Development of Multivariate Failure Threshold with Quantifiable Operation Risks in Machine Prognostics

Xiaodong Jia[1,2], Wenzhe Li[1,2*], Wei Wang[2], Xiang Li[1,2] and Jay Lee[1,2]

[1]*NSF I/UCR Center for Intelligent Maintenance Systems,*
[2]*Department of Mechanical Engineering, University of Cincinnati, PO Box 210072,*

Cincinnati, Ohio 45221-0072, USA

jiaxg@ucmail.uc.edu
li2w6@mai.uc.edu
wang2w3@ucmail.uc.edu
li5xi@ucmail.uc.edu
lj2@ucmail.uc.edu

ABSTRACT

Prognostics and Health Management (PHM) is attracting the attention from both academia and industry due to its great potential to enhance the resilience and responsiveness of the equipment to the potential operation risks. In literature, many methodologies are proposed to predict the Remaining Useful Life (RUL) of the equipment. However, there are two major challenges that limit the practicality of these methodologies. 1) How to generate a quantifiable Health Indicator (HI) to represent the operation risks? 2) How to define a reasonable failure threshold to predict RUL? To answer these two questions, this paper proposes a novel methodology for failure threshold determination with quantifiable operation risk in machine prognostics. In the proposed methodology, Fisher distance and Mann-Kendall (MK) test are firstly used to extract useful sensors based on which HI is estimated by applying Principle Component Analysis (PCA). Then, Rao-Blackwellized Particle Filter (RBPF) is employed to obtain the HI prediction and the uncertainties. Afterwards, a Bivariate-Weibull-distribution-based risk quantification model is designed to quantify the cumulative risk over time and over the increase of HI. The failure threshold, which is the ending point of the RUL, varies over different users and applications depending on the level of risk they want to tolerate. The validation of the methodology is based on the C-MAPSS data from the PHM data competition 2008 hosted by PHM society. The results validate the effectiveness of the proposed risk quantification method and its potential application on machine prognostics.

1. INTRODUCTION

Remaining Useful Life (RUL) prediction has been extensively studied in the literature, and many different methodologies are proposed to fulfill the prediction task. These methodologies include but not limit to Regression-Based Method (RBM), Random Coefficient Method (RCM), Stochastic Process Method (SPM), State-Space Method (SSM) and Similarity-Based Method (SBM), as being summarized in (Cai, Feng, Li, Hsu, & Lee, 2020; Haoshu Cai, Xiaodong Jia, et al., 2020; H. Cai et al., 2020; Jia et al., 2019).

Based on the literature survey, it is found that the failure threshold in most RUL prediction studies is often set empirically by referring to the engineering practices. There still lack a systematic approach to define failure threshold with quantifiable operation risk. This is important because 1) the predicted Health Indicator (HI) cannot be effectively converted to RUL without an effective failure threshold (N. P. Li, Lei, Lin, & Ding, 2015). 2) a quantifiable failure threshold is the key for a prognostic model to interface with the maintenance planning (Camci, 2014, 2015). This is essentially because the maintenance planning models expects the take the quantified risk P_F as input to calculate the direct and indirect failure cost (Camci, 2014, 2015). Without a tunable failure threshold with quantifiable risk, it is difficult to build friendly interface between prognostic models and maintenance planning models. 3) A failure threshold with quantifiable risk will allow different manufacturers or users

1

to tune the tolerable risk based on their practical needs and application scenarios.

As a branch of survival analysis, multivariate survival analysis models covariates and life indicators as clusters to analysis expected duration of time until machine failure. Commonly seen approaches to model the variation between life indicators and other factors are Cox Proportional Hazards Model (Breslow, 1975), Clayton Copula Model (Clayton, 1978) and competing risks models (Prentice et al., 1978), etc. One major focus of multivariate survival is to address the correlation between life indicators and health indicators, which is sometimes picky for dependency measure selection and background knowledge (Georges, Lamy, Nicolas, Quibel, & Roncalli, 2001). In multivariate survival analysis applications, improper dependency measures can render the life indicator over-dominant in the model.

To address this challenge, this study proposes a risk quantification model to assist the optimization of failure threshold (FT). The proposed model considers the joint effects of life indicator and health indicator, where the life indicators is essentially used in reliability analysis, and the health indicator is derived from sensory readings. The proposed risk model falls somewhat between reliability analysis and HI based failure statistics. Based on the proposed risk model, a systematic methodology for RUL prediction is proposed, and the effects of varying tolerable risk on the predictive distribution RUL are studied and discussed based on the well-known C-MAPSS data about aero-engines degradation (Jia, Huang, Feng, Cai, & Lee, 2018).

The rest of the paper is organized as follows. Section 2 gives an overview of the previous studies about the failure thresholding. Section 3 elaborates the proposed methodology. Section 4 illustrate the benchmarking results based on the public datasets. The conclusion remarks are given in Section 5.
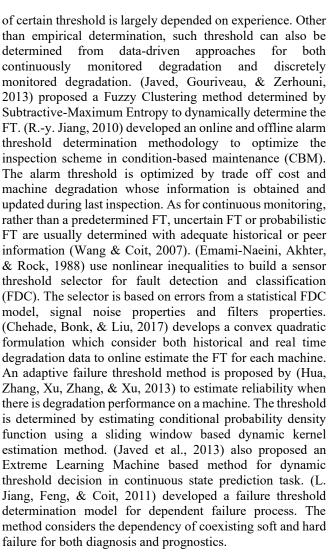
2. REVIEW

In machine prognostics, a unit is often considered to have failed when its health indicator crosses a certain failure threshold(X. Li et al., 2020; Yang et al., 2020). Determination
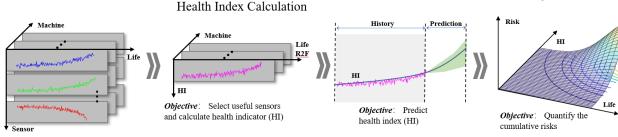
of certain threshold is largely depended on experience. Other than empirical determination, such threshold can also be determined from data-driven approaches for both continuously monitored degradation and discretely monitored degradation. (Javed, Gouriveau, & Zerhouni, 2013) proposed a Fuzzy Clustering method determined by Subtractive-Maximum Entropy to dynamically determine the FT. (R.-y. Jiang, 2010) developed an online and offline alarm threshold determination methodology to optimize the inspection scheme in condition-based maintenance (CBM). The alarm threshold is optimized by trade off cost and machine degradation whose information is obtained and updated during last inspection. As for continuous monitoring, rather than a predetermined FT, uncertain FT or probabilistic FT are usually determined with adequate historical or peer information (Wang & Coit, 2007). (Emami-Naeini, Akhter, & Rock, 1988) use nonlinear inequalities to build a sensor threshold selector for fault detection and classification (FDC). The selector is based on errors from a statistical FDC model, signal noise properties and filters properties. (Chehade, Bonk, & Liu, 2017) develops a convex quadratic formulation which consider both historical and real time degradation data to online estimate the FT for each machine. An adaptive failure threshold method is proposed by (Hua, Zhang, Xu, Zhang, & Xu, 2013) to estimate reliability when there is degradation performance on a machine. The threshold is determined by estimating conditional probability density function using a sliding window based dynamic kernel estimation method. (Javed et al., 2013) also proposed an Extreme Learning Machine based method for dynamic threshold decision in continuous state prediction task. (L. Jiang, Feng, & Coit, 2011) developed a failure threshold determination model for dependent failure process. The method considers the dependency of coexisting soft and hard failure for both diagnosis and prognostics.

The reviewed FT determination methods are hardly based upon risks the operator would directly want to shoulder. Some autonomous FT determination methods are hard to adjust according to additional constraints and some other methods my require large computation resources. To address these challenges and expand the horizon of FT determination in machine prognostics, this study proposes a risk quantification methodology to assist the optimization of FT.

Figure 1. An overview of the proposed methodology

3. METHODOLOGY

3.1. Overview

The proposed methodology includes 3 three major steps, as shown in Figure **1**.

1) Health Indicator (HI) estimation: this step includes the selection of trended features that are useful for RUL prediction and the estimation of the embedded degradation trend. In this study, the feature selection is based on Fisher distance and Mann-Kendall (MK) test. The estimated HI is given by performing Principal Component Analysis (PCA) on the selected sensor readings.

2) HI Prediction: The purpose of HI prediction is to predict the development of HI in future time horizons. Rao-Blackwellized Particle Filter (RBPF) is employed to obtain the prediction and the uncertainties, as shown in Figure **1**.

3) Risk Quantification: The risk quantification model is designed to quantify the cumulative risk over time and over the increase of HI. The mathematical expression of the risk in this study is described as:

$$R = 1 - S(t, h) \qquad (1)$$
$$= 1 - \mathbf{Prob}(L \geq t, HI \geq h)$$

Where L, HI denotes the Life and Health Indicator, respectively, t, h represents the current time and current HI value. $S(t, h)$ is the joint probability of survival when $L = t$ and $HI = h$. The risk R is defined as $1 - S(t, h)$.

In the following discussion, the primary processing steps in the proposed methodology are elaborated.

3.2. HI Estimation

Health indicator of aero-engine system is usually estimated according to sensor selection from monitoring sequences(H. Cai et al., 2020; Xu, Hou, Li, & Zheng, 2018).The results for feature selection are shown in Figure 2. The Fisher distance in Figure 2(a) measures the usefulness of the sensory data to distinguish the healthy and faulty. In this study, the mean of the first 15 operations cycles are specified as healthy, and the last 15 operation cycles prior to the maintenance are defined as faulty. Based on the preliminary screening according to Fisher distance, sensors 2, 3, 4, 11, 17 are identified as useful sensors for detection.

The MK test further evaluates the presence of a gradual trend in the sensory data throughout the machine degradation. If the MK test is passed, there is a trend in the sensory data. In Figure 2(a), the pass rate of the MK test is utilized as an indicator to show the trend-ability of the data, which indicates that sensors 2, 3, 4, 11, 17 are trend-able. Therefore, in this study, the HI estimation is primarily obtained by performing PCA on the sensor 2, 3, 4, 11, and 17(Jia et al., 2019).
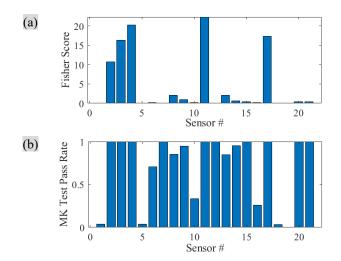
Figure 2. Feature selection based on (a) Fisher distance between healthy and faulty (b) Mann-Kendall test for trend-ability.

3.3. HI Prediction

The RBPF that is specified in Eq.(2) is employed to predict the HI. Compared with other particle filter methods, RBPF is employed because it can depict the prediction uncertainty, and it is efficient to do the computation (Cai et al., 2019).

$$\begin{cases} c_t = c_{t-1} + \epsilon_c \\ b_t = b_{t-1} + \epsilon_b \\ x_t = (x_t - c_t) \cdot e^{b_t \cdot \Delta t} + c_t + \epsilon_u \\ y_t = x_t + \epsilon_y \end{cases} \qquad (2)$$

Where c_t, b_t denotes the samples of the unknown model parameters. The x_t is the state vector, which represents the estimated HI at time t. The y_t is the observed HI at time t. $\epsilon_c, \epsilon_b, \epsilon_u$ and ϵ_y in Eq.(**2**) are assumed to follow Gaussian distribution with zero-mean. The variance of $\epsilon_c, \epsilon_b, \epsilon_u$ and ϵ_y are tuned to the best performance.

The third equation in Eq.(2) regulates an exponential degradation trend, and it is obtained by discretizing the exponential model below:

$$y = a \cdot \exp(b \cdot t) + c \qquad (3)$$

Where a, b, c are the unknown parameters in the exponential model and y is the observed HI at time t.

The application of the RBPF includes the filtering process and the prediction process. The objective of performing filtering on the partial degradation is to estimate the unknown model parameters $\mathbf{\Theta} = \{b, c, x, \sigma_u, \sigma_y\}$. Based on the estimated $\mathbf{\Theta}^t$ at current time t, the prediction can be obtained by extrapolating the state equation (3rd equation) in Eq.(2). σ_b, σ_c, which are standard deviation of b and c, are estimated based on the historical Run-to-Failure (R2F) data.

In the proposed method, the following processing steps are followed:

1) Estimate $\{\Theta_k\}_{k=1,...,M}$ for each historical R2F record, where M denotes the number of historical R2F records.

2) Estimate σ_b, σ_c based on historical data.

3) For each testing unit with partial degradation data, initialize the filtering process by using entries in $\{\Theta_k\}_{k=1,...,M}$. At the end of the filtering process, M different filtering results at current time t are obtained, which is denoted as $\{\Theta_k^t\}_{k=1,...,M}$.

4) Extrapolate the future trend by using entries in $\{\Theta_k^t\}_{k=1,...,M}$ as the starting point of the HI prediction.

To convert the predicted HI distribution into RUL distribution, the failure threshold is required. A systematic way to obtain a failure threshold with quantified risk is still discussed in the literature. The discussion in the next subsection fills in this gap by utilizing a Bivariate Weibull model.

3.4. Risk Quantification

Figure 3. Risk surface given by the Bivariate Weibull model

Figure 3 shows the risk surface and the iso-risk lines that are given by the proposed Bi-variate Weibull model. The iso-risk lines in Figure 3 can be used as a failure threshold by specifying the tolerable risk level. If a 5% risk line is adopted, it implies the user has a 95% probability of not encountering unexpected failure (Generazio, 2007). Typically, the optimal risk level can be obtained based on cost-benefit analysis in practice. Once the optimal risk is obtained, the failure threshold is fixed based on the bivariate model in Figure 3. In this study, the failure threshold is the ending point of the HI prediction.

The definition of the risk is given in Eq.(1). The joint survival in Eq.(1) is obtained from the joint distribution $\text{Prob}(t, h)$, which is a Bivariate Weibull distribution as in below:

$$f(HI, t) = \frac{\gamma_t \gamma_{HI}}{\alpha_t \alpha_{HI}} \left(\frac{HI - \mu_{HI}}{\alpha_{HI}}\right)^{\gamma_{HI}-1} \cdot \qquad (4)$$

$$\left(\frac{t - \mu_t}{\alpha_t}\right)^{\gamma_t - 1} e^{-\left(\frac{HI - \mu_{HI}}{\alpha_{HI}}\right)^{\gamma_{HI}} - \left(\frac{t - \mu_t}{\alpha_t}\right)^{\gamma_t}}$$

The parameters in the Bivariate Weibull density function $\theta = \{\gamma_t, \gamma_{HI}, \alpha_t, \alpha_{HI}, \mu_t, \mu_{HI}\}$ which represent shape, scale and location parameter of life indicator and health indicator respectively. θ is estimated by maximizing the likelihood function from the marginal pdf $f(HI, t)$.

In the Bivariate Weibull distribution model, the life indicator t, and the health indicator h are assumed to be independent. This assumption has significantly simplified the challenge of estimating multivariate PDF (Probability Density Function), and it is valid in most applications.

Weibull distribution is widely used for life distribution in reliability engineering due to its flexibility. It is generally believed dozens of life samples would be sufficient to estimate a 3-parameter life Weibull distribution in practice. More samples would improve such estimation accuracy. Expert experience can also be applied by defining the distribution of HI and life when reference history is insufficient.

Unlike the multivariate survival analysis, the proposed Bivariate risk model considers the life indicator and the health indicator as random variables with equal importance. In multivariate survival analysis, the HIs are modeled as predictors rather than random variables, and this renders the life indicator over-dominant in the model. To give an example, the risk surface and iso-risk lines given by the Cox Proportional Hazard Model, which is a typical model for multivariate survival analysis, are shown in Figure 4. It clearly indicates that the life indicator is over-dominant in the model, and the iso-risk lines cannot be utilized as the failure threshold for RUL prediction.
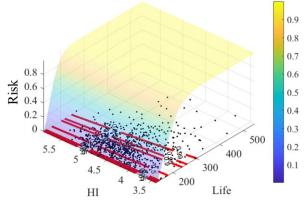

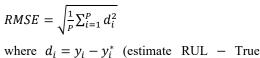Figure 4. Risk surface given by the Cox Regression Model

4. RESULTS & DISCUSSIONS

In this paper, the proposed methodology is illustrated in the aero-engine RUL prediction problem using C-MAPSS dataset. The details of the dataset is summarized in Table 1, where Fan degradation and Higher Pressure Compressor (HPC) degradation are two failure modes in different

datasets. The data processing, HI prediction and risk quantification are deployed sequentially following the procedures from Figure **1**. The prediction performance is validated using RMSE and Score calculated using Eq.(5). As noticed from literature, a constant value R_{early} is assigned as target RUL before degradation period (X. Li, Ding, & Sun, 2018). In this case, R_{early} is set to 125 for result evaluation. The selection of risk level for estimating RUL depends on whether R_{early} is applied or not. When R_{early} is not applied, a 75% risk level is selected for relatively conservative prediction. When R_{early} is applied, the largest prediction RUL is limited to 125 therefore the risk level can be selected as large as possible. In this case 95% risk level is selected when R_{early} is applied. To compare and evaluate the performance of the proposed method, some benchmarked methodologies are summarized from literature: DCNN and LSTM from (X. Li et al., 2018), DBN (Deep Belief Networks), MODBNE (Multi-Objective Deep Belief Networks Ensemble), RF (Random Forest), GB(Gradient Boosting), SVM (Support Vector Machine) and LASSO reported from (Zhang, Lim, Qin, & Tan, 2016), and RULCLIPPER from (Ramasso, 2014).
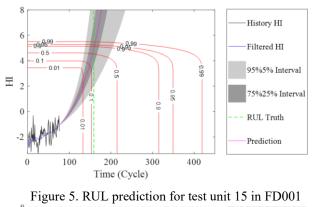
Dataset	Degradation Mode	N_{train}	N_{test}
FD001	HPC	100	100
FD002	HPC	260	259
FD003	HPC, FAN	100	100
FD004	HPC, FAN	249	248
Total	--	709	707

Table 1. Result Comparison with Benchmarks.

$$Score = \begin{cases} \sum_{i=1}^{P} e^{-\left(\frac{d_i}{13}\right)} - 1, \text{for } d_i < 0 \\ \sum_{i=1}^{P} e^{\left(\frac{d_i}{10}\right)} - 1, \text{for } d_i \geq 0 \end{cases} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{P}\sum_{i=1}^{P} d_i^2}$$

where $d_i = y_i - y_i^*$ (estimate RUL − True RUL)

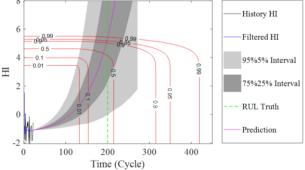Figure 5. RUL prediction for test unit 15 in FD001

Figure 6. RUL prediction for test unit 59 in FD002

Figure 5 and Figure 6 give two examples of the proposed methodology. The significant findings are summarized as following: 1) The filtered HI is a noise-free process; 2) The predicted HI follows an exponential trend; 3) The prediction uncertainty of HI can be effectively converted to the uncertainty of RUL by adopting the iso-risk line as failure threshold. 4) In regard to the RUL distribution, the effects of life indicators are dominant when a failure threshold with

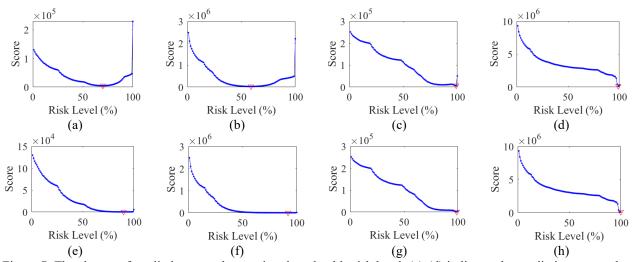|		Proposed Method (Risk=95%)	DCNN	LSTM	RULCLIPPER	Proposed Method (Risk=70%)	MODBNE	DBN	RF	GB	SVM	LASSO
FD001	RMSE	**16.72**	12.61	13.52	13.27	**26.41**	15.04	15.21	17.91	15.67	40.72	19.74
	Score	**592**	273	431	216	**4830**	334	417	479	474	7703	653
FD002	RMSE	**15.51**	22.36	24.42	22.89	**28.86**	25.05	27.12	29.59	29.07	52.99	37.13
	Score	**3837**	14459	10412	2796	**57239**	5585	9031	70465	87280	316483	276923
FD003	RMSE	**24.14**	12.64	13.54	16	**35.08**	12.51	14.71	20.27	16.84	46.32	21.38
	Score	**8014**	284	347	317	**33338**	421	442	711.13	576	22541	1058
FD004	RMSE	**27.97**	23.31	24.21	24.33	**49.1**	28.66	29.88	31.12	29.01	59.96	40.7
	Score	**1629800**	12466	14322	3132	**2704600**	6557	7954	46567	17817	141122	125297
R_{early}		**125**	125	125	125	**NA**	NA	NA	NA	NA	NA	NA

Table 2. Result Comparison with Benchmarks.

Figure 7. The change of prediction score by varying the tolerable risk level. (a)-(d) indicates the prediction score when $R_{early} = 125$. (e)-(h) indicates the prediction when no R_{early} is applied
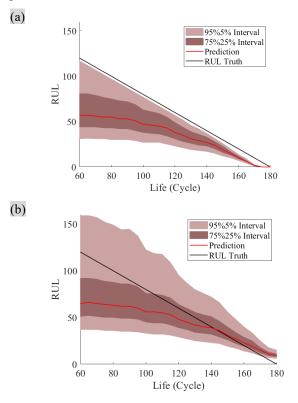
lower risk is chosen. However, the effects of HI on the RUL distribution are dominant when a higher risk is chosen. 5) The RUL ground truth in Figure 5 and Figure 6 indicate that a tolerable risk within [0.9, 0.99] can give rather good accuracy. However, it is important to note that the tolerable risk varies over different applications and over different units.

Figure 7 shows the typical prediction trajectory of a training unit under the risk level 25%, 95% and 50% (where best prediction is achieved). The findings can be summarized as following: 1) With the increasing length of history, the predicted RUL gradually converges to the ground truth; 2) the prediction uncertainty bound is gradually shrinking to a very narrow range; 3) the prediction uncertainty bound is relatively narrow at all prediction horizon when selecting lower risk level compared with higher risk level; 4) the overall predicted RUL under higher risk level is larger than the lower risk level. These findings justify the importance of risk quantification in RUL prediction.

The effects of varying tolerable risk on the prediction score are illustrated in Figure 7. It is found that an optimal score is found for FD001 and FD002 when varying the risk. In contrast, there is no optimal score for FD003 and FD004. This is essentially because FD003 and FD004 contains two degradation types: HPC degradation and FAN degradation, and FD001 and FD002 only have HPC degradation. This finding further implies that the risk model should be different for different failure modes.

The RUL prediction result under the risk level 70% with R_{early} not applied and risk level 95% with R_{early} applied for FD001 to FD004 is compared with benchmarked methods from literature in Table 2. It can be highlighted that 1) the prediction Score and RMSE of FD001 and FD002 very good and are compariable with other benchmarked methodologies; 2) the RMSE of FD002 with R_{early} is 15.51, which is

significantly better than any other benchmarked methods; 3) the general prediction performance under the same risk level are significantly different among FD001 to FD004; 4) the performance of proposed methodology is largely depend on the selection of risk level. By using the proposed risk quantification method, the prediction accuracy can be further optimized.
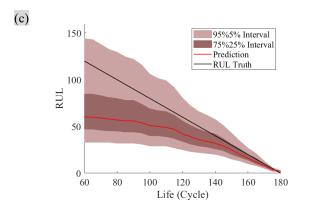
(a)

(b)

(c)

Figure 8. Trajectories of predicted RUL for certain length of operation history of training unit 62 in FD001 at (a) risk level 25% (b) risk level 95% and (c) optimum risk level 50%.

Figure 8 shows the trajectories of predicted RUL under different risk level when different length of operation history is given. The predicted RUL values tend to be smaller at low risk level because the failure threshold is set lower when a low risk level is given which result in early termination of degradation extrapolation. This justify the effectiveness of proposed risk quantification method. The RBPF prediction model is justified by that the uncertainty band gradually shrinks and predicted RUL converge to ground truth with training history increase. It should be addressed that prediction uncertainty and prediction threshold are different where the former is determined by prediction model while the latter is not. Prediction uncertainty is related to the historical data on itself while failure threshold is chosen based on tolerable risk in this case. Failure threshold will determine the end of life (end point of RUL prediction) but will not affect prediction uncertainty bound.
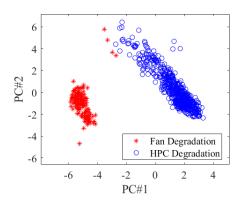

Figure 9. Degradation mode detection of C-MAPSS dataset training samples using PCA+Kmeans
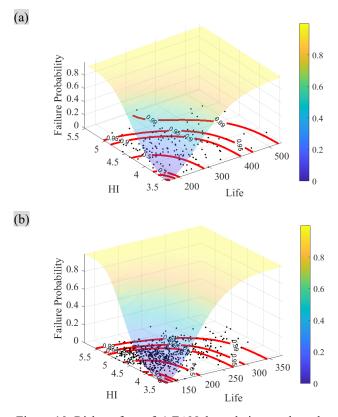
(a)

(b)

Figure 10. Risk surface of a) FAN degradation mode and b) HPC degradation mode given by the Bivariate Weibull model.

In the C-MAPSS case, the degradation mode of training units can be detected and classified using PCA and Kmeans method as shown in Figure 9. Indexed by the FDC result, the risk surface of FAN degradation and HPC degradation can be compared from a) and b). It can be concluded that 1) the risk contour patterns are different between two types of failure; 2) the potential life under FAN degradation is much higher compared with the HPC degradation units under the same risk level; 3) the failure HI for FAN degradation is relatively smaller than the HPC degradation. These suggest that the prediction accuracy based on risk surface trained from all failure modes cannot at least perform high accuracy in predicting FAN degradation units RUL by giving relatively conservative (smaller) RUL estimation. Therefore, the proposed method has very poor prediction performance on FD003 and FD004. This also explains why the optimal risk level for FD003 and FD004 is 99% or 100% since nearly all of the predicted RUL is too small. It is important to build the risk quantification model with regard to different failure modes and conduct FDC before estimating the RUL. Another reason for less prediction accuracy can be the inferring method RBPF is simply randomly initialized. The prediction performance can be further improved by applying some enhanced methodologies for RBPF initialization.

5. CONCLUSION

This paper proposes a Bi-variate model, which can be easily extended to the multivariate case, to define the failure threshold with quantifiable operation risk. The model considers the joint effect of life indicators and health indicators by using a Bi-variate Weibull model. Based on the risk model, a systematic methodology for RUL prediction is proposed, and several important conclusions are achieved:

1) The proposed risk model can effectively convert the predictive distribution of HI into the predictive distribution of RUL.

2) If low risk can be tolerated, the reliability information estimated from life indicator should be primarily considered. If higher risk can be tolerated, the HI based failure threshold will be more meaningful. This indirectly implies the HI based RUL prediction can potentially extend the equipment life in some non-critical applications.

3) The RUL prediction results can reach its optimal performance by varying the tolerable risk in RUL prediction tasks.

4) In RUL prediction, it is crucial to pro-diagnose the equipment failure, and the risk model for different failures should be different.

As for future works, multivariate risk models will be explored by removing the independence assumption in the current model. Besides, adopting the proposed method to interface prognostic models and maintenance planning models will be explored.

REFERENCES

Breslow, N. E. (1975). Analysis of Survival Data under the Proportional Hazards Model. *International Statistical Review / Revue Internationale de Statistique, 43*(1), 45-57. doi:10.2307/1402659

Cai, H., Feng, J., Li, W., Hsu, Y.-M., & Lee, J. (2020). Similarity-based Particle Filter for Remaining Useful Life prediction with enhanced performance. *Applied Soft Computing*, 106474.

Cai, H., Jia, X., Feng, J., Li, W., Hsu, Y.-M., & Lee, J. (2020). Gaussian Process Regression for numerical wind speed prediction enhancement. *Renewable Energy, 146*, 2112-2123. doi:10.1016/j.renene.2019.08.018

Cai, H., Jia, X., Feng, J., Li, W., Pahren, L., & Lee, J. (2020). A similarity based methodology for machine prognostics by using kernel two sample test. *ISA Trans*. doi:10.1016/j.isatra.2020.03.007

Cai, H., Jia, X., Feng, J., Yang, Q., Hsu, Y.-M., Chen, Y., & Lee, J. (2019). A combined filtering strategy for short term and long term wind speed prediction with improved accuracy. *Renewable Energy, 136*, 1082-1090.

Camci, F. (2014). The travelling maintainer problem: integration of condition-based maintenance with the travelling salesman problem. *Journal of the Operational Research Society, 65*(9), 1423-1436.

Camci, F. (2015). Maintenance scheduling of geographically distributed assets with prognostics information. *European Journal of Operational Research, 245*(2), 506-516.

Chehade, A., Bonk, S., & Liu, K. (2017). Sensory-based failure threshold estimation for remaining useful life prediction. *IEEE Transactions on Reliability, 66*(3), 939-949.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika, 65*(1), 141-151.

Emami-Naeini, A., Akhter, M. M., & Rock, S. M. (1988). Effect of model uncertainty on failure detection: the threshold selector. *IEEE Transactions on Automatic Control, 33*(12), 1106-1115.

Generazio, E. R. (2007). Directed Design of Experiments (DOE) for Determining Probability of Detection (POD) Capability of NDE Systems (DOEPOD).

Georges, P., Lamy, A.-G., Nicolas, E., Quibel, G., & Roncalli, T. (2001). Multivariate survival modelling: a unified approach with copulas. *Available at SSRN 1032559*.

Hua, C., Zhang, Q., Xu, G., Zhang, Y., & Xu, T. (2013). Performance reliability estimation method based on adaptive failure threshold. *Mechanical Systems and Signal Processing, 36*(2), 505-519.

Javed, K., Gouriveau, R., & Zerhouni, N. (2013). *Novel failure prognostics approach with dynamic thresholds for machine degradation.* Paper presented at the IECON 2013-39th Annual Conference of the IEEE Industrial Electronics Society.

Jia, X., Cai, H., Hsu, Y., Li, W., Feng, J., & Lee, J. (2019). *A Novel Similarity-based Method for Remaining Useful Life Prediction Using Kernel Two Sample Test.* Paper presented at the Proceedings of the Annual Conference of the PHM Society.

Jia, X., Huang, B., Feng, J., Cai, H., & Lee, J. (2018). *A Review of PHM Data Competitions from 2008 to 2017.* Paper presented at the PHM Society Conference.

Jiang, L., Feng, Q., & Coit, D. W. (2011). *Reliability analysis for dependent failure processes and dependent failure threshold.* Paper presented at the 2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering.

Jiang, R.-y. (2010). Optimization of alarm threshold and sequential inspection scheme. *Reliability Engineering & System Safety, 95*(3), 208-215.

Li, N. P., Lei, Y. G., Lin, J., & Ding, S. X. (2015). An Improved Exponential Model for Predicting

Remaining Useful Life of Rolling Element Bearings. *IEEE Transactions on Industrial Electronics, 62*(12), 7762-7773. doi:10.1109/Tie.2015.2455055

Li, X., Ding, Q., & Sun, J.-Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety, 172*, 1-11.

Li, X., Jia, X., Wang, Y.-L., Yang, S.-J., Zhao, H.-D., & Lee, J. (2020). Industrial Remaining Useful Life Prediction by Partial Observation Using Deep Learning with Supervised Attention. *IEEE/ASME Transactions on Mechatronics*.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. T., & Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 541-554.

Ramasso, E. (2014). Investigating computational geometry for failure prognostics.

Wang, P., & Coit, D. W. (2007). *Reliability and degradation modeling with random or uncertain failure threshold.* Paper presented at the 2007 Annual Reliability and Maintainability Symposium.

Xu, Y., Hou, W., Li, W., & Zheng, N. (2018). *Aero-Engine Gas-path Fault Diagnosis Based on Spatial Structural Characteristics of QAR Data.* Paper presented at the 2018 Annual Reliability and Maintainability Symposium (RAMS).

Yang, Q., Jia, X., Li, X., Feng, J., Li, W., & Lee, J. (2020). Evaluating Feature Selection and Anomaly Detection Methods of Hard Drive Failure Prediction. *IEEE Transactions on Reliability*.

Zhang, C., Lim, P., Qin, A. K., & Tan, K. C. (2016). Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. *IEEE transactions on neural networks and learning systems, 28*(10), 2306-2318.