Highlights

Estimation of Remaining Useful Life of Rolling Element Bearings based on the Adaptive Kernel Kalman Filter

Z. Li, R. Zhu, T. Verwimp, H. Wen, K. Gryllias

- A methodology for the RUL estimation of REBs is proposed
- The methodology is validated on three datasets
- The proposed methodology demonstrates promising results

Estimation of Remaining Useful Life of Rolling Element Bearings based on the Adaptive Kernel Kalman Filter

Z. Li[a,b], R. Zhu[a,b], T. Verwimp[a,b], H. Wen[a,b], K. Gryllias[a,b]

[a] *KU Leuven, Department of Mechanical Engineering, LMSD - Mecha(tro)nic System Dynamics, Celestijnenlaan 300, Heverlee, B-3001, Belgium*
[b] *Flanders Make@KU Leuven, Belgium*

Abstract

Estimating the Remaining Useful Life (RUL) of Rolling Element Bearings (REBs) is a very challenging task, complicating the optimal scheduling of shutdowns and maintenance operations in industry. In recent years, a number of prognostic methodologies have been proposed, mainly categorized into 3 groups: physical model-based, AI-based, and statistical-based methodologies. Statistical based methodologies, including the Kalman filter and its variants, are commonly used in RUL estimation due to their explainability and efficiency. However, their reliance on specific Health Indicators (HIs) often restricts their generalization capabilities for different scenarios. Additionally they often face accuracy and robustness issues due to the nonlinearity of the degradation trends in rotating machines. To overcome these limitations, this study introduces a prognostic methodology based on the Adaptive Kernel Kalman Filter (AKKF), which integrates HI extraction, anomaly and failure thresholds setting, parameters estimation in the data and kernel space, and uncertainty quantification. The proposed methodology is applied to three bearing degradation datasets: an in-house dataset captured on the KU Leuven gearbox prognostics test rig, and two publicly available datasets from the University of Ferrara (UNIFE) and Xi'an Jiaotong University (XJTU). The performance of the methodology is evaluated using different metrics and is compared with the Extended Kalman Filter (EKF). The results from the aforementioned three datasets indicate that the AKKF-based methodology is promising to be used for the highly nonlinear degradation trends of REBs, achieving good results. Moreover, several issues, including the HI selection, the degradation model selection and the filter selection are discussed in detail.

Keywords:

Rolling Element Bearings, Estimation of Remaining Useful Life, Adaptive Kernel Kalman Filter, Prognostics, Run-to-failure tests

Nomenclature

AT	Anomaly Threshold
EoL	End of Life
$f(\cdot)$	State function
F_{thres}	Failure threshold
Γ_{n-1}	Transformation matrix for particle draw step
$\mathcal{G}_{\mathbf{yy}}$	Gram matrix for output space
FT	Failure Threshold
$h(\cdot)$	Measurement function
HI	Health Indicator
I	Identity matrix
κ	Regularization parameter for matrix inversion
λ	Stabilization parameter for matrix inversion
MM	Moving Median
$MMAD$	Moving Median Absolute Deviation
$\phi_{\mathbf{x}}(\mathbf{x})$	Feature mapping of $\mathbf{x}$
$\mathcal{H}_{\mathbf{x}}$	Reproducing Kernel Hilbert Space (RKHS) for data space $\mathcal{X}$
$\hat{\mu}_X$	Empirical Kernel Mean Embedding (KME) of X
$RUL(t)$	Remaining Useful Life at time t
S_n	Kernel covariance matrix at step n
σ	Bandwidth of the Gaussian kernel
T	Transition matrix in kernel space
T_{life}	Time at life end when HI reaches failure threshold
V	Transition residual
$\mathbf{w}_n$	Kernel weight vector at step n
$\mathbf{x}_n$	Hidden states at step n
$\hat{\mathbf{x}}_{n-1}$	Approximate states at step $n-1$
$X^{\mathcal{D}}_{n-1}$	States in the data space at step $n-1$
$\hat{\Sigma}_{n-1}$	Approximate state covariance at step $n-1$
$\langle\cdot\rangle$	Inner product operator
$\otimes$	Tensor product
Φ	Collection of feature mappings for the states
Υ	Collection of feature mappings for the measurements

$\mathbf{y}_n$	Observation at step n
$^+$	Indication of a posterior estimation
$^-$	Indication of a prior estimation
$k_{\mathbf{x}}(\mathbf{x}, \mathbf{x}^{'})$	Kernel function
$\hat{\mathcal{C}}_{XY}$	Empirical covariance in RKHS
$\mathcal{K}_n$	Adaptive Kernel Kalman Filter (AKKF) gain at step n

1. Introduction

Rolling Element Bearings (REBs) are key components in rotating machinery being the interface between stationary and rotating parts. Bearing failures account for 40% of total electrical motor failures [1], as shown in Fig. 1. Therefore, monitoring the health of REBs and scheduling timely maintenance is crucial to mitigate time waste, economic losses, and potential hazards [2]. Condition monitoring (CM) plays a crucial role in ensuring the reliability of REBs [3, 4, 5, 6]. The main type of CM consists of three stages, i.e., fault or anomaly detection [7], fault diagnosis [8] and fault prognosis or estimation of the Remaining Useful Life (RUL) [9]. Within CM, prognostics is the final step, presenting significant challenges due to the complexity, uncertainties, and non-linearities of the degradation trends in REBs. Moreover there are different definitions for the RUL estimation for different systems or components [10, 11] based on the varying demands of the industry. In this study, the RUL estimation is defined as how long REBs can be used at a given time point [11] after a defect has occured. Therefore, RUL estimation can also be explained as prediction of the future degradation trend.

Over the years, various methodologies have been developed in the field of prognostics. These methodologies generally fall into three categories, that is, approaches based on physics models [12], methods driven by Artificial Intelligence (AI) [9, 13, 14] and statistical methodologies [15, 16, 17]. Physics model-based approaches can reveal the failure mechanism and consequently could provide an accurate prediction. However, the construction of a physical model requires detailed physical parameters, which is generally computationally consuming and limits the generality and applicability of the approach for RUL estimation. Moreover AI methods aim to model complex relationships between extracted features and degradation patterns [6]. These methods require big data to train the model and are relatively time-consuming as they usually have a large number of parameters to estimate, making the

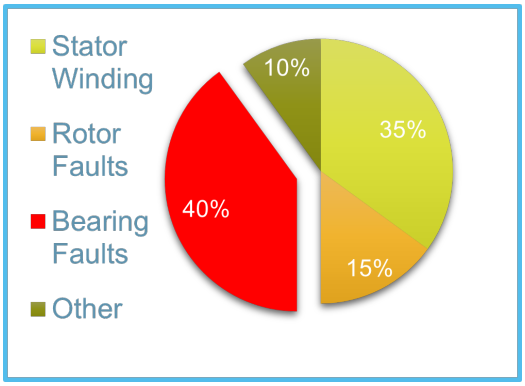


Figure 1: Distribution of main failures of electrical motors [1]

implementation in the industry still rather difficult. Moreover, model interpretability is crucial for gaining trust in the results of CM. In contrast, statistical approaches, such as Kalman Filter (KF) and its variants (for instance, Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), Switching Kalman filter (SKF), etc.), Wiener processes [18], and Markov models [19], offer greater explainability, rooted in empirical functions and probability theory. More specifically, Singleton et al. proposed a stochastic modeling based approach to predict the RUL of REBs by introducing the EKF [20], where different vibration features, i.e., the Rényi entropy of time-frequency distributions, can be tracked in a physically explainable way. Lim et al. [21] reported a Switching Kalman Filter (SKF) for model estimation and life prediction, where multiple degradation models are employed and the most probable model is inferred from the CM data using Bayesian estimation before it is applied for prediction. Furthermore, Cui et al. [22] combined UKF and SKF and proposed a novel Switched Unscented Kalman Filter (SUKF), which can handle the declining trend in the CM data where the SKF algorithm may perform inadequately. The authors of [23] proposed a novel methodology, which incorporates health indicator (HI) selection, a robust anomaly detection technique (Support Vector Data Description), combined with the multi-step estimation method Moving Horizon Estimation (MHE), to estimate the RUL of rolling element bearings. The results indicate that the performance of the MHE is better than traditional KF as MHE considers not only the status of the previous step but also the status of several prior steps.

While the methodologies discussed earlier offer promising results, they

are generally dependent on complex or specialized HIs, which limits their universality in practical applications. The optimal HI for a given scenario may vary [24], even under identical operating conditions and equipment, due to factors such as variability in material properties, installation discrepancies, and the type of faults encountered. This dependency on specific HIs complicates the applicability of these methods across different components and systems, as the most effective HI can change due to these unforeseen variables. Furthermore, the full potential of the KF as an estimator has not yet been completely realized. The KF determines the optimal solution by balancing observed values with theoretical predictions, implying its effectiveness should, in theory, be independent of HI selection if the nonlinear system is properly designed. However, its effectiveness diminishes in complex nonlinear systems where linearization techniques are necessary to mitigate errors and enhance stability. Lastly, the uncertainty of the estimation results needs to be emphasized. Sometimes, this is more important than the accuracy of the RUL estimates, as it directly affects maintenance decision-making.

To address the aforementioned challenges, a simple and generalizing feature, the Root Mean Square (RMS), is selected as the HI for RUL estimation in this study. RMS is both stable and broadly applicable across most vibration signals, making it suitable for implementation in a wide range of rotating machinery, as it effectively represents the system's power. However, RMS may introduce fluctuations during RUL estimation, which will affect the stability and accuracy of the estimation results. To mitigate this issue, a methodology based on Adaptive Kernel Kalman Filter (AKKF) is proposed. This approach employs a kernel-based technique to effectively handle nonlinearities inherent in the degradation process. Specifically, the kernel facilitates the capturing and processing of critical features of the nonlinear system within a high-dimensional kernel space, enabling more stable and accurate estimation of the REBs' degradation state. Furthermore, the Monte Carlo method is employed to sample the estimation parameters, thereby providing quantified uncertainty and enhancing the reliability of the predictions. The efficacy of the proposed methodology is validated in multiple datasets, demonstrating its high stability and generalization capability.

The remainder of the paper is structured as follows: Section 2 introduces the Adaptive Kernel Kalman Filter (AKKF) along with the relevant theoretical background. Section 3 provides a comprehensive overview of the RUL estimation framework. Moreover, three case studies are presented in Section 4. The first dataset originates from the in-house KU Leuven Gearbox Prog-

nostics Test Rig, while the second and third datasets have been made publicly available respectively from the University of Ferrara and the Xi'an JiaTong University. For performance comparison purposes, the EKF is implemented, and various metrics are utilized to thoroughly analyze the performance of the proposed methodology. Section 5 offers an in-depth analysis of the methodology, drawing on the results from the case studies. The paper concludes with a summary of the key findings in Section 6.

2. Adaptive Kernel Kalman Filter - AKKF

2.1. Empirical Kernel Mean Embedding

A system can be described by a dynamic model in the data space, as shown in Eqs. (1) and (2). Eq. (1) describes that the hidden states of a system evolving in time follow a certain function with noise, which is called the state function or the progress function. Eq. (2), known as the measurement function, depicts how the hidden states are observed. The KF is used to estimate the most optimal states by considering the state noise and the measurement noise.

$$\mathbf{x}_n = f(\mathbf{x}_{n-1}, \mathbf{u}_n) \tag{1}$$

$$\mathbf{y}_n = h(\mathbf{x}_n, \mathbf{v}_n) \tag{2}$$

Here, $\mathbf{x}_n$ represents the hidden states at step n, $n = 1, 2, ..., N$, while y_n denotes the observation at step n. The classic KF can only handle linear problems. Based on the literature [25, 26, 27, 28], it can be assumed that:

Assumption 1: In the kernel feature spaces (i.e. the Reproducing Kernel Hilbert Space (RKHS)), non-linear estimations can be solved using linear operations.

Therefore, to perform the calculation in a higher-dimensional space (in this study, the RKHS is implemented), the features depicted in the data space (for the RUL estimation, these are the parameters of the state function and the observations) need to be mapped into the kernel space. The Gaussian kernel function, which has an infinite feature dimension, is selected in this work to map the features into the RKHS. For an instance $\mathbf{x}$ of a random variable X in the data space $\mathcal{X}$, the feature mapping is denoted as $\phi_{\mathbf{x}}(\mathbf{x}) = k_{\mathbf{x}}(\mathbf{x}, \cdot) \in \mathcal{H}_{\mathbf{x}}$. $\mathcal{H}_{\mathbf{x}}$ represents a RKHS for the data space $\mathcal{X}$. The function $k_{\mathbf{x}}(\mathbf{x}, \mathbf{x}^{'})$ is a kernel function. Specifically, a Gaussian kernel function is given by $\langle \phi_{\mathbf{x}}(\mathbf{x}), \phi^{'}_{\mathbf{x}}(\mathbf{x}^{'}) \rangle_{\mathcal{H}_{\mathbf{x}}} = k_{\mathbf{x}}(\mathbf{x}, \mathbf{x}^{'}) = e^{-\frac{1}{\sigma^2}||(\mathbf{x}-\mathbf{x}^{'}||^2}$, where the σ is the bandwidth of the Gaussian kernel and $\langle \cdot \rangle$ is the operator of the inner product.

Considering the probability distribution function (pdf) $p(x)$ of the instance $\mathbf{x}$, its Empirical Kernel Mean Embedding (KME) is approximated as the average of the samples' feature mapping, viz.

$$\hat{\mu}_X = \frac{1}{M}\sum_{i=1}^{M}\phi_{\mathbf{x}}(\mathbf{x}^{\{i\}}) \tag{3}$$

where M represents the number of samples. The symbol $\hat{}$ is the average operator. Moreover the joint pdf $p(\mathbf{x}, \mathbf{y})$ can be approximated as:

$$\hat{\mathcal{C}}_{XY} = \frac{1}{M}\sum_{i=1}^{M}\phi_{\mathbf{x}}(\mathbf{x}^{\{i\}}) \otimes \phi_{\mathbf{y}}(\mathbf{y}^{\{i\}}) \tag{4}$$

where $\otimes$ denotes the tensor product. For feature mapping using kernel functions, the equation $\langle\phi_{\mathbf{x}}(\mathbf{x}), \phi_{\mathbf{y}}(\mathbf{y})\rangle_{\mathcal{H}_{\mathbf{x}}} = k_{\mathbf{x}}(\mathbf{x}, \mathbf{x}^{'})k_{\mathbf{y}}(\mathbf{y}, \mathbf{y}^{'})$ is satisfied. For the sake of simplicity, let $\Phi = [\phi_{\mathbf{x}}(\mathbf{x}^{\{1\}}), \phi_{\mathbf{x}}(\mathbf{x}^{\{2\}}), ..., \phi_{\mathbf{x}}(\mathbf{x}^{\{M\}})]$ and $\Upsilon = [\phi_{\mathbf{y}}(\mathbf{y}^{\{1\}}), \phi_{\mathbf{y}}(\mathbf{y}^{\{2\}}), ..., \phi_{\mathbf{y}}(\mathbf{y}^{\{M\}})]$. Subsequently, Eq. (4) becomes:

$$\hat{\mathcal{C}}_{XY} = \frac{1}{M}\Phi\Upsilon^{\mathrm{T}} \tag{5}$$

It should be noted that for the Gaussian kernel the kernel trick needs to be implemented since there is no explicit mapping function. Similarly, the KME of a conditional pdf is shown in Eq. (6).

$$\hat{\mu}_{X|\mathbf{y}} = \hat{\mathcal{C}}_{X|Y}\phi(\mathbf{y}) \tag{6}$$

where $\hat{\mathcal{C}}_{X|Y}$ is the estimator of the conditional operator. It can be derived as:

$$\hat{\mathcal{C}}_{X|Y} = \hat{\mathcal{C}}_{XY}(\hat{\mathcal{C}}_{YY} + \kappa I)^{-1} \tag{7}$$

where κ is a hyper-parameter used to stabilise the matrix inversion operation, while I is the identity matrix. By substituting Eq. (5) into Eq. (7), it can be written as:

$$\hat{\mathcal{C}}_{X|Y} = \hat{\mathcal{C}}_{XY}(\hat{\mathcal{C}}_{YY} + \kappa I)^{-1} = \Phi(\mathcal{G}_{\mathbf{yy}} + \kappa I)^{-1}\Upsilon^{\mathrm{T}} \tag{8}$$

where $\mathcal{G}_{\mathbf{yy}} = \Upsilon^{\mathrm{T}}\Upsilon$ is called the Gram matrix. So, Eq. (6) can be re-written as:

$$\hat{\mu}_{X|\mathbf{y}} = \Phi(\mathcal{G}_{\mathbf{yy}} + \kappa I)^{-1}(\Upsilon^{\mathrm{T}}\phi(\mathbf{y})) \tag{9}$$

In the above equation there are three terms Φ, $(\mathcal{G}_{\mathbf{yy}}+\kappa I)^{-1}$, and $\Upsilon^{\mathrm{T}}\phi(\mathbf{y})$. The first 2 terms can be seen as a kernel weight vector, i.e., $\mathbf{w} = (\mathcal{G}_{\mathbf{yy}}+\kappa I)^{-1}\mathcal{G}_{:,\mathbf{y}}$. And the third term, $\Upsilon^{\mathrm{T}}\phi(\mathbf{y})$, can be written as $[k_{\mathbf{y}}(\mathbf{y}^{\{1\}},\mathbf{y})], [k_{\mathbf{y}}(\mathbf{y}^{\{2\}},\mathbf{y})], ..., [k_{\mathbf{y}}(\mathbf{y}^{\{M\}},\mathbf{y})]^{\mathrm{T}}$, which is the kernel function denoted as $\mathcal{G}_{:,\mathbf{y}}$. Therefore, Eq. (9) shows that in the RKHS, the KME of the conditional distribution can be explained as a weighted feature mapping of different particles. Since Eq. (9) is only related to the linear matrix operation, the KF rule can be implemented in the RKHS. Finally, the following equation is derived as:

$$\hat{\mu}_{X|\mathbf{y}} = \Phi\mathbf{w} \tag{10}$$

2.2. Kernel Kalman Rule

Next, the Kalman Rule can be extended into the kernel space, which is the Kernel Kalman Rule (KKR). In the KF, the mean value and the pdf of the Gaussian distribution can be predicted by the state function and corrected by the information obtained from the observation according to the Minimization of the Mean Squared Error (MMSE). Similarly, the KKR [28] employs the KF for the kernel weight vector and the kernel covariance matrix using the mapping of the states and the observations.

For clarity, the symbol $^+$ is denoted as the posterior estimation for each variable, while the symbol $^-$ indicates the prior estimation. Therefore, the KKR is described by the following equations at the step n:

$$\begin{cases} \mathbf{w}_n^- = T\mathbf{w}_{n-1}^+ & (a) \\ S_n^- = TS_{n-1}^+T^{\mathrm{T}} + V & (b) \\ \mathbf{w}_n^+ = \mathbf{w}_n^- + \mathcal{K}_n(\mathcal{G}_{:,\mathbf{y}_n} - \mathcal{G}_{\mathbf{yy}}\mathbf{w}_n^-) & (c) \\ S_n^+ = S_n^- - \mathcal{K}_n\mathcal{G}_{\mathbf{yy}}S_n^- & (d) \\ \mathcal{K}_n = S_n^-(\mathcal{G}_{\mathbf{yy}}S_n^- + \kappa I)^{-1} & (e) \end{cases} \tag{11}$$

where $\mathbf{w}_n$ and S_n represent the kernel weight vector and the kernel covariance matrix in the kernel space, respectively. T is the transition matrix in the kernel space, and V is the transition residual. $\mathcal{G}_{:,\mathbf{y}}$ and $\mathcal{G}_{\mathbf{yy}}$ are the kernel function and the Gram matrix mentioned before. $\mathcal{K}_n$ is the AKKF gain in the kernel space. Traditionally, the transition matrix T needs to be determined by training data [29], which is not effective for online RUL estimation. Therefore, the adaptive KKR [28], where a particle draw step is added for the kernel adaption after the update step, is implemented in this study. According to this approach, the kernel weight vector and the

kernel covariance matrix are re-estimated based on the particle drawn from the updated distribution.

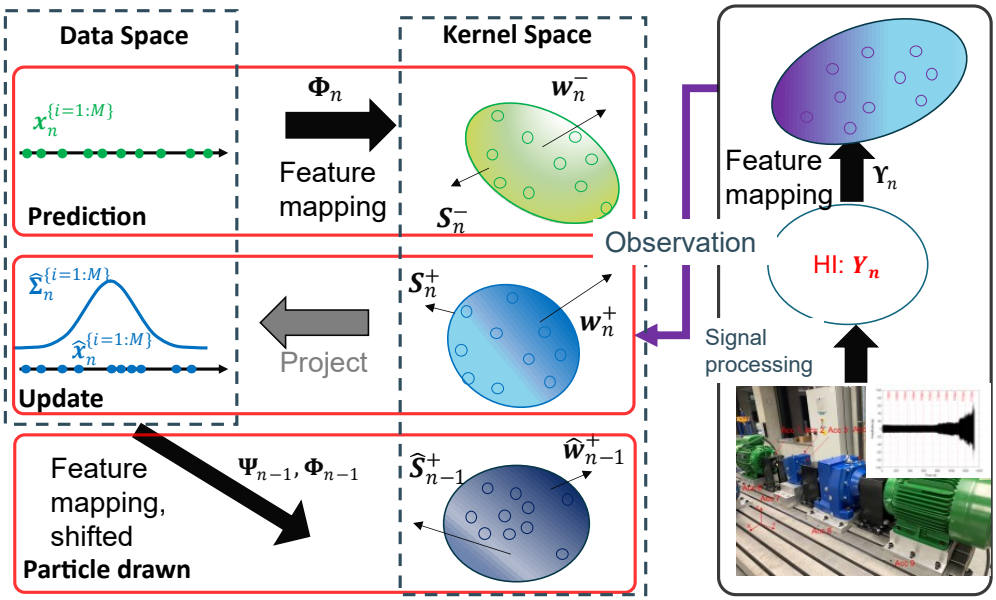


Figure 2: The diagram of the AKKF.

2.3. Adaptive Kernel Kalman Filter Realization

As depicted in Fig. 2, the AKKF is divided into 3 steps, Prediction, Update and Particle Draw, respectively. The entire procedure forms a loop, with the Particle Draw step as a potential starting point to clarify the process.

(1) Particle drawn step

Let's assume a state $X_{n-1}^{\mathcal{D}} = [\mathbf{x}_{n-1}^{\{1\}}, \mathbf{x}_{n-1}^{\{2\}}, ..., \mathbf{x}_{n-1}^{\{M\}}]$ in the data space and its mapping $\Phi_{n-1} = [\phi_{\mathbf{x}}(\mathbf{x}_{n-1}^{\{1\}}), \phi_{\mathbf{x}}(\mathbf{x}_{n-1}^{\{2\}}), ..., \phi_{\mathbf{x}}(\mathbf{x}_{n-1}^{\{M\}})]$ in the kernel space, where M represents the number of particles. Based on the Eqs. (3) and (5), the posterior KME and its covariance can be calculated at step $n-1$ in the kernel space as follows:

$$\hat{\mu}_{\mathbf{X}_{n-1}}^{+} = \Phi_{n-1}\mathbf{w}_{n-1}^{+} \tag{12}$$

$$\hat{\mathcal{C}}_{\mathbf{X}_{n-1}\mathbf{X}_{n-1}} = \Phi_{n-1}S_{n-1}^{+}\Phi_{n-1}^{\mathrm{T}} \tag{13}$$

Thus, the new particles of the states $\hat{X}^{\mathcal{D}}_{n-1} = [\hat{\mathbf{x}}^{\{1\}}_{n-1}, \hat{\mathbf{x}}^{\{2\}}_{n-1}, ..., \hat{\mathbf{x}}^{\{M\}}_{n-1}]$ in the data space for a Gaussian kernel can be drawn from the approximated distribution $\mathcal{N}(\hat{\mathbf{x}}_{n-1}, \hat{\Sigma}_{n-1})$:

$$\hat{\mathbf{x}}_{n-1} = X^{\mathcal{D}}_{n-1}\mathbf{w}^{+}_{n-1} \tag{14}$$

$$\hat{\Sigma}_{n-1} = X^{\mathcal{D}}_{n-1}S^{+}_{n-1}X^{\mathcal{D}\mathrm{T}}_{n-1} \tag{15}$$

where the new feature mapping can be denoted as $\Psi_{n-1} = [\phi_{\mathbf{x}}(\hat{\mathbf{x}}^{\{1\}}_{n-1}), \phi_{\mathbf{x}}(\hat{\mathbf{x}}^{\{2\}}_{n-1}), ..., \phi_{\mathbf{x}}(\hat{\mathbf{x}}^{\{M\}}_{n-1})]$. Then, the adapted kernel weight vector $\mathbf{w}^{+}_{n-1}$ and the adapted kernel covariance matrix S^{+}_{n-1} can be calculated as:

$$\hat{\mathbf{w}}^{+}_{n-1} = \Gamma_{n-1}\mathbf{w}^{+}_{n-1} \tag{16}$$

$$\hat{S}^{+}_{n-1} = \Gamma_{n-1}S^{+}_{n-1}\Gamma^{\mathrm{T}}_{n-1} \tag{17}$$

where $\Gamma_{n-1} = (\Psi^{\mathrm{T}}_{n-1}\Psi_{n-1} + \lambda I)^{-1}\Psi^{\mathrm{T}}_{n-1}\Phi_{n-1}$. Similar to κ mentioned in previous part, λ is a parameter to stabilize the matrix inversion. Eqs. (16) and (17) can be derived by the assumption that $\Psi_{n-1}\hat{\mathbf{w}}^{+}_{n-1} = \Phi_{n-1}\mathbf{w}^{+}_{n-1}$, which means that the posterior KME should be equal before and after new particles are drawn. Therefore, the kernel weight vector and the kernel covariance matrix will be adapted instead of the training transition matrix T.

Fig. 2 depicts this particle drawn process. The kernel weight and the covariance are adapted by the mapping of the updated state particles Ψ_{n-1} and the mapping of the previous state particles Φ_{n-1} in the kernel space using Eqs. (16) and (17).

(2) Prediction step

Based on Eq. (1), the states in the data space at step n can be calculated as:

$$\mathbf{x}^{\{i\}}_{n} = f(\hat{\mathbf{x}}^{\{i\}}_{n-1}, \mathbf{u}_n), i = 1, 2, ..., M \tag{18}$$

where the feature mapping $\Phi_n = \left[\phi_{\mathbf{x}}(\mathbf{x}^{\{1\}}_{n}), \phi_{\mathbf{x}}(\mathbf{x}^{\{2\}}_{n}), ..., \phi_{\mathbf{x}}(\mathbf{x}^{\{M\}}_{n})\right]$ in the kernel space can be obtained by the particles of the states $X^{\mathcal{D}}_{n} = [\mathbf{x}^{\{1\}}_{n}, \mathbf{x}^{\{2\}}_{n}, ..., \mathbf{x}^{\{M\}}_{n-1}]$ in the data space. Then, these particles of the states can be projected into the kernel space. Meanwhile, the kernel weight vector and the kernel covariance matrix can be calculated as:

$$\mathbf{w}^{-}_{n} = \hat{\mathbf{w}}^{+}_{n-1} \tag{19}$$

$$S_n^- = \hat{S}_{n-1}^+ + V_n \tag{20}$$

where V_n is the transition residual matrix [28] and it can be calculated by:

$$V_n = \frac{1}{M}[(\Phi_n^{\mathrm{T}}\Phi_n + \lambda I)^{-1}\Phi_n^{\mathrm{T}}\Phi_n - I][\Phi_n^{\mathrm{T}}\Phi_n + \lambda I)^{-1}\Phi_n^{\mathrm{T}}\Phi_n - I]^{\mathrm{T}} \tag{21}$$

As shown in Fig. 2, the green particles at time step n are propagated from the last timestamp via the state function. These green particles are then mapped into the kernel space to calculate the kernel weight covariance.

(3) Update step

For the Update step n, the measurement function (Eq. (2)) is used in the data space:

$$\mathbf{y}_n^{\{i\}} = h(\mathbf{x}_n^{\{i\}}, \mathbf{v}_n), i = 1, 2, ..., M \tag{22}$$

Similar to the Prediction step, the observations are mapped into the kernel space, where the feature mapping would be $\Upsilon_n = [\phi_{\mathbf{y}}(\mathbf{y}_n^{\{1\}}), \phi_{\mathbf{y}}(\mathbf{y}_n^{\{2\}}), ..., \phi_{\mathbf{y}}(\mathbf{y}_n^{\{M\}})]$. Then Eq. (11) (c)-(e) can be employed to update the kernel weight vector and the kernel covariance matrix, which are:

$$\mathbf{w}_n^+ = \mathbf{w}_n^- + \mathcal{K}_n(\mathcal{G}_{:,\mathbf{y}_n} - \mathcal{G}_{\mathbf{yy}}\mathbf{w}_n^-) \tag{23}$$

$$S_n^+ = S_n^- - \mathcal{K}_n\mathcal{G}_{\mathbf{yy}}S_n^- \tag{24}$$

$$\mathcal{K}_n = S_n^-(\mathcal{G}_{\mathbf{yy}}S_n^- + \kappa I)^{-1} \tag{25}$$

$\mathcal{G}_{:,\mathbf{y}}$ and $\mathcal{G}_{\mathbf{yy}}$ are the kernel function and the Gram matrix mentioned before, respectively.

As shown in Fig. 2, the observation information processed from the raw vibration data is used to update the predicted kernel weight vector and the corresponding covariance matrix in the kernel space. Then new updated state distribution is approximated using the updated kernel weight and covariance. New particles are drawn from this new distribution and stored for the next loop.

Finally, the AKKF algorithm can be employed for the RUL estimation of REBs. In the next section, the proposed RUL estimation framework is summarized.

3. Proposed framework of RUL estimation

In order to estimate the RUL using the AKKF, two commonly used assumptions are considered:

Assumption 2: The degradation curve of REBs follows exponential equations.

Assumption 3: RUL is calculated based on the equation:

$$RUL(t) = EoL - t \tag{26}$$

$$EoL = solve(f(\boldsymbol{\theta_p}; T_{life}) = F_{thres}, T_{life}) \tag{27}$$

where t is the current time. EoL represents the End of Life. It is calculated by solving the equation of $f(\boldsymbol{\theta_p}; T_{life}) = F_{thres}$, where $f(\boldsymbol{\theta_p}; T_{life})$ represents a degradation model of the bearing based on parameters $\boldsymbol{\theta_p}$, and F_{thres} is the failure threshold. T_{life} represents the time at which the HI reaches the failure threshold. Consequently, the RUL is assumed to be negatively correlated to the current time t.

Fig. 3 illustrates the RUL estimation for a system, represented over time. Observed HI values are shown as blue circles, with a black solid line indicating the filtered values derived from these observations using an estimator. The black dashed line projects the estimated future values of the HI with a confidence interval represented by black dash-dot lines. In addition to the observations, state functions need to be designed for estimators. With the state functions, Eq. (27) is used for RUL calculations.

Single exponential and double exponential equations are employed for the state function in this study, as they are commonly used in bearing RUL estimation and demonstrate better fitting to degradation curves [20]. The detailed equations and their implementations (discrete versions) can be found in Appendix A. It should be noted that for the design of a dynamic system (state functions and measurement functions), observability ensures that the internal states (in this study, the estimated parameters of degradation functions) can be accurately estimated from output measurements, which is essential for effective monitoring, control, and prognostics of dynamic systems from the observations. Thus, the arrangement of the vector form of the state and the measurement functions needs to be designed carefully. Various versions of the exponential equations are introduced in Appendix A to address this concern. Two different degradation functions and their corresponding discrete versions implemented in this study are listed in Table 1. In

model **M1**, the term $x_{n,2}$ represents the parameter b of the single exponential equation. The parameter a can be calculated based on the term $x_{n,1}$ and the parameter b. In model **M2**, the term $x_{n,3}$ represents the parameter b of the double exponential equation. The parameters a and c can be calculated based on the terms $x_{n,1}$, $x_{n,2}$, and the parameter b. The parameter d in the model **M2** is fixed. In this study, the parameter d of **M2** is assigned a value of 0.001.

Table 1: The degradation equations

Models	Degradation behavior	State and measurement functions
Single Exponential M1	$HI(n) = ae^{bn}$	$\mathbf{x}_n^{M1} = \begin{bmatrix} x_{n-1,1}e^{x_{n-1,2}n} \\ x_{n,2} = x_{n-1,2} \end{bmatrix}$ $\mathbf{y}_n^{M1} = x_{n,1}$
Double Exponential M2	$HI(n) = ae^{bn} + ce^{dn}$	$\mathbf{x}_n^{M2} = \begin{bmatrix} x_{n-1,1}e^{x_{n-1,3}n} \\ x_{n-1,2}e^{d_{fixed}n} \\ x_{n-1,3} \end{bmatrix}$ $\mathbf{y}_n^{M2} = x_{n,1} + x_{n,2}$

Taking into consideration the previously discussed aspects, the flowchart presented in Fig. 4 summarizes the framework on how to implement the AKKF and estimate the RUL.

Step 0 ***Initialization:*** The parameters are initialized. The Anomaly threshold and the Failure threshold are determined by the health stage information, i.e., $AT = Max(MM(Healthy\ stage) + k_{AT} \times MMAD(Healthy\ stage))$, $FT = k_{FT} \times Min(MM(Healthy\ stage))$, respectively, where MM denotes the Moving Median and $MMAD$ represents the Moving Median Absolute Deviation. The maximum selection for the anomaly threshold eliminates the influence of the outliers. Besides, the initial particle distribution, the initial kernel weight vector and matrix, the initial mapping and the initial residual matrix are also specified.

Step 1 ***Particle Extraction:*** The particles of the parameters of the degradation functions are extracted at the time stamp $n-1$, where the parameters used in the exponential functions are assumed to be Gaus-

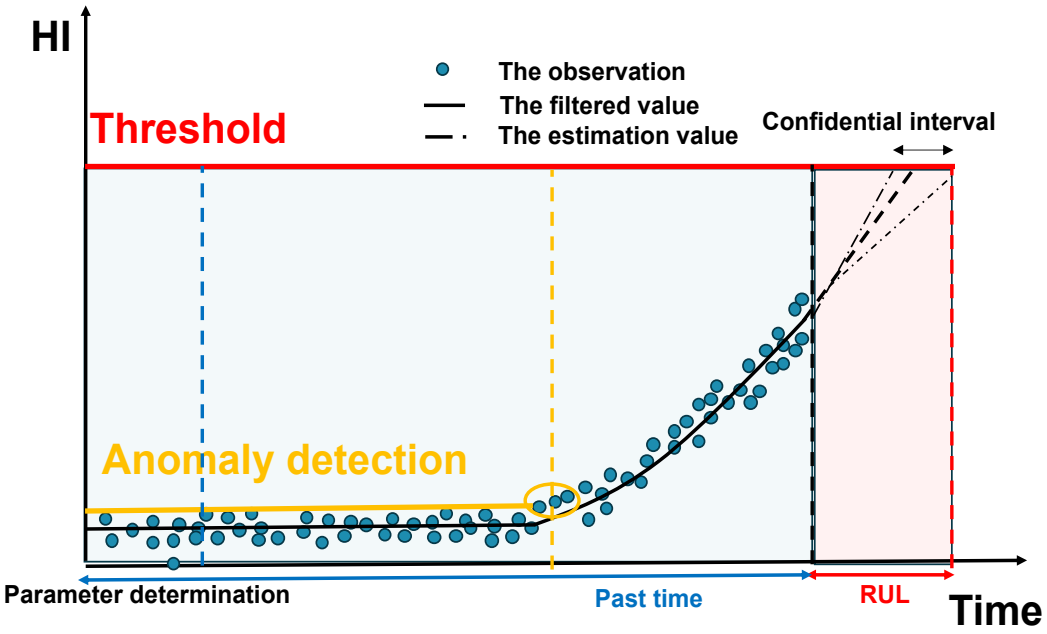


Figure 3: The diagram of the RUL estimation.

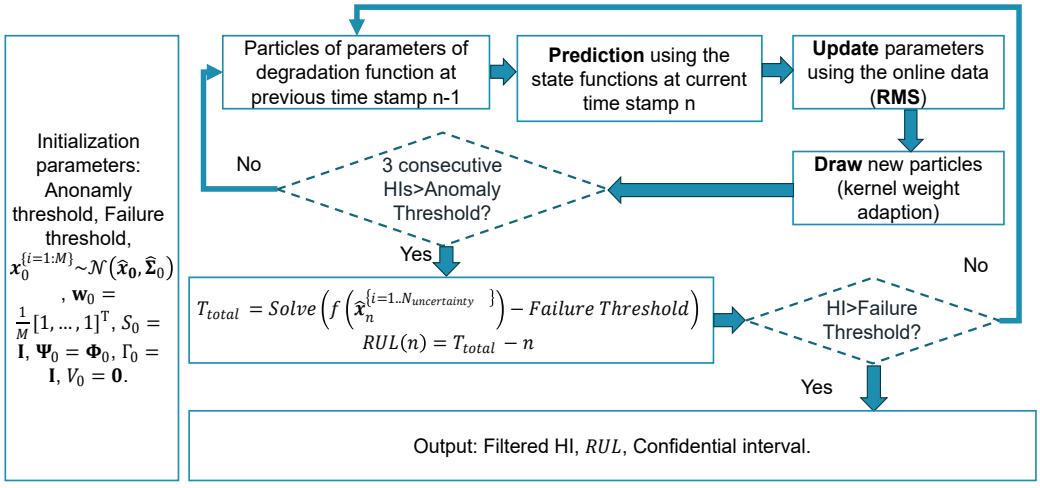


Figure 4: The flowchart of the proposed methodology.

sian. These particles will be delivered into the assumed degradation functions in the next step.

Step 2 ***Parameters Propagation:*** Parameters at current time stamp n are propagated based on Eqs. (18) to (21). The state parameters are calculated in the data space. The kernel weight vector and the corresponding covariance matrix are calculated in the kernel space.

Step 3 ***Data Mapping and Update:*** In this step, the data acquired at the current time stamp is mapped into the kernel space, where the KKR is implemented.

Step 4 ***New State Approximation:*** The new state particles can be approximated using the updated kernel weight and the covariance in this step.

Step 5 ***Anomaly check:*** A check if the observation (i.e., HI) is larger than the anomaly threshold is realised. As shown in Fig. 3, if three consecutive values of the HI are larger than the anomaly threshold, then the RUL estimation starts. If no, the procedure returns to **Step 1**.

Step 6 ***RUL Estimation:*** The total life is estimated based on the state particles drawn from the updated state distribution in the data space. RUL and its confidence interval can be then determined as the current time is known.

Step 7 ***Failure Check:*** A check if the observation is larger than the failure threshold is realized. If no, the process returns to **Step 1**. Otherwise, the filtered HI, the estimated RUL, and the confidence interval are output.

4. Application of the methodology and analysis of results

In this study, three datasets are used for the validation and evaluation of the proposed methodology. The first dataset has been captured at the KU Leuven Gearbox Prognostics Test Rig (KUL-GBPTR), where a bearing incurred damage during an Accelerated Life Test. . The second dataset [12] and the third dataset [30] have been provided by University of Ferrara (UNIFE) and Xi'an Jiaotong University (XJTU). It should be mentioned that RMS, which is a representative and stable indicator, is selected as the

HI for estimation in this study. Further discussion on this choice is provided in the subsequent section.

4.1. KU Leuven Gearbox Prognostics Test Rig dataset

4.1.1. KUL-GBPTR description

The KUL-GBPTR, comprising two 22 kW AC motors is presented in Fig. 5. A bearing (SKF 6205) mounted on the input shaft of the test gearbox sustained damage during an accelerated life test. Two tri-axial accelerometers (acc 2 and 3) were mounted at the input of the gearbox. The sampling frequency was set equal to 102.4 kHz in order to capture high-frequency information as the fault frequencies would modulate the natural frequencies which are generally high. Data were acquired in 10-second intervals every 10 minutes, resulting in a frequency resolution of 0.1 Hz. The speed of the speed motor was 2600 RPM, while the load motor was set to 50 Nm. The signals acquired from accelerometer 3 (Y direction) exhibited larger vibration amplitudes compared to the other signals, as shown in Fig. 6. Consequently, signals captured with the accelerometer 3 (Y direction) are used in this section.

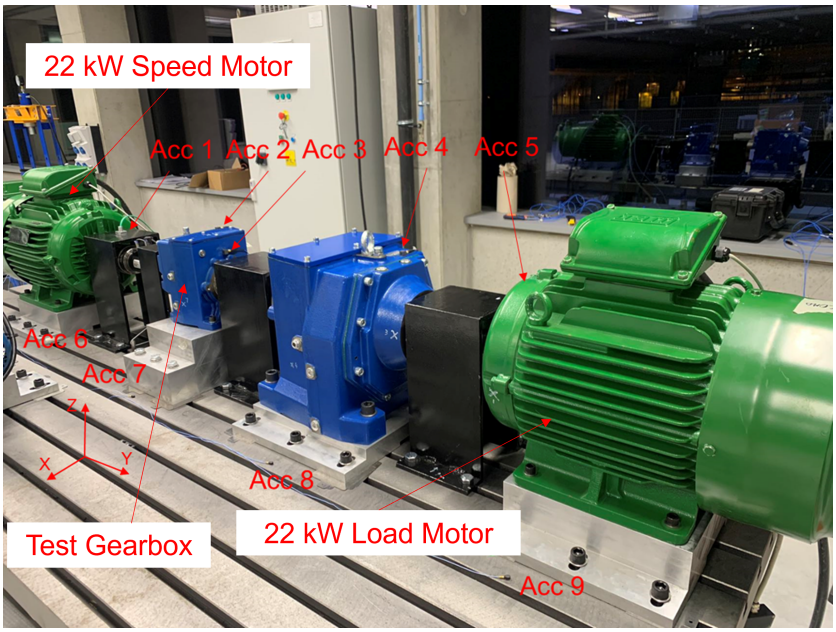
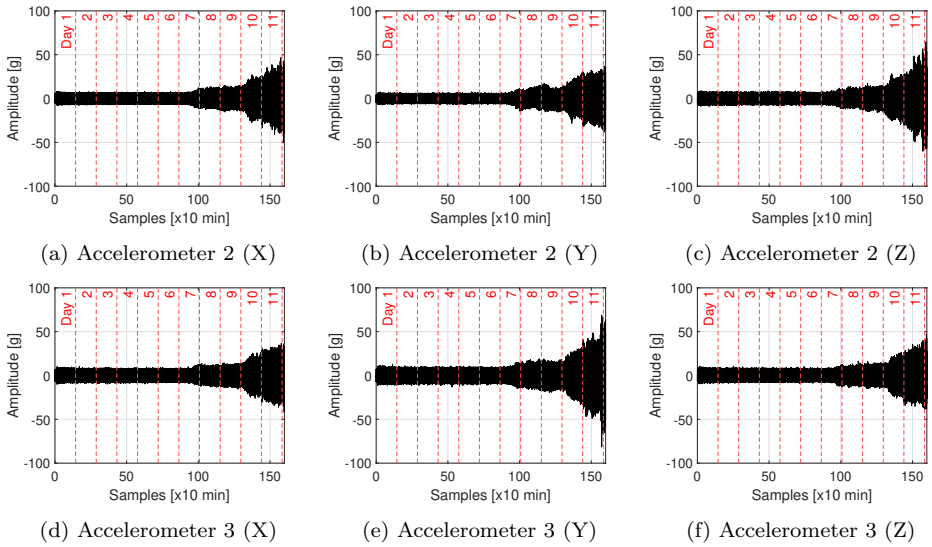


Figure 5: KU Leuven Gearbox Prognostics Test Rig



(a) Accelerometer 2 (X) (b) Accelerometer 2 (Y) (c) Accelerometer 2 (Z)

(d) Accelerometer 3 (X) (e) Accelerometer 3 (Y) (f) Accelerometer 3 (Z)

Figure 6: The raw vibration signals. There are 1592 samples in total, 1 sample was selected for every 10-sample to represent the vibration level during this period.

4.1.2. Ablation studies for key parameters

In this subsection, effects of the key parameters of the proposed method on the performance are evaluated.

Firstly, the number of particles must be determined, taking into account the balance between computational complexity and accuracy. Given that the double exponential model involves more parameters, it is employed for the determination of the number of particles. Table 2 presents the optimal particle number as 500 comprehensively considering the computational time and the accuracy. The data processing has been realised in a 12th Gen Intel(R) Core(TM) i7-12800H 2.40 GHz using MATLAB 2024b. This study employs four metrics, provided in detail in Appendix C. It is important to note that the closer the Cumulative Relative Accuracy (CRA) is to 1, the more accurate the estimator is.

Next, the initial guess $\hat{x}_0$ and $\hat{\Sigma}_0$ values are studied based on the single exponential model. Theoretically, methodologies are expected to be insensitive to the initial values. To evaluate the robustness of the proposed methodology against variations in initial values, different sets of initial values are assigned and analyzed, as shown in Table 3 and Table 4. The RUL results obtained with these varying initial values are reported in Fig. 7. In the subplot (a),

Table 2: The metrics of different particles

Particle number	20	50	100	200	**500**	2000
Matlab Runtime (s)	0.64	2.74	9.92	38.95	324.18	1.15E+04
CRA	-78.82	-8.87	0.03	0.65	**0.76**	0.80
MAE (×10 mins)	9018.45	220.28	70.42	44.56	**39.49**	30.55
RMSE(×10 mins)	16114.72	292.24	80.28	54.15	**53.37**	48.73
Score	0	0.35	0.48	0.70	**0.81**	0.83

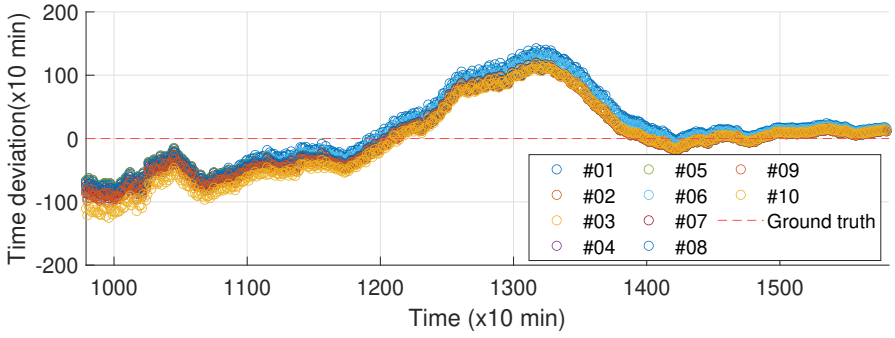
RUL estimation deviations from the actual RUL (ground truth) are calculated and visualized. It is shown that the proposed methodology is generally insensitive to the initial values, as most of the curves are closely aligned or essentially overlap. Additionally, two curves showing the greatest discrepancy are selected, and the results of their differences of the RUL estimations are presented in subplot (b). The maximum bias is 2.63 (x10 mins) at the beginning of the degradation stage. And then the bias gets closer to 0 as time evolves.

Table 3: Cases for different initial values of the hidden state

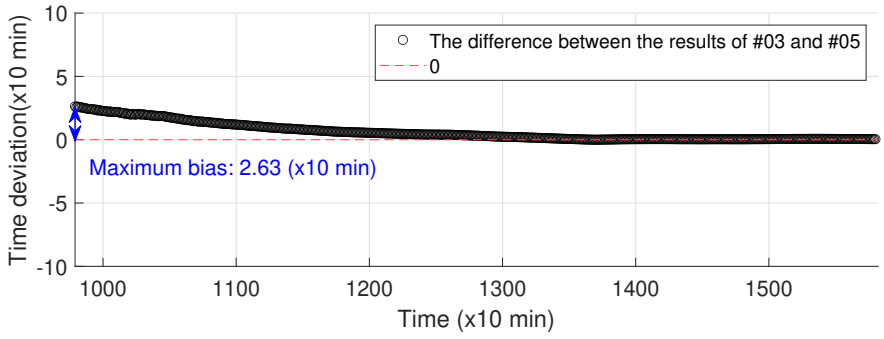
	#01	#02	#03	#04	#05
$\hat{x}_0$	$\begin{bmatrix} HI(1) \\ 1 \end{bmatrix}$	$\begin{bmatrix} HI(1) \\ 1E-1 \end{bmatrix}$	$\begin{bmatrix} HI(1) \\ 1E-2 \end{bmatrix}$	$\begin{bmatrix} HI(1) \\ 1E-3 \end{bmatrix}$	$\begin{bmatrix} HI(1) \\ 1E-5 \end{bmatrix}$
$\hat{\Sigma}_0$	$\begin{bmatrix} 1E-2 & 0 \\ 0 & 1E-6 \end{bmatrix}$				

Table 4: Cases for different initial values of the state covariance

	#06	#07	#08	#09	#10
$\hat{x}_0$	$\begin{bmatrix} HI(1) \\ 1E-2 \end{bmatrix}$				
$\hat{\Sigma}_0$	$\begin{bmatrix} 1E-2 & 0 \\ 0 & 1E-2 \end{bmatrix}$	$\begin{bmatrix} 1E-2 & 0 \\ 0 & 1E-4 \end{bmatrix}$	$\begin{bmatrix} 1E-2 & 0 \\ 0 & 1E-6 \end{bmatrix}$	$\begin{bmatrix} 1E-4 & 0 \\ 0 & 1E-8 \end{bmatrix}$	$\begin{bmatrix} 1E-6 & 0 \\ 0 & 1E-1 \end{bmatrix}$



(a)



(b)

Figure 7: . The influence of different initial values to the results. (a) The RUL results deviation from the ground truth for different initial values (b) The RUL results difference between #03 and #05.

Finally, the Gaussian kernel bandwidth is studied as this is the most important parameter for the kernel methods. Since the parameters of the state functions and observations must be mapped into the kernel space, a thorough examination of these two pivotal bandwidth values, σ_x and σ_y for the states and observations respectively, is conducted. The results are summarized in Table 5. The bandwidth of the kernel influences the degree of smoothing applied in the methodology. A larger bandwidth results in more pronounced smoothing, leading to greater loss of detail but more effective noise reduction. Table 5 indicates that the model keeps less details and removes more noise for the observations in the kernel space. This is expected since the RUL calculation relies on the state parameters estimation, which are thereby assigned smaller bandwidth. Similarly with the initial values analysis, Fig. 8 reports the results of different kernel width. It demonstrates that different kernel bandwidths indeed affect the results. However, similar trends are observable across various curves, particularly in the later stages of degradation. Around 1400 (x10 minutes), all curves essentially converge and overlap. In this case, #15 was selected as the kernel width. Nevertheless, it is worth noting that further investigation into the automatic selection of kernel bandwidth could be beneficial for enhancing accuracy.

Table 5: Cases for different kernel width

	#11	#12	#13	#14	#15	#16	#17	#18	#19
σ_x	2	5	10		10		2	5	20
σ_y		1E2		10	1E3	1E4		1E3	

4.1.3. RUL estimation and discussion of results

Following the procedure outlined in Fig. 4, new values of the parameters are estimated based on the updated observations over time. Subsequently, the RUL is calculated using these new values of the parameters. Fig. 9 shows the dynamic process of the RUL estimation. The filtered values appear smoother from the HI perspective, as shown in Fig. 9(a), (b) and (c). On the RUL side, the uncertainty decreases as the bearing degrades. At 1200 ($\times$10 mins), the standard deviation of the estimated RUL is 21.84 ($\times$10 mins). Over time, the standard deviation becomes 13.02 ($\times$10 mins) at 1400 ($\times$10 mins). Finally, the standard deviation decreases to 4.48 ($\times$10

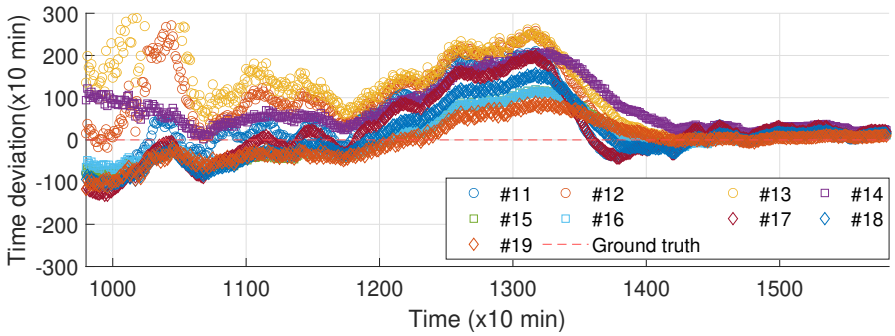
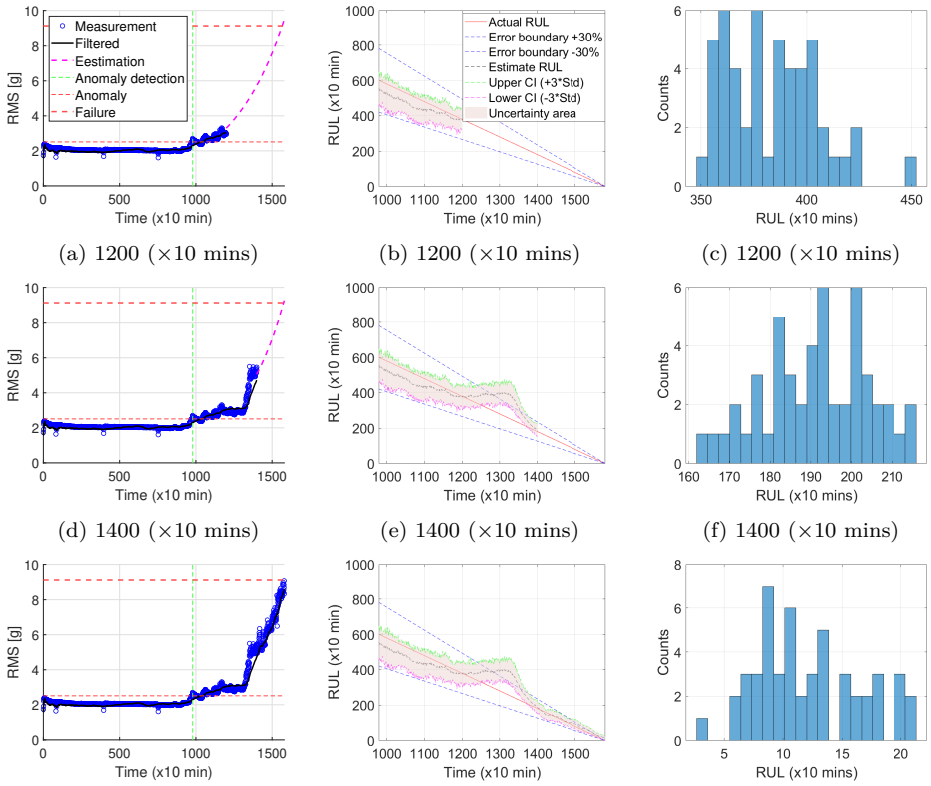


Figure 8: The RUL results deviation from the ground truth for different initial kernel bandwidth.

mins) at the end of life. It should be noted that the histograms of the RUL show that the samples of the RUL at different times do not fit the Gaussian distribution anymore. Fig. 10 demonstrates the evolution of the estimated parameters of the model **M2** as the process advances. The parameter a for model **M2** gradually decreases as time evolves and stabilizes at a certain level after about 1400 ($\times$10 mins). The parameter b exhibits fluctuation at the initial stage from 0 to approximately 400 ($\times$10 mins), while the c is also unstable at the beginning. These fluctuations are attributed to the initial parameters guess. Then, these 2 parameters, b and c, gradually decrease and increase, respectively, until about 1400 ($\times$10 mins). Eventually, these 2 parameters get relatively stable after 1400 ($\times$10 mins), which is similar with the parameter a. This trend indicates that as the bearing approaches the EoL, the parameters become more stable, leading to stabilized estimated RUL, which is an inherent feature of the Kalman Filter.

Fig. 11 presents a comparison of the estimated RUL for different estimators and degradation models. For the same estimator, they have similar RUL estimations. For the same degradation models, the AKKF shows superior results as the estimated RUL is closer to the actual life. In general, all estimators approximately converge to 0. However, the AKKF gains better accuracy at the initial fault stage (approximately from 979 ($\times$ 10 mins) to 1400 ($\times$ 10 mins)). It should be noted that even if it is assumed that the parameters' distributions of the exponential function are Gaussians, the RUL's distributions are not. That is why the Monte-Carlo method is used to approximate the RUL distribution. Table 6 shows the comparison of the estimated RUL using different estimators. It indicates that the double exponential model (**M2**)

combined with the AKKF has a slightly better performance. It should be noted that from CRA, MAE and RMSE side, the difference between AKKF (M1) and AKKF (M2) is minor. But from the perspective of the score, the AKKF (M2) has better performance in terms of reliability as the score represents the percentage of the estimations that fall within a reliable area. Detailed discussion will be in the next subsection as there similar results can be observed.



(a) 1200 (×10 mins) (b) 1200 (×10 mins) (c) 1200 (×10 mins)

(d) 1400 (×10 mins) (e) 1400 (×10 mins) (f) 1400 (×10 mins)

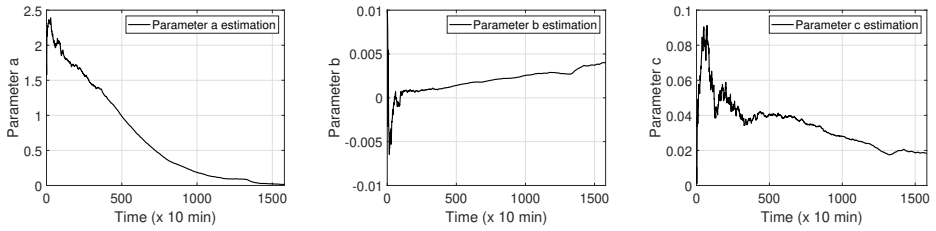
(g) End of Life (×10 mins) (h) End of Life (×10 mins) (i) End of Life (×10 mins)

Figure 9: The dynamic process of RUL estimation. (a), (d) and (g) show the filtered HI for the past time and estimation of the HI at the current time; (b), (e) and (h) are the RUL estimations; (c), (f) and (i) are the histograms of the RUL at the current time.

4.2. University of Ferrara dataset

4.2.1. University of Ferrara test rig description

The experimental test rig, as shown in Fig. 12, consists of an electric motor, an inverter to manage the motor's operation, a lever system to provide



(a) The estimated parameter a (b) The estimated parameter b (c) The estimated parameter c

Figure 10: The estimated parameters of model **M2**

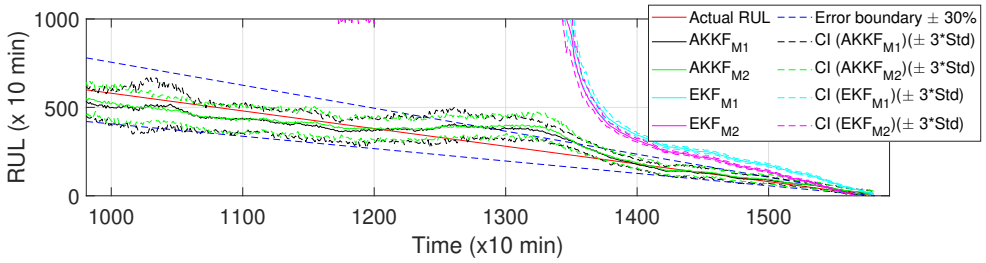


Figure 11: The comparison of the estimated RUL of the bearing of KUL-GBPTR

Table 6: The metrics comparison of the estimated RUL

	AKKF (M1)	AKKF (M2)	EKF (M1)	EKF (M2)
CRA	**0.78**	0.76	-1.91	-1.60
MAE (×10 mins)	39.66	**39.49**	928.23	844.49
RMSE(×10 mins)	**51.53**	53.37	1246.08	1186.09
Score	0.72	**0.81**	0.00	0.02

load, two roller support bearings, and a test bearing [12]. The three bearings are mounted on a shaft connected to the motor through a flexible coupling. The operating conditions are listed in Table 7. Additionally, the acquisition duration was 5 seconds, where the sampling interval was 5 minutes. More details can be found in [12].

Table 7: Description of the operating conditions of University of Ferrara dataset

Condition (C)	Load (kN)	Speed (rpm)	Bearing code
C1	5	2400	CB11
C2	4.7	2400	CB21
C3	4	2400	CB31, CB32, CB33
C4	3	2400	CB41

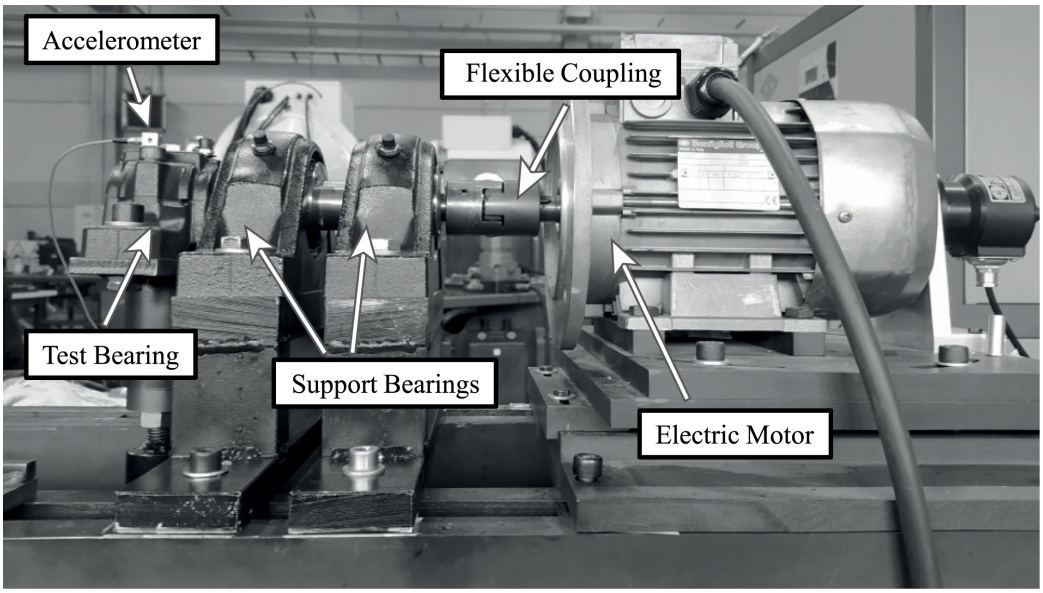
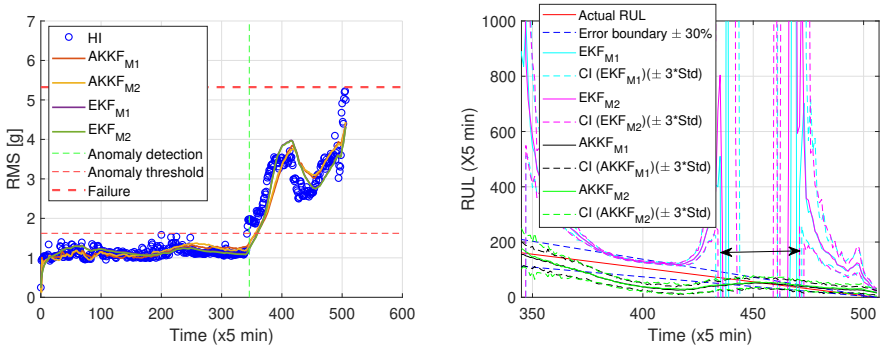


Figure 12: The test rig of University of Ferrara

4.2.2. RUL estimation and discussion of results

The RUL estimation results for CB11 are illustrated in Fig. 13. The anomaly detection method accurately avoided false alarms around the 200 (x5 mins) time stamp, as demonstrated in Fig. 13(a), highlighting its robustness

and precision in differentiating normal fluctuations from actual anomalies. From RUL estimation perspective, it is clear that around 420-480 (x5 mins), the EKF estimator loses its functionality, as indicated by the double black arrow in Fig. 13(b), due to the fluctuations of the HI at this stage. However, both the EKF-based and AKKF-based estimators converged closer to the ground truth in the final stages. Furthermore, Fig. 14 reports how each model's predictions fluctuate near the actual RUL. Compared with EKF-based methods, the proposed methodology demonstrates better robustness and accuracy, as the estimated RUL consistently maintains smaller deviations from the actual RUL and exhibit less fluctuation over time.



(a) (b)

Figure 13: The filtered HI and RUL comparison for CB11 of Ferrara dataset: (a) filtered HI and (b) RUL estimation.

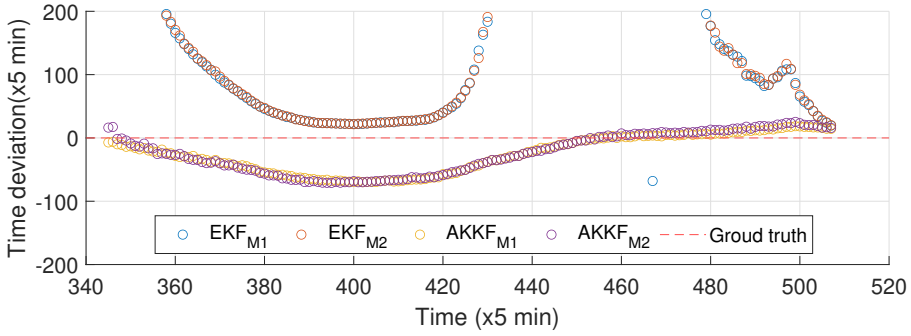


Figure 14: The time deviation from the estimated RUL to the actual RUL for the RUL estimation of the CB11 of Ferrara dataset.

For the remaining 5 bearings, their RUL estimations are presented in Fig. 15. It is evident that all of the models converge to the similar points (close to 0 at the late stages). However, EKF-based model shows big fluctuation at certain stage for different bearings (e.g., from around 1965 to 1975 (x5 mins), marked within the double black arrow of the CB32, as shown in subplot (c)). The proposed methodology provides relatively accurate and stable estimation according to the results. Furthermore, similarly with the previous subsection, different metrics are employed to evaluate the results, as reported in Table 7. It should be noted that from the perspective of the score, the **M2** model may shows slightly better results. This is why different metrics are employed in this study. The score indicates the number of points that fall within the error band. As previously discussed, it suggests the model is more reliable as the estimated RUL may always fluctuate close to the ground truth within a certain area. Overall, AKKF-based with the **M1** model achieves the highest performance in this dataset.

Table 8: The metrics comparison of the estimated RUL by the AKKF-based and EKF-based methodology for Ferrara dataset

	AKKF (M1)				AKKF (M2)				EKF (M1)				EKF (M2)			
	CRA	MAE	RMSE	Score	CRA	MAE	RMSE	Score	CRA	MAE	RMSE	Score	CRA	MAE	RMSE	Score
CB11	0.30	30.80	38.96	0.30	0.18	32.61	40.09	0.37	-5.81	308.29	693.07	0.19	-17.80	5502.46	730.49	0.19
CB21	0.31	7.50	9.20	0.17	0.41	5.37	6.54	0.54	-0.95	16.48	23.25	0.20	-0.92	16.81	24.14	0.17
CB31	0.56	5.64	9.78	0.49	0.21	15.18	17.59	0.00	-0.27	23.32	34.30	0.00	-20.62	225.57	238.04	0.00
CB32	0.03	16.48	19.51	0.27	0.02	17.22	20.95	0.37	-3.39	79.95	151.31	0.05	-10.56	184.42	797.47	0.03
CB33	0.61	15.43	18.23	0.10	0.63	11.33	13.09	0.42	-1.67	145.90	410.66	0.00	-1.26	1.16E+02	243.90	0.00
CB41	0.45	19.53	24.63	0.16	0.40	20.88	25.78	0.21	-2.83	68.65	136.37	0.18	-2.41	60.35	100.31	0.18
Average	**0.38**	**15.90**	**20.05**	0.25	0.31	17.10	20.67	**0.32**	-2.49	107.10	241.49	0.10	-8.93	1017.61	355.73	0.09

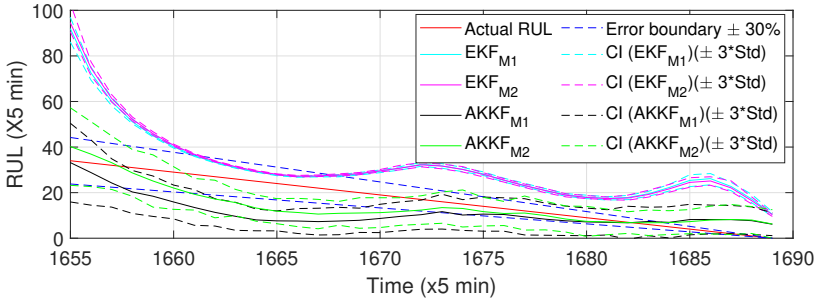
4.3. Xi'an Jiaotong University (XJTU) dataset

4.3.1. Xi'an Jiaotong University (XJTU) test rig description

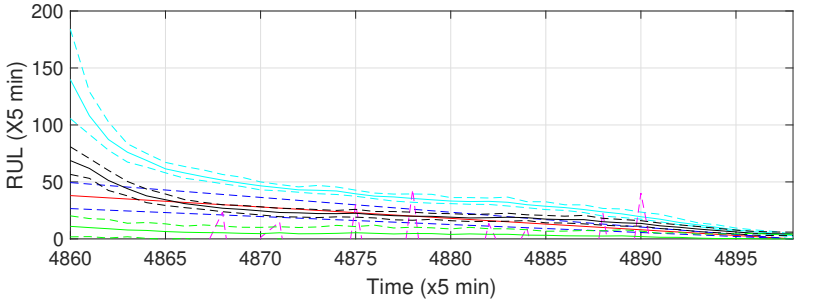
The XJTU dataset [30] is also used to validate the proposed methodology. The test rig of XJTU has 2 supporting bearings and 1 test bearing. The test bearing was loaded by a hydraulic actuator, as shown in Fig. 16. The operation conditions of the XJTU dataset are listed in Table 9. The sampling frequency of the vibration data captured by the acquisition system was set at 25.6 kHz. The acquisition duration was 1.28 seconds, where the sampling interval was 1 minute. More details can be found in [30].

4.3.2. RUL estimation and discussion of results

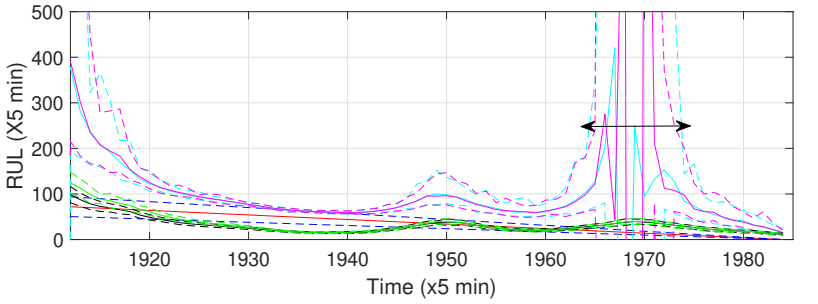
A comprehensive comparison of results obtained by the AKKF-based methodology and the EKF-based methodology is conducted, as presented


Actual RUL
EKF$_{M1}$
EKF$_{M2}$
AKKF$_{M1}$
AKKF$_{M2}$
Error boundary ± 30%
CI (EKF$_{M1}$)(± 3*Std)
CI (EKF$_{M2}$)(± 3*Std)
CI (AKKF$_{M1}$)(± 3*Std)
CI (AKKF$_{M2}$)(± 3*Std)
RUL (X5 min)
Time (x5 min)

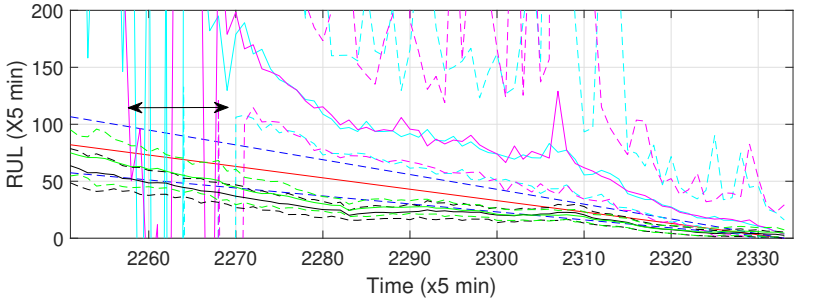
(a)


RUL (X5 min)
Time (x5 min)

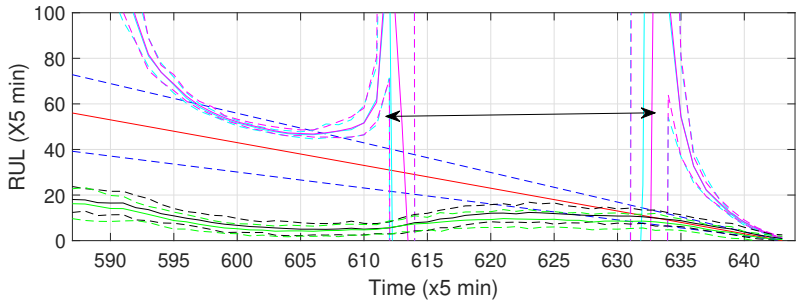
(b)


RUL (X5 min)
Time (x5 min)

(c)


RUL (X5 min)
Time (x5 min)

(d)



(e)

Figure 15: The RUL comparison for different models of Ferrara dataset. (a) CB21, (b) CB31, (c) CB32, (d) CB33 and (e) CB41.

in Table 10. Generally, the AKKF outperforms the EKF when employing the same model. For the same filter, the results are more complicated. Model **M2** shows better performance combined with EKF. While the results of model **M1** closely resemble those of model **M2** for the AKKF, it shows a marginally better performance. Moreover, it is notable that the results of all estimators are relatively disappointing for CB 4, 6, 11, 13, 14. Consequently, a more in-depth investigation into these scenarios is conducted in the subsequent section.

The results of CB 3, CB 5, CB 7, CB 12 and CB 15 are shown in Fig. 17. For CB 3 (Fig. 17(a) and (b)), an anomaly is detected at 60 ($\times$1 min). The fault gradually increases after the anomaly time, so the estimators get suitable results. Fig. 17(b) shows that at the EoL, the results from all estimators converge approximately to 0, while the AKKF performs better at the initial fault stage, which is similar to the bearing of the KUL-GBPTR. It can be found that at approximately 130 ($\times$1 min) and 145 ($\times$1 min), the HI observations decrease and increase again. Correspondingly, the RUL estimated by all estimators also fluctuate at these time stamps. It can be clearly seen that the AKKF is more stable compared with the EKF which exhibits peaks when there is a small fluctuation. Fig. 17(c) and (d) are the estimations for CB 5. For all the estimators, the estimated RUL converge to 0 at the end of life, but the AKKF performs better at the initial estimation stage. For CB 7, as shown in Fig. 17(e) and (f), the AKKF is more stable for the RUL estimation. However, it should be mentioned that at the EoL, model **M2** using EKF performs better than others, as it closely approaches 0. The results for CB 10 (Fig. 17(g) and (h)) also indicates that the AKKF is more stable, as

high peaks can be found when the HI observations suddenly change in the EKF-based results. The EKF encounters challenges with the data obtained from CB 12. As shown in Fig. 17(i) and (j), the EKF loses efficacy when the HI is relatively noisy, while the AKKF gains more satisfactory results. It should be highlighted that even CB 12 presents high-noise, but still, the EKF and AKKF successfully smooth the HI curve, which might be used for the health stage determination based on this feature.

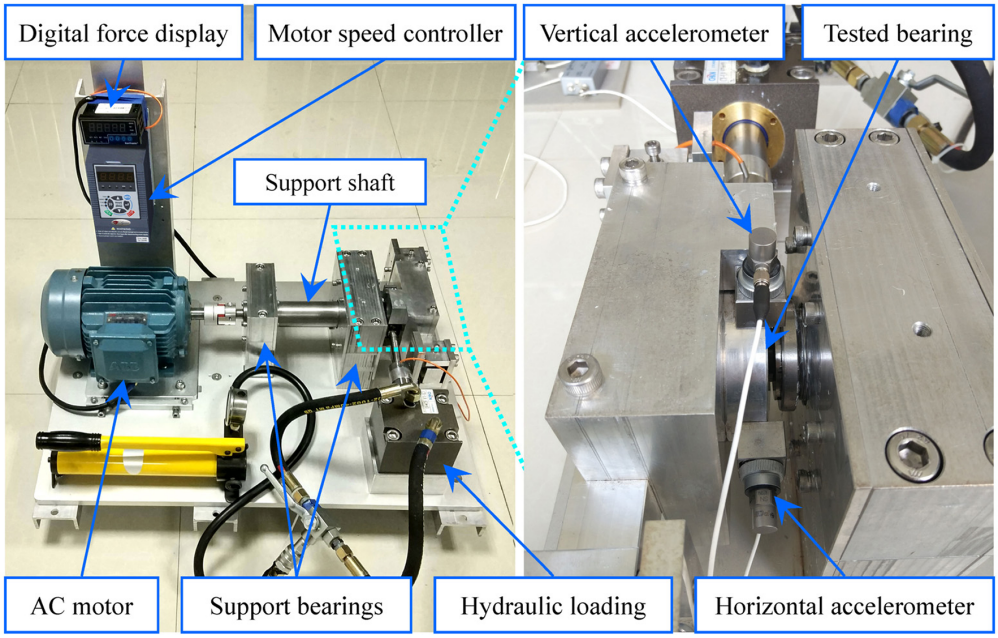


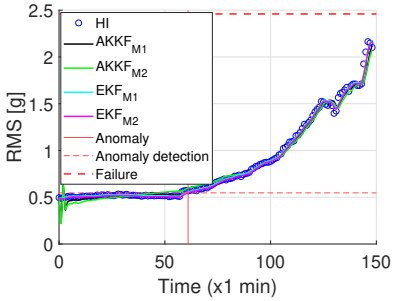
Figure 16: The XJTU test rig [30]

Table 9: The operation conditions of XJTU dataset

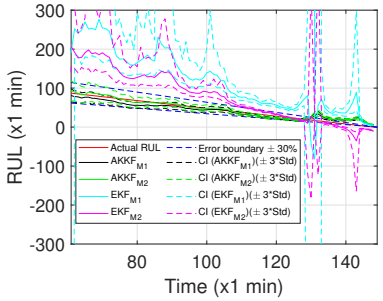
Conditions	Load (kN)	Speed (rpm)	Bearing code
C1	12	2100	CB1-CB5
C2	11	2250	CB6-CB10
C3	10	2400	CB11-CB15

4.4. Summary

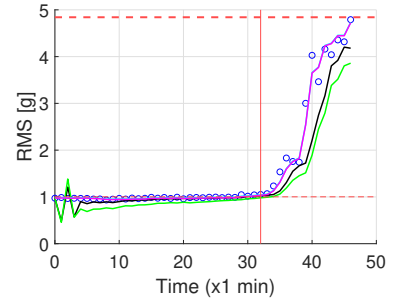
This subsection recapitulates the key outcomes of Section 4 as follows:


HI
AKKF$_{M1}$
AKKF$_{M2}$
EKF$_{M1}$
EKF$_{M2}$
Anomaly
Anomaly detection
Failure
RMS [g]
Time (x1 min)

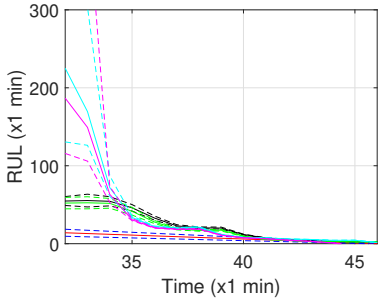
(a)


Actual RUL
AKKF$_{M1}$
AKKF$_{M2}$
EKF$_{M1}$
EKF$_{M2}$
Error boundary ± 30%
CI (AKKF$_{M1}$)(± 3*Std)
CI (AKKF$_{M2}$)(± 3*Std)
CI (EKF$_{M1}$)(± 3*Std)
CI (EKF$_{M2}$)(± 3*Std)
RUL (x1 min)
Time (x1 min)

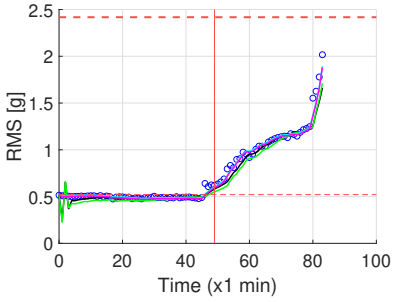
(b)


RMS [g]
Time (x1 min)

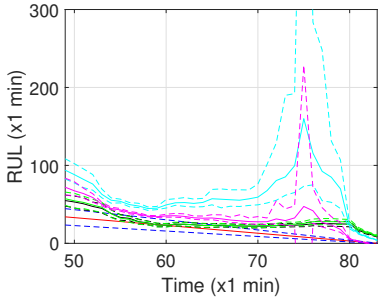
(c)


RUL (x1 min)
Time (x1 min)

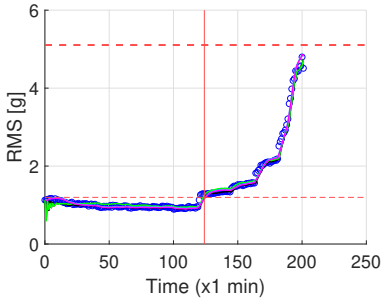
(d)


RMS [g]
Time (x1 min)

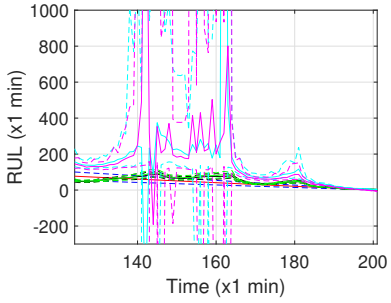
(e)


RUL (x1 min)
Time (x1 min)

(f)


RMS [g]
Time (x1 min)

(g)


RUL (x1 min)
Time (x1 min)

(h)

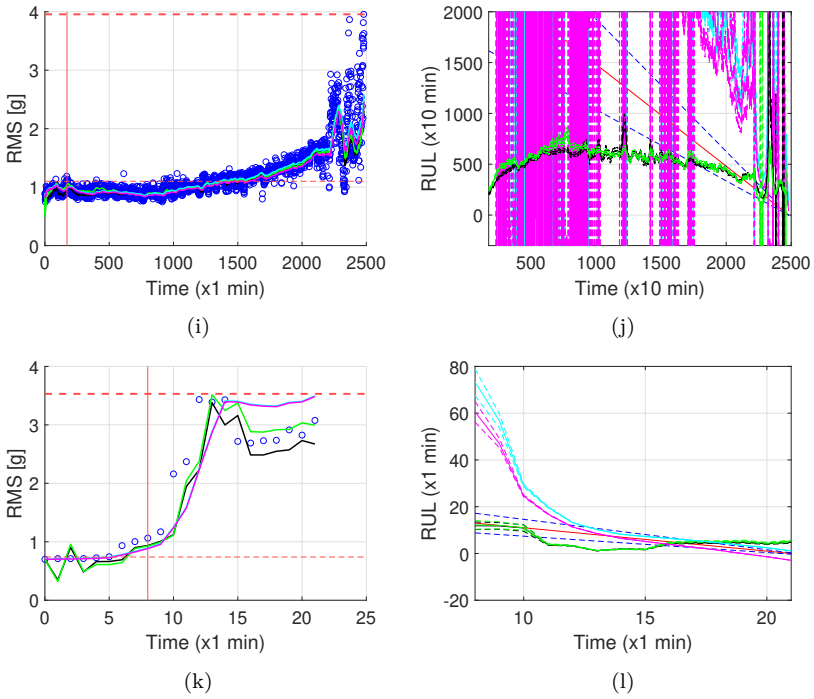


Figure 17: The filtered HI and RUL comparison for XJTU dataset. (a) and (b) for CB 3; (c) and (d) for CB 5; (e) and (f) for CB 7; (g) and (h) for CB 10; (i) and (j) for CB 12; (k) and (l) for CB 15. The unit is $\times 1$ min as the acquisition interval is 1 min.

Table 10: The metrics comparison of the estimated RUL by the AKKF-based and EKF-based methodology for XJTU dataset

	AKKF (M1)				AKKF (M2)				EKF (M1)				EKF (M2)			
	CRA	MAE	RMSE	Score	CRA	MAE	RMSE	Score	CRA	MAE	RMSE	Score	CRA	MAE	RMSE	Score
CB1	-0.30	15.17	22.50	0.38	-0.51	17.57	25.86	0.38	-4.47	72.27	130.51	0.00	-2.52	40.79	66.61	0.19
CB2	-1.44	23.41	24.99	0.00	-1.41	22.53	23.93	0.00	-2.82	51.07	60.48	0.00	-1.19	34.16	41.87	0.00
CB3	0.37	6.89	7.87	0.81	0.35	5.17	6.80	0.80	-1.38	80.21	96.75	0.00	-0.11	41.35	55.28	0.19
CB4	-12.04	71.51	75.52	0.00	-12.93	80.40	83.90	0.00	-122.69	993.88	1879.15	0.00	-75.94	241.43	410.61	0.03
CB5	-0.66	17.34	23.95	0.07	-0.40	15.12	21.73	0.20	-2.75	48.06	94.20	0.00	-2.03	39.76	77.58	0.27
Average	**-2.82**	**26.87**	**30.96**	0.25	-2.98	28.16	32.44	**0.27**	-26.82	249.10	452.22	0.00	-16.36	79.50	130.39	0.14
CB6	-114.96	426.49	438.57	0.00	-115.19	428.99	439.07	0.00	-117.60	456.00	456.00	0.00	-117.60	456.00	456.00	0.00
CB7	-0.86	10.98	13.02	0.31	-1.06	11.87	14.32	0.31	-4.32	46.43	55.35	0.00	-0.47	17.61	20.49	0.03
CB8	-1.45	60.27	93.64	0.39	-1.98	66.54	95.09	0.30	-6.32	620.16	1177.07	0.04	-3.14	263.41	464.51	0.13
CB9	-8.18	28.48	31.34	0.00	-8.67	29.39	32.51	0.00	-393.69	1055.22	3651.86	0.00	-705.61	610.91	1987.89	0.00
CB10	0.52	15.21	19.03	0.46	0.47	17.29	21.87	0.41	-3.23	170.16	626.40	0.00	-1.96	136.03	319.00	0.05
Average	**-24.99**	**108.29**	**119.12**	**0.23**	-25.29	110.81	120.57	0.21	-105.03	469.59	1193.34	0.01	-165.75	296.79	649.58	0.04
CB11	-88.86	1174.67	1514.15	0.00	-96.45	1573.89	1851.46	0.00	-101.75	2073.44	2188.38	0.00	-102.19	2121.99	2264.18	0.00
CB12	-8.38	748.42	972.23	0.25	-7.12	703.54	911.31	0.23	-11.19	8035.34	15491.13	0.00	-6.51	5578.58	9546.35	0.01
CB13	-584.03	571.03	571.20	0.00	-655.40	617.20	618.44	0.00	-11257.02	7826.16	8545.55	0.00	-4141.33	4061.08	4061.89	0.00
CB14	-65.93	722.30	742.12	0.00	-47.21	601.23	645.92	0.00	-156.37	1080.11	1196.89	0.00	-157.09	1079.65	1200.33	0.00
CB15	0.21	3.11	3.82	0.43	0.13	3.12	3.84	0.43	-0.18	15.46	29.11	0.29	-0.08	12.22	23.01	0.36
Average	**-149.40**	**643.91**	**760.70**	**0.14**	-161.21	699.79	806.20	0.13	-2305.30	3806.10	5490.21	0.06	-881.44	2570.70	3419.15	0.07

(1) Three datasets were used to validate the proposed methodology, demonstrating its universality.

(2) The hyper-parameters of the proposed methodology were thoroughly analyzed using the KUL-GBPRTR dataset.

(3) Overall, the combination of the single exponential model with the AKKF consistently delivered superior performance, optimally balancing computational efficiency and accuracy.

5. Discussion

There are three main considerations in the process of the RUL estimation: the selection of the HI, the establishment of the degradation models (equations describing the degradation behaviour of bearings), and the selection of the filter.

From the HI perspective, various indicators in the time and frequency domains can be used. In the case of the bearing in the KUL-GBPTR dataset, several state-of-the-art methodologies were employed to extract HIs. As illustrated in Fig. 18(a), the RMS exhibits the best performance for the RUL estimation as it demonstrates a clear trend, making it suitable for prognostics within these selected HIs. Comparing Fig. 18(b) with (d), the sum of Enhanced Envelope Spectra (EES) [31] (Fig. 18(b)) at different characteristic frequencies (CFs) displays better trendability. However, the fluctuation

gradually increases as the fault develops. Conversely, the sparsity (L2/L1 norm near the CFs, Fig. 18(d)) enables easy identification of the fault categories. It can be observed that the amplitude increases at the BPFI, while the curves of the amplitude at other CFs remain stable throughout the bearing's lifespan, indicating an inner race fault for the bearing of KUL-GBPTR. Fig. 18(c) demonstrates that the Gini Index of EES can be very useful for clustering the health stage of the bearing, revealing three distinct stages for this bearing. Additionally, Fig. 18(e) and (f) show the Gini Index of the SE and the quasi-arithmetic mean (QAM) [32] are sensitive to the anomalies, where those peaks are the restarts of the test rig. The detailed formulas are demonstrated in Appendix B. It should be mentioned here, again, the HIs are dependent on the machine where the REBs are installed and many other factors. Thus, for different test rigs, the optimal HIs might be different.

The establishment of the degradation models primarily focuses on the accuracy of equations capable of describing the degradation trend. Ensuring that the designed system is observable is crucial. If the system is not observable, the hidden state might not be uniquely determined, and the filter could become unstable. Additionally, based on the results obtained in all of the cases studied above, the single exponential model gains better RUL estimation compared to the double exponential model in combination with the AKKF. Although having more parameters can offer greater flexibility in modeling the degradation curve, it also increases complexity and sensitivity to noise.

In the process of filter selection, the results suggest that the AKKF performs better, especially when combined with the KME and KKR. On the other hand, the EKF linearizes the state and measurement functions using the Jacobian matrix, which can lead to inaccurate estimation when the model is highly nonlinear. Therefore, in RUL estimation, the choice of filter can be critical, particularly when dealing with models that exhibit high nonlinearity.

As mentioned in the last section, some bearings in the XJTU dataset yielded unsatisfactory results. This section explores the issue using CB 4 and CB 13, while other bearings share similar problems. As shown in Fig. 19 (a), the anomaly threshold is set at 0.4894 g, with the corresponding anomaly time being 89 ($\times$1 min). The failure threshold is 1.8779 g at 120 ($\times$1 min). For CB 13, the anomaly threshold is set at 0.6646 g and the corresponding anomaly time is 343 ($\times$1 min). Its stop threshold is 2.7076 g at 345 ($\times$1 min). The problem is that the fault develops too quickly for the model to capture it accurately. This issue relates to the thresholds set for both measurement and

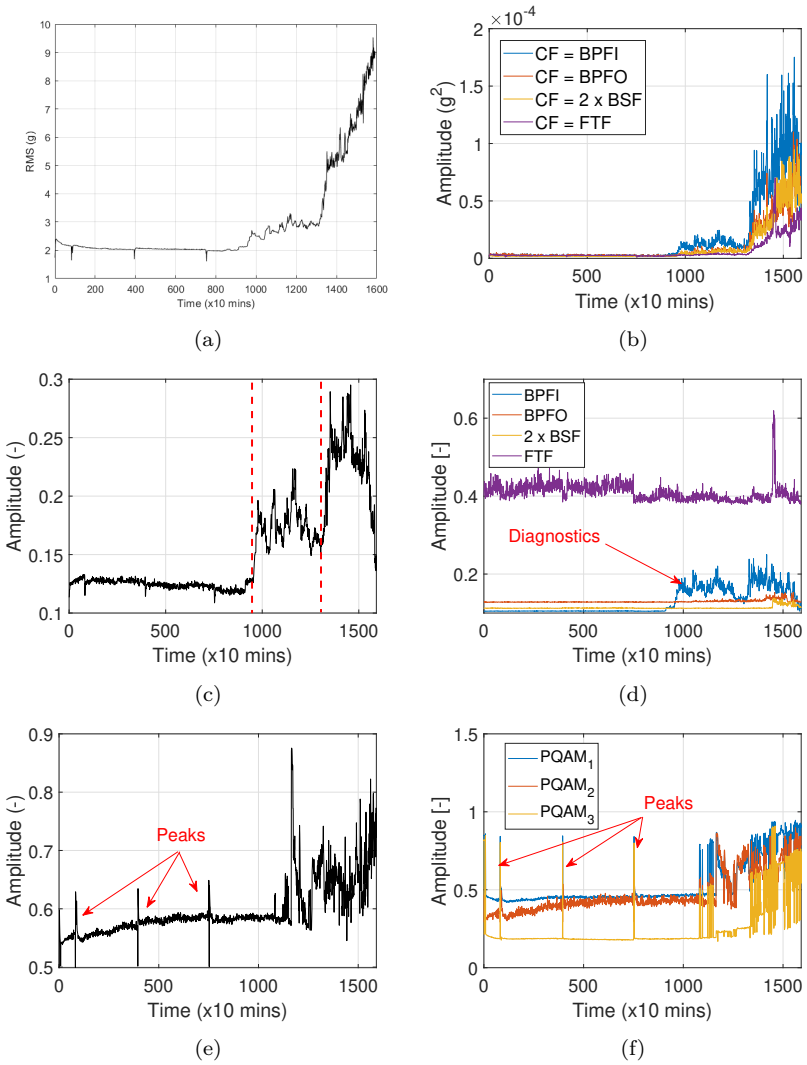
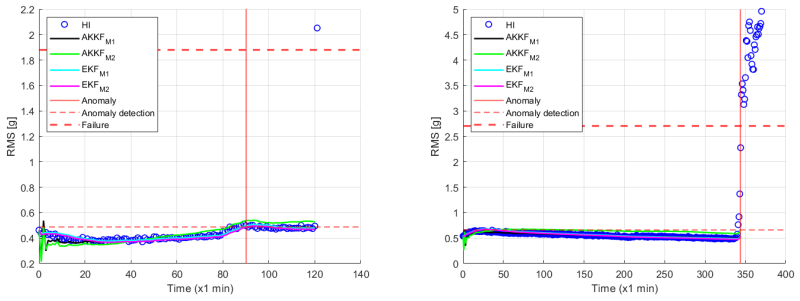


Figure 18: HIs. (a) Root Mean Square (RMS) in the time domain, (b) The sum of EES at 1 x CF, 2 x CF and 3 x CF (Characteristic Frequency). (c) Gini Index of the EES, (d) The L2 / L1 norm of the EES at CF * 0.98-CF*1.02, (e) Gini Index of the Squared Envelope (SE) of the signal which is filtered by the selected optimal band using the spectral kurtosis, (f) The General power function-based quasi-arithmetic mean (QAM) of SE.

estimation, as well as the estimator's requirement for sufficient data to accurately capture the degradation trend. One potential solution for this problem is to divide the bearing degradation curve into different stages, allowing the model to use different models at different stages instead of capturing transient changes in the bearing. In other words, different degradation models might better fit the curve at different stages. However, it's important to note that clustering should be performed online to be useful for prognostics.



(a) (b)

Figure 19: The filtered HI and the observation for CB 4 and 13, (a) CB 4 and (b) CB 13.

Additionally, some assumptions made in this study aim to simplify the estimation process, yet they may prompt further questions that warrant deeper investigation. For example, assuming that the RUL follows a linear relationship with time could be debatable. From a material perspective, if a bearing is in a healthy state, the actual RUL should theoretically be infinite. Conversely, as the bearing defect worsens, the actual RUL should decrease at an accelerating rate.

6. Conclusions

In this study, a methodology for estimating the RUL of REBs based on the AKKF has been introduced, with potential applicability to a broader range of rotating machinery. Three datasets have been analyzed, and an in-depth discussion of the estimation results has been conducted. The conclusions drawn from the aforementioned results and discussions are as follows:

1. **Proposed Methodology**: A methodology has been proposed for estimating the RUL of bearings based on the AKKF. This methodology involves the formulation of the state function, the measurement function

and the observation, enabling it to be employed for online estimation, which enhances its practicality for industry applications.

2. **Applications**: The methododology was aplied, validated and evaluated on three datasets: the KUL-GBPTR dataset, the publicly available University of Ferrara dataset and the publicly available XJTU dataset. Comparative analysis revealed that the AKKF outperforms the EKF, particularly benefiting from the incorporation of KME and KKR, thus enhancing the filter's robustness and accuracy for addressing nonlinear issues.
3. **Degradation models**: Two types of exponential equations were used in different filters. The single exponential model combined with the AKKF generally exhibited the best performance. The Monte Carlo method was employed to sample the estimation parameters, thereby providing quantified uncertainty and enhancing the reliability of the predictions.
4. **Health indicators**: The selection of HIs is elucidated, considering that different HIs possess varying capabilities in anomaly detection, diagnostics, and prognostics. Both estimators necessitate a gradual degradation rather than a sudden fault to avoid yielding unsatisfactory results, where the sudden fault prognostics should be discussed further.

Acknowledgments

The authors acknowledge the financial support of the Chinese Scholarship Council and the KU Leuven Internal Funds.

Appendix A. Implementation of equations

To discretize the equations, Let's start from the single exponential model $M1$:

$$\mathbf{x}_n = a_n e^{b_n n} + \mathbf{u}_n \tag{A.1}$$

$$\mathbf{x}_{n-1} = a_{n-1} e^{b_{n-1}(n-1)} + \mathbf{u}_{n-1} \tag{A.2}$$

where $a_{n-1} = a_n = a, b_{n-1} = b_n = b, \mathbf{u}_{n-1} = \mathbf{u}_n = \mathbf{u}$. Then the following equation can be obtained:

$$\mathbf{x}_n = a e^{bn} + \mathbf{u}_n = a e^{b(n-1)} e^b + \mathbf{u} = \mathbf{x}_{n-1} e^b + \mathbf{u} \tag{A.3}$$

Thus, the above equation can be written as a vector form to implement in the computer. For the state function:

$$\mathbf{x}_n^{ver1,M1} = \begin{bmatrix} x_{n-1,1}e^{x_{n-1,2}} \\ x_{n-1,2} \end{bmatrix} + \mathbf{u} \tag{A.4}$$

For the measurement function:

$$\mathbf{y}_n^{ver1,M1} = x_{n,1} + \mathbf{v} \tag{A.5}$$

In the EKF, the equations need to be linearized, where the Jacobi matrix can be implemented. Here the Jacobi matrix of the single exponential equation is demonstrated since it is more clear and the double exponential one can be derived similarly. Based on Eqs. (1) and (2), they can be linearized as (for simplicity, the noise terms are removed):

$$\mathbf{x}_n = \mathcal{F}_n \mathbf{x}_{n-1} \tag{A.6}$$

$$\mathbf{y}_n = \mathcal{H}_n \mathbf{x}_n \tag{A.7}$$

where $\mathcal{F}_n$ and $\mathcal{H}_n$ are Jacobi matrices. They can be derived as:

$$\mathcal{F}_n^{ver1,M1} = \frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} e^{x_{n-1,2}} & x_{n-1,1}e^{x_{n-1},2} \\ 0 & 1 \end{bmatrix} \tag{A.8}$$

$$\mathcal{H}_n^{ver1,M1} = \frac{\partial h}{\partial \mathbf{x}} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{A.9}$$

Then, the observability matrix is introduced as follows:

$$rank(\mathcal{O}_{matrix}) = rank(\begin{bmatrix} \mathcal{H}_n \\ \mathcal{H}_n\mathcal{F}_n \\ ... \\ \mathcal{H}_n\mathcal{F}_n^{dimension-1} \end{bmatrix}) = dimension \tag{A.10}$$

which indicates that if the observability matrix $\mathcal{O}_{matrix}$ is full-rank, the system is observable. So, the Equation A.8 and A.9 can be taken into the Equation A.10:

$$rank(\mathcal{O}_{matrix}) = rank(\begin{bmatrix} 1 & 0 \\ e^{x_{n-1,2}} & x_{n-1,1}e^{x_{n-1,2}} \end{bmatrix}) = 2, \textbf{if } x_{n-1,1} \neq 0 \tag{A.11}$$

Another version $ver2$, which is commonly used in the research, follows:

$$\mathbf{x}_n^{ver2,M1} = \begin{bmatrix} x_{n-1,2}e^{x_{n-1,3}(n-1)} \\ x_{n-1,2} \\ x_{n-1,3} \end{bmatrix} + \mathbf{u} \tag{A.12}$$

$$\mathbf{y}_n^{ver2,M1} = x_{n,1} + \mathbf{v} \tag{A.13}$$

The rank of the observability matrix is 2.

Next, different configurations for the double exponential model $M2$ can be obtained similarly to the single exponential equation. In this section, five different versions are illustrated and explored to assess their observability.

$$\mathbf{x}_n^{ver1,M2} = \begin{bmatrix} x_{n-1,1}e^{x_{n-1,2}} + x_{n-1,3}e^{x_{n-1,4}n}(1 - e^{(x_{n-1,2} - x_{n-1,4})}) \\ x_{n-1,2} \\ x_{n-1,3} \\ x_{n-1,4} \end{bmatrix} + \mathbf{u} \tag{A.14}$$

$$\mathbf{y}_n^{ver1,M2} = x_{n,1} + \mathbf{v} \tag{A.15}$$

$$\mathbf{x}_n^{ver2,M2} = \begin{bmatrix} x_{n-1,1}e^{x_{n-1,3}} \\ x_{n-1,2}e^{x_{n-1,4}} \\ x_{n-1,3} \\ x_{n-1,4} \end{bmatrix} + \mathbf{u} \tag{A.16}$$

$$\mathbf{y}_n^{ver2,M2} = x_{n,1} + x_{n,2} + \mathbf{v} \tag{A.17}$$

$$\mathbf{x}_n^{ver3,M2} = \begin{bmatrix} x_{n-1,1} \\ x_{n-1,2} \\ x_{n-1,3} \\ x_{n-1,4} \end{bmatrix} + \mathbf{u} \tag{A.18}$$

$$\mathbf{y}_n^{ver3,M2} = x_{n,1}e^{x_{n,2}n} + x_{n,3}e^{x_{n,4}n} + \mathbf{v} \tag{A.19}$$

$$\mathbf{x}_n^{ver4,M2} = \begin{bmatrix} x_{n-1,2}e^{x_{n-1,3}n} + x_{n-1,4} * e^{x_{n-1,5}n} \\ x_{n-1,2} \\ x_{n-1,3} \\ x_{n-1,4} \\ x_{n-1,5} \end{bmatrix} + \mathbf{u} \tag{A.20}$$

$$\mathbf{y}_n^{ver4,M2} = x_{n,1} + \mathbf{v} \tag{A.21}$$

The rank of each version's observability can be calculated based on the above definition. Viz: $rank(\mathcal{O}_{matrix})^{ver1} = 2$, $rank(\mathcal{O}_{matrix})^{ver2} = 3$, $rank(\mathcal{O}_{matrix})^{ver3} = 1$ and $rank(\mathcal{O}_{matrix})^{ver4} = 2$. Here, the parameter d is proposed to be fixed in the double exponential equation to ensure the observability of the system. Otherwise, inferring all of the system's states is not feasible. The final version, whose observability matrix is full-rank, which is implemented in this study is shown in Equation A.22 and A.23. It should be noted the fixed parameter $d \neq 0$.

$$\mathbf{x}_n^{ver5,M2} = \begin{bmatrix} x_{n-1,1}e^{x_{n-1,3}} \\ x_{n-1,2}e^{d_{fixed}} \\ x_{n-1,3} \end{bmatrix} + \mathbf{u} \tag{A.22}$$

$$\mathbf{y}_n^{ver5,M2} = x_{n,1} + x_{n,2} + \mathbf{v} \tag{A.23}$$

Appendix B. HIs extraction

Here, the signal $x[i]$ is used to represent a piece of signal in the time domain which has K points ($i = 1, 2, ..., K$). $SE^{AS}_{filtered}[i]$ and EES are the SE of the filtered signal using the optimal band selected by the spectral kurtosis and the EES [33] of the signal, respectively. The HIs and their definition can be found in Table B1.

Appendix C. The metrics

In this study, four metrics are used for the performance determination, including Cumulative Relative Accuracy (CRA), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Score. Their definition and explanation are listed in the Table C1.

References

[1] S. Sharma, W. Abed, R. Sutton, B. Subudhi, Corrosion fault diagnosis of rolling element bearing under constant and variable load and speed conditions, IFAC-PapersOnLine 48 (30) (2015) 49–54, 9th IFAC Symposium on Control of Power and Energy Systems CPES 2015. doi:https://doi.org/10.1016/j.ifacol.2015.12.352.
URL https://www.sciencedirect.com/science/article/pii/S2405896315029948

Table B1: The HIs' formulas and corresponding explanations

HI	Definition	Explanation
RMS	$\sqrt{\frac{1}{n}\sum_i^K x_i^2}$	The arithmetic mean of the squares of a sequence of data, which shows the energy level.
The sum of EES	$\sum_{z=1}^{3} max(EES(0.98 * z * CF : 1.02 * z * CF))$	The integration over the instantaneous frequency axis showing the hidden periodicities of second-order cyclostationarity
Gini Index of EES	$\frac{1}{2} - 2\sum_{i=1}^{K}((\frac{K-i+\frac{1}{2}}{K})(\frac{EES[i]}{\|EES[i]\|_{L1}}))$	For characterizing the impulsiveness of repetitive transients
The L2 / L1 norm of the EES	$\frac{\|EES(0.98*CF:1.02*CF)\|_{L2}}{\|EES(0.98*CF:1.02*CF)\|_{L1}}$	The sparsity of the EES within the fault frequencies search band.
QAM1 of SE	$1 - \frac{PQAM_{(SE_{filtered}^{AS}+0.005)^{-1}}}{PQAM_{(SE_{filtered}^{AS})^1}}$	The extension of the sparsity measures (L2/L1, Gini Index, Kurtosis, etc.) which can be reformulated using different ratios of the QAM.
QAM2 of SE	$1 - \frac{PQAM_{(SE_{filtered}^{AS})^1}}{PQAM_{(SE_{filtered}^{AS}+0.005)^2}}$	
QAM3 of SE	$1 - \frac{PQAM_{(SE_{filtered}^{AS}+0.005)^{-2}}}{PQAM_{(SE_{filtered}^{AS}+0.005)^{-1}}}$	

*$PAQM_{(x+q)^p} = ((\sum_{i=1}^{K} x[i]+q)^p/K)^{\frac{1}{p}} - q$. For example, $QAM1 = 1 - \frac{((\sum_{i=1}^{K} SE_{filtered}^{AS}[i]+0.005)^{-1}/K)^{\frac{1}{-1}} - 0.005}{((\sum_{i=1}^{K} SE_{filtered}^{AS}[i])^1/K)^{\frac{1}{1}}}$.

Table C1: The metrics

Metrics	Definition	Description
CRA	$\sum_{i=Anomalytime}^{EoL} \omega_i RA_i, (RA_i = 1 - \frac{\left\|RUL_{actual}(i) - RUL_{prediction}(i)\right\|}{RUL_{actual}(i)}, \omega(i) = \frac{i}{\sum_{Anomalytime}^{EoL} i})$	CRA can comprehensively assess the accuracy of a prognostics approach by aggregating the relative prediction accuracies at all inspection time.
RMSE	$\sqrt{\frac{\sum_{Anomalytime}^{EoL}(RUL_{actual}(i) - RUL_{prediction}(i))^2}{EoL - Anomalytime}}$	RMSE is used to describe the error between the estimated RUL and ground truth RUL
MAE	$\frac{\sum_{Anomalytime}^{EoL}\left\|RUL_{actual}(i) - RUL_{prediction}(i)\right\|}{EoL - Anomalytime}$	Similar with RMSE
Score	The quantities of the estimated RUL points which are within the error area/(EoL-Anomaly time)	Score indicates how many points are within the error area

[2] Y. Wen, M. Fashiar Rahman, H. Xu, T.-L. B. Tseng, Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective, Measurement 187 (2022) 110276. doi:https://doi.org/10.1016/j.measurement.2021.110276.
URL https://www.sciencedirect.com/science/article/pii/S0263224121011805

[3] K. Javed, R. Gouriveau, N. Zerhouni, State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels, Mechanical Systems and Signal Processing 94 (2017) 214–236. doi:https://doi.org/10.1016/j.ymssp.2017.01.050.
URL https://www.sciencedirect.com/science/article/pii/S0888327017300468

[4] Y. Pan, T. Wu, Y. Jing, Z. Han, Y. Lei, Remaining useful life prediction of lubrication oil by integrating multi-source knowledge and multi-indicator data, Mechanical Systems and Signal Processing 191 (2023) 110174. doi:https://doi.org/10.1016/j.ymssp.2023.110174.
URL https://www.sciencedirect.com/science/article/pii/S088832702300081X

[5] N. Eleftheroglou, G. Galanopoulos, T. Loutas, Similarity learning hidden semi-markov model for adaptive prognostics of composite structures, Reliability Engineering & System Safety 243 (2024) 109808. doi:https://doi.org/10.1016/j.ress.2023.109808.
URL https://www.sciencedirect.com/science/article/pii/S0951832023007226

[6] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: A systematic review from data acquisition to rul prediction, Mechanical Systems and Signal Processing 104 (2018) 799–834. doi:https://doi.org/10.1016/j.ymssp.2017.11.016.
URL https://www.sciencedirect.com/science/article/pii/S0888327017305988

[7] M. Ma, L. Fu, Z. Zhai, R.-B. Sun, Transformer based kalman filter with em algorithm for time series prediction and anomaly detection of complex systems, Measurement 229 (2024) 114378.

doi:https://doi.org/10.1016/j.measurement.2024.114378.
URL https://www.sciencedirect.com/science/article/pii/S0263224124002628

[8] X. Xu, W. Li, M. Zhao, H. Hu, Mobile device-based bearing diagnostics with varying speeds, Measurement 200 (2022) 111639. doi:https://doi.org/10.1016/j.measurement.2022.111639.
URL https://www.sciencedirect.com/science/article/pii/S0263224122008491

[9] X. Ma, B. Yan, H. Wang, H. Liao, A hybrid prognostic method for rotating machinery under time-varying operating conditions by fusing direct and indirect degradation characteristics, Measurement 214 (2023) 112831. doi:https://doi.org/10.1016/j.measurement.2023.112831.
URL https://www.sciencedirect.com/science/article/pii/S0263224123003950

[10] J. Sikorska, M. Hodkiewicz, L. Ma, Prognostic modelling options for remaining useful life estimation by industry, Mechanical Systems and Signal Processing 25 (5) (2011) 1803–1836. doi:https://doi.org/10.1016/j.ymssp.2010.11.018.
URL https://www.sciencedirect.com/science/article/pii/S0888327010004218

[11] M. L. Baptista, E. M. Henriques, H. Prendinger, Classification prognostics approaches in aviation, Measurement 182 (2021) 109756. doi:https://doi.org/10.1016/j.measurement.2021.109756.
URL https://www.sciencedirect.com/science/article/pii/S0263224121007132

[12] A. Gabrielli, M. Battarra, E. Mucchi, G. Dalpiaz, Physics-based prognostics of rolling-element bearings: The equivalent damaged volume algorithm, Mechanical Systems and Signal Processing 215 (2024) 111435. doi:https://doi.org/10.1016/j.ymssp.2024.111435.
URL https://www.sciencedirect.com/science/article/pii/S0888327024003339

[13] W. Zhang, X. Li, X. Li, Deep learning-based prognostic approach for lithium-ion batteries with adaptive time-series pre-

diction and on-line validation, Measurement 164 (2020) 108052. doi:https://doi.org/10.1016/j.measurement.2020.108052.
URL https://www.sciencedirect.com/science/article/pii/S026322412030590X

[14] Y. Wang, J. Wu, Y. Cheng, J. Wang, K. Hu, Memory-enhanced hybrid deep learning networks for remaining useful life prognostics of mechanical equipment, Measurement 187 (2022) 110354. doi:https://doi.org/10.1016/j.measurement.2021.110354.
URL https://www.sciencedirect.com/science/article/pii/S0263224121012483

[15] D. Wang, J. zhen Kong, F. Yang, Y. Zhao, K.-L. Tsui, Battery prognostics at different operating conditions, Measurement 151 (2020) 107182. doi:https://doi.org/10.1016/j.measurement.2019.107182.
URL https://www.sciencedirect.com/science/article/pii/S0263224119310486

[16] X. Sun, K. Zhong, M. Han, A hybrid prognostic strategy with unscented particle filter and optimized multiple kernel relevance vector machine for lithium-ion battery, Measurement 170 (2021) 108679. doi:https://doi.org/10.1016/j.measurement.2020.108679.
URL https://www.sciencedirect.com/science/article/pii/S0263224120311908

[17] M. Ahwiadi, W. Wang, An enhanced particle filter technology for battery system state estimation and rul prediction, Measurement 191 (2022) 110817. doi:https://doi.org/10.1016/j.measurement.2022.110817.
URL https://www.sciencedirect.com/science/article/pii/S0263224122001130

[18] N. Li, Y. Lei, T. Yan, N. Li, T. Han, A wiener-process-model-based method for remaining useful life prediction considering unit-to-unit variability, IEEE Transactions on Industrial Electronics 66 (3) (2019) 2092–2101. doi:10.1109/TIE.2018.2838078.

[19] E. Soave, G. D'Elia, G. Dalpiaz, Prognostics of rotating machines through generalized gaussian hidden markov models, Mechanical Systems and Signal Processing 185 (2023) 109767.

doi:https://doi.org/10.1016/j.ymssp.2022.109767.
URL https://www.sciencedirect.com/science/article/pii/S0888327022008354

[20] R. K. Singleton, E. G. Strangas, S. Aviyente, Extended kalman filtering for remaining-useful-life estimation of bearings, IEEE Transactions on Industrial Electronics 62 (3) (2015) 1781–1790. doi:10.1109/TIE.2014.2336616.

[21] C. K. R. Lim, D. Mba, Switching kalman filter for failure prognostic, Mechanical Systems and Signal Processing 52-53 (2015) 426–435. doi:https://doi.org/10.1016/j.ymssp.2014.08.006.
URL https://www.sciencedirect.com/science/article/pii/S0888327014003227

[22] L. Cui, X. Wang, Y. Xu, H. Jiang, J. Zhou, A novel switching unscented kalman filter method for remaining useful life prediction of rolling bearing, Measurement 135 (2019) 678–684. doi:https://doi.org/10.1016/j.measurement.2018.12.028.
URL https://www.sciencedirect.com/science/article/pii/S0263224118311734

[23] J. Qi, R. Zhu, C. Liu, A. Mauricio, K. Gryllias, Anomaly detection and multi-step estimation based remaining useful life prediction for rolling element bearings, Mechanical Systems and Signal Processing 206 (2024) 110910. doi:https://doi.org/10.1016/j.ymssp.2023.110910.
URL https://www.sciencedirect.com/science/article/pii/S088832702300818X

[24] J. Antoni, K. Kestel, C. Peeters, Q. Leclère, F. Girardin, T. Ooijevaar, J. Helsen, On the design of optimal health indicators for early fault detection and their statistical thresholds, Mechanical Systems and Signal Processing 218 (2024) 111518. doi:https://doi.org/10.1016/j.ymssp.2024.111518.
URL https://www.sciencedirect.com/science/article/pii/S0888327024004163

[25] L. Song, K. Fukumizu, A. Gretton, Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric inference

in graphical models, IEEE Signal Processing Magazine 30 (4) (2013) 98–111. doi:10.1109/MSP.2013.2252713.

[26] K. Fukumizu, L. Song, A. Gretton, Kernel bayes' rule: Bayesian inference with positive definite kernels, Journal of Machine Learning Research 14 (118) (2013) 3753–3783.
URL http://jmlr.org/papers/v14/fukumizu13a.html

[27] M. Sun, M. E. Davies, I. Proudler, J. R. Hopgood, Adaptive kernel kalman filter, in: 2021 Sensor Signal Processing for Defence Conference (SSPD), 2021, pp. 1–5. doi:10.1109/SSPD51364.2021.9541455.

[28] M. Sun, M. E. Davies, I. K. Proudler, J. R. Hopgood, Adaptive kernel kalman filter, IEEE Transactions on Signal Processing 71 (2023) 713–726. doi:10.1109/TSP.2023.3250829.

[29] G. H. W. Gebhardt, A. G. Kupcsik, G. Neumann, The kernel kalman rule, Machine Learning 108 (2019) 2113 – 2157.
URL https://api.semanticscholar.org/CorpusID:190643191

[30] B. Wang, Y. Lei, N. Li, N. Li, A hybrid prognostics approach for estimating remaining useful life of rolling element bearings, IEEE Transactions on Reliability 69 (1) (2020) 401–412. doi:10.1109/TR.2018.2882682.

[31] J. Antoni, G. Xin, N. Hamzaoui, Fast computation of the spectral correlation, Mechanical Systems and Signal Processing 92 (2017) 248–277. doi:https://doi.org/10.1016/j.ymssp.2017.01.011.
URL https://www.sciencedirect.com/science/article/pii/S0888327017300134

[32] B. Hou, D. Wang, T. Xia, Y. Wang, Y. Zhao, K.-L. Tsui, Investigations on quasi-arithmetic means for machine condition monitoring, Mechanical Systems and Signal Processing 151 (2021) 107451. doi:https://doi.org/10.1016/j.ymssp.2020.107451.
URL https://www.sciencedirect.com/science/article/pii/S0888327020308372

[33] A. Mauricio, K. Gryllias, Cyclostationary-based multiband envelope spectra extraction for bearing diagnostics: The combined improved envelope spectrum, Mechanical Systems and Signal Processing 149 (2021)

107150. doi:https://doi.org/10.1016/j.ymssp.2020.107150.
URL https://www.sciencedirect.com/science/article/pii/S0888327020305367