

Novel failure prognostics approach with dynamic thresholds for machine degradation

Kamran Javed, Rafael Gouriveau, Noureddine Zerhouni
 FEMTO - ST Institute, 24 rue Alain Savary, 25000 Besançon, France
 Email: raphael.gouriveau@femto-st.fr

Abstract—Estimating remaining useful life (RUL) of critical machinery is a challenging task. It is achieved through essential steps of data acquisition, data pre-processing and prognostics modeling. To estimate RUL of a degrading machinery, prognostics modeling phase requires precise knowledge about failure threshold (FT) (or failure definition). Practically, degrading machinery can have different levels (states) of degradation before failure, and prognostics can be quite complicated or even impossible when there is absence of prior knowledge about actual states of degrading machinery or FT.

In this paper a novel approach is proposed to improve failure prognostics. In brief, the proposed prognostics model integrates two new algorithms, namely, a Summation Wavelet Extreme Learning Machine (SWELM) and Subtractive-Maximum Entropy Fuzzy Clustering (S-MEFC) to predict degrading behavior, automatically identify the states of degrading machinery, and to dynamically assign FT. Indeed, for practical reasons there is no interest in assuming FT for RUL estimation. The effectiveness of the approach is judged by applying it to real dataset in order to estimate future breakdown of a real machinery.

I. INTRODUCTION

Prognostics and Health Management (PHM) is an emerging discipline which links studies of failure mechanisms and life cycle management [1]. Predicting future behavior of complex machinery is a complicated challenge that is experienced in many industrial applications covered by the community of PHM. Mainly, PHM discipline aims at making use of past, present and future information of an equipment in order to assess its degradation, diagnose faults, predict and manage its failures [2]. Considering such activities, Fig. 1 shows complete PHM cycle which is composed of three phases i.e. observe, analyze and act. Within analysis phase, prognostics appears as a key task with future capabilities, that are inevitable for critical machinery to shift from *fail-fix* to *predict-prevent* maintenance.

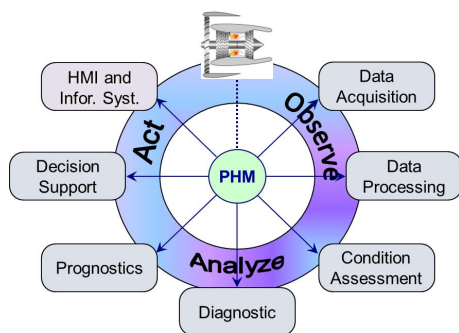


Fig. 1. Prognostics and health management cycle

According to ISO:13381-1, prognostics is defined as the “estimation of time to failure and risk for one or more existing and future failure modes” [3]. It is thereby a promising activity that benefits in form of planning, safety, availability and maintenance cost reduction [4]. This paper deals with data-driven approach (mainly from Artificial Intelligence category) [5], and aims at improving classical approach to estimate RULs. Mainly, data-driven approaches have an advantage of better applicability, when there is absence of prior knowledge or human experts. Therefore they can be seen as black box models that learn system’s behavior directly from data and do not require any specific knowledge about the system [6], [7]. For data-driven approaches, the aspect of failure prediction can be viewed as set of activities that must be performed to facilitate decision making and maintenance tasks [6] (Fig. 2).

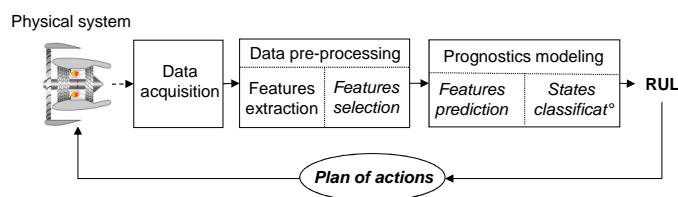


Fig. 2. Frame of data driven PHM

Leaving aside data-acquisition and data preprocessing steps, developments of this paper aim at improving the prognostics model. Practically, machine degradation can be represented by continuous and discrete states, where continuous state shows value of degrading signal while discrete states depict fault modes. Therefore, prognostics modeling requires an efficient tool to predict continuous states of a signal and a classifier to estimate states of degrading machinery [6], [8]. However, prognostics can be quite challenging when few knowledge or previous experiences on degradation process are available [9], or in absence of prior knowledge about fault modes and FT (unlabeled data).

The use of discrete states not only avoids classical way of assuming thresholds for continuous degrading signals but can also benefit in reducing uncertainty of RUL estimates due to imprecise thresholds based on assumptions. This idea was initially proposed in [10], and was further improved in [11]. However, to assign FT (at faulty state), practically, these works have a common drawback: the number of states (fault modes) are pre-assumed, which is not the case for real machinery, because each machine can have different levels of degradation before failure.

According to all this, main issues of prognostics modeling addressed in this paper can be pointed out as follows.

- How to build an accurate prediction model?
- How to manage unlabeled data?
- How to set FT without prior knowledge?

Considering such issues, improvements to prognostics modeling are made by following contributions. In brief, for continuous state prediction task, a new algorithm for neural network is proposed, namely the Summation Wavelet Extreme Learning Machine (SWELM), that is accurate enough and with better applicability. In short, SWELM is an improved variant of an extreme learning machine (ELM) algorithm [12], [13], combined with wavelet theory [14]. Secondly, for discrete state estimation task, another algorithm is proposed called as, Subtractive-Maximum Entropy Fuzzy Clustering (S-MEFC). S-MEFC algorithm takes the advantages of Subtractive Clustering (SC) algorithm [15], and of a fuzzy clustering approach by means of Maximum Entropy Inference (MEI) [16]. S-MEFC enables managing unlabeled multi-dimensional data to automatically determine fault modes (or states) by assigning FT in an unsupervised manner. Finally, RUL estimates for prognostics can be achieved by integrating SWELM and S-MEFC for simultaneous continuous state prediction and discrete state estimation.

The paper is organized as follows. Section II discusses issues of existing thresholding techniques and highlights the importance of dynamic threshold assignment. Section III presents proposed approach for prognostics which includes details about SWELM, S-MEFC and RUL estimation. In section IV the entire proposition is illustrated and discussed on simulation data of Turbofan engines from NASA. Finally, section V concludes this work.

II. DYNAMIC THRESHOLD: PROBLEM STATEMENT

In a classical manner, RUL estimation is based on the study of one-dimensional signal: where RUL is computed between critical time to initiate prediction t_c and the time at which degrading signal passes the FT (θ) or a limit on damage level, i.e., system enters in potential dangerous mode (Fig. 3). FT does not necessarily indicate complete failure of the system, but beyond which risk of functionality loss [17]. Uncertainty of RUL estimation is due to prediction and FT. Most importantly, it is not only the question of performing good predictions but also stopping prediction process at right time, which can result on-time RUL estimates rather than early or late estimates. However, early predictions are preferable than late predictions which may cause catastrophic consequences. Specification of FT is a critical issue and there is lack of stan-

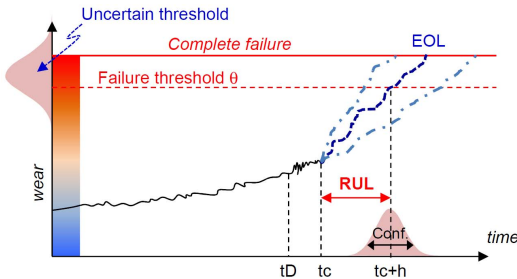


Fig. 3. Failure threshold and RUL uncertainty

dard approaches in defining threshold [7], [18]. Assessment of system's health requires the threshold to be tuned precisely which can be a practical limitation especially in situations when the signal does not represent any physical meaning [11]. In addition, the use of multi-dimensional degrading signals to achieve reliable RUL estimates makes thresholding techniques more complicated.

The use of discrete states not only avoids classical way of assuming thresholds for continuous states of signals but can also benefit in reducing uncertainty of RUL estimates. This idea was initially proposed in [10], and was further improved in [11]. However, works presented in [6], [8], [10], [11] have a common drawback, that the number of discrete states are pre-assumed for continuous observations. RUL estimates are achieved accordingly. Practically it is not the case for real machinery, because behavior of each machine differs according to operational environment. It can be quite challenging or even an impossible task for previous approaches that pre-assume fault modes or state to assign threshold. Therefore, during operation each machine can have different states (fault modes) or degradation levels towards failure or faulty condition [19], [20]. In other words, different failure mechanisms have different trending parameter which requires FT to be assigned dynamically according to a particular case.

III. PROPOSED APPROACH FOR PROGNOSTICS

In PHM applications, prognostics approaches are used to compute long term predictions of continuous observations followed by FT in order to detect faulty state and to finally estimate the RUL of the degrading system. However, real industrial systems are intrinsically not "perfect" and the usefulness of gathered data is highly dependent on the variability of phenomena, sensor nonlinearity, etc. This all makes difficult for understanding (and modeling) of complex and uncertain behavior of real system [12]. The proposed approach integrates two new algorithms, namely, SWELM (neural network) and S-MEFC (hybrid of density based and fuzzy clustering) to perform simultaneous prediction and state estimation for RUL estimates. Details about proposed algorithms and RUL estimation procedure are presented in following sections.

A. SWELM for continuous state prediction

The accuracy of a prognostics system is related to its ability to predict degrading behavior of an equipment. Within data-driven approaches, artificial neural networks (ANN's) are a special case of adaptive networks. They are widely used among machine learning methods in PHM domain [21]. Although there are several ANN based approaches, relatively a new (one-pass) algorithm that requires only one parameter for tuning was proposed recently for a single layer feed forward neural networks (SLFN), namely the Extreme learning Machine (ELM) (see [13] for details). According to author's knowledge ELM has not been applied to prognostics for RUL estimation task. Obviously, no algorithm is perfect, and apart from fast learning, performances of ELM are closely related to parameter initialization step as well as to the type of activation functions used to transform non-linearity. Such issues can increase the complexity of model and may lead to ill-condition [12]. To face those problems, we propose the SWELM that is an improved variant of ELM that combines

ANN and wavelet theory [14]. SWELM appears to be an effective solution to build approximation or prediction model for prognostics. Like ELM, SWELM is also a tuning free one pass approach for SLFN. Similarly, in SWELM approach, hidden node parameters are not only independent of training but also each other.

1) *Structure and mathematical formalization:* The basic structure of SWELM is composed of three fully connected layers of neurons (input, hidden and output layers). However, main differences between the structure of ELM and SWELM are the following.

- To implement non-linear transformations in a better manner, each hidden neuron holds a parallel conjunction of two distinct activation functions (f_1 and f_2) rather than a single activation function. The output from an hidden node is the average value after performing transformation from dual activations ($f = (f_1 + f_2) / 2$) (see Fig. 4).
- Secondly, to improve the convergence of algorithm, an inverse hyperbolic sine (Eq. (1), [22]) and a Morlet wavelet (Eq. (2), [23]) are used as dual activation functions.

$$f_1 = \theta(t) = \operatorname{arcsinh}(t) = \int_0^t \frac{dx}{(1+x^2)^{1/2}} \quad (1)$$

$$f_2 = \psi(t) = \cos(5t) e^{(-0.5t^2)} \quad (2)$$

Let note n and m the numbers of inputs and outputs, N the number of learning data samples (x_i, t_i) , where $i \in [1 \dots N]$, $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathfrak{R}^n$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in \mathfrak{R}^m$, and \tilde{N} the number of hidden nodes, each one activation functions f_1 and f_2 . For each sample j , the output o_j is mathematically expressed as:

$$\sum_{k=1}^{\tilde{N}} \beta_k \bar{f}[(\theta, \psi)(w_k \cdot x_j + b_k)] = o_j, j = 1, 2, \dots, N \quad (3)$$

where $w_k = [w_{k1}, w_{k2}, \dots, w_{kn}]^T \in \mathfrak{R}^n$, is an input weight vector connecting the k^{th} hidden neuron to the input layer neurons, $(w_k \cdot x_j)$ is the inner product of weights and inputs, and $b_k \in \mathfrak{R}$ is the bias of k^{th} neuron of hidden layer. Also, \bar{f} shows the average output from two different activation functions i.e., an inverse hyperbolic sine activation function θ and a Morlet wavelet activation function ψ . Finally, $\beta_k = [\beta_{k1}, \beta_{k2}, \dots, \beta_{km}]^T \in \mathfrak{R}^m$, is the weight vector to connect the k^{th} neuron of hidden layer and output neurons.

In order to minimize the difference between network output o_j and given target t_j , $\sum_{j=1}^{\tilde{N}} \|o_j - t_j\| = 0$, there exist β_k , w_k and b_k such that:

$$\sum_{k=1}^{\tilde{N}} \beta_k \bar{f}[(\theta, \psi)(w_k \cdot x_j + b_k)] = t_j, j = 1, 2, \dots, N \quad (4)$$

which can be expressed in matrix form as,

$$H_{avg} \beta = T \quad (5)$$

where H_{avg} is a $[N \times \tilde{N}]$ matrix expressed as,

$$H_{avg}(w_1, \dots, w_{\tilde{N}}, x_1, \dots, x_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}) =$$

$$\bar{f}(\theta, \psi) \begin{bmatrix} (w_1 \cdot x_1 + b_1) & \dots & (w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \dots & \vdots \\ (w_1 \cdot x_N + b_1) & \dots & (w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix} \quad (6)$$

and,

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \quad (7)$$

Finally, the least square solution of the linear system defined in Eq. (5), with minimum norm (magnitude) of output weights β is:

$$\hat{\beta} = H_{avg}^\dagger T \quad (8)$$

2) *Learning scheme:* Main learning phase derives from Eq. (5) and (6). However, it is desired to take care of parameters initialization task and to provide a better starting point to algorithm. Two types of parameters have to be considered: the ones from the wavelets (dilation and translation) adapted by an heuristic procedure [24], and the ones from the SLFN (weights and bias for input to hidden layer nodes), initialized by Nguyen Widrow (NW) procedure [25]. Details of the learning scheme are synthesized in Fig. 4.

B. S-MEFC for discrete state estimation

The discrete states estimation approach involves multi-dimensional signals from degrading machinery to assign precise threshold, which is recent in PHM community [10], [11]. This strategy facilitates prognostics, when there is absence of prior knowledge about actual states of the degrading system (unlabeled data) as well as failure definition (FT). Indeed, such problems can lead to poor prognostics results. Our previous developments also deal with FT issue and manage unlabeled data in an unsupervised manner by applying fuzzy C-means clustering approach to estimate degrading states [6], [8]. However the methods proposed in [6], [8], [10], [11] have a common drawback that number of states are pre-assumed, where in reality it is contrary (see section II). Moreover, these approaches also face other drawbacks of human choice of parameters that can limit their applicability. Works devoting to clustering analysis of multi-dimensional signals (or time series) are relatively scant for prognostics. Certainly, it is difficult to achieve due to issues related to real monitoring data, i.e., sensor noise and high variability, absence of prior knowledge, dimension and scales as well as size of data, etc.

1) *S-MEFC algorithm:* Clustering is necessary to identify structure in an unlabeled data. Methods for clustering can be classified in to five major categories: partitioning methods, hierarchical methods, density based methods and model based [26]. However, each method has its pros and cons, that are beyond the scope of this paper. S-MEFC algorithm takes benefits of density based Subtractive Clustering (SC) algorithm [15] and of a fuzzy clustering approach by means of Maximum Entropy Inference (MEI) [16].

S-MEFC algorithm is summarized in Alg. 1. In brief, consider a training dataset containing n unlabeled samples denoted as:

$$\ell_D = \{x_i\}_{i=1}^n \quad (9)$$

where $x_i \in \mathfrak{R}^N$ is a N-dimensional real measurements vector. SC approach is used to automatically determine clusters

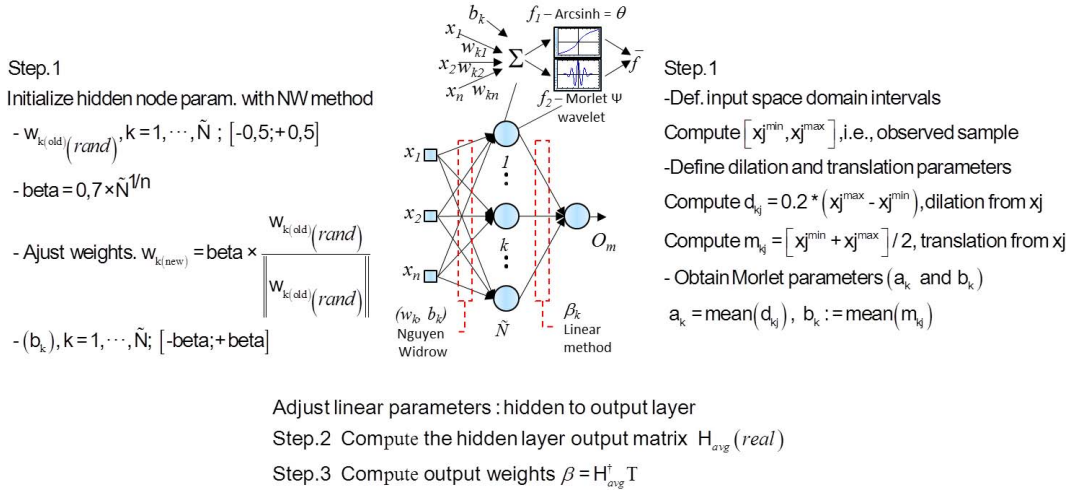


Fig. 4. SWELM structure and algorithm

(states) c in multi-dimensional data, and the centers belonging to each cluster $V = \{v_j\}_{j=1}^c$ (see [15]). SC is one-pass approach that requires only one parameter to be tuned, i.e. radius of neighborhood ra . However, the center positions are not optimized and need to be adjusted [27].

For this task fuzzy clustering approach by means of MEI is used. The obtained centers, V from SC serve the need of MEI clustering algorithm, that avoids any random initialization of centers. To optimize centers positions and to assign membership to each data point in a particular cluster, given σ a fuzziness parameter (set by user), the algorithm runs in an iterative manner until termination criteria (ϵ) is met. The MEI based fuzzy partition matrix can be represented as $U = [\mu_{ij}]_{c \times n}$, where μ_{ij} represents the membership degree of i^{th} object in j^{th} cluster. It should be noted that the key component in clustering to measure similarity between two data being compared [28]. This can strongly influence shape of clustering. In our case we applied Standardized Euclidean distance D_{SE} metric. It is very similar to Euclidean distance except that every dimension is divided by its standard deviation. This leads to better clustering than would be achieved with Euclidean distance when each dimension has different scales. Let x_i, v_j each be a N -dimensional vector. The Standardized Euclidean distance is computed as:

$$D_{SE} = \sqrt{\sum_{k=1}^N (1/s_k^2) (x_{ik} - v_{jk})^2} \quad (10)$$

C. RUL estimation

The flow diagram of failure prognostics using dynamic threshold assignment is shown in Fig. 5.

1) *Learn predictor and classifier*: Consider training dataset containing multi-dimensional signal (sensor measurement/feature) \mathbf{LFi} , which can be from multiple learning instances \mathbf{L} . According to dimension of data, \mathbf{N} univariate predictors \mathbf{Pi} are build using SWELM and trained with data from multiple instances \mathbf{L} . Simultaneously, for each learning instance containing multi-dimensional signals \mathbf{Fi} , a classifier

Algorithm 1 S-MEFC

- Require** - Learning dataset Eq. (9)
- Fix $ra, \epsilon, \sigma > 0$
Ensure - Cluster centers V and fuzzy partitioning U

learning procedure

- 1: Obtain initial cluster centers V_{old} using SC
- 2: Compute fuzzy partition matrix U using MEI

$$\mu_{ij} = \frac{e^{D_{ij}^2/2\sigma}}{\sum_{k=1}^c e^{D_{ik}^2/2\sigma}} \quad \forall i, j \quad (11)$$

- 3: Adjust cluster centers V_{new}

$$V_j = \frac{\sum_{i=1}^n \mu_{ij} \cdot x_i}{\sum_{i=1}^n \mu_{ij}} \quad \forall j \quad (12)$$

- 4: Repeat step 2 and 3 until termination criteria is met

$$\|V_{new} - V_{old}\| < \epsilon \quad (13)$$

(unsupervised) is build using S-MEFC. Each classifier \mathbf{CL} can represent different cluster or states related to particular learning instance. It should be noted that prior learning, data should be pre-processed to remove noisy part. Different filtering techniques can be used: moving average, Loess filter, etc.

2) *Simultaneous prediction and state estimation*: In the off-line phase, \mathbf{N} univariate prediction models and \mathbf{CL} classifiers are build. Let a new test instance \mathbf{TFi} containing partial monitoring data form multi-dimensional signal up to some critical time \mathbf{tc} (or piece of trajectories up to \mathbf{tc}), from where prognostics should be initiated. Prior to RUL estimation task, threshold is dynamically assigned by looking at distance similarity (euclidean) among classifiers and indexes of test data \mathbf{TFi} (eg. multi-dimensional data from index= 1: \mathbf{tc}). Note that each index of \mathbf{TF} can have similarity to a particular \mathbf{CL} . Suppose that, most of the indexes \mathbf{TF} match classifier \mathbf{CLs} that was built in learning phase. Now consider that the total states (cluster) in

CLs are 4; which means 4^{th} state is faulty, and threshold is set to that. When the matched CLs is identified, the current state at tc is also determined by distance similarity between tc states of CLs. After assigning threshold and identification of current state, RUL estimation task is initiated. Finally, N predictors are used to perform multi-step ahead prediction (MSP) in an iterative manner [21], and simultaneously discrete states are estimated by distance similarity (euclidean), thanks to best match CLs (see Fig. 6). The MSP continues until final discrete state reaches threshold point i.e, transition from degrading (d) state to faulty (f) occurs (see Eq. (14)).

$$\text{transition } d \xrightarrow{\text{State } f} RUL = S_{d \rightarrow f}(t+h) - tc \quad (14)$$

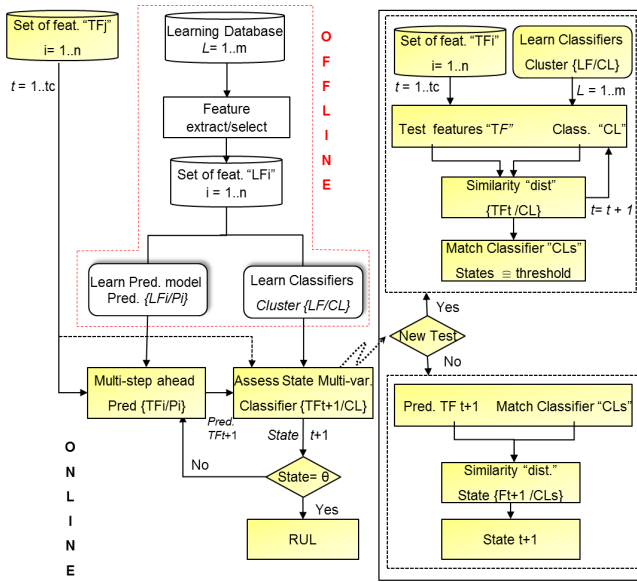


Fig. 5. Failure prognostics using dynamic threshold

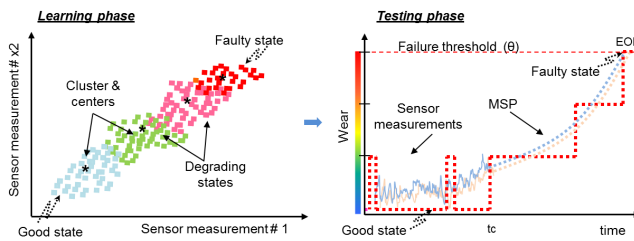


Fig. 6. Multi-variate data online state estimation

IV. EXPERIMENTS, RESULTS AND DISCUSSION

A. Tests outline

1) *Turbofan data*: To demonstrate dynamic threshold assignment procedure, we consider challenge data set of diagnostics and prognostics of machine faults from first international conference of PHM [29]. This data consists of multi-variate time series signals (26 features) from different degrading instances and contaminated with measurement noise. Each engine begins from a normal state but, due to some fault occurrence, starts to degrade. The fault magnitude increases with

time until failure state takes place. In order to train and test the prognostics model, we used text files “*train – FD001.txt*” composed of 100 training instances (with different temporal length or life span) (Fig. 7), and “*test – FD001.txt*” also with 100 instances (but only first 15 are used for testing). It should be noted that the test data are composed of pieces of trajectories and remaining life span is unknown. Each instance either for train or test, consists of 26 variables, from which only 8 are used accordingly to [11].

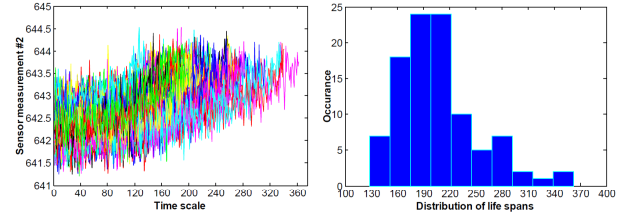


Fig. 7. Sensor measurement from 100 engines (training data) and life spans

2) *Simulation setting and performance evaluation*: According to dimension of data, SWELM was used to build 8 univariate predictors for MSP. Where the topology of each SWELM model was set as, 3 regressors input node, 5 hidden nodes and 1 output node. In case of state estimation with S-MEFC model, the neighborhood radius $ra = 0.4$ was set and fuzziness parameter was set as $\sigma = 0.35$.

For results, estimated RUL are compared with actual RUL provided in the file “*rul – FD001.txt*”. Most importantly for a given test instance TF up to critical time tc , an interval $I = [-10, 13]$, is considered to assess RUL estimates as on-time, early or late, see [11], [29]. Unfortunately the authors who use such data do not clearly mention error of RUL estimates (for each test) which prevents benchmarking of the approach.

B. prognostics results

To estimate RULs, SWELM and S-MEFC model were learned with **100** instances and tests were performed on **15**. It should be noted that both learning or testing datasets were used with actual scales, without normalization. Table I summarizes performances for all tests initiated at tc , estimated current state at tc and the thresholds that are assigned dynamically by distance similarity procedure.

TABLE I. RUL ESTIMATES

Test	tc	Current state	threshold	Act. RUL	Estim.RUL
1	31	2	4	112	112
2	49	2	3	98	54
3	126	3	7	69	68
4	106	2	7	82	80
5	98	1	7	91	100
6	105	1	7	93	108
7	160	2	5	91	114
8	166	1	7	95	102
9	55	1	7	111	105
10	192	3	5	96	68
11	83	3	7	97	67
12	217	1	6	124	131
13	195	2	6	95	92
14	46	2	7	107	81
15	76	3	5	83	106
Mean % error					15,5 %

Among **15** tests instances, RUL estimates of **8** cases fall in interval $I = [-10, 13]$, i.e., on-time predictions. The amount of early predictions is **4**, and amount of late predictions is **3**. It should be noted that early predictions are preferable than late predictions that can cause catastrophic situations [29]. Overall, our proposed method performed prognostics with **15.5%** error. For illustration, the RUL estimation results, with simultaneous prediction (by SWELM) and state estimation (by S-MEFC) involving multi-dimensional signals from the first test instance (engine), are shown in Fig. 8.

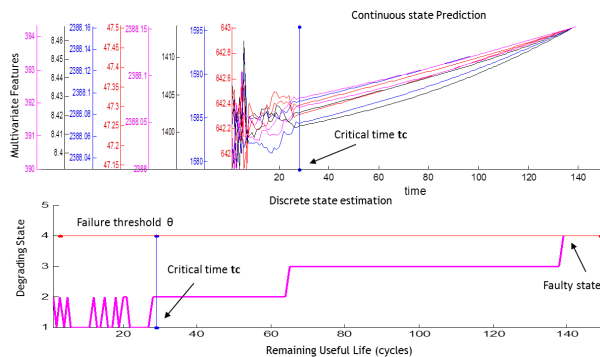


Fig. 8. RUL estimation - Test 1

V. CONCLUSION

The main objective of this paper is to improve data-driven prognostics by introducing a new approach to assign failure thresholds and without any assumptions. Practically, the states of degrading machinery can be continuous which represent value of degrading signal or discrete which depict fault modes. In this context, RUL estimates are achieved through simultaneous prediction and state estimation by integrating SWELM and S-MEFC algorithms. The method shows improved applicability; to manage unlabeled multi-dimensional data, trajectories of different lengths, complexity, and human choice of parameters. Developments are further improved to see overall performance on complete data.

REFERENCES

- [1] S. Uckun, K. Goebel, and P. Lucas, "Standardizing research methods for prognostics," in *International Conference on PHM*, 2008, pp. 1–10.
- [2] E. Zio, "Prognostics and health management of industrial equipment," in *Diagnostics and Prognostics of Engineering Systems: Methods and Techniques*. Seifedine Kadry ed., 2012, pp. 333–356.
- [3] ISO13381-1, *Condition monitoring and diagnostics of machines - prognostics - Part1: General guidelines*. International Standard, ISO, 2004.
- [4] T. Brotherton, G. Jahns, J. Jacobs, and D. Wroblewski, "Prognosis of faults in gas turbine engines," in *Proceedings of the IEEE Aerospace Conference*, vol. 6, 2000.
- [5] A. Heng, S. Zhang, A.-C.-C. Tan, and J. Matwew, "Rotating machinery prognostic: State of the art, challenges and opportunities," *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 724–739, 2009.
- [6] K. Javed, R. Gouriveau, R. Zemouri, and N. Zerhouni, "Features selection procedure for prognostics: An approach based on predictability," in *8th IFAC Int. Symposium on Fault Detection, Supervision and Safety for Tech. Proc., SAFEPROCESS'12*, Mexico, 2012.
- [7] M. Pecht and R. Jaai, "A prognostics and health management roadmap for information and electronics-rich systems," *Microelectronics Reliability*, vol. 50, pp. 317–323, 2010.
- [8] K. Javed, R. Gouriveau, R. Zemouri, and N. Zerhouni, "Improving data-driven prognostics by assessing predictability of features," in *Annual Conf. of the PHM Society*, 2011.
- [9] M. El-Koujok, R. Gouriveau, and N. Zerhouni, "Reducing arbitrary choices in model building for prognostics: An approach by applying parsimony principle on an evolving neuro-fuzzy system," *Microelectronics Reliability*, vol. 51, pp. 310–320, 2011.
- [10] E. Ramasso and R. Gouriveau, "Prognostics in switching systems: Evidential markovian classification of real-time neuro-fuzzy predictions," in *Prognostics and Health Management Conference*, Macao, China, 2010.
- [11] E. Ramasso, M. Rombaut, and N. Zerhouni, "Joint prediction of continuous and discrete states in time-series based on belief functions," *IEEE Tran. on Cybernetics, Part B: Cybernetics*, vol. 43, no. 1, pp. 37–50, 2013.
- [12] K. Javed, R. Gouriveau, N. Zerhouni, R. Zemouri, and X. Li, "Robust, reliable and applicable tool wear monitoring and prognostic: approach based on an improved-extreme learning machine," in *IEEE Conf. on Prognostics and Health Management*, Denver, CO, USA, 2012.
- [13] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Int. Joint Conf. on Neural Networks*, Budapest, Hungary, 2004.
- [14] A. Banakar and M. F. Azeem, "Artificial wavelet neural network and its application in neuro-fuzzy models," *Applied Soft Computing*, vol. 8, no. 4, pp. 1463–1485, 2008.
- [15] S. L. Chiu, "Fuzzy model identification based on cluster estimation," *Journal of intelligent & Fuzzy systems*, vol. 2, no. 3, pp. 267–278, 1994.
- [16] R.-P. Li and M. Mukaidono, "A maximum-entropy approach to fuzzy clustering," in *Proc. of the Int. Joint Conf. 4th IEEE Int. Conf. Fuzzy/2nd Internat. Fuzzy Eng. Symp., Japan*, vol. 4, 1995, pp. 2227–2232.
- [17] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher, "Metrics for evaluating performance of prognostic techniques," in *International Conference on PHM*, 2008, pp. 1–17.
- [18] S. Kumar and M. Pecht, "Modeling approaches for prognostics and health management of electronics," *Int. Jour. of Performability Engineering*, vol. 6, no. 5, p. 467, 2010.
- [19] B. H. Nystad, G. Gola, and J. E. Hulsund, "Lifetime models for remaining useful life estimation with randomly distributed failure thresholds," in *First european conf. of the prognostics and health management society*, vol. 3, 2012.
- [20] S. Cheng and M. Pecht, "A fusion prognostics method for remaining useful life prediction of electronic products," in *IEEE Conf. on Automation Science and Engineering, India*, 2009, pp. 102–107.
- [21] R. Gouriveau and N. Zerhouni, "Connexionist-systems-based long term prediction approaches for prognostics," *IEEE Tran. on Reliability*, vol. 61, no. 4, pp. 909–920, 2012.
- [22] M.-B. Li, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "Fully complex extreme learning machine," *Neurocomputing*, vol. 68, pp. 306–314, 2005.
- [23] A. Pourtaghi and M. Lotfollahi-Yaghin, "Wavenet ability assessment in comparison to ann for predicting the maximum surface settlement caused by tunneling," *Tunneling and Underground Space Technology*, vol. 28, pp. 257–271, 2012.
- [24] Y. Oussar and G. Dreyfus, "Initialization by selection for wavelet network training," *Neurocomputing*, vol. 34, no. 1-4, pp. 131–143, 2000.
- [25] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," in *Proceedings of the Int. Joint Conference on Neural Networks IJCNN*, San Diego, CA, USA, 1990.
- [26] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*. Morgan kaufmann, 2006.
- [27] K. Bataineh, M. Naji, and M. Saqer, "A comparison study between various fuzzy clustering algorithms," *Jordan Journal of Mechanical & Industrial Engineering*, vol. 5, no. 4, pp. 335–343, 2011.
- [28] T. Warren Liao, "Clustering of time series data-a survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.
- [29] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Prognostics and Health Management, 2008. International Conference on*. IEEE, 2008, pp. 1–9.