

Spherical-Dynamic Time Warping - A New Method for Similarity-Based Remaining Useful Life Prediction

Xiaochuan Li^{a1*}, Shuiqing Xu^{a2}, Yingjie Yang^b, Tianran Lin^c, David Mba^d, Chuan Li^e

^{a1}School of Electrical Engineering and Automation, Hefei University of Technology, Hefei, China. Email: xiaochuan.li@hfut.edu.cn

^{a2}School of Electrical Engineering and Automation, Hefei University of Technology, Hefei, China. Email: xsqanhui@hfut.edu.cn

^bSchool of Computer, Engineering and Media, De Montfort University, Leicester, LE1 9BH, UK. Email: yyang@dmu.ac.uk

^cFaculty of Mechanical and Automotive Engineering, Qingdao University of Technology, Qingdao, China. Email: trlin@qut.edu.cn

^dFaculty of Computing, Engineering and the Built Environment, Birmingham City University, Birmingham, UK. Email: david.mba@bcu.ac.uk

^eSchool of Mechanical Engineering, Dongguan University of Technology. Email: chuanli@dgut.edu.cn

Abstract

Machinery prognostics and health management (PHM) plays a key role in the reliable and efficient operation of industrial processes. With the emerging big data era, data-driven prognostic methods which avoid considering complicated system models have attracted growing research interest. Among many data-driven models, similarity-based prediction methods have been popular due to their strong interpretability and relatively simple implementation process. Nevertheless, when quantifying the similarity between two trajectories, most existing similarity measures neglect the nonlinearity of the distance measurement at different degradation stages and degradation alignments with timing difference, which may not be sufficient to retrieve the most suitable trajectories for remaining useful life (RUL) prediction. To overcome these limitations, a spherical-Dynamic Time Warping (spherical-DTW) algorithm is put forward to find an optimal match between the test and training trajectories at the retrieval step. Dynamic Time Warping allows degradation alignments with timing difference through stretching or compressing the trajectories with regard to time, thereby the data in similar degradation levels can be well aligned across different units. Moreover, a newly defined nonlinear spherical distance method is introduced and incorporated into the retrieval process to account for the nonlinearity of the damage propagation process. The significance of this study is that the newly proposed spherical-DTW algorithm goes one step further to consider the nonlinearity of fault evolutions and allow degradation pattern alignments with timing difference when performing similarity-based prognostics. Two run-to-failure cases, involving a real-world industrial compressor failure case and a gas turbine engine failure dataset, are investigated to demonstrate the effectiveness and superiority of the proposed algorithm.

Keywords: Condition monitoring; Prognosis; Dynamic time warping; spherical distance.

1. Introduction

Modern mechanical systems, such as gas turbine engines, compressors, wind turbines, etc., are usually featured by adverse working conditions, demanding service performance and high failure costs, and are subject to performance degradation and mechanical failures (X. Li et al., 2022). Thus, it is of great importance to precisely detect the potential malfunctions of the machines and accurately predict their remaining useful life (RUL), which can assist the decision-making process so that timely maintenance can be carried out before catastrophic failure occur (J. Wu et al., 2018). Prognostics and health management (PHM) strategy

involves two major tasks, namely fault detection and fault prognostics. While fault detection mainly deals with the detection and isolation of faults that have incipient indications, fault prognostics focus on the estimation of time left before the system reaches the safe operational limit. The currently PHM methods can be generally categorized into three classes, namely model-based, data-driven and hybrid approaches (X. Li et al., 2021). Among them, data-driven methods have been the most popular in both academics and industries in recent years due to the fact that they are easy to implement and are data-dependent instead of physics knowledge-dependent, and there are many research papers written on this topic (Lei et al., 2018; X. Li et al., 2021; Loukopoulos et al., 2019; T. Wang et al., 2008b).

Among many data-driven models, similarity-based prediction methods have been gaining growing attention due to the great merits of easy implementation and their strong interpretability. Apart from similarity-based methods, Many state-of-the-art data-driven prognostic methods, including deep learning methods (Ding et al., 2023; Listou Ellefsen et al., 2019; Ren et al., 2022; Sun et al., 2023; Tian et al., 2023) and machine learning methods (Loutas et al., 2019; G. Ma et al., 2023) involve a relatively complex model design and training process in order to build mappings between the sensory measurements (inputs) and the RUL (outputs). The mapping process is usually carried out through parameter tuning, modelling training and testing, and the resulting models have been treated as black boxes: their mechanism of operation remains unknown. Statistical methods such as Wiener process, has also been applied to describe the nonlinearity and stochasticity of the system degradation (Y. Zhang et al., 2022; Z. Zhang et al., 2017). Wiener process model is suitable for prognostic tasks where limited data are available. In Wiener process modelling, much more attention has been paid to the drift coefficient function rather than the diffusion coefficient function in both the academic and engineering fields. Among all the drift coefficient functions, the linear model, power-law model and the exponential-law model (Zhai & Ye, 2017; Z. Zhang et al., 2017) gain the most popularity in the scientific community when nonlinear degradation processes were modelled. Unlike the similarity methods that we proposed, these fixed model structure will struggle to harness the degradation features with

complicated nonlinear behaviors. Similarity-based prognostics, on the other hand, is an approach to solve new problems by retrieving stored 'cases' describing similar prior problem-solving episodes and adapting their solutions to fit new problems. The idea is to extract proper health indicators (HIs) from the running machine that can reflect the health status of the system, and then to find the top few training HIs which possesses similar degradation trends with the test HI and use their failure times to predict the RUL of the test machine. The RUL prediction procedure generally consists of three main steps: instance retrieval from the training HIs, prediction through local model and aggregation of local predictions. Similarity-based prognostics' application in machinery RUL prediction has gained popularity in recent years. Similarity calculation and similarity matching rules are two important tasks in similarity-based prognostics. In similarity calculation, the similarity value of two given HIs segments is calculated using a distance measure. Liu et al. (2007) proposed to use the Mahalanobis distance and kernel function mapping to conduct the similarity calculation between two multi-dimensional data vectors. Euclidean distance was adopted as a distance measure in similarity-based prognostics and was applied to engineered systems in (T. Wang et al., 2008a). Similarly, in (Loukopoulos et al., 2018) a similarity-based K-Nearest Neighbours Regression method was employed to predict the RUL of an industrial compressor, where the Euclidean distance was utilized for similarity calculation. Based on Euclidean distance, the pointwise difference and the shape factor were considered in the similarity calculations in (Zeming et al., 2019). In (H. Wang et al., 2020), Multiple similarity calculations are combined to obtain a comprehensive similarity value. The Euclidean distance and cross-correlation similarity measure function were combined to form a new similarity measure function. In (Hou et al., 2020), the effect of using similarity to improve the prediction accuracy of deep supervised learning was investigated, where Euclidean distance was used to measure the distance between the original input and the reconstructed HIs. In (W. Ma et al., 2022), a physical-oriented multi-scale feature similarity was calculated for rotating machine performance assessment. In (Yu et al., 2020), the cosine distance, measured the spatial distance in the high-dimensional space, was employed to evaluate the similarity

between two HI segments. In (Y. Liu et al., 2019), Manhattan distance was considered for similarity calculation as well. In (H. Cai et al., 2020; Jia et al., 2019), a statistical distance measure called maximum mean discrepancy (MMD) was utilized to measure the distance between two HI segments, which evaluates the average embedding distance of the two distributions in kernel space. To cope with the noise and linear transformation of HI segments, the normalized correlation coefficient, originated from the image processing techniques, was applied to the similarity calculation of multi-dimensional condition monitoring data in (X. Wu et al., 2020). Another important task in similarity-based prognostics is to choose a proper similarity matching rule so that the most similar HIs can be found for similarity calculation. A number of studies have been conducted to design similarity matching rules in order to improve the overall prognostic performance in the presence of challenges such as limited HI instances. In (Kim et al., 2020), a dynamic time warping (DTW) technique was utilized for measuring the similarity between data so that a number of virtual run-to-failure data can be created for RUL prediction. DTW is a time sequence matching rule considering the time delay. Different methods have been proposed to improve the existing similarity matching rules. In (Huang et al., 2019; Yu et al., 2020, 2021), the similarity calculations of different timestamps are considered as of the same importance. A zero-centering rule was proposed to solve the problem that similar instances might be missed during the similarity evaluation due to the different initial degradation states of the machines in (Yu et al., 2020). A new similarity matching rule which checks the degradation consistency of the predicted referential instances was proposed in (Xue et al., 2022a), which can eliminate the prediction referential instances that result in marked prediction errors. In (Gu & Chen, 2018), a customized similarity matching which considers assigning unequal weights for different timestamps was proposed. It gives more weights to the similarity calculations of timestamps after the maintenance. Similarity matching procedure was conducted for different degradation states to improve the prognostics performance in (H. Wang et al., 2020). Based on the assumption that systems that have undergone similar events will probably degrade similarly in the future, the authors of (Lyu et al., 2020) put forward a similarity matching rule

considering the position of timestamps and system degradation rate. Recently, a review paper was written on the topic of similarity-based prognostics (Xue et al., 2022b). Besides, many existing similarity-based prediction papers have considered the nonlinear features associated with prognostics, such as (Bleakie & Djurdjanovic, 2013; J. Liu et al., 2007; Xue et al., 2022b).

Nevertheless, many of the existing research neglected two important research questions in terms of using similarity-based prognostics: 1) In many of the studies reviewed, only one-to-one point pattern matching was considered, making it difficult for degradation patterns with timing difference to be matched; 2) The degradation nonlinearity (i.e. the damage severity and its propagation speed of mechanical systems are often vary with time) was rarely considered in the design of a distance measure, making it difficult to accurately capture the underlying degradation process. To address these challenges, we put forward a new method called spherical-DTW for similarity-based prognostics.

Nevertheless, many of the existing research neglected two important research questions in terms of using similarity-based prognostics. 1) a research question that is often neglected by researchers in the field of PHM is the nonlinearity of the propagation of damage. This nonlinearity mainly manifests in the variable deterioration rate of a unit at different time steps. To be specific, according to the previous studies of damage propagation modelling, including the popular damage propagation modelling of gas turbine engines conducted by NASA (Saxena et al., 2008), the observation of similar degradation trends in practice (Goebel et al., 2007), fatigue crack models of compressor valves (Dusil & Appell, 1976), degradation modelling of rolling element bearings (N. Li et al., 2015), despite local variations, the overall rate of damage propagation show an increasing trend as the equipment approaches its failure point. Therefore, data in late cycles should be in general accorded more importance when performing distance measurement since the failure is more likely to occur by the end of life due to the ascended degradation rate. 2) In many of the studies reviewed, only one-to-one point pattern matching was considered, making it difficult for degradation patterns with timing difference to be matched. Successful similarity matching is based on the ability to

accurately match degradation trajectories despite mis-alignment issues caused by timing difference and variations in deterioration speed.

To address these challenges, we put forward a new method called spherical-DTW for similarity-based prognostics. In this method, we define a nonlinear distance measure called spherical distance to capture the nonlinear degradation trend in the instance retrieval step by according more importance to the late observations. The spherical distance was incorporated into the dynamic time warping (DTW) (Q. Cai et al., 2015; Sakoe & Hiba, 1978) technique to further allow pattern matching with timing difference through stretching or compressing the HIs along the time axis. It is worth noting that the spherical distance is for the first time defined for prognostics and tailored for the purpose of capturing the nonlinear damage propagation of mechanical systems in this study. Although the DTW technique has already used for measuring the similarity in existing work, this study proposes the first attempt to explore its capability of matching degradation patterns with timing difference and investigates whether this merit makes DTW a better distance than the traditional one-to-one point pattern matching distances. DTW is adopted to conduct a similarity measure between train trajectories and the test trajectory in the retrieval step. And the train trajectories are stretched or compressed along the time axis to allow many-to-one point matching to achieve a reasonable fit, making it a promising tool to replace the traditional methods. Then the spherical distance was incorporated into the DTW technique to form the new distance measure spherical-DTW. We tested the proposed method on both simulation and real-world data.

The remainder of this paper is organized as follows. The definitions of spherical distance, the proof of its nonlinearity and the definition of spherical-DTW is provided in Section 2. In Section 3, two case studies regarding turbine engine faults and reciprocating compressor failures are used for validation and comparison purposes. Finally, some conclusions are given in Section 4.

2. Methodology

2.1 spherical Distance

In this article, we propose to quantify the nonlinearity embedded in the degradation of an equipment using the concept of spherical distance which is based on our previous work (Yang & Chiclana, 2009). spherical distance is a nonlinear distance method that is used to measure the distance between two intuitionistic fuzzy sets. In this study, we extend the concept of spherical distance to model mechanical systems with nonlinear damage propagation rate.

In the following subsections, the definition of intuitionistic fuzzy sets and spherical distance are firstly provided. The nonlinearity of spherical distance is also demonstrated mathematically and geometrically afterwards.

2.1.1 Definition of spherical distance

To intuitively understand the concept of spherical distance, let's assume that the human behavior can be classified as worst, poor, acceptance, good and perfect. Using fuzzy sets, one can assign their corresponding fuzzy membership (this can be interpreted as the score achieved in an exam) as 0, 0.25, 0.5, 0.75 and 1. Obviously, the semantic distance between good and perfect is greater than the semantic distance between acceptance and good since more efforts are required to be made in the former case. In other words, although the fuzzy membership distance between good and perfect (0.25 in this case) equals the fuzzy membership distance between acceptance and good, the spherical distance between good and perfect should be accorded a greater value to correctly reflect the semantic distance. Given a degradation trajectory $X = \{x_{t_i}: t_i \in T\}$, with $T = \{t_1, t_2, \dots, t_n\}$ being the sampling time steps. Obviously, the sampling points $\{t_1, t_2, \dots, t_n\}$ represent the degree of closeness of the equipment to its failure point, with t_1 being the least close point to the end of life and t_n being the closest point to the failure. We can assign their corresponding fuzzy membership as $U = \{u_{t_1}, u_{t_2}, \dots, u_{t_n}\} = \{0, \frac{1}{n-1}, \dots, 1\}$, with value 0 corresponding to t_1 and value 1 corresponding to t_n . From the perspective of damage propagation, the semantic distance between t_n and t_{n-1} should be greater than the semantic distance between t_2 and t_1 since the

degradation rate within the interval $[t_{n-1}, t_n]$ is meant to be greater than that of the interval $[t_1, t_2]$. If a linear distance method is adopted to represent the semantic distance stated above, then the difference between t_n and t_{n-1} equals the semantic distance between t_2 and t_1 , namely, $u_{t_2} - u_{t_1} = u_{t_n} - u_{t_{n-1}}$. This is clearly not able to achieve our goals. Therefore, we propose spherical distance that is nonlinear with respect to the change of the membership degrees $U = \{u_{t_1}, u_{t_2}, \dots, u_{t_n}\}$ to quantify the nonlinearity embedded in the degradation of an equipment.

To better understand the concept of spherical distance, the definition of intuitionistic fuzzy sets is provided firstly:

DEFINITION 1. (Intuitionistic fuzzy set) *An intuitionistic fuzzy set A of the universe of discourse U is given by*

$$A = \{\langle u, \mu_A(u), \nu_A(u) \rangle | u \in U\} \quad (1)$$

where

$$\mu_A: U \rightarrow [0,1], \quad \nu_A: U \rightarrow [0,1],$$

and

$$0 \leq \mu_A(u) + \nu_A(u) \leq 1 \quad \forall u \in U.$$

where $\mu_A(u)$ and $\nu_A(u)$ denote the degree of membership and degree of non-membership of u to A , respectively.

Another parameter that is related to intuitionistic fuzzy sets is the hesitancy degree, $\tau_A(u) = 1 - \mu_A(u) - \nu_A(u)$, which denotes the hesitance to the membership of u to A . Based on the definition of intuitionistic fuzzy sets, we have

$$\mu_A(u) + \nu_A(u) + \tau_A(u) = 1 \quad (2)$$

which is equivalent to

$$x^2 + y^2 + z^2 = 1 \quad (3)$$

where

$$x^2 = \mu_A(u), \quad y^2 = \nu_A(u), \quad z^2 = \tau_A(u)$$

Obviously, this is not the only interpretation of the degree of membership, degree of non-membership and hesitance of u to A . However, there is no special reason to discriminate $\mu_A(u)$, $\nu_A(u)$ and $\tau_A(u)$. Hence a nonlinear transformation to a unit sphere, as illustrated in Fig. 1, is adopted in this work.

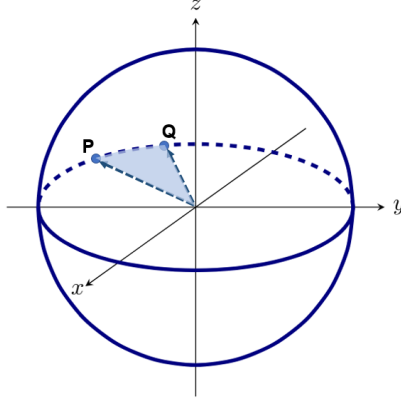


Figure 1. Illustration of the nonlinear transformation to unit sphere

Follow this transformation, the distance between two intuitionistic fuzzy sets can be defined as the spherical distance between their corresponding points on a unit spherical surface. This distance is represented by the arc of the greatest circle passing through both points. In fig. 1, assume points P and Q are the corresponding points of two intuitionistic fuzzy sets A and B on a unit sphere, the spherical distance between P and Q is given as follows

$$d_s(P, Q) = \arccos \left\{ 1 - \frac{1}{2} \left[(x_P - x_Q)^2 + (y_P - y_Q)^2 + (z_P - z_Q)^2 \right] \right\} \quad (4)$$

where the coordinate values x_P, y_P, z_P and x_Q, y_Q, z_Q are determined by the degree of membership, degree of non-membership and hesitance of the intuitionistic fuzzy sets.

DEFINITION 2. (spherical distance): Let $A = \{(u_i, \mu_A(u_i), \nu_A(u_i)) | u_i \in U\}$ and $B = \{(u_i, \mu_B(u_i), \nu_B(u_i)) | u_i \in U\}$ be two intuitionistic fuzzy sets of the universe of discourse $U = \{u_1, u_2, \dots, u_n\}$, their spherical distance between A and B is

$$d_s(A, B) = \frac{2}{\pi} \sum_{i=1}^n \arccos \left\{ 1 - \frac{1}{2} \left[\frac{(\sqrt{\mu_A(u_i)} - \sqrt{\mu_B(u_i)})^2 + (\sqrt{v_A(u_i)} - \sqrt{v_B(u_i)})^2}{(\sqrt{\tau_A(u_i)} - \sqrt{\tau_B(u_i)})^2} + \right] \right\} \quad (5)$$

In definition 2, the constant term $\frac{2}{\pi}$ is adopted to limit the distance values in the range $[0, 1]$, and n is the total number of memberships.

Since $\mu_A(u) + v_A(u) + \tau_A(u) = 1$ and $\mu_B(u) + v_B(u) + \tau_B(u) = 1$, the spherical distance can also be defined as

$$d_s(A, B) = \frac{2}{\pi} \sum_{i=1}^n \arccos \left\{ 1 - \frac{1}{2} \left[\sqrt{\mu_A(u_i)\mu_B(u_i)} + \sqrt{v_A(u_i)v_B(u_i)} + \sqrt{\tau_A(u_i)\tau_B(u_i)} \right] \right\} \quad (6)$$

Further, we have the normalized spherical distance:

$$d_{ns}(A, B) = \frac{2}{n\pi} \sum_{i=1}^n \arccos \left\{ 1 - \frac{1}{2} \left[\sqrt{\mu_A(u_i)\mu_B(u_i)} + \sqrt{v_A(u_i)v_B(u_i)} + \sqrt{\tau_A(u_i)\tau_B(u_i)} \right] \right\} \quad (7)$$

where $0 \leq d_{ns}(A, B) \leq 1$.

2.1.2 Nonlinear characteristic of spherical distance and its visualization

Spherical distance is a nonlinear distance measure, as for its nonlinear characteristic, the following properties hold:

LEMMA 1. Let $A = \{(u_i, \mu_A(u_i), v_A(u_i)) | u_i \in U\}$, $B = \{(u_i, \mu_B(u_i), v_B(u_i)) | u_i \in U\}$, $C = \{(u_i, \mu_C(u_i), v_C(u_i)) | u_i \in U\}$ be three intuitionistic fuzzy sets of the universe of discourse $U = \{u_1, u_2, \dots, u_n\}$. Let $D = \{d_1, d_2, \dots, d_n\}$ and $E = \{e_1, e_2, \dots, e_n\}$ be two sets of constants (real positive numbers) with the following conditions being satisfied:

$$|\mu_B(u_i) - \mu_A(u_i)| = d_i, |v_B(u_i) - v_A(u_i)| = e_i \quad (8)$$

$$|\mu_C(u_i) - \mu_A(u_i)| = d_i, |v_C(u_i) - v_A(u_i)| = e_i \quad (9)$$

If C is an extreme crisp set, namely $\mu_C(u_i) \vee v_C(u_i) = 1$ and $\mu_C(u_i) \wedge v_C(u_i) = 0$ for all $u_i \in U$, then the following properties hold:

$$d_{ns}(A, B) < d_{ns}(A, C) \quad (10)$$

Proof.

Let C being the extreme crisp set with full non-memberships, then we have

$$\mu_C(u_i) = 1, v_C(u_i) = 0, \tau_C(u_i) = 0 \quad (11)$$

Since $|\mu_C(u_i) - \mu_A(u_i)| = d_i, |v_C(u_i) - v_A(u_i)| = e_i$, then

$$\mu_A(u_i) = 1 - d_i, v_A(u_i) = e_i, \tau_A(u_i) = d_i - e_i \quad (12)$$

From $|\mu_B(u_i) - \mu_A(u_i)| = d_i, |v_B(u_i) - v_A(u_i)| = e_i$, we have

$$\mu_B(u_i) = 1 - 2d_i, v_B(u_i) = 2e_i, \tau_B(u_i) = 2(d_i - e_i) \quad (13)$$

Then

$$\begin{aligned} d_{ns}(A, B) &= \frac{2}{\pi} \sum_{i=1}^n \arccos \left\{ 1 - \frac{1}{2} \left[\sqrt{2e_i^2 + \sqrt{(1-d_i)(1-2d_i)}} + \sqrt{2(d_i - e_i)^2} \right] \right\} \\ &= \frac{2}{\pi} \sum_{i=1}^n \arccos \left\{ \sqrt{2d_i} + \sqrt{(1-d_i)(1-2d_i)} \right\} \end{aligned} \quad (14)$$

$$d_{ns}(A, C) = \frac{2}{\pi} \sum_{i=1}^n \arccos \left\{ \sqrt{1-d_i} \right\} \quad (15)$$

Apparently, the following conditions are satisfied:

$$\sqrt{2d_i} + \sqrt{(1-d_i)(1-2d_i)} > \sqrt{1-d_i} \quad (16)$$

Therefore

$$d_{ns}(A, C) > d_{ns}(A, B) \quad (17)$$

The proof of C being the extreme crisp set with full non-memberships is provided, and the case in which C being the extreme crisp set with full memberships can be proved in a similar way.

Lemma 1 shows that with the same difference in memberships and non-memberships, the spherical distance between an extreme crisp set and an intuitionistic fuzzy set is always greater than the distance between two other intuitionistic fuzzy sets. This coincides with the nonlinearity of the damage propagation rate in that the degradation rate near the failure point is assumed to be larger than that at an earlier time.

Based on the nonlinear characteristic of the spherical distance as stated in LEMMA 1, if we regard the sampling steps $T = \{t_1, t_2, \dots, t_n\}$ of a degradation trajectory $X = \{x_{t_i} : t_i \in T\}$ as intuitionistic fuzzy sets and model the overall degradation rate over interval $[t_A, t_B]$ with its corresponding spherical distance $d_{ns}(t_A, t_B)$, then LEMMA 1 indicates that with the same partition/interval length, the deterioration rate over a close-to-failure time interval is always

greater than that of an earlier time interval. This conclusion is consistent with the nonlinearity of damage propagation of mechanical systems that we mentioned previously.

To further provide a geometric and intuitive understanding of the proposed spherical distance, we demonstrate how the spherical distance varies with respect to $\mu_B(u_i)$ with fixed values of $\mu_A(u_i)$ in Fig. 2a. To be specific, the changes of spherical distance with respect to $\mu_B(u_i)$ when $\mu_A(u_i)$ is fixed to 0.3, 0.5 and 0.7 are illustrated in Fig. 2a. Clearly, the spherical distance increases abruptly when the membership $\mu_B(u_i)$ approaches extreme crisp values 0 and 1, while the changes are relatively smooth yet still nonlinear when intermediate membership values are taken. Due to this nonlinear characteristic, the proper selection of membership values $U = \{u_{t_1}, u_{t_2}, \dots, u_{t_n}\}$ for sampling points $\{t_1, t_2, \dots, t_n\}$ plays a vital role in modelling damage propagation. Fig. 2b shows how the spherical distance between two consecutive sampling times (i.e. $d_{ns}(u_{t_1}, u_{t_2}), d_{ns}(u_{t_2}, u_{t_3}), \dots, d_{ns}(u_{t_{n-1}}, u_{t_n})$) changes with respect to different selections of u_{t_1} and u_{t_n} with n being fixed to 30. To be specific, u_{t_1} and u_{t_n} are set to 0.5 and 1, 0.6 and 1, 0.8 and 1, respectively. It is evident that the spherical distance varies with the different initiations of u_{t_1} and u_{t_n} , but the trend of spherical distance remains similar in that d_{ns} changes sharply near the failure point and slightly for early fault showing an overall growing trend. This observation agrees with our human perception about the quality change against quantity in that more efforts need to be made for a student to achieve a higher score. Similarly, this conclusion agrees with our knowledge of mechanical system deterioration in that the damage propagation rate shows an overall growing trend, rendering more emphasis placed on late data samples.

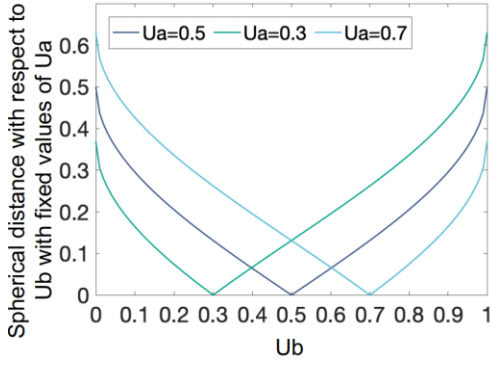
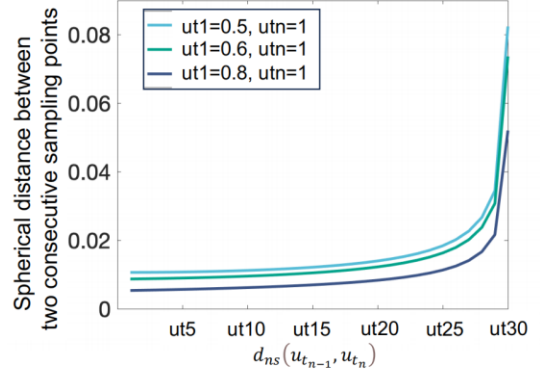
(a) How spherical distance changes when U_a is fixed(b) How spherical distance changes with different selections of ut_1 and ut_n

Figure 2. Illustration of the nonlinear characteristic of spherical distance

2.2 Dynamic Time Warping

Dynamic time warping was firstly proposed in (Sakoe & Hiba, 1978) for spoken word recognition. It uses a dynamic programming approach to align a continuous speech time series and a specific word template so that some optimization criteria can be reached. Given two time series, $X = \{x_{t_1}, x_{t_2}, \dots, x_{t_n}\}$ and $Y = \{y_{t_1}, y_{t_2}, \dots, y_{t_m}\}$, they can be arranged to form a $n \times m$ grid, in which each grid point $(x_{t_{\pi_x(k)}}, y_{t_{\pi_y(k)}})$ corresponds to an alignment between elements $x_{t_{\pi_x(k)}}$ and $y_{t_{\pi_y(k)}}$, where π denotes a complete warping path. A warping path π is a sequence of indices that define an alignment. In Fig. 3 demonstrates a possible warping path between time series X and Y . In the figure, for instance $\pi(3)$ corresponds to the alignment between x_{t_3} and y_{t_2} (here, $k = 3$, $\pi_x(3) = 3$ and $\pi_y(3) = 2$).

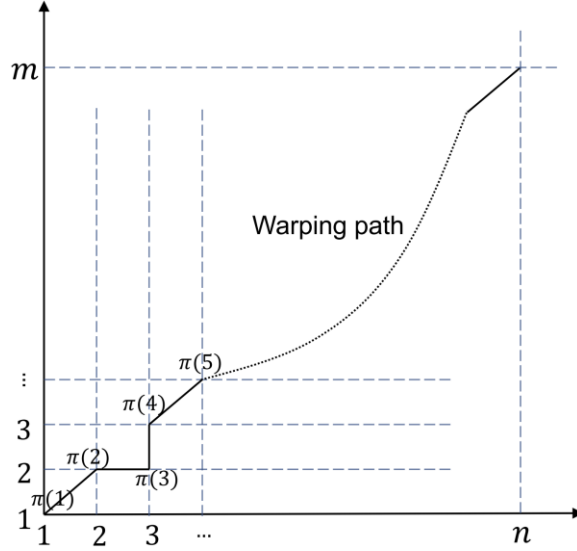


Figure 3. A warping path example

In order to measure the difference between two time series X and Y , a distance measure

$$d_E(x, y) = \left\| x_{t_{\pi_x(k)}} - y_{t_{\pi_y(k)}} \right\| \quad (18)$$

is employed, where $\|\cdot\|$ denotes the 2-norm distance. Then the dynamic time warping problem can be formulated as a minimization over potential warping paths based on the cumulative distance for each elemental warping path step w_k as

$$d_E(X, Y) = \min_{X, Y} \sum_{k=1}^h \left\| x_{t_{\pi_x(k)}} - y_{t_{\pi_y(k)}} \right\| \quad (19)$$

According to (Sakoe & Hiba, 1978), the warping paths must satisfy the following restrictions:

- 1) Monotonic conditions

$$\pi_x(k) \leq \pi_x(k+1) \text{ and } \pi_y(k) \leq \pi_y(k+1)$$

- 2) Continuity conditions

$$\pi_x(k+1) - \pi_x(k) \leq 1 \text{ and } \pi_y(k+1) - \pi_y(k) \leq 1$$

- 3) Boundary conditions

$$\pi_x(1) = 1 \text{ and } \pi_y(1) = 1, \text{ and } \pi_x(h) = n \text{ and } \pi_y(h) = m$$

- 4) Adjustment window conditions

$$|\pi_x(k) - \pi_y(k)| \leq l$$

where l is called window length and takes an appropriate positive integer. It is used to prevent the case in which too excessive timing difference is induced when matching an element in X onto an element in Y . The optimization problem defined above can be solved recursively via the following equation

$$d_E(i, j) = \|x_{t_i} - y_{t_j}\| + \min \begin{cases} d_E(i-1, j) \\ d_E(i-1, j-1) \\ d_E(i, j-1) \end{cases} \quad (20)$$

where $i = 1, 2, \dots, n$ and $j = \max(1, i-l), \dots, \min(m, i+l)$. This formulation calculates the cumulative distance which is the sum of the distance between the current point $\|x_{t_i} - y_{t_j}\|$ and the minimum of the cumulative of distances of the neighbouring elements. When the recursion is finished, the solution to Eq. (19) can be found in $d_E(n, m)$, that is, the minimum Euclidean distance between the time-warped versions of sequences X and Y .

Once the optimal warping path is found, the corresponding cumulative distance $d_E(X, Y)$ can be utilized as a score to describe the fitness between the testing trajectory and the training trajectory. Different from the traditional pointwise distance measures, DTW quantifies the degree of fitness achievable by stretching or compressing the time series X and Y in regard to time, allowing degradation patterns to be matched in spite of time difference. An illustration is provided in Fig. 4 where a point-wise distance measure between two time series is shown on the left and the DTW distance is shown on the right. For the point-wise distance, each vertical dashed line corresponds to a time stamp, and only the alignments between points captured at the same time stamp are allowed. In other words, there was no stretching or compressing of the time series with respect to time. On the contrary, DTW distance allows multiple-to-one point alignment as well as alignments with timing difference, making it easier for similar degradation patterns to be matched.

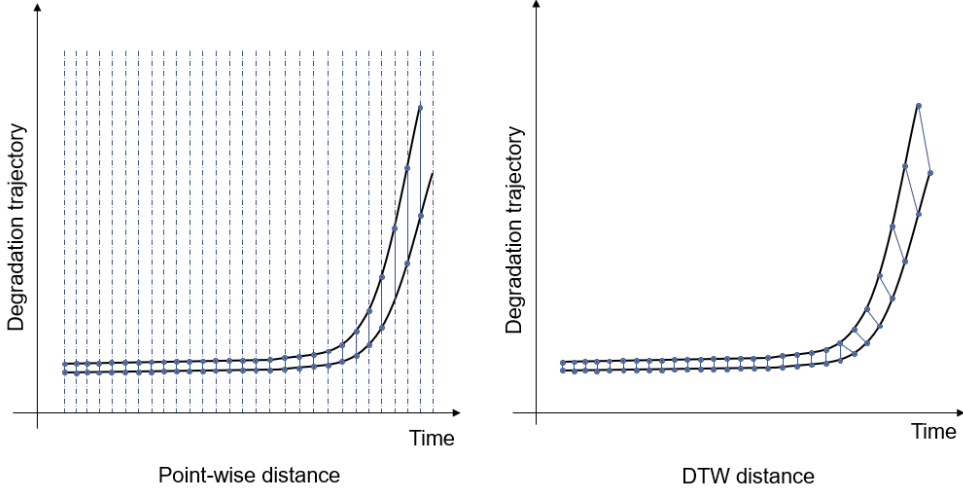


Figure 4. Illustration of the differences between DTW and point-wise distance measures

2.3 Spherical-Dynamic Time Warping Based Prognostics

In this paper, we put forward a new distance measure called spherical-Dynamic time warping (spherical-DTW) for similarity-based prognostics which jointly considers the nonlinearity of damage propagation and the trajectory alignment with timing difference. Given two time series, $X = \{x_{t_1}, x_{t_2}, \dots, x_{t_n}\}$ and $Y = \{y_{t_1}, y_{t_2}, \dots, y_{t_m}\}$, their spherical-DTW distance is formulated as

$$d(X, Y) = \sum_{k=1}^h \left\| x_{t_{\pi_x(k)}} - y_{t_{\pi_y(k)}} \right\| \cdot d_{ns}(u_{k-1}, u_k) \quad (21)$$

where $\pi(k)$ denotes the k th element of the optimal warping path $\pi = \{\pi(1), \pi(2), \dots, \pi(h)\}$, and there are in total h number of elements in W . Obviously, $\sum_{k=1}^h \|\pi(k)\|$ itself represents the accumulative DTW distance between time series X and Y over all data alignments. Restricted by the monotonicity conditions of the DTW distance, the points $\pi(k)$ on the warping path must be monotonically ordered with respect to time. Therefore, it is reasonable to assign a series of membership values $U = \{u_0, u_1, \dots, u_h\}$ to the alignments $\{\pi(1), \pi(2), \dots, \pi(h)\}$, where u_0 and u_h are defined as the lower and upper bounds of the membership sequence U , respectively. That is, we treat each element $\pi(k)$ of the warping

path π as an intuitionistic fuzzy set which contains only one membership u_k , and utilize the spherical distance $d_{ns}(u_{k-1}, u_k)$ to represent the degradation rate associated with $\pi(k)$. With proper selections of the lower and upper bounds u_0 and u_h , the nonlinear characteristics associated with the spherical distances $\{d_{ns}(u_0, u_1), d_{ns}(u_1, u_2), \dots, d_{ns}(u_{h-1}, u_h)\}$ (as illustrated in Fig. 2(b)) can well capture the underline system behaviour associated with the deterioration process. In other words, the spherical distance $d_{ns}(u_{k-1}, u_k)$ is actually introduced here as a weighting coefficient to allow the spherical-DTW distance $d(X, Y)$ measure the nonlinear characteristics of the degradation process by according more importance to the late observations where the failure is more likely to occur.

The proposed framework for prognostics is illustrated in Fig. 5. The run-to-failure health indicators (HIs) are firstly obtained from the multi-sensor data using the Principal Component Analysis (PCA) method. The basic idea behind the subsequent similarity-based prognostic method is to find the top few training HI segments which possess similar degradation trend that the test HI has and use their failure times to predict the RUL of the test HI. Suppose a set of HIs showing patterns of run-to-failure evolution are contained in a library $\Sigma = \{X_1, X_2, \dots, X_m, \dots, X_M\}$. Meaning that there are M number of run-to-failure HIs in the training library Σ , and each single HI is denoted as $X_m = \{x_{t_1}, x_{t_2}, \dots, x_{t_N}\}$. For a test HI segment X' with $X' = \{x'_1, x'_2, \dots, x'_{t'_n}, \dots, x'_{t'_N}\}$, assume that $x'_{t'_n}$ represents the value of the HI at the current operating time t'_n , and the length of the up-to-date test HI segment X' is t'_N . Then each one of the training HIs in library Σ , namely X_m , can be truncated into $t_N - t'_N$ pieces of HI segments with the length of each segment being t'_N and the start time of the segment taking the numbers $t_n = [1, 2, \dots, t_N - t'_N + 1]$, which is denoted as $X'_m = \{x'_{t_n}, x'_{t_n+1}, \dots, x'_{t_n+t'_N}\}$. To have a comprehensive evaluation of the prognostic performance of the proposed method and to demonstrate its usability in practice, in this study we tested on each failure case its prediction accuracy for various prediction start time rather than a randomly chosen prediction start time. When performing similarity matching, if the prediction start time is close to the actual failure time of the test HI, then the number of data points included in the test HI $X' = \{x'_{t_1}, x'_{t_2}, \dots, x'_{t'_N}\}$ is large. The adoption of sliding

windows helps to filter out those referential HIs whose failure times are too early, avoiding an under-estimated RUL of the test case.

To measure the similarity between the truncated segments X'_m and the up-to-date test HI X' , the spherical-DTW method is considered:

$$d(X', X'_m) \quad (22)$$

Based on the calculated spherical-DTW values, the most similar HI segment \hat{X}'_m , among the total $t_N - t'_N$ number of HI segments resulted from the complete HI X_m , can be easily determined. The RUL of the test instance with respect to X_m is computed as:

$$RUL_m = RUL(\hat{X}'_m) \quad (23)$$

where $RUL(\hat{X}'_m)$ represents the remaining time until \hat{X}'_m reaches the predefined failure threshold. Similarly, the most similar truncated HI segment \hat{X}'_m and its corresponding remaining life RUL_m of other training instances can also be determined. When computing the RUL of the test failure case, the RUL of the HI segments can be directly added up. But in this study, the RUL was calculated by the weighted result of the referential HI segments. The RUL of the test instance is finally estimated as a weighted sum of the top few training HIs as:

$$RUL = \sum_{m=1}^T w_m RUL_m \quad (24)$$

where T is the number of most similar HIs considered for calculated the RUL, and it is set to 3 in this study. The weighting coefficient w_m was calculated according to the spherical-DTW value of X'_m as:

$$w_m = \frac{1}{\sum_{m=1}^T \frac{1}{d(X', X'_m)}} \quad (25)$$

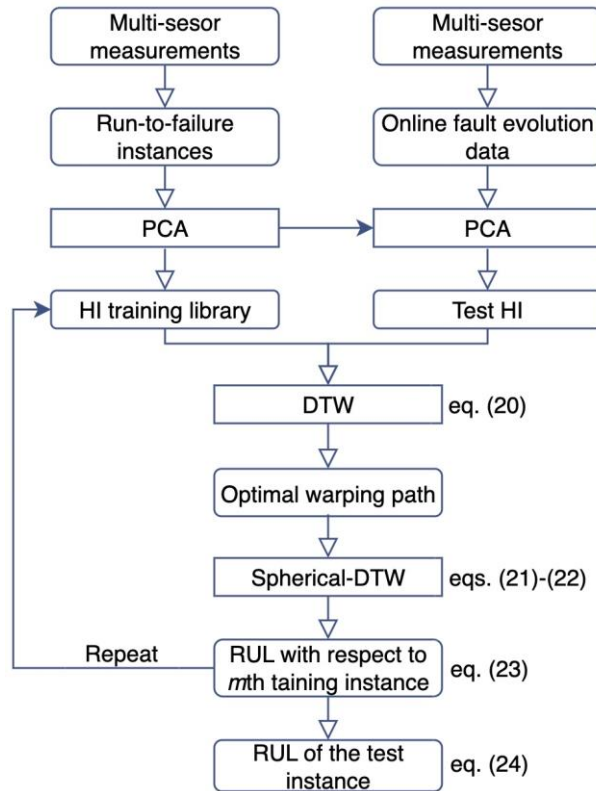


Figure 5. Flowchart of the proposed spherical-DTW based prognostics framework.

3. Case Studies

3.1 Case study 1 - Industrial Case Study

The first case study involves data captured online from a running two-stage, four-cylinder, double-acting reciprocating compressor. Within a period of one and a half year, the machine experienced 13 failures that all took place in cylinder number four. According to the maintenance records, all 13 failures were caused by improper sealing induced leakage that was originated from a missing piece from the outer structure of the valve plate due to fracture. To summarize, the failure mode was valve leakage due to broken valve plate. The valves were supplied by the same manufacturer with the same model and specifications, and the failed valves were discharge valves either on the head end (HE) or the crank end (CE). It is worth noting that the compressor carried out various tasks and operated under varying operating conditions.

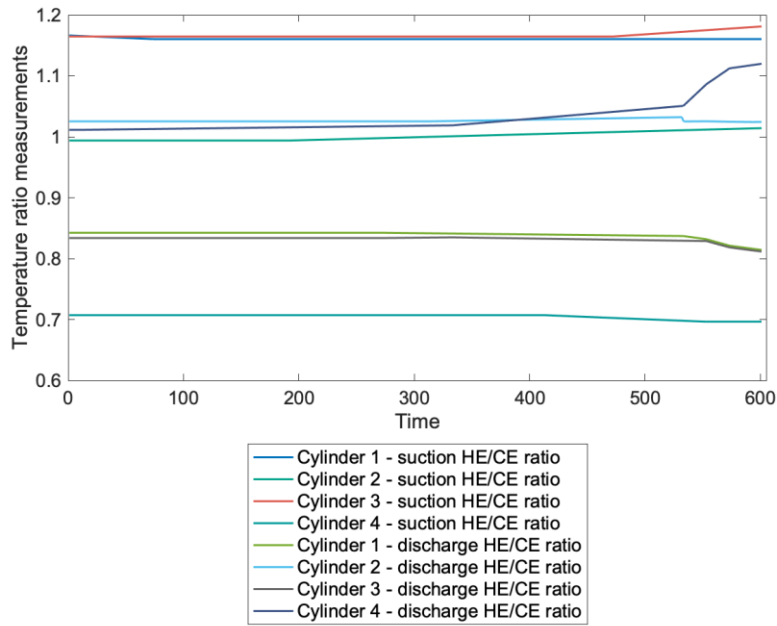


Figure 6. Temperature ratios captured from the compressor for failure case No.4 (unit of horizontal axis: seconds)

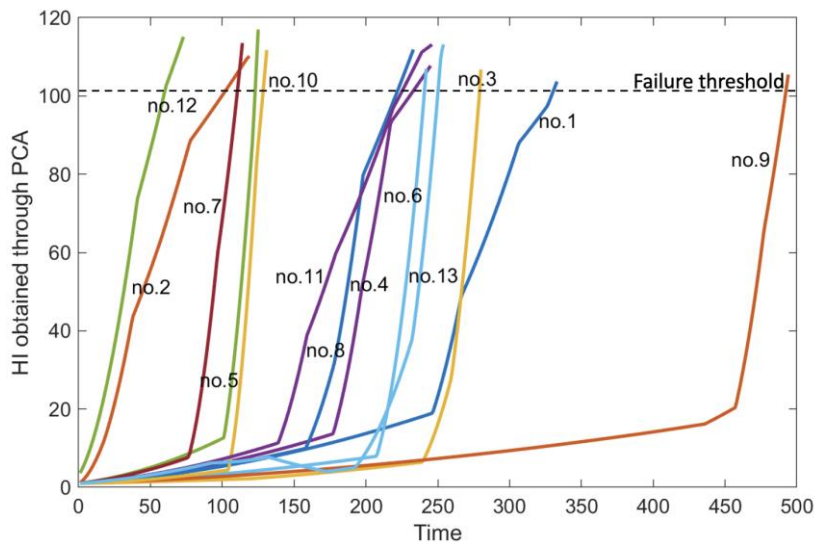


Figure 7. Run-to-failure HIs for the 13 valve failure cases (unit of horizontal axis: seconds)

Only temperature sensors were placed on the compressor, but the sensor type and topology cannot be provided due to confidentiality reasons. The raw data was pre-processed by the data acquisition system and the temperature ratios T_{HE}/T_{CE} were utilized as the system

output. Since there were eight valves on the machine, the total number of outputs were eight: one temperature ratio per valve, four valves per cylinder. Fig. 6 illustrates the resultant T_{HE}/T_{CE} temperature ratios right before the failure for case No.4. Obviously, the ratios of the temperatures start to rise a few minutes prior to the malfunction. In the prognostics process, the 8 temperature ratios were treated as the multidimensional input to build a PCA model prior to fault detection. The healthy datasets used to build the PCA model contains information from various operating conditions throughout the year and a half period, enabling the constructed HIs to well represent the health status of the system. The trained PCA model is employed to construct online HIs when the run-to-failure data becomes available. As mentioned previously, thirteen historical pieces of data were available, that contained healthy, degradation and failure status. Fig. 7 demonstrates the constructed HIs for all 13 failure cases. The HI values exhibit an increasing trend as time progresses, with the late stages data showing a faster growing trend than the early fault. In addition, the failure times vary in duration, indicating the stochastic nature of fault evolution. This phenomenon may also be attributed to the various operating conditions and tasks that the machine have involved in during which the data was recorded, where various gas compositions may influence differently the rate a crack propagates on valve plate. Having similar failure thresholds for different HIs is appealing for use in prognostics. That is why PCA was adopted in this study. It is worth noting that the HIs were extracted using the Hotelling's T2 index calculated based on PCA, because it demonstrates the highly desirable characteristic of having very similar failure values (end-of-life) at all cases (see Fig. 7), having their end-of-life values distributed around 103. This simplifies a great deal the failure threshold selection at a value around 103 for all failure cases. In comparison, the Q index which can be generated from the remaining principal components left by the Hotelling's T2 index will have its end-of-life values vary significantly and is therefore not preferable for prognostics. Interested readers are referred to our previous publication (Loutas et al., 2019).

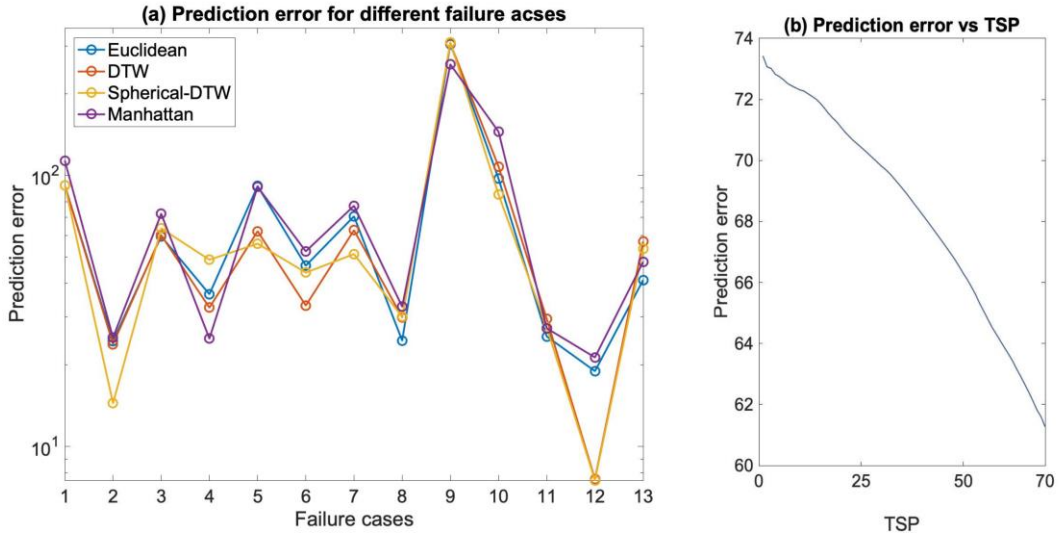


Figure 8. Overall RUL prediction performance of spherical-DTW verified on the reciprocating compressor

Fig.8 demonstrates the overall RUL estimation performance of the proposed spherical-DTW framework for the reciprocating compressor failures. To be specific, Fig. 8(a) illustrates the averaged prediction error for all 13 failure cases for time-to-start-prediction (TSP) ranging from 1 to 70 seconds. Mean absolute error (MAE) was employed to quantify the estimation errors in this study since the focus is how far the predicted RUL deviate from the true value. The prediction errors were averaged with respect to different TSPs in order to show the generalization capacity of the proposed method, and the latest TSP was set to 70 because the length of the shortest failure instance was 70.

In terms of how the prediction errors were calculated, the following descriptions are provided. Failure HIs as demonstrated in Fig.7 were served as the composite components of the HI training library. Following the steps as illustrated in Fig. 5, the HI of Case 1 was firstly marked as unknown and the remaining 12 cases were treated as training cases. We firstly set the TSP to 1, which means just the first point on the HI of case 1 will be used to search for the best matching HIs. We calculate the spherical-DTW distance between the first point on the HI of case 1 with each candidate segment on the remaining 12 cases, and then find the

top three similar HIs with the least spherical-DTW values. These HIs will be used to compute the HI of case 1 according to eqs. 23-24. Moreover, we gradually increase the value of TSP from 1 to 70 to imitate the situation that in an online monitoring case, the known sensory measurements are continuously accumulating. We follow the same steps as we did when the TSP was set to 1 to compute the resulting RUL for TSP ranging from 1 to 70 and the final prediction errors as shown in Fig 8 (a) were averaged with respect to these different TSPs. We repeat the above procedures for cases 2 to 13 to yield the corresponding prediction errors.

It can be seen from Fig. 8(a) that, except for failure No.9, the predictions were in general close to the actual RULs. The reason for this excessive average prediction error is that case 9 is the right outlier failure case (see Fig. 7), and the similar cases found by the algorithm are rather far away from the trajectory 9, leading to underestimated RULs. The limited number of run-to-failure trajectories can also lead to this large prediction error, but the situation can be improved when more failure data is captured. Besides, the excessive average prediction errors happened not only for spherical-DTW but all prognostic methods compared. This coincides with the findings in our previous study (Loutas et al., 2019). Fig. 8(a) also shows that the proposed spherical-DTW method can approach the actual RUL, at least in most cases. For cases 2, 5, 7, 10 and 12, spherical-DTW achieved the lowest prediction errors. For cases 1, 3 and 11, yet not the best performing method, spherical-DTW was able to provide a similar prediction performance and deviated little from the best performing method. Surprisingly, the proposed method achieved the lowest prediction error for the left outlier failure case (i.e. case 12), demonstrating its ability to overcome the diversified trajectories. Figure 8 only shows the prediction results of spherical-DTW and a few selected distance measures. To demonstrate a thorough and clear comparison of the proposed method with existing frequently used distances, more numerical comparison results are presented in Table I.

Fig. 8(b) demonstrates the trend of the prediction accuracy with respect to the value of the TSP, from which it can be seen that a larger TSP results in a higher prediction accuracy. This is because the similarity prognostic method learns more about the degradation trajectory

when the TSP moves towards the actual time of failure, and this is actually a highly desirable attribute of a prognostic algorithm.

To show the superiority of the proposed distance for similarity matching, the comparisons between the proposed methodology and some classical distances for similarity-based prognostics are discussed as follows. Firstly, some famous distance measures including DTW, Euclidean, Manhattan, Frechet, locality in-between polylines (LIP), edit distance on real sequences (EDR) and Hausdorff distance were employed to predict the RUL of the machine. To quantitatively evaluate their prognostic abilities, the mean absolute error (MAE) and root mean square error (RMSE) values were utilized as performance metrics. The comparison results of different distance methods are summarized in Table I. Spherical-DTW achieved the lowest prediction error with respect to both MAE and RMSE in this case. Obviously, DTW method leverages its advantages to allow similarity matching with timing differences, achieving a much smaller prediction error than other distance measures like Euclidean, EDR and Manhattan methods. The proposed spherical-DTW distance further turns the DTW method into a nonlinear distance measure, which utilizes the spherical distance to model the nonlinear fault propagations of the system under study. As a result, the spherical-DTW managed to achieve the lowest prediction error when compared with its linear counterparts. It can be concluded that the proposed distance is more effective in terms of finding the similar historical degradation trajectories for RUL prediction.

Table I. RUL prediction performance of spherical-DTW and its counterparts

	Spherical-DTW	DTW	Euclidean	Manhattan	Frechet	LIP	EDR	Hausdorff
MAE	68.36	69.68	71.89	76.01	79.69	78.08	118.44	79.70
RMSE	68.88	70.09	72.15	76.23	88.05	85.16	124.32	88.06

To give a detailed view of the overall prediction performance, the probability density functions of the predicted RULs against various TSPs (ranging from 1 to the failure time) are estimated using a kernel density estimation method and are demonstrated in Fig.9. Due to page limit, only 12 out of 13 failure cases were plotted. The vertical dashed line in each subplot demonstrates the true failure time of the corresponding failure case. For instance, the

true failure time in the first subplot is around 350. Taking the predicted RULs for TSPs ranging from 1 to the machine's true failure time as inputs, each probability density was estimated through the Kernel density estimation (Botev et al., 2010) method. Each density curve therefore reflects the overall RUL predictions against the failure case's true failure time for different TSP values, representing the general prediction accuracy of the evaluated method. spherical-DTW and its counterparts were compared in Fig. 9. The black arrow marks on the density curve in cases 1, 2, 11, 12 and 13 highlight the superiority of spherical-DTW over other prognostic methods. For instance, spherical-DTW resulted in more accurate predictions around the true failure time in case 1 when compared with its counterparts. Besides, the density curve of spherical-DTW intersects with the black dashed line (true failure time) at a higher value in cases 2, 11 and 12, indicating that more density samples lie close to the true failure time. In comparison, other methods seemed to have most of the density samples lie to the left of the true failure time, resulting in under-estimated RUL predictions. In case 13, spherical-DTW was able to have its density curve peaks at the true failure time meaning that it had the highest rate of accurate predictions. In comparison, other methods had most of their density samples deviate from the true failure time and the number of accurate predictions is small. Furthermore, spherical-DTW presented similar prediction densities when compared with other distance measures in cases 3, 4, 6, 7 and 10, which means that all methods demonstrated similar failure time prediction ability when the TSP moves from the early failure stage to imminent failure.

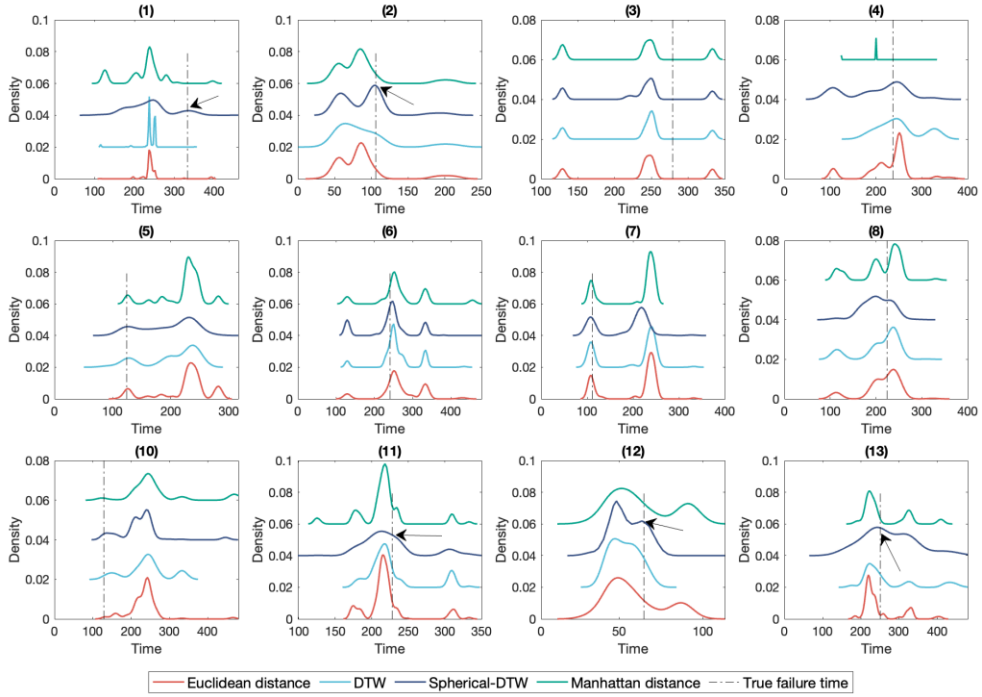


Figure 9. Kernel density of the predicted RUL for various TSPs

Coincides with the results shown in Fig. 8(a) and (b), the spherical-DTW method were able to achieve an overall better prediction performance. As mentioned previously, with a proper selection of the lower and upper bounds, u_0 and u_h , of the membership values $U = \{u_0, u_1, \dots, u_h\}$, can the nonlinear characteristics of the spherical-DTW well represent the damage propagations. Fig.10 illustrates the training errors with respect to different selections of lower bound u_0 and upper bound u_h , with $u_0 = 0.6$ and $u_h = 0.7$ yielding the lowest training error. Hence $u_0 = 0.6$ and $u_h = 0.7$ was selected in this study. The training process were carried out using the first 30 time stamps as the training data, namely, TSP ranges from 1 to 30.

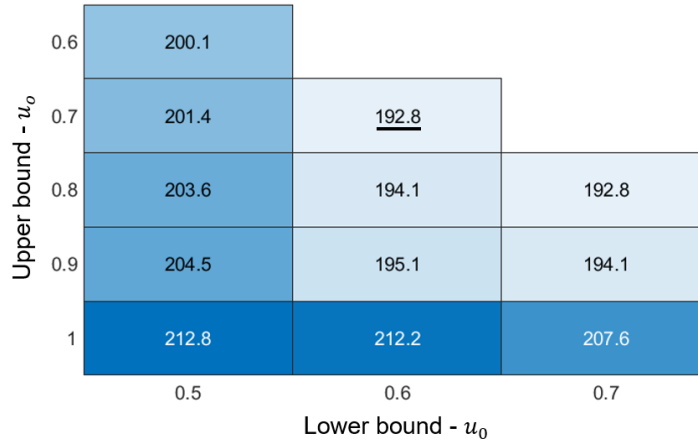


Figure 10. Training error with respect to different selections of lower bound u_0 and upper bound u_h

The hyperparameters in the proposed model mainly involve the fuzzy membership values. As shown in fig. 10, the determination of fuzzy membership values affects the prediction error in several ways. To be specific, for a fixed value of u_0 , the larger the upper bound u_h is, the larger the prediction error is. Similarly, for a fixed value of u_h , the smaller the lower bound u_0 is, the larger the prediction error is. This means the difference between the fuzzy membership values u_0 and u_h need to be small (e.g. 0.1) in order to have smooth rather than rapid changing spherical distance values to ensure a good approximation to the nonlinear characteristics of the system's degradation. Besides, when the difference between u_0 and u_h is fixed, the optimal value of u_0 that leads to the smallest prediction error can be found via a cross validation process like we did in fig. 10. When the fuzzy membership interval $[u_0, u_h]$ is close to 1, the corresponding spherical distance values involve sharp changes, but when the interval $[u_0, u_h]$ is far from 1, the spherical distance values $\{d_{ns}(u_0, u_1), d_{ns}(u_1, u_2), \dots, d_{ns}(u_{h-1}, u_h)\}$ involve smoother transitions. In this study, the optimal parameters are $u_0 = 0.6$ and $u_h = 0.7$, meaning that a smoother transitions between spherical distance values yields better approximations to the nonlinear deterioration characteristics. We observed that the errors with the lower and upper bound set (0.7,0.8) has

the same training error. We finally set the bound set to $(0.6, 0.7)$ in order to have spherical distances that involve smoother changes.

Taking failure cases 2 and 7 as examples, fig. 11 shows the detailed prediction results of the proposed method at different starting points. In both cases, the prediction error demonstrates a general descending trend as the starting point approaches the actual failure time. This is easily understandable because when the TSP is moving towards the failure time, more run-to-failure data becomes available hence more degradation information can be employed to perform the similarity matching.

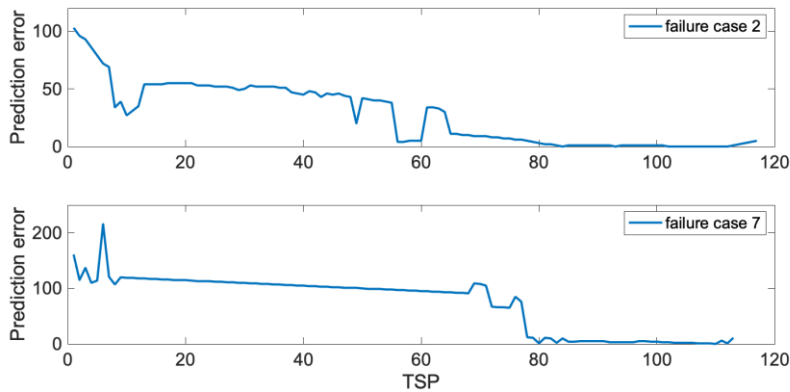


Figure 11. Estimation error with respect to different prediction start points

3.2 Case study 2 – C-MAPSS engine failure dataset

To demonstrate the generalization capacity of the proposed spherical-DTW method for similarity-based prognostics, its application to data generated from the Commercial Modular Aero Propulsion System Simulation (C-MAPSS) program (Ramasso & Saxena, 2014) are presented in this section. The data are simulated for a fleet of 100 engines of the same type, containing one hundred run-to-failure cases. The engines were operating normally at the initial condition and then developed a fault at some point and then the fault increased in magnitude until system failure occurred. Similar to (X. Li et al., 2022), 9 complementary variables among the total 26 variables found in the dataset (corresponding to variables 2, 3, 4, 7, 11, 12, 15, 20 and 21) were selected to construct the HIs through the PCA method. For

details regarding the HI construction process, readers are referred to (Loutas et al., 2019). All 100 failure cases were utilized to verify the proposed method.

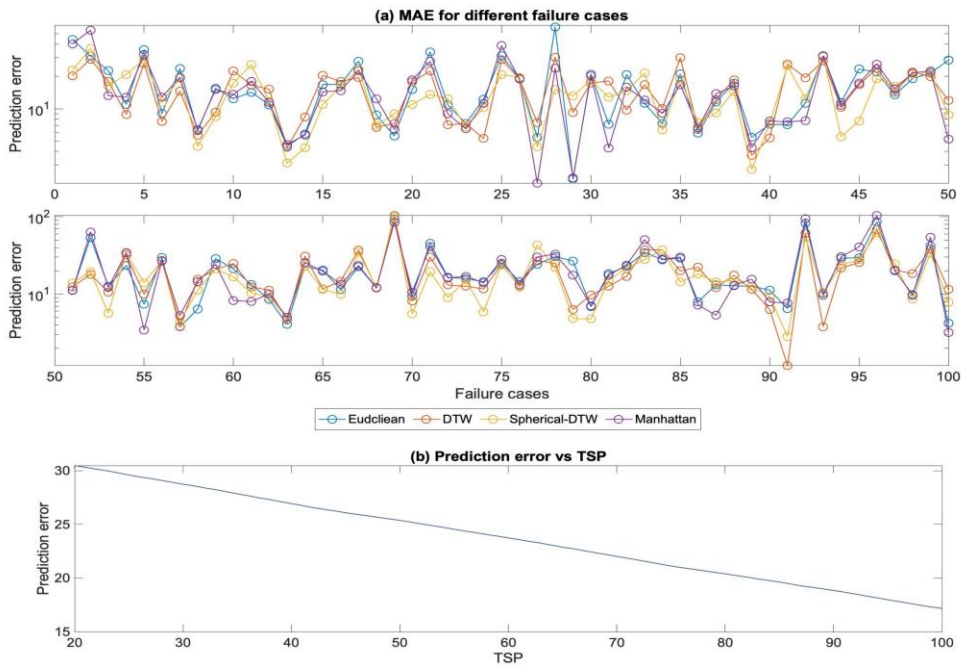


Figure 12. Overall RUL prediction performance of spherical-DTW verified on the CMAPSS data

The overall RUL prediction performance of the proposed spherical-DTW method is illustrated in Fig.12. The upper two subplots show the averaged prediction error for all 100 failure cases for TSP ranging from 1 to 100 minutes. Overall, the prediction errors are smaller than that of the industrial case due to the fact that the simulated failure data contain less stochastic degradations, and the total number of failure cases is apparently larger than that of the valve failure dataset. The spherical-DTW method was able to achieve the lowest averaged prediction error for a total of 40 out of 100 cases as shown in fig. 12 (a). In comparison, the number of the lowest-prediction-error cases for DTW, Euclidean and Manhattan distance were 28, 15 and 17 out of 100 cases, respectively. These results indicate that the spherical-DTW distance was able to achieve the overall highest prediction accuracy thanks to its

approximation to the nonlinear evolution of the system and DTW's similarity matching ability. Besides, the proposed method was able to provide a relatively accurate RUL prediction for the left outlier failure case (i.e. case 39). It is because there is a large amount of failure trajectories near the HI of the left outlier case. The lower bound u_0 and upper bound u_o of membership U were the same as the industrial case, namely $u_0 = 0.6$ and $u_o = 0.7$. The subplot on the bottom shows the variations of prediction error with respect to TSP. Clearly, larger TSP means more data to be included in the similarity matching, and hence a higher prediction accuracy can be obtained.

The comparisons between the proposed method and some typical distances including DTW, Euclidean, Manhattan, Frechet, LIP, EDR and Hausdorff distance were carried out in terms of both MAE and RMSE values. The comparison results of different distance methods are summarized in Table II. It can be noted that spherical-DTW obtained the lowest MAE prediction error, showing obvious competitiveness compared with other typical similarity distances for the RUL estimations of C-MAPSS turbofan datasets. DTW shows a better performance than most of the distances compared, highlighting the importance of degradation pattern matching with timing difference. The inclusion of spherical distance further enhanced the capability of the DTW to capture the nonlinear growing trend of the fault evolution in the retrieval step, thereby allowing the lowest MAE. Although the RMSE value of spherical-DTW is not always the lowest (e.g. the C-MAPSS dataset), the proposed method still achieved a lower RMSE level than a majority of the popular used distance measures. More importantly, the proposed distance shows good overall prediction capability and generalization capability for all the datasets considered in this study. The proposed method was able to achieve an overall better prediction result mainly because it leverages the ability of the spherical distance to imitate the nonlinearity of a degradation process and the capability of DTW to find close degradation patterns from similar HI trajectories.

The detailed predictions of 12 failure cases are also presented in the form of kernel densities in Fig. 13. It can be seen that most of the RUL estimations were accurately predicted by spherical-DTW. The black arrow marks on the density curve in cases 1, 15, 37, 44, 57, 65,

91 and 99 highlight the superiority of spherical-DTW over other prognostic methods compared. In these cases, the density curve of spherical-DTW have its highest peak value at the true failure time and has most of its density samples lie close to the peak, indicating that a majority of the predictions are corrected made at or near the true failure time. In comparison, other methods either have a broader probability distribution or have a majority of the density samples lie far away from the true failure time. Besides, spherical-DTW presented similar density curves in cases 85 and 74, demonstrating similar prediction abilities when the TSP moves from the early failure stage to imminent failure. The proposed method was also compared with the BiLSTM-ED method (Yu et al., 2019) in terms of score (S) and accuracy (A). The S and A levels are 306 and 59%, respectively, for the spherical-DTW method; In comparison, the S and A levels for the BiLSTM-ED method are 273 and 57%, respectively. It is worth noting that the focus of this study is on the similarity matching process where a new distance is proposed for similarity-based prognostics, while the BiLSTM-ED method features using a deep learning model to construct HIs for similarity-based prognostics. The prognostic performance of our model can be further improved by including a sophisticated HI construction tool, and this will be considered in our future studies.

Table II. RUL prediction performance of spherical-DTW and its counterparts

	Spherical-DTW	DTW	Euclidean	Manhattan	Frechet	LIP	EDR	Hausdorff
MAE	18.85	19.97	21.25	21.47	20.86	18.92	28.75	22.40
RMSE	26.19	27.29	28.08	28.45	27.28	20.72	26.61	20.67

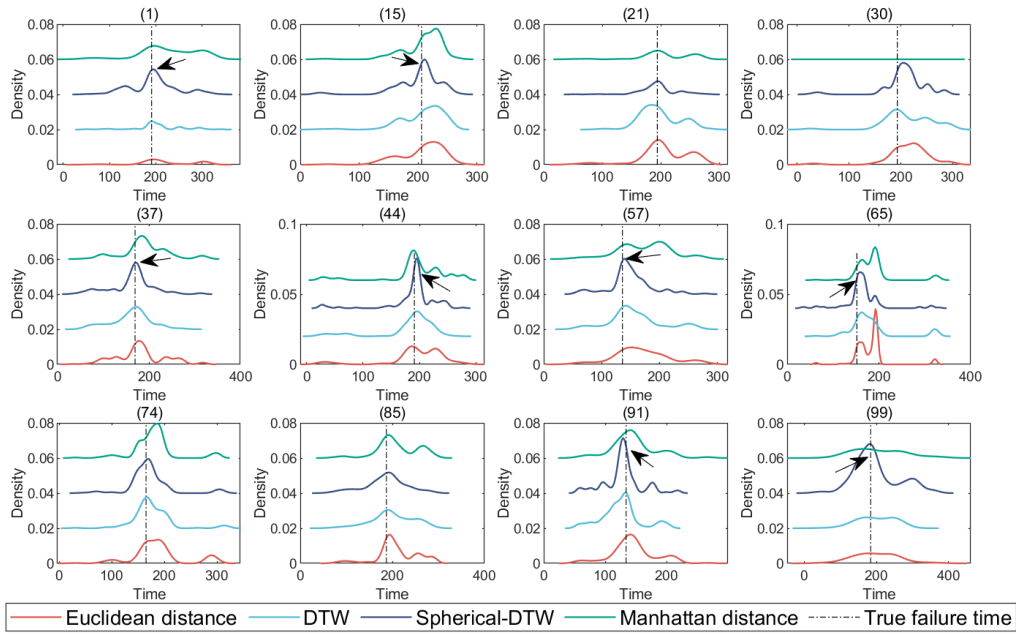


Figure 13. Kernel density of the predicted RUL for various TSPs

4. Conclusion

In this work, a newly designed distance measure - spherical-DTW was put forward for similarity-based prognostics. It jointly considers the nonlinearity of damage propagation and the trajectory alignment with timing difference in the retrieval step, making itself a promising tool for degradation matching based prognostics in the presence of nonlinear fault evolutions and various operating conditions. The definition of spherical distance, proof of its nonlinear characteristic as well as the details for implementing the proposed method were presented. Two case studies were provided to demonstrate the effectiveness of the method in predicting RUL for mechanical systems. The first case study is based on a real-world reciprocating compressor, where valve failures occurred thirteen times during the time period the machine was carrying out various duties. The second case study involves a simulation of gas turbine engine failures, and multi-sensor run-to-failure measurements were recorded. Statistics were carried out to showcase the effectiveness of the proposed DTW-spherical method in comparison to its counterparts. Results indicated that the DTW algorithm outperformed the traditional pointwise distance measure by allowing degradation pattern matching with timing

difference and the inclusion of spherical distance further improved the prediction performance through modelling nonlinear degradation behaviours. Further studies will be carried out to investigate the optimal way to determine the upper and lower bounds of the spherical distance in the presence of large volumes of data. Furthermore, the performance of similarity-based prognostic methods is greatly dependent on the availability of run-to-failure data. If similar degradation curves cannot be found, then the prediction accuracy would be low. This coincides with the prediction results in fig. 8 where excessive errors are observed for case 9 given that case 9 is the right outlier failure case among all 13 historical HIs, hence similar degradation curves are relatively more difficult to be found. The effectiveness of the proposed method and other similarity-based prognostic method can be improved if the amount of failure data grows.

References

- Bleakie, A., & Djurdjanovic, D. (2013). Analytical approach to similarity-based prediction of manufacturing system performance. *Computers in Industry*, *64*(6), 625–633. <https://doi.org/10.1016/j.compind.2013.02.013>
- Botev, B. Z. I., Grotowski, J. F., & Kroese, D. P. (2010). Kernel density estimation via diffusion. *The Annals of Statistics*, *38*(5), 2916–2957. <https://doi.org/10.1214/10-AOS799>
- 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 Journal*, *94*, 106474. <https://doi.org/10.1016/j.asoc.2020.106474>
- Cai, Q., Chen, L., & Sun, J. (2015). Piecewise statistic approximation based similarity measure for time series. *Knowledge-Based Systems*, *85*, 181–195. <https://doi.org/10.1016/j.knosys.2015.05.005>
- Ding, Y., Jia, M., Cao, Y., Ding, P., Zhao, X., & Lee, C. G. (2023). Domain generalization via adversarial out-domain augmentation for remaining useful life prediction of bearings under unseen conditions. *Knowledge-Based Systems*, *261*, 110199.

<https://doi.org/10.1016/j.knosys.2022.110199>

Dusil, R., & Appell, B. (1976). Fatigue and Fracture Mechanics Properties of Valve Steels. *International Compressor Engineering Conference School*.

<https://doi.org/https://docs.lib.purdue.edu/icec/174>

Goebel, K., Qiu, H., Eklund, N., & Yan, W. (2007). Modeling propagation of gas path damage. *IEEE Aerospace Conference Proceedings*, 1–8.

<https://doi.org/10.1109/AERO.2007.352835>

Gu, M., & Chen, Y. (2018). A multi-indicator modeling method for similarity-based residual useful life estimation with two selection processes. *International Journal of System Assurance Engineering and Management*, 9(5), 987–998.

<https://doi.org/10.1007/s13198-018-0708-y>

Hou, M., Pi, D., & Li, B. (2020). Similarity-based deep learning approach for remaining useful life prediction. *Measurement: Journal of the International Measurement Confederation*, 159, 107788. <https://doi.org/10.1016/j.measurement.2020.107788>

Huang, C. G., Huang, H. Z., Peng, W., & Huang, T. (2019). Improved trajectory similarity-based approach for turbofan engine prognostics. *Journal of Mechanical Science and Technology*, 33(10), 4877–4890. <https://doi.org/10.1007/s12206-019-0928-3>

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. *Phm 2019, September*, 1–9. <https://doi.org/10.36001/phmconf.2019.v11i1.788>

Kim, S., Kim, N. H., & Choi, J. H. (2020). Prediction of remaining useful life by data augmentation technique based on dynamic time warping. *Mechanical Systems and Signal Processing*, 136, 106486. <https://doi.org/10.1016/j.ymssp.2019.106486>

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, 104, 799–834. <https://doi.org/10.1016/j.ymssp.2017.11.016>

Li, N., Lei, Y., 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. <https://doi.org/10.1109/TIE.2015.2455055>
- Li, X., Lin, T., Yang, Y., Mba, D., & Loukopoulos, P. (2022). Index similarity assisted particle filter for early failure time prediction with applications to turbofan engines and compressors. *Expert Systems With Applications*, 207(June), 118008. <https://doi.org/10.1016/j.eswa.2022.118008>
- Li, X., Mba, D., Lin, T., Yang, Y., & Loukopoulos, P. (2021). Just-in-time learning based probabilistic gradient boosting tree for valve failure prognostics. *Mechanical Systems and Signal Processing*, 150, 107253. <https://doi.org/10.1016/j.ymsp.2020.107253>
- Li, X., Zhang, W., Ma, H., Luo, Z., & Li, X. (2021). Degradation Alignment in Remaining Useful Life Prediction Using Deep Cycle-Consistent Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 1–12. <https://doi.org/10.1109/TNNLS.2021.3070840>
- Listou Ellefsen, A., Bjørlykhaug, E., Æsøy, V., Ushakov, S., & Zhang, H. (2019). Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture. *Reliability Engineering and System Safety*, 183(June 2018), 240–251. <https://doi.org/10.1016/j.ress.2018.11.027>
- Liu, J., Djurdjanovic, D., Ni, J., Casoetto, N., & Lee, J. (2007). Similarity based method for manufacturing process performance prediction and diagnosis. *Computers in Industry*, 58(6), 558–566. <https://doi.org/10.1016/j.compind.2006.12.004>
- Liu, Y., Hu, X., & Zhang, W. (2019). Remaining useful life prediction based on health index similarity. *Reliability Engineering and System Safety*, 185, 502–510. <https://doi.org/10.1016/j.ress.2019.02.002>
- Loukopoulos, P., Pilidis, P., Bennett, I., Zolkiewski, G., Xiaochuan, L., & Mba, D. (2018). Abrupt fault remaining useful life estimation using measurements from a reciprocating compressor valve failure. *Mechanical Systems and Signal Processing*.
- Loukopoulos, P., Zolkiewski, G., Bennett, I., Sampath, S., Pilidis, P., Li, X., & Mba, D. (2019). Abrupt fault remaining useful life estimation using measurements from a reciprocating compressor valve failure. *Mechanical Systems and Signal Processing*,

121, 359–372. <https://doi.org/10.1016/j.ymsp.2018.09.033>

- Loutas, T., Eleftheroglou, N., Georgoulas, G., Loukopoulos, P., Mba, D., & Bennett, I. (2019). Valve Failure Prognostics In Reciprocating Compressors Utilizing Temperature Measurements, PCA-based Data Fusion And Probabilistic Algorithms. *IEEE Transactions on Industrial Electronics*, 1–1. <https://doi.org/10.1109/TIE.2019.2926048>
- Lyu, J., Ying, R., Lu, N., & Zhang, B. (2020). Remaining useful life estimation with multiple local similarities. *Engineering Applications of Artificial Intelligence*, 95(January), 103849. <https://doi.org/10.1016/j.engappai.2020.103849>
- Ma, G., Wang, Z., Liu, W., Fang, J., Zhang, Y., Ding, H., & Yuan, Y. (2023). A two-stage integrated method for early prediction of remaining useful life of lithium-ion batteries. *Knowledge-Based Systems*, 259, 110012. <https://doi.org/10.1016/j.knosys.2022.110012>
- Ma, W., Guo, L., Gao, H., Yu, Y., & Qian, M. (2022). A health indicator construction method based on self-attention convolutional autoencoder for rotating machine performance assessment. *Measurement: Journal of the International Measurement Confederation*, 204(September), 112108. <https://doi.org/10.1016/j.measurement.2022.112108>
- Ramasso, E., & Saxena, A. (2014). Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets. *International Journal of Prognostics and Health Management*, ISSN2153-2648, 1–15.
- Ren, L., Qin, H., Xie, Z., Li, B., & Xu, K. (2022). Aero-Engine Remaining Useful Life Estimation Based on Multi-Head Networks. *IEEE Transactions on Instrumentation and Measurement*, 71, 1–10. <https://doi.org/10.1109/TIM.2022.3149094>
- Sakoe, H., & Hiba, S. (1978). Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1).
- Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling

- for aircraft engine run-to-failure simulation. *2008 International Conference on Prognostics and Health Management, PHM 2008*.
<https://doi.org/10.1109/PHM.2008.4711414>
- Sun, J., Yan, Z., Han, Y., Zhu, X., & Yang, C. (2023). Deep Learning Framework for Gas Turbine Performance Digital Twin and Degradation Prognostics from Airline Operator Perspective. *Reliability Engineering and System Safety*, 109404.
<https://doi.org/10.1016/j.ress.2023.109404>
- Tian, J., Jiang, Y., Zhang, J., Wu, S., & Luo, H. (2023). A Novel Transfer Ensemble Learning Framework for Remaining Useful Life Prediction under Multiple Working Conditions. *IEEE Transactions on Instrumentation and Measurement*, 72, 1–11.
<https://doi.org/10.1109/TIM.2023.3273676>
- Wang, H., Ni, G., Chen, J., & Qu, J. (2020). Research on rolling bearing state health monitoring and life prediction based on PCA and Internet of things with multi-sensor. *Measurement: Journal of the International Measurement Confederation*, 157, 107657.
<https://doi.org/10.1016/j.measurement.2020.107657>
- Wang, T., Yu, J., Siegel, D., & Lee, J. (2008a). A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems. *Prognostics and Health Management, 2008. PHM 2008. International Conference On*, 1–6.
<https://doi.org/10.1109/PHM.2008.4711421>
- Wang, T., Yu, J., Siegel, D., & Lee, J. (2008b). A similarity based prognostic approach for remaining useful life estimation of engineered systems. *International Conference on Prognostics and Health Management*, 4–9.
- Wu, J., Wu, C., Cao, S., Or, S. W., Deng, C., & Shao, X. (2018). Degradation Data-Driven Time-To-Failure Prognostics Approach for Rolling Element Bearings in Electrical Machines. *IEEE Transactions on Industrial Electronics*, 66(1), 529–539.
<https://doi.org/10.1109/TIE.2018.2811366>
- Wu, X., Xu, L., Wang, J., Yang, D., Li, F., & Li, X. (2020). A prognostic-based dynamic optimization strategy for a degraded solid oxide fuel cell. *Sustainable Energy*

- Technologies and Assessments*, 39(March). <https://doi.org/10.1016/j.seta.2020.100682>
- Xue, B., Xu, F., Huang, X., Xu, Z., & Zhang, X. (2022). Improved similarity based prognostics method for turbine engine degradation with degradation consistency test. *Applied Intelligence*, 52(9), 10181–10201. <https://doi.org/10.1007/s10489-021-03034-6>
- Xue, B., Xu, H., Huang, X., Zhu, K., Xu, Z., & Pei, H. (2022). Similarity - based prediction method for machinery remaining useful life : A review. In *The International Journal of Advanced Manufacturing Technology*. Springer London. <https://doi.org/10.1007/s00170-022-09280-3>
- Yang, Y., & Chiclana, F. (2009). Intuitionistic Fuzzy Sets: Spherical Representation and Distances. *International Journal of Intelligent Systems*, 24, 399–420. <https://doi.org/10.1002/int>
- Yu, W., Kim, I. Y., & Mechefske, C. (2019). Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme. *Mechanical Systems and Signal Processing*, 129, 764–780. <https://doi.org/10.1016/j.ymssp.2019.05.005>
- Yu, W., Kim, I. Y., & Mechefske, C. (2020). An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme. *Reliability Engineering and System Safety*, 199(March), 106926. <https://doi.org/10.1016/j.res.2020.106926>
- Yu, W., Kim, I. Y., & Mechefske, C. (2021). Analysis of different RNN autoencoder variants for time series classification and machine prognostics. *Mechanical Systems and Signal Processing*, 149, 107322. <https://doi.org/10.1016/j.ymssp.2020.107322>
- Zeming, L., Jianmin, G., & Hongquan, J. (2019). A maintenance support framework based on dynamic reliability and remaining useful life. *Measurement: Journal of the International Measurement Confederation*, 147, 106835. <https://doi.org/10.1016/j.measurement.2019.07.063>
- Zhai, Q., & Ye, Z. S. (2017). RUL Prediction of Deteriorating Products Using an Adaptive Wiener Process Model. *IEEE Transactions on Industrial Informatics*, 13(6), 2911–

2921. <https://doi.org/10.1109/TII.2017.2684821>

Zhang, Y., Yang, Y., Li, H., Xiu, X., & Liu, W. (2022). A Data-Driven Modeling Method for Stochastic Nonlinear Degradation Process With Application to RUL Estimation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(6), 3847–3858. <https://doi.org/10.1109/TSMC.2021.3073052>

Zhang, Z., Hu, C., Si, X., Zhang, J., & Zheng, J. (2017). Stochastic degradation process modeling and remaining useful life estimation with flexible random-effects. *Journal of the Franklin Institute*, 354(6), 2477–2499. <https://doi.org/10.1016/j.jfranklin.2016.06.039>