Optimized Age Dependent Clustering Algorithm for Prognosis: A Case Study on Gas Turbines

Ali Mahmoodian\textsuperscript{1}, Mohammad Durali\textsuperscript{2}*, Tooraj Abbasian Najafabadi\textsuperscript{3} Mahmood Saadat\textsuperscript{2}

1. PhD candidate, Mechanical Engineering, Sharif University of tech, Tehran, P.O. box 11155-9567, Iran
2. Professor, Mechanical Engineering, Sharif University of tech, Tehran, P.O. box 11155-9567, Iran
3. Assistant professor, ECE faculty, Tehran University, P.O. box 14395-515, Iran

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, neural fitting tool is used. Considering age based clustering concept, determination of main elements of the ADC model is discussed. Genetic algorithm (GA) is used to find the elements of the optimal model. Lastly, fuzzy technique is applied to modify the clustering. The efficacy of the proposed method is demonstrated with a case study on the health monitoring of some turbofan engines. The results show that the concept of clustering even without optimization processes is efficient even for the simplest form of performance. However, by optimizing structure elements and fuzzy clustering, the prognosis accuracy increases up to 71%. The effectiveness of ADC prognosis is proven in comparison with other methods.

Keywords: Age dependent classification; Health monitoring; Prognosis; Genetic Algorithm; Prognostics

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 RUL [1]. As the maintenance in various domains is changing to predictive paradigm, a great portion of research is conducted for health management, prognostics and risk assessment [2-4]. E-maintenance frameworks within Prognosis models are set up on industrial systems. 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]. If a perfect system model is not accessible, the data-driven prognostics method is employed to estimate the RUL. Most efforts focusing on data-driven approaches appear to reflect the difficulty of defining both damage and failure criterion in model-based approaches. Data-driven approaches depend on the availability of run-to-failure data [5]. Therefore, research has been performed on damage propagation modeling for gas turbines that could produce the required data for developing prognostics algorithms [6, 7].

With increase of sensing technologies and data centers, large volumes of data is being collected for PHM purposes. However, the first challenge is how to map the conditions between a complex system with its level of drop. The other challenge in a PHM process is designing the prediction module to estimate 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], combining neural networks and wavelet theory [12] and fusion of prognostics algorithms [13] have been used as prediction tools. Yu et al. have developed a prognostics system based on logistic regression and a state-space-model for engine RUL prediction [14]. Simon et al. have compared the accuracy of estimation and computational effort of variants of the Kalman Filter (KF) like linearized KF, extended KF and unscented KF for aircraft engine health estimation [15]. Lu et al. presented a nonlinear underdetermined state estimation method based on extended KF and showed that the proposed method results in a significant enhancement in relation to underdetermined estimation accuracy and robustness [16]. Son et al. proposed a constrained KF for prediction in the condition that monitoring signals are noisy [17]. In addition to Kalman Filter, 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 methodology combining 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, perfect working state and failure state. Whereas, deterioration process of many real-world cases includes different phases of health conditions, leading to a multistage degradation process [23].

Most of the researches focusing on multistage degradation process suffer from some restrictions that are explained as following. Subsequently, the motivation and contribution of this article are described.

First, multistage structures have usually considered a Markovian deterioration process, with constant transition rate and consequently ignoring the aging and device deterioration over time [24,25]. In other models, such as 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, an algorithm based on age-clustering 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 such as neural network [28], linear or nonlinear regression [29], particle filter [30], support vector machine (SVM) [31, 32] etc. and the main focus 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 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, but yields significant performance improvements.

The outline of this paper is as following. In section 2 The ADC prognostics framework and its implementation are described. In section 3 a case study is presented to demonstrate the application of the results of this paper. This paper ends with the summary and conclusions 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 Fig. 1. The proposed framework is implemented as per the following phases.

a) Clustering

The time to initiate prediction of each test unit is indicated with $tc$. For a prognosis problem with a number of test units, there is an interval of $tc$ [$\min(tc):\max(tc)$]. This interval should be divided to some smaller partitions in the clustering phase. The width of the $i^{th}$ cluster is defined as eq. (1):

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

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

in which $tn$ 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 which should be determined through an optimization process. More details are discussed in the following sections.

b) Reproduction

In the next phase, train data set is reproduced $n$ times for $n$ different age spans. To reach this goal, $n$ observation time points ($to_i$, $i=1:n$) are considered at which feature extraction of train data set is performed. Each observation point is
relevant to a certain cluster \((t_0_i \text{ for } i^{th} \text{ cluster})\). In train data set, for \(i=1:n\), units with life lengths less than \(t_0_i\) are withdrawn; the others are trimmed to \(t_0_i\). 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 \(t_0_i\) is the middle point of \(i^{th}\) interval \([t_{i-1}, t_i]\). However for 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. Appendix A illustrates an example that 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, LOLIMOT etc. is trained in order 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

Until now, the main components of an ADC framework are composed and it is 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 \((t_c)\). The clustering technique is discussed in the following.

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

2-2. Clustering technique

Clustering can be done in 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 \(t_c\) is between the minimum and maximum age of that cluster”.

The main drawback with this approach is that the data points around the clusters boundaries may fall into cluster \((i)\), while they have a higher maturity to 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 such as 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). The number of clusters is an important element of ADC framework and affects other elements. Low number of clusters may decrease the model accuracy, while 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 cluster widths are equal, but usually it is not the best solution. The portions with low dynamic and clear trend can be modeled with 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 health indicator of train data set is observed for each cluster. Observation time of the $i^{th}$ 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 ADC framework are executed, and those elements which best satisfy the selection criteria are selected. Different evolutionary optimization methods such as genetic algorithm, simulated annealing, etc. can be used to find the optimal elements. The objective function may be various prognosis measures like accuracy, mean absolute error, squared absolute error, etc. If fuzzy clustering technique is used, the parameters of membership functions can also be determined by optimization process.

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, there are five extremely popular turbofan engines datasets that have been used in more than 70 publications. Among these datasets, dataset #1 was greatly used (70% of the cases) compared to all other datasets [36]. In dataset #1 two subsets are available: 1) train data set, 2) test data set. The whole data is from a fleet of engines of the same type. The train data is composed from multiple units operated until failure occurred. Other units have operated to different levels of destruction. The data consists of 21 measurements listed in table 1, which are measured during every flight cycle. The goal is to predict the RUL of test units.

3-2 **Health Indicator**
Considering health indicator (HI) design for a data driven PHM process, different methods are used to map the sensor information to the health status of the system. Direct use of all sensor data [37], feature selection [38] and multisensors fusion [39] has been proposed by researchers. 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. The feature selection considers some features that are better predictable and can contribute to prognostic modeling [38]. The effectiveness of the data fusion method has been examined in [42] and it has been used in numerous researches [13]. However, 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 fusion of multi-sensors information represents the HI of the device under study. Data processing and feature extraction is performed through a methodology 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, 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 process for four sensors of engine #1 are indicated in Fig. 3. For more details on this section, we refer the reader to references [42, 44].

3-3 Prognostic measures

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

\[ err = RUL_s - RUL_t \]  \hspace{1cm} (2)

In PHM context, sometimes it is desirable to predict early compared to predicting late. Therefore, the asymmetric interval \( I = [-10, +13] \) around the true RUL is considered to evaluate the performance as shown in Fig. 4. The accuracy measure is the percentage of test units that their RUL estimation falls within the interval \( I \) [45]. This criterion is considered as a severe one compared to the ones defined in the literature [46].

To evaluate the performance of ADC method more accurately, mean square error (MSE) and mean absolute error (MAE) are measured too as equations (3) and (4) [47]:

\[ e_{mse} = \frac{1}{N} \sqrt{\sum_{i=1}^{N} err^2} \]  \hspace{1cm} (3)

\[ e_{mae} = \frac{1}{N} \sum_{i=1}^{N} |err| \]  \hspace{1cm} (4)

3-4 Implementation of ADC framework
The steps taken here to implement the ADC algorithm are based on the framework given in figures 1 and 2. At first, four phases of ADC implantation are performed with initial values for structure elements. Subsequently, optimization is done and the final ADC model is completed.

a) Clustering

Histogram of prediction start time \((t_c)\) for test units is shown in Fig. 5. The minimum and the maximum values are 31 and 303 cycles. The initial values for structure elements are listed in table 2. To select the initial parameters, cluster widths are set equal which is the simplest form to start the algorithm. Also, the midpoints of clusters are considered as the observation points. For the number of clusters, it is shown in following sections 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.

b) Reproduction

The train data set is reproduced for each cluster. For the first cluster, all engines with lives longer than respective observation cycle (54 cycles from table 3), are stopped at the 54th cycle. Then data process is performed and health indicator \(HI_{tr}(1)\) is extracted. For the second cluster, all engines with lives longer than 100 cycles \((t_{o2}=100 \text{ from table } 3)\) are stopped at the 100th cycle and similarly \(HI_{tr}(2)\) is computed. The scheme is repeated for all clusters. The HI signals derived from fusion of multi-sensors from train data set engines before clustering are shown in Fig. 6. The observation times are indicated with dashed lines. HI signals after clustering are shown in appendix B.

Train data set includes 100 engines. All engines are used in the first cluster, so the respective \(HI_{tr}(1)\) vector dimension is 54×100 \((t_{o1}=54 \text{ cycles, Number of reproduction (1) members}=100 \text{ engines})\). For the second cluster, all train data are used too and \(HI_{tr}(2)\) vector is 100×100 \((t_{o2}=100 \text{ cycles, Number of reproduction (2) members}=100 \text{ engines})\). For the third cluster the lives of 4 engines are smaller than 145 cycles and are withdrawn; so the \(HI_{tr}(3)\) vector is 145×96 \((t_{o3}=145 \text{ cycles, Number of reproduction (3) members}=96 \text{ engines})\). Through similar analysis, dimensions of HI vectors for all clusters are determined and summarized in table 4.

c) Mapping

Mapping phase is performed through neural fitting tool of MATLAB. A forward multilayer perceptron (MLP) with backward propagation training algorithm is proposed in this research with neural network structure. The training algorithm is Bayesian. The developed ANN model consists of two layers with hidden sigmoid neurons and linear output neurons as shown in Fig. 7. Two features of each HI signal are inputted to MLP and the output is engine RUL [48, 49]. The summary of the formation of six ANNs is presented in table 5. The training algorithm for all networks is Bayesian regularization. Regression R values indicate 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 fusion of multi-sensors for test data set are shown in Fig. 8. Table 6 shows details of RUL estimation for 15 engines. For each engine, $tc$ is considered firstly. Then the appropriate cluster is determined according to the clustering technique. For simplicity, non-fuzzy rule is 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 ADC method using initial values for structure elements are as in table 7. The measures are calculated for all 100 test units.

3-5 Optimization

In previous section, ADC was implemented with a set of initial values and simple assumptions for the framework, and the prognosis measures were evaluated. In this section, the structure elements are optimized and the final ADC model is built. For this an evaluation data set is required. This data set is used in place of main test data set in middle-phase tests of algorithm to tune its parameters. The evaluation data set is composed from train data set, in a way that train units are stopped some cycles before failure at which prediction should initiate.

It is possible to optimize whole elements simultaneously which leads to numerous possible set of elements and prolonged run time. However, in order to minimize optimization time, the optimum number of clusters is determined at first and subsequently the other elements are optimized. For this purpose, the prognosis measures are calculated while the number of clusters is changed and all other conditions and assumptions are unchanged. As shown in Fig. 9, all prognostic measures stay almost unchanged for $n>2$ for the present case study. Therefore, a low number of clusters is chosen, (for instance 4 clusters) to avoid costly computations in the next optimization steps. Successively, the clusters width ($CW$) and the observation points ($to$) are optimized by genetic algorithm. For this computation the OPTIMTOOL of MATLAB is used. The objective is 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

In order to rectify clustering, fuzzy technique is used [50]. Trapezoidal-shaped membership functions are used for clustering. The fuzzy clustering scheme used in the present paper is shown in Fig. 10.

3-7 Results and discussion

The prognostic measures for different methods are summarized in table 10. All indexes in ADC models are better than ANN without clustering. The prognosis measures are improved significantly with clustering, although the core
algorithm in ADC models is still ANN. The reason being, without age-based clustering, it is absolutely probable that some young units are used for prognosis of an aged or a middle-aged unit, or life prediction of a young unit is performed according to data mining of some old units which in either case reduces the accuracy. It is worth noting that although the present case study uses ANN, the methodology developed in this paper is not limited to deal with ANN, but can also be extended to other estimation methods for improving the results.

To our knowledge, few papers have used full testing dataset. Among them are 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 is performed as indicated in table 11. The accuracy of the proposed ADC approach is better than other approaches. Figure 11 shows the prognostics results of 100 engines for optimized ADC model with fuzzy clustering method.

Finally, prognosis results of engines in different cycles are studied in this paper. The actual RUL value and the RUL estimate are shown in Fig. 12 for engines #97-100. The results show that the RUL estimate is reasonably close to the actual remaining life (especially in older ages). In general, the results indirectly support the hypothesis that, age clustering method lead to acceptable results in prognosis.

4. Conclusion and summary

This research proposed a prognosis algorithm based on ADC. A four-phase prognosis 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, afterwards neural fitting tool was used to build n prediction modules using reproduction sets of train data, and finally cluster selection and RUL estimation of each test unit using the relevant prediction module was completed. Initially, simple assumptions were used to determine the elements of the ADC model. Successively, optimization of the initial prognosis model was considered. The effect of 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 using simple assumptions (53%) was better than the classic methods one (29%), although the core algorithm in both cases was ANN. This is one of the important results of the present study that proves effectiveness of ADC concept. In the present article, artificial neural network 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.

Through an optimization process, results enhancement of the ADC model was achieved (65% and 71% accuracy for non-fuzzy and fuzzy optimized ADC). The optimization goal was maximizing accuracy and the results indicated acceptable accuracy compared to other methods.
One advantage of this algorithm is that the core prediction tool is a conventional methods and the ADC algorithm rectifies that method with more reliable results. Another advantage of the proposed method is that it can deal more reliably with real situations.

Nomenclature

\[
\begin{align*}
T_c & \quad \text{Start time of prediction for a system} \\
CW & \quad \text{Cluster width} \\
tr & \quad (\text{related to}) \text{ Train data set} \\
to & \quad \text{Certain observation points at which the health indicator is observed for each cluster in the training phase.} \\
N & \quad \text{Number of clusters} \\
HI & \quad \text{Health Indicator} \\
\text{ANN} & \quad \text{Artificial Neural Network} \\
\text{PM (i)} & \quad \text{The Prediction Module for the } i^{th} \text{ cluster} \\
\text{RUL} & \quad \text{Remained Useful Life} \\
\text{ADC} & \quad \text{Age Dependent Clustering} \\
\text{PHM} & \quad \text{Prognostics and Health Monitoring} \\
\text{MAE} & \quad \text{Mean Absolute Error} \\
\text{MSE} & \quad \text{Mean Squared Error}
\end{align*}
\]

References

30. Mul Hassan M, Danish F, Yousuf WB, Khan TM. “Comparison of different life distribution schemes for prediction of crack
42. Diallo ON. “A data analytics approach to gas turbine prognostics and health management” (Doctoral dissertation, Georgia Institute of Technology).(2010)
Biographies

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

Mohmmad Durali received his PhD in Aerospace Engineering from MIT (1980). Now, he is a Professor of Mechanical Engineering Department in 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 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 Ph.D. degree in Mechanical Engineering from Sharif University of Technology (2002). Now, he is a professor of Mechanical Engineering Department in Sharif University. His research interest is new engineering design methods, dynamic systems optimization, Vehicle Dynamics & Chassis Design.

Appendix A

Figure a-1 illustrates an example that train data is reproduced 3 times for 3 age clusters (n=3).

Appendix B

Figure a-2 shows HI signals after clustering.
Table captions

Table 1. A List of sensors and Measurements [6]
Table 2. Initial values for structure elements
Table 3. Summary of engines clustering into 6 groups
Table 4. Summary of six reproductions of train data set
Table 5. Summary of the formation of the prediction modules
Table 6. Details of RUL estimation for 15 units (initial ADC model)
Table 7. Summary of prognostic measures for ADC model with initial values
Table 8. Optimized values for structure elements
Table 9. Summary of optimum clustering scheme
Table 10. Summary of prognostic measures
Table 11. Comparison of accuracy for different methods

Figure captions

Fig. 1. The ADC framework
Fig. 2. The summary of the elements selection approach.
Fig. 3. Different stages of data process for Engine #1 [44]
Fig. 4. Prognostic measure [45]
Fig. 5. Histogram of prediction start time ($t_c$) for test units
Fig. 6. HI signals for train data set before clustering (dashed lines indicate observation cycles)
Fig. 7. MLP framework
Fig. 8. HI derived from fusion of multi-sensors for 100 test data
Fig. 9. Prognostic measures via. number of clusters
Fig. 10. The fuzzy clustering scheme used in the present paper
Fig. 11. Results of engine prognostics with fuzzy optimized ADC model
Fig. 12: Prognosis results of engines #97-100 in different cycles
Fig. a-1. Illustration of reproduction of a train data set to 3 clusters
Fig. a-2. HI signals for train data after clustering
### Table 1. A List of sensors and Measurements [6]

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>T24</td>
<td>Total temperature at LPC outlet</td>
</tr>
<tr>
<td>T30</td>
<td>Total temperature at HPC outlet</td>
</tr>
<tr>
<td>P30</td>
<td>Total pressure at HPC outlet</td>
</tr>
<tr>
<td>Nc</td>
<td>Physical core speed</td>
</tr>
<tr>
<td>Pr</td>
<td>Engine pressure ratio (P50 / P2)</td>
</tr>
<tr>
<td>Phi</td>
<td>Ratio of fuel flow to Ps30</td>
</tr>
<tr>
<td>BPR</td>
<td>Bypass ratio</td>
</tr>
<tr>
<td>BE</td>
<td>Bleed enthalpy</td>
</tr>
<tr>
<td>T50</td>
<td>Total temperature at LPT outlet</td>
</tr>
<tr>
<td>Ps30</td>
<td>static pressure at HPC outlet</td>
</tr>
<tr>
<td>farB</td>
<td>Burner Fuel air ratio</td>
</tr>
</tbody>
</table>

### Table 2. Initial values for structure elements

<table>
<thead>
<tr>
<th>Structure elements</th>
<th>Initial value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of clusters</td>
<td>6</td>
</tr>
<tr>
<td>Cluster widths (cycle)</td>
<td>[45, 45, 45, 45, 45, 45]</td>
</tr>
<tr>
<td>Observation points (cycle)</td>
<td>[54, 100, 145, 190, 236, 281]</td>
</tr>
</tbody>
</table>

### Table 3. Summary of engines clustering into 6 groups

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Observation time (to)</th>
<th>age interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>54th cycle</td>
<td>31-76cycles</td>
</tr>
<tr>
<td>II</td>
<td>100th cycle</td>
<td>77-122cycles</td>
</tr>
<tr>
<td>III</td>
<td>145th cycle</td>
<td>123-167cycles</td>
</tr>
<tr>
<td>IV</td>
<td>190th cycle</td>
<td>168-212cycles</td>
</tr>
<tr>
<td>V</td>
<td>236th cycle</td>
<td>213-258cycles</td>
</tr>
<tr>
<td>VI</td>
<td>281th cycle</td>
<td>259-303cycles</td>
</tr>
</tbody>
</table>
Table 4. Summary of six reproductions of train data set

<table>
<thead>
<tr>
<th>Reproduction</th>
<th>Number of members</th>
<th>Vector</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>100</td>
<td>$H_{1r}(1)$</td>
<td>54×100</td>
</tr>
<tr>
<td>II</td>
<td>100</td>
<td>$H_{1r}(2)$</td>
<td>100×100</td>
</tr>
<tr>
<td>III</td>
<td>96</td>
<td>$H_{1r}(3)$</td>
<td>145×96</td>
</tr>
<tr>
<td>IV</td>
<td>62</td>
<td>$H_{1r}(4)$</td>
<td>190×62</td>
</tr>
<tr>
<td>V</td>
<td>19</td>
<td>$H_{1r}(5)$</td>
<td>236×19</td>
</tr>
<tr>
<td>VI</td>
<td>8</td>
<td>$H_{1r}(6)$</td>
<td>281×8</td>
</tr>
</tbody>
</table>

Table 5. Summary of the formation of the prediction modules

<table>
<thead>
<tr>
<th>ANN</th>
<th>Regression R Values of Net</th>
<th>Input</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>27%</td>
<td>2x100 matrix, representing 2 features of 100 HI signals</td>
<td>1x100 matrix, representing RULs of 100 engines</td>
</tr>
<tr>
<td>II</td>
<td>59%</td>
<td>2x100 matrix, representing 2 features of 100 HI signals</td>
<td>1x100 matrix, representing RULs of 100 engines</td>
</tr>
<tr>
<td>III</td>
<td>85%</td>
<td>2x96 matrix, representing 2 features of 96 HI signals</td>
<td>1x96 matrix, representing RULs of 96 engines</td>
</tr>
<tr>
<td>IV</td>
<td>99%</td>
<td>2x62 matrix, representing 2 features of 62 HI signals</td>
<td>1x62 matrix, representing RULs of 62 engines</td>
</tr>
<tr>
<td>V</td>
<td>99%</td>
<td>2x19 matrix, representing 2 features of 19 HI signals</td>
<td>1x19 matrix, representing RULs of 19 engines</td>
</tr>
<tr>
<td>VI</td>
<td>99%</td>
<td>2x8 matrix, representing 2 features of 8 HI signals</td>
<td>1x8 matrix, representing RULs of 8 engines</td>
</tr>
</tbody>
</table>
Table 6. Details of RUL estimation for 15 units (initial ADC model)

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

Table 7. Summary of prognostic measures for ADC model with initial values

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (%)</td>
<td>53</td>
</tr>
<tr>
<td>MAE</td>
<td>14.9</td>
</tr>
<tr>
<td>MSE</td>
<td>1.9</td>
</tr>
</tbody>
</table>

Table 8. Optimized values for structure elements

<table>
<thead>
<tr>
<th>Structure elements</th>
<th>Best value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of clusters</td>
<td>4</td>
</tr>
<tr>
<td>Cluster widths(cycle)</td>
<td>[44 49 70 104]</td>
</tr>
<tr>
<td>Observation cycles</td>
<td>[53 100 159 208]</td>
</tr>
</tbody>
</table>
Table 9. Summary of optimum clustering scheme

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Observation cycle (to)</th>
<th>age interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>53th cycle</td>
<td>31-75 cycles</td>
</tr>
<tr>
<td>II</td>
<td>100th cycle</td>
<td>76-125 cycles</td>
</tr>
<tr>
<td>III</td>
<td>159th cycle</td>
<td>126-196 cycles</td>
</tr>
<tr>
<td>IV</td>
<td>208th cycle</td>
<td>199-303 cycles</td>
</tr>
</tbody>
</table>

Table 10. Summary of prognostic measures

<table>
<thead>
<tr>
<th>Method</th>
<th>Remark</th>
<th>Accuracy (%)</th>
<th>MAE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>Without clustering</td>
<td>29</td>
<td>27.5</td>
<td>4.17</td>
</tr>
<tr>
<td>Initial ADC model</td>
<td>Simplest form</td>
<td>53</td>
<td>14.9</td>
<td>1.9</td>
</tr>
<tr>
<td>Optimized ADC Model</td>
<td>Non-fuzzy clustering</td>
<td>65</td>
<td>14.9</td>
<td>2</td>
</tr>
<tr>
<td>Final ADC model</td>
<td>Fuzzy clustering</td>
<td>71</td>
<td>12.8</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Table 11. Comparison of accuracy for different methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Remark</th>
<th>Correct %</th>
<th>Early %</th>
<th>Late %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuzzy optimized ADC model</td>
<td>Tested on 100 test units.</td>
<td>71</td>
<td>23</td>
<td>6</td>
</tr>
<tr>
<td>Ramasso [45]</td>
<td>Tested on 100 test units.</td>
<td>67</td>
<td>Nan</td>
<td>Nan</td>
</tr>
<tr>
<td>Non-fuzzy ADC model</td>
<td>Tested on 100 test units.</td>
<td>65</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>Khelif et al.[52]</td>
<td>Tested on 100 test units.</td>
<td>54</td>
<td>18</td>
<td>28</td>
</tr>
<tr>
<td>Ramasso et al. [51]</td>
<td>Tested on 100 test units.</td>
<td>53</td>
<td>36</td>
<td>11</td>
</tr>
<tr>
<td>Javed et al. [46]</td>
<td>Tested only on 15 test units.</td>
<td>53</td>
<td>27</td>
<td>20</td>
</tr>
<tr>
<td>Wang et al. [13]</td>
<td>Sensor selection proposed by the authors.</td>
<td>44</td>
<td>19</td>
<td>37</td>
</tr>
<tr>
<td>Wang et al. [13]</td>
<td>Sensor selection used in [51]</td>
<td>50</td>
<td>19</td>
<td>31</td>
</tr>
</tbody>
</table>
Fig. 1. The ADC framework
Fig. 2. The summary of the elements selection approach.
(a) Normalized sensors
Fig. 3. Different stages of data process for Engine #1 [44]

Fig. 4. Prognostic measure [45]
Fig. 5. Histogram of prediction start time ($t_c$) for test units

Fig. 6. HI signals for train data set before clustering (dashed lines indicate observation cycles)
Fig. 7. MLP framework

Fig. 8. HI derived from fusion of multi-sensors for 100 test data
Fig. 9. Prognostic measures via. number of clusters

Fig. 10. The fuzzy clustering scheme used in the present paper
(a) Part 1: motor numbers 1:35

(b) Part 2: motor numbers 36:70
Part 3: motor numbers 71:100

Fig. 11. Results of engine prognostics with fuzzy optimized ADC model
Fig. 12: Prognosis results of engines #97-100 in different cycles
Fig. a-1. Illustration of reproduction of a train data set to 3 clusters
Fig. a-2. HI signals for train data after clustering