

Concurrent Estimation of Remaining Useful Life for Multiple Faults in an Ion Etch Mill: A Data-driven Approach

Kuldeep Singh¹, Balaji Selvanathan², Kalyani Zope³, Sri Harsha Nistala⁴, Venkataramana Runkana⁵

^{1,2,3,4,5}*TCS Research, Tata Research Development and Design Centre, Tata Consultancy Services Ltd., Pune, Maharashtra, 411013, India*

kuldeep.s5@tcs.com

balaji.selvanathan@tcs.com

kalyani.zope@tcs.com

sriharsha.nistala@tcs.com

venkat.runkana@tcs.com

ABSTRACT

The 2018 PHM Data Challenge posed the problem of estimating Remaining Useful Life (RUL) for multiple faults in ion etch mills. As with any industrial system, run-to-failure data for the mills is not directly available and the mills experience more than one fault at the same time. We propose a novel data-driven methodology to address these challenges and develop a workflow that can be used for concurrent estimation of RUL for multiple faults in ion etch mills in real time. In the proposed approach, operational data of the ion etch mill is used to build a machine learning model for predicting a health score of the mill and to create a library of truncated degradation curves for each fault. These are then utilized for RUL predictions using Dynamic Time Warping (DTW) curve matching. Application of the proposed approach to test and validation datasets provided during the data challenge showed reasonable agreement between RUL predictions and the ground truth. The approach proposed here can be extended to other industrial systems and equipment for which historical operational data and failure information is available. This framework will help optimize health management and pave the way for predictive maintenance of industrial equipment.

1. INTRODUCTION

Process anomalies and equipment failure are major areas of concern for manufacturing and process industries. Process

anomalies may be triggered due to disturbances in normal operating conditions and/or due to operator response to process disturbances. If such anomalies are not detected and arrested at an early stage, they may lead to abnormal events (e.g. uncontrollable operation) or accidents. Similarly, faults in mechanical equipment such as blowers, pumps, compressors, valves, etc. that occur due to aging, wear and tear, fatigue or abnormal changes in the operating environment may lead to failure if the faults are not detected and addressed as early as possible.

Industries typically follow a preventive maintenance strategy wherein repairs and replacement of components in equipment are carried out at periodic intervals (e.g. every 4 months) that are decided on the basis of historical failure data of assets. Preventive maintenance, however, is not an optimum strategy as perfectly healthy components may be replaced during maintenance and equipment failure may occur in the gap between two maintenance tasks. This shortcoming led industries to turn to predictive maintenance or Condition Based Maintenance (CBM) wherein the health of equipment is assessed using sensor data and maintenance is scheduled only when the equipment health falls below a certain threshold. While predictive maintenance is not an entirely new concept (Jardine et al., 2006), there is renewed interest in the subject due to the emergence of the Industrial Internet of Things (IIoT) and the availability of enormous amount of data collected by equipment and process sensors, mobile and wireless logs, software logs, cameras, microphones and wireless sensor networks at a high frequency (Qin, 2014).

Predictive maintenance consists of the following key steps – continuous monitoring of equipment using sensor data, detection and diagnosis (i.e. classification or localization) of faults, and estimation of RUL of the faulty component.

Kuldeep Singh et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Reliable and accurate prediction of RUL of components and physical systems helps in scheduling maintenance activities optimally and in managing the health of the system. In this context, the problem posed by the 2018 PHM Data Challenge (PHM Data challenge, 2018) is very pertinent to manufacturing and process industries.

The 2018 PHM Data Challenge involved analyzing the fault behavior of Ion Mill Etch Tools (IMET) and predicting the time-to-failure or RUL of the tools due to three different faults. In an IMET, high intensity beams of charged ions are used to etch metallic, non-metallic or semi-conductor wafers mounted on a rotating fixture in a vacuum enclosure. It contains a cooling system known as ‘FlowCool’ that is used for cooling the wafers during the etching process. The cooling system passes helium gas behind the wafer at a specified flow rate and the heated helium gas is indirectly cooled by a separate water cooling system (PHM Data challenge, 2018). Any failure in the FlowCool system could damage the wafer during etching. Hence, a prognostics and health management system would be useful for predicting the failure of the FlowCool system a priori so that appropriate control actions or maintenance activities can be undertaken to prevent wafer damage. The objective of the data challenge is to predict the RUL of the FlowCool system due to three different faults, namely FlowCool Pressure Dropped Below Limit, FlowCool Pressure Too High and FlowCool Leak.

This year’s data challenge is different from previous data challenges viz. PHM’08, PHM’10, IEEE’12 and IEEE’14 (Huang et al., 2017) in two significant ways. Firstly, while run-to-failure data from healthy state to failure of the system was made available for training purposes in the previous data challenges, regular operational data of the IMET is made available in this year’s data challenge. Secondly, while previous data challenges dealt with only one type of fault/failure, this year’s data challenge deals with three types of faults where one or more faults could occur simultaneously at any given time.

We propose a novel data-driven approach to tackle non-availability of run-to-failure data and the presence of multiple faults in the system for concurrent estimation of RUL for multiple faults. The data-driven approach was chosen due to the abundance of IMET operational data. In our approach, multiple health score models were built from continuous operational data and were used to create a library of truncated degradation curves for each fault. The estimation of RUL for each fault was then carried out using a curve matching technique similar to the one proposed by Wang et al. (2008). According to (Wang et al., 2008), the RUL of the test unit is read as the RUL of the training unit from the point where the degradation pattern of the test unit has the best match with the degradation patterns of the training units.

The rest of the paper is organized as follows: The objective of the 2018 PHM Data Challenge and the dataset provided are discussed in Section 2. Insights from data exploration are

provided in Section 3. The RUL estimation methodology is described in Section 4 and the RUL prediction results are presented in Section 5. Finally, conclusions and future work are discussed in Section 6.

2. DETAILS OF THE DATA CHALLENGE

As mentioned in Section 1, in an IMET, for which the dataset was provided, the FlowCool system could be affected by three faults, namely, FlowCool Pressure Dropped Below Limit, FlowCool Pressure Too High and FlowCool Leak. One, two or all the three faults could occur simultaneously at any given time. The objective of the data challenge is to predict the time remaining until the occurrence of next fault, that is, to predict the RUL of the tool, for each of the above mentioned three faults. The RUL prediction should be done by building a model from time series sensor data collected from various IMETs operating under various conditions and settings. The predictions of time-to-failure at a specific time should only use time-series data from current and past times (PHM Data challenge, 2018).

The training dataset consists of multivariate time series that are collected from 20 different IMETs during operation. Each tool may be at a different level of degradation with respect to the three faults and this is unknown. Training data also consists of the times of occurrences of faults and the category of faults for each of the 20 IMETs. It is understood that the initiation of the fault could have happened much earlier than the provided fault time. Along with the fault time information, the Time-To-Failure (TTF) or RUL that indicates the time remaining until the next failure at each time step is also provided. This serves as the ground truth for training.

Time series data from 5 IMETs (a subset of the 20 IMETs from the training data) without the RUL information was also made available for testing the models developed using the training data. Validation data provided towards the end of the competition also consisted of time series data from the same 5 IMETs.

The predicted RULs are evaluated using a score computed as per the rules shown in Table 1. Two scores, namely, primary and secondary, are used to assess the submitted RUL (SUB) predictions by comparing them with the ground truth RUL (GT) of the validation data. The secondary score is similar in nature to the primary score, but it penalizes the false positives and false negatives more, as shown in Table 1. Both primary and secondary scores have a sub-score for RUL prediction at each time step. The sub-scores for each prediction are summed and divided by the total number of time steps in the validation data to arrive at the final score.

Table 1. Rules for computation of score.

Ground Truth TTF (GT)	Submission TTF (SUB)	Primary Score	Secondary Score
Number	Number	$e^{(0.001*GT)} * GT - SUB $	$0.1 * (GT - SUB) * (GT - SUB)$
Not-a-Number (NaN)	Number	$e^{(0.001*SUB)} * SUB$	$\frac{5}{ SUB + 3}$
Number	NaN	$e^{(0.001*GT)} * GT$	$20 * \frac{1}{e^{ GT +0.1}}$
NaN	NaN	0	0

3. DATA EXPLORATION

The multivariate time series data in training, test and validation datasets consists of 23 parameters, 5 of which are categorical and the rest 18 are numeric variables. Important categorical variables include Wafer ID, Recipe, and Recipe Step. IMETs work in a batch operation mode, that is, the tool is stopped between etching of two different wafers. Wafer ID refers to the wafer that is being etched at any given point in time. While ‘Recipe’ refers to the combination of settings that are used for etching a wafer, ‘Recipe Step’ refers to the process step in a particular recipe. Analysis of recipes and recipe steps from all the training files revealed that there are 347 unique recipes and that a given recipe number (e.g. 300) can have different number of recipe steps. There are 1568 unique recipe-recipe step combinations across all training tools. One wafer is subjected only to one etch recipe in one batch. The same wafer may be subjected to different etch recipes at different times to achieve the desired etch pattern. On the basis of these observations, we divide the IMET operational data into ‘wafer-level data sequences’ using the Wafer ID. These sequences were used for RUL estimation.

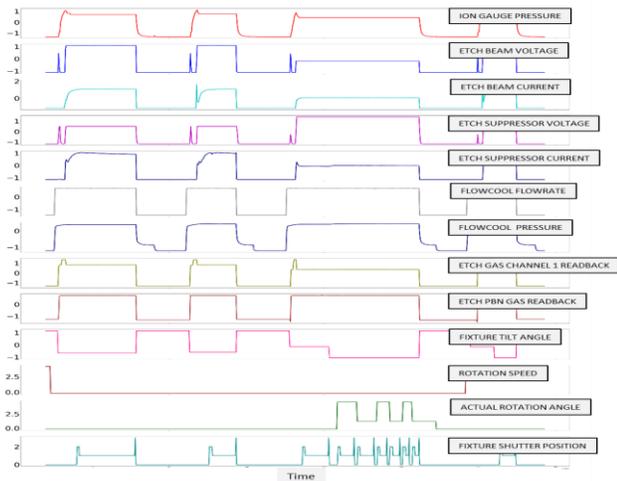


Figure 1. Sample trends of selected parameters for tool #01_M02

Fig. 1 shows the trends of the numeric variables across three wafers for the tool 02_M01. Analysis of the trends revealed the sequence of operations in an IMET: Circulation of FlowCool is started following which the flow of argon into the ion source and the Particle Beam Neutralizer (PBN) assembly, and the vacuum of the ion mill chamber is started. When the chamber is sufficiently close to vacuum, etch beam current and etch suppressor are passed and the etching of the wafer begins. The currents, voltages, fixture tilt angle, fixture shutter position, and flow rate and pressure of FlowCool vary during the etching process depending on the recipe step that is in operation. The rotation speed was found to be constant for a majority of the time. At the end of the etching process, when the tool is switched off, the currents, voltages, argon flow rates and the FlowCool flow rate drop almost instantaneously. The FlowCool pressure and the vacuum in the chamber decrease slowly after the tool is switched off. Any deviation from the established sequence of operations may be indicative of faults in the tool.

Based on our understanding of the ion etching process and working of ion etch mill tools, we have selected 13 out of the 18 parameters such as Ion Gauge Pressure, Etch Beam and Etch Suppressor Voltage and Currents, FlowCool Flow rate and Pressure, Fixture Tilt Angle and Rotation Speed as important process parameters for RUL estimation. The remaining 5 parameters were ignored as they are counter variables.

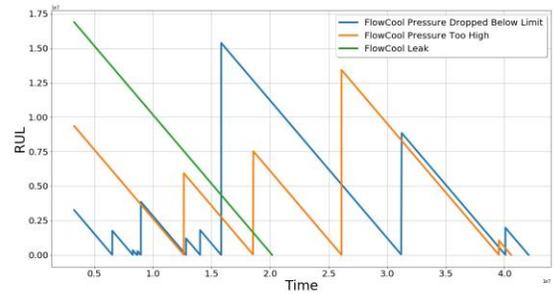


Figure 2. Trends of ground truth RUL for tool #01_M01

As mentioned in Section 2, multiple faults could exist in an IMET at the same time. The ground truth RUL trends for all the faults for tool #01_M01 are shown in Fig. 2 and it can be observed that RUL predictions exist for more than one fault at any given point of time. Therefore, our approach provides for concurrent estimation of RUL for multiple faults.

4. RUL ESTIMATION METHODOLOGY

The RUL estimation approach consist of two phases, viz. training and testing phase. Training was carried out using operational data only from those IMETs for which test datasets were provided (Table 2). Training was carried out separately for each IMET. This is because the operational

behaviour of two IMETs could be different as each IMET is handling different recipes. Merging data from various IMETs could lead to contamination of operational behaviour. For this reason, operational data from each individual IMET was used separately for training.

Table 2. Datasets used for training.

Datasets used for Training	Test Datasets
01_M02	01_M02
02_M02	02_M02
03_M01	03_M01
04_M01	04_M01
06_M01	06_M01

4.1 Training Phase

The sequence of steps followed in the training phase are shown in Fig. 3 and explained below.

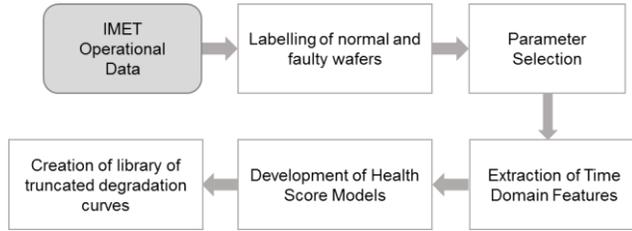


Figure 3. Sequence of steps in the training phase

1) Labelling of normal and faulty wafers

For each IMET used for training, the operational data was divided into ‘wafer-level data sequences’ based on Wafer ID as mentioned in Section 3. The wafers were categorised as normal, Fault #1 (FlowCool Leak), Fault #2 (FlowCool pressure dropped below limit) and Fault #3 (FlowCool pressure too high). Wafers for which RUL data for any of the three faults is not available were considered normal. For each fault type, few wafers before a given fault time period are labelled as faulty (according to corresponding fault type). Some wafers were not used for training.

2) Parameter Selection

As discussed in Section 3, 13 important process parameters were selected for RUL estimation. This selection was based on our understanding of the ion etching process and analysis of variable trends of all the available parameters.

3) Extraction of Time Domain Features

Preliminary attempts to build health score regression models using raw sensor data did not result in models with a good accuracy. Hence, time domain features such as mean, standard deviation, peak, Root Mean Square (RMS), kurtosis, skewness, crest factor and shape factor were extracted to derive underlying information from the raw data. A fixed

window size of 100 instances with a window shift of 50% was used. Windowing and extraction of time domain features was done for all labelled wafers. Time domain features were computed for each of the 13 selected parameters. Each instance of time domain features takes the corresponding label of its wafer. All the 104 features (13 parameters \times 8 features) were used for building health score models.

4) Development of Health Score Models

Regression models for health score of the IMET were built using the 104 features by labelling the faulty wafers as **1** and the normal wafers as **0**. Rather than having a single score model for all faults, we have modelled each fault separately to aid in concurrent RUL prediction for multiple faults. Thus, three health score models exist for each tool. Instead of assuming the form of the degradation curve and choosing the form of the health score model, several machine learning models – linear as well as nonlinear – were trained in order to arrive at the best health score model for each fault.

The following models were trained: Logistic Regression model, Generalized Linear Models (GLM) with Gaussian and Gamma families, Multivariate Adaptive Regression Splines (MARS), Support Vector Regression (SVR), Random Forest (RF), MultiLayer Perceptron (MLP) model and Gradient Boosted Machine. Of these, the Random Forest model, with its capability to handle imbalanced data, was found to have the highest model accuracy and lowest mean square error for all the faults across all IMETs. The RF health score model has the following form;

$$y_i^k = f_k(X_i) \quad (1)$$

where y_i^k is the health score corresponding to k^{th} fault and i^{th} window, X_i is the 1×104 matrix of time domain features for i^{th} window, f_k is the functional form of the RF model relating X_i to y_i^k .

The health score is a bounded value that is indicative of the health of the IMET with respect to each fault. A value closer to 0 indicates healthy/normal state while a value closer to 1 indicates faulty state. Further, an increase of health score from 0 towards 1 indicates an increasing faulty trend.

5) Creation of library of truncated degradation curves

The RF health score models for each fault were used to predict health score for every instance of time domain features extracted for all the wafers in an IMET (normal wafers, faulty wafers and the wafers not included in training). The health score values thus obtained from all the wafers were smoothed using exponential weighted moving average technique to obtain continuous health score curves similar to the one shown in Fig. 4.

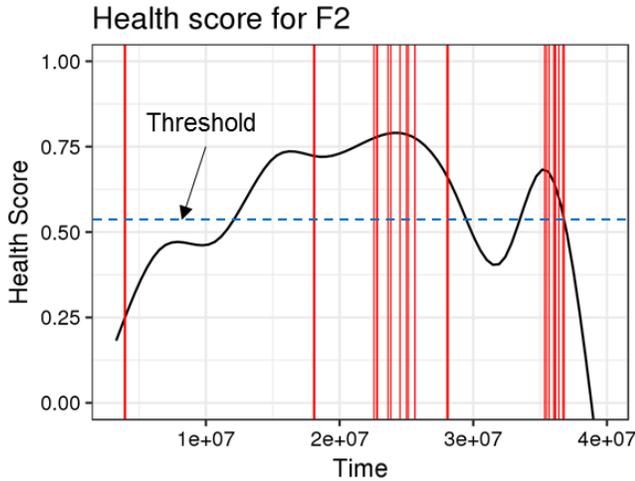


Figure 4. Health score curve for fault #2 in tool #01_M02 (Vertical red lines indicate failure times due to fault #2)

Based on visual analysis of health score curves, a threshold (T) on health score was chosen, beyond which the signature of the fault was considered significant. It can be seen from Fig. 4 that health score is higher than the threshold (0.55) at most of the failure times compared to the rest of the time. For each fault, all the sequences of health scores above the threshold and their corresponding RULs were extracted and stored. These form the library of ‘truncated degradation curves’ for a given fault and IMET. These are called truncated degradation curves as they do not start from a health score of zero because such complete run-to-failures are not always available in operational data. The truncated degradation curves for the three faults obtained for tool # 01_M02 are shown in Fig. 5, 6 and 7.

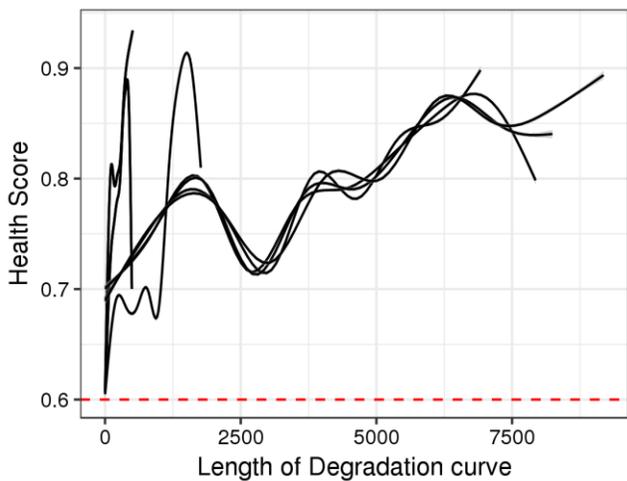


Figure 5. Degradation curves for fault #1 for tool #01_M02

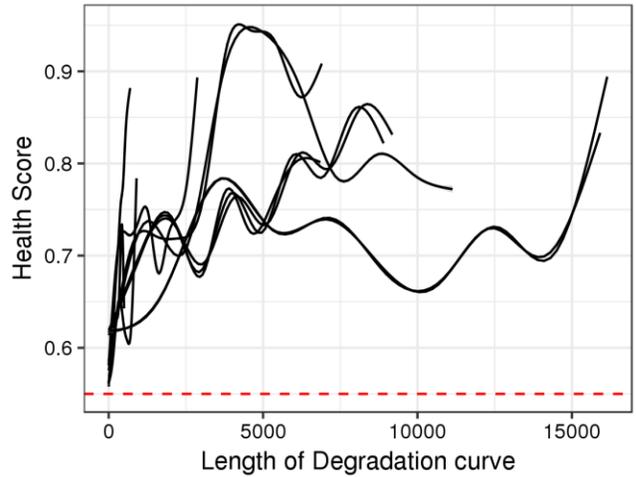


Figure 6. Degradation curves for fault #2 for tool #01_M02

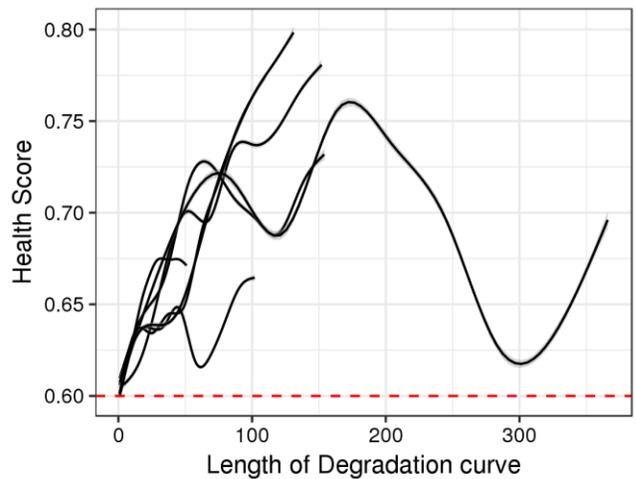


Figure 7. Degradation curves for fault #3 for tool #01_M02

It can be observed that fault #3 propagates faster than fault #1 which propagates faster than fault #2.

4.1. Testing Phase

The sequence of steps followed in the testing phase are shown in Fig. 8 and described below. This is used for estimation of RUL on test dataset or validation dataset for each of the faults at any given time.

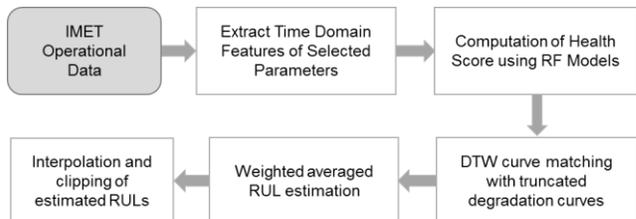


Figure 8. Sequence of steps in the testing phase

1) Extraction of Time Domain Features of Selected Parameters

The IMET operational data from the testing/validation dataset was divided into ‘wafer-level data sequences’, and time domain features for 13 parameters were extracted for each window in the wafer as mentioned in the training phase. In this case, wafers were not labelled and all wafers were considered for RUL estimation.

2) Computation of Health Score using RF Models

The RF health score models, one for each fault, were used to predict the health scores using the time domain features. Each model will signify the health of the IMET with respect to the corresponding fault at any given time. If the health score overshoots the threshold for N consecutive number of times for a particular fault, it is considered as initiation of a particular fault and RUL estimation for the fault is triggered. Here, N is a hyper-parameter that is tuned (e.g. 5 to 7) for each fault and each tool to improve the RUL prediction score on test data.

3) DTW Curve Matching with Truncated Degradation Curves

The health score sequence that has overshoot the threshold N number of times is extracted and its closeness to the library of truncated degradation curves was computed using the Dynamic Time Warping (DTW) distance. The matching was performed using the fixed length-sliding window approach over each degradation curve in the library. Fig. 9 shows a faulty test score sequence overlaid on the library of degradation curves while computing the DTW distance. The DTW distance is chosen over traditional Euclidean distance as DTW is better suited to compare the shapes of two curves (Gu & Jin, 2006). The window for which the DTW distance between faulty test score sequence and a degradation curve is considered the best match, and its corresponding RUL value in the library is taken as the RUL estimate for the faulty test score sequence. Hence, for every degradation curve in the library, we get the closest distance and a corresponding RUL estimate.

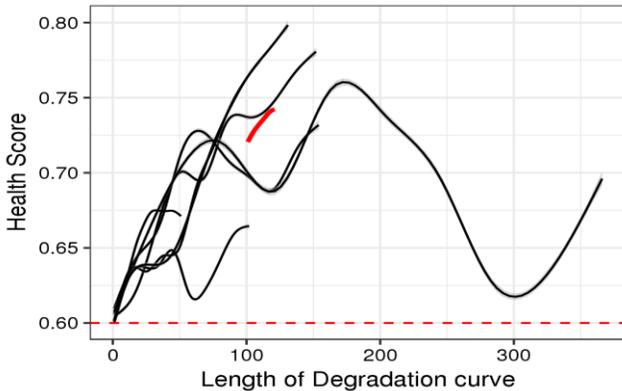


Figure 9. Faulty test score sequence (red) overlaid on the degradation curves for fault #3 for tool #01_M02

4) Weighted Average RUL Estimation

The final RUL estimate was calculated as the weighted average of the individual RUL estimates using Eq. (2). Weights were assumed to be inversely proportional to the closest distances.

$$RUL|_t = \frac{\sum_i^n \left(\frac{1}{d_i} \times RUL_i\right)}{\sum_i^n \left(\frac{1}{d_i}\right)} \quad (2)$$

where $RUL|_t$ is the final RUL estimate at time t , RUL_i is the RUL estimate obtained from i^{th} degradation curve, d_i is the closest distance obtained from the i^{th} degradation curve.

This process was repeated as we progress through all the windows of the testing/validation data. RULs corresponding to each window and each fault were thus estimated.

5) Interpolation and Clipping of estimated RULs

Since the RULs were estimated at a window level instead of the instance level, there will be gaps in the RUL estimates at various times. In order to fill these gaps, we have performed linear interpolation between two discrete times at which the RUL estimates are available, provided the two consecutive values of time are within twice the distance of the length of faulty test score sequence, that is, $2N$.

We have noticed that RUL is over-predicted in some cases during cross validation with training data. This is handled by clipping the RUL predictions for a particular fault to the limit of $0.75 \times (\max(RUL) - \min(RUL))$.

5. RESULTS & DISCUSSION

The RUL estimation methodology proposed in Section 4 is initially validated on the training dataset by using 80% of it for training and 20% of it for testing, and found to give reasonable RUL predictions on the 20% testing set. The models developed during the training phase and the approach were applied for predicting RULs for the 5 files in the test and validation datasets provided during the competition. The scores for these RUL predictions calculated using the rules mentioned in Section 2 are shown in Table 3.

Table 3. RUL prediction scores on test and validation datasets.

Scoring function	Score on Test dataset	Score on Validation dataset
Primary	~100	96.45
Secondary	-	1.78e+06
Average	-	8.92e+05

It can be seen that while the primary scores on the test and validation datasets are reasonable, the secondary score on the validation dataset is quite high. While this may be partially due to the ‘squared error’ form of the secondary score, it could also be due to erroneous RUL predictions. To verify this, the ground truth (revealed after the competition was closed) and the predicted RULs for the 5 files in the test dataset were compared (as ground truth of validation dataset is not disclosed) and shown in Fig. 10 for fault #3 for tool #03_M01 in the test dataset.

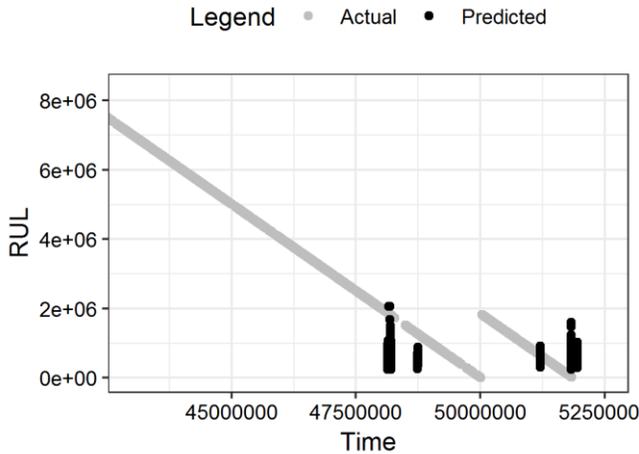


Figure 10. Comparison of ground truth and predicted RUL values for fault #3 for tool #03_M01 in the test dataset

It can be seen from Fig.10 that RUL predictions are not far from the actual failure time, possibly because the signature is too weak and far from the actual failure, to be reflected in the process parameters leading to meaningful RUL estimation. On the other hand, while RUL predictions are closer to the actual failure time, they do not follow the ground truth consistently. Similar observations can be made for other faults and other tools. This indicates that there is scope for improvement in RUL predictions. Optimizing the modeling approach, using operational data from multiple IMETs and setting a reasonable horizon for RUL estimation are some of the potential improvements. We, however, feel that the biggest improvement would be possible by selecting sensors closer and more relevant to the problem at hand. In this case, all three faults are related to the FlowCool system and there are only two parameters viz. flow rate and pressure of helium in the FlowCool circuit. Even though the performance of the FlowCool system is influenced by the performance of the ion etch mill, sensors related to the water system used to cool helium in the FlowCool circuit or those related to the FlowCool pump may exhibit better signatures of FlowCool system failure. The accuracy of RUL estimation may improve significantly by including these variables in the analysis.

6. CONCLUSIONS

We propose a novel data-driven methodology for estimation of RULs in IMETs. This approach addresses the challenges of absence of run-to-failure and the presence of multiple faults simultaneously in the system. Based on the understanding of the process and the equipment, 13 important parameters were selected to characterize normal and faulty behaviour of the IMET.

Operational data of the IMET was used to build a health score model and to create a library of truncated degradation curves for each fault. These libraries were utilized for obtaining estimates of RUL using DTW curve matching. The proposed approach was applied to test as well as validation datasets; the estimated RULs for both these datasets were found to be in reasonable agreement with the ground truth. The current approach may be improved by optimizing parameters such as smoothing factor and score threshold, utilizing data from multiple ion etch mills and through the use of deep learning techniques. Further improvement in RUL estimates may be realized by including sensors from the water cooling system and the FlowCool pump.

REFERENCES

- Gu, J., & Jin, X. (2006, September). A simple approximation for dynamic time warping search in large time series database. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 841-848). Springer, Berlin, Heidelberg.
- Huang, B., Di, Y., Jin, C., & Lee, J. (2017). Review of data-driven prognostics and health management techniques: Lessons learned from PHM data challenge competitions. *Machine Failure Prevention Technology 2017*, Virginia Beach, VA, USA
- Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483-1510.
- Qin, S. J. (2014). Process data analytics in the era of big data. *AIChE Journal*, 60(9), 3092-3100.
- Wang, T., Yu, J., Siegel, D., & Lee, J. (2008, October). A similarity-based prognostics approach for remaining useful life estimation of engineered systems. *International Conference on Prognostics and Health Management 2008*, Denver, CO, USA
- 2018 PHM Data Challenge, [Online], <https://www.phmsociety.org/events/conference/phm/18/data-challenge>