Scientia Iranica B (2021) 28(3), 1245–1258

SCIENTIA IRANICA

Sharif University of Technology
Scientia Iranica
Transactions B: Mechanical Engineering
http://scientiairanica.sharif.edu

Optimized age dependent clustering algorithm for prognosis: A case study on gas turbines

A. Mahmoodian[a], M. Durali[a,*], T. Abbasian Najafabadi[b], and M. Saadat Foumani[a]

a. *Department of Mechanical Engineering, Sharif University of Technology, Tehran, P.O. Box 11155-9567, Iran.*
b. *Faculty of ECE, Tehran University, Tehran, P.O. Box 14395-515, Iran.*

Received 24 June 2019; received in revised form 12 March 2020; accepted 13 July 2020

KEYWORDS
Age-dependent classification;
Health monitoring;
Prognosis;
Genetic algorithm;
Prognostics.

Abstract. This paper proposes an Age-Dependent Clustering (ADC) structure to be used for prognostics. To achieve this aim, a step-by-step methodology is introduced, that includes clustering, reproduction, mapping, and finally estimation of Remaining Useful Life (RUL). In the mapping step, a neural fitting tool is used. To clarify the age-based clustering concept, the main elements of the ADC model is discussed. A Genetic algorithm (GA) is used to find the elements of the optimal model. Lastly, the fuzzy technique is applied to modify the clustering. By investigating a case study on the health monitoring of some turbofan engines, the efficacy of the proposed method is demonstrated. The results showed that the concept of clustering without optimization processes is efficient even for the simplest form of performance. However, by optimizing structure elements and fuzzy clustering, the prognosis accuracy increased up to 71%. The effectiveness of ADC prognosis is proven in comparison with other methods.

1. Introduction

The prognostics concept, in simple words, is the process of identifying the early signs of failure and degradation in a component and subsequently predicting its Remaining Useful Life (RUL) [1]. As today the maintenance framework in various domains is changing to the predictive paradigm, a great portion of the recent researches is dedicated to the areas such as health management, prognostics, and risk assessment [2–4]. In the industrial systems, the E-maintenance frameworks are used within Prognosis models . The condition of a machine can be monitored at any time through an efficient Prognostics and Health Management (PHM) system. This brings significant cost savings through optimization of maintenance planning and elimination of needless preventative maintenance. Furthermore, the maintenance strategy is transformed from traditional fail-and-fix methods to predict-and-prevent practices. Nowadays, numerous prognostics methods have been developed for economic and operational purposes.

Prognostics methods are categorized into three classes: model-based, data-driven, and hybrid approaches [2]. In case where a perfect system model is not accessible, the data-driven prognostics method is employed to estimate the RUL. Most efforts focus on data-driven approaches and it seems that they reflect the difficulty of defining both damage and failure criteria in model-based approaches. Data-driven approaches depend on the availability of run-to-failure data [5]. Therefore, the damage propagation modeling

*. *Corresponding author. Tel.: +98 21 66165514*
E-mail addresses: mahmoodian@mech.sharif.edu (A. Mahmooidan); durali@sharif.edu (M. Durali); najafabadi@ut.ac.ir (T. Abbasain Najafabadi); m_saadat@sharif.edu (M. Saadat Foumani)

doi: 10.24200/sci.2020.53863.3459

for gas turbines came to be the subject of recent researches, the results of which could provide the required data for developing prognostics algorithms[6,7].

Today as a result of the development of sensing technology and data centres, the collection of large volume data for PHM purposes is made possible. However, the first challenge is how to map the conditions of a complex system which experience a descending trend. The other challenge in a PHM process is designing the prediction module to estimate the RUL of a system after observing a particular disorder. Methods from nonlinear filters [8], dynamic wavelet neural network [9], similarity-based approach [10], to network-based particle filtering [11], a combined method in which the neural networks and wavelet theory are combined [12], and fusion of prognostics algorithms [13] have been used as prediction tools. Yu has developed a prognostics system based on logistic regression and a state-space-model for engine RUL prediction [14]. Simon has compared the accuracy of estimation and computational effort of variants of the Kalman Filter (KF) including linearized KF, extended KF, and unscented KF for aircraft engine health estimation [15]. Based on extended KF, Lu et al. presented a nonlinear underdetermined state estimation method and showed that the proposed method results in a significant enhancement in terms of underdetermined estimation accuracy and robustness [16]. Son et al. proposed a constrained KF for prediction in the condition that monitoring signals are noisy [17]. Besides KF, the particle filter is an attractive approach for degradation prediction [18,19].

A review of health monitoring, diagnostics, and prognostics of mechanical machines is done in [20]. In recent years, researchers have suggested novel methods for prognostics algorithms, including pattern-based method which is the result of the combination of the logical analysis of data and Kaplan-Meier estimator. [21], Bayesian hierarchical model [22], etc. However, in conventional prognosis algorithms, two levels of health states are often assumed including the perfect working state and failure state. It is while, deterioration process of many real-world cases includes different phases of health conditions, leading to a multistage degradation process [23].

Most of the researches which focused on the multistage degradation process suffers from some restrictions that are explained in the following section. Subsequently, the motivation and contribution of this article are described.

First, multistage structures have usually considered a Markovian deterioration process, with a constant transition rate, consequently the parameters such as aging and device deterioration over time are ignored [24,25]. In other models, such as the semi-Markov explicit-duration process [26], the limitation is that the transition between stages is not affected by the age of the device. To overcome the above challenges, based on age-clustering an algorithm is presented in this paper.

The second limitation of current studies is the problem of prediction model selection [27]. Usually, it is assumed that the prediction tool is already known: for example, in this regard the tools such as neural network [28], linear or nonlinear regression [29], particle filter [30], Support Vector Machine (SVM) [31,32] could be t referred to. The focus of the present study is on the parameter selection of the selected tool. The contribution made in this paper is to present a procedure for optimized age-based clustering of data that can be combined with many conventional prediction methods such as Artificial Neural Network (ANN), Local Linear Model Trees (LOLIMOT) [33], regression, etc. to improve their results. In the present article, a flexible structure for prognosis is introduced that uses ordinary and available predictive tools in place of complex methods, while resulting in significant performance improvements.

This paper is organized as follows: In Section 2 the Age-Dependent Clustering (ADC) prognostics framework and its implementation are described To demonstrate the application of the results of this paper, a case study is presented in Section 3. Finally the summary and conclusions are presented in Section 4.

2. ADC algorithm

In this section, the ADC framework and structure elements are briefly discussed.

2.1. ADC framework

This section briefly describes the fundamentals of the ADC framework as shown in Figure 1. The proposed framework includes the following phases:

a) **Clustering.** The time to initiate the prediction of each test unit is indicated with tc. For a prognosis problem with some test units, there is an interval of tc $[\min(tc) : \max(tc)]$. This interval should be divided into some smaller partitions in the clustering phase. The width of the ith cluster is defined by Eq. (1):

$$CW_i = t_i - t_{i-1}, \tag{1}$$

$$\sum_{i=1}^{n} CW_i = \max(tc) - \min(tc).$$

tn is $\max(tc)$ and $t0$ is $\min(tc)$. The number of clusters (n) and the width of each cluster (CW_i) are two elements of structure that should be determined through an optimization process. More details are discussed in the following sections.

b) **Reproduction.** In the next phase, the train data set is reproduced n times for n different age spans.

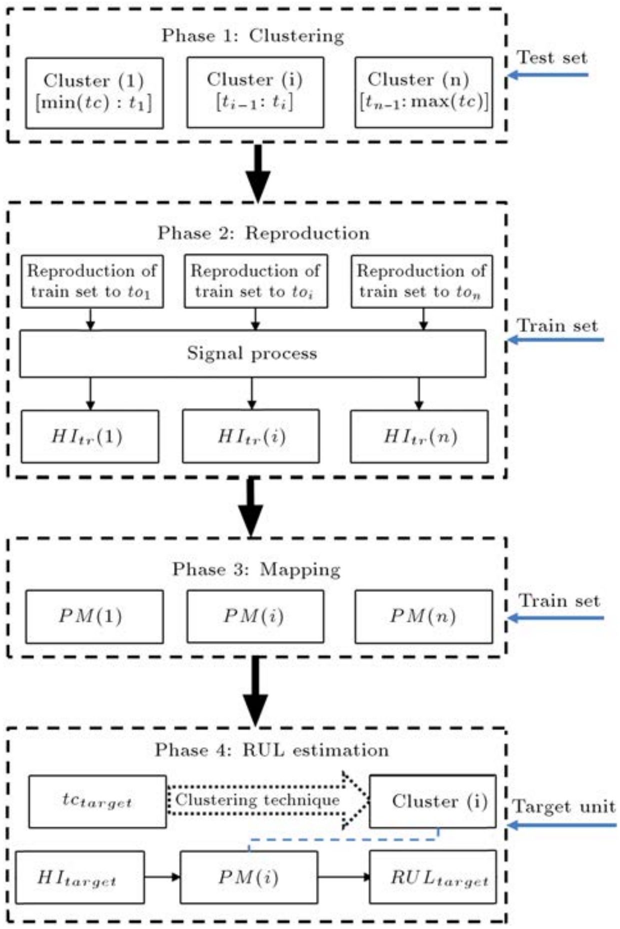

Figure 1. The Age-Dependent Clustering (ADC) framework.

To this end, n observation time points (to_i, i= $1 : n$) are considered and at these points the feature extraction of train data set is performed. Each observation point is relevant to a certain cluster (to_i for ith cluster). In the train data set, for $i = 1 : n$, units with life lengths less than to_i are withdrawn; the others are trimmed to to_i. Each reproduction of train data would be matched with the relevant cluster if its observation point is well selected. For simplicity, it may be assumed that to_i is the middle point of ith interval $[t_{i-1} : t_i]$. However, to obtain better results, observation points should be determined through an optimization process. In the last step of this phase, signal processing of units in each reproduction is performed and n vectors of Health Indicator (HI) are produced. The example presented in Appendix A indicates that the train data is reproduced 3 times for 3 age clusters ($n = 3$).

c) **Mapping.** In the previous phase, train data was reproduced n times and n**HI** vectors were produced. Now in the third phase, a learning algorithm or a fitting tool like ANN, regression, or LOLIMOT is trained to map between **HI** and RULs of n reproduction of train data sets. At the end of this phase n prediction modules are developed for n clusters.

d) **RUL estimation.** Up to this point, the main components of an ADC framework are provided and now they are ready to be used. So, for a target test device, in the first step, the appropriate cluster is selected according to its age when prediction initiates (tc). The clustering technique is discussed in the following.

In the next step, signal processing and HI calculation (such as train units) for the target device is performed. Finally, the features of the HI of the target units are inputted into the appropriate prediction module and the device RUL is estimated.

2.2. Clustering technique

Clustering can be done in a classic manner so that each unit belongs to a specific group. In this way, clustering follows a simple rule: "a unit belongs to a cluster if its tc is between the minimum and maximum age of that cluster".

The main drawback of this approach is that the data points around the boundaries of the clusters may fall into the cluster (i), while they have a higher maturity compared to the cluster ($i + 1$) or cluster ($i - 1$). To solve this problem, it is possible to apply fuzzy clustering techniques so that each data point belongs to one cluster to some degree specified by a membership function. There are various membership functions including sigmoidal, trapezoidal-shaped, Z-shaped, S-shaped, etc. [34].

2.3. ADC elements

The elements of the ADC framework are described in the following:

a) **Number of clusters (n).** This element represents how many age intervals the test data set should be partitioned to. It is conceptually similar to the number of states in a multistate degradation model (Moghaddas et al. (2014) [23]). The number of clusters is an important element of the ADC framework and affects other elements. The low number of clusters may decrease the model accuracy, while the crowded clustering not only does not improve the accuracy but also increases processing load.

b) **Cluster Width (CW).** This element defines the width of each cluster. For simplicity, it may be assumed that all CWs are equal, but usually, it is not the best solution. The portions with the low dynamic and clear trend can be modeled with a low number of clusters having long widths. However, for higher dynamic, more partitions with narrower width are suitable.

c) **Observation time (to).** This element refers to the time points at which the HI of the train data set is observed for each cluster. Observation time of the ith cluster is a point between t_i and t_{i-1}. Failure to select appropriate points for observation of train data may result in poor fitting and unreliable prediction modules for each cluster.

2.4. Elements selection

ADC structure elements depend on the conditions of the test data set and should be customized for a problem. To reach this goal, the elements are compared using a prognostic measure. Figure 2 shows the summary of the elements selection approach. Four phases in the ADC framework are executed, and those elements which best satisfy the selection criteria are selected. Different evolutionary optimization methods such as genetic algorithm, and simulated annealing, can be used to find the optimal elements. The objective function may be various prognosis measures like accuracy, Mean Absolute Error (MAE), squared absolute error. If the fuzzy clustering technique is used, the parameters of membership functions can also be determined by the optimization process.

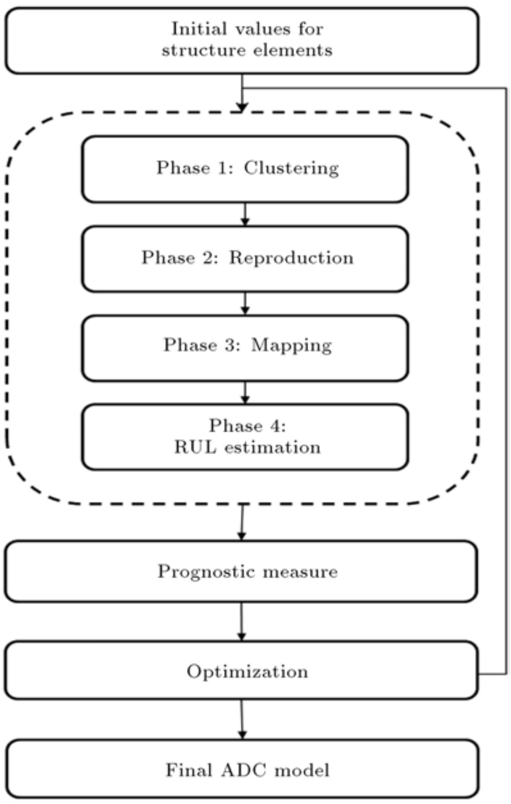

Figure 2. The summary of the elements selection approach.

3. Case study

To illustrate the outcome of this article on prognostics and health monitoring, a case study on turbofan engines from NASA's prognostics information repository is performed [35]. In this section, the effectiveness of the proposed model is focused on.

3.1. Data description

Among numerous datasets collected by NASA's prognostic center of excellence, five extremely popular turbofan engine datasets have been used in more than 70 studies. Among these datasets, dataset #1was used more (70% of the cases) compared to all other datasets [36]. The dataset #1 includes two subsets: 1) train data set, and 2) test data set. The whole data is from a fleet of engines of the same type. The train data is composed of multiple units which operated until failure occurred. Other units operated until destruction occurred at different levels. The data consists of 21 measurements listed in Table 1, which are measured during every flight cycle. The goal was to predict the RUL of test units.

3.2. Health indicator

Considering HI design for a data- driven PHM process, different methods are used to map the sensor information of the system health status. Previously, the method such as direct use of all sensor data [37], feature selection [38], and multi-sensors fusion [39] were the subject of the study of some researchers and they proposed these methods for mapping the sensor information. The simplest method is to use available sensor data without any selection or fusion; for instance, in [40,41] health parameters are flow capacity and isentropic efficiency. In the feature selection, some features that are better predictable and can contribute to prognostic modeling are considered [38]. The data fusion method is used in numerous researches [13] and its effectiveness has been examined in [42]. However,

Table 1. A list of sensors and measurements [6].

Symbol	Description
T24	Total temperature at LPC outlet
T30	Total temperature at HPC outlet
P30	Total pressure at HPC outlet
Nc	Physical core speed
Pr	Engine pressure ratio (P50 / P2)
Phi	Ratio of fuel flow to Ps30
BPR	Bypass ratio
BE	Bleed enthalpy
T50	Total temperature at LPT outlet
Ps30	Static pressure at HPC outlet
farB	Burner fuel air ratio

besides these methods the different fusion methods have been used such as Principal Component Analysis (PCA), and multistream deep recurrent neural network [43].

In the present paper, a reconstructed signal from the fusion of multi-sensors information represents the HI of the device under study. Data processing and feature extraction is performed through a method developed by Diallo [42]. In the first step, data normalization is essential for putting all different types of sensors readings in the same order of magnitude. After normalizing, the signal noise must be removed. In the next step, a fusion of multi-sensors information is performed and a representative signal is calculated. The reconstructed signal for each device is called HI. Finally, the characteristics of two damage indicators SAD (Sum of Absolute Differences) and SSD (Sum of Square Differences) of HI are computed as the Health Indicator Features (HIF). Different stages of data processing for four sensors of engine #1 are indicated in Figure 3. More details are presented in references [42,44].

3.3. Prognostic measures

To measure the prognosis, the error is defined for a given prediction by Eq. (2):

$$err = RUL_t - RUL_s. \tag{2}$$

Regarding PHM, sometimes the early prediction is preferred to late prediction. Therefore, the asymmetric interval $I = [-10, +13]$ around the true RUL is considered to evaluate the performance as shown in Figure 4. The accuracy measure is the percentage of test units and the RUL estimation of this test units falls within the interval I [45]. This criterion appear to be more severe, compared to those defined in the literature [46].

To evaluate the performance of the ADC method more accurately, Mean Square Error (MSE) and Mean Absolute Error (MAE) are measured too by equations (3) and (4) [47]:

$$e_{mse} = \frac{1}{N}\sqrt{\sum_{t=1}^{N} err^2}, \tag{3}$$

$$e_{mae} = \frac{1}{N}\sum_{t=1}^{N} |err|. \tag{4}$$

3.4. Implementation of ADC framework

Based on the illustrated framework (Figures 1 and 2) the steps for ADC algorithm implementation are determined. At first, four phases of ADC implementation are performed with initial values for structure elements. Subsequently, the optimization is done and the final ADC model is completed.

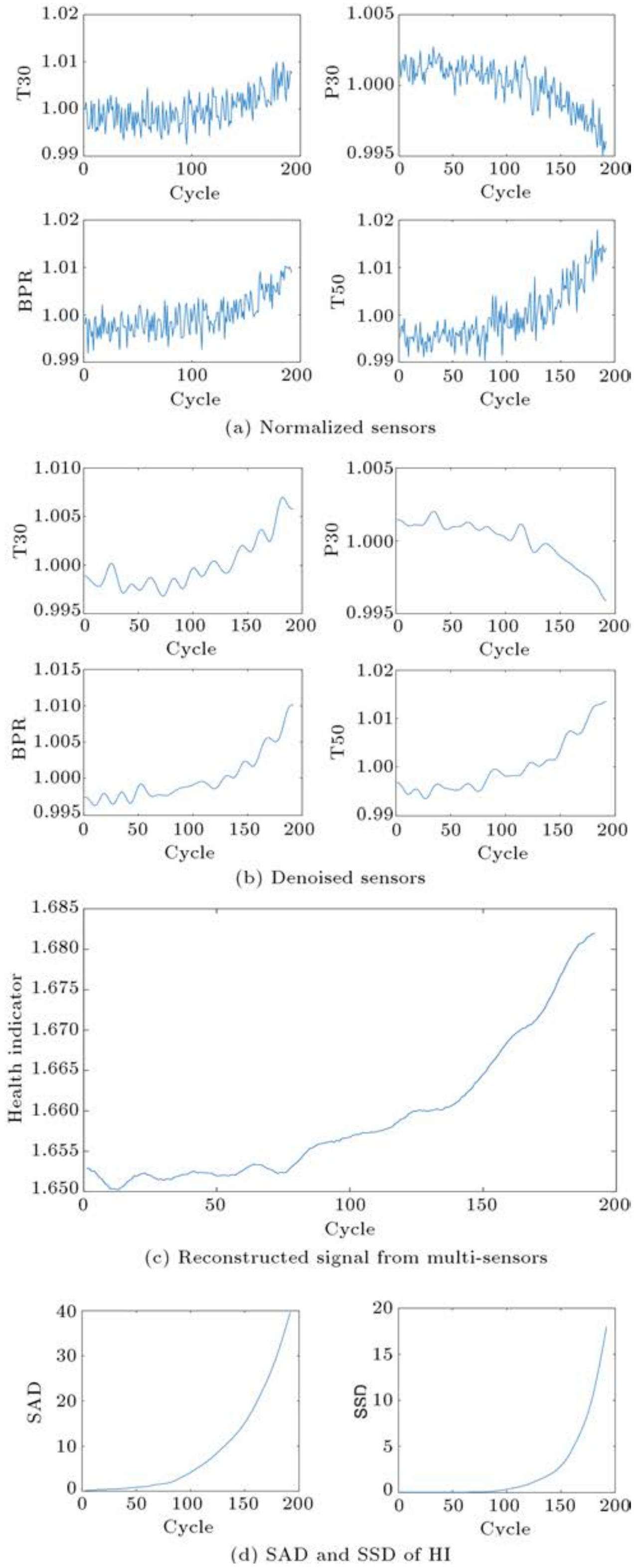

Figure 3. Different stages of data process for engine #1 [44].

a) **Clustering.** The histogram of the prediction start time (tc) for test units is shown in Figure 5. The minimum and the maximum values are 31 and 303

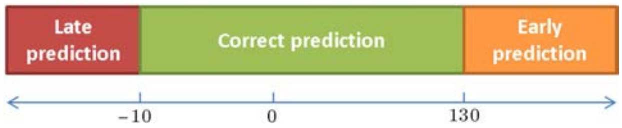

Figure 4. Prognostic measure [45].

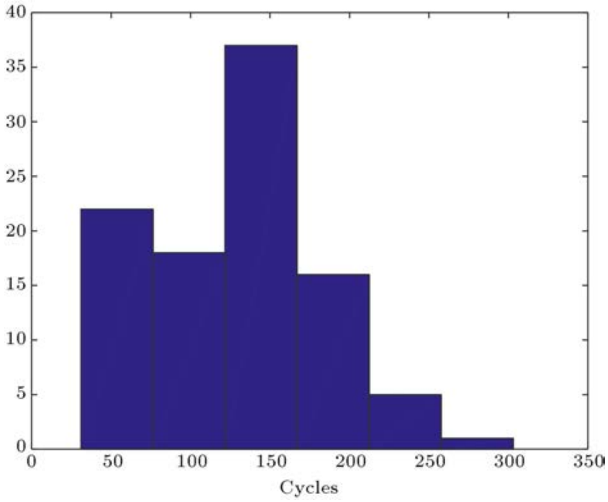

Figure 5. Histogram of prediction start time (tc) for test units.

Table 2. Initial values for structure elements.

Structure elements	Initial value
Number of clusters	6
Cluster widths (cycle)	[45, 45, 45, 45, 45, 45]
Observation points (cycle)	[54, 100, 145, 190, 236, 281]

Table 3. Summary of engines clustering into 6 groups.

Cluster	Observation time (to)	Age interval
I	54th cycle	31–76 cycles
II	100th cycle	77–122 cycles
III	145th cycle	123–167 cycles
IV	190th cycle	168–212 cycles
V	236th cycle	213–258 cycles
VI	281th cycle	259–303 cycles

cycles. The initial values for structure elements are listed in Table 2. To select the initial parameters, CWs are set equal , this is the simplest form to start the algorithm. Also, the midpoints of clusters are taken as the observation points. Regarding the number of clusters, in following sections it is shown that all prognostic measures stay almost unchanged for $n > 2$ for the present case study. Therefore, a low number of clusters is chosen (6 clusters) to avoid costly computations. For these initial values, the clustering scheme is summarized in Table 3.

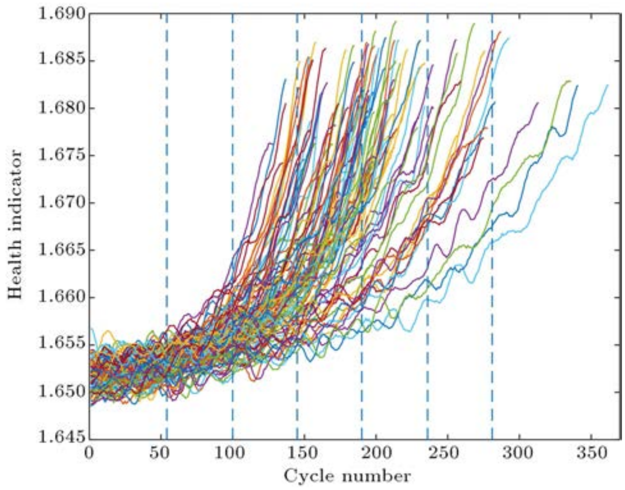

Figure 6. Health Indicator (HI) signals for train data set before clustering (dashed lines indicate observation cycles).

b) **Reproduction.** The train data set was reproduced for each cluster. For the first cluster, all engines with lives longer than the respective observation cycle (54 cycles from Table 3), were stopped at the 54th cycle. Then data process was performed and health indicator HI_{tr} (1) was extracted. For the second cluster, all engines with lives longer than 100 cycles ($to_2 = 100$ from Table 3) were stopped at the 100th cycle, and similarly HI_{tr} (2) was computed. The scheme is repeated for all clusters. The HI signals of train data resulted from multi-sensors fusion before clustering are shown in Figure 6. The observation times are indicated with dashed lines. HI signals after clustering are shown in Appendix B.

Train data set included 100 engines. All engines were used in the first cluster, so the respective $\mathbf{HI}_{tr}\mathbf{(1)}$ vector dimension was 54×100 ($to_1 = 54$ cycles, number of reproduction (1) members = 100 engines). For the second cluster, all train data are used too and the $\mathbf{HI}_{tr}\mathbf{(2)}$ vector was 100×100 ($to_2 = 100$ cycles, number of reproduction (2) members = 100 engines). For the third cluster, the lives of 4 engines were smaller than 145 cycles; so the $\mathbf{HI}_{tr}\mathbf{(3)}$ vector was 145×96 ($to_3 = 145$ cycles, number of reproduction (3) members=96 engines). Through similar analysis, dimensions of **HI** vectors for all clusters are determined and they are summarized in Table 4.

c) **Mapping.** The mapping phase was performed using the neural fitting tool of MATLAB. In this study a forward Multilayer Perceptron (MLP) with a backward propagation training algorithm with neural network structure is proposed. The training algorithm was Bayesian. The developed ANN model consisted of two layers with hidden sigmoid neurons and linear output neurons as shown in Fig-

Table 4. Summary of six reproductions of train data set.

Reproduction	Number of members	Vector	Dimension
I	100	$\mathbf{HI}_{tr}(1)$	54×100
II	100	$\mathbf{HI}_{tr}(2)$	100×100
III	96	$\mathbf{HI}_{tr}(3)$	145×96
IV	62	$\mathbf{HI}_{tr}(4)$	190×62
V	19	$\mathbf{HI}_{tr}(5)$	236×19
VI	8	$\mathbf{HI}_{tr}(6)$	281×8

Table 5. Summary of the formation of the prediction modules.

ANN	Regression R values of net	Input	Target
I	27%	2×100 matrix, representing 2 features of 100 HI signals	1×100 matrix, representing RULs of 100 engines
II	59%	2×100 matrix, representing 2 features of 100 HI signals	1×100 matrix, representing RULs of 100 engines
III	85%	2×96 matrix, representing 2 features of 96 HI signals	1×96 matrix, representing RULs of 96 engines
IV	99%	2×62 matrix, representing 2 features of 62 HI signals	1×62 matrix, representing RULs of 62 engines
V	99%	2×19 matrix, representing 2 features of 19 HI signals	1×19 matrix, representing RULs of 19 engines
VI	99%	2×8 matrix, representing 2 features of 8 HI signals	1×8 matrix, representing RULs of 8 engines

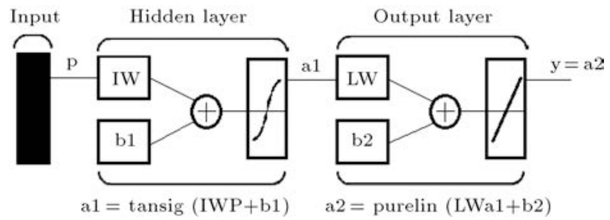

Figure 7. Multilayer Perception (MLP) framework.

ure 7. Two features of each HI signal were inputted to MLP and the output was engine RUL [48,49]. The summary of the formation of six ANNs is presented in Table 5. The training algorithm for all networks was Bayesian regularization. Regression R values indicated the correlation between outputs and targets of the network. An R-value of 100% means a close relationship.

d) **RUL estimation.** The HI signals derived from the fusion of multi-sensors for the test data set are shown in Figure 8. Table 6 shows the details of RUL estimation for 15 engines. For each engine, tc was considered first. Then the appropriate cluster was determined according to the clustering technique. For simplicity, a non-fuzzy rule was applied in this phase: "a unit belongs to a cluster if its tc is between the minimum and maximum age of that cluster". Subsequently, the relevant ANN is used and the RUL is estimated.

The prognostic measures for the ADC method are presented in Table 7. The initial values for structure elements are used in this method and the measures are calculated for all 100 test units.

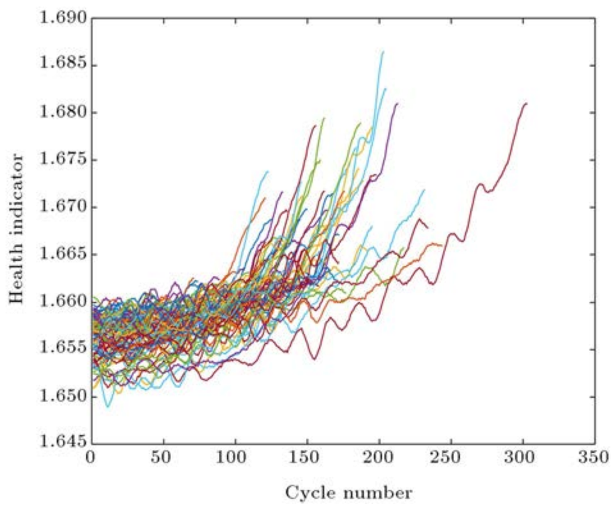

Figure 8. Health Indicator (HI) derived from fusion of multi-sensors for 100 test data.

3.5. Optimization

In the previous section, ADC was implemented with a set of initial values and simple assumptions regarding the framework, and the prognosis measures were evaluated. In this section, the structural elements are optimized and the final ADC model is built. For this, an evaluation data set is required. To tune its Parameters, in middle-phase tests of the algorithm, this data set was used in place of the main test data

Table 6. Details of Remaining Useful Life (RUL) estimation for 15 units (initial ADC model).

Engine	*tc*	Cluster	Estimated RUL	Actual RUL	Error
1	31	1	110	112	–2
2	49	1	92	98	–6
3	126	3	67	69	–2
4	106	2	104	82	22
5	98	2	113	91	22
6	105	2	109	93	16
7	160	3	99	91	8
8	166	3	101	95	6
9	55	1	160	111	49
10	192	4	72	96	–24
11	83	2	129	97	32
12	217	5	64	124	–60
13	195	4	75	95	–20
14	46	1	95	107	–12
15	76	1	102	83	19

Table 7. Summary of prognostic measures for Age-Dependent Clustering (ADC) model with initial values.

Measure	Value
Accuracy (%)	53
MAE	14.9
MSE	1.9

set. The evaluation data set includes train data set, therefore the train units are stopped some cycles before failure, where the prediction should initiate.

It is possible to optimize whole elements simultaneously which leads to numerous possible sets of elements and prolonged run time. However, to minimize optimization time, the optimum number of clusters was determined at first and subsequently, the other elements were optimized. For this purpose, the prognosis measures were calculated while the number of clusters was changed and all other conditions and assumptions remained unchanged. As shown in Figure 9, all prognostic measures remained almost unchanged for $n > 2$ for the present case study. Therefore, a low number of clusters was chosen (for instance 4 clusters) to avoid costly computations in the next optimization steps. Subsequently, the CW and the observation points (to) are optimized by a GA. For this computation, the OPTIMTOOL of MATLAB was used. The objective was to maximize the accuracy. The optimized variables are listed in Table 8. The optimal clustering scheme is summarized in Table 9.

3.6. Clustering technique

To rectify clustering, the fuzzy technique was used [50].

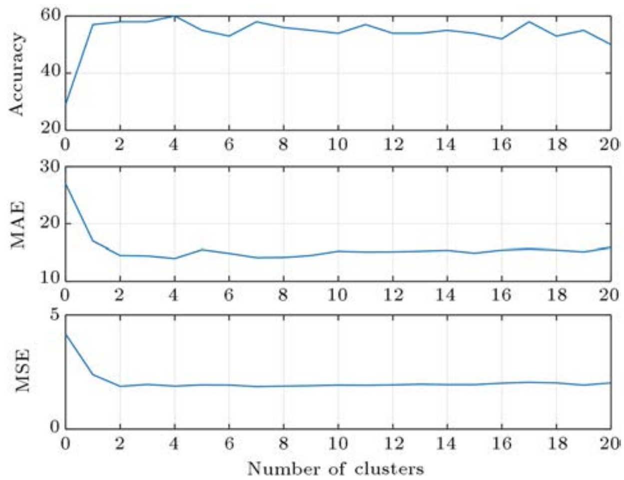

Figure 9. Prognostic measures via number of clusters.

Table 8. Optimized values for structure elements.

Structure elements	Best value
Number of clusters	4
Cluster widths (cycle)	[44 49 70 104]
Observation cycles	[53 100 159 208]

Trapezoidal-shaped membership functions were used for clustering. The fuzzy clustering scheme used in the present paper is shown in Figure 10.

3.7. Results and discussion

The prognostic measures for different methods are summarized in Table 10. All indexes of ADC models were better than ANN without clustering. The

Table 9. Summary of optimum clustering scheme.

Cluster	Observation cycle (to)	Age interval
I	53th cycle	31–75 cycles
II	100th cycle	76–125 cycles
III	159th cycle	126–196 cycles
IV	208th cycle	199–303 cycles

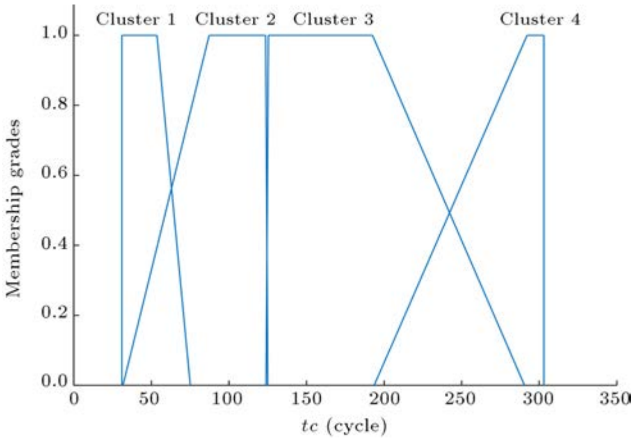

Figure 10. The fuzzy clustering scheme used in the present paper.

prognosis measures were improved significantly with clustering, although the core algorithm in ADC models was still ANN. This may be followed from the following reasons: 1) Probably, some young units were used for the prognosis of an aged or a middle-aged unit, without age-based clustering, and 2) It may be that the life prediction of a young unit was performed according to data mining of some old units. Both of these cases lead to the accuracy reduction. It is worthy of consideration that the method herein developed was the ANN, however, this does not imply that the method presented in this study is limited to ANN, rather it has the capability to be applied in other estimation methods to improve the results.

To our knowledge, the full testing dataset is considered in a few number of studies. Some examples of the application of full testing dataset are the EVIPRO algorithm [51], similarity-instance- based approach [52], and RULCLIPPER algorithm [46] in which the prognosis was measured by accuracy criteria. To evaluate the effectiveness of the ADC algorithm, a comparison with other approaches was performed as indicated in Table 11. The accuracy of the proposed ADC approach was better than other approaches. Figure 11 shows the prognosis results of 100 engines for the optimized ADC model with fuzzy clustering method.

Finally, prognosis results of engines in different cycles are studied. The actual RUL value and the RUL estimate for engines #97 − 100 are shown in Figure 12. The results showed that the RUL estimate was reasonably close to the actual remaining life (especially in older ages). In general, the results indirectly supported the hypothesis that, the age clustering method leads to acceptable results in prognosis.

4. Conclusion and summary

Based on Age-Dependent Clustering (ADC), this study proposes a prognosis algorithm. A four-phase prognosis

Table 10. Summary of prognostic measures.

Method	Remark	Accuracy (%)	MAE	MSE
ANN	Without clustering	29	27.5	4.17
Initial ADC model	Simplest form	53	14.9	1.9
Optimized ADC model	Non-fuzzy clustering	65	14.9	2
Final ADC model	Fuzzy clustering	71	12.8	1.8

Table 11. Comparison of accuracy for different methods.

Method	Remark	Correct (%)	Early (%)	Late (%)
Fuzzy optimized ADC model	Tested on 100 test units	71	23	6
Ramasso [45]	Tested on 100 test units	67	Nan	Nan
Non-fuzzy ADC model	Tested on 100 test units	65	19	16
Khelif et al. [52]	Tested on 100 test units	54	18	28
Ramasso et al. [51]	Tested on 100 test units	53	36	11
Javed et al. [46]	Tested only on 15 test units	53	27	20
Xu et al. [13]	Sensor selection proposed by the authors	44	19	37
Xu et al. [13]	Sensor selection used in [51]	50	19	31

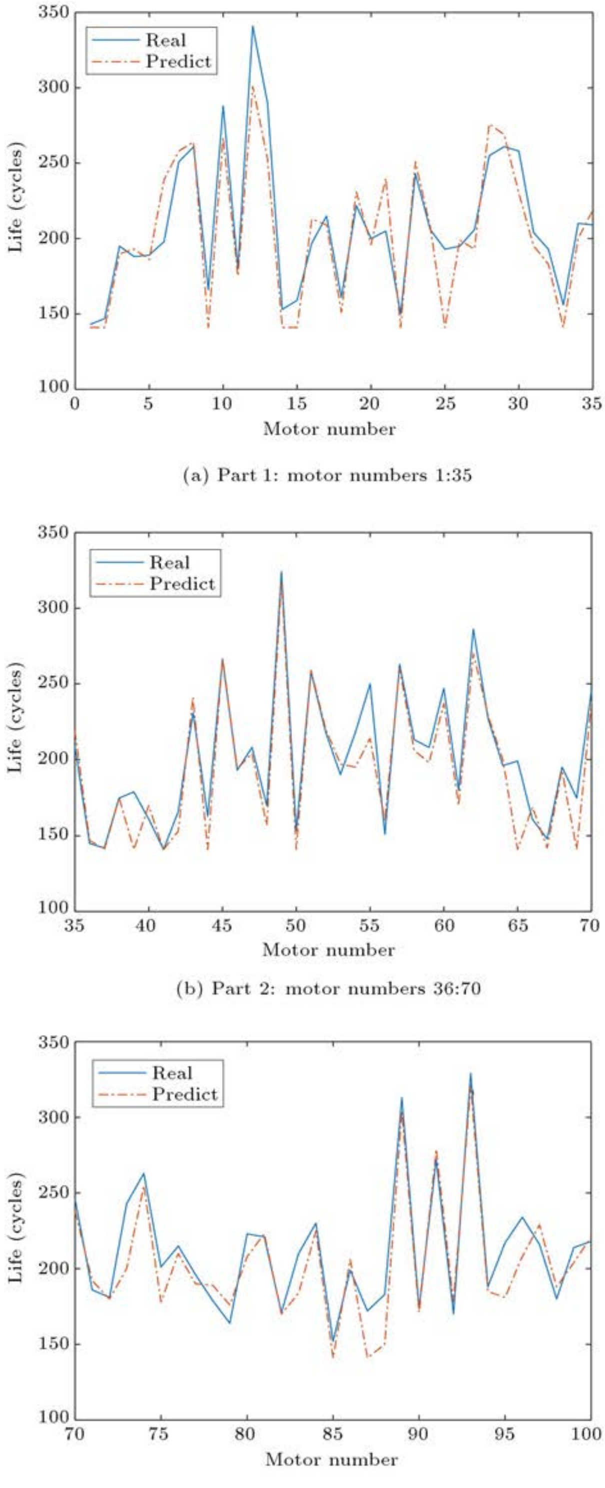

Figure 11. Results of engine prognostics with fuzzy optimized Age-Dependent Clustering (ADC) model.

framework was introduced. In the first phase, clustering of test data set was performed, then reproduction of train data set for each cluster was done, afterward neural fitting tool was used to build n prediction modules using reproduction sets of train data, and finally, cluster selection and Remaining Useful Life (RUL) estimation of each test unit was completed by using the relevant prediction module. Initially, simple assumptions were used to determine the elements of the ADC model. Subsequently, the optimization of the initial prognosis model was considered. The effect of the increasing number of clusters was also studied. A genetic algorithm was applied to optimize the width of clusters and observation points and finally, fuzzy clustering technique was used to increase prognosis accuracy.

The accuracy of the initial ADC model in which the simple assumptions were used (53%) was better than the classic methods (29%), although the core algorithm in both cases was Artificial Neural Network (ANN). This is one of the important results of the present study that proves the effectiveness of the ADC concept. In the present study, the ANN was used as the core prediction tool. The ADC idea can be used and combined with different classic prognostics methods. The core prediction tool may be any of the other existing methods.

The ADC model optimization caused its results to be improved (65% and 71% accuracy for non-fuzzy and fuzzy optimized ADC). The aim of this study was to maximize accuracy through optimization process and the results indicated acceptable accuracy compared to other methods.

One advantage of this algorithm is that the core prediction tool is a conventional method and the ADC algorithm rectifies this method with more reliable results. Another advantage of the proposed method follows from its reliability characteristic which enables us to apply it in real situations.

Nomenclature

tc	Start time of prediction for a system
CW	Cluster Width
tr	(related to) train data set
to	Certain observation points at which the health indicator is observed for each cluster in the training phase
N	Number of clusters
HI	Health Indicator
ANN	Artificial Neural Network
PM(i)	Prediction Module for the ith cluster
RUL	Remained Useful Life
ADC	Age- Dependent Clustering
PHM	Prognostics and Health Monitoring
MAE	Mean Absolute Error
MSE	Mean Squared Error

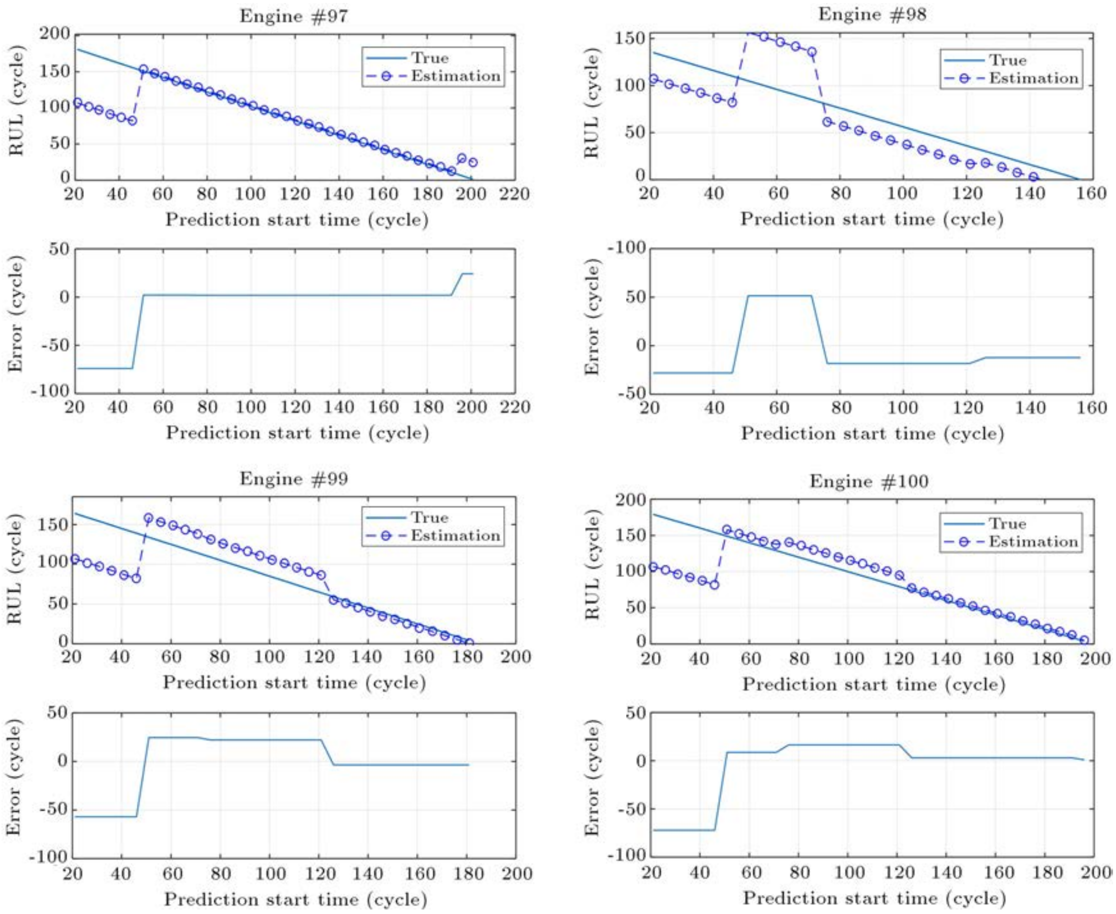

Figure 12. Prognosis results of engines #97 – 100 in different cycles.

References

1. Gao, H., Wang, A., Bai, G., Wei, C., and Fei, C. "Substructure-based distributed collaborative probabilistic analysis method for low-cycle fatigue damage assessment of turbine blade-disk", *Aerospace Science and Technology*, **1**(79), pp. 636–46 (2018).

2. Tahir, M., Aslam, M., Hussain, Z., Abid, M., and Haider Bhatti, S. "Bayesian analysis of heterogeneous doubly censored lifetime data using the 3-component mixture of Rayleigh distributions: A Monte Carlo simulation study", *Scientia Iranica*, Transactions on Industrial Engineering (E), **26**, pp. 1789–1808 (2018).

3. Ding, S., Wang, Z., Qiu, T., Zhang, G., Li, G., and Zhou, Y. "Probabilistic failure risk assessment for aeroengine disks considering a transient process", *Aerospace Science and Technology*, **1**(78), pp. 696–707 (2018).

4. Amezquita-Sanchez, J.P. and Adeli, H. "Feature extraction and classification techniques for health monitoring of structures", *Scientia Iranica, Transactions A, Civil Engineering*, **22**(6), pp. 1931–1940 (2015).

5. Jardine, A.K., Lin, D., and Banjevic, D. "A review on machinery diagnostics and prognostics implementing condition-based maintenance", *Mechanical Systems and Signal Processing*, **20**(7), pp. 1483–1510 (2006).

6. Saxena, A., Goebel, K., Simon, D., and Eklund, N. "Damage propagation modeling for aircraft engine run-to-failure simulation", In *Prognostics and Health Management, PHM 2008, International Conference*, pp. 1–9 (2008).

7. Amezquita-Sanchez, J.P. and Adeli, H. "Optimal tuner selection using Kalman filter for a real-time modular gas turbine model", *Transaction A, Civil Engineering*, **22**(6), p. 1940 (2015).

8. Zhao, W. "A probabilistic approach for prognostics of complex rotary machinery systems", PhD Thesis, University of Cincinnati (2015).

9. Vachtsevanos, G., Wang, P., and Khiripet, N. "Prognostication: algorithms and performance assessment methodologies", In *ATP Fall National Meeting Condition-Based Maintenance Workshop*, pp. 15–17, San Jose, California (1999).

10. Li, L.L., Ma, D.J., and Li, Z.G. "Residual useful life estimation by a data-driven similarity-based approach", *Quality and Reliability Engineering International*, **33**(2), pp. 231–9 (2017).

11. Baraldi, P., Compare, M., Sauco, S., and Zio, E. "Ensemble neural network-based particle filtering for prognostics", *Mechanical Systems and Signal Processing*, **41**(1), pp. 288–300 (2013).

12. Javed, K., Gouriveau, R., and Zerhouni, N. "SW-ELM: A summation wavelet extreme learning machine algorithm with a priori parameter initialization", *Neurocomputing*, **10**(123), pp. 299–307 (2014).

13. Xu, J., Wang, Y., and Xu, L. "PHM-oriented integrated fusion prognostics for aircraft engines based on sensor data", *IEEE Sensors Journal*, **14**(4), pp. 1124–1132 (2014).

14. Yu, J. "Aircraft engine health prognostics based on logistic regression with penalization regularization and state-space-based degradation framework", *Aerospace Science and Technology*, **68**, pp. 345–361 (2017).

15. Simon, D. "A comparison of filtering approaches for aircraft engine health estimation", *Aerospace Science and Technology*, **12**(4), pp. 276–84 (2008).

16. Lu, F., Ju, H., and Huang, J. "An improved extended Kalman filter with inequality constraints for gas turbine engine health monitoring", *Aerospace Science and Technology*, **58**, pp. 36–47 (2016).

17. Son, J., Zhou, S., Sankavaram, C., Du, X., and Zhang, Y. "Remaining useful life prediction based on noisy condition monitoring signals using constrained Kalman filter", *Reliability Engineering & System Safety*, **152**, pp. 38–50 (2016).

18. Zhou, D., Wu, Y., Gao, F., Breaz, E., Ravey, A., and Miraoui, A. "Degradation prediction of PEM fuel cell stack based on multiphysical aging model with particle filter approach", *IEEE Transactions on Industry Applications*, **53**(4), pp. 4041–4052 (2017).

19. Ahsan, S., Lemma, T.A., and Muhammad, M. "Prognosis of gas turbine remaining useful life using particle filter approach", *Mate-rialwissenschaft und Werkstofftechnik*, **50**(3), pp. 336–345 (2019).

20. Tahan, M., Tsoutsanis, E., Muhammad, M., and Karim, Z.A. "Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: A review", *Applied Energy*, **15**, pp. 122–144 (2017).

21. Ragab, A., Yacout, S., Ouali, M.S., and Osman, H. "Pattern-based prognostic methodology for condition-based maintenance using selected and weighted survival curves", *Quality and Reliability Engineering International*, **33**(8), pp. 1753–1772 (2017).

22. Losi, E., Venturini, M., and Manservigi, L. "Gas turbine health state prognostics by means of Bayesian hierarchical models", *Journal of Engineering for Gas Turbines and Power*, **1**, pp. 141–148 (2019).

23. Moghaddass, R. and Zuo, M.J. "An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process", *Reliability Engineering & System Safety*, **124**, pp. 92–104 (2014).

24. Huang, C.C. and Yuan, J. "A two-stage preventive maintenance policy for a multi-state deterioration system", *Reliability Engineering & System Safety*, **95**(11), pp. 1255–1260 (2010).

25. Soro, I.W., Nourelfath, M., and Aït-Kadi, D. "Performance evaluation of multi-state degraded systems with minimal repairs and imperfect preventive maintenance", *Reliability Engineering & System Safety*, **95**(2), pp. 65–69 (2010).

26. Dong, M. and He, D. "Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis", *European Journal of Operational Research*, **178**(3), pp. 858–878 (2007).

27. Nguyen, K.T., Fouladirad, M., and Grall A. "Model selection for degradation modeling and prognosis with health monitoring data", *Reliability Engineering & System Safety*, **169**, pp. 10–16 (2018).

28. De Giorgi, M.G., Ficarella, A., and De Carlo, L. "Jet engine degradation prognostic using artificial neural networks", *Aircraft Engineering and Aerospace Technology*, **92**(3), pp. 296–303 (2019).

29. Lu, F., Wu, J., Huang, J., and Qiu, X. "Aircraft engine degradation prognostics based on logistic regression and novel OS-ELM algorithm", *Aerospace Science and Technology*, **84**, pp. 661–671 (2019).

30. M ul Hassan, M., Danish, F., Yousuf, W.B., and Khan, T.M. "Comparison of different life distribution schemes for prediction of crack propagation in an aircraft wing", *Engineering Failure Analysis*, **1**(96), pp. 241–254 (2019).

31. Huang, H.Z., Wang, H.K., Li, Y.F., Zhang, L., and Liu, Z. "Support vector machine based estimation of remaining useful life: current research status and future trends", *Journal of Mechanical Science and Technology*, **29**(1), pp. 151–163 (2015).

32. Goebel, K., Saha, B., and Saxena, A. "A comparison of three data-driven techniques for prognostics", In *62nd Meeting of the Society for Machinery Failure Prevention Technology (mfpt)*, pp. 119–131 (2008).

33. Razavi, S.A., Najafabadi, T.A., and Mahmoodian, A. "A prognosis methodology based on enhanced lolimot algorithm using historical data", In *2019 Prognostics and System Health Management Conference (PHM-Paris)*, pp. 35–38 (2019).

34. Zarandi, M.F., Faraji, M.R., and Karbasian, M. "An exponential cluster validity index for fuzzy clustering with crisp and fuzzy data", *Scientia Iranica, Transactions E, Industrial Engineering*, **17**(2), pp. 90–95 (2010).

35. Saxena, A. and Goebel, K. "C-MAPSS data set, NASA Ames prognostics data repository", last retrieved from http://ti.arc.nasa.gov/project/ prognostic-d atarepository,NASA Ames, Moffett Field, CA (2008).

36. Ramasso, E. and Saxena, A. "Review and analysis of algorithmic approaches developed for prognostics on CMAPSS dataset", In *Annual Conference of the Prognostics and Health Management Society*, pp. 128–134 (2014).

37. Rezvani, K., Maia, N.M., and Sabour, M.H. "A comparison of some methods for structural damage detection", *Scientia Iranica, Transactions B, Mechanical Engineering*, **25**(3), pp. 1312–1322 (2018).

38. Javed, K., Gouriveau, R., Zemouri, R., and Zerhouni, N. "Features selection procedure for prognostics: An approach based on predictability", *Reliability Engineering & System Safety*, **15**, pp. 165–175 (2017).

39. Le Son, K., Fouladirad, M., Barros, A., Levrat, E., and Iung, B. "Remaining useful life estimation based on stochastic deterioration models: A comparative study", *Reliability Engineering & System Safety*, **12**, pp. 165–175 (2013).

40. Mohammadi, E. and Montazeri-Gh, M. "Simulation of full and part-load performance deterioration of industrial two-shaft gas turbine", *Journal of Engineering for Gas Turbines and Power*, **136**(9), pp. 602–609 (2014).

41. Li, Y.G. and Nilkitsaranont, P. "Gas turbine performance prognostic for condition-based maintenance", *Applied Energy*, **86**(10), pp. 2152–2161 (2009).

42. Diallo, O.N. "A data analytics approach to gas turbine prognostics and health management", Doctoral dissertation, Georgia Institute of Technology (2010).

43. Su, Y., Tao, F., Jin, J., Wang, T., Wang, Q., and Wang, L. "Failure prognosis of complex equipment with multistream deep recurrent neural network", *Journal of Computing and Information Science in Engineering*, **20**(2) (2020).

44. Mahmoodian, A., Durali, M., and Saadat, M. "Investigating different structures for mapping sensor information of a complex mechanical system to its health status", *Proceedings of the 26th ISME Conference*, pp. 110–114 (2018).

45. Ramasso, E. "Investigating computational geometry for failure prognostics", *International Journal of Prognostics and Health Management*, **5**(1), pp. 165–178 (2014).

46. Javed, K., Gouriveau, R., and Zerhouni, N. "Novel failure prognostics approach with dynamic thresholds for machine degradation", *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE*, pp. 4404–4409 (2013).

47. Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., and Schwabacher, M. "Metrics for evaluating performance of prognostic techniques in prognostics and health management", *PHM 2008, International Conference*, pp. 1–17 (2008).

48. Mahmoodian, A., Durali, M., and Saadat, M. "A novel prognostic model of performance degradation based on fusion of current and historical predictions (FCHP)", *Proceedings of the 6th GTC Conference*, pp. 210–212 (2018).

49. Vazirizade, M., Bakhshi, A., and Bahar, O. "Online nonlinear structural damage detection using Hilbert Huang transform and artificial neural networks", *Scientia Iranica, Transactions A, Civil Engineering*, **26**(3), pp. 180–188 (2019).

50. Zarandi, M.F., Faraji, M.R., and Karbasian, M. "An exponential cluster validity index for fuzzy clustering with crisp and fuzzy data", *Scientia Iranica, Transactions E, Industrial Engineering*, **17**(2), pp. 54–69 (2010).

51. Ramasso, E., Rombaut, M., and Zerhouni, N. "Joint prediction of observations and states in time-series: a partially supervised prognostics approach based on belief functions and knn. Networks", *International Conference on PHM*, pp. 11–17 (2013).

52. Khelif, R., Malinowski, S., Chebel-Morello, B., and Zerhouni, N. "RUL prediction based on a new similarity-instance based approach. In Industrial Electronics (ISIE)", *2014 IEEE 23rd International Symposium*, pp. 2463–2468 (2014).

Appendix A

The example illustrated in Figure A.1 indicates that the train is reproduced 3 times for 3 age clusters (n=3)

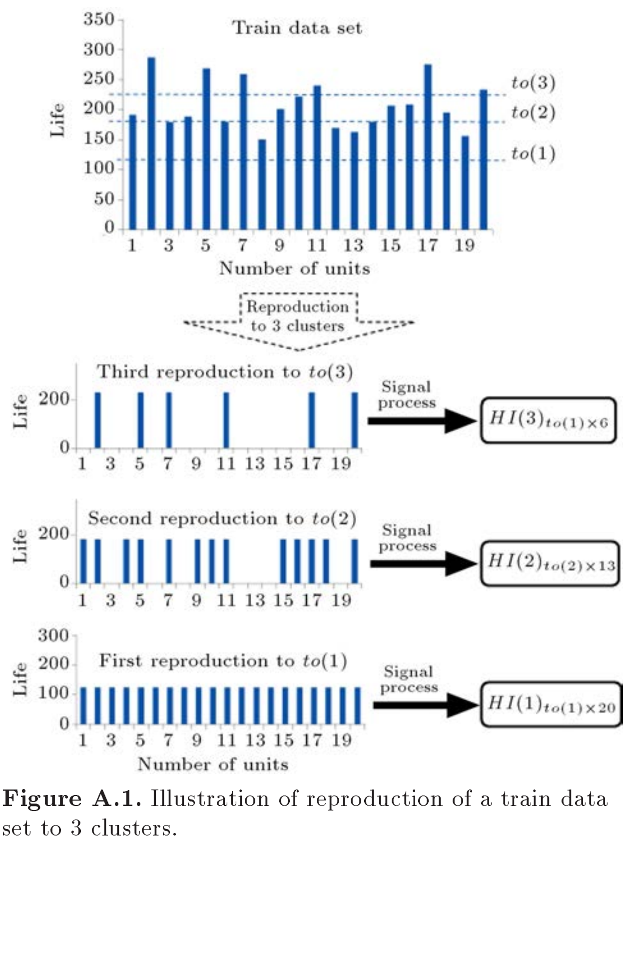

Figure A.1. Illustration of reproduction of a train data set to 3 clusters.

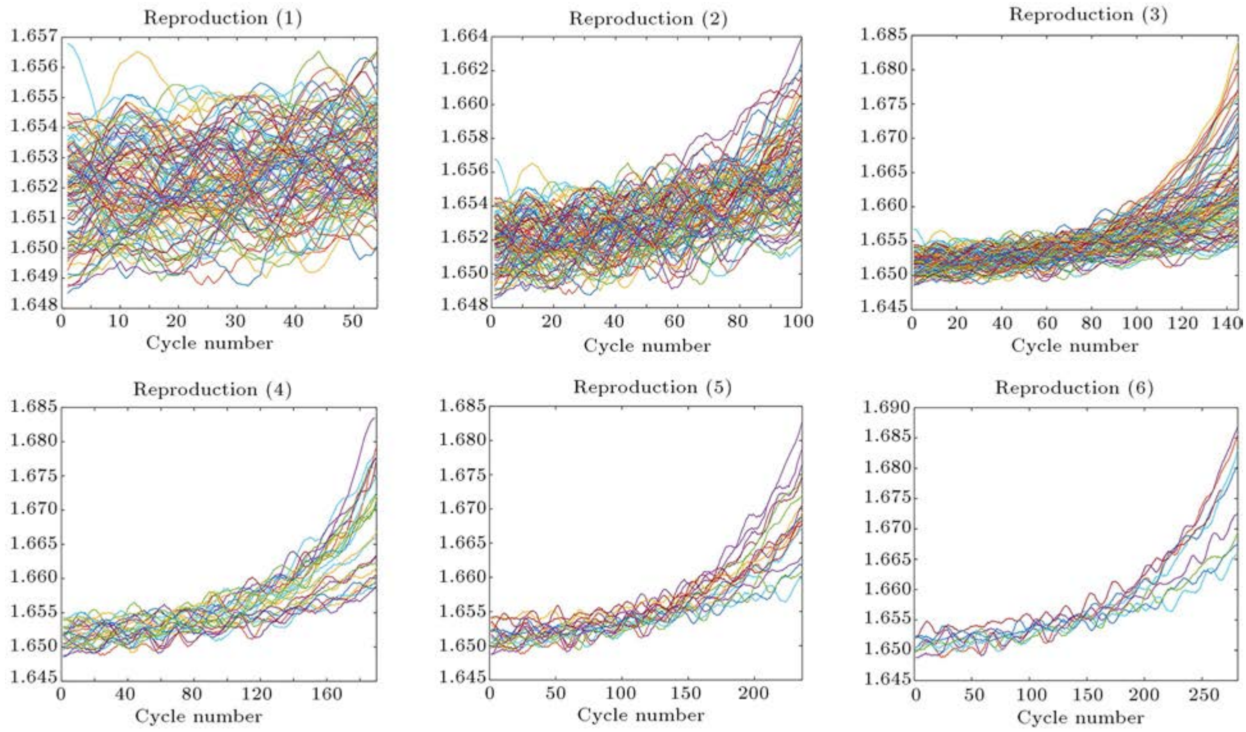

Figure B.1. Health Indicator (HI) signals for train data after clustering.

Appendix B

Here, signals after clustering are shown in Figure B.1.

Biographies

Ali Mahmoodian is a PhD candidate of Mechanical Engineering at Sharif University. He received his BSc degree from KNTU (2010) and his MSc degree in Mechanical Engineering from Tehran University (2007). His research interest is Control, Management, and Prognosis.

Mohammad Durali received his PhD in Aerospace Engineering from MIT (1980). Now, he is a Professor of the Mechanical Engineering Department at Sharif University of Technology. His research interest is Design and Automation, System Dynamics & Modelling, Propulsion Systems.

Tooraj Abbasian Najafabadi received his PhD in Electrical Engineering-Control from the University of Tehran (2010) and now he is an Assistant Professor of Electrical Engineering in Tehran University. His research interest is control and automation.

Mahmoud Saadat Foumani received his PhD degree in Mechanical Engineering from Sharif University of Technology (2002). Now, he is a Professor of the Mechanical Engineering Department at Sharif University. His research interest is new engineering design methods, dynamic systems optimization, vehicle dynamics and chassis design.