journal of the American Statistical association*, v. 89, n. 428, p. 1303-1313, (1994).

[6] EFRON, B.; TIBSHIRANI, R. J. *An introduction to the bootstrap*. New York: Chapman e all/CRC. 1993.

[7] MORETTIN, P. A., *ONDAS E ONDALETAS. Da Análise de Fourier à Análise de ondaletas de Séries Temporais*. Editora da Universidade de São Paulo, São Paulo, (2014).

[8] PERCIVAL, D.B. & WALDEN, A.T. *Wavelet methods for time series analysis.* Cambridge University Press, New York. (2006).

[9] KECMAN, V. *Support vector machines: an introduction. In Studies in Fuzziness and Soft Computing*, Wang, L. (ed.), v. 177, p. 1-47. Berlin Heidelberg: Springer-Verlag. (2005).

[10] VINOD, H. D.; LOPEZ-DE-LACALLE, J. *Maximum Entropy Bootstrap for Time Series: The meboot R Package. Journal of Statistical Software*, v. 29, n. 5, p. 1–19, 2009).

[11] CARPENTER, J.; BITHELL, J. *Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. Statistics in Medicine*, v. 19, n. 9, p. 1141–1164, (2000).

[12] SUTRISNO, E., OH, H., VASAN, A. S. S., E PECHT, M., *Estimation of remaining useful ball bearings using data driven methodologies. 2012 IEEE Conference on Prognostics and Health Management*, 2, 1-7, (2012).