Methodology on Establishing Multivariate Failure Thresholds for Improved Remaining Useful Life Prediction in PHM

Wenzhe Li [1,2], Xiaodong Jia [1,2*], Yuan-Ming Hsu[1,2], Youwen Liu[2] and Jay Lee[1,2]

[1]*Intelligent Maintenance Systems (IMS) Center,*

[2]*Department of Mechanical Engineering, University of Cincinnati, PO Box 210072,*

Cincinnati, Ohio 45221-0072, USA

li2w6@mai.uc.edu
jiaxg@ucmail.uc.edu

ABSTRACT

Prognostic approaches commonly try to predict the Remaining Useful Life (RUL) based on machine health status by either directly establish a mapping or setting up a failure threshold to determine the End-of-Life (EoL). On the one hand, determining a failure threshold is crucial but subjective for most reported cases. Machine operation risks, which are intuitional but difficult to quantify, can be used to bridge the gap between prediction and determining a multivariate failure threshold. On the other hand, historical machine life information is rarely considered together with the condition indicators for such prognostic tasks. Building multivariate failure thresholds based on quantifiable operation risks for prognostic tasks is the general topic that is rarely studied due to the following challenges: 1) How to quantify operation risks under multiple variables? 2) How to determine the multivariate failure thresholds? 3) How to make reliable extrapolations of future conditions? To address these questions, as the extension of our previous work (Jia, Li, Wang, Li, & Lee, 2020), this paper proposes 1) a Gaussian Copula model-based risk quantification method to determine multivariate failure thresholds, and 2) a Similarity enhanced Blackwellized Particle Filter (RBPF) to predict future system conditions. Two examples of establishing tri-variate and bi-variate failure thresholds are given. The proposed methodology is validated on the aero-engine RUL prediction task based on the C-MAPSS dataset from the PHM society data competition 2008. The result suggests that the proposed methodology has better explainability and practicability with comparable prediction capability.

1. INTRODUCTION

Prognostics and Health Management (PHM) methodologies and techniques have been widely studied in academia and practiced by the industry in recent years. The prognostics seek to predict the RUL based on machine operating status described by measurements from condition monitoring (CM). One popular way is to build a direct mapping (e.g., a deep learning regression model) from machine status to RUL(Khan & Yairi, 2018; Y. Wang, Zhao, & Addepalli, 2020). The other primary method type, such as Stochastic Process Method (SPM), Similarity-Based Method (SBM), and State-Space Method (SSM), could primarily rely on a failure threshold to determine the EoL (Cai, Feng, Li, Hsu, & Lee, 2020; Haoshu Cai, Xiaodong Jia, et al., 2020; H. Cai et al., 2020; Jia et al., 2019). The study to resolve systematically setting up failure thresholds for prognostic tasks are rarely found since, in most cases, such threshold is set empirically by human experience or based upon some simple statistic distributions (P. Wang & Coit, 2007). As emphasized in our previous work (Jia et al., 2020), operation risk is an essential criterion for critical systems and essential input of maintenance planning. However, it is not easy to be quantified based upon operating conditions. It would be practical for practitioners to set up a failure threshold to determine EoL based on a quantifiable operation risk.

Multivariate survival analysis could be applied to quantify operation risk based on reliability and operating condition, which generally takes life and its covariates to analyze the expected duration of time until machine failure. Effective models include multivariate distribution, competing risks models (Prentice et al., 1978), Cox Proportional Hazards Model (Breslow, 1975), and Copula Model (Clayton, 1978). Our previous work (Jia et al., 2020) proved that COX Proportional Hazard Model has the over-dominant issue and multivariate Weibull distribution has the independence

assumption of input variables. Since machine life can be correlated to some indicator of aging over time, the independence assumption does not always hold.

To address these challenges, this study proposes 1) a Gaussian Copula model-based risk quantification method to determine multivariate failure thresholds, and 2) a Similarity enhanced Blackwellized Particle Filter (RBPF) to predict future system conditions. The proposed Gaussian Copula Model could establish the joint effects of life and multiple variables. Life is essentially used in reliability analysis, and other variables can be health indicators or raw sensory readings. Based on the proposed risk model, a systematic methodology for RUL prediction is proposed. Two cases of establishing bi-variate and tri-variate failure thresholds are demonstrated and discussed. The proposed RUL prediction methodology is validated on the aero-engine RUL prediction task based on the C-MAPSS dataset from the PHM society data competition 2008.

The rest of the paper is organized as follows. Section 2 gives a literature review of the Copula Model and its application. Section 3 elaborates the proposed methodology. Section 4 illustrates the benchmarking results based on the public datasets. The conclusions are given in Section 5.

2. LITERATURE REVIEW

As most widely used to depict dependence relations in economics (Trivedi & Zimmer, 2006), clinical chemistry and finance (Cherubini, Luciano, & Vecchiato, 2004), quantifying risks in insurance (McNeil, Frey, & Embrechts, 2015), and establishing process control chart (Busababodhin & Amphanthong, 2016), Copula models create a link between multivariate joint distributions and univariate marginal distributions (Glidden, 2000; Nelsen, 2007; Patton, 2012). Regular Copula Models can be classified into T-student Copula (from which Gaussian Copula is derived) and Archimedean Copulas, which include Frank, Gumbel, Clayton (Clayton, 1978), and Joe Copula distinguished by different Copula functions(Kumar, 2019; Nelsen, 2007). As the most prominent Copula model, Gaussian Copula could easily expand to high dimensionality (more than three) compared with the most asymmetric bivariate Copulas. However, Gaussian Copula could only model linear correlation and not account for tail dependence (Klein, Kneib, Marra, & Radice, 2020). Such tail dependence can be modeled using Clayton in bivariate cases. In a general categorization scale, copula models can be classified as fully parametric, semiparametric (Chen & Fan, 2006) and nonparametric (Genest & Segers, 2009). Maximum Likelihood Estimation (MLE) or multi-stage MLE can be directly used to estimate model parameters for the parametric models such as Gaussian Copula and T-student Copula (Lindskog, 2000). As for parametric models such as Clayton and Frank, numerical optimization of log-likelihood function

is expected to estimate model parameters(Bouyé, Durrleman, Nikeghbali, Riboulet, & Roncalli, 2000).

Bivariate Copula is more common and abundantly studied than high-dimensional Copula models, limiting the choice when multiple variables exist. A suitable Copula model is generally case-sensitive (Busababodhin & Amphanthong, 2016). However, it still provides a powerful tool to build the joint influence of life and other health indicators towards risk quantification. As reviewed in (Jia et al., 2020), methodologies of determining failure threshold rarely take risk into consideration. Some proposed methods are very complex and inflexible (Jiang, 2010) (Javed, Gouriveau, & Zerhouni, 2013). To address these challenges, this study proposes a Gaussian Copula model-based risk quantification methodology to assist in establishing multivariate failure threshold for prognostic tasks.

3. METHODOLOGY
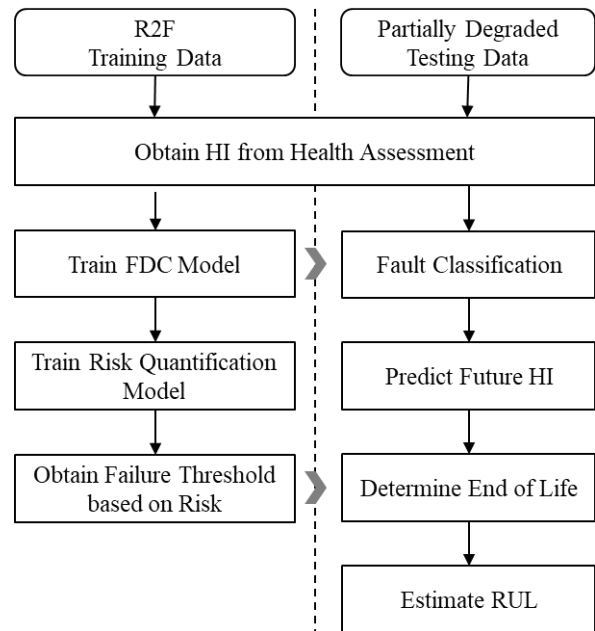
3.1. Methodology Overview

Figure 1. Methodology workflow of predicting RUL of engines based on multivariate failure threshold and Similarity enhanced RBPF

The workflow of the proposed methodology and its application on C-MAPSS Engine RUL prediction is shown in Figure 1.

1) **Obtain HI from health assessment**: As suggested in (Jia et al., 2020), Principal Component Analysis is conducted on Run to Failure (R2F) readings of sensor #2, #3, #4, #11, #17, which are selected using Fisher distance and Mann-Kendall test. The obtained 1-

dimensional principal component is treated as the health indicator.

2) **Train Fault Detection and Classification Model:** The C-MAPSS dataset consists of two failure types: FAN degradation and High Pressure Compressor (HPC) degradation, which can be easily clustered by the K-means method. The obtained label is used for training a Fault Detection and Classification (FDC) model, which is further applied to determine the failure threshold model for testing data.

3) **Train Risk Quantification Model:** The risk quantification model calculates the operation risk given the machine operation time and HI. Two models are trained for FAN degradation and HPC degradation failure mode, respectively. The multivariate failure threshold can be obtained based on the given risk value. The detailed procedure and discussion will be given in Section 3.2.

4) **Predict Future HI:** This step extrapolates the partially degraded HI trajectories from the testing samples using a similarity enhanced RBPF. The future HI values are estimated together with the prediction uncertainties. The EoL can be determined by applying the risk-based failure threshold. Details of HI prediction are given in section 3.3.

3.2. Train Risk Quantification Model

The system operation risk can be addressed in terms of failure time (reliability) and its covariates (other indicators of operation status) in the presence of dependence. Dependent failure times are typical when system components degrade over time. As for some critical systems, safe is defined as when the machine is reliable in terms of time and healthy in terms of functionally. Risk can be therefore defined as not being able to survive and remain healthy. The mathematical expression of the risk in this study is described in Eq.(**1**).

$$R = 1 - S(t, x_1, \ldots, x_n)$$
$$= 1 - \mathbf{Prob}(L \geq t, X_1 \geq x_1, \ldots, X_n \geq x_n) \quad (1)$$

Where L denotes the Life and $X_1, \ldots, X_n$ are indicators (crucial process measurement) of operation status, respectively, $t, x_1, \ldots, x_n$ represents the current time and current indicators' value. $S(\cdot)$ is the survival function meaning the joint probability of survival. The risk R is defined as $1 - S(\cdot)$.

Variables' dependencies affect ways of calculating the joint probability of survival are classified. The joint cumulative distribution function (d.f.) can be calculated simply as the product of the marginal cumulative d.f. when all variables are independent, which is however, not always valid. The degradation tendency of some components could have strong correlations with machine life. Some degradation pattern can also be reflected by different sensors. The potential complex

dependencies make the independent assumption of indicators and time not always hold when calculating the joint probability. Copula model provides a simple but effective solution to calculate such complex joint probability.

Let H be an (n) dimensional d.f. with continuous marginal d.f. $H_{X1}, H_{X2}, \ldots, H_{Xn}$, then the copula model of H can be defined as Eq. (**2**). $C(H_{X1}(x_1), H_{X2}(x_2), \ldots, H_{Xn}(x_n))$ is a multivariate d.f. with margins $H_{X1}, H_{X2}, \ldots, H_{Xn}$. To construct such a multivariate d.f. using a copula model, one can 1) specify a marginal univariate distribution for each variable and 2) choose a copula to provide a correlation structure between variables.

$$H(x_1, x_2, \ldots, x_n)$$
$$= C(H_{X1}(x_1), H_{X2}(x_2), \ldots, H_{Xn}(x_n)) \quad (2)$$

In this case, the Gaussian copula model is used to construct the joint distribution due to its simplicity (Gaussian assumption) and high dimensional capability (for tri-variate and higher). The general form of Gaussian copula is shown in Eq.(**3**), in which ρ is the correlation parameter. To train and test a Gaussian copula model, the probability-integral transformations (to uniformity) should first be performed to map the raw variables into the unit cube $[0,1]^n$ (Lindskog, 2000). Afterward, the model parameter θ can be estimated based on S samples by maximizing Eq.(**4**) using MLE. In this case f_n is the density function of Gaussian distribution for each variable. The obtained model can be future used to calculate any joint probability as shown in Eq.(**5**) according to the Sklar properties (Sklar, 1996)

$$C(u_1, u_2, \ldots, u_n; \rho)$$
$$= \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n)) \quad (3)$$

$$\ell(\theta) = \sum_{s=1}^{S} \sum_{n=1}^{N} \ln f_n(x_n^s; \theta_n) +$$
$$\sum_{s=1}^{S} \ln C\big(\Phi_{X1}(x_1^s; \theta_1), \Phi_{X2}(x_2^s; \theta_2), \ldots, \Phi_{XN}(x_N^s; \theta_N)\big) \quad (4)$$

$$C_n(u, v) = C(1, \ldots, 1, u, v, 1, \ldots, 1) \quad (5)$$

To illustrate the training procedure and discuss the key features of the proposed risk quantification model, a tri-variate example is shown based on life, sensor #2 and sensor #4 (denoted as 'Sensor 1' and 'Sensor 2' for simplicity) from C-MAPSS dataset. These two representative sensors are selected based on the distribution disparity at healthy and faulty operation status as shown in Figure 2 (a) and (b). The faulty data generally follow a normal distribution.

$$R = 1 - \mathbf{Prob}(L \geq t, X_1 \geq x_1, X_2 \geq x_2)$$
$$= H_L(t) + H_{X1}(x_1) + H_{X2}(x_2)$$
$$-H_{X1,X2}(x_1, x_2) - H_{X1,L}(x_1, t) - H_{X2,L}(x_2, L) \quad (6)$$
$$+H_{X1,X2,L}(x_1, x_2, t)$$

The operation risk can be calculated using Eq.(**6**) in which $H_L(t) = \mathbf{Prob}(L < t)$ is the marginal cumulative d.f. of life at t. $H_{X1,X2}(x_1, x_2)$ is the joint cumulative d.f. of X_1 and X_2 at (x_1, x_2). $H_{X1,X2,L}(x_1, x_2, t)$ is the joint cumulative d.f. of L, X_1 and X_2 at (t, x_1, x_2).

The sensor readings at EoL are collected and used to train the tri-variate Gaussian copula model. Based on the copula model, the marginal d.f. and bivariate joint cumulative d.f. can be obtained according to Eq.(**5**). The risk value can be therefore calculated given a vector of input $[x_1, x_2, t]$. Based on this risk quantification model, a multivariate failure threshold appeared as a (n) dimensional surface at different risk values from $[0,1]$. Figure 4 shows the failure threshold of FAN degradation at risk level 0.1, 0.3, 0.5, 0.7, 0.9, 0.99 respectively. The EoL can be easily determined for an extrapolated degradation trace.
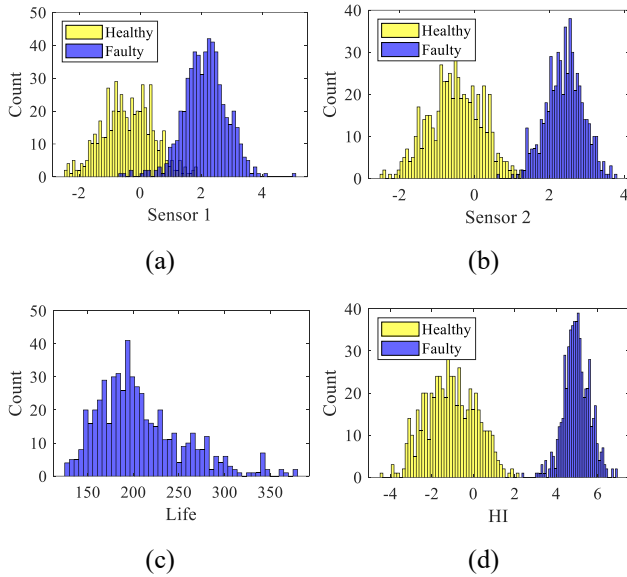

(a)

(b)

(c)

(d)

Figure 2. The distribution of data used for training the risk quantification models, where (a) and (b) are distributions of healthy and faulty data for Sensor 1 and Sensor 2; (c) is the distribution of EoL, and (d) is the distribution of HI at healthy and faulty.

Several conclusions can be drawn from Figure 3: 1) As the tolerable risk increase, the threshold surface expands in all directions, covering more EoL data used for training; 2) the shape of the threshold surface is dominated by one variable on the rims and jointly determined by all variables at the central area; 3) The percentage of training EoL data within the threshold is lower than the risk value, indicating even at the risk of 0.99, there is still substantial historical machines alive. This is caused by the strict definition of risk which is the contrary event of 'healthy and alive'. This definition suits the critical system or critical failure mode where one failure could lead to dangerous consequences. It can also be addressed that each sensor is equally weighted with life in

this model, which can be too strict for some cases. The proposed risk quantification method could be a powerful tool to determine multivariate failure threshold when considered sensors are equally critical and indicate the system's health status in a certain manner. Otherwise, to avoid such tight assumptions, applications could consider health assessment in advance and build a bi-variate risk quantification model rather than a high-dimensional risk quantification model.
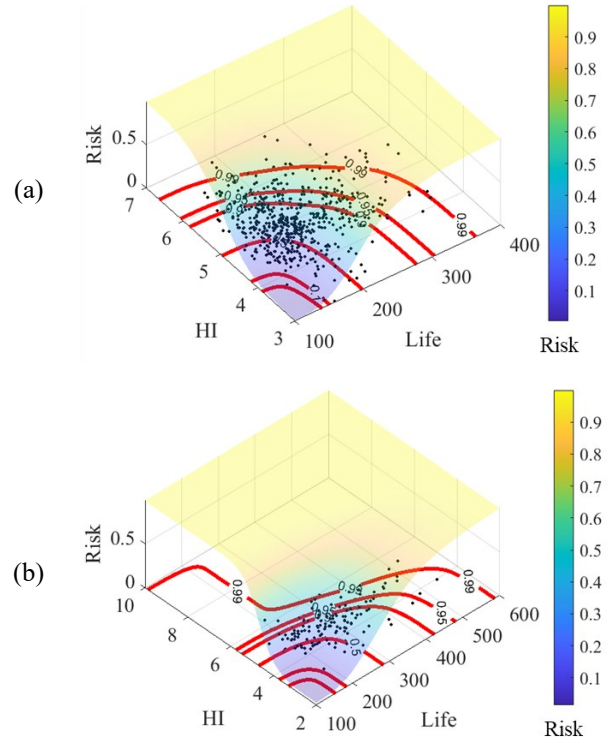

Figure 3. Risk quantification model of a) FAN degradation mode and b) HPC degradation mode given by the Gaussian Copula model.

In practice, the standard of choosing tolerable risk varies from case to case and largely depends on expert experience. When historical R2F data is available, a suitable risk level could be determined by the cross-validation strategy, where the data is split in multiple folds for training and validating the risk quantification model. Since the EoL information is available, practitioners could quickly obtain the corresponding risk level for further use. To apply this methodology for data competition tasks like the C-MAPSS RUL prediction in the case study, the risk level could be tuned as a hyperparameter for better prediction accuracy. In this case, 75% is selected to achieve the most balanced prediction performance on all four datasets.

To obtain suitable failure thresholds for the C-MAPSS RUL prediction, this study built a bi-variate risk quantification model based on HI and life, as shown in Figure 2 (c) and (d).
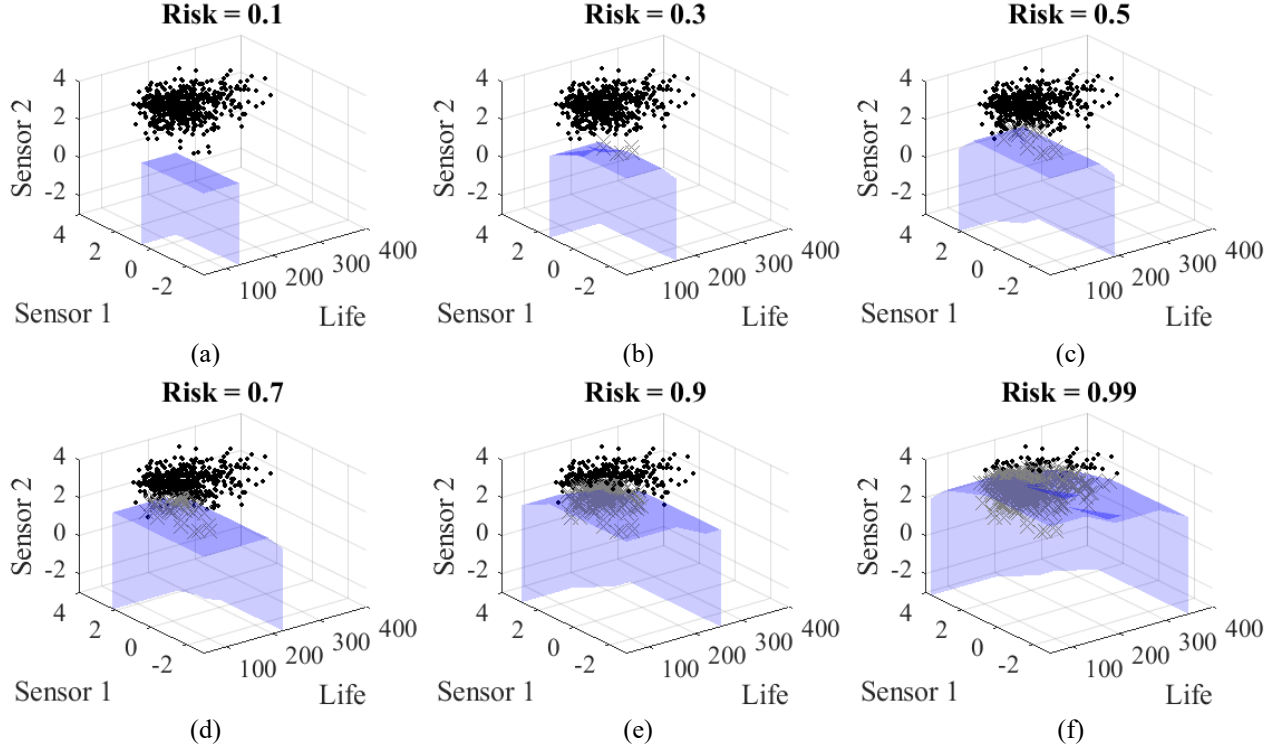
Figure 4. The illustration of obtained tri-variate (Sensor 1, Sensor 2, Life) failure threshold at risk level (a) 0.1, (b) 0.3, (c) 0.5, (d) 0.7, (e) 0.9, (f) 0.99. The black dots are reference EoL data beyond the current threshold and grey crosses are reference EoL data within the current threshold

It could also benefit the model since the faulty distribution of HI is much thinner and separated from the healthy distribution compared with Sensor 1 and Sensor 2. The obtained 2-d failure threshold for FAN degradation and HPC degradation are shown in Figure 3 (a) and (b). The HPC degradation has a relatively longer tail in terms of life.

3.3. RUL Prediction using Similarity Enhanced RBPF

RBPF is used in the proposed work for making robust extrapolation of future HI with uncertainty based on partially degraded data. Compared with the traditional Particle Filter and Kalman Filter, RBPF is capable of carrying out analytical marginalization to enhance filtering and extrapolating performance (Särkkä, Vehtari, & Lampinen, 2007).

$$y = a \cdot \exp(b \cdot t) + c \tag{7}$$

$$\begin{cases} c_t = c_{t-1} + \epsilon_c \\ b_t = b_{t-1} + \epsilon_b \\ x_t = (x_t - c_t) \cdot e^{b_t \cdot \Delta t} + c_t + \epsilon_u \\ y_t = x_t + \epsilon_y \end{cases} \tag{8}$$

Eq.(7) is built to regulate an exponential degradation trend where a, b, c are the unknown parameters and y is the observed HI at time t. The RBPF can be therefore built as Eq.(8) to predict HI, in which x_t is estimated HI at t, b_t, c_t are model parameters at t. The $\epsilon_c, \epsilon_b, \epsilon_u$ and ϵ_y in Eq.(8)

follow Gaussian distribution with zero-mean and whose variances are tuned to the best model performance.

The implementation of the RBPF includes two steps: filtering and extrapolating. During the filtering step, the unknown model parameters $\Theta = \{b, c, x, \sigma_u, \sigma_y\}$ are estimated based on the partial degradation. Based on the estimated Θ^t at current time t, the prediction can be obtained by extrapolating the state equation (3rd equation) in Eq.(8).

For each testing sample, the initialization of Θ could significantly impact extrapolation integrity and uncertainty. To make a reasonable initialization, a similarity matching based approach is introduced to obtain critical RBPF parameters b, c and σ_b, σ_c. The workflow of the proposed Similarity enhanced RBPF is as following:

1) Estimate $\{\boldsymbol{\eta}_s\}_{s=1,\dots,S}$ for each historical R2F record using MLE, where $\boldsymbol{\eta}_s = \{b_s, c_s\}$ and S denotes the number of historical R2F records.

2) For each partially degraded testing sample, matching a similar R2F trace and record its $\boldsymbol{\eta}_s$. The R2F trace is considered similar when a window's Euclidian distance to the testing data is smaller than a small number ε. A set of K similar parameters $\{\boldsymbol{\eta}_s\}_{s=1,\dots,K}$ is obtained for each sample.

3) Estimate b, c, σ_b, σ_c based on $\{\boldsymbol{\eta}_s\}$ and obtain initial Θ

4) For each testing unit with partial degradation data, initialize N particles with Θ. At the end of the filtering process, N different filtering results at current time t are obtained, which is denoted as $\left\{\Theta_{s.p}^{t}\right\}_{p=1,...,N}$

5) Extrapolate the future trend by using entries in $\left\{\Theta_{s.p}^{t}\right\}_{p=1,...,N}$ as the starting point of the HI prediction. And use the failure threshold to determine EoL.
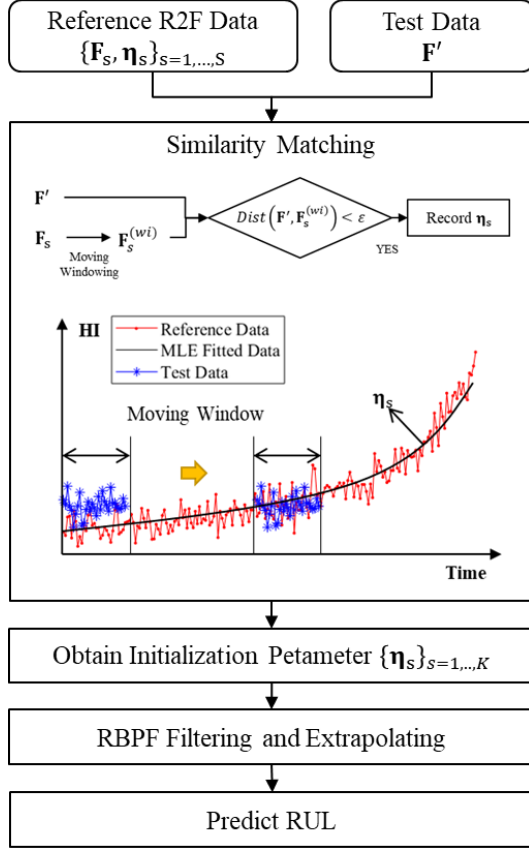

Figure 5. Workflow of RUL prediction based on similarity enhanced RBPF.

4. RESULTS & DISCUSSIONS

The proposed methodology for risk quantification and failure threshold determination is validated on the RUL prediction task on the C-MAPSS dataset. The prediction performance of proposed methods shows superiority over benchmarked methods from some recent literature in terms of Root Mean Squared Error (RMSE) and Score calculated using Eq.(9). The benchmarked methodologies include deep Convolutional Neural Network (DCNN) and Long Short-Term Memory (LSTM) from (Li, Ding, & Sun, 2018), Deep Belief Networks (DBN), Multi-Objective Deep Belief Networks Ensemble (MODBNE), Random Forest (RF), Gradient Boosting (GB), Support Vector Machine (SVM) and Least absolute shrinkage and selection operator (LASSO) reported from (Zhang, Lim, Qin, & Tan, 2016), and

RULCLIPPER from (Ramasso, 2014). As noticed by literature, a constant value R_{early} (set to 125) is assigned as target RUL before the degradation period. When R_{early} is applied, the largest prediction RUL is limited to 125. For illustration, the prediction result of 75% risk level is summarized in Table 2 and Table 3 together with the obtained best result.

Dataset	Failure Mode	Conditions	N_{train}	N_{test}
FD001	HPC	1	100	100
FD002	HPC	6	260	259
FD003	HPC, FAN	1	100	100
FD004	HPC, FAN	6	249	248
Total	--	--	709	707

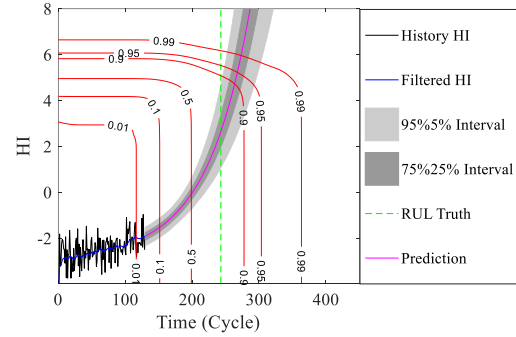Table 1. Summary of C-MAPSS Dataset

Figure 6. RUL prediction for test unit 23 in FD001, which is a FAN degradation
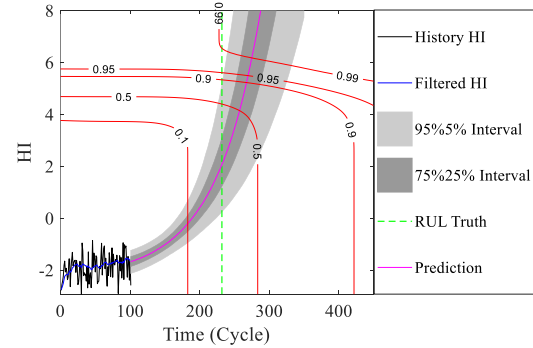

Figure 7. RUL prediction for test unit 400 in FD003, which is a HPC degradation

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

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

(9)

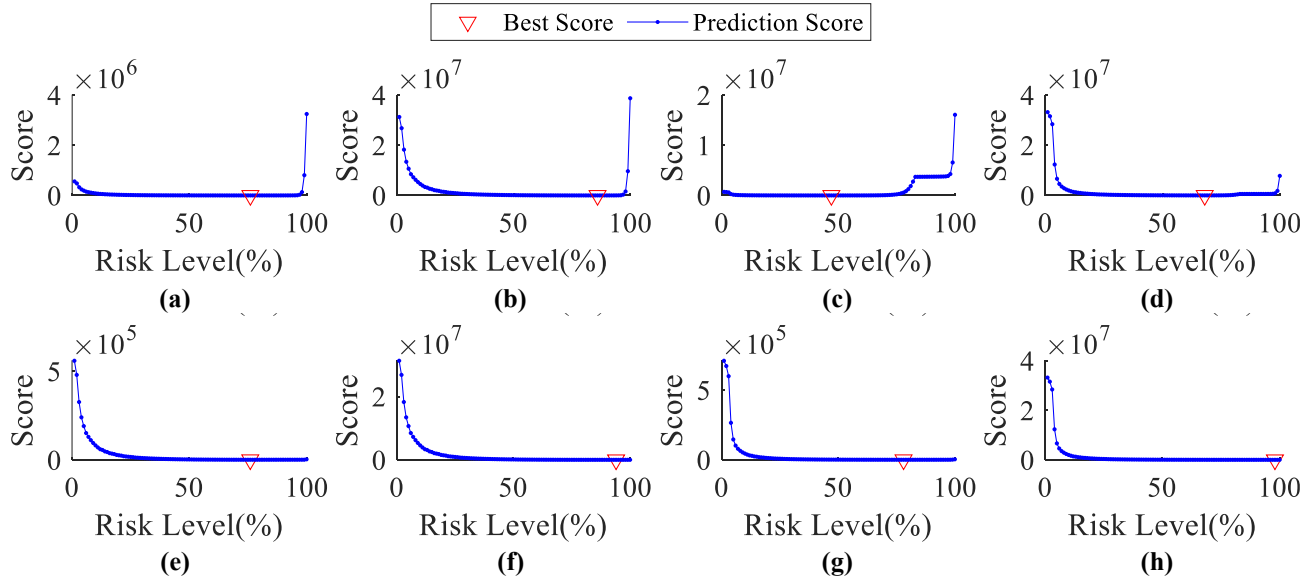where $d_i = y_i - y_i^*$ (estimate RUL − True RUL)

Figure 8. The change of prediction score by varying the tolerable risk level. (a)-(d) indicates the prediction score for FD001 to FD004 when no R_{early} is applied. (e)-(h) indicates the score for FD001 to FD004 when $R_{early} = 125$

		Proposed Method (Best | Risk)		Proposed Method (Risk: 75%)	MODBNE	DBN	RF	GB	SVM	LASSO
FD001	RMSE	**17.44**	**76%**	17.65	15.04	15.21	17.91	15.67	40.72	19.74
	Score	**415.39**		420.40	334	417	479	474	7703	653
FD002	RMSE	**23.61**	**86%**	29.09	25.05	27.12	29.59	29.07	52.99	37.13
	Score	**3869.2**		9356.27	5585	9031	70465	87280	316483	276923
FD003	RMSE	**25.5**	**47%**	30.97	12.51	14.71	20.27	16.84	46.32	21.38
	Score	**2073.4**		222670.86	421	442	711.13	576	22541	1058
FD004	RMSE	**33.2**	**68%**	31.53	28.66	29.88	31.12	29.01	59.96	40.7
	Score	**18436**		42880.46	6557	7954	46567	17817	141122	125297
R_{early}		**NA**		NA	NA	NA	NA	NA	NA	NA

Table 2. Prediction Result Comparison with Benchmark Methods (R_{early} Not Applied)

		Proposed Method (Best | Risk)		Proposed Method (Risk: 75%)	DCNN	LSTM	RULCLIPPER
FD001	RMSE	**15.63**	**76%**	15.79	12.61	13.52	13.27
	Score	**415.39**		420.40	273	431	216
FD002	RMSE	**14.9**	**94%**	14.88	22.36	24.42	22.89
	Score	**3971.5**		9356.27	14459	10412	2796
FD003	RMSE	**15.67**	**78%**	15.85	12.64	13.54	16
	Score	**478.5**		485.05	284	347	317
FD004	RMSE	**17.67**	**98%**	17.53	23.31	24.21	24.33
	Score	**5671**		9145.20	12466	14322	3132
R_{early}		**125**		125	125	125	125

Table 3. Prediction Result Comparison with Benchmark Methods (R_{early} Applied)

Figure 6 and Figure **7** illustrate how the proposed bi-variate failure threshold determines the EoL and predict RUL using the RBPF based extrapolations. Some significant findings are as follows: 1) the failure threshold determination model of FAN and HPC degradation have a significant difference in terms of life, which justify the importance of FDC before quantifying operation risks; 2) the extrapolated HI has reasonable exponential shape and uncertainty range, suggesting the effectiveness and necessity of applying the similarity enhanced RBPF; 3) the prediction uncertainty of HI can be easily converted to RUL prediction uncertainty with the failure thresholds as contour lines; 4) the EoL point has a high change appearing in the complex joint, which shows the necessity of depicting the correlation of HI and life and justified the capability of copula model for such tasks; 5) the ground truths can be spotted in the rational middle of all risk levels, indicating the effectiveness of the proposed methodology.

The effects of varying tolerable risk on the prediction score are illustrated in Figure 8. It can be seen from (a)-(d) when R_{early} is not applied, an optimal score is found for FD001 to FD004 when the risk levels are 76%, 86%, 47%, 68%. This suggests that practitioners can flexibly obtain such a useful failure threshold based on practice operation risk. This risk value can statistically reflect the probability of survival under healthy status. Comparatively, the optimal scores are found on relatively higher risk levels when R_{early} is applied. This is because R_{early} generally removes the penalty of scores when a longer RUL is predicted, which could accept a higher risk level. The risk level for each dataset to achieve the best score is summarized in Table 2 and Table 3 which compare the RMSE and score performance of the proposed method against benchmark methods with $R_{early} = 0$ and 125.

As seen from Table 2 where R_{early} is not applied, the proposed model has comparable score performance compared with the best among other methods on FD001 and FD002. It can be highlighted that the prediction Score of FD002 is 3869.2, which is significantly better than all the other methods. When R_{early} is applied, as suggested in Table 3, the proposed methodology holds comparable score performance with all the benchmarked deep learning methods throughout FD001 to FD004. The major advantage of the proposed methodology over deep learning is its explainability, practicability, and lower requirement on data volume. The overall performance on each dataset is much better in RMSE and score compared with our previous work in (Jia et al., 2020). However, the proposed methodology does not show better accuracy on FD001 and FD003 than the benchmarked deep learning methods. This is because the training data for building the risk quantification model and the similarity base are unbalanced where more historical records are available under multiple operation conditions (6 conditions in FD002 and FD004). This could introduce a bias towards multiple conditions where the machine degrades

faster. Therefore, the overall prediction of RUL on FD001 and FD003 is conservative. Generally, the score performance is slightly worse on FD003 and FD004 than FD001 and FD002 due to the combination of multiple failure modes and operation conditions. The FDC model's misclassification could select wrong model for failure threshold determination, causing conservative or radical prediction results. The assumption of exponential degradation could make some parameters very sensitive to addressing the HI curve before the degradation started. This could also cause radical prediction on the HPC failure model when selecting a relatively larger risk level. The proposed method can be future improved using a more accurate FDC model and complex degradation function for RBPF extrapolation. Building a local model for each dataset could also increase prediction performance by addressing the data imbalance issue. However, this would increase modeling complexity and reduce the data used for training each model.

5. CONCLUSION

This paper proposes a Gaussian Copula model based multivariate risk quantification model to determine the failure threshold for RUL prediction tasks. A systematic methodology for RUL prediction is proposed based on the failure threshold determination method and the similarity enhanced RBPF extrapolator. The work discussed the pros and cons of the proposed risk quantification method with a tri-variate and a bi-variate example. The obtained bi-variate failure threshold is further validated on the aero-engine RUL prediction task using the C-MAPSS dataset. Several essential conclusions are achieved:

1) The proposed risk quantification model can easily quantify machine operation risk based on multiple input factors without variables' independence assumption.
2) The proposed failure threshold can effectively convert the predictive distribution of HI into the predictive distribution of RUL.
3) The tolerable risk can be flexibly tuned accordingly. If low risk can be tolerated, the life indicator dominates the threshold. If higher risk can be tolerated, the health indicator dominates the threshold.
4) The similarity enhanced RBPF could make extrapolation more reliable. It can still work well with limited reference data.
5) For a particular case, The RUL prediction performance could reach its optimal by varying the tolerable risk in RUL prediction tasks.

As for future work, the bivariate Copula model can be updated to those with better tail character (such as Clayton, Joe, etc.) to enhance the tails dependence of life. The definition of risk can be further studied to suit for the non-essential case with multiple variables. The proposed methodology can be further validated on the newly published C-MAPSS 2 dataset from the 2021 PHM Conference Data Challenge.

REFERENCES

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., & Roncalli, T. (2000). Copulas for finance-a reading guide and some applications. *Available at SSRN 1032533*.

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

Busababodhin, P., & Amphanthong, P. (2016). Copula modelling for multivariate statistical process control: a review. *Communications for Statistical Applications and Methods, 23*(6), 497-515.

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

Chen, X., & Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics, 130*(2), 307-335.

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula methods in finance*: John Wiley & Sons.

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.

Genest, C., & Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics, 37*(5B), 2990-3022.

Glidden, D. V. (2000). A two-stage estimator of the dependence parameter for the Clayton-Oakes model. *Lifetime Data Analysis, 6*(2), 141-156.

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., Li, W., Wang, W., Li, X., & Lee, J. (2020). *Development of Multivariate Failure Threshold with Quantifiable Operation Risks in Machine Prognostics.* Paper presented at the Annual Conference of the PHM Society.

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

Khan, S., & Yairi, T. (2018). A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing, 107*, 241-265.

Klein, N., Kneib, T., Marra, G., & Radice, R. (2020). Bayesian mixed binary-continuous copula regression with an application to childhood undernutrition. In *Flexible Bayesian Regression Modelling* (pp. 121-152): Elsevier.

Kumar, P. (2019). Copula functions and applications in engineering. In *Logistics, supply chain and financial predictive analytics* (pp. 195-209): Springer.

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.

Lindskog, F. (2000). Modelling dependence with copulas and applications to risk management. *Swiss Federal Institute of Technology Zurich*.

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: concepts, techniques and tools-revised edition*: Princeton university press.

Nelsen, R. B. (2007). *An introduction to copulas*: Springer Science & Business Media.

Patton, A. J. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis, 110*, 4-18.

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.

Särkkä, S., Vehtari, A., & Lampinen, J. (2007). Rao-Blackwellized particle filter for multiple target tracking. *Information Fusion, 8*(1), 2-15.

Sklar, A. (1996). Random variables, distribution functions, and copulas: a personal look backward and forward. *Lecture notes-monograph series*, 1-14.

Trivedi, P., & Zimmer, D. (2006). Copula modelling in econometrics: Introduction to practitioners. In: Wiley, New York.

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.

Wang, Y., Zhao, Y., & Addepalli, S. (2020). Remaining useful life prediction using deep learning approaches: A review. *Procedia Manufacturing, 49*, 81-88.

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.