

Estimation of machinery's remaining useful life in the presence of non-Gaussian noise by using a robust extended Kalman filter

Hamid Shiri ^{*a}, Pawel Zimroz^a, Agnieszka Wyłomańska^b, Radosław Zimroz^{a, a}

^a*Faculty of Geoengineering, Mining and Geology, Wrocław University of Science and Technology, Na Grobli 15, 50-421 Wrocław, Poland*

^b*Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wrocław University of Science and Technology, Wyspiańskiego 27, 50-370 Wrocław, Poland*

Abstract

Estimation of the remaining useful life (RUL) of industrial machinery is essential for condition-based maintenance (CBM). While numerous papers have explored this issues, challenges arise as machinery often works in non-stationary conditions, particularly in harsh environments (like mining machines, wind turbines, helicopters, etc). The data collected from such environments are affected by non-Gaussian noise, posing difficulties for traditional approaches to non-linear state estimation or prediction. The widely used extended Kalman filter (EKF) suffers from the non-Gaussian noise effect due to its recursive minimum L2-norm filtering. To address these issues, we propose a robust EKF based on the maximum correntropy criterion. This method effectively estimates the RUL of the time-varying degradation process in the presence of non-Gaussian noise, also enabling confidence interval computation for uncertainty management. The efficiency of our approach was confirmed through application to simulated and benchmark data sets, outperforming Kalman filter-based methods for both simulated and real-world scenarios.

Keywords: prognostics, remaining useful life (RUL), non-linear degradation, extended Kalman filter, robust methods, non-Gaussian noise.

Nomenclature

Nomenclature		KF	Kalman Filter
CBM	Condition-based maintenance	MAE	Mean absolute erro
CI	confidence interval	MCEKF	Maximum correntropy-based Extended Kalman Filter
CKF	Cubature Kalman filter	MCSKF	maximum correntropy criterion switching Kalman filter
CM	Condition monitoring	ML	Machine learning
EKF	Extended Kalman Filter	MSE	Mean square error
EOL	End of the life	PDF	Probability density function
FPT	First prediction time	PF	Particle filter
HI	Health index	PHM	Prognostics and health management
IMS	Intelligent maintenance system data set		

*Corresponding author.

**Corresponding author, ...@pwr.edu.pl

Email address: hamid.shiri@pwr.edu.pl (Hamid Shiri *)

QKF	Quadratic Kalman filter	SMC	Sequential Monte Carlo
RMS	Root mean square	TRB	tapered roller bearing
RUL	Remaining useful life	UKF	Unscented Kalman filter

1. Introduction

Condition-based maintenance (CBM) revolutionises maintenance strategies by continuous monitoring of equipment and systems in real time, allowing precise maintenance decisions based on their current condition instead of rigid schedules. By leveraging advanced sensors and data analysis, the CBM rapidly identifies deviations from normal operation, ensuring optimal asset health assessment. This enables organisations to minimise downtime, extend the life of equipment, and reduce maintenance costs, leading to unparalleled operational efficiency in diverse industries. In the context of CBM, the role of RUL or prognostics takes centre stage, providing critical insights into an asset’s operational longevity based on its current state and performance trends. RUL serves as an early warning system for potential failures, empowering proactive and cost-effective maintenance planning that further reduces downtime and maximises asset lifespan. This data-driven methodology equips organizations to make informed maintenance choices, strategically prioritize vital components, and refine maintenance resource allocation, ultimately bolstering operational efficiency and curbing maintenance expenditures. Prognostic approaches can be categorised into three main groups: the data-driven approach (including the machine learning (ML) based approach [1, 2], and the statistical model-based approach [3–5]), physics-based approach [6, 7], and hybrid approach [8, 9].

Physics-based model approaches try to explain the degradation process by taking advantage of the physics of this process based on damage and fracture mechanics [10]. This family of approaches could provide accurate results. However, the exact physics knowledge of complex systems is not always available or too expensive to extract [11]. Therefore, in real applications, the use of physics-based model approaches is subject to significant restrictions.

The data-driven approach aims to build a model of the degradation process by using historical data, particularly valuable when examining intricate engineering systems such as wind turbines, aircraft, and mining machines. Within this approach, there are two prominent subclasses: the machine learning-based approach and the statistical model-based approach, each offering distinct advantages and prerequisites. The machine learning-based approach serves as a powerful tool for modelling, segmenting, and predicting data from complex systems with degradation processes that defy easy correlation with physics or statistics. Nevertheless, these methods demand a substantial volume of historical degradation data for effective training, a resource that often proves elusive for many practical equipment scenarios. In contrast, statistical model-based approaches require comparatively less degradation data and do not require advanced knowledge of equipment mechanics. Moreover, they exhibit considerable potential to accommodate degradation uncertainties. When opting for a statistical model-based approach, the selection of an appropriate model that accurately characterises the degradation process becomes crucial. In summary, while both machine learning and statistical model-based approaches are integral to data-driven methodologies, the former thrives in complex, data-rich environments but demands more historical data, whereas the latter excels in scenarios with limited data and uncertainty considerations, prioritising the choice of an effective degradation model. The hybrid approach aims to benefit from the advantages of both physics-based and data-driven approaches by combining them. More detailed information on the hybrid approach can be found in the literature [8, 9].

Kalman filter (KF) is one of the well-known methods to perform state estimation for engineering purposes, which has been used regularly for diagnostics and prognostics [12–15]. However, most of the real cases, especially in the area of prognostics, are non-linear, which creates a limitation for using KF. To solve the system’s non-linear state estimation problem, many extensions of KF, such as EKF [16], quadratic KF (QKF) [17], unscented KF (UKF) [18], cubature KF (CKF) [19], have been proposed in the literature. EKF is computationally efficient and simpler to implement rather than other mentioned filters, however, might not accurately handle with highly non-linear systems. The fundamental idea of EKF is to linearise the system using the first-order term of the Taylor series expansion of non-linear functions.

KF family usually uses second-order statistics and can perform optimal solutions when the noise is Gaussian distributed. However, most of machines in industries work in harsh environments [20]. Due to this, the data collected from these machines are usually affected by noise with non-Gaussian characteristics, which can pose a great challenge for the classic KF family approach. We should clarify that in this paper we consider heavy-tailed noise, which is a specific type of non-Gaussian noise characterised by the presence of extreme or outlying values in a data set, often leading to a distribution with thicker tails than the Gaussian (normal) distribution. Non-Gaussian noise, unlike the more typical Gaussian noise, is frequently encountered in practical applications, such as monitoring the health of mining equipment and wind turbines. For instance, when assessing the condition of mining machinery dealing with the screening of falling ore, the noise in the data often exhibits non-Gaussian characteristics due to irregular patterns and unexpected behaviors during the sieving process. Similarly, in the context of wind turbines, turbulence in the surrounding wind flow can introduce non-Gaussian noise into the collected data. Additionally, in scenarios involving electrical devices or electromagnetic fields, non-Gaussian noise may arise from the unpredictable behavior of electromagnetic interference or fluctuations in power supply. By acknowledging the existence of non-Gaussian noise originating from sources like falling ore in sieving screens, wind turbulence, and electromagnetic fields, we can better adapt our diagnostic and prognostic methods to accommodate these unique statistical properties.

Recently, many methods have been developed to handle the effect of non-Gaussian noise in classic KF. A part of these approaches is developed based on considering the non-Gaussian distribution for noise, which is fulfilled for distributions like alpha-stable distributions, Student's t distribution, etc [21, 22]. Nevertheless, finding the analytic solutions for these distributions is a hard task and a high computational cost is also required. Another approach uses the Gaussian mixture distribution to approximate non-Gaussian noise [23–26]. The high computational cost can be mentioned as the main weakness of this family of methods. The third family of approaches that attempts to address a non-Gaussian noise problem is sequential Monte Carlo (SMC) sampling, which can approximate any distribution [27]. The particle filter (PF) can be the best example of this family of methods [24, 28]. As with previous methods, the high computational cost is the most significant disadvantage of this family of approaches.

The computational cost is a key parameter in the selection of methods in the subject of prognostics and RUL estimation, mainly because it is needed to process a large amount of data. To treat the high computational cost and the non-Gaussian noise effect, some papers have proposed a replacement for the mean-square error (MSE) criterion (inherent to the standard KF). For example, based on the definition of corr entropy which is a similarity measure (for introduction to topic, see [29]), some research used the correntropy criterion to drive KF [12, 30–33]. Since the correntropy criterion uses a higher-order statistic of the noise instead of second-order statistics, the performance of state estimation will increase in comparison to standard KF in the presence of noise with non-Gaussian characteristics. In [12, 34], Shiri et al. introduced a novel approach that utilises the maximum correntropy switching Kalman filter (MCSKF) for state detection when encountering non-Gaussian noise in a system. Their findings highlight the promise of MCSKF for online state detection, although their research primarily focusses on machine state detection without addressing predictive applications. It is important to note that the methods proposed in their work rely on a linear KF framework, which may not be universally applicable due to its linear assumptions in many practical scenarios.

In this study, we aim to enhance the accuracy of RUL predictions and offer probabilistic outcomes to tackle the challenges posed by dynamic, non-linear degradation processes combined with noise exhibiting non-Gaussian characteristics. To achieve this, we introduce a robust EKF that is based on the maximum correntropy criteria, i.e. the maximum correntropy EKF (MCEKF). This approach seeks to address the complexities of the RUL estimation problem by considering both the uncertainties in the degradation process and the non-Gaussian characteristics of the noise.

The main contribution of this paper can be summarized as follows:

- a robust version of EKF is driven based on maximum correntropy criteria to deal with non-linear degradation processes combined with noise with non-Gaussian characteristics;
- comparative simulations with the other filters are executed by Monte Carlo analysis in the presence of Gaussian and non-Gaussian noise to indicate the dominance of MCEKF;

- the proposed MCEKF is carried out on three benchmark real data sets to estimate RUL and verify the efficacy of the approach in real environments.

The rest of the paper is organised as follows. In Section 2, the lifetime curve models are examined based on the literature. In Section 3, the theory for the critical parts of the processing methods is described. Then, in Section 4, the results for the simulated data are shown. Ultimately, the results for three benchmark data sets obtained using the proposed approach are demonstrated in Section 5 with an expression of all intermediate steps. Conclusions are formed in Section 6.

2. Lifetime curve model

In the PHM community, there are few statistical models that are used to describe the degradation process. These models are usually made up of the deterministic part (which is trying to qualify the global trend of the degradation process) and the random part (which is trying to consider the uncertainty of the degradation process). They are selected based on the behaviour of the degradation process with different deterministic and random trends [20]. In this work, we selected a time-varying model with three different regimes. The first regime has a constant deterministic trend, which refers to the healthy state of the system (the machine is working stable). The second regime has a linear growing trend which is assigned to the degradation of the machine. The last regime has an exponential trend, which is attributed to the critical state of the degradation machine. Moreover, the FPT starts from the last regime in this model.

Also, as discussed in the literature [20], modelling of the random part is very important to reflect the uncertainty. In some models it is assumed that the variance of the noise is constant and the distribution of the noise is Gaussian [35–37], see Fig. 1. However, Zulawinski et al. [20] demonstrated that the random part may be non-homogeneous (has a time-varying scale), and the noise distribution is far from the Gaussian distribution. Therefore, we used a model with non-stationary characteristics and non-Gaussian noise; see Fig. 2. The scale (variance) of the random part changes during the degradation process based on the regime in which we are in.

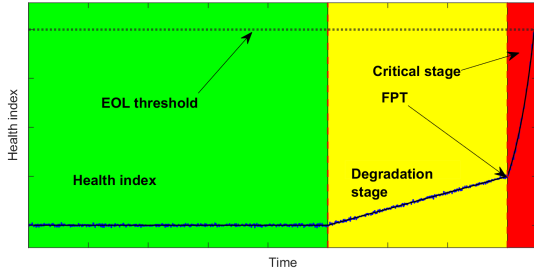


Figure 1: The long-term degradation model with three stages.

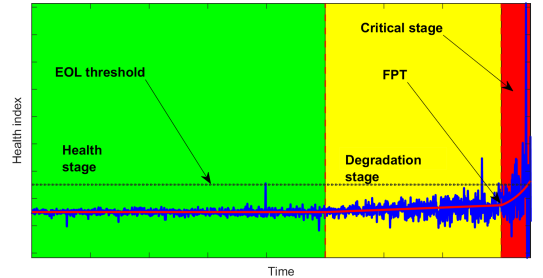


Figure 2: The proposed long-term degradation model with three stages.

It is important to note that our study operates under the assumption of a known FPT or first point of the critical stage, from which our prediction methodology commences. Due to this fact, there is no valuable information on the healthy stage for prediction, and starting the prediction from the degradation stage will lead to over- or underestimation of RUL. It is worth highlighting that the accurate identification of the FPT is a crucial issue in the context of the estimation of the RUL, bearing a significant effect on the eventual results. However, we must clarify that the exploration of FPT detection lies beyond the scope of this current work, which is centred on the prediction task. Consequently, we leverage the findings derived from our previous research endeavours, where the dedicated focus was directed toward FPT detection within real-world data sets. For those interested in an in-depth analysis of this facet, we kindly direct them to our previous comprehensive publications that explore extensively this subject [2, 12, 38, 39].

In the present study, we build upon the results of our prior work [12], which introduced a robust methodology

based on the switching maximum correntropy Kalman filter (SMCKF) to address the challenges of threshold problem (selecting the threshold for warning and alarm not for EOL) and online diagnostics, especially in the presence of non-Gaussian noise, utilising condition monitoring (CM) data. This approach leverages a set of dynamic system models to elucidate various stages of degradation, employing robust Bayesian estimation techniques. Due to its foundation in dynamic behaviour modelling, this approach obviates the need for a fixed threshold in the diagnostic process. Therefore, we can characterise our current paper as an extension of our previous work, with a specific focus on prediction tasks and addressing nonlinearity.

3. Methodology

This section outlines the methodology grounded in a degradation process for the estimation of RUL. The RUL estimation workflow, based on our proposed approach, is illustrated in Fig. 3. The data obtained, such as sensor-recorded vibration data, serves as a means to monitor the machine’s health condition. Through signal processing techniques, HI is constructed from the measured signals, effectively representing the machine’s health status. Depending on the evolving degradation patterns of HI, the life of the machinery is categorised into three distinct stages: health, degradation, and the critical stage, where the FPT is identified as the initial entry into the critical stage. In the critical stage, we defined a state-space degradation model, and its parameters are estimated using the MCEKF and updated in real time as new HI values become available. An EOL threshold is defined, enabling the calculation of the PDF of the estimated RUL based on the constructed degradation model and the EOL threshold.

Furthermore, the details of the state space degradation model and the theoretical basis of this process, which includes the derivation of the EKF and MCEKF, are described in the following subsections.

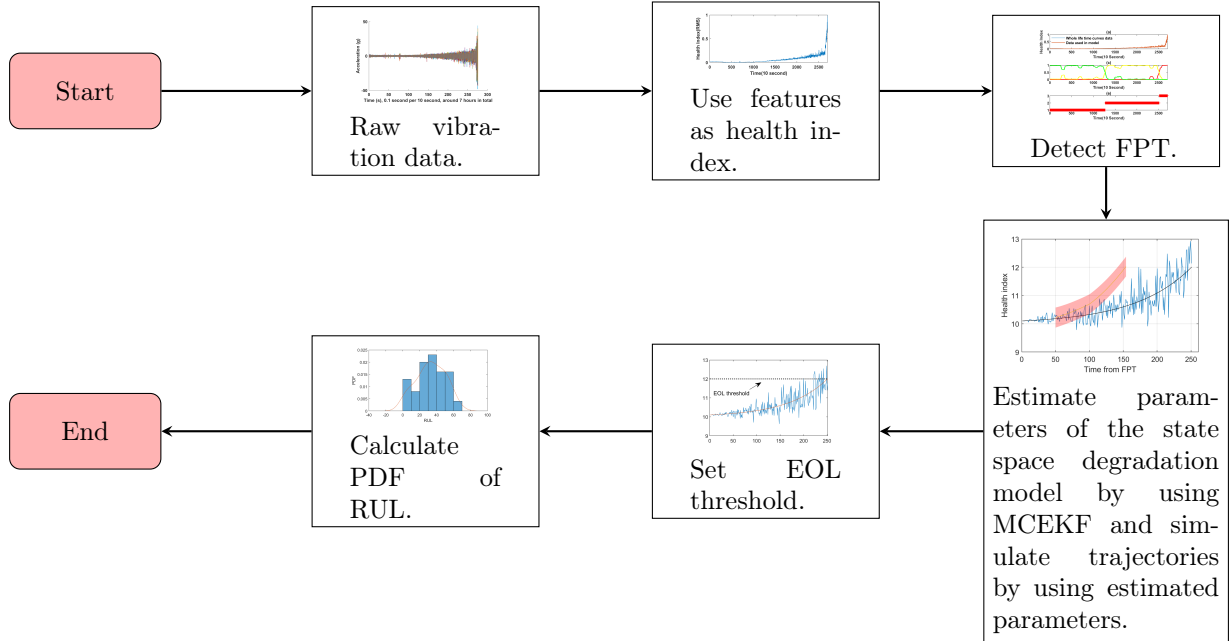


Figure 3: Complete procedure for derive distribution of RUL

3.1. RUL

RUL is a critical metric in the field of machinery prognostics, representing the duration for which a machine is expected to operate efficiently before the need for repair or replacement. Fig. 4 illustrates the concept of RUL. Our methodology focusses on training a predictive model using historical data exclusively. Subsequently, we employed this model to forecast future values of the degradation process, accompanied

by CI. The estimated degradation curve is then examined to identify the point in time at which it initially crosses a predetermined EOL threshold, denoted t_{EOL} . The predicted RUL is calculated as the difference between this failure time t_{EOL} and the time of the last observation in the training data set (t_0), such that the predicted RUL = $t_{\text{EOL}} - t_0$.

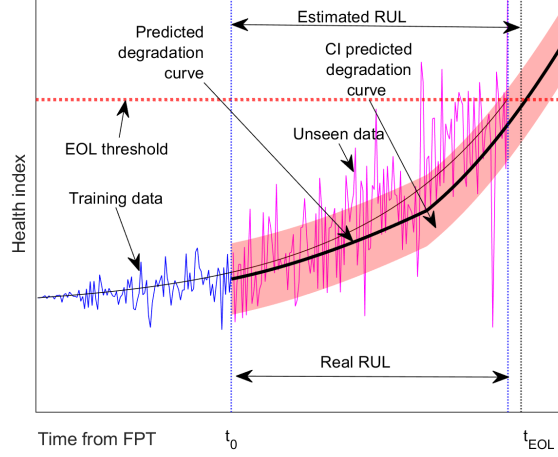


Figure 4: Complete procedure for derive distribution of RUL.

3.2. State space degradation model

As confirmed by the findings in [20], the exponential model consistently demonstrates a better global regression performance compared to polynomial and other conventional models. Consequently, we have opted to adopt the ensuing exponential model as the degradation pattern designated for our analysis of the third regime of the degradation process (critical stage):

$$y_t = a \exp(bt) + c, \quad (1)$$

where observation y_t corresponds to the HI, t is the discrete time and a, b, c represent the parameters of the model that need to be initialized with real data.

Recognising the intricate and time sensitive nature of the degradation process, this study postulates that the model parameters a_t, b_t, c_t , are subject to temporal variations. This assumption aligns closely with the dynamics of the degradation process in the real world. Let the dynamic state \mathbf{x}_t be defined as a column vector $\mathbf{x}_t = [a_t, b_t, c_t]^T$. Furthermore, recognising the impact of external noise and inherent uncertainty, we introduce the subsequent state space model:

$$\begin{cases} \mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{q}_t \\ y_t = a_t \exp(b_t t) + c_t + m_t \end{cases}, \quad (2)$$

where, the process noise is denoted as \mathbf{q}_t , presumed to be uncorrelated zero-mean white noise. This noise factor is characterised by the covariance matrix \mathbf{Q}_t . Besides, we assume m_t to be a scalar noise, not necessarily Gaussian, to suit heavy tailed noise case inherent in the real data (her it has no longer the interpretation of the measurement noise). The noise m_t is characterized with a time changing scale parameter M_t . Keeping an analogy with the degradation model from Eq. (5), we map the relationship between HI and the parameters featured in Eq. (2) denoting it with the non-linear functions \mathbf{f} and h as follows:

$$\mathbf{f}(\mathbf{x}_{t-1}) = \mathbf{x}_{t-1}, \quad (3)$$

$$h(\mathbf{x}_t) = a_t \exp(b_t t) + c_t. \quad (4)$$

Furthermore, we set the initial state \mathbf{x}_0 and the initial values for \mathbf{Q}_0 and M_0 as well as initial value $\mathbf{P}_{0|0}$ for the covariance matrix $\mathbf{P}_{t|t}$ which is described in the next subsection. In the following two subsections we describe EKF and introduce MCEKF considering more general notation. Please note, that in general approach y_t would be replaced with a vector \mathbf{y}_t , the scalar function h with a vector function \mathbf{h} and the scale parameter M_t would be replaced with the matrix \mathbf{M}_t which refers to the covariance matrix in the Gaussian case.

3.3. Extended Kalman filter

The EKF works based on the use of Taylor series to transform non-linear filtering problems into linear forms. The discrete non-linear state-space representation of the model can be presented as follows.

$$\begin{aligned}\mathbf{x}_t &= \mathbf{f}(\mathbf{x}_{t-1}) + \mathbf{q}_t, \\ \mathbf{y}_t &= \mathbf{h}(\mathbf{x}_t) + \mathbf{m}_t,\end{aligned}\tag{5}$$

where \mathbf{x}_t is the unknown dynamic state at time t and \mathbf{y}_t is its observation, respectively – both of them are in general vectors of selected size. Furthermore, \mathbf{q}_t and \mathbf{m}_t represent processes and measurement noise that are considered Gaussian distributed: $\mathbf{q}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$, $\mathbf{m}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{M}_t)$. Furthermore, \mathbf{f} and \mathbf{h} are non-linear state transition function and measurement function, respectively. Let us denote by $\mathbf{P}_{t|t-1}$ the prior estimate covariance matrix:

$$\mathbf{P}_{t|t-1} = E[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})^T]\tag{6}$$

and with $\mathbf{P}_{t|t}$ the posterior estimate covariance matrix:

$$\mathbf{P}_{t|t} = E[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t})^T].\tag{7}$$

The hat operator here means the estimation procedure made by KF conditioned on the specific time (normally on the current step t or the previous step $t-1$).

The EKF can be summarised as follows with the two steps iterated subsequently, namely the prediction step:

$$\begin{aligned}\hat{\mathbf{x}}_{t|t-1} &= \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1}), \\ \hat{\mathbf{P}}_{t|t-1} &= \mathbf{A}_t \hat{\mathbf{P}}_{t-1|t-1} \mathbf{A}_t^T + \mathbf{Q}_t,\end{aligned}\tag{8}$$

and the update step:

$$\begin{aligned}\hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1})), \\ \hat{\mathbf{P}}_{t|t} &= (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \hat{\mathbf{P}}_{t|t-1} (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t)^T + \mathbf{K}_t \mathbf{M}_t \mathbf{K}_t^T.\end{aligned}\tag{9}$$

where T means matrix transposition, \mathbf{I} is identity matrix of the proper size. The term \mathbf{K}_t in the above equations is called Kalman gain and for the standard EKF its optimal value is given by the formula:

$$\mathbf{K}_t = \hat{\mathbf{P}}_{t|t-1} \mathbf{H}_t^T (\mathbf{H}_t \hat{\mathbf{P}}_{t|t-1} \mathbf{H}_t^T + \mathbf{M}_t)^{-1},\tag{10}$$

where \mathbf{A}_t and \mathbf{H}_t are extracted by calculating the Jacobian matrices \mathbf{f} and \mathbf{h} , respectively, as follows:

$$\mathbf{A}_t = \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_{t-1}}\tag{11}$$

$$\mathbf{H}_t = \left. \frac{\partial \mathbf{h}}{\partial \mathbf{x}} \right|_{\mathbf{x}_t}\tag{12}$$

3.4. Maximum correntropy Extended Kalman filter

In the following, for the model expressed with Eq. (5) another form of an objective function is introduced. Basis for this modification is the maximum correntropy criterion. In this work, we use Gaussian kernel function, and thus the correntropy of two random vectors \mathbf{X} and \mathbf{Y} is defined as following [29]:

$$C(\mathbf{X}, \mathbf{Y}) = E[G(\|\mathbf{X} - \mathbf{Y}\|)].\tag{13}$$

where $G_\sigma(u) = \exp(-\frac{u^2}{2\sigma^2})$ and σ is the size of the kernel. The maximum correntropy criterion states that the distance (measured with Euclidean norm) between two random vectors is minimised when the correntropy, defined as the expected value of an arbitrary kernel function, is maximised. The benefit of using correntropy is that, unlike the squared error cost function, in its Maclaurin series expansion it has terms of higher order than second and is insensitive to outliers (as was described in [31]). Based on this criterion, there are different possible formulations of the cost function for MCEKF as a discrete sum (sample mean) of kernel function values. In [40] it is summed through all eigenvalues of both \mathbf{M}_t and $\mathbf{P}_{t|t-1}$. We, on the other hand, define it as a sum of two values of the kernel function, similar to [31] and our previous work [12]:

$$J_{MC} = \frac{1}{2}G_\sigma(\|\mathbf{y}_t - \mathbf{h}(\mathbf{x}_t)\|_{\mathbf{M}_t^{-1}}) + \frac{1}{2}G_\sigma(\|\mathbf{x}_t - \mathbf{f}(\hat{\mathbf{x}}_{t-1|t-1})\|_{\mathbf{P}_{t|t-1}^{-1}}). \quad (14)$$

Note that it takes the two standardised noise terms of the same model that EKF does in the previous section, but it combines them in a different way.

Using this cost function, compared to using a quadratic cost function from the standard KF, makes the estimation more robust in the presence of non-Gaussian noise, which is the case when the real observation process does not necessarily fulfil the assumptions of the KF model (see [31]). The formulation of MCEKF can be derived by finding \mathbf{x}_t that maximizes the objective function J_{MC} given by Eq. (14). The resulting prediction and update steps are driven by the same equations as for the standard EKF; see Eq. (8), (9) and (11), (12) but the optimal Kalman gain for the MCEKF is given with the formula:

$$\mathbf{K}_t = \hat{\mathbf{P}}_{t|t-1} \lambda_t \mathbf{H}_t^T (\mathbf{H}_t \hat{\mathbf{P}}_{t|t-1} \lambda_t \mathbf{H}_t^T + \mathbf{M}_t)^{-1}, \quad (15)$$

where

$$\lambda_t = G_\sigma(\|\mathbf{y}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1})\|_{\mathbf{M}_t^{-1}}). \quad (16)$$

3.5. Discussion of the hyperparameters setting

In the state estimation procedure using MCEKF, when a significant outlier appears, the term $\mathbf{y}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1})$ in Eq. (9) (usually called innovation or prefit residual) diverges, but \mathbf{K}_t controls the divergence of the estimator $\hat{\mathbf{x}}_{t|t}$. See [30–33] for more details on the procedure of driving equations and stability.

Also, we should clarify that, in the MCEKF, the σ parameter (kernel size) controls how sensitive the filter is to differences between predicted and measured states. A larger σ makes the filter more tolerant to deviations, while a smaller σ makes it more sensitive to small differences, which could help to reject outliers but might also be more affected by noise. Therefore, selecting a proper value of σ is crucial for balancing noise robustness and sensitivity to data deviations. It often requires some experimentation or domain knowledge to determine the best σ value for a given application.

In addition, the measurement error, denoted as m_t , is determined by calculating the variance of stationary measurements acquired when the machinery is in a healthy operational state. It is important to note that the specific constant value of $M_t = M$ may vary among individual cases. The assumption of stationary measurements from well-functioning machinery, as applied in this study, is rigorously validated by analysing data from multiple cases, thus confirming its practical applicability.

As discussed in [41], the process error, represented as \mathbf{q}_t , encapsulates the uncertainty associated with the filter's ability to model real-world dynamics. A lower process error enhances the filter's fitting of condition monitoring (CM) data, making the model more responsive to delicate changes, although potentially making it susceptible to overfitting. In our approach, the process error, denoted as $\mathbf{q}_t = \mathbf{q}$, is fine-tuned for the MCEKF model based on historical data from similar defect cases and is assumed to be consistent across cases. In this study, an arbitrary covariance matrix is employed, denoted $\mathbf{P}_{0|0}$.

With the establishment of these matrices and the definition of the initial parameters, the MCEKF model for the prognostic evaluation of bearing degradation is formulated.

3.6. EOL

The EOL in machinery prognostics is a critical parameter, and selecting the right threshold is paramount for effective maintenance and operation. It marks the point when a machine is expected to become unreliable, which makes it crucial for scheduling preventive maintenance, cost reduction, safety, resource optimisation, reliability, and environmental impact. The key to success lies in data-driven approaches, such as predictive maintenance techniques and thorough data analysis, which enable organisations to accurately identify and set appropriate EOL thresholds for their machinery. It is worth noting that choosing an appropriate EOL value for real data sets presents a considerable challenge, and this has been the subject of extensive research over the past decade [37, 42]. The complexity of this task endures. In our study, we adopted a random selection approach for the EOL value. This choice comes from the flexibility of our proposed method, which is not restricted to predefined EOL values. It can be easily computed for a range of EOL values, offering versatility in its application.

Additionally, in Appendix [Appendix B](#), pseudocodes detailing our proposed methodology are provided for reference.

4. Simulation and results

4.1. Model of the degradation curve

In the following section according to [20], the HI data is generated. The simulation is restricted only to the third regime starting from the FPT as was argued in Section 2. As was discussed there, the HI is constructed from two parts:

$$HI(t) = D(t) + R(t), \quad (17)$$

where $D(t)$ and $R(t)$ are as deterministic and random parts of the HI, respectively. The deterministic part in Eq. (17) is described as follows:

$$D(t) = a_1 \exp(b_1 t) + c_1, \quad (18)$$

where a_1 , b_1 , and c_1 are constant values that are used to build the exponential function that corresponds to the critical stage. In the simulation part, we consider Gaussian and Student's t distributions for the random component of the degradation process (for more details about Student's t distribution, see [Appendix A](#)). The time-varying random component corresponding to $R(t)$ is generated in the following way:

$$R(t) = SC(t)\tilde{R}(t), \quad (19)$$

where $SC(t)$ corresponds to the scale of the random part (variance) which is constructed as follows:

$$SC(t) = a_2 \exp(b_2 t), \quad (20)$$

where values a_2 , b_2 are constant.

The $\tilde{R}(t)$ function represents a series of independent identically distributed random variables (iid). For the Gaussian distribution case, we consider $\tilde{R}(t) \sim \mathcal{N}(0, 1)$, and for the Student's t case, we have $\tilde{R}(t) \sim t(\nu)$ where ν is the parameter called the number of degrees of freedom.

4.2. Results for simulated signal

In this subsection, we apply our proposed methodology to simulated data based on Subsection 4.1. To show the performance of the proposed method and make a comparison with other filters from the KF family, we applied the classic EKF and UKF as well-known modifications of the classical KF for non-linear state estimation. For more details on EKF and UKF we encourage the reader to read [43, 44].

4.2.1. Degradation process in the presence of the Gaussian noise

The simulated degradation curve (based on the previous subsection) for the last regime in the presence of Gaussian noise and exponential time-varying scale is shown in Fig. 5. Panel (a) presents the health index plus deterministic trends. The random component is demonstrated in panel (b), as it can be seen; the scale of the random component grows exponentially over time. In the end, in panels (c) and (d), the scale and generated Gaussian noise used for the simulated degradation curve are shown, respectively.

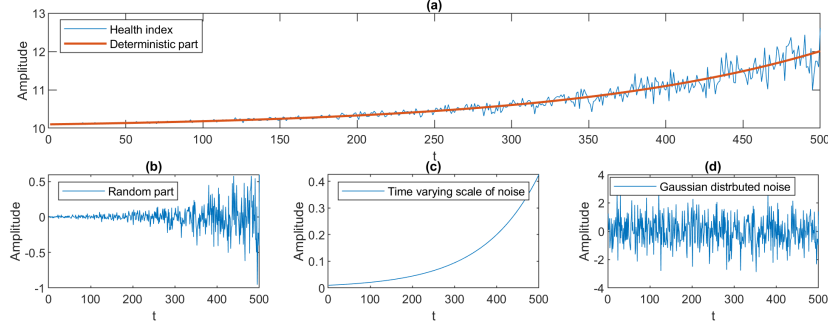


Figure 5: Simulated health index in the presence of Gaussian noise.

The final results of applying the proposed method to the simulated data are presented in Fig. 6. The top panel illustrates the degradation curves and the EOL level. The bottom panel shows the estimated mean of RUL using UKF, EKF, and the proposed MCEKF. The black dash line and the purple area show the real RUL value and $\pm 20\%$ accuracy bound (for each data point (in this paper is shown by t) leading up to the end of life, the difference between the EOL point and that data point is calculated. This difference represents the real RUL at that point in time, which is shown in the figure with a black dashed line for all points after FPT). Additionally, it is important to note that the time sequence in the lower panel originates from the FPT. We used data from the FPT onwards to the specific time point t to iteratively refine the model and derive accurate estimates of the RUL.

As could be expected, when the number of data is less, the accuracy of the estimated RUL by using all filters is not acceptable, and according to the bottom panel with UKF and EKF the RUL is underestimated, and with MCEKF the RUL is overestimated. After a while, around $t = 180$, it can be seen that MCEKF is approaching the accuracy bound and approximately remains in the accuracy bound until the end of the life, while the performance of UKF and EKF has improved with time. The results are close to the accuracy bound. However, the estimated RULs until $t = 430$ are not entirely within the accuracy bound, and after that, the performance is approximately equal to the MCEKF results and acceptable. Furthermore, it should be mentioned that the kernel size for MCEKF was determined as $\sigma = 0.5$ through a trial and error process, which serves as a crucial hyperparameter for the MCEKF algorithm. The selected σ value lies within the range of zero to one, and it is noteworthy that when the σ value is selected much big, the trends of both the MCEKF and the EKF align. It is important to highlight that selecting a lower value of σ enhances the robustness of the model against outlier data points.

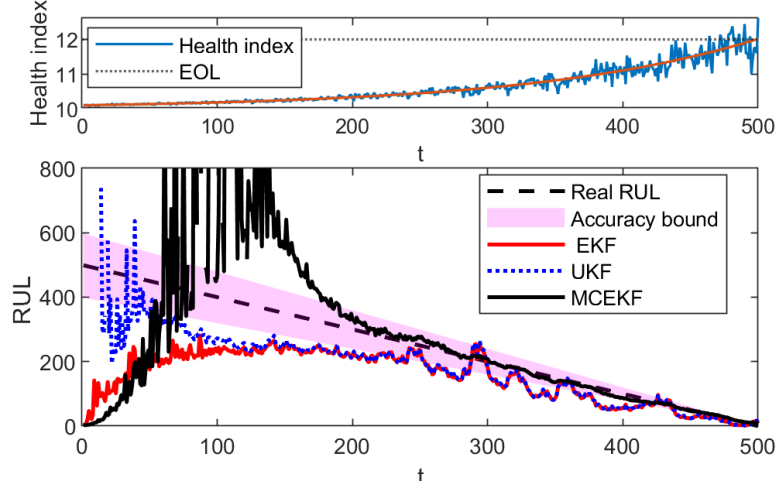


Figure 6: Final results of predicted RUL in the presence of the Gaussian noise, top panel: health index plus EOL, bottom panel: predicting RUL results performed by EKF, UKF, and proposed MCEKF and 20% accuracy bound.

The probability density function (PDF) portraying the estimated RUL values of the MCEKF technique is shown in Fig. 7. Evidently, in the initial stages, where the data set contains an insufficient number of data points, the PDF of the estimated RUL does not yield favourable outcomes. However, a discernible improvement becomes apparent as the size of the data set increases, allowing the PDF to progressively enclose the real RUL. In particular, beyond the temporal threshold of $t = 200$, the mean of the PDF closely aligns with the actual RUL values. Moreover, the 90% confidence interval (CI) related to the estimated RUL, derived from 100 simulated trajectories using the MCEKF model, is presented in Fig. 8. As it can be seen; after $t = 200$, the RUL values are in the scope of the 90% CI. This notable alignment between the real RUL and the 90% CI serves as evidence for the efficacy of our proposed model, especially in the context of its robustness against Gaussian noise influence.

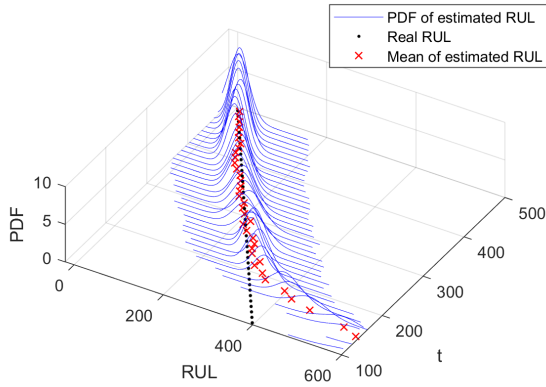


Figure 7: PDF of estimated RUL in the presence of the Gaussian noise.

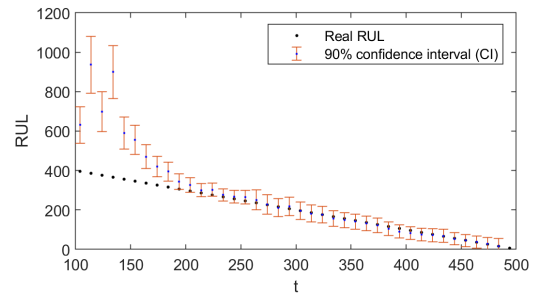


Figure 8: The 90% CI of estimated RUL by using MCEKF for simulated health index in the presence of Gaussian noise.

4.2.2. Degradation process in the presence of the non-Gaussian noise

The simulated degradation curve for the last regime based on Subsection 4.1 in the presence of non-Gaussian noise (Student's t with $\nu = 2.1$) and a time-varying exponential scale is shown in Fig. 9. Panel (a) presents the health index plus deterministic trends. The random component is demonstrated in panel (b), as it can be seen that the scale of the random component is exponentially growing over time. In the

end, in panels (c) and (d), the scale and generated non-Gaussian noise used for the simulated degradation curve are shown.

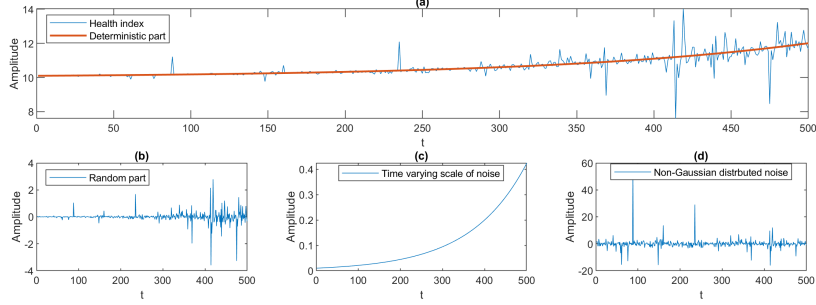


Figure 9: Simulated health index in the presence of non-Gaussian noise.

The final results of applying the proposed method to the simulated data are shown in Fig. 10. The top panel shows the degradation curves and the level of EOL. In the top panel, a few strong outliers (related to non-Gaussian noise) around $t = 240, 380, 430, 480$ are present. It can be expected that they may affect the non-robust methods. The bottom panel shows the estimated RUL using EKF, UKF, and the proposed MCEKF. The black dashed line and the purple area show the real RUL value $\pm 20\%$ accuracy bound to predict RUL. As previously, when the number of data is less, the accuracy of the estimated RUL for all filters tested is low, according to a bottom panel EKF and UKF underestimated RUL and MCEKF overestimated RUL. After a while, around $t = 150$, it can be seen that MCEKF is near the accuracy bound and approximately remains in the accuracy bound until the end of the life, while the performance of UKF and EKF has improved over time. The results are close to the accuracy bound. However, until $t = 180$ the obtained trajectories are not located in the accuracy bound and after that, the performance is going to be better, however, they are much worse in comparison to the MCEKF result. Also, it can be clearly seen in the case where the noise has non-Gaussian characteristics (the top panel), the performance of the UKF and EKF are dramatically reduced (see the circles in top and bottom panel). Furthermore, when examining the results from the UKF and EKF, around $t = 430$, it is evident that this outlier contributes to the diverges, leading the filter to provide results with negative amplitude. This deviation ultimately impacts the RUL estimation, suggesting an increase towards infinity at that specific time. At the same time, the proposed MCEKF could handle non-Gaussian noise problems and remains in the accuracy bound. It should be noted that for this case, we selected the kernel size of MCEKF equal to $\sigma = 0.45$ by trial and error as in the previous case.

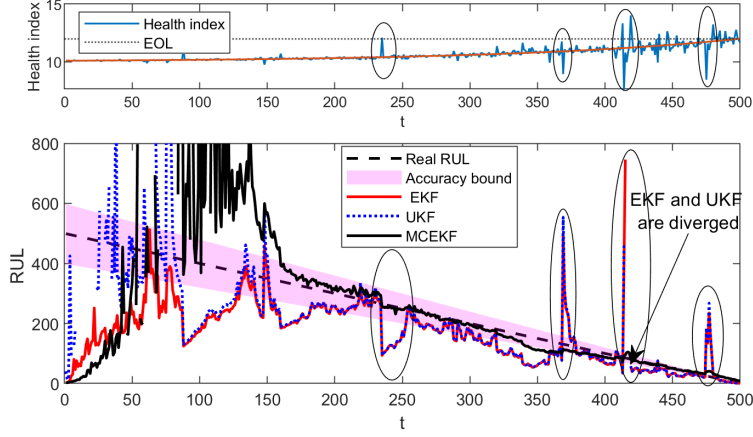


Figure 10: Final results of predicted RUL in the presence of the non-Gaussian noise, top panel: health index plus EOL, bottom panel: the mean of the predicted RUL performed by EKF, UKF, and proposed MCEKF and 20% accuracy bound.

The PDF portraying the estimated RUL values from the MCEKF technique is shown in Fig. 11. As in the previous case in the initial stages, where the data set contains an insufficient number of data points, the PDF of the estimated RUL does not yield favourable outcomes. However, a discernible improvement becomes apparent as the data set size increases, allowing the PDF to progressively enclose the real RUL. Furthermore, the 90% CI related to the estimated RUL, derived from 100 simulated trajectories using the MCEKF model, is presented in Fig. 12. As it can be seen; after $t = 200$, the RUL values are within the scope of the 90% CI. This notable alignment between the real RUL and the 90% CI serves as evidence for the efficacy of our proposed model, notably in the context of its robustness against non-Gaussian noise influence.

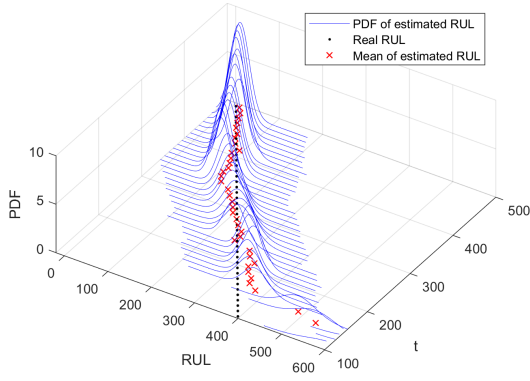


Figure 11: Estimate PDF of RUL in the presence of the non-Gaussian noise.

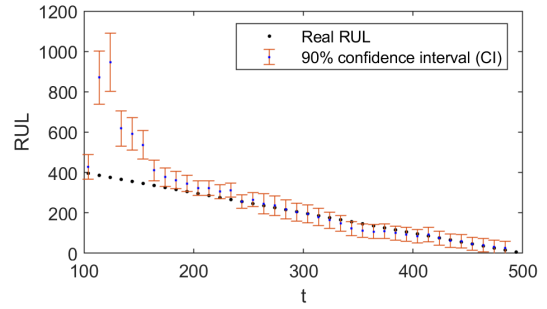


Figure 12: The 90% CI of estimated RUL by using MCEKF for simulated health index in the presence of non-Gaussian noise.

To confirm the performance of the proposed method, we repeated this procedure for 100 simulated cases in the presence of noise with Gaussian characteristics and presented the results using box plots, see Fig. 13. We used the mean absolute error (MAE) to compare the performance of the methods; see Eq. (21). It should be noted that a lower MAE value means better performance.

$$MAE = \frac{1}{N} \sum_{t=1}^N |RUL(t) - \widehat{RUL}(t)|, \quad (21)$$

where N is the number of all the data points in which the RUL is calculated, $\widehat{RUL}(t)$ is the estimated RUL at the time (data point) t , and $RUL(t) = N - t$ is the real RUL.

As it can be seen in Fig. 13, the median of the MAE for MCEKF has a lower value; also, the first and third quantiles of MCEKF have fewer MAE, which means that it worked better to predict RUL. After that, EKF and UKF have a lower median value for MAE, respectively. It should be noted that the kernel size for MCEKF is elected equally to $\sigma = 0.81$.

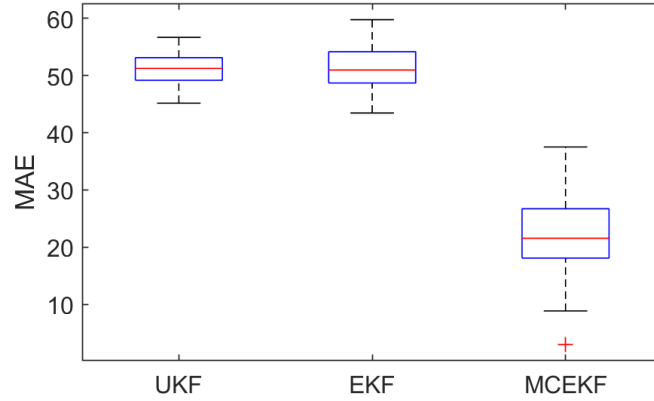


Figure 13: MAE based on 100 simulated cases in the presence of the Gaussian noise.

To confirm the performance of the proposed method, we repeated this procedure for 100 simulated cases in the presence of non-Gaussian noise (Student's t with $\nu = 2.1$) and presented the results using box plots, see Fig. 14 as previously. As it can be seen in Fig. 14, the median of the MAE for MCEKF has a lower value; also, the first and third quantiles of MCEKF have a lower MAE, meaning that it worked better to predict RUL. After that EKF and UKF have a lower median value for MAE, respectively. It should be noted that the kernel size for MCEKF is elected equally to $\sigma = 0.56$.

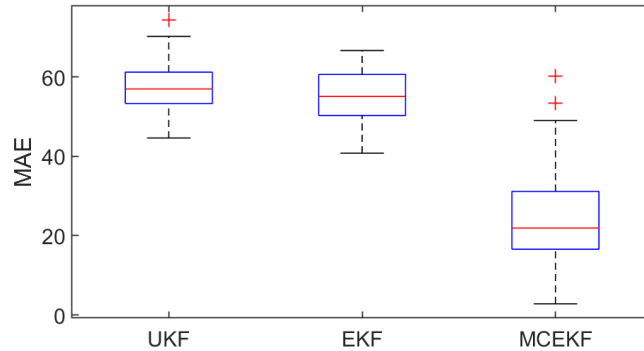


Figure 14: MAE based on 100 simulated cases in the presence of the non-Gaussian noise (Student's t distributed noise with $\nu = 2.1$).

Furthermore, to clear up the impact of varying levels of non-Gaussianity on the estimated RUL derived from these filters, we have conducted a simulation involving 100 trajectories for different levels of the degree

of freedom ν of Student's t distribution. Subsequently, these methodologies were employed for the estimation of the RUL, and the MAE was computed by contrasting the actual RUL with the mean of the predicted RUL. The results of these calculated median MAE values are illustrated in Fig. 15. It is evident from the results that the MCEKF consistently exhibits lower MAE values across different degrees of non-Gaussian noise. This observation underscores the robustness of MCEKF when faced with non-Gaussian noise scenarios.

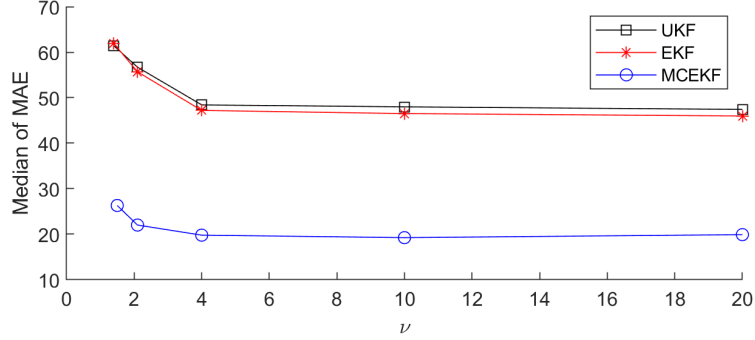


Figure 15: Median of MAE based on 100 simulated cases in the presence of different levels of the non-Gaussian noise (Student's t distributed noise with different level of the ν value).

The procedure was replicated on various kernel sizes, denoted as σ , to assess its impact on the results. The median of the MAE was calculated for 100 simulated trajectories in the presence of Gaussian noise. The results of the calculations are presented in Fig. 16. As can be seen, the median of MAE has the lowest value when the $0.3 < \sigma < 0.5$. However, this growing median of MAE is not substantial enough to infer the high sensitivity of the proposed method to σ in the presence of Gaussian noise, particularly when for the high value of σ . On the other hand, we repeated the previous procedure in the presence of non-Gaussian noise (Student's t with $\nu = 2.1$). The results are shown in Fig. 17. As can be seen, the median of MAE has the lowest value when the $0.3 < \sigma < 0.5$. Also, we should note that the median of MAE is smaller for Gaussian, and for non-Gaussian is growing faster. It means that the selection of σ for non-Gaussian is more critical (which seems to be obvious).

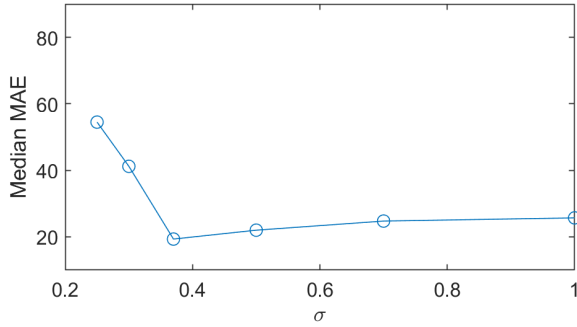


Figure 16: Median of MAE based on 100 simulated cases in the presence of Gaussian noise with different kernel size σ for MCEKF.

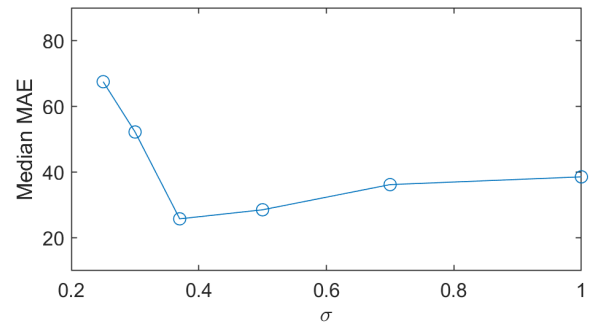


Figure 17: Median of MAE based on 100 simulated cases in the presence of non-Gaussian noise (Student's t distributed noise with $\nu = 2.1$) with different kernel size σ for MCEKF.

5. Result for real data sets

In this section, we explain the results that arise from the implementation of the methodology suggested in real-world data sets. Specifically, we investigated the efficacy of our approach in three distinct data

sets. The first and second data sets originate from the FEMTO and IMS benchmarks, which are renowned for their extensive use as a benchmark in various studies. In particular, these data sets present evidence of a Gaussian distribution. The last data set under scrutiny pertains to wind turbine bearing data and notably encompasses solid components that do not stick to the Gaussian distribution. The existence of such non-Gaussian behaviour introduces an additional layer of complexity to our investigation, warranting an in-depth analysis of the model performance in such scenarios. Also, note that for each data set, all initial parameters of the filters were selected as the same values.

5.1. Data sets description

5.1.1. FEMTO data set

This data set was made publicly available during the IEEE International conference on PHM in 2012 and was generously provided by the Franche-Comté Electronics Mechanics Thermal Science and Optics-Sciences and Technologies Institute [45]. Comprising 17 run-to-failure data sets of rolling element bearings acquired from the PRONOSTIA platform, see Fig. 18, this data set was utilised to perform accelerated degradation tests, effectively simulating several years' worth of bearing wear in a few hours. This was achieved by applying a high-level radial force that exceeded the maximum dynamic load of the bearings. Throughout the tests, the rotational speed of the bearings remained constant. Data acquisition was facilitated using two accelerometers and a thermocouple, capturing vibration signals and bearing temperatures. Bearing failure was deemed to have occurred when the amplitude of the vibration signal exceeded 20 g.

Notably, this data set represents natural-occurring bearing degradation, devoid of any pre-seeded faults or artificial interventions. Compounding the complexity of the data set is the limited availability of only two training units for each operational condition. Furthermore, diverse fault patterns and lifetimes characterise different units, even when subjected to the same conditions. This heterogeneity between the training and testing units exacerbates the challenge of predicting RUL.

Vibration signals within the data set exhibit a relatively low frequency resolution, as each sample spans a time length of 0.1 s, resulting in a frequency resolution of 10 Hz. Consequently, conventional fault diagnostic methods that rely on detailed frequency analysis are not suitable for this data set.

Furthermore, the data set exemplifies the propensity of incipient faults in one component to propagate to other components through frequent contact, resulting in the simultaneous occurrence of various fault patterns, as visually represented in Fig. 19. This distinctive behaviour underscores the suitability of the data set for addressing multi-fault prognostic challenges.

This data set is commonly used for health index construction [46–53], health stage evaluation [2, 54–58], and predicting RUL [59–64].

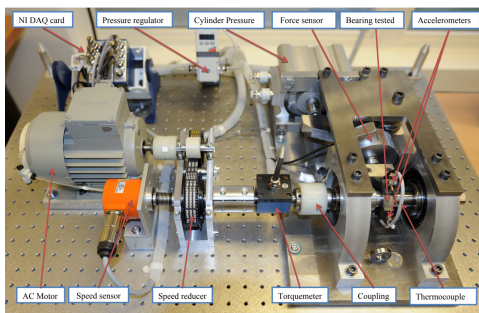


Figure 18: FEMTO test rigs [45].



Figure 19: Different faults of rolling element bearings [45].

FEMTO case study

In this paper, we apply our proposed methodology to estimate the RUL of the bearing labeled as 1–1 within the FEMTO data set. The data in this data set have been gathered under specific operating conditions, precisely at 1800 rpm and with an applied load of 4000 N, as shown in Fig. 20.

For each set of vibration data, we calculate the RMS, which is subsequently used as the health index for our analysis.

Our choice of this specific case study was predicated on the compelling characteristics that it exhibits, rendering it an exemplary illustration of a three-stage degradation model, marked by predominantly monotonic behaviour. Furthermore, as indicated in [65], the random component of this case study closely approximates a Gaussian distribution. This aspect lends substantial support to our objective of demonstrating the efficacy of our proposed model in a real world scenario where the underlying distribution aligns with a Gaussian distribution.

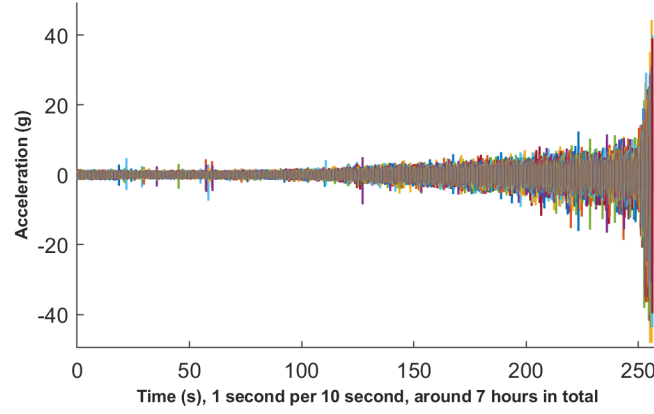


Figure 20: Raw vibration data for case study bearing 1–1 .

5.1.2. IMS data set

The data set under consideration was provided by the centre for intelligent maintenance systems (IMS) at the University of Cincinnati [66], and it has been made available through the NASA prognostic data repository of NASA [67]. This data set comprises three distinct subsets of bearing degradation tests. Each test involved the installation of four Rexnord ZA-2115 double-row bearings on a common shaft, see Fig. 21. Accelerometers were placed on bearing housings to capture vibration signals and an oil circulation system was engineered to ensure bearing lubrication. For debris collection, a magnetic plug was integrated into the oil feedback pipe. The execution of the test was designed to be adaptive, with an electrical switch that triggers cessation once the accumulation of debris exceeded a predetermined threshold. Following the conclusion of these tests, the bearings underwent a detailed examination and the observed fault patterns were meticulously recorded.

Vibration signals exhibit high frequency resolution, making them well suited for fault diagnosis through frequency analysis techniques. According to the data instructions [67], each file consists of 20,480 data samples sampled at 20 kHz. Consequently, operators can derive valuable frequency-domain features to monitor the degradation progression of specific bearing components, such as rollers, outer race, and inner race. At the conclusion of the tests, detailed fault patterns for each bearing were observed and recorded, as depicted in Fig. 22 offering a rich source of information for researchers to explore the relationships between different fault patterns and their corresponding degradation trends.

Moreover, note that this data set has been extensively used in various publications such as segmentation [68, 69], RUL prediction [70–74], and condition monitoring [66].

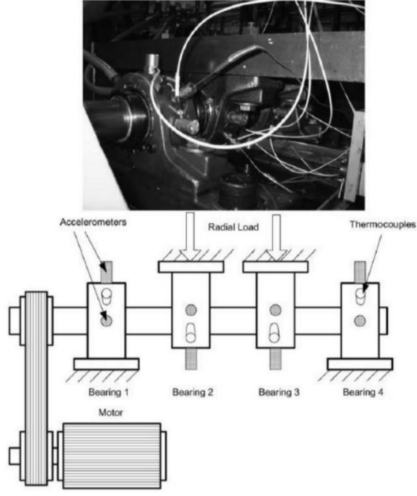


Figure 21: IMS test rig [66].

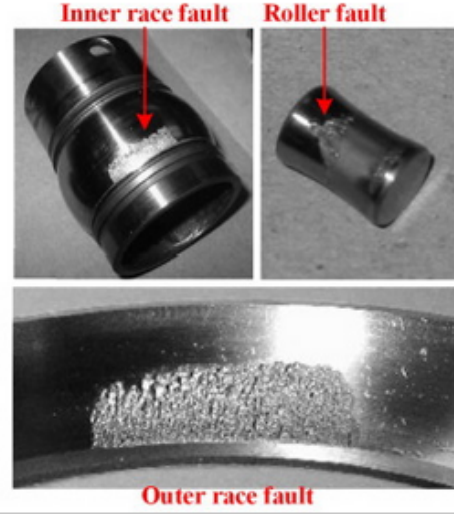


Figure 22: Different faults of rolling element bearings [66].

IMS case study

In this part, we employ the IMS data set, which comprises three distinct subsets. For our study, we specifically focus on Subset 3 and select the vibration of bearing number 3 as our case study, see Fig. 23. Subset 3 captures data over a recording duration spanning from April 4, 2004, at 09:27:46 to April 4, 2004, at 19:01:57. This data set encompasses vibration data, which were collected from four channels. The channel arrangement is as follows: bearing1-channel1; bearing2-channel2; bearing3-channel3; bearing4-channel4. Data are recorded at a 10-minute interval in ASCII file format. The experiment concludes with an outer race failure event on bearing number 3, see Fig. 22.

Our choice of this particular case within the IMS data set was underpinned by its distinctive features, most notably the remarkably brief transition between the degradation and critical stages, which might even warrant classification as a two-stage model rather than a multistage model. This specific case also provides an additional scenario where the noise distribution closely approximates a Gaussian distribution, further enhancing its relevance for our study.

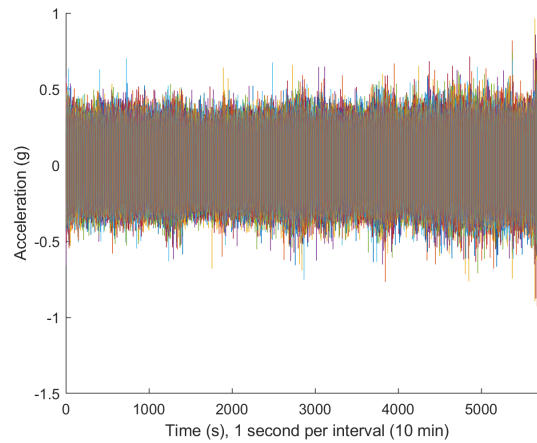


Figure 23: Raw vibration data for IMS subset 3 bearing 3 .

5.1.3. Wind turbine data set

The wind turbine data set comprises sensor data collected from the high-speed bearing shaft of a 2.2 MW wind turbine, visually depicted in Fig. 24. This data set encompasses measurements of the bearing's inner race energy, sampled at 10-minute intervals over a period of approximately 50 days. In particular, the bearing under investigation shows an inner race fault and is identified as SKF 32222 J2 tapered roller bearings. This particular tapered roller bearing (TRB) boasts a 200 mm outer diameter, a bore of 110 mm, and a total length of 56 mm, featuring 20 rolling elements set at a 16° taper angle and weighing approximately 20 pounds. The bearing in question is supported by two pillow blocks and is equipped with a load cell to measure the force applied to the bearing. The wind turbine operates at variable speeds ranging from 2 to 100 Hz, with a load cell capable of measuring loads up to 1000 pounds, although most tests were conducted at 150 or 300 pounds. To mitigate the risk of catastrophic gearbox failure, the highest applied test load was maintained at 50% of the rated power. Further comprehensive information on this data set can be found in [75].

Remarkably, this data set, which we refer to as the wind turbine data set, has received considerable attention in recent years as a topic of interest for prognostics of RUL in various research endeavours, exemplified by works such as [75–78].

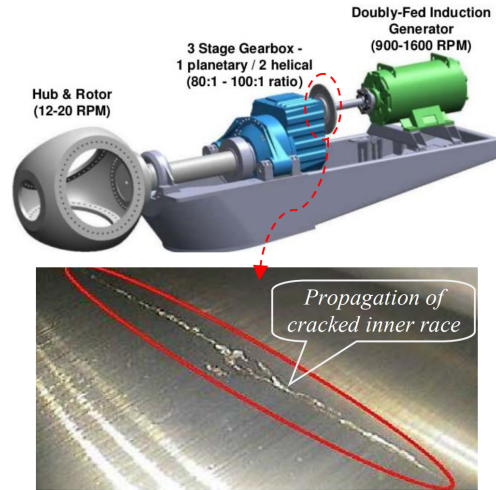


Figure 24: Wind turbine test rigs [75].

Wind turbine data set case study

For the application of our proposed approach to the wind turbine data set, the inner race energy is chosen judiciously as the representative health index. For more details on the procedure of extraction of health indexes, please see [79]. In particular, as shown in Fig. 37, this data set is characterised by a conspicuous presence of outliers. Consequently, it is prudent to categorise this data set as possessing a trend punctuated by non-Gaussian noise.

Furthermore, the health index exhibits fluctuations, which can be attributed to dynamic factors such as load variations or even phenomena such as self-healing. This complex behaviour, observed within real-world scenarios, presents a big challenge when it comes to the tasks of segmentation and RUL prediction. The presence of such intricate dynamics underscores the complexity inherent in accurately characterising and forecasting RUL in practical applications.

5.2. Results

5.2.1. Results for FEMTO data set

In this part, the proposed method is used to estimate the RUL of the The FEMTO data set. Panel (a) in Fig. 25 shows the complete degradation curves for this case study. Also, the small window shows the area after FPT, which is used to estimate RUL. It should be noted that several papers focused on detecting FPT in the literature, including our previous papers [2, 12, 38, 39], in which we used different methods. In this article, we utilise the results of [12] that are shown in panel (b) in Fig. 25. Therefore, here we focus on the estimation of the RUL that is starting from this point that is already identified.

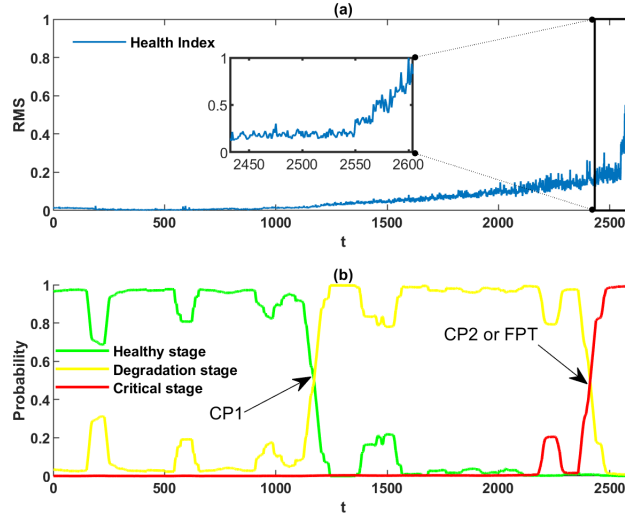


Figure 25: (a) Health index (RMS) of FEMTO data set bearing1-1 (b) FPT detected that is provided in this [12].

Fig. 26, shows the results of applying the proposed approach to the FEMTO data set. Top panel illustrates the degradation curves and the level of EOL. The bottom panel shows the estimated results using EKF, UKF, and the proposed MCEKF. The black dashed line and the purple area show the real RUL value and the $\pm 20\%$ accuracy bound for predicting RUL. As expected, when the number of data is small, the accuracy of the estimated RUL by using all filters is not acceptable. Based on bottom panel, EKF, UKF, and MCEKF first overestimated RUL. After a while, around $t = 30$, it can be seen that all methods are reaching the accuracy bound and remain approximately inside until the end of life. In addition, the selected parameters for MCEKF are presented in Table 1 and the parameters of EKF and UKF are selected as the same value as MCEKF.

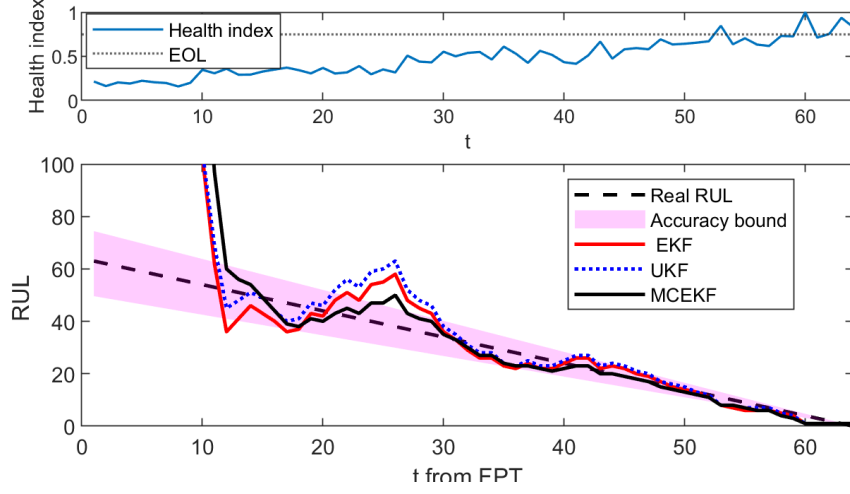


Figure 26: Final results of estimated RUL for FEMTO data set, top panel: health index plus EOL, bottom panel: predicted RUL results performed by EKF, UKF, and proposed MCEKF and 20% accuracy bound.

The PDF showing the estimated RUL values from the MCEKF technique is shown in Fig. 27. Evidently, in the initial stages, where the data set contains an insufficient number of data points, the PDF of the estimated RUL does not yield favourable results. However, a discernible improvement becomes apparent as the data set size increases, allowing the PDF to progressively enclose the RUL distribution. In particular, beyond the temporal threshold of $t = 30$, the mean of the PDF closely aligns with the actual value of the RUL. Furthermore, the 90% CI related to the estimated RUL, derived from 100 simulated trajectories using the MCEKF model, is presented in Fig. 28. As can be seen; after $t = 30$, the RUL values are within the scope of the 90% CI.

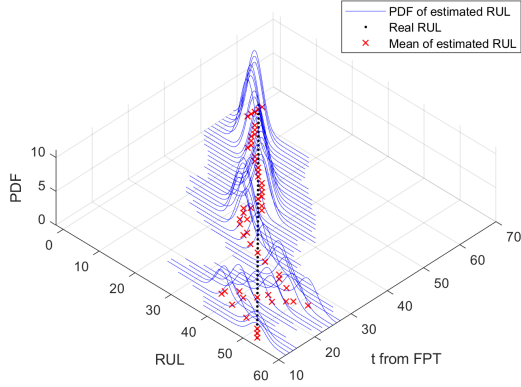


Figure 27: PDF of estimated RUL by using MCEKF for FEMTO data set.

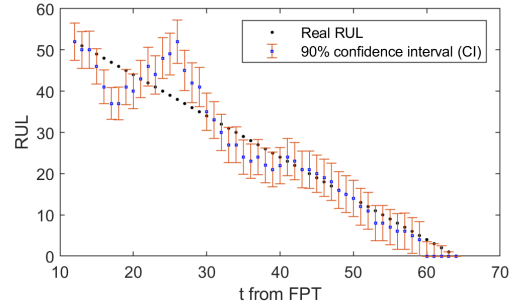


Figure 28: The 90% CI of estimated RUL by using MCEKF for FEMTO data set.

To illustrate the sensitivity of our proposed methodology to the hyperparameter σ (kernel bandwidth for the MCEKF), we conducted an analysis employing different ranges of σ . The results related to the estimated RUL are shown in Fig. 29. In particular, in this scenario, the selection of a higher value of σ yields more accurate results compared to opting for a lower value σ . To reinforce this observation, we present the MAE computed between the real RUL and the estimated RUL using our proposed method in Fig. 30, further affirming the findings from the preceding Fig. 29.

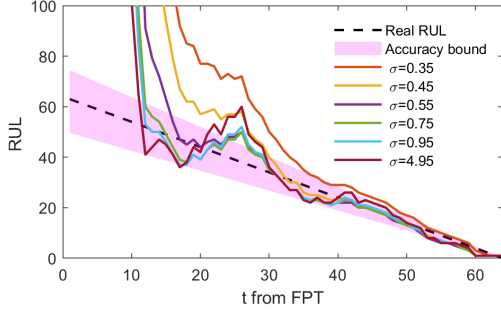


Figure 29: Predicted RUL results performed by proposed MCEKF with different value σ and 20% accuracy bound.

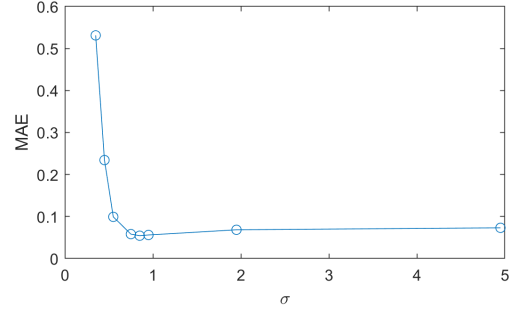


Figure 30: MAE between the predicted RUL performed by proposed MCEKF with different value σ and real RUL.

Table 1: Selected parameters value for MCEKF in case of FEMTO data set.

MCEKF parameters	std of Q_t	std of M_t	σ
FEMTO data set	[0.00001, 0.00001, 0.00001]	0.0038	0.87

5.2.2. Results for IMS data set

In this part, the proposed method is used to estimate the RUL of IMS data set. Panel (a) in Fig. 31 shows the complete degradation curves for this case study. Also, the small window shows the area after FPT, which is used to estimate RUL. It should be noted that several papers focused on the detection of FPT in the literature, including our previous papers [2, 12, 38, 39], in which we used different methods to identify this point. So in this paper we used the results of [12] that is shown in panel (b) in Fig. 31. Therefore, this paper only focused on the estimation of RUL from this point.

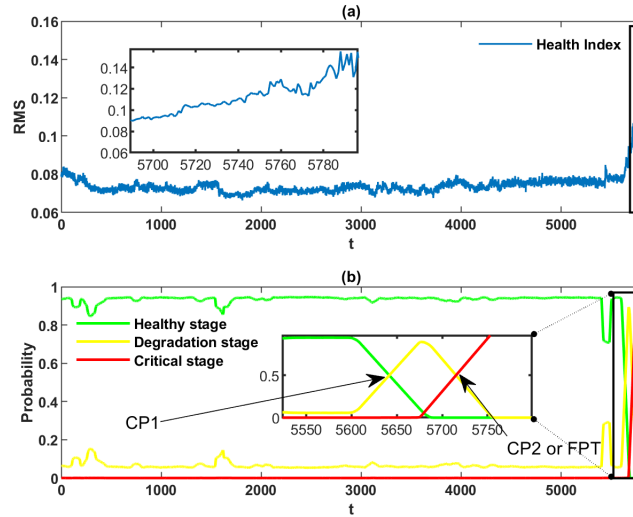


Figure 31: (a) Health index (RMS) of IMS data set subset 3 bearing3 (b) FPT detected by using a method that is introduced in this [12].

The results of applying all of the methods on the IMS data set are presented in Fig. 32. Top panel illustrates the degradation curves and EOL level. Bottom panel shows the estimated results using EKF, UKF, and proposed MCEKF. The black dash line and the purple area show the real RUL value $\pm 20\%$

accuracy bound for predicting RUL. As expected, when the number of data is small, the accuracy of the estimated RUL by using all filters is not acceptable. Based on bottom panel, EKF, UKF, and MCEKF overestimated RUL at first. After a while, around $t = 30$, it can be seen that MCEKF is coming to the accuracy bound and remaining approximately in the accuracy bound until the end of the life, while we can see that UKF and EKF could not provide accurate results. Furthermore, we provide the details of the parameters chosen for the MCEKF in Table. 2. It is important to note that, for the fair comparative analysis, the parameters of the EKF and the UKF have been set to the same values as those used in the MCEKF.

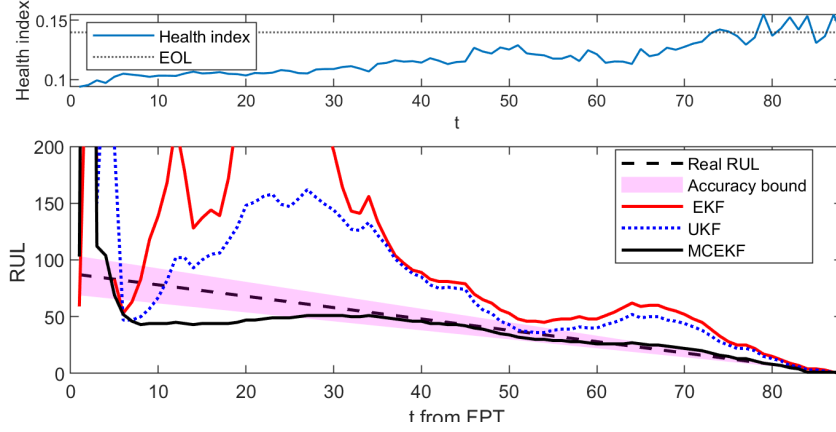


Figure 32: Final results of estimated RUL for IMS data set, top panel: health index plus EOL, bottom panel: predicted RUL results performed by EKF, UKF, and proposed MCEKF and 20% accuracy bound.

The PDF of the estimated RUL values obtained using the MCEKF technique is shown in Fig. 33. Evidently, in the initial stages, where the data set contains an insufficient number of data points, the PDF of the estimated RUL does not yield favourable results. However, a discernible improvement becomes apparent as the data set size increases, allowing the PDF to progressively enclose the real RUL. In particular, beyond the temporal threshold of $t = 30$, the mean of the PDF aligns closely with the actual RUL value. Furthermore, the 90% CI related to the estimated RUL, derived from 100 simulated trajectories using the MCEKF model, is presented in Fig. 34. As it can be seen; after $t = 30$, the RUL values are in the scope of the 90% CI.

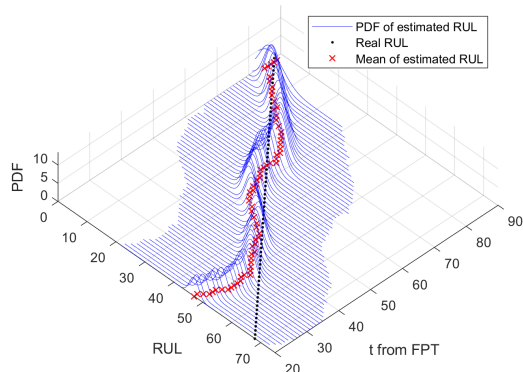


Figure 33: PDF of estimated RUL by using MCEKF for IMS data set.

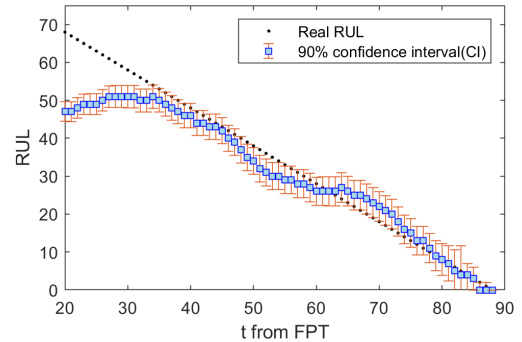


Figure 34: The 90% CI of estimated RUL by using MCEKF for IMS data set.

To illustrate the sensitivity of our proposed methodology to the hyperparameter σ (kernel bandwidth for the MCEKF), we conducted an analysis employing different ranges of σ . The results pertaining to the estimated RUL are shown in Fig. 35. In particular, in this scenario, the selection of a σ value between 0.25 – 0.4 yields more accurate results compared to opting for a lower or higher value σ . To reinforce this observation, we present the MAE calculated between the real RUL and the estimated RUL using our proposed method in Fig. 36, further affirming the findings of the preceding Fig. 35.

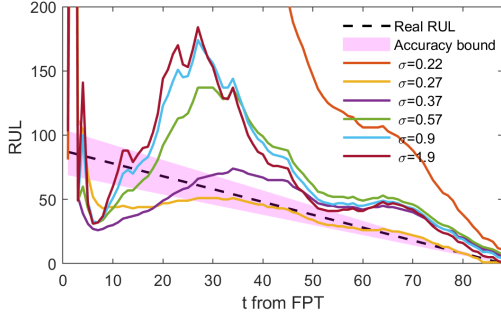


Figure 35: Predicted RUL results performed by proposed MCEKF with different value σ and 20% accuracy bound.

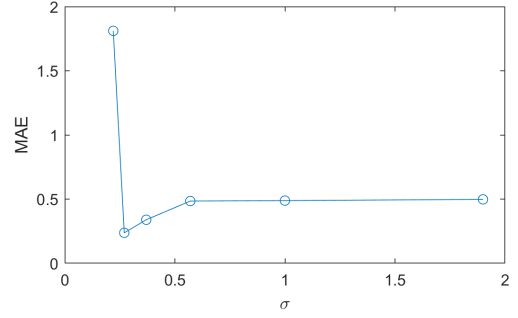


Figure 36: MAE between the predicted RUL performed by proposed MCEKF with different value σ and real RUL.

Table 2: Selected parameters value for MCEKF in case of IMS data set.

MCEKF parameters	std of Q_t	std of M_t	σ
IMS data set	[0.00001, 0.00001, 0.00001]	0.0038	0.27

5.2.3. Results for wind turbine data set

In this part, the proposed method is used to estimate the RUL of the wind turbine data set. Panel (a) in Fig. 37 shows the complete degradation curves for this case study. Also, the small window shows the area after FPT, which is used to estimate RUL. It should be noted that several papers focused on detecting FPT in the literature for this data set, including our previous papers [2, 12, 38, 39], in which we used different methods. In this article, we utilise the results of [12] that are shown in panel (b) in Fig. 37. Therefore, here we focus on the estimation of the RUL that is starting from this point

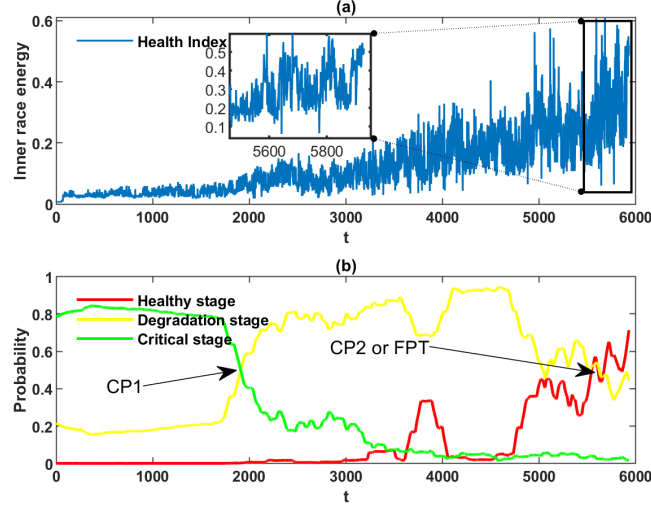


Figure 37: (a) Wind turbine health index (inner race energy), (b) FPT detected that is provided in this [12].

Fig. 38, shows the results of applying all the methods to the wind turbine data set. Top panel illustrates the degradation curves and the level of EOL. Bottom panel shows the estimated results using EKF, UKF, and the proposed MCEKF. The black dash line and the purple area show the real RUL value $\pm 20\%$ accuracy bound to predict RUL. As expected, for scenarios where the data set is limited, the RUL estimates by all filtering methodologies exhibit sub-optimal accuracy. Analysing bottom panel reveals initial results for both EKF, UKF, and MCEKF to overestimate RUL. In particular, a noticeable convergence towards the accuracy bond can be discerned around $t = 180$. However, it is imperative to acknowledge that due to the non-linear trajectory of the health index, achieving precise results remains a challenge.

Subsequently, the results of all methodologies as time progresses are shown. As we can seen; after $t = 300$, the mean RUL estimated by MCEKF closely aligns with the accuracy bound, while the other results produced by EKF and UKF exhibit less accurate results. It should be noted that around $t = 310$ and $t = 440$, the RUL estimates derived from EKF and UKF appear to be susceptible to the influence of non-Gaussian noise. In contrast, the MCEKF, as anticipated, reduced the impact of such non-Gaussian noise. In this study, we have carefully selected the parameters for the MCEKF applicable to the case study at hand. These parameter choices are detailed in Table. 3. Importantly, we would like to highlight that, for consistent comparison, the measurement noise and process noise parameters have been set to the same values for the EKF, the UKF, and the MCEKF.

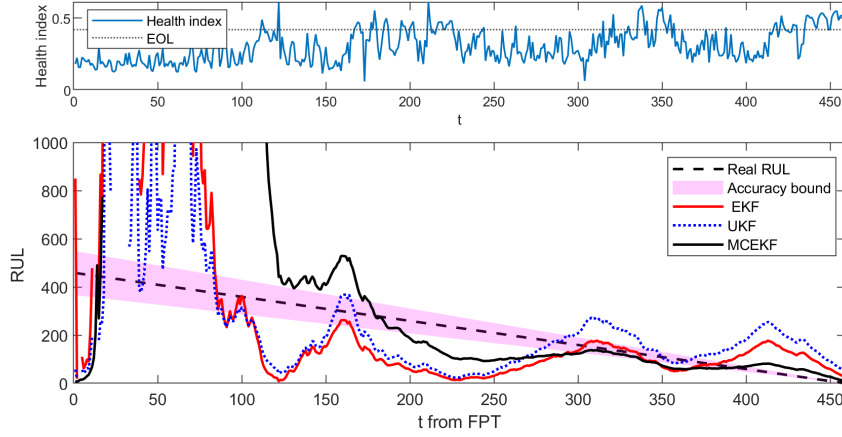


Figure 38: Final results of estimated RUL for wind turbine data set, top panel: health index plus EOL, bottom panel: predicted RUL results performed by EKF, UKF, and proposed MCEKF and 20% accuracy bound.

The PDF of the estimated RUL values from the MCEKF approach is shown in Fig. 39. As in previous cases, in the initial stages, where the data set contains an insufficient number of data points, the PDF of the estimated RUL does not yield favourable outcomes. However, a discernible improvement becomes apparent as the data set size increases, allowing the PDF to progressively enclose the RUL distribution. In particular, beyond the temporal threshold of $t = 250$, the PDF closely matches the actual RUL value. Moreover, the 90% CI related to the estimated RUL, derived from 100 simulated trajectories using the MCEKF model, is shown in Fig. 40. As it can be seen; after $t = 260$, the RUL values are in the scope of the 90% CI. This notable alignment between the real RUL and the 90% CI serves as evidence for the efficacy of our proposed model, notably in the context of its robustness against non-Gaussian noise influence and non-linear trend of health index.

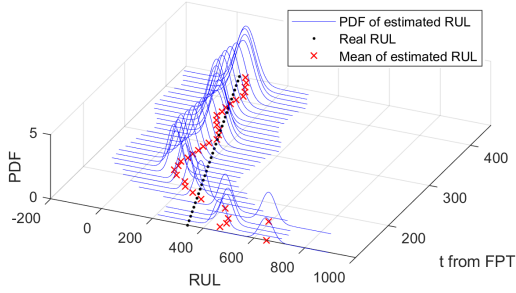


Figure 39: PDF of estimated RUL by using MCEKF for wind turbine data set.

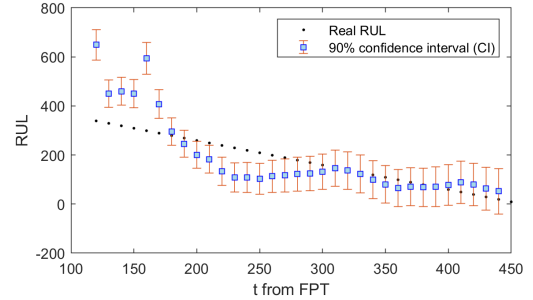


Figure 40: The 90% CI of estimated RUL by using MCEKF for wind turbine data set.

To illustrate the sensitivity of our proposed methodology to the hyperparameter σ (kernel bandwidth for the MCEKF), we conducted an analysis employing different ranges of σ . The results pertaining to the estimated RUL are shown in Fig. 41. In particular, in this scenario, the selection of a σ value between 0.25 – 0.4 yields more accurate results compared to opting for a lower or higher value σ . To reinforce this observation, we present the MAE calculated between the real RUL and the estimated RUL using our proposed method in Fig. 42, further affirming the findings of the previous Fig. 41.

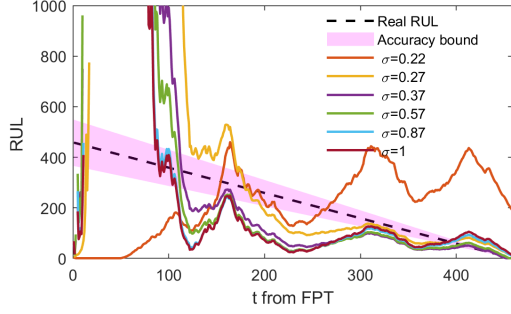


Figure 41: Predicted RUL results performed by proposed MCEKF with different value σ and 20% accuracy bound.

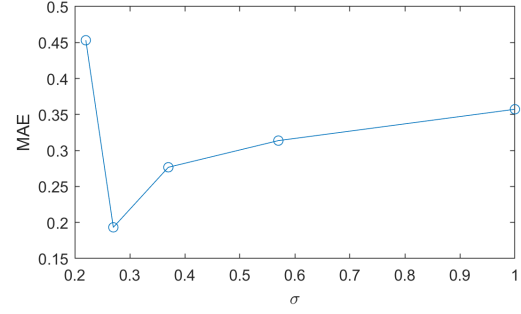


Figure 42: MAE between the predicted RUL performed by proposed MCEKF with different value σ and real RUL.

Table 3: Selected parameters value for MCEKF in case of wind turbine data set.

MCEKF parameters	std of Q_t	std of M_t	σ
wind turbine data set	[0.00001, 0.00001, 0.00001]	0.0265	0.28

5.3. Discussion

The results presented for three real data sets confirm the efficacy of our proposed models. In particular, the results obtained through the EKF, UKF, and proposed MCEKF for the FEMTO data set exhibit acceptable and satisfactory performance. This can be attributed to the predominantly monotonic behaviour of the data set. Furthermore, our earlier research endeavours have established that the noise distribution for this particular case study approximates a Gaussian distribution.

On the other hand, the observations derived from the wind turbine data set align with our expectations, showcasing a markedly distinct behaviour. This data set's dynamics is characterised by non-linear behaviour, devoid of the monotonic trends observed in the FEMTO data set or IMS data set. The non-monotonic behaviour can be attributed to factors such as load variations, self-healing phenomena, and changing environmental conditions. Notably, the presence of sparse spikes further reinforces the indication of a non-Gaussian noise distribution.

Nevertheless, the performance of our proposed method eclipses that of both EKF and UKF for the wind turbine data set, while the mean estimated RUL generated by all methods oscillates between overestimation and underestimation over time. It is crucial to note that RUL estimation is a multifaceted endeavour, not merely a single prediction point task. As exemplified by the wind turbine data set, while the mean estimated RUL may exhibit a disparity from the actual RUL, the crucial aspect is that the 90% CI successfully enclose the RUL range. This confirms the robustness of our proposed approach in the face of the data set's intricate dynamics and non-Gaussian noise influence.

For a complete understanding of the performance of the method, we present the normalised MAE between the actual RUL and the mean estimated RUL across all data sets using Eq. (21), as illustrated in Fig. 43. In particular, this figure demonstrates that the proposed MCEKF exhibits a significantly reduced MAE compared to the EKF and the UKF.

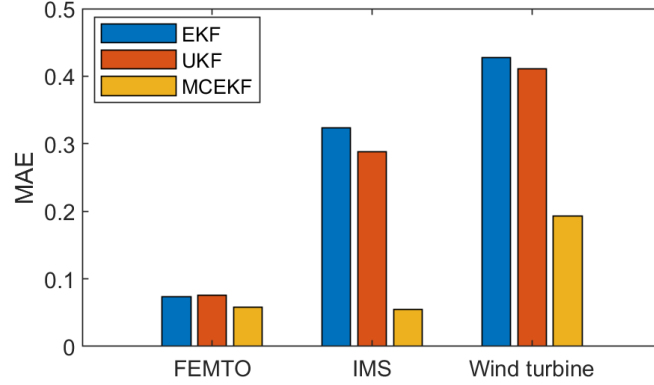


Figure 43: Calculated MAE for all data sets.

Indeed, while the results presented here underscore the substantial potential of our proposed model for RUL estimation, we remain cognisant of a pivotal distinction within the KF family. This family mostly deals with the deterministic part of the degradation process, where the PDF of estimated RUL mostly relates to uncertainties in the deterministic part of the degradation process.

In light of this understanding, our future endeavours will strategically align with this paradigm. We propose the development of a hybrid model that integrated the strengths inherent in the MCEKF with stochastic methodologies designed to specifically enclose the stochastic aspects of the degradation process. We expect that this integration would effectively leverage the precision of MCEKF’s deterministic estimation with the power of stochastic methods in modelling the random part and thereby will provide a more comprehensive and refined framework for RUL estimation. This approach may have the potential to further enhance the accuracy and robustness of RUL estimation while taking into account both deterministic and stochastic elements.

5.3.1. Computational cost

Table 4 demonstrates the computational cost of all methods for real data sets. The methods are implemented with Matlab 2022b, and the hardware properties of the system for this implementation are as follows: Processor: Intel (R) Core (TM) i7-10750H CPU @ 2.60GHz 2.59 GHz and Ram: 32.0 GB. As shown in Table 4, the computational costs of EKF, UKF and MCEKF are low. However, the UKF is the fastest method.

Table 4: Computational cost of methods.

	EKF	UKF	MCEKF
IMS data set	0.348865 (s)	0.319379 (s)	0.386541 (s)
FEMTO data set	0.792177 (s)	0.578108 (s)	1.303608 (s)
wind turbine data set	9.063914 (s)	4.871542 (s)	25.141721 (s)

6. Conclusions

In this study, we introduced a robust MCEKF approach, guided by the maximisation of correntropy criterion, to estimate the RUL of mechanical systems operating in the midst of noise possessing non-Gaussian characteristics. Most of the machines work in harsh environments. Usually, the data acquired from them is affected by noise with non-Gaussian characteristics, which can challenge traditional classic methods that assume that the noise has a Gaussian distribution; we have demonstrated the effect on the performance of the RUL application.

Furthermore, a health index data model is presented, spanning from FPT to EOL, incorporating an exponential deterministic trend and an exponential time-varying scale along with non-Gaussian noise (specifically, Student's t distribution). This model, inspired by our previous work, effectively captures the degradation process and enables the generation of artificial data sets for simulation purposes.

The efficacy of the proposed approach proves instrumental in addressing the challenges posed by non-stationary and non-Gaussian noise inherent in the process. Furthermore, to address the uncertainties inherent in the degradation process, a PDF of the estimated RUL is established.

In order to showcase the superiority of the proposed MCEKF-based approach, comprehensive comparative simulations are performed through Monte Carlo analysis. These simulations, conducted under conditions encompassing both Gaussian and non-Gaussian noise, consistently underscore the dominance of the MCEKF method in handling complex and challenging noise profiles.

The ultimate validation of the MCEKF method is carried out through its application to three benchmark real-world data sets. This real-world deployment serves a dual purpose: to confirm the effectiveness of the proposed approach in accurate RUL estimation and to confirm its applicability and efficacy within authentic operational environments. The method's demonstrated superiority through comparative simulations, coupled with its successful deployment on real-world data sets, attests to its potential as an invaluable tool in the realm of mechanical systems prognostics and health management. Furthermore, the computational cost of the proposed method is acceptable and has massive potential to be used for actual applications. The biggest drawback of this approach is that the measurement noise part of the processes has been tuned manually. In future work, the solution to address it will be investigated in detail.

Data Availability Statement

Conflicts of interest

The authors declare no conflict of interest.

Bibliography

- [1] A. Diez-Olivan, J. Del Ser, D. Galar, B. Sierra, Data fusion and machine learning for industrial prognosis: trends and perspectives towards industry 4.0, *Information Fusion* 50 (2019) 92–111.
- [2] F. Moosavi, H. Shiri, J. Wodecki, A. Wylomańska, R. Zimroz, Application of machine learning tools for long-term diagnostic feature data segmentation, *Applied Sciences* 12 (13) (2022) 6766.
- [3] X.-S. Si, W. Wang, C.-H. Hu, D.-H. Zhou, Remaining useful life estimation—a review on the statistical data driven approaches, *European Journal of Operational Research* 213 (1) (2011) 1–14.
- [4] Z.-S. Ye, M. Xie, Stochastic modelling and analysis of degradation for highly reliable products, *Applied Stochastic Models in Business and Industry* 31 (1) (2015) 16–32.
- [5] D. Kucharczyk, A. Wylomańska, J. Obuchowski, R. Zimroz, M. Madziarz, Stochastic modelling as a tool for seismic signals segmentation, *Shock and Vibration* 2016 (2016).
- [6] A. Heng, S. Zhang, A. C. Tan, J. Mathew, Rotating machinery prognostics: State of the art, challenges and opportunities, *Mechanical Systems and Signal Processing* 23 (3) (2009) 724–739.
- [7] M. S. Kan, A. C. Tan, J. Mathew, A review on prognostic techniques for non-stationary and non-linear rotating systems, *Mechanical Systems and Signal Processing* 62 (2015) 1–20.
- [8] L. Liao, F. Köttig, Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction, *IEEE Transactions on Reliability* 63 (1) (2014) 191–207.
- [9] Z. Zhao, J. Wu, T. Li, C. Sun, R. Yan, X. Chen, Challenges and opportunities of ai-enabled monitoring, diagnosis & prognosis: A review, *Chinese Journal of Mechanical Engineering* 34 (1) (2021) 1–29.
- [10] 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.
- [11] J. Xiong, O. Fink, J. Zhou, Y. Ma, Controlled physics-informed data generation for deep learning-based remaining useful life prediction under unseen operation conditions, *Mechanical Systems and Signal Processing* 197 (2023) 110359.
- [12] H. Shiri, P. Zimroz, J. Wodecki, A. Wylomańska, R. Zimroz, K. Szabat, Using long-term condition monitoring data with non-gaussian noise for online diagnostics, *Mechanical Systems and Signal Processing* 200 (2023) 110472.
- [13] Y. Li, K. Feng, X. Liang, M. J. Zuo, A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved Vold-Kalman filter and multi-scale sample entropy, *Journal of Sound and Vibration* 439 (2019) 271–286.

- [14] M. Baptista, E. M. Henriques, I. P. de Medeiros, J. P. Malere, C. L. Nascimento Jr, H. Prendinger, Remaining useful life estimation in aeronautics: combining data-driven and Kalman filtering, *Reliability Engineering & System Safety* 184 (2019) 228–239.
- [15] C. K. R. Lim, D. Mba, Switching kalman filter for failure prognostic, *Mechanical Systems and Signal Processing* 52 (2015) 426–435.
- [16] Y. Bar-Shalom, X. R. Li, T. Kirubarajan, Estimation with applications to tracking and navigation: theory algorithms and software, John Wiley & Sons, 2001.
- [17] I. Arasaratnam, S. Haykin, R. J. Elliott, Discrete-time nonlinear filtering algorithms using gauss–hermite quadrature, *Proceedings of the IEEE* 95 (5) (2007) 953–977.
- [18] S. J. Julier, J. K. Uhlmann, Unscented filtering and nonlinear estimation, *Proceedings of the IEEE* 92 (3) (2004) 401–422.
- [19] I. Arasaratnam, S. Haykin, Cubature kalman filters, *IEEE Transactions on automatic control* 54 (6) (2009) 1254–1269.
- [20] W. Żuławiński, K. Maraj-Zygmunt, H. Shiri, A. Wylomańska, R. Zimroz, Framework for stochastic modelling of long-term non-homogeneous data with non-gaussian characteristics for machine condition prognosis, *Mechanical Systems and Signal Processing* 184 (2023) 109677.
- [21] H. W. Sorenson, D. L. Alspach, Recursive Bayesian estimation using Gaussian sums, *Automatica* 7 (4) (1971) 465–479.
- [22] A. Harvey, A. Luati, Filtering with heavy tails, *Journal of the American Statistical Association* 109 (507) (2014) 1112–1122.
- [23] D. Magill, Optimal adaptive estimation of sampled stochastic processes, *IEEE Transactions on Automatic Control* 10 (4) (1965) 434–439.
- [24] I. Bilik, J. Tabrikian, MMSE-based filtering in presence of non-Gaussian system and measurement noise, *IEEE Transactions on Aerospace and Electronic Systems* 46 (3) (2010) 1153–1170.
- [25] D. Alspach, H. Sorenson, Nonlinear Bayesian estimation using Gaussian sum approximations, *IEEE Transactions on Automatic Control* 17 (4) (1972) 439–448.
- [26] J. H. Kotecha, P. M. Djuric, Gaussian sum particle filtering, *IEEE Transactions on Signal Processing* 51 (10) (2003) 2602–2612.
- [27] A. Doucet, N. De Freitas, N. J. Gordon, et al., Sequential Monte Carlo methods in practice, Vol. 1, Springer, 2001.
- [28] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, J. Miguez, Particle filtering, *IEEE Signal Processing Magazine* 20 (5) (2003) 19–38.
- [29] W. Liu, P. Pokharel, J. Principe, Correntropy: A localized similarity measure, in: The 2006 IEEE International Joint Conference on Neural Network Proceedings, 2006, pp. 4919–4924. [doi:10.1109/IJCNN.2006.247192](https://doi.org/10.1109/IJCNN.2006.247192).
- [30] B. Chen, X. Liu, H. Zhao, J. C. Principe, Maximum correntropy Kalman filter, *Automatica* 76 (2017) 70–77.
- [31] R. Izanloo, S. A. Fakoorian, H. S. Yazdi, D. Simon, Kalman filtering based on the maximum correntropy criterion in the presence of non-Gaussian noise, in: 2016 Annual Conference on Information Science and Systems (CISS), IEEE, 2016, pp. 500–505.
- [32] G. Wang, R. Xue, J. Wang, A distributed maximum correntropy Kalman filter, *Signal Processing* 160 (2019) 247–251.
- [33] G. Wang, Y. Zhang, X. Wang, Iterated maximum correntropy unscented Kalman filters for non-Gaussian systems, *Signal Processing* 163 (2019) 87–94.
- [34] H. Shiri, J. Wodecki, R. Zimroz, Robust switching kalman filter for diagnostics of long-term condition monitoring data in the presence of non-gaussian noise, in: IOP Conference Series: Earth and Environmental Science, Vol. 1189, IOP Publishing, 2023, p. 012007.
- [35] C. Lim, D. Mba, Switching kalman filter for failure prognostic, *Mechanical Systems and Signal Processing* 52–53 (1) (2015) 426 – 435.
- [36] B. Zhang, S. Zhang, W. Li, Bearing performance degradation assessment using long short-term memory recurrent network, *Computers in Industry* 106 (2019) 14–29.
- [37] Q. Li, Z. Ma, H. Li, X. Liu, X. Guan, P. Tian, Remaining useful life prediction of mechanical system based on performance evaluation and geometric fractional lévy stable motion with adaptive nonlinear drift, *Mechanical Systems and Signal Processing* 184 (2023) 109679.
- [38] H. Shiri, P. Zimroz, J. Wodecki, A. Wylomańska, R. Zimroz, Data-driven segmentation of long term condition monitoring data in the presence of heavy-tailed distributed noise with finite-variance, *Mechanical Systems and Signal Processing* 205 (2023) 110833.
- [39] J. Janczura, T. Barszcz, R. Zimroz, A. Wylomańska, Machine condition change detection based on data segmentation using a three-regime, -stable hidden markov model, *Measurement* (2023) 113399.
- [40] B. Chen, X. Liu, H. Zhao, J. C. Principe, Maximum correntropy kalman filter (2015). [arXiv:1509.04580](https://arxiv.org/abs/1509.04580).
- [41] P. Zarchan, H. Musoff, Polynomial kalman filters, *Fundamentals of Kalman Filtering: A Practical Approach* (2005) 156.
- [42] M. Arrinda, M. Oyarbide, H. Macicior, E. Muxika, H. Popp, M. Jahn, B. Ganey, I. Cendoya, Application dependent end-of-life threshold definition methodology for batteries in electric vehicles, *Batteries* 7 (1) (2021) 12.
- [43] E. A. Wan, R. Van Der Merwe, The unscented kalman filter for nonlinear estimation, in: Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373), IEEE, 2000, pp. 153–158.
- [44] M. I. Ribeiro, Kalman and extended kalman filters: Concept, derivation and properties, *Institute for Systems and Robotics* 43 (46) (2004) 3736–3741.
- [45] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, C. Varnier, Pronostia: An experimental platform for bearings accelerated degradation tests., in: IEEE International Conference on Prognostics and Health Management, PHM’12., IEEE Catalog Number: CPF12PHM-CDR, 2012, pp. 1–8.
- [46] A. Mosallam, K. Medjaher, N. Zerhouni, Time series trending for condition assessment and prognostics, *Journal of manufacturing technology management* (2014).

- [47] T. H. Loutas, D. Roulias, G. Georgoulas, Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression, *IEEE Transactions on Reliability* 62 (4) (2013) 821–832.
- [48] K. Javed, R. Gouriveau, N. Zerhouni, P. Nectoux, Enabling health monitoring approach based on vibration data for accurate prognostics, *IEEE Transactions on industrial electronics* 62 (1) (2014) 647–656.
- [49] 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) (2014) 1781–1790.
- [50] B. Zhang, L. Zhang, J. Xu, Degradation feature selection for remaining useful life prediction of rolling element bearings, *Quality and Reliability Engineering International* 32 (2) (2016) 547–554.
- [51] S. Hong, Z. Zhou, E. Zio, W. Wang, An adaptive method for health trend prediction of rotating bearings, *Digital Signal Processing* 35 (2014) 117–123.
- [52] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, J. Dybala, A model-based method for remaining useful life prediction of machinery, *IEEE Transactions on reliability* 65 (3) (2016) 1314–1326.
- [53] Y. Nie, J. Wan, Estimation of remaining useful life of bearings using sparse representation method, in: *2015 Prognostics and System Health Management Conference (PHM)*, IEEE, 2015, pp. 1–6.
- [54] Z. Liu, M. J. Zuo, Y. Qin, Remaining useful life prediction of rolling element bearings based on health state assessment, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 230 (2) (2016) 314–330.
- [55] J. K. Kimotho, C. Sondermann-Wölke, T. Meyer, W. Sextro, Machinery prognostic method based on multi-class support vector machines and hybrid differential evolution–particle swarm optimization, *Chemical Engineering Transactions* 33 (2013).
- [56] D. Zurita, J. A. Carino, M. Delgado, J. A. Ortega, Distributed neuro-fuzzy feature forecasting approach for condition monitoring, in: *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, IEEE, 2014, pp. 1–8.
- [57] L. Guo, H. Gao, H. Huang, X. He, S. Li, Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring, *Shock and Vibration* 2016 (2016).
- [58] X. Jin, Y. Sun, Z. Que, Y. Wang, T. W. Chow, Anomaly detection and fault prognosis for bearings, *IEEE Transactions on Instrumentation and Measurement* 65 (9) (2016) 2046–2054.
- [59] H. Li, Y. Wang, Rolling bearing reliability estimation based on logistic regression model, in: *2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE)*, IEEE, 2013, pp. 1730–1733.
- [60] Z. Huang, Z. Xu, X. Ke, W. Wang, Y. Sun, Remaining useful life prediction for an adaptive skew-wiener process model, *Mechanical Systems and Signal Processing* 87 (2017) 294–306.
- [61] Y. Wang, Y. Peng, Y. Zi, X. Jin, K.-L. Tsui, A two-stage data-driven-based prognostic approach for bearing degradation problem, *IEEE Transactions on industrial informatics* 12 (3) (2016) 924–932.
- [62] Y. Pan, M. J. Er, X. Li, H. Yu, R. Gouriveau, Machine health condition prediction via online dynamic fuzzy neural networks, *Engineering Applications of Artificial Intelligence* 35 (2014) 105–113.
- [63] L. Wang, L. Zhang, X.-z. Wang, Reliability estimation and remaining useful lifetime prediction for bearing based on proportional hazard model, *Journal of Central South University* 22 (12) (2015) 4625–4633.
- [64] L. Xiao, X. Chen, X. Zhang, M. Liu, A novel approach for bearing remaining useful life estimation under neither failure nor suspension histories condition, *Journal of Intelligent Manufacturing* 28 (8) (2017) 1893–1914.
- [65] W. Zulawinski, K. Maraj-Zygmata, H. Shiri, A. Wylomanska, R. Zimroz, Framework for stochastic modelling of long-term non-homogeneous data with non-Gaussian characteristics for machine condition prognosis, *Mechanical Systems and Signal Processing* 184 (2023) 109677.
- [66] H. Qiu, J. Lee, J. Lin, G. Yu, Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, *Journal of sound and vibration* 289 (4-5) (2006) 1066–1090.
- [67] J. Lee, H. Qiu, G. Yu, J. Lin, et al., Rexnord technical services: Bearing data set, Moffett Field, CA: IMS, Univ. Cincinnati. NASA Ames Prognostics Data Repository, NASA Ames (2007).
- [68] A. Soualhi, H. Razik, G. Clerc, D. D. Doan, Prognosis of bearing failures using hidden markov models and the adaptive neuro-fuzzy inference system, *IEEE Transactions on Industrial Electronics* 61 (6) (2013) 2864–2874.
- [69] J. B. Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, F. Fnaiech, Accurate bearing remaining useful life prediction based on weibull distribution and artificial neural network, *Mechanical Systems and Signal Processing* 56 (2015) 150–172.
- [70] Y. Zhou, A. Kumar, C. Parkash, G. Vashishtha, H. Tang, J. Xiang, A novel entropy-based sparsity measure for prognosis of bearing defects and development of a sparsogram to select sensitive filtering band of an axial piston pump, *Measurement* 203 (2022) 111997.
- [71] A. Kumar, C. Parkash, G. Vashishtha, H. Tang, P. Kundu, J. Xiang, State-space modeling and novel entropy-based health indicator for dynamic degradation monitoring of rolling element bearing, *Reliability Engineering & System Safety* 221 (2022) 108356.
- [72] A. Kumar, C. Gandhi, Y. Zhou, R. Kumar, J. Xiang, Latest developments in gear defect diagnosis and prognosis: A review, *Measurement* 158 (2020) 107735.
- [73] Q. Li, S. Y. Liang, J. Yang, B. Li, Long range dependence prognostics for bearing vibration intensity chaotic time series, *Entropy* 18 (1) (2016) 23.
- [74] Y. Qian, R. Yan, R. X. Gao, A multi-time scale approach to remaining useful life prediction in rolling bearing, *Mechanical Systems and Signal Processing* 83 (2017) 549–567.
- [75] E. Bechhoefer, R. Schlanbusch, Generalized prognostics algorithm using Kalman smoother, *IFAC-PapersOnLine* 48 (21) (2015) 97–104.
- [76] L. Saidi, J. B. Ali, E. Bechhoefer, M. Benbouzid, Wind turbine high-speed shaft bearings health prognosis through a

- spectral kurtosis-derived indices and svr, *Applied Acoustics* 120 (2017) 1–8.
- [77] L. Saidi, E. Bechhoefer, J. B. Ali, M. Benbouzid, Wind turbine high-speed shaft bearing degradation analysis for run-to-failure testing using spectral kurtosis, in: 2015 16th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), IEEE, 2015, pp. 267–272.
- [78] J. B. Ali, L. Saidi, S. Harrath, E. Bechhoefer, M. Benbouzid, Online automatic diagnosis of wind turbine bearings progressive degradations under real experimental conditions based on unsupervised machine learning, *Applied Acoustics* 132 (2018) 167–181.
- [79] E. Bechhoefer, R. Schlanbusch, Generalized prognostics algorithm using kalman smoother, *IFAC-PapersOnLine* 48 (21) (2015) 97–104.

Appendix A. Student's t distribution

The PDF of the (unscaled) Student's t distribution is the following:

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\frac{\nu}{2} \sqrt{\nu\pi}} \frac{1}{(1+\frac{x^2}{\nu})^{\frac{\nu+1}{2}}}, \quad (\text{A.1})$$

where ν is called the number degrees of freedom and $\Gamma(\cdot)$ is the gamma function. It should be noted that: the mean of the Student's t distribution is defined if $\nu > 1$ as $\mu = 0$; the variance of the Student's t distribution is defined if $\nu > 2$ as $\sigma = \frac{\nu}{\nu-2}$.

Appendix B. Pseudocodes for proposed method

Subsequently, to enhance clarity and facilitate understanding of the RUL estimation process by the MCEKF, we propose the following pseudocodes:

Algorithm 1 RUL Estimation with MCEKF

- 1: **Input:** Observations y_t for discrete time t
 - 2: **Output:** Predicted RUL, Probability
 - 3: **Initialize State and Covariance Matrix:**
 - 4: $\mathbf{x}_0 \leftarrow$ Initial state vector $[a_0, b_0, c_0]^T$
 - 5: $\mathbf{P}_0 \leftarrow$ Initial covariance matrix
 - 6: **Initialize Process and Measurement Noise Covariance and maximum correntropy bandwidth:**
 - 7: $\mathbf{Q}_t \leftarrow$ Process noise covariance matrix
 - 8: $M_t \leftarrow$ Time-changing scale parameter for measurement noise
 - 9: $\sigma \leftarrow$ Bandwidth for MCEKF
 - 10: **for** t **in** observations **do**
 - 11: **MCEKF Update:**
 - 12: $\mathbf{x}_t, \mathbf{P}_t \leftarrow$ MCEKF-Update($\mathbf{x}_{t-1}, \mathbf{P}_{t-1}, y_t, \mathbf{Q}_t, M_t, \sigma$)
 - 13: **Prediction of Next Observation:**
 - 14: $y_t \leftarrow h(\mathbf{x}_t)$
 - 15: **Prediction of RUL:**
 - 16: $t_{\text{EOL}} \leftarrow$ Time when degradation curve crosses EOL threshold
 - 17: $t_0 \leftarrow$ Time of the last observation in the training data set
 - 18: **Calculate Probability of Predicted RUL:**
 - 19: $P(\text{Predicted RUL}) \leftarrow$ CalculateProbability($\mathbf{x}_t, \mathbf{P}_t, t_{\text{EOL}}, t_0$)
 - 20: **return** Predicted RUL, $P(\text{Predicted RUL})$
 - 21: **end for**
-

Algorithm 2 MCEKF

```
1: Initialize state estimate  $\hat{\mathbf{x}}_0$  and covariance matrix  $\mathbf{P}_0$ 
2: for each time step  $t$  do
3:   Prediction:
4:    $\hat{\mathbf{x}}_{t|t-1} = \mathbf{f}(\hat{\mathbf{x}}_{t-1})$  ▷ Predicted state
5:    $\mathbf{A}_t = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{t-1})$  ▷ Jacobian of process function
6:    $\mathbf{P}_{t|t-1} = \mathbf{A}_t \mathbf{P}_{t-1} \mathbf{A}_t^T + \mathbf{Q}_t$  ▷ Predicted covariance
7:   Update:
8:    $\mathbf{H}_t = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{t|t-1})$  ▷ Jacobian of measurement function
9:    $\mathbf{z}_t = \mathbf{y}_t - \mathbf{h}(\hat{\mathbf{x}}_{t|t-1})$  ▷ Measurement residual
10:   $\lambda_t = G_\sigma \left( \|\mathbf{z}_t\|_{\mathbf{M}_t^{-1}} \right)$  ▷ Scale factor
11:   $\mathbf{K}_t = \hat{\mathbf{P}}_{t|t-1} \lambda_t \mathbf{H}_t^T \left( \mathbf{H}_t \hat{\mathbf{P}}_{t|t-1} \lambda_t \mathbf{H}_t^T + \mathbf{M}_t \right)^{-1}$  ▷ Kalman gain
12:   $\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t \mathbf{z}_t$  ▷ Updated state estimate
13:   $\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \hat{\mathbf{P}}_{t|t-1}$  ▷ Updated covariance
14: end for
```
