Remaining Useful Life Estimation for Railway Gearbox Bearings Using Machine Learning

Lodiana Beqiri, Zeinab Bakhshi, Sasikumar Punnekkat, and Antonio Cicchetti

Mälardalen University, Sweden
{lodiana.beqiri, zeinab.bakhshi, sasikumar.punnekkat, antonio.cicchetti}@mdu.se

Abstract. Gearbox bearing maintenance is one of the major overhaul cost items for railway electric propulsion systems. They are continuously exposed to challenging working conditions, which compromise their performance and reliability. Various maintenance strategies have been introduced over time to improve the operational efficiency of such components, while lowering the cost of their maintenance. One of these is predictive maintenance, which makes use of previous historical data to estimate a component's remaining useful life (RUL). This paper introduces a machine learning-based method for calculating the RUL of railway gearbox bearings. The method uses unlabeled mechanical vibration signals from gearbox bearings to detect patterns of increased bearing wear and predict the component's residual life span. We combined a data smoothing method, a change point algorithm to set thresholds, and regression models for prediction. The proposed method has been validated using real-world gearbox data provided by our industrial partner, Alstom Transport AB in Sweden. The results are promising, particularly with respect to the predicted failure time. Our model predicted the failure to occur on day 330, while the gearbox bearing's actual lifespan was 337 days. The deviation of just 7 days is a significant result, since an earlier RUL prediction value is usually preferable to avoid unexpected failure during operations. Additionally, we plan to further enhance the prediction model by including more data representing failing bearing patterns.

Keywords: Railway · Gearbox bearing · Predictive maintenance · Remaining useful life · Machine learning.

1 Introduction

The maintenance of a railway system plays an important role in ensuring its safety, dependability, and efficiency [1]. Train reliability is a daily requirement for millions of people, and as such, it is a perpetual challenge for all vehicle manufacturers. The reliability expectation is met by using electrical and mechanical components such as robust pantographs, transformers, and an optimized propulsion system. These components are subject to significantly demanding operation conditions, and to preserve their operational performance, effective maintenance strategies are essential. In the context of predictive maintenance, safe and efficient train operation [2] is relied on accurate estimation of RUL of railway components. In this respect, traditional methods, like model-based prediction, leverage complex models, such as non-linear ones [3] or temperature models [4], which can potentially impact the accuracy of the prediction [5]. Machine learning (ML) techniques, on the other hand, offer increasingly popular alternatives that provide improved efficiency and accuracy [2, 6, 7]. ML models can utilize sensor data and other operational parameters to predict the remaining lifespan of bearings, enabling proactive maintenance and minimizing downtime. Data analytics, feature extraction, and ML techniques have shown promising potentials for predicting component failures and estimating RUL [8, 5]. The use of ML techniques for RUL estimation has been explored in various fields, such as wind turbines [9, 10, 5], where high operation and maintenance costs make it essential to predict component failures. However, these techniques have been rarely applied to train propulsion systems or their components. In fact, although there are some similarities between gearbox bearings in different contexts, there exist significant differences. Train propulsion systems operate under specific operational conditions and encounter various environmental variables, including temperature, vibration levels, humidity, etc. As a result, the wear patterns and degradation mechanisms exhibited by train propulsion systems differ from those observed in other domains, influencing the various methods used to assess them. Typically, vibration data from propulsion system gearbox bearings are analyzed using techniques such as Fourier or time-frequency analysis to detect anomalous patterns associated with bearing problems. There data can then be fed to further steps of anomaly detection system, such as, ML algorithms to classify and locate bearing problems.

The current study focuses on the challenges presented by train propulsion system gearbox bearings and their wear. Due to the limitation in obtaining real-world gearbox bearing data, existing works usually depend on simulated data or controlled operating condition data in laboratory settings. Instead, this paper presents a method that has been validated using real data from a train propulsion system with a maximum speed of 300 km/hour given by our industrial partner Alstom. The approach proposes a preprocessing phase that uses low pass filtering to reduce oscillations in raw data [11] and increase RUL estimation accuracy. The obtained data is employed in a regression model to predict RUL. The proposed techniques also includes a change point algorithm, necessary to derive thresholds for assessing degradation trends. By going into more details, the process begins with an analysis of sensor data acquired from a real-world propulsion system. A combination of one class support vector machine (One-class SVM) and interpolation is used, allowing signal outlier identification and trend analysis. In the case of deteriorating trends, a change point technique, Pruned Exact Linear Time (PELT) is applied to the data to identify the signal's variation instances. Based on the variation points thresholds are built. To develop a prediction model, regression models such as the polynomial and the exponential ones are created and trained on the data. The RUL of the bearings is then determined by using the best-performing model. The obtained prediction results are promising, as our prediction model is very close to the actual bearing failure, with only a 7-day difference [1]. In other words, in real-life condition the bearing lasted an additional 7 days before failing. In fact, such a margin would prevent downtime due to in-service failures, while at the same time avoiding excessive waste of remaining lifespan. Moreover, in contrast to previous research, the proposed approach enables the provision of explanations about how thresholds for degradation trends are established through the application of a change point algorithm without the use of domain-specific knowledge. The remainder of the paper is structured as follows: Section 2 provides background information about the research effort, including railway maintenance, the propulsion system, and estimation of RUL. Moreover, Section 3 discusses existing related works and the proposed solutions. Section 4 presents an overview of the adopted research methodology and discusses the data preparation, exploratory data analysis, and prediction models. A summary of the obtained results and findings is illustrated in Section 5, while Section 6 provides conclusive remarks and discusses the possibilities for further development.

2 Background

2.1 Railway Maintenance

Railway maintenance focuses on boosting operating availability and safety of its components, while reducing expenses and downtime [12], and detecting potential issues. To achieve these goals, maintenance should be systematic, with thorough planning and continual monitoring of different components conditions. The maintenance activities are broadly categorized into: reactive, preventive, predictive [13, 14], as further discussed below. Historically, train maintenance has been **reactive**, also known as run-to-failure maintenance. This technique entails simply examining and repairing equipment after it has failed. This is the most basic but least effective strategy, as the cost of interventions and accompanying downtime after failure will be prohibitively expensive, including the potential growth of safety concerns. **Preventive maintenance** is planned and scheduled to reduce the chance of equipment failure while also enhancing production efficiency; in particular, inspections and replacement of components on specific pieces of equipment are performed on a regular basis. Even if better than the reactive approach, preventive maintenance is still more expensive than the predictive one, since while most failure issues are avoided, there exists a high chance of carrying out unneeded remedial activities. **Predictive maintenance** seeks to estimate the failure time of a system or its components based on experience and/or historical data and replace them before they fail. By predicting the need for maintenance in advance, this strategy also helps in improving maintenance planning, which takes time and resources. **Condition-based**

[1] Our model forecasts the failure to occur on day 330, and in reality, the gearbox bearing lasts for 337 days.

maintenance (CBM) is a form of predictive maintenance that shifts the scope of inspections towards changes that could indicate possible failures rather than performing general inspections at regular intervals. In particular, CBM aims at detecting early symptoms of oncoming failures and hence predicting the need for maintenance; typically, it employs sensors measuring variables that may affect the machine's efficiency. Moreover, to assess when/if a defect is detected CBM leverages thresholds to preclude unnecessary replacements and carry out maintenance activities only when required.

2.2 The Propulsion System

The propulsion system of a train generates the required power and force to propel the train, allowing it to move and assuring efficient transportation. The propulsion mechanism comprises of a traction motor attached to the bogie and a gearbox linked to the wheel axle. The method proposed in this paper is validated using data from a train propulsion system provided by the world-leading train manufacturer Alstom. There are eight carriages on the train where the data is collected from, four of which are traction cars. Each of these cars has two bogies. A bogie is a train component that sits beneath the train's body and holds and links all of the locomotive's parts, including axles and wheels. Each bogie has two axles, each with its own traction motor and gearbox. Figure 1 depicts a simplified representation of components in the electric propulsion system analyzed in this study, such as the traction motor, gearbox, and the respective sensor placement. Gearbox bearings have an important role in the functioning of the train propulsion system. Bearings are key components that facilitate the smooth operation and transmission of power within the gearbox. They support the rotating shafts and gears, ensuring proper alignment and reducing friction, thus enabling efficient power transfer. Bearings must withstand frequent movement, varied speeds, and severe loads while retaining performance. They are, however, prone to wear, fatigue, lubrication contamination, and other potential damage. Contact fatigue is the most common cause of bearing failures [15]. Other factors include oxidation, fatigue on the rolling elements, and misalignment of bearings during installation [16]. Many challenges exist in maintaining and ensuring the performance of propulsion system gearbox bearings. Understanding the challenges is crucial for designing effective maintenance and optimizing their performance which ensures the train runs efficiently.

2.3 Estimation of Remaining Useful Life (RUL)

RUL estimate plays a pivotal role as an aspect of predictive maintenance, contributing to the effectiveness of maintenance procedures. RUL for gearbox bearings has attracted considerable attention in the literature, not only in the railway industry, but also in other manufacturers who rely on machinery [11],[17],[5],[18],[19]. Models such as similarity, degradation, and survival models have emerged as tools for forecasting the remaining lifespan of essential railway components.

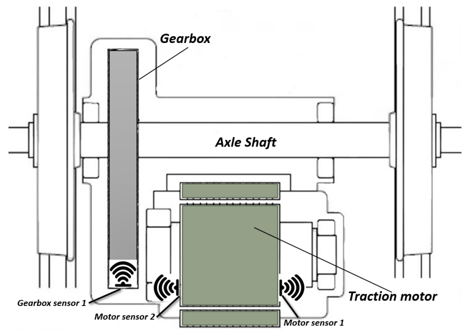

Fig. 1. Components within the Electric Propulsion System

These models aim to anticipate the RUL based on criteria such as wear, deterioration patterns, and historical data. *Similarity models* are based on the RUL forecast of a testing machine based on a historical comparison of known behavior of other similar machines. They use run-to-failure data describing the degradation profile. *Degradation models* extrapolate previous behavior to predict future conditions. If the condition indicator is known to signal failure, regression models are adopted, and the remaining time calculated till some predetermined threshold is reached. In this paper we leverage a degradation model to estimate the RUL. *Survival analysis* is a method for analyzing data based on the time it takes for an event to occur and estimates the probability distribution of failure.

Machine learning approaches and data-driven techniques like as regression models, neural networks, and support vector machines can be used to analyse historical data and trends of bearing degradation. In the following section, we will discuss some of the prior research done in the aforementioned context.

3 Related Work

Carvalho et al. [20] conducted a systematic literature review on predictive maintenance using ML techniques. They explored the equipment types studied and the ML methods used, concluding that there is an increasing trend in using ML for predictive maintenanced, which helps reduce the cost of unnecessary equipment replacement in various applications. Based on Carvalho et al.'s work, we identified related works that explored different ML methods. Amruthnath et al. [21] focused on early failure identification on vibration data from an exhaust fan. They evaluated various algorithms such as T^2 statistics, PCA, hierarchical clustering, K-means, and fuzzy C-means clustering. The authors suggested that clustering techniques can be a cost-effective solution when maintenance costs are high. By monitoring machine health regularly using clustering, expenses on machine maintenance can be saved until a critical level is reached. In another study [22], the same authors proposed an unsupervised learning approach for fault class prediction and detection in a predictive maintenance system. They utilized Gaussian Mixture Models (GMM) and K-means algorithms to forecast the machine state and achieved an 82.96% accuracy for error prediction. Kundu et al. [23] proposed a method for predicting the RUL of rolling bearings using a combination of K-means clustering and change point detection algorithms. By identifying failure patterns in the data, the authors improved the accuracy of their RUL predictions. They suggested that their method could be extended to other applications of predicting RUL and state interference where changing a state shows degradation of the bearing. This method is useful for calculating the probability of shifting from a healthy state to damage by using a state matrix. Hong et al. [24] proposed a method for predicting the RUL of bearings using Gaussian Process Regression (GPR). They utilized RMS, kurtosis, and crest factor as input features to construct the minimum quantization error (MQE) through a self-organizing map (SOM). The authors found that using a composite kernel improved the prediction accuracy and reduced the variance of the RUL in comparison to using a single kernel, highlighting the importance of kernel selection in GPR for RUL prediction in machine health monitoring. Elasha et al. [11] proposed a bearing prognosis approach that used regression and back-propagation neural networks to estimate the RUL of high-speed shaft bearings of a wind turbine. They demonstrated the effectiveness of the regression model in improving the predictive performance of the neural network model, with the proposed ANN model exhibiting strong performance in predicting the remaining useful life of a bearing. Li et al. [25] improved the exponential model and utilized particle filtering to eliminate random faults in bearing degradation process. Their method was demonstrated on four tests and outperformed the original exponential model used in their previous work in predicting RUL of rolling element bearings. While this study enhanced prediction accuracy of the exponential model by selecting optimal FPT and minimizing random errors, failure threshold remains subjective and few studies have been done to determine them dynamically in RUL prediction. The authors in [26] developed a method for estimating the RUL of rolling element bearings in induction motors using dynamic regression models. They used a gradient-based approach to build failure thresholds and developed the time to start prediction (TSP) metric to detect the onset of bearing degradation, after which the trend in the bearing health indicator should be continuously monitored to estimate the RUL. The proposed methodology was evaluated on run-to-failure data, nevertheless, the authors state that further study is required to confirm its efficacy since it was limited to a single dataset.

The studies in this subsection focus on unsupervised learning for fault detection, RUL prediction, and improving exponential model accuracy. Techniques like clustering algorithms, change point detection, Gaussian process regression, adaptive first prediction time selection, and particle filtering can save costs by reducing equipment replacements and improving machine uptime [27, 28]. However, selecting the appropriate technique depends on the equipment, data, and maintenance goals [20].

4 Methodology

This section describes the methodology used in the course of the work. The approach includes exploratory data analysis, trend analysis, and the use of one-class SVM and interpolation to identify and handle the outliers. Following the deployment of the prediction model, PELT is used to detect changes in signal trend and establish state-definers thresholds. The prediction model is applied to the signals that are identified as degrading. Using the training data, multiple regression models are trained, and the best model is chosen to forecast the RUL of gearbox bearings [2]. The whole procedure is presented in Figure 2, and the subsections that follow describe each step.

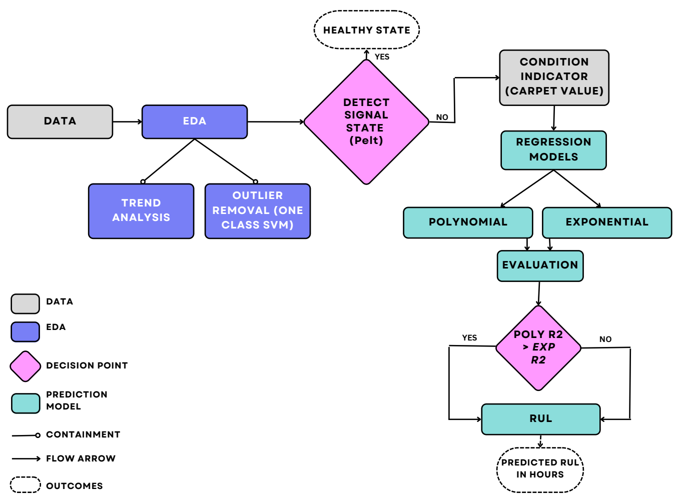

Fig. 2. Proposed Methodology

4.1 Dataset

Accelerometers are mounted on the motor and gearbox casings to capture vibration signals. Two sensors are installed in the traction motor, and two more are installed in the same location in the gearbox, also shown in Figure 1. The sensors are programmed to transmit data to a controller unit, which is then analyzed to determine the prognosis. Data was collected from a train over a 10-month period, and specifically from the four traction cars on the train, with the goal of identifying concerning patterns. The prediction model was constructed using

[2] While the source code remains proprietary, the GitHub repository includes a pseudocode representation `https://github.com/lodianabeqiri/BearingRUL-Estimation-Pseudocode`

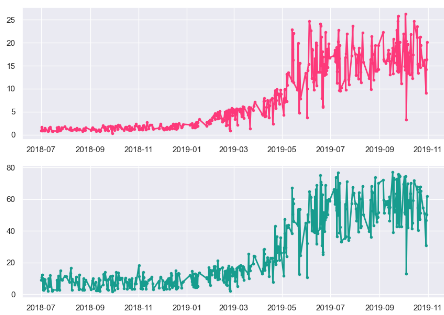

Fig. 3. Carpet and Max Values: Damaged Gearbox Bearing

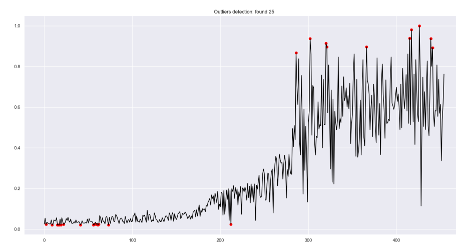

Fig. 4. Outlier Detection

data from another train's damaged gearbox bearing. The gearbox was removed for inspection by Alstom engineers, and a problem with the transmission's bearing was discovered. The signals received through the sensors were further filtered by the engineers to produce two features called as the carpet value and the maximum value. The carpet value reflects the energy level of the signal, while the maximum values represent the signal's peak values. If there is no damage, the carpet values can be used to reveal the amount of vibration, or energy inside the motor and gearbox bearing. Based on domain expertise, when there is bearings deterioration, the carpet value rises as the damage worsens. On the other hand, the maximum values do not always imply component damage, since they depend on both external noise and component degradation. For these reasons, the carpet value is employed as a prediction indication in this paper to uncover data variation associated to failure.

4.2 Exploratory Data Analysis (Eda)

Eda is a crucial step in comprehending the data, making it easier to spot trends and anomalies. The data has been partitioned into training and test sets in a 80/20 proportion. The MinMaxScaler normalization technique has been used on the training set for transforming the numerical values to a common scale without distorting the values range or removing information. The test data has been used to evaluate the model. Figure 3 depicts the maximum value in green and the carpet value in pink from the damaged bearing [3].

Visualization is a useful tool for understanding the data trends. A time series trend refers to the pattern or direction of change that can be observed across time. The moving average and Bollinger Bands were examined in the analysis. Bollinger Bands are standard deviation envelopes that appear above and below a

[3] GitHub repository `https://github.com/lodianabeqiri/RULforBearings_images` contains all the figures presented in this paper.

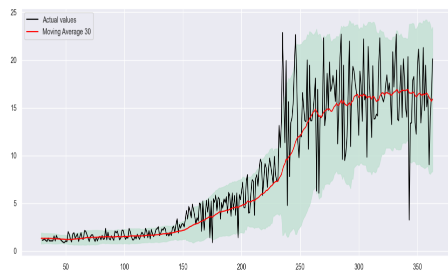

Fig. 5. Trend Analysis: Damaged Gearbox Bearing

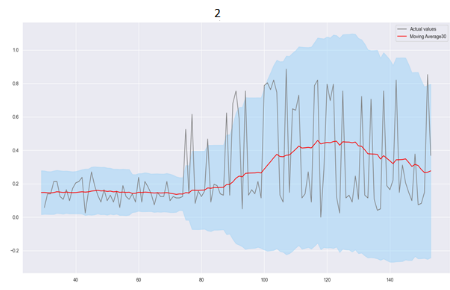

Fig. 6. Trend Analysis: Signal 2

simple moving average. A moving average represents the average value of preceding data points without weighting. The trend of the failed bearing is illustrated in Figure 5. The moving average with a 30-day time window is represented by the red line, while the green zone depicts the Bollinger Bands. The signal does not vary significantly at first, but when the degradation process begins, the values begin to rise. The signal appears to have a stagnant pattern with consistently high values as it approaches the failure phase. Figure 6 illustrates another signal, which shows a slight shift in the trend with a considerable increase in values.

4.3 Outliers

Outliers are observations, also known as abnormalities, that do not fit with the rest of the data. Summary statistics, including mean and variance, can be affected by outliers. Traditional deterministic methods are often applied in practice for outlier identification, such as displays of the distribution and labeling each observation over or below a specific threshold as an outlier. One-class classification is a subfield of ML that focuses on identifying outliers. In this paper we use one-class SVM. It is a SVM variant that captures the density of the majority class and classifies outliers as examples at the density function's extremes. It learns the distribution's bounds referred to as "support" and can thus classify any points outside the boundary as outliers. The algorithm parameter include *nu* that is used to fine-tune the trade-off between overfitting and generalization, parameter *gamma* and the *kernel function*. The decision boundary will be increasingly "linear" as the gamma increases, and the more complex the model, and the greater the risk of overfitting. The *kernel function* changes the training set of data so that a non-linear decision surface can be translated into a linear equation in a larger number of dimension spaces. Figure 4 depicts the outliers as red dots discovered by one-class SVM with the optimized parameters: nu=0.05, kernel="rbf", gamma=0.01. Data distribution could be drastically altered by removing outliers. Therefore linear interpolation was chosen to estimate the missing value by directly linking points in the ascending order.

4.4 Models

This subsection provides an overview of the models used. We examine the rationale for selecting these models and provide a brief explanation of each.

Regression analysis is a type of curve fitting optimization problem, where the objective is to find the best line or curve that fits the data in a way that minimizes the difference between the predicted and actual values. We investigated polynomial and exponential regression models and compared their results to determine the best fit. *Polynomial regression* is a version of linear regression in which a polynomial equation is used to describe the data in order to capture the curvilinear relationship between the independent and dependent variables. The polynomial equation of degree n is represented as:

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_n x^n + \epsilon \tag{1}$$

where y is the dependent variable, x is the independent variable, $\theta_0, \theta_1, \ldots, \theta_n$ are the regression coefficients or weights, ϵ is the error term, and n is the degree of the polynomial. The 1-degree polynomial is a simple linear regression, therefore the degree value must be greater than 1. If the n value is low, the model will struggle to fit the data properly, and if it is high, the model will easily overfit the data.
Exponential regression is the process of determining the best exponential function equation for a set of data. The exponential equation is given as:

$$y = \theta_0 e^{\theta_1 x} + \epsilon \tag{2}$$

where y is the dependent variable, x is the independent variable, θ_0 and θ_1 are the exponential regression coefficients, and ϵ is the error term.
PELT is a change point algorithm that can be used to detect performance declines [29]. There is no unique definition for the term "change point ". They can be regarded as time series points with statistical characteristics that differ from the data distribution. For a given cost function, penalty score, and model, PELT is used to locate the change points in a data set by computing the segmentation of the data that minimizes the cost function. The algorithm uses the pruning rule where many indices are deleted, resulting in a significant reduction in computational cost while maintaining the ability to determine the best segmentation. To find multiple change points, PELT is first applied to the entire dataset and then iteratively and independently applied to each partition until no change points are found.

The bearing degradation process due to measurement noise is vulnerable to a variety of fluctuations, which may affect the model's ability to evaluate the degradation trend [11]. In this case, the data is smoothed before being used as input to the prediction models. The *Savitzky-Golay* is a low pass filter that smooths out data with certain oscillations using a polynomial function, resulting in a signal that is easier to understand and analyze. The technique is repeated for all data points, yielding a new set of data points that closely resembles the original data.

The following paragraphs discuss the practical application of the discussed models on the data. We investigate the process of optimizing these models to ensure their effectiveness in capturing data patterns and discuss their outcomes.

4.5 Setting the Thresholds

The PELT algorithm locates the points in the damaged gearbox bearing carpet value where there is a change or an obvious rising trend. These identified points are used as reference thresholds to assess the health of other signals. The premise behind this approach is that if any other signals have degraded or failed, they may exhibit a similar pattern to the damaged bearing. This assumption was made due to a paucity of data on different failure behavior, and we relied on known failures to detect the similar failure trend. The PELT model was fed with the carpet value and a penalty score and detects the variation point of the signal to build the thresholds. The best parameters for PELT were the Gaussian kernel as a model with a penalty score of 10 and as the cost function the constrained sum of approximation errors. To enhance the signal, the Savgol filter was applied using a window size of 51 and a polynomial order of 3. Subsequently, the filtered signal was passed to PELT. Three vertical lines in Figure 7 represent the detected thresholds for illustration reasons. The red line represents the degradation threshold, and the degradation zone extends from the red to the blue line. When the signal exceed the blue line, it reaches failure. These thresholds are used to determine whether we proceed to compute the RUL for other signal. Based on the thresholds, most of the signals that were evaluated for each car were categorized as healthy. The method was then applied to the damaged bearing signal, the result is that the signal has already failed, and no RUL calculation is required.

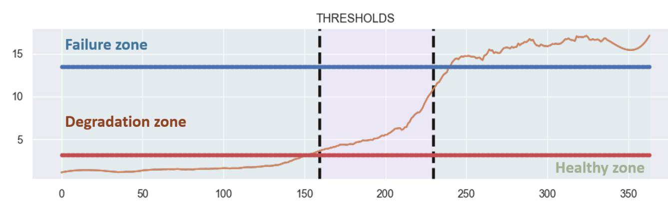

Fig. 7. Thresholds

However, when compared to signal 2 presented in section 4.2 it is classified as degraded, and for this signal the RUL can be determined. Figure 8 shows in green the filtered signal 2. The x-axis represents the time at which the signal was acquired, while the y-axis represents the carpet value of the signal. The signal exhibits an increasing pattern that begins after 2000 hours and ends before 5000 hours. Unlike the damaged signal in Figure 3 , which has a growing tendency over time until failure, the signal 2 is susceptible to the "healing phenomenon" [30]. After the values increase, possibly due to a defect, the signal returns to low values. For instance this could happen as a result of newly formed surface defects caused by the rolling elements of the bearings. Assuming signal 2 will eventually fail after recovery, we employed extrapolation using numpy polyfit in python to forecast future data points and observe the potential trend the signal might follow. Extrapolation is the process of assuming values outside the range of the currently available data by using data from existing points. It is important to

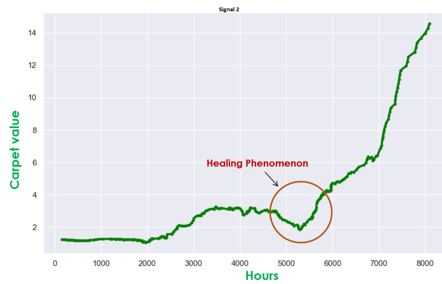

Fig. 8. Signal 2 Extrapolation

note that extrapolation might result in inaccurate results, owing to the variety of gearbox bearing degradation patterns. However, in scenarios with a sufficient number of samples exhibiting the same failure pattern, or multiple failure cases for more complex extrapolation tasks, insights into the signal's future behavior become possible. We conclude that there is insufficient data to calculate the RUL for signal 2.

4.6 RUL

To determine the RUL, three main functions were developed: *actual or true RUL*, another to estimate the *RUL*, and yet another to calculate the *prediction error*. The *actual RUL* is defined as the time elapsed between the true failure of the bearing and the actual time we consider for each point on the training set. The RUL is calculated in terms of hours. We used the last instance of our provided data as the time when the bearing failed, which was 8092 hours in the training data. For the *estimated RUL*, we used the polynomial regression that best fitted the data, calculating the estimated RUL as the difference between the estimated time of bearing's failure and the time determined by our polynomial function for each data point on training set by projecting the point to the hour axes. The estimated time is to be around 7928 hours in the training set by extrapolating the data and finding this value when the data exceeded the failure threshold. Finally, the *prediction error* is defined to determine how well the prediction model performed, as the difference between the actual and predicted RULs.

5 Results

The polynomial and exponential functions applied to the damaged bearing signal are shown in Figures 9 and 10, respectively. Figures 11 shows the model constants and assessment metrics for each regression model. Evaluation tables indicate that the polynomial function exhibited the lowest Root Mean Square Error (RMSE) and the highest R squared (R2) value, indicating the best fit to the data. Consequently, we employ polynomial regression to calculate the RUL. Similar results can be observed when it comes to the test data.

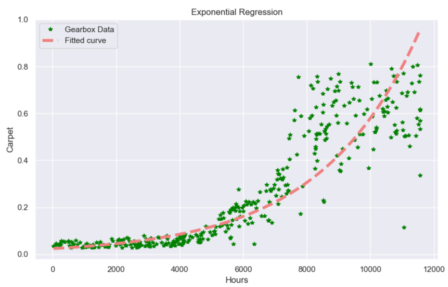

Fig. 9. Exponential Function

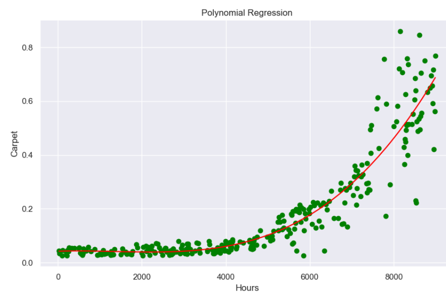

Fig. 10. Polynomial Function

TRAINING SET

REGRESSION MODELS	Model coefficients		RMSE	R Squared
CARPET VALUE				
Exponential Carpet	-3.918663348186325	0.0003624022446393382	**0.01960712346455304**	**0.7866537673654433**
Polynomial	[0.00000000e+00 4.69008693e-06 -6.07368172e-09 1.50563032e-12]		**0.018956361781946**	**0.8716962090304008**

Fig. 11. Assessment Metrics and Regression Coefficients in Training Data

The graph in Figure 12 compares the estimated RUL from polynomial regression, which is visualized in green, to the actual RUL, which is represented by the diagonal black line. The time at which the fitted carpet value exceeded the defined threshold as determined by PELT was then used to extrapolate the expected failure time. Considering only the time the signal entered the degradation phase until it reached the failure threshold (at 8092 hours or 337 days), the actual RUL of the signal is calculated to be 2979 hours. As a result, the signal has 2979 hours until it fails as soon as it enters the degradation phase. Meanwhile, extrapolating the estimated RUL yields 7928 hours or 330 days, implying that the signal has 2812 hours left when it enters the degradation phase. The difference between the actual life and the estimated time left for the bearing to function properly is 168 hours, or 7 days. The green line in Figure 12 regressing the estimated life time underestimates the signal's life by a few hours; however, some time intervals between 1000 and 1500 hours are comparable to the true RUL. The prediction error, mean square error (MSE) is the distance between estimated line and the actual line. Since the estimated values can be less or greater than actual values, a simple sum of differences can be zero and this lead to the incorrect conclusion that the forecast is correct. As we square and use RMSE, all errors are positive, and the mean is positive, indicating that there is some difference between the estimates and the actual. The calculated RMSE is 210. While no RMSE value is universally correct, a lower mean indicates a more accurate forecast.

5.1 Threats to Validity

As any other experimental work, also this paper needs to take into consideration threats to validity [31]. Internal and construct validity have to deal with the

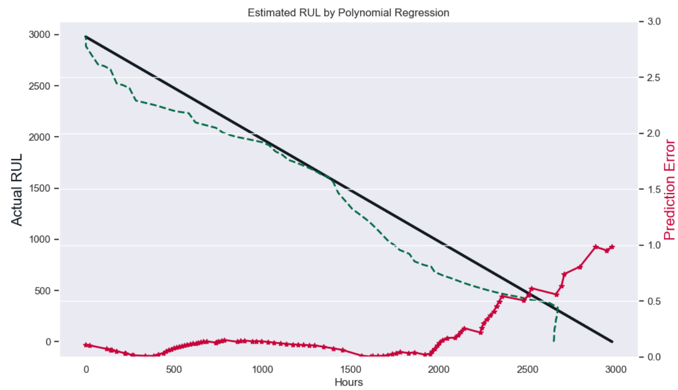

Fig. 12. Regression Model for Predicting RUL

set-up of the experiments and the potential bias of the involved researchers. In this respect, the developed techniques use standard data cleaning and analysis techniques and no refinement procedure, e.g. for outliers and thresholds, has been adopted in accordance with (railway) domain experts. When it comes to conclusion validity, the availability of bearing data is limited to a single failure case. This limitation does not allow us to make conclusions about the precision of the estimation algorithm with an adequate level of confidence. Nonetheless, the approach is generic into distinguishing between healthy and unhealthy states by utilizing domain knowledge about signal characteristics. While the approach accurately detects the states, it is worth mentioning that by integrating more data depicting failure bearing patterns, the precision of the predictions can be improved.

6 Conclusion

A RUL prediction method has been developed by leveraging real-world data from gearbox bearings. Initially, the data was analyzed to identify increasing trends, and the one-class SVM was used to eliminate any outliers. Moreover, the PELT algorithm identified changes in signal properties, allowing the construction of valid degradation thresholds without prior knowledge of them. Two regression models were trained and compared, with polynomial regression achieving a higher R2 value of 0.87 compared to exponential regression. RUL prediction utilized the polynomial function and anticipated the failure with 7 days in advance to the real failure. Furthermore, it was observed that adding data filtering techniques to our model, such as a low pass filter, significantly improves the performance and helps to smooth out the fluctuations in the gearbox degradation trend. The method, however, is dependent on a certain known degradation trend, resulting in restricted data for defining the thresholds. More data on similar behaviors will allow for the construction of more rigorous degradation criteria.

References

[1] Marco Macchi et al. "Maintenance management of railway infrastructures based on reliability analysis". In: *Reliability Engineering & System Safety* 104 (2012), pp. 71–83.

[2] Dechen Yao et al. "Remaining useful life prediction of roller bearings based on improved 1D-CNN and simple recurrent unit". In: *Measurement* 175 (2021), p. 109166.

[3] Mengnan Cao et al. "Study of wind turbine fault diagnosis based on unscented Kalman filter and SCADA data". In: *Energies* 9.10 (2016), p. 847.

[4] Yingning Qiu et al. "Applying thermophysics for wind turbine drivetrain fault diagnosis using SCADA data". In: *IET Renewable Power Generation* 10.5 (2016), pp. 661–668.

[5] James Carroll et al. "Wind turbine gearbox failure and remaining useful life prediction using machine learning techniques". In: *Wind Energy* 22.3 (2019), pp. 360–375.

[6] Yu Zang et al. "Hybrid remaining useful life prediction method. A case study on railway D-cables". In: *Reliability Engineering & System Safety* 213 (2021).

[7] Mengru Hou, Dechang Pi, and Bingrong Li. "Similarity-based deep learning approach for remaining useful life prediction". In: *Measurement* 159 (2020).

[8] ASAE Zaher et al. "Online wind turbine fault detection through automated SCADA data analysis". In: *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* 12.6 (2009), pp. 574–593.

[9] Yanhui Feng et al. "Monitoring wind turbine gearboxes". In: *Wind Energy* 16.5 (2013), pp. 728–740.

[10] Pramod Bangalore et al. "An artificial neural network-based condition monitoring method for wind turbines, with application to the monitoring of the gearbox". In: *Wind Energy* 20.8 (2017), pp. 1421–1438.

[11] Faris Elasha et al. "Prognosis of a wind turbine gearbox bearing using supervised machine learning". In: *Sensors* 19.14 (2019), p. 3092.

[12] Weiping Shao, Yongping Hao, et al. "Study on preventive maintenance strategies of filling equipment based on reliability-cantered maintenance". In: *Tehnički vjesnik* 28.2 (2021), pp. 689–697.

[13] Gian Antonio Susto et al. "Machine learning for predictive maintenance: A multiple classifier approach". In: *IEEE transactions on industrial informatics* 11.3 (2014), pp. 812–820.

[14] Jiawei Xie et al. "Systematic literature review on data-driven models for predictive maintenance of railway track: Implications in geotechnical engineering". In: *Geosciences* 10.11 (2020), p. 425.

[15] Farshid Sadeghi et al. "A review of rolling contact fatigue". In: (2009).

[16] Han Peng et al. "A Review of Research on Wind Turbine Bearings' Failure Analysis and Fault Diagnosis". In: *Lubricants* 11.1 (2022), p. 14.

[17] Milad Rezamand et al. "An integrated feature-based failure prognosis method for wind turbine bearings". In: *IEEE/ASME Transactions on Mechatronics* 25.3 (2020), pp. 1468–1478.

[18] Wei Teng et al. "Prognosis of the remaining useful life of bearings in a wind turbine gearbox". In: *Energies* 10.1 (2016), p. 32.

[19] Mohamed Elforjani and Suliman Shanbr. "Prognosis of bearing acoustic emission signals using supervised machine learning". In: *IEEE Transactions on industrial electronics* 65.7 (2017), pp. 5864–5871.

[20] Thyago P Carvalho et al. "A systematic literature review of machine learning methods applied to predictive maintenance". In: *Computers & Industrial Engineering* 137 (2019), p. 106024.

[21] Nagdev Amruthnath and Tarun Gupta. "A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance". In: *2018 5th international conference on industrial engineering and applications (ICIEA)*. IEEE. 2018, pp. 355–361.

[22] Nagdev Amruthnath and Tarun Gupta. "Fault class prediction in unsupervised learning using model-based clustering approach". In: *2018 International Conference on Information and Computer Technologies (ICICT)*. IEEE. 2018.

[23] Pradeep Kundu, Seema Chopra, and Bhupesh K Lad. "Multiple failure behaviors identification and remaining useful life prediction of ball bearings". In: *Journal of Intelligent Manufacturing* 30 (2019), pp. 1795–1807.

[24] Sheng Hong et al. "Bearing remaining life prediction using Gaussian process regression with composite kernel functions". In: *Journal of Vibroengineering* 17.2 (2015), pp. 695–704.

[25] Naipeng Li et al. "An improved exponential model for predicting remaining useful life of rolling element bearings". In: *IEEE Transactions on Industrial Electronics* 62.12 (2015), pp. 7762–7773.

[26] Wasim Ahmad et al. "A reliable technique for remaining useful life estimation of rolling element bearings using dynamic regression models". In: *Reliability Engineering & System Safety* 184 (2019), pp. 67–76.

[27] Hashem M Hashemian. "State-of-the-art predictive maintenance techniques". In: *IEEE Transactions on Instrumentation and measurement* 60.1 (2010), pp. 226–236.

[28] Sujata Butte, AR Prashanth, and Sainath Patil. "Machine learning based predictive maintenance strategy: a super learning approach with deep neural networks". In: *2018 IEEE Workshop on Microelectronics and Electron Devices (WMED)*. IEEE. 2018, pp. 1–5.

[29] G Dorcas Wambui, Gichuhi Anthony Waititu, and Anthony Wanjoya. "The power of the pruned exact linear time (PELT) test in multiple changepoint detection". In: *American Journal of Theoretical and Applied Statistics* 4.6 (2015).

[30] Mohamad-Ali Mortada, Soumaya Yacout, and Aouni Lakis. "Diagnosis of rotor bearings using logical analysis of data". In: *Journal of Quality in Maintenance Engineering* 17.4 (2011), pp. 371–397.

[31] Claes Wohlin et al. *Experimentation in software engineering*. Springer Science & Business Media, 2012.