Prognostics of Rolling Element Bearings based on Entropy indicators and Particle Filtering

Junyu QI[1, 2], Alexandre MAURICIO[1, 2], Konstantinos GRYLLIAS[1, 2]

[1]Department of Mechanical Engineering, KU Leuven,
[2]Dynamics of Mechanical and Mechatronic Systems, Flanders Make
Celestijnenlaan 300, BOX 2420, 3001 Leuven, Belgium
Junyu.qi@kuleuven.be, alex.ricardomauricio@kuleuven.be, konstantinos.gryllias@kuleuven.be

Abstract

Rolling element bearings' damage is the main cause for unexpected breakdown in rotating machinery. Therefore there is a continuous industrial interest on condition monitoring of bearings, targeting towards the development and proposal of robust diagnostic techniques which can detect accurately, robustly and early the generation of the fault. On the other hand, industry is not interested only in the proper damage detection and identification of faults but is mainly targeting towards the robust estimation of the Remaining Useful Life (RUL) of machine elements. The proper estimation of the RUL could be linked directly with the maintenance planning and warehouse organization, providing immediately profits in terms of employees health and safety, environmental protection and continuous production. A plethora of diagnostics and prognostics indicators have been proposed during the last decade focusing towards the accurate representation and tracking of the health state of bearings and other machine elements. However, in certain cases (e.g. nonstationary operating conditions), the classic techniques for bearings prognostics (e.g. statistical analysis, frequency analysis and time-frequency analysis) underperform due to the high noise influence or the high machines' complexity. Therefore the classical diagnostics indicators may identify the fault quite late and fail to identify properly the RUL.

In this paper, prognostics indicators based on the measurement of disorder (e.g. entropy) are used in order to track the degradation process of the machinery. An advanced health indicator, termed Info-entropy, is built based on the Spectral Entropy, the Envelope Spectral Entropy and the Spectral Negentropy. It is compared with the Spectral Kurtosis, the Kurtosis and the RMS based on the criterion of trendability. The indicators are estimated on the well-known bearing dataset Prognostia and Particle filtering is used in order to estimate the RUL of bearings.

1 Introduction

In the era of Industry 4.0, Condition Monitoring (CM) is currently attracting unprecedented attention. Rotating machinery systems and components, such as Rolling Element Bearings (REBs), are utilized frequently in almost each mechanical equipment, including electric motors, wind turbines and compressors, etc. The breakdown of large amount of machineries happens due to the failure of bearings, which are mainly caused by misalignment, corrosion, insufficient lubrication and oil contamination, resulting further in accelerated wear in the surface. As the damage accumulates in the continuous operation, the functionality of REBs reduces through the stages of mild (stage 1), moderate (stage 2) and profound fault (stage 3), as presented in Figure 1.

Usually, stage 1 endures for long time, and as the fault reaches to a certain level, the degradation rate starts accelerating and as a result the stage 2 lasts for shorter time compared to stage 1. In stage 3, the functionality reduces to the End of Life (EoL) incredibly fast. Concerning the change of functionality, a number of maintenance concepts have been proposed in the last decades, i.e., the proactive maintenance (to stage 1), preventative maintenance (to stage 2) and reactive maintenance (to stage 3). The reactive

maintenance (fault detection and diagnosis) have been explored widely and a number of systematic, automatic and accurate methods are successfully established. However, the period of the reactive maintenance usually sustains shortly till the EoL, as stage 3 shown in Figure 1, and it requires expensive cost. On one hand, the hazardous performance of the bearings influences the normal operating condition of the equipment, which is related with the continuous production quality. On the other hand, it may lead to accidental breakdown of machinery. Preventative maintenance is localized in stage 2, while the defect is still in the impending or the incipient stage and the functionality can be corrected by proper intervention, e.g. cleanness, lubrication and alignment. The proactive maintenance in stage 1 costs less expensive, where the defect incubates slowly and the REB exhibits an acceptable performance. Therefore, if the life of in-service components is able to be predicted in the early time (stage 1), the maintenance strategy and actions can be optimally planned. Compared with traditional CM techniques, the predictive maintenance is recently catching particular attention. The state of the art of methodologies [1] includes mainly three directions, i.e. the Physics model based Prognostics (PbP), the Data-driven based Prognostics (DbP) and the Hybrid-approach based Prognostics (HbP).

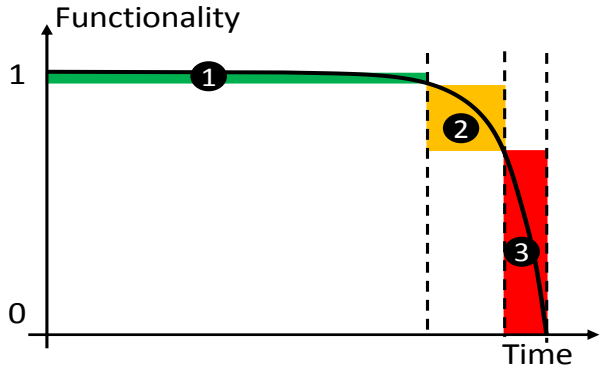

Figure 1: Degradation process and maintenance strategy

The PbP usually describes the degradation phenomenon in an explicit mathematical formulation. The establishment of a physical Finite Element Model (FEM) depends on the available sufficient information, e.g. the geometry, the load and the material, etc. In the past, some physics models have been developed, e.g. spall progression models, crack-growth model, gas path model, mixed lubrication model, etc. [2]. Theoretically, the PbP is built systematically and well adaptive to the degradation evolution process. However, the modeling in the system level requires complete domain knowledge and is expensive to be developed, therefore, the PbP is usually applied to a specific component.

With the advantage of flexibility, applicability and economy, the DbP is preferred by the industrial domain. Unlike the difficulty of building the physics model, an analytic model is directly derived from massive data. In general, the DbP is classified as Machine Learning (ML) and Statistical Model Based (SMB) techniques [3]. ML learns the hidden information from big data. Instead of an explicit model, ML works like a black box and till now often the output is hardly interpretable. In order to reduce the computation cost and enhance the knowledge understanding, the direction of SMB is considered in this paper, e.g. Kalman filter (KF), Particle Filter (PF), linear regression, Hidden Markov Model (HMM), etc. An analytic model, e.g. the exponential model [4], is learned from training data, and further is updated with testing data.

Medjaher et al. [5] presented a method of bearing prognostics based on the regression method. Yet, it is not an optimal solution for the stochastic degradation progress. Alternatively, the Bayesian Recursive Estimation (BRE) is able to represent the model, considering the uncertainty. For example, HMM is used for the RUL prediction, considering different failure modes [6]. However, the probability transition matrix between the model's states need to be presupposed. Qi et al. [7] employed the Unscented Kalman filter (UKF) for the bearing RUL prediction. Additionally, in order to cope with the limitation of the single model, the Switching Kalman Filter was investigated for the prognostics of a helicopter bearing. [8]. PF achieves advantageous performance than KF versions without the restrictions of Gaussian noise, system linearity or existence of low nonlinearities. Li et al. [9] predicted bearing RUL combining the PF with an improved exponential model. If Health Indicator (HI) exhibits a degradation pattern with multiple degradation rates,

the single statistical model is probably not adequate anymore. Therefore, a multi-model PF is introduced to track the linear and nonlinear stage, and the corresponding mode is recognized automatically by the calculation of the likelihood [10].

The result of SMB is usually dependent on several factors, e.g., the statistical model, the HI, the failure threshold, etc. To extract high quality HI, different signal processing techniques have been developed in the domain of time, frequency and time-frequency. However, many classic indicators are frequently deficient in the real applications. Ahmad et al. [11] proposed an improved HI removing the fluctuation based on the Linear Rectification Technique (LRT). In order to track the degradation in the best way, some researchers focused on the change in the specific frequency band. Singleton et al. [12] extracted a HI over the specific frequency band of the Choi-Williams Transform. Soualhi et al. [13] built a HI based on the Hilbert Huang Transform, decomposing the signal into several Intrinsic Mode Functions (IMF) in different frequency bands. However, an optimal HI with a remarkable tendency is usually challengeable to attain in reality, especially the tendency in the full lifetime. As a result, classic HIs, such as RMS and Kurtosis, may fail to track the trend in case of high noise, low impulsiveness and varying operating condition. The problem of low trendability triggers the proposal of start prediction point [9], [11] and the RUL prediction by the start prediction point is usually starting in the very late stage, e.g. in stage 3 of Figure 1. Moreover, it also influences the statistical model selection in PF.

In order to enhance the prediction performance, a new HI Info-entropy is proposed for the bearing prognostics in this paper. Following the logic of [12]-[13], Info-entropy assesses the chaos level for a specific frequency band and is composed of different entropy indicators in the frequency domain, i.e. the Spectral Entropy (SpecE), the Spectral Flatness (SpecF) and the Spectral Negentropy (SpecNegE). Based on the criterion of trendability, the best entropy indicators from a certain frequency band is selected. Then, the Info-entropy is collaborated with a classic exponential model and PF. The main contributions of the proposed methodology for SMB prognostics are summarized as follows.

1) Through the chaos measurement in multiple frequency bands, Info-entropy recognizes the hidden signature of the vibration signal of the REB. It provides a new perspective for the HI extraction compared to the traditional impulsiveness based features. Even in cases of ignoring the characteristic bearing frequencies, the low frequency resolution and the repetitive transients, the Info-entropy could detect the inherent change.

2) The in-service machinery degrades with time and its corresponding functionality reduces also, as can be seen in Figure 1. The Info-entropy is sensitive to frequency distributions, demonstrating the characteristic of high trendabillity. With the full lifetime degradation tracking, the aforementioned concept of start prediction point is no longer the practical bottleneck. Consequently, the use of Info-entropy allows for the start of prediction from the initial moment and potentially is useful for long-term RUL prediction.

3) Likewise, as Info-entropy present a high quality of trendability, the corresponding statistical model selection could be simplified. Compared with the complex model, e.g. for the non-monotonic trend, the actual calculation rested on the simple exponential model, becomes faster and efficient.

The rest of paper is organized as follows. The mathematical theory of PF and the construction of Info-entropy is explained in Section 2. The proposed methodology is detailed step by step in Section 3. In Section 4 and 5, the experimental dataset and the RUL prediction results are described. Finally, the paper is concluded in Section 6.

2 Mathematical theory

The theory of Particle filter and the proposed HI based on the Info-entropy are explained in this Section.

2.1 Particle filter

As a kind of BRE method, PF uses a number of particles to estimate the posterior distribution of a stochastic process. It is particularly useful for the estimation of a nonlinear dynamic system with non-Gaussian noise. The principles of PF is based on the Bayesian theory and sampling strategy. Bayesian theory includes steps of the prediction and update. In the prediction step, the priori Probability Density Function (PDF) of the k_{th} predicted state x_k is gained with the posteriori knowledge $p(x_k|z_{1:k-1})$. Then, the posteriori PDF $p(x_k|z_{1:k})$ is updated when the k_{th} measurement $z_k$is available. According to Monte Carlo Sampling (MCS), the state x_k is approximated by a set of particles $x_k^i, i = 1 \dots N$, which are sampled from a certain probability distribution. The posteriori probability of Bayesian estimation is able to be approximated

with MCS, $p(x_k|z_{1:k}) \approx \frac{1}{N}\sum_{i=1}^{N}\delta(x_k - x_k^i)$, where δ denotes the Dirac delta function. Each particle has the same weight $\frac{1}{N}$. However, particles degeneracy arises in the sequential importance sampling. The amount of effective particles decreases after some certain recursive steps. Hence, the resampling duplicates the high weight particles and replaces the lower ones. The importance distribution $q(x_k|x_{k-1}, z_k)$ is generally chosen as $p(x_k^i|x_{k-1}^i)$, then $w_k^i \propto w_{k-1}^i p(z_k|x_k^i)$.

PF mainly consists of three steps, the particles generation, the weight calculation and the resampling.

1) Particles generation

As the initial step, the amount (N) of particles are firstly generated from a priori distribution and then are transmitted through the propagation model, $x_k^i \sim p(x_k|x_{k-1}^i), i = 1 \ldots N$.

2) Weight calculation

The likelihood of each particle x_k^i is calculated with the measurement z_k, $w_k^i = p(z_k|x_k^i)$, and is normalized as $\breve{w}_k^i = \frac{w_k^i}{\sum_{i=1}^{N} w_k^i}$.

3) Resampling

Being the most frequently used resampling method, the Multinomial Resampling (MR) compares the weight of particles $\breve{w}_k^i$ with a random threshold between 0 and 1, and only the particles, which are above the threshold, are kept. It is worthy to mention, that the Systematic Resampling (SR) is utilized to replace the classic MR and to enhance the PF performance. Unlike the MR method, SR divides the whole interval into M subspaces $\mathrm{U}^{\mathrm{i}} = ((\mathrm{i}-1)/\mathrm{M}, \mathrm{i}/\mathrm{M}), \mathrm{i} = 1, \ldots, \mathrm{M}$, and the particles are taken with a corresponding random threshold in the U^i, higher particle diversity is thus guaranteed.

2.2 Info-entropy extraction

Many classic HIs are extracted based on the impulsiveness of the signal, e.g., the statistical indicators, the Spectral Kurtosis, frequency indicators extracted by the envelope spectrum, after the application of Hilbert transform, etc. Nevertheless, the fault information is easily affected by the noise, non-stationarities and repetitive transients. Moreover, the behavior of rotating machinery performs physically nonlinear, due to the instantaneous variation of the friction, damping and loading conditions. Thus, the chaos assessment is taken into account. In this aspect, entropy is the versatile measurement of disorder, unevenness of distribution and complexity. The concept of entropy, different entropy indicators and the info-entropy will be explained in this Section.

2.2.1 Entropy

As property of thermodynamics, the entropy of an isolated system never decreases. It always evolves toward the thermodynamic equilibrium, which means each of the microstates S_i inside the system has an equal probability p_i.

2.2.2 Entropy indicators

A number of variants of entropy have been proposed in the time series, e.g. the Approximate Entropy, the Sample Entropy, the Kolmogorov Entropy and the Permutation Entropy. Yet, the assessment in the time domain is often insufficient. The frequency information is explored here, as the compensation to capture the latent variation. Thus, the Spectral Entropy, the Spectral Flatness and the Spectral NegEntropy, i.e., the NegEntropy in the Squared Envelope (NegE-SE) and the Squared Envelope Spectrum (NegE-SES) are introduced.

SpecE has been proposed to measure the amplitude distribution of the Fourier spectrum of a signal $x(t)$. It is defined as the same form of entropy:

$$SpecE = -\sum_{i=1}^{L} \mathrm{p_i} log_2\,(\mathrm{p_i}) \tag{1}$$

with,

$$p_i = \frac{X(i)}{\sum_{i=1}^{L} X(i)} \tag{2}$$

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \tag{3}$$

When the spectral amplitudes distribute evenly, SpecE yields a large value. On the contrary, it outputs a small value if the amplitudes are concentrated around a specific region. To avoid the influence of the data length, the SpecE is usually normalized as in Equation 4, and ranges between 0 and 1.

$$SpecE = \frac{-\sum_{i=1}^{L} \mathrm{p_i} log_2\,(\mathrm{p_i})}{log_2(L)} \tag{4}$$

SpecF, also called Wiener entropy, has been initially used for the whiteness characterization of speech analysis [14]. It is calculated by the ratio between the geometric mean and the arithmetic mean of the power spectrum, as presented in equation 5.

$$SpecF = \frac{\exp\left(\frac{1}{L}\sum_{i=1}^{L} \ln\left(|X(i)|^2\right)\right)}{\frac{1}{L}\sum_{i=1}^{L} \ln\left(|X(i)|^2\right)} \tag{5}$$

Equivalent to the SpecE, SpecF approaches to 1 if the spectrum uniformly scatters and to 0 if the spectrum is spiky.

As the opposite of entropy, NegEntropy is used as a measure of distance from normality. The Spectral NegEntropy in the frequency band $(f - 0.5 \cdot \Delta f, f + 0.5 \cdot \Delta f)$ is defined in [15] as:

$$\Delta \mathrm{I}_{\varepsilon}(\mathrm{f}, \Delta \mathrm{f}) = \langle \frac{\varepsilon_{\mathrm{x}}(\mathrm{t,f,\Delta f})^2}{\langle \varepsilon_{\mathrm{x}}(\mathrm{t,f,\Delta f})^2 \rangle} \ln\left(\frac{\varepsilon_{\mathrm{x}}(\mathrm{t,f,\Delta f})^2}{\langle \varepsilon_{\mathrm{x}}(\mathrm{t,f,\Delta f})^2 \rangle}\right) \rangle \tag{6}$$

where, ε_x is the SE of signal $x(t)$, $\langle * \rangle$ indicates the average operator and ΔI_ε represents the NegEntropy in the SE (NegE-SE). Similarly, NegE-SES is written as:

$$\Delta \mathrm{I}_{\mathrm{E}}(\mathrm{f}, \Delta \mathrm{f}) = \langle \frac{\mathrm{E_x(t,f,\Delta f)^2}}{\langle \mathrm{E_x(t,f,\Delta f)^2} \rangle} \ln\left(\frac{\mathrm{E_x(t,f,\Delta f)^2}}{\langle \mathrm{E_x(t,f,\Delta f)^2} \rangle}\right) \rangle \tag{7}$$

where, ΔI_E is the NegEntropy in the SES, and E_x denotes SES of $x(t)$. In addition, to overcome the shortcomings of NegE-SE and NegE-SES, the weighted average Spectral NegEntropy (NegE-Ave) has been considered:

$$\Delta \mathrm{I}_{1/2}(\mathrm{f}, \Delta \mathrm{f}) = (\Delta \mathrm{I}_{\varepsilon}(\mathrm{f}, \Delta \mathrm{f}) + \Delta \mathrm{I}_{\mathrm{E}}(\mathrm{f}, \Delta \mathrm{f}))/2 \tag{8}$$

SpecNegE has reverse behavior with Spectral Entropy. If vibration signal shows more impulsiveness or repetitive transient, it returns a higher value.

2.2.3 Info-entropy

In order to capture the informative event of machinery, spectral NegEntropy has been adopted in Infogram [15]. The signal is firstly decomposed into different frequency bands using a 1/3-binary tree. Then, the demodulation frequency band is localized by the maximal spectral NegEntropy, which is caused from the faulty component.

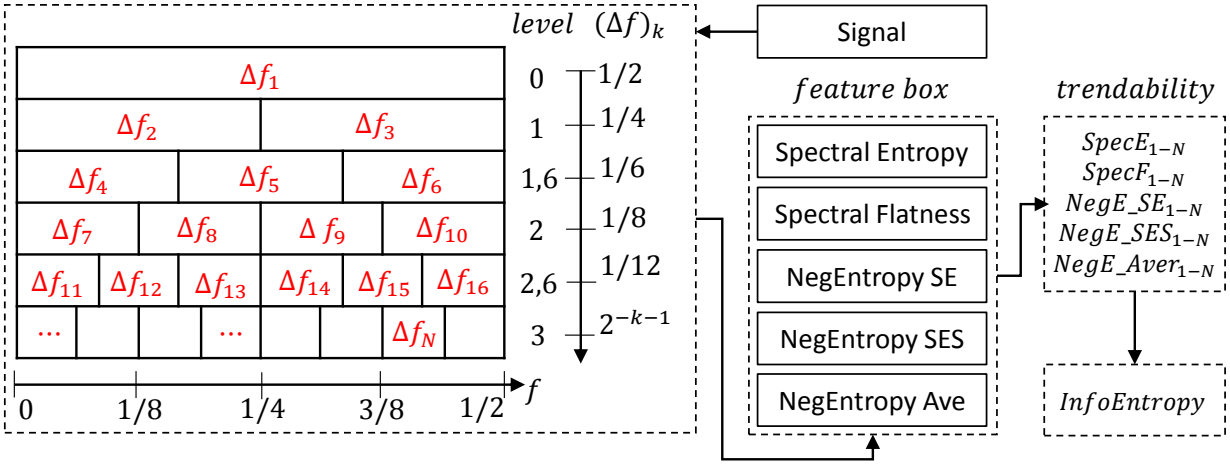

Figure 2: The illustration of Info-entropy proposition

The proposed Info-entropy is structured as presented in Figure 2. At first, the decomposed signals in different frequency bands Δf_1, Δf_2, ⋯, Δf_N are calculated by five kinds of Spectral Entropy and NegEntropy. Then, each entropy indicator has an amount of N values and is collected in a feature box, i.e., $SpecE_{1-N}$,

$SpecF_{1-N}$, $NegE\_SE_{1-N}$, $NegE\_SES_{1-N}$ and $NegE\_Ave_{1-N}$. In order to catch the degradation evolution for the specific operating condition, the anticipant feature is selected based on the trendability, as shown in Equation 9. In this way, the most sensitive and informative frequency band is localized. Finally, the selected Info-entropy can be any of the feature box with an optimal score, based on Equation 10, depicting well the whole degradation process.

$$R = \frac{n(\sum xy)-(\sum x)(\sum y)}{\sqrt{(n\sum x^2-(\sum x)^2)(n\sum y^2-(\sum y)^2)}} \quad (9)$$

where x and y represent, respectively, the time index and the feature. n indicates the length of the feature. The trendability measures the correlation between the time and the feature. The value R varies between -1 and 1. The final score of the HI is expressed in the form of equation (10). The perfect score is equal to 0 and signifies high trendability for the full lifetime.

$$Score = 1 - abs(R) \quad (10)$$

3 Proposed methodology

The proposed methodology is detailed in Figure 3. Firstly, the data is recorded in a certain operating condition. However, the vibration signals are easily polluted during the experiment, by e.g. the deterministic composition of gear transmission, anomalies, noise etc. and different preprocessing techniques are widely derived to enhance the bearing signal. In order to solve the problem of outliers, thresholding using the Standard Deviation (SD) is often used. But SD is problematic to the extreme values, e.g.to infinite valueσ. Compared to the criterion of SD, the application of three Median Absolute Deviations (MAD) is more robust to detect and remove anomalies. Therefore, the Moving MAD (MMAD) is proposed as α pre-processing technique. The MAD is calculated w.r.t the D sequence points. The samples beyond the threshold Thr_1 (e.g. $Thr_1 = 3 \cdot MAD$) are recognized as outliers and αρε replaced with Thr_1. Moreover, α prewhitening technique is suggested before calculating the Spectral NegEntropy [15].

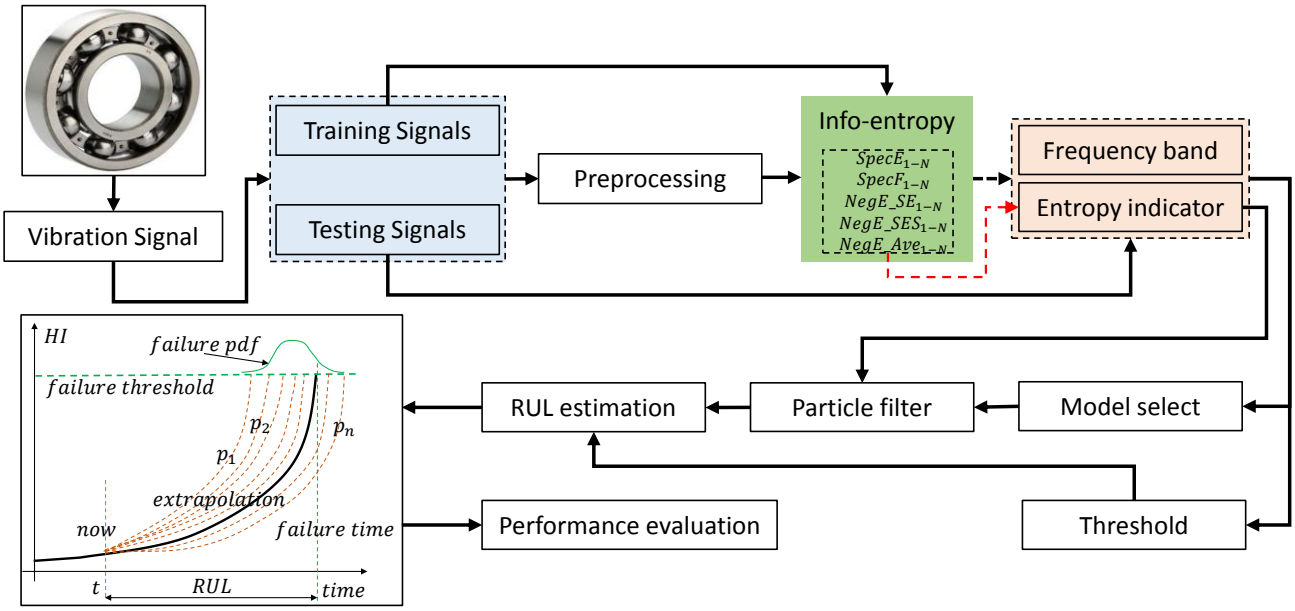

Figure 3: The flowchart of proposed methodology

Then, following the explanation of Info-entropy in Figure 2, the training dataset is utilized to localize the most informative frequency band and the corresponding entropy indicator in the feature box with the best score is selected. Based on the extracted HI, the statistical model is easily determined to represent the degradation process of the bearing. If the model parameter b_i indicates the degradation path of the i_{th} bearing B_i, the similarity level between HIs is measured by the ratio between the slowest degradation b_{min} w.r.t the closest path b_k.

$$\text{ratio} = \frac{b_{min}}{b_k} \quad (9)$$

In case, the Info-entropy of all the members of training sets $U(B_i)$ present high similarity (e.g. $ratio > \epsilon$, ϵ is the threshold of similarity level), the failure threshold thr is defined as the average value of the maximum or minimum of the His and the initial model parameters pr_0 are calculated by the average of the fitted values from $U(B_i)$. Otherwise, when the HIs show low similarity in the degradation path (e.g. $ratio < \epsilon$), the thr and pr_0 are set based on the closest HIs. They depend on the two bearings B_{n1} and B_{n2} of $U(B_i)$ if the starting point HI_0 of testing data is localized in the middle of their HI_0. Yet, when HI_0 falls outside the ranges of HI_0 of $U(B_i)$, thr and pr_0 are chosen based on the HI with the closest HI_0.

Finally, after the selection of the statistical model, the initial parameters and the failure threshold, PF starts to update the model parameters with the availability of new measurements. The RUL is calculated recursively by the extrapolation to the thr. As a set of particles is used, the RUL is then expressed as the probability of distribution. The RUL result is evaluated with three classic metrics, i.e., the Cumulative Relative Accuracy (CRA), the Convergence (C_m) and the Error (Err) [16].

The Relative Accuracy (RA) evaluates the error of the RUL prediction in relation to the actual RUL at the specific time. In order to measure the error at multiple time instances, the CRA aggregates the RA of a given time span.

$$\mathrm{CRA}_\lambda = \frac{1}{|l_\lambda|}\sum_{l_\lambda} \mathrm{w_k RA_k} \tag{11}$$

where, l_λ is the set of all time indexes, w_k and RA_k are, respectively, the weight factor and the RA at the time step k. The CRA approaches to 1 in the case of the ideal RUL prediction.

By the measurement of the Euclidean distance, convergence expresses how fast the prediction goes close to the actual RUL.

$$\mathrm{C_m} = \sqrt{(\mathrm{x_c} - \mathrm{t_p})^2 + \mathrm{y_c}^2} \tag{12}$$

where x_c and y_c are the centroid coordinates of the area under the prediction error, such as the accuracy and t_p is the first inspection time. A smaller distance shows that the predicted RUL converges faster to the actual RUL.

In addition, the percentage error is also frequently used, and is defined as follows:

$$\mathrm{Err} = \frac{\mathrm{RUL_{re} - RUL_{pr}}}{\mathrm{RUL_{re}}} \times 100\% \tag{13}$$

where, RUL_{re} and RUL_{pr} are the actual and the predicted RUL, respectively. The percentage is equal to 0 when the predicted RUL coincides with the actual RUL.

4 Experimental degradation data

The proposed methodology is evaluated on the experimental dataset PRONOSTIA [17]. The setup consists of a motor, a custom gearbox, two support bearings and the test bearing, which bears load from a pneumatic actuator. The experiment is implemented under three constant operating conditions (radial load and rotating speed). The data is measured with a sampling frequency of 25.6 kHz for a duration of 0.1 seconds every 10 seconds. Acceleration signals are recorded simultaneously in the horizontal and vertical direction. The EoL in the experiment is defined based on the vibration amplitude. The measurement stops when the vibration at one of the two measurement directions exceeds 20g in a sequence of samples. In the end, seven measurements are implemented for the first two conditions and three measurements for the condition 3.

5 Results and discussion

Based on the experimental dataset and the established procedure of the proposed methodology, the results of the RUL prediction are discussed. In this section, the experimental dataset is firstly preprocessed before calculating the Info-entropy. In the view of trendability, the extracted HI is compared with some

existing papers. Then, the statistical model, the initial parameters and the failure threshold are decided. Finally, the RUL prediction results are presented.

5.1 Data preprocessing

Following the explanation given in section 3, the preprocessing is important to remove the environmental influences. After visual inspection, the time wave of the vibration signals contain outliers and unsymmetric points. The window length D in MMAD is empirically chosen equal to 50. After cleanness, the samples against the degradation behavior are efficiently separated from the normality. In addition, a prewhitening technique based on the linear prediction is utilized before the Spectral NegEntropy. The order is set equal to 50.

5.2 Info-entropy extraction

Based on the description of Section 4, seven bearings have been measured in the first operating condition. In order to demonstrate the priority of the proposed HI in comparison with other state of the art papers [9], [11], only the same 4 bearings (bearing 1-1, bearing 1-3, bearing 1-4 and bearing 1-7) in the first operating conditions are predicted in this paper.

After conditioning of the vibration signals, the Info-entropy is firstly calculated on the six training sets. The attained informative frequency band and the entropy indicator are then employed into the rest testing data. In the process of Info-entropy calculation, five entropy indicators, i.e. $SpecE_{1-N}$, $SpecF_{1-N}$, $NegE\_SE_{1-N}$, $NegE\_SES_{1-N}$ and $NegE\_Ave_{1-N}$ are simultaneously achieved. In this paper, only the example of NegE-SE of the bearing 1-1 is presented in Figure 4. The NegE-SE in each frequency band performs an increasing tendency and it coincides with the degradation evolution and the accumulation of damage. The Info-entropy is able to find the highest trendability for all training datasets in certain frequency band, which represents the characteristic information of the same operating condition.

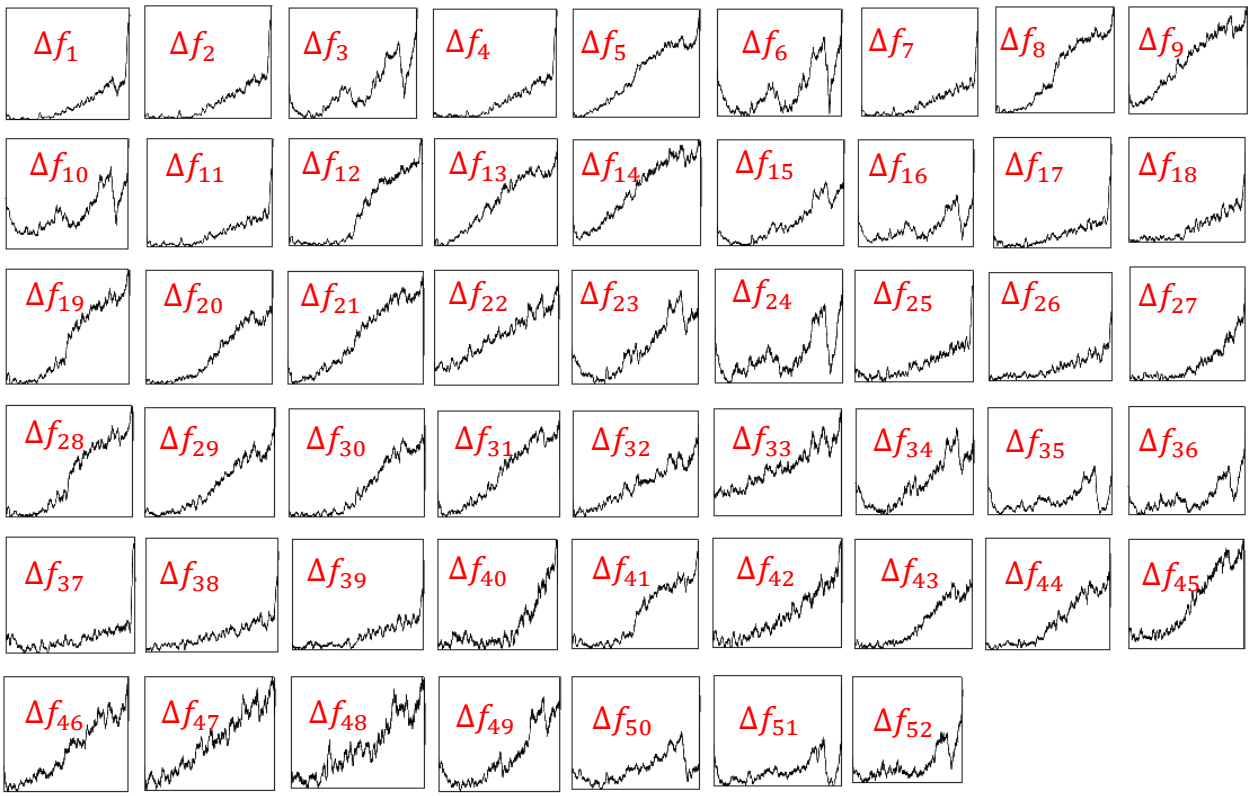

Figure 4: Spectral NegE-SE of decomposed exemplary signal (bearing 1-1)

	Bearing 1-1	Bearing 1-3	Bearing 1-4	Bearing 1-7
RMS	0.3387	0.4676	0.2707	0.6127
Kurtosis	0.3576	0.1863	0.3633	0.2786

SK	0.1840	0.1692	0.5237	0.5131
Info-entropy	0.0215	0.0467	0.1778	0.0428

Table 1: Trendability comparison of different His

The extracted feature in the frequency bands $\Delta f_1 - \Delta f_{52}$ in Figure 4 is generated by a 1/3 binary-tree, where the decomposition level L_v is empirically set equal to 4. Theoretically, the larger level L_v is able to capture more information with more decomposed signals. However, it requires also more computation cost. Based on the score of Equation 15, NegE-SE is selected as the Info-entropy among five different entropy indicators over the frequency band Δf_5, i.e., 4267 Hz-8533 Hz. The NegE-SE of four to be tested bearings are shown in Figure 5d. To highlight the superiority of the proposed HI, two statistical features Kurtosis and RMS for the same bearings in [9] and the Spectral Kurtosis (SK) are introduced as a comparison, and their trends are shown together in the Figure 5.

Analysing Figure 5a, RMS indicates the stable trend in the early stage. Especially for bearing 1-4 and 1-7, it suddenly reaches to failure after the long term 'healthy' running. Compared with RMS, Kurtosis in Figure 5b is sensitive to the impulsiveness and the slope of degradation is more distinct, e.g. in bearing 1-7. However, its trend drops in the late stage, where the fault size may accumulate and the vibration signal becomes less impulsive than the early stage fault. SK in Figure 5c has similar results as Kurtosis, but with slight higher slop. It is obvious that Info-entropy tracks well the full lifetime degradation for the four bearings. The slope of each bearing is high and cluster together. In addition, their corresponding scores based on the Equation 10 are listed in Table 1. The values of Kurtosis of bearings 1-3 and 1-7 approach closer to 0 than RMS. However, the SK of the bearing 1-1 and 1-3 triumphs over Kurtosis. It should be mentioned that the Info-entropy reveals the best score compared to the other indicators for each bearing. Therefore, with the advantage of trendability, the Info-entropy will be applied in the next section.

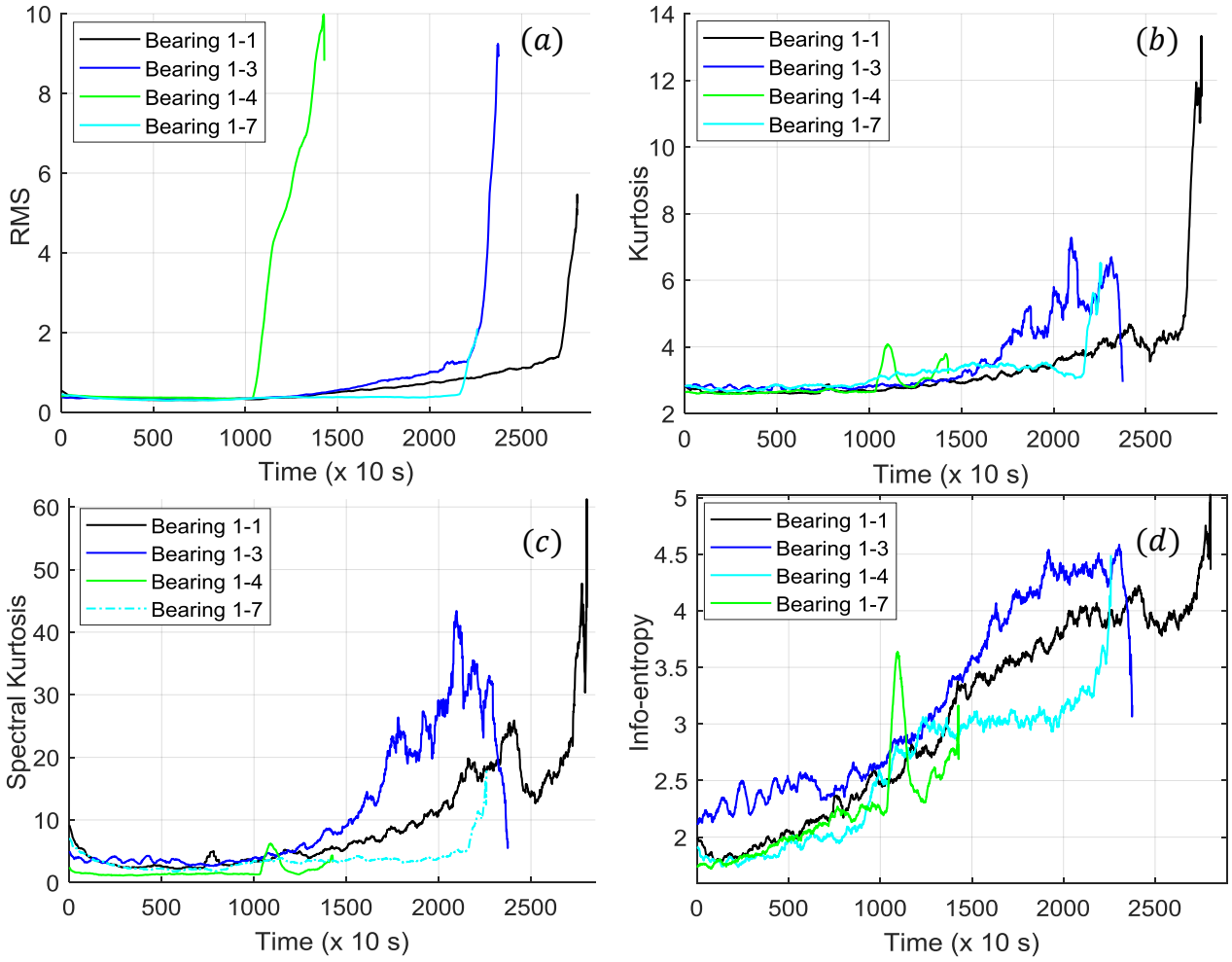

Figure 5: Comparison of different HIs

5.3 RUL results

5.3.1 Parameters definition

From the extracted Info-entropy, the degradation trend of each bearing generally follows an exponential path. Hereby, a classic exponential model is thus chosen for PF:

$$f(t) = ae^{b \cdot t} \tag{14}$$

where, a and b are the model parameters.

The failure threshold and the initial parameters are realized based on the strategy in Section 3. The threshold of similarity level ϵ is set equal to 0.6. The ratio is lower than 0.6 for the bearings in condition 1. In addition, as known, Spectral Entropy and Spectral Flatness range between 0 and 1. To facilitate the implementation, Spectral NegEntropy is normalized by the maximal value of HIs from the training dataset.

5.3.2 Results

After the selection of the HI, the statistical model and the parameters definition, the RUL of the aforementioned four bearings (length $Len = 1400, 1802, 1139, 1502$, duration $d = Len \times 10\ s$) is predicted based on the proposed methodology. In order to evaluate the effectiveness of different algorithms, the PF based methods, e.g., the Classic Particle Filter (CPF) with MR strategy and the Enhanced Particle Filter (EPF) with SR, are compared with other SMB methods, i.e. the CKF, the EKF, the UKF and the simple regression (REG). Finally, their performance are evaluated using the performance metrics, e.g. the CRA, the C_m and the Err.

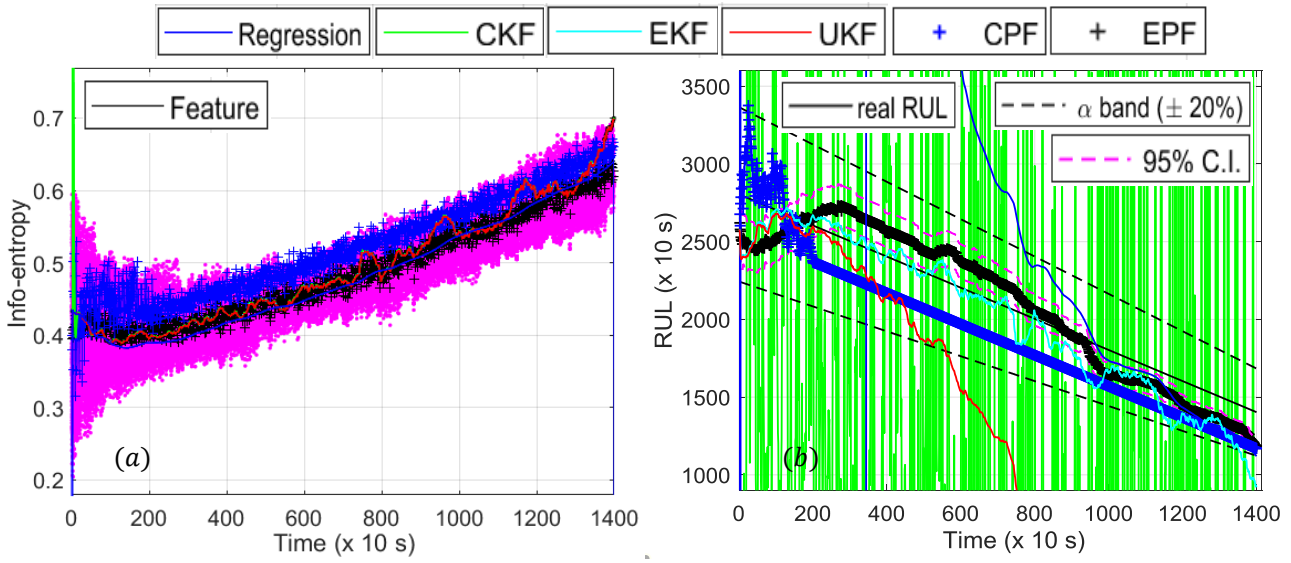

Figure 6: Feature estimation (links) and RUL estimation (right) by different methods for bearing 1-1

The results of bearing 1-1 are presented in Figure 6. In the Figure 6a, the feature is estimated perfectly using the KF versions (i.e. CKF, EKF and UKF). The particle filter generates a number of particles to approximate the posterior distribution of the state in each step. The mean value of the CPF and the EPF is plotted in the Plus (+) sign. From the Figure 6a, the PF tracks the general trend of the feature till the end, and the EPF approximates closer to the feature than CPF. By updating the model parameters in each step, the RUL result is able to be calculated by the extrapolation to the predefined threshold. The corresponding RUL estimation is presented in Figure 6b. The CKF estimates the RUL with severe fluctuation without convergence, while EKF and UKF predict in a similar manner before 200 (x 10 s). Afterwards, UKF diverges from the actual RUL, probably because of the random generated sigma points. However, the EKF follows almost the actual RUL before 900 (x 10 s) and then deviates to the outside of the (+/-) 20% error band. As comparison, the EPF performs the comparable prediction as EKF before 900 (x 10 s) and much closer to the actual RUL than the simple regression. Between 900 (x 10 s) and the end, the EPF goes slightly far from the actual RUL and has a better performance than EKF. It should be noticed that the full lifetime prediction by EPF falls within the error band. Moreover, the uncertainty of the RUL prediction is evaluated with the 95% confidence band, starting in wide uncertainty and decreasing with the time passing by. In

contrast, the performance of CPF seems reliable with the stable distance to the actual RUL line, yet, it has the severe problem of particle degeneracy after 200 (x 10s).

The result of the second example bearing 1-4 is presented in Figure 7. Similarly with the results in Figure 6, the feature estimation of KFs demonstrates better than PFs in Figure 7a, especially for the part after 1000 (x 10 s). In Figure 7b, the performance of different algorithms is shown. The worst results are achieved by the CKF. UKF demonstrates good results before 400 (x 10 s) but then diverges to the outside of the error band. The RUL prediction by EKF seems more reliable than REG, CKF and UKF, but with large biased error before 1000 (x 10 s). Then, it converges fast to the actual RUL and diverges again in the very late stage. Analogously to the Figure 6b, CPF predicts in the form of a straight line after 180 (x 10 s). Although the observed prediction error is small, yet, it has a serious problem of particle degeneracy. As comparison, the EPF performs better results than EKF and REG before 1000 (x 10s), and converges with a biased error w.r.t the actual RUL. Looking back to the Info-entropy in Figure 5d, the failure threshold is set based on the bearing 1-7, which is much higher than the real life of bearing 1-4. This is the reason that the prediction of EPF and EKF has a biased error in this case. Therefore, it can be concluded that obviously the failure threshold set has a critical influence to the final RUL estimation.

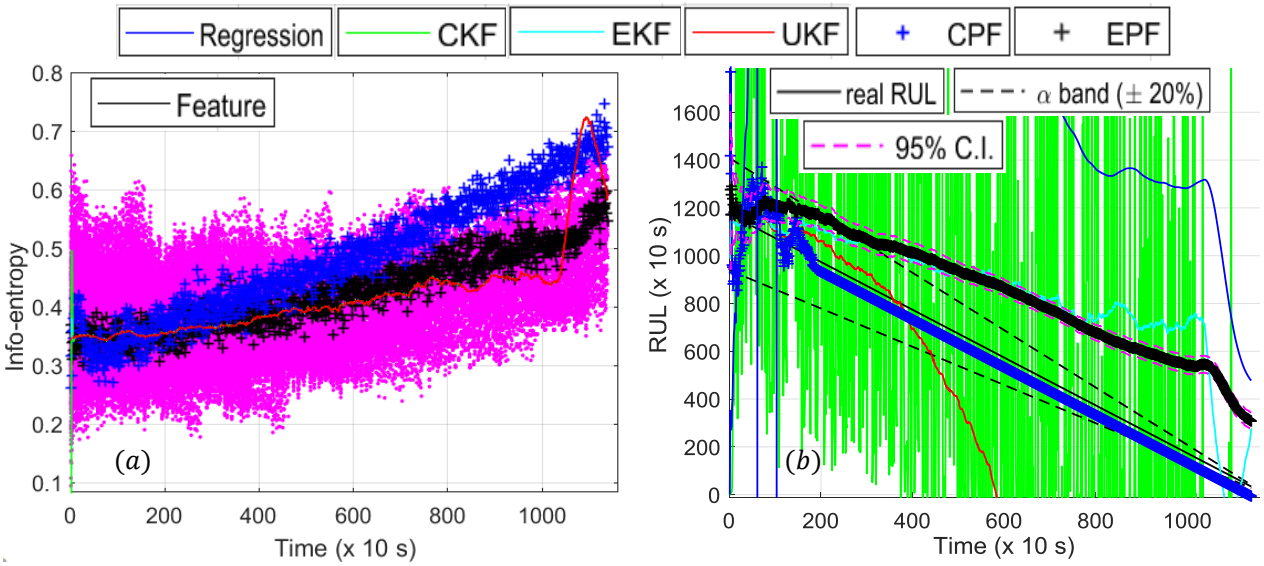

Figure 7: Feature estimation (links) and RUL estimation (right) by different methods for bearing 1-4

In addition to the visual depiction, the performance of these four bearings are presented in Table 2. For all bearings, the CRA of the EPF is the best ,closest to the 1, while the EKF performs better than the rest. With the huge fluctuation, CKF gains a negative CRA value for all bearings here. In the view of the Err, EPF performs the best score almost in total, with the exception of bearing 1-4. REG shows advantage than CKF, EKF and UKF in some bearings as it is calculated however based on the all history data in each step. Moreover, the convergence of EPF is not always best at CRA and Err and it may be surpassed by the CPF and CKF. However, the CPF is easily caught with the problem of degeneracy and the CKF usually has the larger prediction error.

Unlike the RUL prediction using the concept of start prediction point in [9], [11]-[13], with its help the good RUL results are possible to be achieved at the very late stage, i.e. the stage 3 of Figure 1. Using the proposed HI, the RUL in this paper is predicted from the beginning. The prediction of some bearings performs even better than [9] and [11], for example, the CRA of bearing 1-1 has better score than an exponential model in [9] and the bearing 1-3 and 1-7 has smaller error than the LRT technique in [11]. Therefore the proposed Info-entropy is able to track the degradation very early and can achieve good results in the long terms. The prognostics in the proposed methodology becomes more meaningful than state of the art papers, because the maintenance operators have plenty of time to schedule the repair plan, further avoiding a hazardous machinery breakdown.

6 Conclusion

In this paper, a statistical model based prognostics methodology for the RUL estimation of rolling element bearings is proposed. The methods of HI construction, statistical model selection and RUL

estimation are explained in detail. Traditionally, a HI with a high trendability is expected, being capable to reflect the full lifetime degradation. However, the degradation status is commonly not easily tracked, especially in case of less impulsiveness and low frequency resolution. Many classic signal processing techniques may fail to extract high quality HI, which is critical to the selection of the failure threshold, of the model and even of the concept of start prediction point. In order to overcome these problems, entropy, as a chaos measurement, is considered. An HI Info-entropy is constructed based on the entropy indicators in the frequency domain, i.e., the Spectral Entropy, the Spectral Flatness and the Spectral NegEntropy. After the signal decomposition, the most informative frequency band is localized and the entropy indicator that follows the optimal degradation trend is selected.

Through the validation on the experimental dataset, Info-entropy is effective to capture the bearing degradation trend in the full lifetime. It provides a new perspective to construct the high trendability HI in a complex operation and further simplify the prognostics process, e.g. the model selection, the elimination of the concept of the start prediction point. In the end, the Info-entropy is combined with a classic exponential model, and is incorporated with six different methods, i.e. mathematical regression, classic Kalman filter, Extended Kalman filter, Unscented Kalman filter and Particle filter. Through the analysis, systematic resampling based Particle filter is efficient to better enhance the particle diversity than the classic resampling method and it achieves more accurate prediction results than others.

		Bearing 1-1	Bearing 1-3	Bearing 1-4	Bearing 1-7
REG	CRA	-0,07	0,02	-3,69	-5,41
	Err	20,20	-10,19	-1308,02	5,84
	C_m	912,92	855,53	356,43	1110,79
CKF	CRA	-1,59	-7,26	-8,12	-1,77
	Err	617,46	727,73	411,52	649,89
	C_m	167,54	357,92	3125,91	176,38
EKF	CRA	0,91	0,30	-0,30	0,35
	Err	34,38	-36,49	-717,90	-52,45
	C_m	681,80	892,43	410,18	804,69
UKF	CRA	-0,08	-3,46	-10,70	-0,86
	Err	496,72	546,77	283,19	546,74
	C_m	224,40	767,07	7740,37	391,44
CPF	CRA	0,87	0,64	0,81	0,34
	Err	16,64	64,57	119,43	101,15
	C_m	687,63	848,54	541,94	671,61
EPF	CRA	0,93	0,73	-0,50	0,44
	Err	15,52	-0,34	-808,25	-4,37
	C_m	696,84	949,01	407,38	792,10
PF [9]	CRA	0,87	0,76	0,87	0,93
	Err	/	/	/	/
	C_m	8,99	4,28	275,40	290,20
LRT [11]	CRA	0,94	/	/	/
	Err	/	7,00	0	19,00
	C_m	/	/	/	/

Table 2: Comparison of RUL estimation by different methods

References

[1] K. Javed, R. Gouriveau and N. Zerhouni, "State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels," *Mech. Syst. Signal Process.*, vol. 94, pp. 214–236, 2017.

[2] A. Cubillo, S. Perinpanayagam and M. Esperon-Miguez, "A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery," *ADV. MECH. ENG.*, vol. 8, no. 8, pp. 1-21, 2016.

[3] B. Sun, S. Zeng, R. Kang and M. G. Pecht, “Benefits and Challenges of System Prognostics,” *IEEE Trans. Rel.*, vol. 61, no. 2, pp. 323-335, Jun. 2012.
[4] N. Gebraeel, M. Lawley, R. Li and J. Ryan, “Residual Life Distribution from Component Degradation Signals: A Bayesian Approach,” *IIE Trans.*, vol. 37, no. 6, pp. 543-557, 2005.
[5] K. Medjaher, N. Zerhouni, and J. Baklouti, “Data-driven prognostics based on health indicator construction: Application to PRONOSTIA’s data,” in *Proc. Eur. Control Conf.*, 2013, pp. 1451–1456.
[6] D. Tobon-Mejia, K. Medjaher, N. Zerhouni, G. Tripot, “Hidden Markov models for failure diagnostic and prognostic.,” in *Proc. Prognostics Syst. Health Manage. Conf.*, Shenzhen, China, 2011 May, pp. 1–8.
[7] J. Qi, A. Mauricio, M. Sarrazin, K. Janssens and K. Gryllias, “Remaining Useful Life Prediction of Rolling Element Bearings Based on Unscented Kalman Filter,” in *Proc. Int. Conf. on Condition Monitoring of Machinery in Non-stationary operations*, Santander, Spain, 2018 June 20-22, pp. 123–138.
[8] C. K. R. Lim and D. Mba, “Switching Kalman filter for failure prognostic,” *Mech. Syst. Signal Process.*, vol. 52-53, pp. 426-435, Feb. 2015.
[9] N. Li, Y. Lei, J. Lin, and S. X. Ding, “An improved exponential model for predicting remaining useful life of rolling element bearings,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7762–7773, Dec. 2015.
[10] P. Wang, R. Qiang and R. X. Gao, “Multi-Mode Particle Filter for Bearing Remaining Life Prediction.,” in *Proc. Int. Manufacturing Science and Engineering Conf.*, Texas, USA, 2018 June.
[11] W. Ahmad, S. Khan and J. Kim, “A Hybrid Prognostics Technique for Rolling Element Bearings Using Adaptive Predictive Models.,” *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, Feb. 2018.
[12] R. K. Singleton, E. G. Strangas, and S. Aviyente, “Extended Kalman filtering for remaining-useful-life estimation of bearings,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1781–1790, Mar. 2015.
[13] A. Soualhi, K. Medjaher and N. Zerhouni, “Bearing Health monitoring based on Hilbert-Huang Transform, Support Vector Machine and Regression.,” *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 52–62, Jan. 2015.
[14] A. Gray and J. Markel, “A Spectral-Flatness Measure for Studying the Autocorrelation Method of Linear Prediction of Speech Analysis.,” *IEEE Trans Acoust.*, vol. 22, no. 3, pp. 207-217, Jun. 1974.
[15] J. Antoni, “The infogram: Entropic evidence of the signature of repetitive transients.,” *Mech. Syst. Signal Process.*, vol. 74, Nov. 2018.
[16] A. Saxena, J. Celaya, B. Saha, S. Saha and K. Goebel, “On Applying the Prognostic Performance Metrics.,” in *Proc. Prognostics Health Manage. Conf.*, San Diego, USA, 2009 September 27- October 1, pp. 1-16.
[17] P. Nectoux, R. Couriveau, K. Medjaher and E. Ramasso, “PRONOSTIA: An experimental platform for bearings accelerated degradation tests.,” in *Proc. IEEE Int. Conf. on Prognostics Health Manage.*, Denver, USA, 2012 June, pp. 1-8.