A Similarity-based and Model-based Fusion Prognostics Framework for Remaining Useful Life Prediction

Xiaochuan Li
Faculty of Computing, Engineering and Media
De Montfort University
Leicester, UK
xiaochuan.li@dmu.ac.uk

David Mba
[1]Faculty of Computing, Engineering and Media
De Montfort University
Leicester, UK
[2]Department of Mechanical Engineering
University of Nigeria
Nsukka, Nigeria

Tianran Lin
School of Mechanical and Automotive Engineering
Qingdao University of Technology
Qingdao, China

***Abstract*—In this work, a hybrid prognostic framework which interfaces a model-based prognostic method, namely particle filter, with a similarity-based prognostic method is proposed. The proposed framework consists of automatic determination of predication start point, sensor fusion, and prognostics steps that lead to accurate remaining useful life (RUL) estimations. This approach first applies the canonical variate analysis (CVA) approach for determining the prediction start time and constructing the prognostic health indicators (HIs). The similarity-based method is then employed together with the model-based particle filter (PF) algorithm to improve the predictive performance in terms of reducing the uncertainty of RUL and improving the prediction accuracy. The proposed framework can automatically construct HIs that are suitable for RUL prediction and offer higher prediction accuracy and lower uncertainty boundaries than traditional model-based PF methods. Our proposed approach is successfully applied on aircraft turbofan engines RUL prediction.**

Keywords-canonical variate analysis; particle filter; similarity-based prognostics; hybrid method; RUL prediction

I. INTRODUCTION

Unexpected downtime caused by machinery failures is costly and can incur large economic losses and security threats. In light of this industrial challenge, numerous predictive maintenance programs/approaches have been put forward in recent years to predict equipment failures and to ensure maintenance is carried out only when necessary. However, it is difficult and costly to carry out RUL prediction when equipment is under normal conditions since little information about the degradation trend can be found during this stage. To make a prognostic framework suitable for online monitoring, it is essential to include a module which can automatically determine prediction start time such that the RUL prediction is implemented only after certain failures are detected. Widely used data-driven process monitoring tools, known in general as multivariate statistical process monitoring (MSPM) techniques, have been widely used for detecting process abnormalities. Principal component analysis (PCA) [1], independent component analysis (ICA) [2] and their improved versions [3]–[6] are representative methods in early studies for industrial process monitoring. CVA [7] is also a MSPM technique, and it was shown to be superior to other dimension reduction techniques for monitoring dynamic process under varying operational conditions [5]. In this study, a CVA based HI is adopted together with kernel density estimation (KDE) to automatically determine the prediction start time. Additionally, CVA is employed in this work to generate degradation features based on which system failure time can be estimated.

Various data-driven methods are available in the literature for the prediction of degradation trajectories and RUL, including exponential regression [8], deep learning networks [9], grey models [10]. Generally, large amounts of historical data are expected to ascertain accurate prognostics. In the case of the trajectory-based RUL prediction, the inherent model error can accumulate over time leading to inaccurate RUL prediction results. Furthermore, the predictions made by a single trained predictor is deterministic, and do not have an uncertainty interval. In order to solve these problems, the model-based PF [11] method is employed to realize the joint prediction of RUL and the associated confidence levels, thereby allowing uncertainties generated during the prediction to be taken into consideration. However, one drawback of PF is that its predictive results greatly rely on the potentially imperfect fault evolution model, which may lead to error accumulation when it

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

comes to long-term prediction. As a result, the prediction results obtained at early degradation stages often demonstrate large deviations from the actual remnant life, and the associated uncertainty boundaries are extreme high. In order to tackle the problems brought by the model-based PF method, a similarity-based trajectory prediction method is employed to "guide" the prediction phrase of the PF method by constantly providing the model with future "measurements". This provides the possibility to update the inaccurate particles and their associated weights. The future values of the of the test engine's HI are predicted based on the weighted sum of the most similar trajectories found in all reference engines. In this way, the predictive accuracy at early degradation stages can be greatly improved and the uncertainty boundaries can be lowered compared with model-based methods. On the other hand, it is certainly possible to predict the testing degradation trajectory by finding the closest degradation trajectories from the historical training library. However, The RUL predicted using this method is deterministic and hence does not have an associated uncertainty interval. The proposed method is tested on the C-MAPSS simulation data sets.

II. METHODOLOGY

A. CVA for the determination of prediction start time

Fault detection in real rotating machinery using CVA was introduced in [7]. Although fault detection is not the main purpose of this study, it provides the starting point of RUL prediction in the sense that the prognostic analysis can be triggered once anomalies have been detected. A brief introduction of CVA-based diagnostic method is given in the following subsections.

Given two data sets $y_{1,t} \in \mathcal{R}^n$ and $y_{2,t} \in \mathcal{R}^n$, CVA aims to find the projections $z_{1,t} = K \cdot y_{1,t}$ and $z_{2,t} = G \cdot y_{2,t}$ such that the correlation between $z_{1,t}$ and $z_{2,t}$ is maximized. $z_{1,t}$ and $z_{2,t}$ are also known as canonical variates (CVs). The past and future column vectors $y_{a,t} \in \mathcal{R}^{na}$ and $y_{b,t} \in \mathcal{R}^{nb}$ can be formed as:

$$y_{a,t} = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-a} \end{bmatrix} \in \mathcal{R}^{na} \tag{1}$$

$$y_{b,t} = \begin{bmatrix} y_{t} \\ y_{t+1} \\ \vdots \\ y_{t+b-1} \end{bmatrix} \in \mathcal{R}^{nb} \tag{2}$$

where a and b are the size of the past and future windows of data, respectively. Then the past and future Hankel matrices $\hat{Y}_a$ and $\hat{Y}_b$ can be formed as:

$$\hat{Y}_a = [\, \hat{y}_{a,t+1}, \hat{y}_{a,t+2}, \dots, \hat{y}_{a,t+N}] \in \mathcal{R}^{na \times N} \tag{3}$$

$$\hat{Y}_b = [\hat{y}_{b,t+1}, \hat{y}_{b,t+2}, \dots, \hat{y}_{b,t+N}] \in \mathcal{R}^{nb \times N} \tag{4}$$

where $\hat{y}_{a,t}$ and $\hat{y}_{b,t}$ are the normalized past and future vectors, respectively. $N = M - a - b + 1$ represents the number of $y_{a,t}$ contained in the Hankel matrices $\hat{Y}_a$. The sample covariance matrices $\sum_{a,a}$, $\sum_{b,b}$ and $\sum_{a,b}$ can be obtained as

$$\sum_{a,a} = \hat{Y}_a \hat{Y}_a{}^T/(N-1)\ ; \quad \sum_{b,b} = \hat{Y}_b \hat{Y}_b{}^T/(N-1)\ ; \quad \sum_{b,a} = \hat{Y}_b \hat{Y}_a{}^T/(N-1) \tag{5}$$

CVA aims to maximize the correlation between the linear combinations of the projections $z_{1,t} = K \cdot y_{1,t}$ and $z_{2,t} = G \cdot y_{2,t}$. The algebraic solution is obtained by performing singular value decomposition (SVD) on the scaled matrix $\mathcal{H}$:

$$\mathcal{H} = \sum_{b,b}{}^{-1/2} \sum_{b,a} \sum_{a,a}{}^{-1/2} = U \sum V^T \tag{6}$$

The canonical residuals are given by the difference between the first q projections in the past and the future space:

$$r_t = L_q^T \hat{y}_{b,t} - \sum_q J_q^T \hat{y}_{a,t} \tag{7}$$

where $L_q^T \hat{y}_{b,t}$ is the first q future projections, and $L^T = \sum_{b,b}^{-1/2} U_q^T$. Similarly, $J_q^T \hat{y}_{a,t}$ is the first q past projections, and $J^T = \sum_{a,a}^{-1/2} V_q^T$. $\sum_q = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_q)$ is the diagonal matrix of the descending first q singular values.

In this paper, the determination of prediction start time is carried out by comparing a CVA distinction-based index T_c with a statistically determined fault threshold.

$$T_c = \frac{T^2}{\sigma^{T^2}} + \frac{Q}{\sigma^Q} + \frac{T_d}{\sigma^{T_d}} \tag{8}$$

$$T^2 = z_t^T z_t, z_t = V_q^T \Sigma_{a,a}^{-1/2} \hat{y}_{a,t} \tag{9}$$

$$Q = e_t^T e_t, e_t = V_{na-q}^T \Sigma_{a,a}^{-1/2} \hat{y}_{a,t} \tag{10}$$

$$T_d = f\big(c(r_t - 0)^T \mathrm{S}^{-1}(r_t - 0)\big) = \frac{|c(r_t-0)^T \mathrm{S}^{-1}(r_t-0)|}{|c|[(r_t-0)^T \mathrm{S}^{-1} S \mathrm{S}^{-1}(r_t-0)]} = [(r_t)^T \mathrm{S}^{-1}(r_t)]^{1/2} = [r_t^T (I - \sum \sum^T) r_t]^{1/2} \tag{11}$$

where σ^{T^2}, σ^Q and σ^{T_d} are the threshold of Hotelling's T^2 and Q statistics [12], and, T_d index, respectively. c demotes a normalizing constant, and $\mathrm{S} = I - \sum \sum^T$ denotes the covariance matrix of the training data. σ^{T^2}, σ^Q and σ^{T_d} are calculated using the kernel density estimation method as proposed in [13].

B. Trajectory prediction based on HI similarity

The CVA method described in the previous subsection is not only capable of determining the prediction start time but also constructing the HIs needed for subsequent prognostic analysis. The HIs are constructed as per equations (1)-(11) using the multivariate observations obtained from the machine. In other words, the T_c index served as the HI of the system under study. This HI is constructed for each individual machine, and hence a HI training library based on historical run-to-failure data can be built. One can predict the degradation trajectory of an engine by computing a weighted sum of the degradation HIs of the similar HI portions in all referential engines. This method requires neither a time-consuming training process nor sophisticated modelling to extrapolate the trajectory based on similarity between HIs. The HIs obtained from the sensory measurements are employed to perform similarity matching between the test and referential engines. The proposed approach searches for the HIs in the training library that is the closest to the current degradation trajectory of the test engine up to time k. The value of the trajectory at $k + \Delta k$ is predicted by the trajectories of the

best-matching trajectories at time $k + \Delta k$. The values of the best-matching trajectories at time $k + \Delta k$ can be assigned a weight according to the similarity between the testing trajectory and all best-matching historical trajectories. The similarity is measured by the Euclidean distance.

C. Enhancement of exponential regression using PF

The main idea of using PF for prognostics is to use a state transition model to predict the future state of a system:

$$x_k = g(x_1, x_2, \dots, x_k) + u_k \quad (12)$$

where g represents the state transition function, x_k denotes the system state at time k, and w_k is a noise term. u_k can be determined by the state transition model's modelling errors during the state tracking phase. In this work, we use the exponential regression model to describe the state evolution as formulated in (12).

Mathematically speaking, during the state tracking phase, PF uses a weighted set of random particles to approximate the posterior marginal density of the system state:

$$p(x_k|y_{1:k}) \approx \sum_{j=1}^{N} \pi_k^j \delta(x_k - x_k^j) \quad (13)$$

where N is the number of particles, and $\sum_{j=1}^{N} \pi_k^j = 1$. When new measurements become available, the weights π_k^j are updated in line with the principle of importance sampling:

$$\pi_k^j \propto \pi_{k-1}^j p(y_k|x_k^j) \quad (14)$$

According to [14], [15], a canonical representation of the measurement model can be used when the system state x_k is not directly measurable:

$$y_k = x_k + v_k \quad (15)$$

where both y_k and x_k represent the system state (in this case the HI obtained via CVA), and v_k is a noise term. After the state tracking phase, state prediction is implemented by projecting the current particle population based on the last estimated particles and the associated weights along all possible pathways.

In summary, Fig. 1 illustrates a flowchart of the proposed prognostic method. Once the prediction start time is determined by the CVA method, the hybrid prognostic framework will be triggered. The future value of the degradation trajectory of the testing engine is predicted through the best-matching historical HIs. The RUL is predicted by applying the PF-enhanced model-based method to the estimated trajectory.

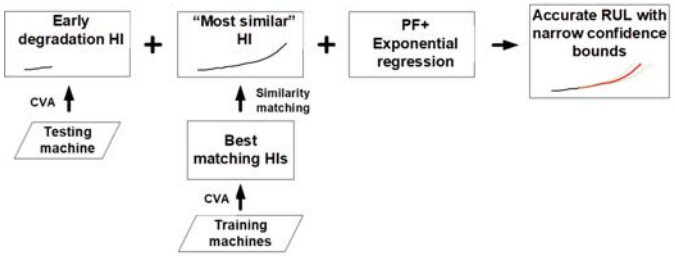

Figure 1. Flowchart of the proposed prognostic framework

III. INDUSTRIAL APPLICATION

A. Data collection

The data used in this experiment are from the C-MAPSS simulation program provided by the prognostics center of excellence of National Aeronautics and Space Administration (NASA), which aims to solve the lack of run-to-failure test data for data-driven prognostics [34]. The dataset contains information of the residual life of multiple engines in a homogeneous state. These engines are in normal state at the beginning, and then failure occurs at a certain moment and show obvious degradation of performance, and the degradation continues to accumulate until the system failure. Because of the degradation process integrity of the training data set, this paper selects samples in first group of training data set for demonstration. The first group of training data set is the assembly of High-Pressure Compressor (HPC) failure generated in the same operating state, including the monitoring samples of 100 compressors. 8 complementary variables among the total 26 variables found in the dataset were selected to construct the degradation HIs. The selected variables are variables no. 7,8,9,11,13,15,17 and 18.

B. Determination of prediction start time and RUL prediction

For the 8 performance parameters, CVA is used to extract the main performance features that represent both health state and degradation trend of the engines. CVA can also reduce the data's dimensionality. A training dataset obtained from different machines at normal operation conditions are used to train the CVA model. By doing this, the various operating conditions of the fleet of are fully accommodated. The control limit for healthy operational conditions is determined using the Kernel density estimation method [26] with a 99% confidence level. The trained CVA model is then employed to calculate a HI for each of the testing engine.

Then the similarity-based trajectory prediction method as proposed in Section II was applied to predict at future times the testing engine's HI values. A portion of the testing HI is compared with all the referential HIs in the training library using a similarity measure (the similarity measure was chosen as the Euclidean distance in this study). The predicted HI was calculated as the weighted sum of the ten best-matching trajectories. Finally, the merged HI is continuously fed into the particle filter as new "measurements" after the end of the state training phase, allowing the particles and the associated weights to be updated during state prediction. The failure threshold that the PF required to calculated the RUL was determined by the probability method introduced in [17]. The similarity measure was chosen as the Euclidean distance in this study.

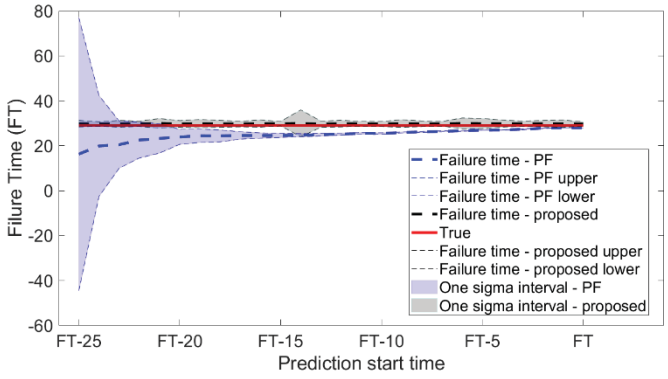

Figure 2. Flowchart of the proposed prognostic framework

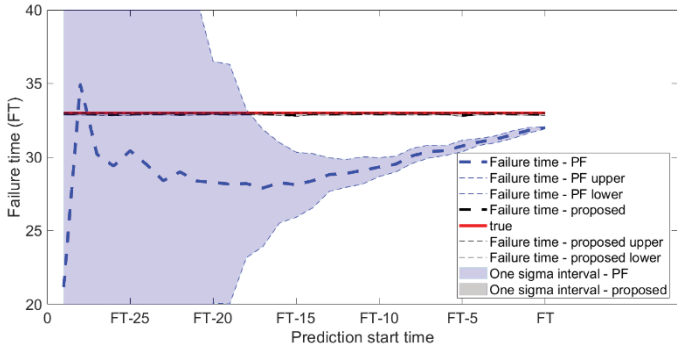

Figure 3. Flowchart of the proposed prognostic framework

Figs. 2 and 3 showed two exemplary results of the RUL prediction results for testing series #1 and #2 for predictions starting at different starting points for the proposed method and the standard PF algorithm, respectively. The areas between the dashed black curves denote the one standard deviation tolerance interval, which was calculated assuming the particles follow a normal distribution. It is observable from the figure that the predicted failure time using the proposed method is not only centered closer to the actual failure time at all times but also has narrower confidence boundaries, especially at early degradation stages. It is also obvious that the confidence boundaries of the proposed approach are much lower than those of the PF, especially at early degradation stages.

The proposed approach was tested on 8 engines. The predictive accuracy of the proposed method and the PF-enhanced exponential regression method were compared and the results were demonstrated in Table I. The predictive accuracy and uncertainty level of the proposed framework is shown to be superior to the standard PF.

TABLE I. PREDICTIVE PERFORMANCE

Failure cases	MAD		RMSD		MSD	
#1	4.17	0.83	4.93	0.84	4.80	1.64
#2	4.37	0.82	4.99	0.83	3.65	0.12
#3	4.75	0.40	5.67	0.41	10.68	2.71
#4	2.78	1	3.69	1	46.95	0.00049
#5	3.68	0.075	4.15	0.078	97.1	0.069
#6	2.03	2.91	3.22	2.75	9.84	0.44
#7	18.37	1.71	28.08	1.72	342.46	0.37
#8	4.09	0.97	8.29	0.98	128.27	0.0037

*For each column: left (PF), right (proposed method)

Prognostic performance metrics are provided, and three evaluation criteria were employed in this work to quantitatively benchmark the predictive accuracy of the methods compared:

1. Root mean square deviation (RMSD):

$$RMSD = \sqrt{\sum_{i=1}^{N}(RUL_{pre,i} - RUL_{true,i})^2/N} \quad (17)$$

2. Mean absolute deviation (MAD):

$$MAD = \sum_{i=1}^{N}|RUL_{pre,i} - RUL_{true,i}|/N \quad (18)$$

3. Mean standard deviation

$$MSD = \sum_{i=1}^{N}\sigma_i/N \quad (19)$$

IV. CONCLUSION

In this work, a hybrid prognostic framework was proposed for RUL prediction of turbofan engines. The effectiveness of this method was validated on the C-MAPSS datasets. By comparing the values of the CVA-based T_c HI with the pre-defined threshold, the prediction start time was automatically determined. The proposed prognostic method demonstrates superior performance than the standard PF algorithm in terms of narrowed confidence boundaries and improved predictive accuracy. Then proposed method can be considered as a substitute of the standard PF-based prognostic approaches when large number of historical failure cases are available. How to determine the failure threshold is a critical issue for many RUL prediction tasks. Based on the current threshold setting, the proposed method outperforms the traditional PF method in terms of predictive accuracy and confidence boundary. How to select an optimal failure threshold will be investigated in future study.

REFERENCES

[1] S. Wook, C. Lee, J. Lee, J. Hyun, and I. Lee, "Fault detection and identification of nonlinear processes based on kernel PCA," *Chemom. Intell. Lab. Syst.*, vol. 75, 2005, pp. 55–67.

[2] J. Fan and Y. Wang, "Fault detection and diagnosis of non-linear non-Gaussian dynamic processes using kernel dynamic independent component analysis," *Inf. Sci. (Ny).*, vol. 259, 2014, pp. 369–379.

[3] W. Li and S. J. Qin, "Consistent dynamic PCA based on errors-in-variables subspace identification," *J. Process Control*, vol. 11, no. 6, 2001, pp. 661–678.

[4] S. Yin, X. Zhu, S. Member, and O. Kaynak, "Improved PLS Focused on Key-Performance- Indicator-Related Fault Diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, 2015, pp. 1651–1658.

[5] C. Ruiz-cárcel, Y. Cao, D. Mba, L. Lao, and R. T. Samuel, "Statistical process monitoring of a multiphase flow facility," *Control Eng. Pract.*, vol. 42, 2015, pp. 74–88.

[6] G. Stefatos and A. Ben Hamza, "Dynamic independent component analysis approach for fault detection and diagnosis," *Expert Syst. Appl.*, vol. 37, no. 12, 2010, pp. 8606–8617.

[7] X. Li, X. Yang, Y. Yang, I. Bennett, A. Collop, and D. Mba, "Canonical variate residuals-based contribution map for slowly

evolving faults," *J. Process Control*, vol. 76, 2019, pp. 87–97.

[8] J. Wu, C. Wu, S. Cao, S. W. Or, C. Deng, and X. Shao, "Degradation Data-Driven Time-To-Failure Prognostics Approach for Rolling Element Bearings in Electrical Machines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 1, 2018, pp. 529–539.

[9] Q. Wu, K. Ding, and B. Huang, "Approach for fault prognosis using recurrent neural network,", 2018, *J. Intell. Manuf.*.

[10] X. Yang, Z. Fang, Y. Yang, D. Mba, and X. Li, "A novel multi-information fusion grey model and its application in wear trend prediction of wind turbines," 2019, *Appl. Math. Model.*.

[11] Z. Chen, "Bayesian filtering: From Kalman filters to particle filters, and beyond," *Statistics (Ber).*, vol. 182, no. 1, 2003, pp. 1–69.

[12] H. Hotelling, "New Light on the Correlation Coefficient and its Transforms," *J. R. Stat. Soc. Ser. B*, vol. 15, no. 2, 1953, pp. 193–232.

[13] P. E. P. Odiowei and C. Yi, "Nonlinear Dynamic Process Monitoring Using Canonical Variate Analysis and Kernel Density Estimations," *IEEE Trans. Ind. Informatics*, vol. 6, no. 1, 2010, pp. 36–45,.

[14] C. Chen, B. Zhang, G. Vachtsevanos, and M. Orchard, "Machine condition prediction based on adaptive neuro-fuzzy and high-order particle filtering," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 9, 2011, pp. 4353–4364.

[15] X. Li, X. Yang, Y. Yang, I. Bennett, and D. Mba, "A novel diagnostic and prognostic framework for incipient fault detection and remaining service life prediction with application to industrial rotating machines," *Appl. Soft Comput.*, 2019, p. 105564.

[16] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *2008 International Conference on Prognostics and Health Management, 2008, PHM 2008*.

[17] L. Liao, "Discovering Prognostic Features Using Genetic Programming in Remaining Useful Life Prediction," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, 2014, pp. 2464–2472.