Remaining Useful Life Estimation for Systems with Non-Trendability Behaviour

Sergey Porotsky and Zigmund Bluvband
ALD Group
Tel-Aviv, Israel
sergey@ald.co.il, zigmund@ald.co.il

***Abstract* - This paper presents an approach and the presented solution of the questions raised in the IEEE PHM 2012 Conference Data Challenge Competition. What was given (known) is the real run-to-failure data of 6 bearings only from the three groups exposed to different operating conditions. One should use this data to estimate the Remaining Useful Life (RUL) of the given set of 11 test bearings. The main feature of the presented data is significant loss of trendability (i.e. "non-trendability") of the defined significant parameters' behavior (horizontal and vertical vibration), thus avoiding the use of well-known supervised learning RUL prediction models. New models have been developed and used; further the Cross-Entropy method has been used for control parameter optimization based on the Cross-Validation procedure.**

The presented solution has been recognized as a Winner in the above mentioned Competition.

The achieved results demonstrate the effectiveness of the approach for the RUL estimation for systems parameters with the non-trendability behavior.

***Keywords* – Cross-Entropy, Cross-Validation, Prognostics, RUL estimation, Remaining Useful Life, Trendability.**

I. INTRODUCTION

Condition-Based Maintenance (CBM) is known to be the most efficient maintenance policy for equipment and components which can be periodically inspected. CBM is usually based on data, collected through condition monitoring. The question is how to predict an item's RUL with the acceptable accuracy and precision (consistency). Prognostics techniques depend on whether the prognostics assessment is based on data obtained from a model or on general historical/statistical data. The first of these two has been termed a "white box" approach (model-based techniques, physics-based), while the second has been called a "black box" approach (data-driven, model-free techniques).

This article describes the development of a data-driven algorithm to predict the RUL – thru the example of a bearing as it degrades from an initial (unknown) state to the state defined as a failure.

In this case six multivariate time series data sets from 3 groups of bearings, exposed to different operating conditions, have been provided as the Learning Set (of the IEEE PHM 2012 Challenge Competition).

Each data set included:

1. time-series of horizontal and vertical vibration measurements for several bearings representing every single bearing where the initial wear and manufacturing conditions were unknown.

2. temperature measurements

Test sets contained examples of units that run some time prior to failure. The 11 test sets are used to predict the RUL of the bearings and evaluate the accuracy of solutions submitted.

This rest of the paper is organized as follows:

- Detailed analysis of the data sets, provided by the Challenge Competition, is presented in section "Experimental Data".
- Section "Data-Driven Prognostics" discusses wide-known data-driven prognostics approaches and reasons for their inappropriateness for the presented data sets.
- The proposed prediction algorithm is presented in section "RUL Estimation Approach".
- Importance of the prediction accuracy and different predictability measures, as absolute and relative accuracy, is discussed in section "Accuracy Criteria".
- Some aspects of the Cross-Validation performing are discussed in section "Training of the proposed model".
- To support the optimization of model control parameters, the Cross-Entropy Global Optimization Method is proposed, brief description of this method is presented in section "Cross-Entropy Method".
- Section "Case study" discusses the output results, as the proposed model is validated using real vibration monitoring data collected in the field from bearings, and a comparative study is performed.
- The final section presents the conclusions.

II. EXPERIMENTAL DATA

The Learning Set includes operational data (time-series of horizontal and vertical vibration measurements) from 6 different bearings, more exactly, 2 bearings for each of 3 groups exposed to different operating conditions [1].

In each Learning Data Set, the unknown bearing was run for a variable time until failure. The lengths of the runs varied, with the minimum run length of 5,150 sec and the maximum length of 28,030 sec.

The Test Set included operational data from 11 different units – five from the first operating conditions, five from the second, and one bearing from the third operating conditions.

In each Test Data Set, the unknown bearing was run for a variable time until the failure, but researchers have got only truncated time-series, i.e. according to some time before failure:

a) The lengths of the truncated (known for researches) runs are essentially varied, with the minimum of 1,720 sec and the maximum of 23,020 sec.
b) Vibration measurements were performed every 10 sec within 0.1 sec slot with period ~ 40 μsec (2,560 measurements per 0.1 sec).
c) The Learning and Test Sets also included temperature measurements (with period 0.1 sec), but only for 10 bearings from all 17. (Here comes the question how to use this data? For the time being we have used temperature measurements time-series only for quality analysis).

The plot in Figure 1 shows typical example of vibration measurements for the first bearing of the first Operating Conditions group in the Learning Set [2]. It is clear, that until last 1/3 of the full time-series the trendability of the vibration behavior is absent. Given this" non-trendability" in the data, it is impossible to easily observe any trends in the data seeking for possible correlation to the degradation in the bearing. Reason for the loss of trendability is the abnormal conditions of bearing working – large radial force applied on the bearing (4,000 N in the first operating conditions, 4,200 N in the second operating conditions and 5,000 N in the third operating conditions). Due to such abnormal conditions, the bearings' time lives were only 1 to 7 hours.

III. DATA-DRIVEN PROGNOSTICS

If case when the measured data are very noisy (as shown on fig. 1), it is impossible to use them directly. Therefore, the first task is to perform de-noising. There are two different ways to perform de-noising: Trendability dependent de-noising and Trendability independent de-noising. Most of the data-driven RUL predicted methods are oriented for Trendability Statistics. Examples of such statistics are following:

- NASA Aircraft Engine Data Base [3]
- NASA Bearing Data Base [3]
- Bearing Data Base of Gould Pumps at a Canadian Kraft Pulp Mill Company [4]
- Etc.

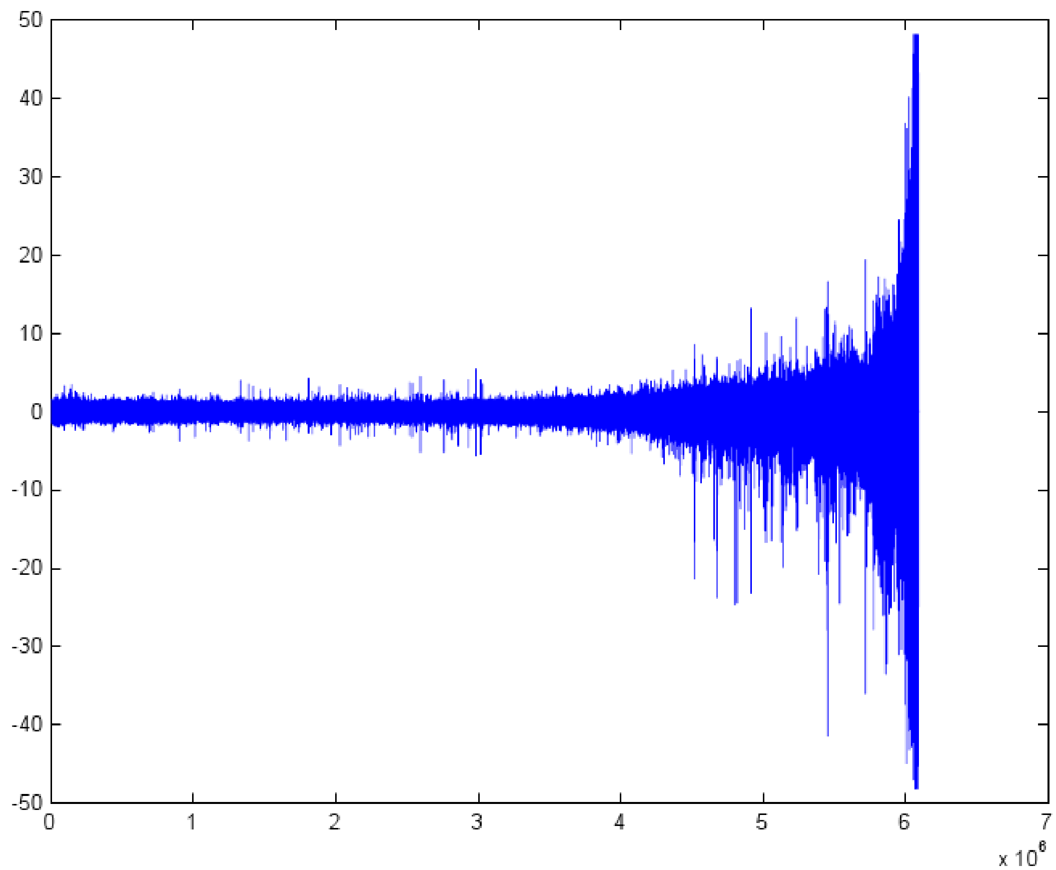

Fig. 1. Vibration changing depending of time

The plot in Figure 2 shows typical example of the Trendability Statistics, see [3], for the Unit number 2 and Sensor number 2 of the data set FD001. For such trendability statistics many approaches are developed. One of the first historical methods, proposed for deteriorating system with single critical parameter (e.g., tires) was based on random process prediction and construction of a conditional failure probability function [5]. In that case the item condition was referred to the clearly defined technical state of the item, identified by known values of critical parameter and corresponding operating time.

In [6] the Similarity-Based Prognostics Approach was used to predict RUL for NASA Aircraft Engine Data Base [3], in [7] Artificial Neural Networks have been considered for Bearing RUL prediction from Data Base [4], and in [8] the Adaptive Neuro-Fuzzy Inference System is used to predict RUL for NASA Bearings Data Base [3].

In case the parameter behavior isn't trendable, but with certain periodicity, some other methods may be used. For example, for data base "CalIt2 Building People Counts Data Set"[9] in [10] Markov-Modulated Poisson Processes models were developed. The plot in Figure 3 shows typical example of the Non-Trendable, but still periodical Statistics.

Unfortunately, for the non-trendable and non-periodical statistics the above mentioned and other wide-known prognostics models are not applicable.

IV. RUL ESTIMATION APPROACH

Preliminary analysis of the given statistics did not pointed out a clear significance of the Temperature measurements for the RUL prediction, due to the following reasons:

- Temperature measurements were performed only for 10 bearings (4 from Learning Data Set and 6 from Test Data Set) from all 17 bearings
- Changing of Temperature are very similar for all 9 bearings (increasing and after this almost constant– plateau), but essentially different for bearing number 1 from first Operational Conditions Group (double increasing and plateau).

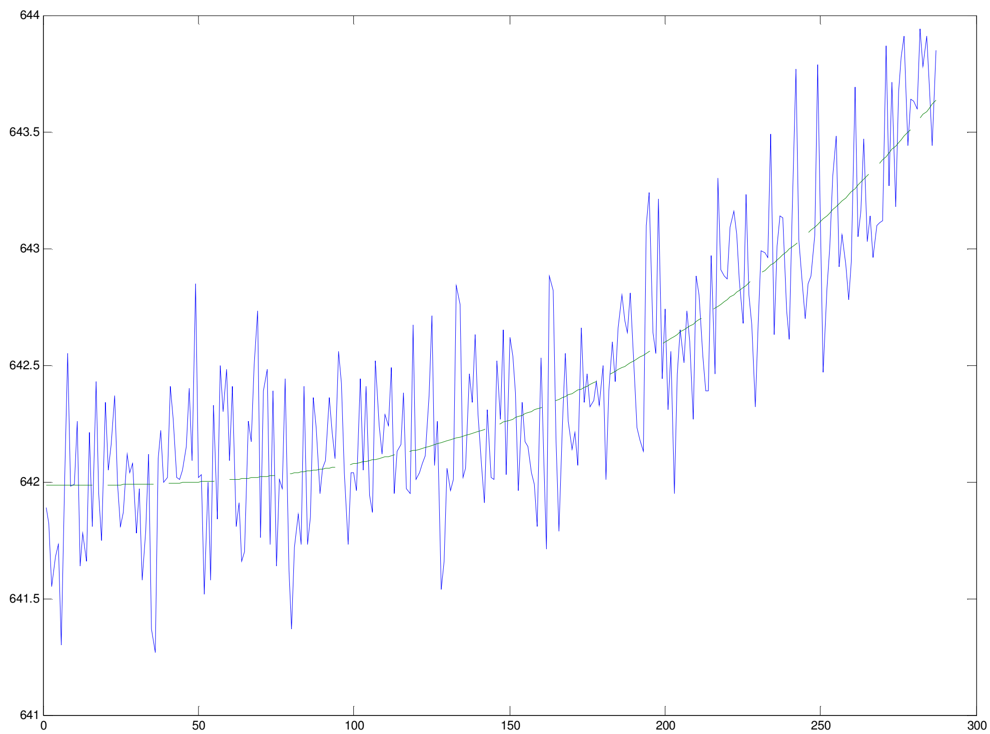

Fig.2. Typical plot of the trendability parameter behavior

Therefore for quantitative analysis we decided to use only vibration parameters. Due to the loss of the trendability, it was impossible to perform de-noising (smoothing) and consequently - RUL prediction directly for vibration parameters. So, instead of directly measured vibration values, we proposed to use accumulated values. In current statistics the vibration values are measured by means of accelerometer as acceleration, in units "g" [2]. Physically the power at the moment t (both for platform and balls) is proportional to the acceleration, and current (instantaneous) degradation is proportional to the power. Thus, accumulated degradation for interval [0…t] is proportional to the accumulated acceleration. Certainly, because the influence of the acceleration on the degradation isn't uniform, we have to take into account following aspects:

- influence of the recent acceleration (e.g., acceleration at the time = current time - 100) is more significant than influence of the older one (e.g., at the time = current time - 1000);
- influence of the large acceleration is more significant than of the low acceleration (e.g., influence of the acceleration = 10g will be 12…15 times stronger than influence of the acceleration = 1g)

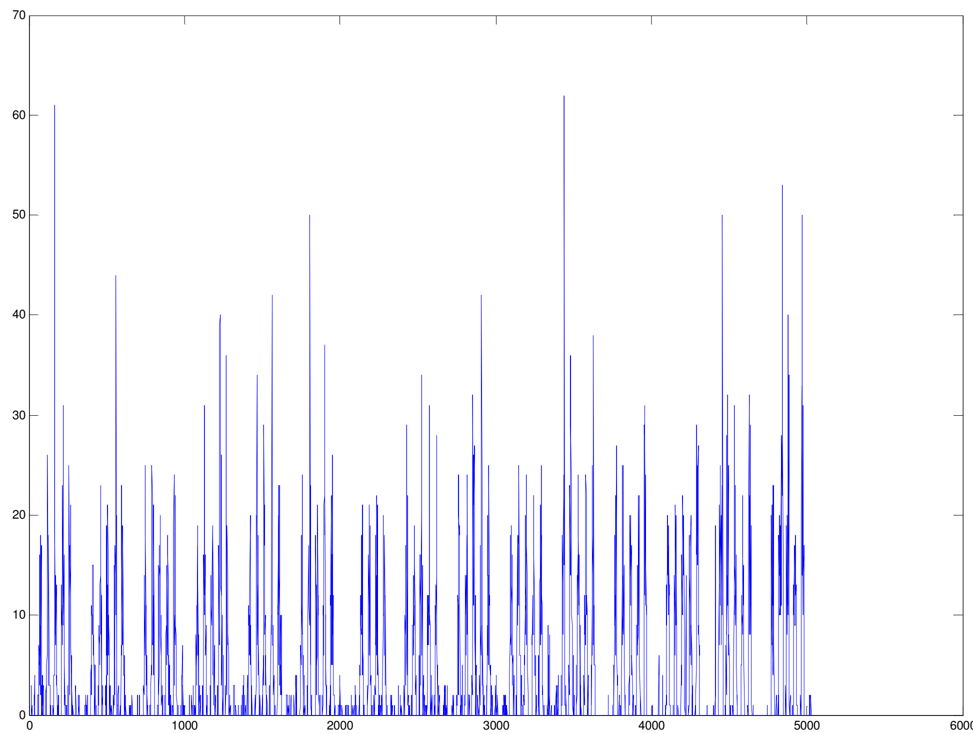

Fig. 3. Number of people entering a building over time

So, for the accumulated degradation of the single bearing at the slot t we propose to use following expression:

$$\mathbf{D(t)} = \sum_{i=1}^{t}\sum_{j=1}^{2560} \mathbf{F(V(i, j))R(i)/2560,}$$

where:

- F is a function to take into account influence of the acceleration value
- R is a function to take into account influence of the time
- t and i are numbers of slots for the considered bearing
- j is number of measurement within slot i (in our case, j = 1…2560)
- V(i, j) is the vibration
- Current time = 10t (in sec)

We didn't attempt to analyze the input data to understand features, presented in the data sets, more clear. So, we have used different types of the F and R functions.
Function R may be following:

- Exponential, $R = e^{(-\alpha(t - i))}$
- Normalized Exponential, $R = te^{(-\alpha(t - i))} / \left(\sum_{i=1}^{t} e^{(-\alpha(t - i))} \right)$
- Polynomial Type 1 , $R = (i/t)^{\alpha}$
- Polynomial Type 2, $R = (t - i + 1)^{\alpha}$
- Normalized Polynomial, etc.

Function F may be following:

- Polynomial , $F = V(i, j)^{\beta}$
- Normalized Polynomial, $F = \left(\sum_{j=1}^{2560} V(i, j)^{\beta} / 2560 \right)^{(1/\beta)}$, etc.

The plot in Figure 4 shows typical example of the Normalized Polynomial Function F behavior depending of slot number - for the $\beta = 3$ and Vertical Vibration of the second bearing from the Learning Data Set of the third operating conditions group .

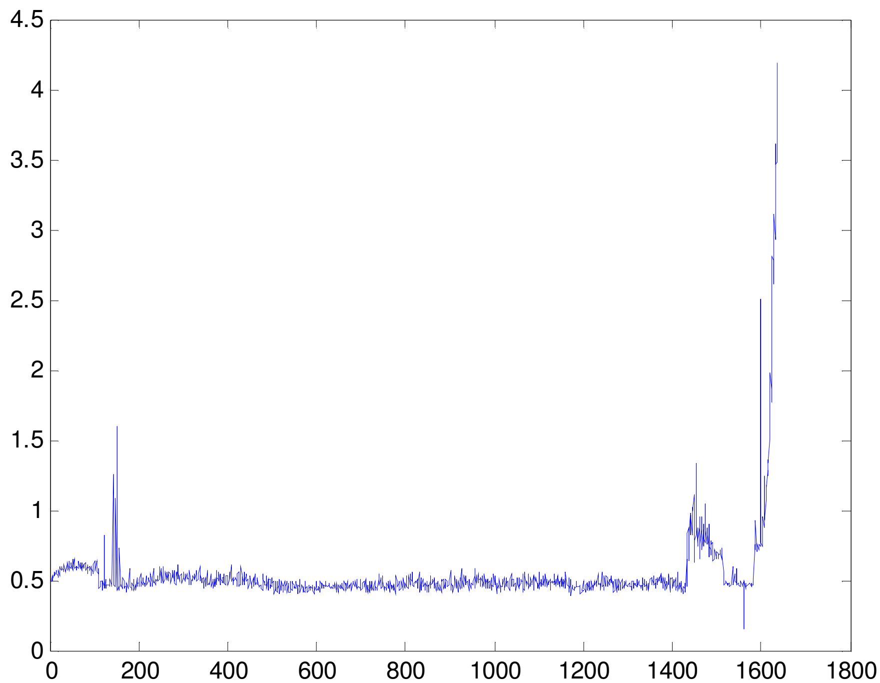

Fig.4. Typical plot of the function F behavior

There are two possibilities to handle measured Horizontal and Vertical Vibration parameters:

- Use an Integral value and perform above proposed accumulation for integral value of the vibration - instead of measured value of the Horizontal and Vertical Vibrations.
- Use one of the wide-known Machine Learning methods for RUL prediction (for example, SVR – Support Vector Regression, RVM – Relevance Vector Machines, etc.) for obtained multi-parameter data - after performing an accumulation separately for the Horizontal and Vertical Vibrations.

Using of the Machine Learning methods assumes a large amount of the input data. For example, the training data set of the PHM-2008 Prognostics Data Challenge included

218 different units [3]. The considered training data of the IEEE PHM 2012 Prognostic Challenge includes much smaller quantity - only 6 bearings from 3 different operating conditions groups. Therefore we should use the first approach. Unfortunately, we could not know what vibration (vertical or horizontal) is more significant and what is the physical picture of the single vibration influence for the Integral vibration.

Due to limited time for the model development and investigation the only following function for Integral vibration was considered:

$V = (wV_h^{\lambda} + (1 - w)V_v^{\lambda})^{1/\lambda}$,

where

- V is Integral value of the vibration
- V_v and V_h are measured values of the vertical and horizontal vibrations
- w is relative weight of the horizontal vibration significance.

Optimal values of types of the F and R functions and control parameters α, β, w and λ were defined by means of Cross-Validation under Learning Set.

There are two possibilities to search values of these control parameters:

- Define common values simultaneously for all operating conditions groups;
- Define individual values separately for single operating conditions groups.

Generally speaking, second approach can provide better accuracy, but it simultaneously leads for over-fitting due to small amount of training bearings in each of the operating conditions groups (two bearings per group). To avoid over-fitting, it is recommended usually to insert additional control parameter – Penalty for Regularization Error, as done at the machine learning methods SVR, RVM, etc. But it can help only if we are using function with high degree of freedom, as RBF (Radial Basis Functions), large-dimension polynomial, etc. The above proposed functions R, F and V have very low amount of control parameters and therefore the use of the Regularization Error couldn't help us to avoid over-fitting. So, to avoid the over-fitting the only approach that could be implemented was the first one.

It was assumed, that for each "k" out of the 3 different operating condition groups, exists such an individual threshold value Thr(k), that for the D value greater than Thr(k) the bearing fails. In another words, for a single bearing from operational condition group k the estimated failure time t is calculated as

min D(t), s.t. D(t) > Thr(k), where Thr(k) is the Accumulated Degradation Threshold for the operational condition group k (k = 1…3).

The optimal values of thresholds were defined by means of Cross-Validation under Learning Set separately for each of the operational condition groups.

V. ACCURACY CRITERIA

Various accuracy metrics have been considered in [11] to evaluate the quality of prognostics (selected prediction model). They may be divided into two classes:

- Metrics, based on absolute accuracy;
- Metrics, based on relative accuracy.

On the PHM-2008 Prognostics Data Challenge the score of the single RUL prediction is defined as the exponential penalty to the absolute prediction error and the score of an algorithm is defined as the total score from all the predictions [12].

On the IEEE PHM 2012 Prognostic Challenge the score of the single RUL prediction is defined as the exponential penalty to the relative prediction error. The score of accuracy for experiment i is defined as follows [13]:

$$\text{Score}[i] = \begin{cases} \exp(\,(-\ln(0.5)(\text{RelEr}[i]/5)\,), & \text{if RelEr}[i] <= 0 \\ \exp(\,(\ln(0.5)(\text{RelEr}[i])/20\,), & \text{if RelEr}[i] > 0 \end{cases}$$

where RelEr[i] is the relative error (in percents) of prediction for test number i, i = 1…11. The final score of all RUL estimates is defined as the mean of all 11 experiment's score.

VI. TRAINING OF THE PROPOSED MODEL

There are several alternatives to perform the cross-validation. General approach is to divide the available data into two groups, one used as the training set and the other as the validation set – and to repeat this procedure several times. Really it is possible only for fully independent data, not for time-series. From our point of view, it is non-correct to use initial part of the single time-series as training set and other part as validation set, because we don't know End-of-Life of the current unit and so we could not know RUL values for initial points of this time-series for current unit (which are points of the training set).

So, for each of the operating conditions groups we have used 2-fold Cross-Validation, i.e. one bearing from single operating conditions group was considered as training and other – as validation, and after this the procedure was repeated with appropriate bearing exchange.

Another question is how to select inspection points (censored, last measured times) for the validation set. It isn't correct to use last measured times from Test Data Set, time cycle number isn't significant parameter – it is rather ratio of time cycle number for full time-series length, but full time-series length is un-known for test time-series!

For Validation Set we proposed to construct four inspection points as relative parts of the full time-series length, with the following relations: 50%, 70%, 85% and 95 % of the bearing End-of-Life time. Consequently, the full Score for the Cross-Validation procedure is calculated as

$$\sum_{k=1}^{3}\sum_{p=1}^{2}\sum_{m=1}^{4} \text{Score(k, p, m in condition, that Training Bearing Index} = 3 - \text{p)/24},$$

where:

- k is the index of the operating conditions group
- p is the index of bearing within learning data set of k-th operating conditions group
- m is the index of the inspection point.

Types of smoothing F and R functions and optimal values of the control parameters α, β, w and λ are selected to maximize Validation Full Score.

VII. CROSS-ENTROPY METHOD

Selection of the types of the smoothing F and R functions was performed by means of simple comparison of the possible alternatives. To select optimal values of the numerical parameters α, β, w and λ by means of simple enumeration is impossible.
On the other hand, to use some methods, based on gradient or pseudo-gradient calculation are also impossible. In such optimization algorithms the initial guesses for the parameters are very crucial, but for many real tasks the Goal Function isn't convex and has many Local Minimums.
For Global Optimization Task one should use some Random Search oriented method – Cross-Entropy Optimization. The method derives its name from the cross-entropy (Kullback-Leibler) distance - a well known measure of "information", which has been successfully applied in diverse fields of engineering and science. Initially the Cross-Entropy method was developed for discrete optimization [14], but later was successfully extended for continuous optimization [15, 16]. It is relatively new random search-oriented approach (in comparison with Genetic Algorithm, implemented as Toolbox in Matlab, or Simulated Annealing Algorithm), but it has provided very good results for several analogous tasks.

VIII. CASE STUDY

The proposed RUL estimation method is tested using the 11 bearings from three operating conditions groups, provided by the IEEE PHM 2012 Prognostic Challenge Competition [1]. A total score of 0.28 was achieved, which is the overall best in the competition. Details of the RUL prediction accuracy are following:

- Within first operating conditions group the RUL accuracy prediction of the three bearings are good, accuracy prediction of the one bearing RUL is moderate, and RUL prediction of the one bearing is wrong.
- Within second operating conditions group the RUL accuracy prediction of the two bearings are good, accuracy prediction of the one bearing RUL is moderate, and RUL predictions of the two bearings are wrong.
- RUL Prediction of the one bearing from third operating conditions was wrong.

IX. CONCLUSIONS

Accurate unit Remaining Useful Life prediction is critical to effective condition based maintenance. In the last ten years a lot of RUL prediction methods were developed, but most of them are applicable only for "with trendability" or "without trendability" (non-trendability) periodical statistics. In this article which describes the case of non-trendability statistics with small amount of units in the learning data set, an advanced model of the data-driven prognostics methods has been presented. The presented suggestion is based on the use of smoothed accumulation of the measured parameters.
The proposed approach is validated using the monitoring data collected in the field from bearings on a FEMTO-ST institute. Experimental results show that the developed model produces satisfactory RUL prediction estimations. To improve accuracy, in future it is supposed to construct more carefully the smoothing functions for measured parameters accumulation, by means of investigation of knowledge of the bearing degradation physical process. Such, the hybrid models will be used instead of the data-driven one.

REFERENCES

[1] IEEE PHM 2012 Prognostic Challenge. Data acquisition parameters and organization of data. http://www.femto-st.fr/f/d/IEEE-Challenge-Data.pdf

[2] IEEE PHM 2012 Prognostic Challenge. PRONOSTIA platform and experiments. http://www.femto-st.fr/f/d/IEEE-Challenge-Platform.pdf

[3] NASA Ames Research Center. Prognostics Center of Excellence Data Repository. http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/

[4] B. Stevens, "EXAKT reduces failures at Canadian Kraft Mill", 2006. http://www.omdec.com/wordpress/wp-content/uploads/2011/03/Case-Kraft-Pulp-Mill.pdf

[5] M. Zule and Z. Bluvband, "RCM policy: new method of residual lifetime prediction according to item's actual operating state", *Reliability, Quality Control and Risk Assessment*, IASTED International Conference, Washington, 1992

[6] T. Wang, J. Yu, D. Siegel and J. Lee, "A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems". *Proceedings of the International Conference on Prognostics and Health Management Society*, Denver, USA, October 2008.

[7] Z. Tian, "An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring". *Journal of Intelligent Manufacturing*, 2009.

[8] K. Javed, R. Gouriveau, R. Zemouri and N. Zerhouni, "Improving data-driven prognostics by assessing predictability of features", *Proceedings of the International Annual Conference of the Prognostics and Health Management Society*, Montreal, Canada, September 2011

[9] UCI Machine Learning Repository. CalIt2 Building People Counts Data Set. http://archive.ics.uci.edu/ml/datasets/CalIt2+Building+People+Counts

[10] A. Ihler, J. Hutchins and P. Smyth, "Adaptive event detection with time-varying Poisson processes", *Proceedings of the 12th ACM SIGKDD Conference (KDD-06)*, August 2006.

[11] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher, "Metrics for Evaluating Performance of Prognostics Techniques," *Proceedings of the International Conference on Prognostics and Health Management Society*, Denver, USA, Oct 2008.

[12] PHM-2008 Prognostics Data Challenge Dataset. http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/

[13] IEEE PHM 2012 Prognostic Challenge. Scoring of results and application procedure. http://www.femto-st.fr/f/d/IEEE-Challenge-Appli.pdf

[14] R. Y. Rubinstein and D. P. Kroese, "The Cross-Entropy Method: A unified approach to Combinatorial Optimization,Monte Carlo Simulation and Machine Learning", Springer-Verlag, 2004

[15] D. P. Kroese, S. Porotsky and R. Y. Rubinstein, "The Cross-Entropy Method for Continuous Multi-Extremal Optimization", *Methodology and Computing in Applied Probability*, 2006, 8(3) : 383–407.

[16] D. P. Kroese, R. Y. Rubinstein, I. Cohen, S. Porotsky and T. Taimre "The Cross-Entropy Method", *Encyclopedia of Operations Research and Management Sciences,* Third edition, Springer-Verlag, 2012.