

# Condition Monitoring of Slow-speed Gear Wear using a Transmission Error-based Approach with Automated Feature Selection

Stefan Sendlbeck<sup>1</sup>, Alexander Fimpel<sup>2</sup>, Benedikt Siewerin<sup>3</sup>, Michael Otto<sup>4</sup>, and Karsten Stahl<sup>5</sup>

<sup>1,3,4,5</sup> *Institute of Machine Elements, Gear Research Centre (FZG), Technical University of Munich (TUM), Munich, 80333, Germany*

*sendlbeck@fzg.mw.tum.de*

*siewerin@fzg.mw.tum.de*

*otto@fzg.mw.tum.de*

*stahl@fzg.mw.tum.de*

<sup>2</sup> *Technical University of Munich (TUM), Munich, 80333, Germany*

*alexander.fimpel@tum.de*

## ABSTRACT

Gear flank changes caused by wear do not only affect the dynamic behavior of gear systems, but they can also compromise the load-carrying capacity of gear teeth up to critical failure. To help avoid unintended consequences like downtime or safety risks, a condition monitoring system needs to be able to estimate the current wear during operation based on available sensor measurements. While many condition monitoring approaches in research rely on vibrational analysis with manual feature engineering, gearboxes running at slow speed do not reveal much excitation information for this purpose. We therefore introduce an approach for slow-speed gear wear monitoring that is based on the dynamic gear transmission error and that contains an automated feature selection process. For this purpose, we extract a large set of features from the preprocessed transmission error samples. Applying combined filter and embedded feature selection methods enables us to automatically identify and remove features with low relevance. The selection process consists of filtering features with no statistical dependence on the target wear value, removing redundant features with a correlation analysis and a recursive feature elimination process with cross-validation based on a random forest regressor. The remaining relevant set of features is the basis for model training and subsequent wear estimation. For this, the present research employed two independent ensemble models, random forest regression and gradient boosted regression trees. To train and test the proposed approach, we conducted slow-speed gear experiments with developing

gear wear on a single-stage spur gear test rig setup. The results of both models show good gear wear estimation performance compared to the actual wear mass loss, even for small quantities. Hence, the proposed transmission error-based approach with automated feature selection is able to quantify the degree of slow-speed wear and offers a possible way for condition monitoring and fault diagnosis.

## 1. INTRODUCTION

Due to the immense advances in information and sensor technology within recent years, machine learning algorithms together with growing computational power have improved numerous applications and offer many new possibilities. Applied to the field of machine condition monitoring, such approaches seem promising. A reliable condition monitoring system with high accuracy is the key for fault diagnosis and predictive maintenance strategies. It not only increases economic efficiency, but also ensures high safety for humans and machines. Therefore, it is necessary to acquire appropriate data, process and evaluate the data, and derive the machine's condition within a certain time and precision. When it comes to gears, slow-speed wear is usually a phenomenon that is undesirable in a larger extent. Gradual material abrasion from the tooth flank can change the dynamic behavior of the entire system and can affect the gears' load-carrying capacity up to critical failure. Therefore, a working condition monitoring system with regard to slow-speed gear wear can prevent such failures and helps to prepare adequate measures.

With the present work, we address condition monitoring possibilities with focus on gear wear at slow rotational speeds. This research aims at correlating the dynamic transmission error with the corresponding degree of accumulated slow-speed wear of pinion and wheel.

Stefan Sendlbeck et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://doi.org/10.36001/IJPHM.2021.v12i2.3026>

Therefore, the idea is to combine a feature-based condition monitoring approach built on the dynamic gear transmission error with an automated feature selection process. While many current condition monitoring systems and research efforts rely on vibration-based analysis, low vibration excitation at slow operating speeds tends to provide not enough information for accurate slow-speed wear diagnosis. However, the latest sensor technology makes angular measurements of high quality possible. Consequently, we instead use the transmission error, which is very accurate at slow speeds and has a rapid reaction time to gear changes, as a basis for analysis. The present approach with a feature elimination process as part of feature selection, on the one hand allows, for interpretability in contrast to various deep learning approaches. On the other hand, it still requires an enormous number of features with a minimum amount of domain knowledge. The approach is validated on multiple gear tests with actual occurring slow-speed wear with detailed “run-to-failure” documentation.

With the present paper, we first provide a review of previous research. Secondly, we elucidate the methods used that contain the experiment setup, the data acquisition, data preprocessing, and the condition monitoring model building and deployment including feature extraction and selection. We then give an insight into the achieved results and discuss them in terms of slow-speed wear condition monitoring.

## 2. REVIEW OF PREVIOUS RESEARCH

Estimation and prognosis of machine health has seen extensive research in academia and industry. Approaches that include machine learning methods have especially gained traction in this past decade, as the application of these advanced methods is made viable with the release of open-source and commercially available software libraries. In specialist and academic literature, two different types of approaches utilizing machine learning for gear condition monitoring systems exist:

- Methodologies with feature engineering
- Methodologies with feature learning (deep learning)

Typically, feature engineering approaches extract features from gear vibration signals. Many different types of feature candidates for this specific purpose have been proposed. These feature candidates can be classified mainly as time domain features, frequency domain features or time-frequency domain features. For gearbox fault diagnosis, for example, statistical features and higher order moments can be calculated from the raw, the derivative and the integral vibration signal as time domain features (Samanta, 2004). Examples for features extracted from the frequency domain of vibration signals are statistical metrics from individual frequency bands (Cerrada et al., 2016). Finally, time-frequency domain features can be calculated, for instance, with the continuous wavelet transform, the discrete wavelet transform, and the wavelet packet transform (Yan, Gao, &

Chen, 2014). Other popular sources of time-frequency domain features are the short time Fourier transform, the Wigner Ville distribution and the Hilbert Huang transform (Soualhi et al., 2018). The features calculated by these signal processing techniques are usually hand-picked and require domain specific knowledge. Generally, these engineered features attempt to capture changes in frequency and amplitude modulation of the vibration signal emitted by the meshing gears. McFadden (1986) developed a mathematical representation of this behavior and in the course of the last few decades, i.e. Wu, Zuo, and Parey (2008) and Feng and Zuo (2012) refined this representation for different types of gearboxes. The extracted features may be used in combination with machine learning methods, for example, with artificial neural networks (Pacheco et al., 2016), support vector machines (Shen, Wang, Kong, & Tse, 2013), random forests and others (Han, Jiang, Zhao, Wang, & Yin, 2018) to diagnose faults in gears and rotating machinery. In some cases, the set of features and the hyperparameters for the machine learning algorithm are selected with the support of metaheuristics, i.e. genetic algorithms (Cerrada et al., 2016; Samanta, 2004).

Deep learning models have seen overwhelming success and enabled performance improvements for notoriously difficult problems in the field of computer vision and natural language processing. The use of these models is mostly in an end-to-end manner, thus making manual feature engineering unnecessary. Inspired by this kind of success, many researchers have introduced a deep learning workflow to various applications of condition monitoring. For example, Janssens et al. (2016) used a convolutional neural network to extract features automatically from vibration signals caused by rolling bearings to identify multiple different faulty bearing conditions. In this study, the authors show that compared to other machine learning methods that require feature engineering, e.g. support vector machines and random forest regressions, the convolutional neural network performs better. Zhao, Yan, Wang, and Mao (2017) took this idea further by introducing bi-directional LSTM (long-short term memory) layers in addition to convolutional layers. LSTM layers are used specifically for temporal modeling in artificial neuronal network architectures. The authors use the proposed network to predict the slow-speed wear condition of machine tools in milling machines. Lee, Kim, Kim, Hur, and Kim (2018) and Yang, Huang, Lu, and Zhong (2018) developed similar methods that incorporate LSTM layers. Approaches also exist that differ from the common combination of convolutional and LSTM layers, for example, by using sparse filtering for feature learning (Lei, Jia, Lin, Xing, & Ding, 2016).

Methodologies to specifically model, simulate and determine wear and wear progression of gears exist in and out of context of condition monitoring and fault diagnosis. Choy, Polyshchuk, Zakrajsek, Handschuh, and Townsend

(1996) presented an analytical approach that models influence of pitting, gear wear and tooth fracture to the vibration signal of gearboxes. The authors imply that this model may also be applicable for gear fault diagnosis. Kuang and Lin (2001) and Ding and Kahraman (2007) use a similar approach. Hu, Smith, Randall, and Peng (2016) developed specialized condition indicators to specifically quantify and assess gear wear based on vibration measurements. According to the authors, these condition indicators may be used in combination with traditional condition indicators and particle analysis. Furthermore, Fromberger et al. (2016) presented different condition monitoring methods for gear fault diagnosis with special focus on utilizing position encoders and transmission error determination. In a study conducted by Chin, Smith, Borghesani, Randall, and Peng (2021), gear wear is determined by assessing changes of the DC component in the transmission error signal. This approach therefore makes use of typically omitted components of the transmission error signal.

The current state of the art offers various approaches to monitor the condition of rotating machines and gears in particular. While the majority of research is based on vibrational data, only few studies consider the transmission error of gears, which seems especially suitable for consideration at slow speeds where the vibration excitation of transmissions is low. In addition, many datasets originate from analytical models for estimating gear wear, whereas actual lubricated slow-speed wear data is rarely used for analysis, although it is very relevant in practice. In terms of algorithms, feature engineering methods, particularly with automated feature selection, have proven to be suitable for various condition monitoring tasks, but their use for gear transmissions especially in combination with slow-speed wear lacks groundwork.

We therefore see the need to evaluate transmission error data from actual lubricated slow-speed wear tests with low amounts of wear in combination with state-of-the-art automated feature engineering and machine learning methods for slow-speed wear condition monitoring. While both feature engineering and feature learning approaches seem to be promising, we will focus on automated feature engineering, due to interpretability advantages and the properties of the dataset available.

### 3. METHODS

With the goal in mind to identify and quantify gear damage, particularly slow-speed wear, we developed an approach and validated it with gear tests. The following sections contain the experiment setup, data acquisition, data processing and condition monitoring methods.

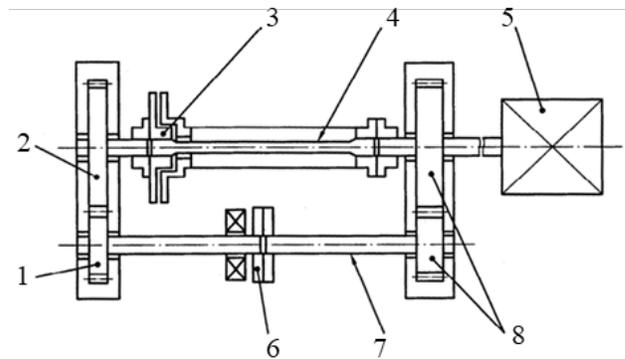
### 3.1. Experiments and Slow-speed Wear Data

For validation, we conducted gear experiments and measured vibrations together with the angular positions of the gear shafts. This allows us to calculate the dynamic transmission error, which is the basis for further processing and for methods presented in this paper in terms of condition monitoring and health evaluation.

#### 3.1.1. Test Rig

For the experiments of slow-speed wear investigation, we used a test rig based on a back-to-back principle according to DIN ISO 14635-1 (2006) (see Figure 1).

The test rig contains a single stage test gearbox with two mating gears, a test pinion (1) and a test wheel (2). The electric motor (5) drives the gearbox via the input shaft (4). Moreover, the test gearbox is part of a mechanical power loop, including a further shaft (7) and a slave gearbox (8) with an identical transmission ratio as in the pair of test gears. The load-carrying clutch (6) enables twisting and subsequent fixing of the shaft (7). Thus, it is possible to apply torque moment to the shafts, thereby loading the gears with a predefined moment. To make sure that damage occurs predominantly on the pair of test gears, the slave gearbox has a higher load-carrying capacity than the test gearbox, i.e. by designing the slave gears with a larger face width.



1	Test pinion	5	Electric motor
2	Test wheel	6	Load-carrying clutch
3	Torque measurement	7	Shaft 1
4	Shaft 2	8	Slave gearbox

Figure 1. FZG standard test rig according to DIN ISO 14635-1 (2006).

In addition to the FZG standard back-to-back test rig (see Figure 1), the test rig used here contains a reduction gearbox interconnected between the electric motor and the slave gearbox. This way, slow operating speeds of the test gears are possible.

### 3.1.2. Experiments

Figure 2 shows the pair of test spur gears used for the slow-speed wear testing. Further gear data are listed in Table 1.

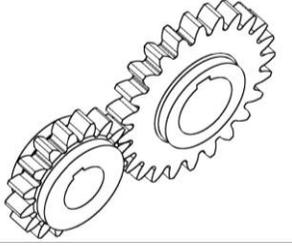


Figure 2. Test gear set.

In total, we conducted three gear experiments, named experiment A, experiment B, and experiment C. The slow-speed wear experiments lasted for approximately 120 h each. Occasional pauses with disassembly and assembly after a defined number of load cycles were necessary to quantify wear. The amount of wear is the accumulated mass loss, determined by the weighing of pinion and wheel. The relative gear position of pinion and wheel did not change due to the weighing process and the lubricant remained in the gearbox. The operating pinion shaft speed for testing was 13.4 rpm, the corresponding pinion shaft load 627 Nm.

For additional details, refer to Siewerin, Dobler, Tobie, and Stahl (2020) and Schultheiss, Tobie, Michaelis, Höhn, and Stahl (2014), who describe the fundamental experiment setup and procedure.

Gear data	Value
Tooth number $z_1 / z_2$	16 / 24
Normal module $m_n$	4.5 mm
Normal pressure angle $\alpha_n$	20°
Working pressure angle $\alpha_w$	22.44°
Center distance $a$	91.5 mm
Face width $b$	14 mm
Helix angle $\beta$	0°
Profile modification factor $x_1 / x_2$	0.1817 / 0.1715
Tip diameter $d_{a1} / d_{a2}$	82.46 mm / 118.36 mm

Table 1. Gear data, see also Siewerin et al. (2020).

### 3.1.3. Slow-speed Wear Results – Ground Truth

Occurring slow-speed wear not only resulted in characteristic marks (see Figure 3), but also led to a mass loss of the gears. The accumulated wear for three different experiments (A, B, and C) with regard to their runtime are displayed in Figure 4 and in the following serve as the ground truth for the condition monitoring approaches.

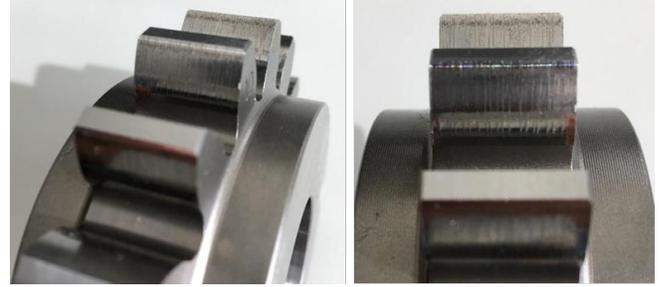


Figure 3. Example for slow-speed wear of test gear pinion.

The distinct differences in the severity of the wear behavior of experiment A, B, and C fundamentally result from different oil or grease used for lubrication. We will not further investigate this lubrication influence in the present work, but refer to Siewerin et al. (2020) for more details. With this paper, we instead focus on investigating the condition monitoring possibilities with regard to wear.

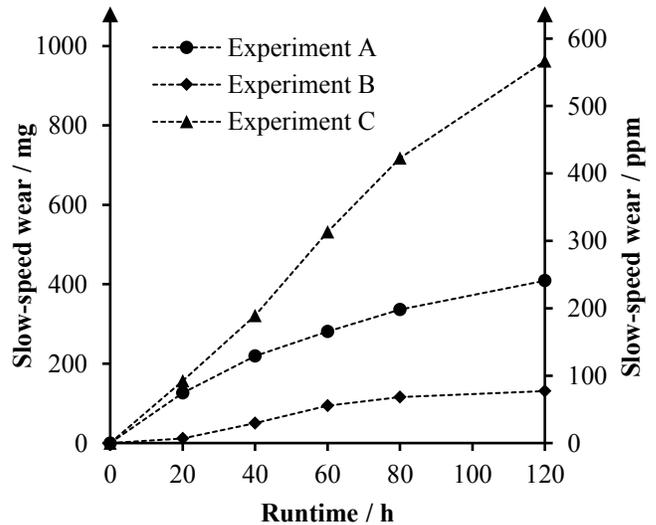


Figure 4. Measured accumulated wear of pinion and wheel for experiment A, experiment B, and experiment C.

### 3.2. Condition Monitoring Workflow

The fundamental workflow we used to develop condition monitoring systems is depicted in Figure 5a). A crucial step in the condition monitoring workflow is model building; hence, it is further specified in Figure 5b). The methodology we developed uses a feature-based model building step, which requires feature engineering. In order to reduce the required amount of domain knowledge and human interaction typically associated with feature engineering, we used a selection process that automatically discovers promising feature candidates, as depicted in Figure 5c). This section aims to provide insights into each individual step and provides background information on the design choices that we made.

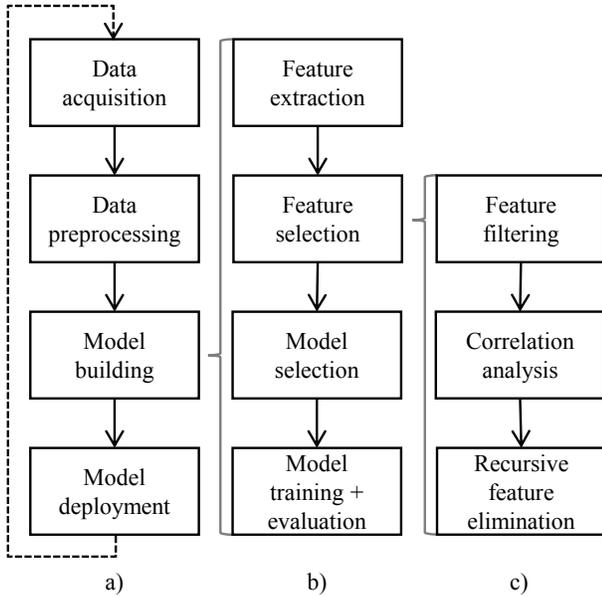


Figure 5. Fundamental workflows for a) condition monitoring, b) model building with feature-based models and c) feature selection.

### 3.2.1. Data Acquisition

The first step in a condition monitoring workflow is data acquisition. Typical methods to acquire data from machines in operation are debris analysis, thermography, performance analysis and vibration analysis (Randall, 2011). The methodology of data acquisition described in the following captures torsional vibration and angular movement. Data acquisition culminates in a dataset consisting of time series that each contain a transmission error signal for a specific slow-speed wear condition of the meshing gears. For this task, we defined a time series containing transmission error as the predictor variable, while the corresponding wear condition is used as the target variable. Wear condition were quantified as a combined weight loss of both pinion and wheel (see Section 3.1.3), which is reasonable because of the same material and heat treatment.

For research in terms of condition monitoring, we added additional measurement equipment to the test rig described in Section 3.1.1. The input and output shaft of the test gearbox each contain an angle measurement sensor (see Figure 6). Therewith, it is possible to measure the absolute angle position of the input shaft as well as the angle position of the output shaft synchronously. The angular sensors work contactless based on optical acquisition. Furthermore, they come with a resolution of 32 bit. The sample rate used in the present experiments was 8.67 kHz.

The angular measurement took place in an intermittent manner every 20 minutes for 60 seconds throughout the entire experiment runtime.

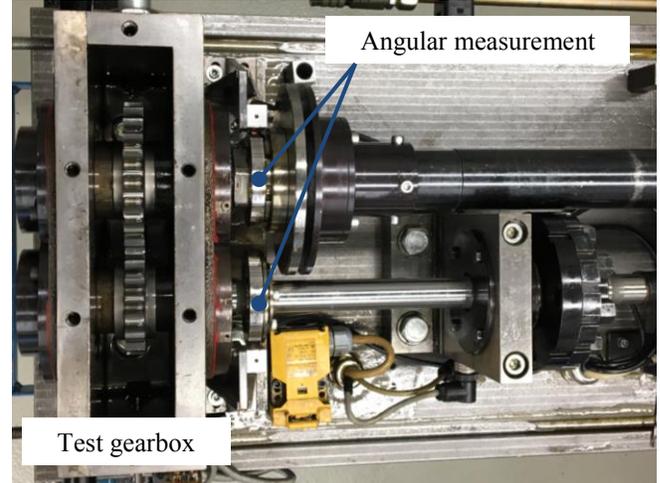


Figure 6. Test gearbox and measurement equipment.

### 3.2.2. Data Preprocessing

Subsequently, the acquired data is preprocessed. Section 3.1.3 contains the determination of the slow-speed wear target variable. We carried out five weight measurements during each experiment, which makes it possible to approximate the wear condition by linear interpolation. The continuous lines in Figure 4 show an approximation of how slow-speed gear wear evolves over time. From this point onwards, we used the relative combined weight loss instead of the absolute combined weight loss to improve comparability among experiments A, B and C. The unit that we used for relative combined weight loss is parts per million (ppm).

In a first preprocessing step, we used the synchronously measured angular data of the shafts to calculate the transmission error  $\Delta\varphi(t)$  of the pair of test gears (see Eq. (1), (Brecher, Gorgels, Hesse, & Hellmann, 2011; Niemann & Winter, 2003)) with  $z_{1,2}$  as the teeth numbers and  $\Delta\varphi_{1,2}(t)$  as the rotation angles of pinion and wheel.

$$\Delta\varphi(t) = \Delta\varphi_2(t) - \frac{z_1}{z_2} \Delta\varphi_1(t) \quad (1)$$

Because of the measurement location outside of the test gearbox, the transmission error includes the transfer path of the sensor to the corresponding gear. However, due to the high stiffness of the shafts compared to the gear mesh and the proximity of the sensors and gears' positions on the shafts, we do not see a considerable influence.

Preprocessing also includes outlier detection, removal and replacement (Aggarwal, 2017). The approach we employed to identify outliers is a modified version of the commonly used z-value test. These so-called robust z-values are calculated using Eq. (2) (Pedregosa et al., 2011, 2020). Notice how, compared to regular calculation of z-values, robust z-values  $z_t^r$  use the median  $\tilde{x}$  of all signal values  $x$

instead of the mean in the numerator and an interquartile range (IQR) between the third quartile  $Q_3$  and the first quartile  $Q_1$  instead of the standard deviation in the denominator. These modifications ensure robustness towards outliers with values that differ from the rest of the signal by several magnitudes. Hence, individual signal values  $x_t$  within a single signal that correspond to a robust z-value  $z_t^r$  higher than 10 are unlikely to be caused by nominal test rig operation. Therefore, these values are subject to either removal or replacement, depending on the temporal location within the signal. This outlier removal procedure is only viable because we assume the underlying data generation process to be approximately stationary.

$$z_t^r = \frac{x_t - \tilde{x}}{Q_3 - Q_1} \quad (2)$$

An exemplary amplitude spectrum of the unfiltered transmission error signal with annotations for selected characteristic frequencies is shown in Figure 7.

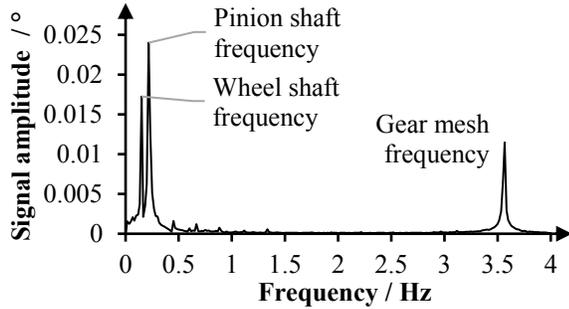


Figure 7. Exemplary excerpt of a transmission error frequency spectrum showing the characteristic frequencies.

Frequency	Value
Gear mesh $f_{gm}$	3.56 Hz
Pinion shaft $f_1$	0.22 Hz
Wheel shaft $f_2$	0.15 Hz
Hunting tooth $f_{ht}$	0.07 Hz

Table 2. Frequencies in the gear transmission system.

Parameter	Value
Filter type	Butterworth
Passband edge frequency	Gear mesh frequency
Stopband edge frequency	Pinion shaft frequency
Maximum loss in passband	1 dB
Minimum attenuation in stopband	40 dB

Table 3. Filter design parameters to remove low frequency components from the transmission error signal.

The exact values for these characteristic frequencies at nominal test rig speed are listed in Table 2. Based on the assumption that low frequency components in the

transmission error signal do not contain relevant information about the wear condition of the meshing gears, we removed these frequencies with a digital high-pass filter. Table 3 contains appropriate filter design parameters to accomplish filtering. Both passband and stopband edge frequencies are dynamically adapted to accommodate minor speed fluctuations of the test rig in order to preserve critical information in the signal.

Figure 8 shows an exemplary excerpt of the high-pass filtered transmission error from the recorded data for one of the tests. Due to the occurring slow-speed wear, the transmission error changes from the beginning to the end of the test. This change is the basis for the fault diagnosis approach presented in this paper.

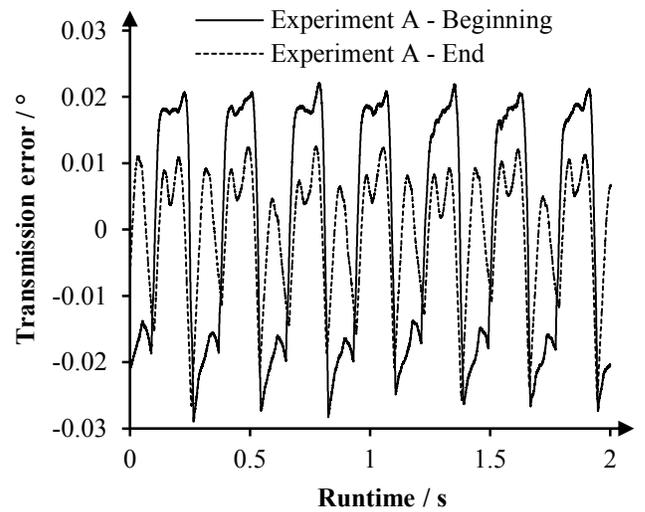


Figure 8. Exemplary transmission error (high-pass filtered) of experiment A at the beginning (no wear) and end of the test (occurring wear) for approximately seven mesh periods.

After filtering, we split each individual measurement signal into multiple sub-series. For meshing gears, multiple natural frequencies exist to divide the signal. Typical split frequencies coincide with those mentioned in Table 2 and their respective harmonics. The fundamental frequency of the transmission error signal matches the hunting tooth frequency  $f_{ht}$ , which is calculated according to Eq. (3) (Mark, 2015; Scheffer & Girdhar, 2004), with  $f_{gm}$  as the gear mesh frequency. For this, the greatest common divisor  $z_{gcd}$  of the number of teeth on the pinion  $z_1$  and the wheel  $z_2$  needs to be determined. Successively, each measurement is split into single hunting tooth cycles in order to capture every possible tooth contact in a sub-series of a measurement.

$$f_{ht} = \frac{f_{gm} \cdot z_{gcd}}{z_1 \cdot z_2} \quad (3)$$

Applied to the existing data, this step yields a total of 1,372 sub-series for experiment A, 1,384 sub-series for

experiment B, and 1,353 sub-series for experiment C. The slight difference in the number of sub-series results from minimally different durations of each experiment.

### 3.2.3. Model Building and Model Deployment

We then used the preprocessed data for model building, which is arguably the most important step in the condition monitoring workflow. For this research project, we defined a model as an input-output mapping between the preprocessed time series containing the transmission error of the meshing gears and the corresponding wear. Generally, there are two different types of models to accomplish this:

- Physics-based models
- Data-driven models

The key difference between the two is that physics-based models derive the target variable from a predefined set of rules, while data-driven approaches derive the rules themselves from the presented data. These “learned” rules can then be applied to new, “unseen” data. At this point, it is important to mention that a mixture of both model types is possible. The scope of our research is limited to data-driven models. Within this scope, we identified three different implementations of data-driven models for time series analysis: instance-based approaches that use dynamic time warping, feature-based approaches and feature learning approaches. Feature-based approaches in particular have the benefit of creating interpretable insight into the data and therefore are the preferred approach. The workflow to create feature-based models that we used is depicted in Figure 5b).

#### Feature Extraction and Selection

The proposed method uses massive feature extraction as a concept by Fulcher and Jones (2014) to create a large set of features from measurement signals. This approach is universally applicable for time series and helps to understand the underlying processes that generate the data. Using features has multiple benefits. It is possible to calculate features from time series with different lengths and, after feature calculation, the amount of data is greatly reduced. This approach of feature extraction does not require extensive domain knowledge, although if useful features are known, they should certainly be included. Typical feature sources that can be used to extract features from time series are (Fulcher, 2017):

- Statistical features such as mean, median, mode, standard deviation, quantiles, skewness and variance
- Features resulting from a discrete Fourier transform and similar concepts, e.g. discrete wavelet transform
- Measures of stationarity
- Parameters from statistical time series models
- Autocorrelation functions

This listing is not exhaustive. The informed reader may have noticed that features that have proven to be useful for gearbox damage evaluation have not been mentioned in particular. Many of these domain-specific statistical features, such as FM0, FM4, and M6A use a signal with removed gear mesh and shaft frequencies as well as their respective harmonics from the vibration signal. Removing the gear mesh frequency harmonics would especially eliminate a significant amount of information from the transmission error signal. For more information on statistical features for gear vibration signals, refer to the work of Samuel and Pines (2005), and more recently the work of Zhu, Nostrand, Spiegel, and Morton (2014), and Sharma and Parey (2016). The result of feature extraction is a design matrix where each row represents a sub-series and each column contains the values for a specific feature. This process yields a design matrix with 5,650 columns and therefore 5,650 features. The features are based on the efficient features of the tsfresh python package v0.15.1 (Christ, Braun, Neuffer, & Kempa-Liehr, 2018).

Since this type of feature extraction requires little domain knowledge, the design matrix likely contains features that do not contain meaningful information that contributes to the prediction of gear wear condition. In fact, these useless features can hamper predictive performance and may increase model-training time. Consequently, feature selection is a crucial step in the proposed workflow. Feature selection methods attempt to retain features with the highest predictive performance while removing all others. Generally, three different approaches to feature selection exist (Guyon & Elisseeff, 2000): wrapper methods, embedded methods and filter methods. With numerous features, wrapper methods quickly become computationally infeasible, which is why they are not further explored in the approach we propose. Both filter and embedded feature selection techniques are used in this proposal.

Only data from experiment C is part of feature selection. At first, we used the so called FRESH algorithm developed by Christ et al. (2018) to filter features from the design matrix that show no statistical dependence on the target variable. This was done by means of hypothesis testing. At first, each feature was checked individually for statistical relevance. For continuous features with continuous target variables, the authors of the FRESH algorithm suggest using the Kendall rank test. Rejecting the hypothesis about the data makes it relevant for target prediction. The rejection threshold  $THOLD_k$  of a p-value is determined by the Benjamini-Yekutieli procedure with Eq. (4) (Benjamini & Yekutieli, 2001). In this equation,  $k$  is the rank of a p-value after sorting all p-values in ascending order and  $m$  is the overall number of hypothesis tests conducted to yield the p-values. Once a p-value at rank  $k$  is lower than the respective threshold  $THOLD_k$ , all p-values with a higher rank indicate rejected hypothesis tests. The Benjamini-Yekutieli procedure attempts to control the so-called false discovery

rate  $FDR$  seen in hypothesis testing. The false discover rate coincides with type I errors in hypothesis testing: “discovery” of dependencies that do not actually exist. We used a false discovery rate of 0.05 for the procedure. Completing this procedure reduces the amount of feature columns in the design matrix from 5,650 to 3,038.

$$THOLD_k = \frac{k}{m} \cdot \frac{FDR}{\sum_{i=1}^m \frac{1}{i}} \quad (4)$$

Similar information may be contained among the remaining features, i.e. mean and median are typical examples for features that in practice often have the same value (Christ, Kempa-Liehr, & Feindt, 2017). Removing highly correlated features therefore serves the purpose of removing redundant information. Typically, this problem is addressed with a principal component analysis; however, in order to keep intuitive interpretability high, we propose usage of a simple correlation analysis as suggested by Albon (2018). This method uses the Pearson correlation coefficient to quantify the linear correlation of a pair of features. Highly correlated features are defined as having an absolute correlation coefficient higher than 0.95. If a pair of features is highly correlated, one of the two is removed. We made this decision despite the known downfalls of the Pearson correlation coefficient and descriptive statistics. In our dataset, we found and removed 92 correlated features in this step, reducing the feature count to 2,946.

Finally, feature selection concludes with recursive feature elimination, which Guyon, Weston, Barnhill, and Vapnik (2002) first developed and applied in combination with support vector machines. Recursive feature elimination is applicable with every machine learning algorithm that creates an internal feature ranking.

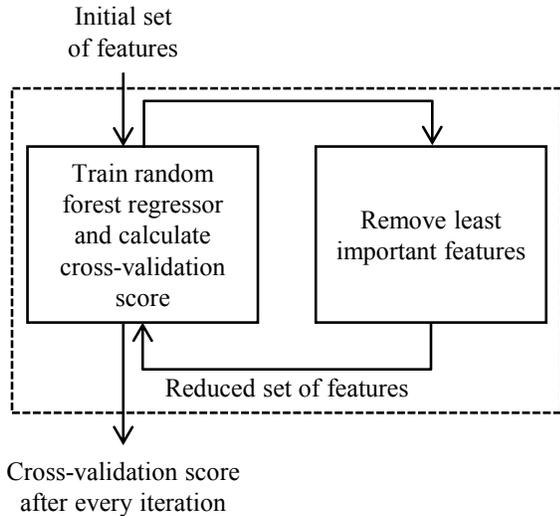


Figure 9. Depiction of recursive feature elimination paired with cross-validation.

The algorithm we used for this proposal is random forest regression; we elucidate more details and benefits of tree-based methods in the paragraph “Model Selection, Training and Evaluation”. The process of recursive feature elimination with cross-validation is depicted in Figure 9. Since the optimal number of features is not known, a cross-validation score is calculated at every step of recursive feature elimination. This helps to evaluate at which point there are enough features to describe the system.

First, we passed an initial dataset with all features to the random forest regressor and a cross-validation score was calculated. Subsequently, the random forest regression model was retrained after removing the features with the lowest feature importance from the initial set of features. This process was repeated until all features were removed. The cross-validation scores were used to monitor model performance during the recursive feature elimination process. Every iteration 0.001 % of the initial number of features was removed. We obtained the final number of features by selecting the smallest number of features before the model performance decreased significantly. By observing the plot in Figure 10, we selected a total of 15 features (see Table 4). This corresponds to the final number of features.

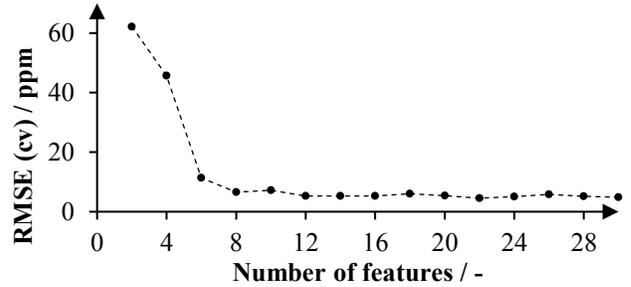


Figure 10. Cross-validated RMSE during recursive feature elimination.

No.	Feature	No.	Feature
1	Skewness	9	Amplitude at 18th order $f_{gm}$
2	Median	10	Quantile 0.3
3	Amplitude at 1st order $f_{gm}$	11	Quantile 0.8
4	Ratio beyond $0.5\sigma$	12	Quantile 0.6
5	Amplitude at 4th order $f_{gm}$	13	Variance
6	Amplitude at 15th order $f_{gm}$	14	Amplitude at 8th order $f_{gm}$
7	Amplitude at 3th order $f_{gm}$	15	Ratio beyond $1\sigma$
8	Amplitude at 25th order $f_{gm}$		

Table 4. Selected set of the 15 highest rated of 5,650 features ordered by descending importance.

### Model Selection, Training and Evaluation

Finally, we used the selected features to train a machine learning model. For the sake of brevity, we are covering model selection, training and evaluation in a single section. Selecting a machine learning algorithm is no simple task. Many different approaches exist ranging from support vector machines, neural networks, linear regression to decision trees and more advanced deep learning techniques. Machine learning algorithms based on decision trees are popular because of multiple reasons (Louppe, 2014):

- Decision trees are interpretable (white box models)
- They work with numerical and categorical variables as input
- Decision trees select features as a part of the tree growing process
- They are robust to errors in the training data
- They are non-parametric

Obviously there are not only upsides to decision trees; their main downside is their lack of accuracy compared to other machine learning methods, as decision trees are generally known as high variance estimators (Murphy, 2012). Decision trees are the basis for more advanced methods such as random forests (Breiman, 2001) and gradient boosting methods (Friedman, 2001). Both methods are classified as ensemble learning and perform well in comparative studies, e.g. those carried out by Caruana and Niculescu-Mizil (2006), Fernández-Delgado, Cernadas, Barro, and Amorim (2014), and more recently Olson, La Cava, Mustahsan, Varik, and Moore (2018). In fact, random forests and gradient boosted trees are among the most popular algorithms used by data scientists and machine learning practitioners (Kaggle, 2019). Because of their widespread use and their generally good performance, we used both, random forest regression (RFR) and gradient boosted regression trees (GBRT) as models in our proposed workflow and compared the results.

Parameter selection and model evaluation were carried out using the data from experiment C, while the test data came from experiments A and B. For this analysis, we treated experiment C as “run-to-failure” experiment that contains all gear wear states we are trying to detect. A 5-fold cross-validation supports the parameter selection, which we applied after shuffling the samples. The metric we used to score each fold is the RMSE. The models we applied use the implementation by the scikit-learn library (Pedregosa et al., 2011). Parametrization is kept simple, since model performance tends to increase with model size and therefore with tree size. Instead of extensive algorithmic hyperparameter tuning, we therefore selected the parameter values by hand. The most important parameter values are

discussed in the following; the remaining parameters are unchanged from the default values provided by scikit-learn.

Both RFR and GBRT models use an ensemble of 500 decision trees, each with an unlimited depth. In order to control overfitting and support generalization, the minimum number of samples required for a node split in a decision tree was set to 5 for both models. The maximum number of features for the growth of each decision tree was capped at 80 % of all available features. Similarly, each decision tree was grown with regard to only 80 % of all available training samples. The gradient boosting method also requires a learning rate that defines how much information each additional tree contributes to the entire model. A typical value for the learning rate is 0.1, which coincides with the value we chose.

Fold no.	RFR	GBRT
RMSE 1st Fold	3.35 ppm	2.75 ppm
RMSE 2nd Fold	4.18 ppm	4.61 ppm
RMSE 3rd Fold	2.21 ppm	1.93 ppm
RMSE 4th Fold	3.20 ppm	3.52 ppm
RMSE 5th Fold	2.08 ppm	1.87 ppm
RMSE average	3.00 ppm	2.94 ppm

Table 5. RMSE values of both decision tree ensembles as a result of cross-validation.

The results of cross-validation of the final model are portrayed in Table 5. A regression plot for both models displays both the ground truth and the predicted values against each other (see Figure 12).

## 4. RESULTS

The results of the proposed workflow follow the model deployment step in Figure 5a). The feature selection process resulted in 15 relevant features. In Figure 11, the six highest rated features according to feature importance are plotted against wear. We improved comparability by min-max normalization of the feature values. The features depicted make up over 90 % of the feature importance determined by the recursive feature elimination algorithm. Three of the top six features are statistical features. While “skewness” and “median” are self-explanatory, the “ratio beyond  $0.5\sigma$ ” quantifies the ratio of signal values that are further than half a standard deviation away from the signal mean. The remaining three features are categorized as amplitudes at specific frequencies in the frequency spectrum. All high-ranking frequencies selected by the feature selection algorithm are harmonics of the gear mesh frequency, namely the first, fourth and fifteenth order of the gear mesh frequency.

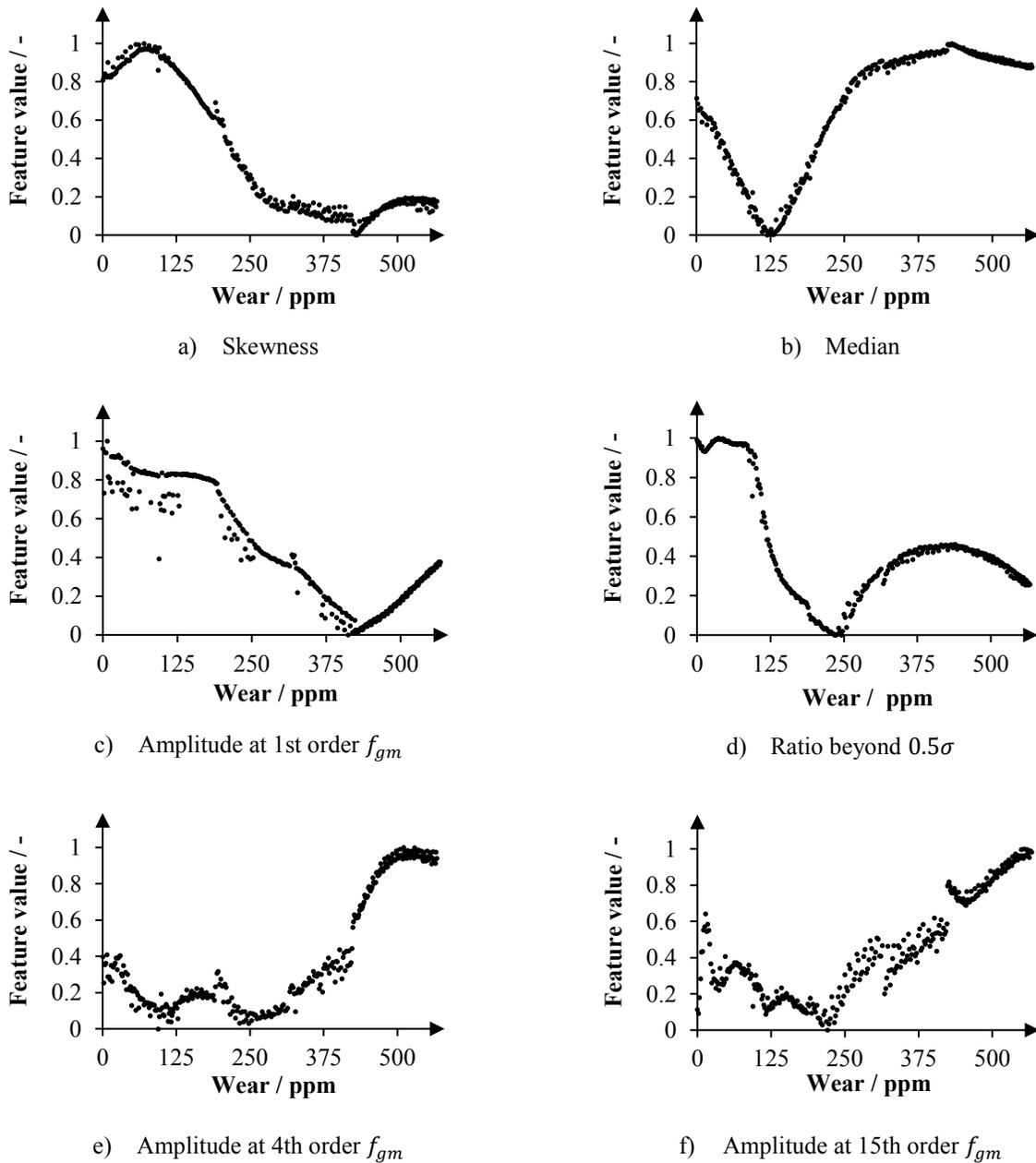
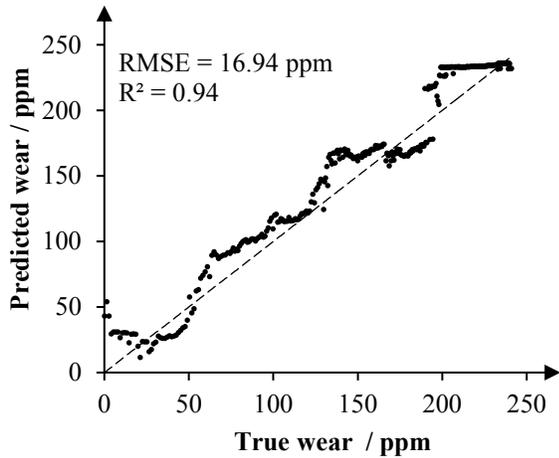


Figure 11. The six highest rated features rated by recursive feature elimination. The features are min-max normalized to improve comparability.

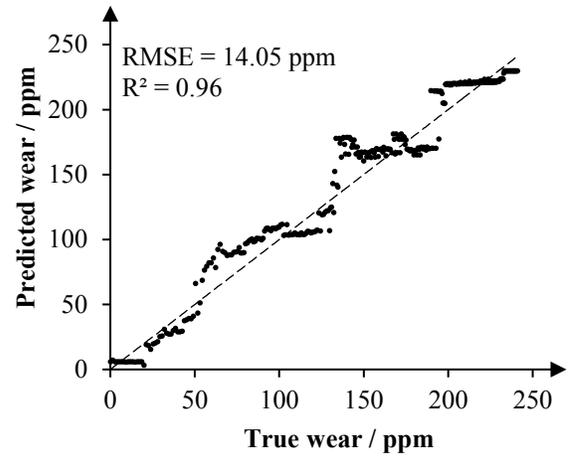
As previously mentioned, we trained and tuned both the RFR and the GBRT only with data from experiment C, which we considered a “run-to-failure” experiment. We used the data from experiment A and experiment B to simulate deployment of the models. The features extracted from experiment A and B correspond to those selected considering only experiment C. Figure 12 depicts the resulting regression plots from the simulated deployment. The vertical axis of each plot displays the slow-speed wear predicted by the model, while the horizontal axis shows the

ground truth. An ideal model would produce the dashed line graph in every plot.

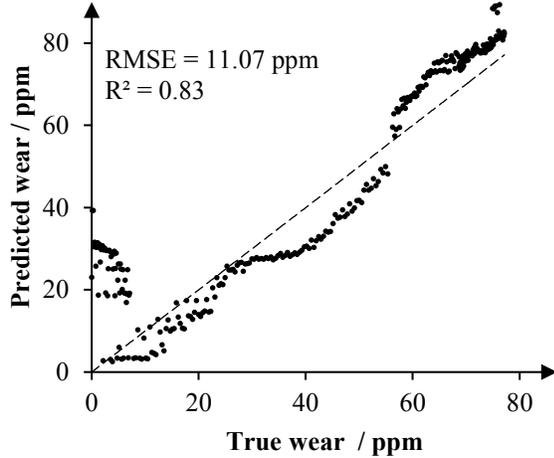
Overall, both models performed well. Goodness of fit, which we determined with the  $R^2$  coefficient of determination and the RMSE for both models, are at acceptable levels. These performance metrics also reveal that the GBRT model slightly outperforms the RFR model. The plots additionally show a resemblance to an approximately piecewise constant function, which is typical for methods based on decision trees.



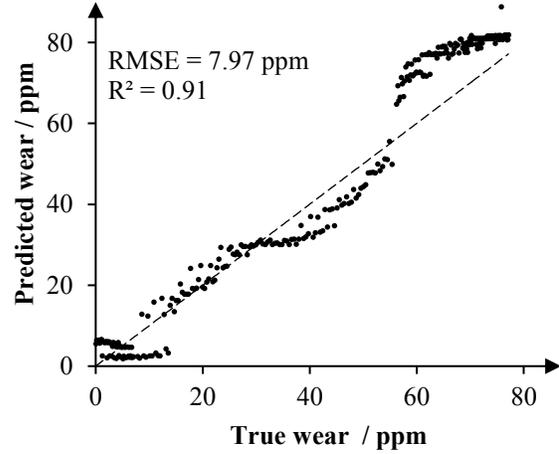
a) RFR – Experiment A



b) GBRT – Experiment A



c) RFR – Experiment B



d) GBRT – Experiment B

Figure 12: Regression results from the test data (experiment A and experiment B).

## 5. DISCUSSION

With this paper, we provide an implementation of our proposed approach to condition monitoring and predictive maintenance tasks with a continuous target variable. Without going into further detail, we would like to note that adaptation of this approach to classification problems with categorical target variables is straightforward. The present approach is mainly data driven, as domain knowledge is solely required in the data acquisition and preprocessing steps. Because gear transmission systems in slow operation do not provide sufficient vibration excitation, we chose to use the dynamic transmission error as the input signal to the condition monitoring workflow in the presented use-case. To demonstrate the viability of this approach, we used a dataset containing transmission error data from spur gears and the respective combined slow-speed wear of both pinion

and wheel. First, the data was high-pass filtered and therefore centered on zero. Subsequently, each signal was split into individual hunting tooth cycles. We chose this particular way of splitting, because the hunting tooth cycle contains all meshing combinations and thereby is a universal approach. However, other signal sub-series lengths, e.g. according to one gear shaft frequency or even according to the gear mesh frequency would surely influence the overall model performance. Subsequently, vast amounts of features were extracted from each individual sub-series and through algorithmic feature selection, only the most important features were retained. Compared to manual feature selection, this approach reduces the human effort and involvement in feature engineering significantly. Compared to feature/representation-learning approaches that are typically implemented with deep neural networks, our proposed approach returns features that may easily be

understood and interpreted by humans. However, models with feature engineering, both in an automated massive or manual way, are only as good as the available feature set. Although we extracted a large set of features, it is not a given that no other supplemental and perhaps more suitable features exist. In particular, expanding the feature set with features specific to gearboxes could help to improve wear estimation capabilities further. If this approach is used to extract and select features from gear vibration data, the addition of gear vibration specific features is crucial. The parameters of the implemented algorithmic feature selection are chosen based on best practices combined with the usage of reasonable computational resources as ulterior motive. Ideally, the extracted set of features and the progression of feature values allows conclusions about the underlying physical principles generating the data. We would like to encourage experimentation and usage of the proposed workflow for condition monitoring tasks with other machinery and sensor arrangements. Finally, we emphasize that the importance of data acquisition, preprocessing, feature extraction and feature selection is pronounced compared to the other steps in the condition monitoring workflow, especially model building.

Obviously, our approach and implementation are not holistic solutions to all condition monitoring tasks. For example, the models we used are based on decision trees, which are known to not extrapolate well. For our dataset and use-case however, this is not a concern, since extrapolation is not necessary with the existence of “run-to-failure” data. Examples of model types that are better suited for extrapolation are linear models and neural networks. Most of the parameters for the ensemble models GBRT and RFR have values that are similar to the defaults provided by scikit-learn. Because slight modifications to the baseline parameters already yield good results, we did not attempt an additional, extensive algorithmic hyperparameter optimization. All downsides aside, methods based on decision trees are white box models. Decision boundaries and the decision-making process of decision trees are therefore traceable without extensive effort. This attribute is likely relevant for applications that require transparency, such as condition monitoring of components that are crucial for safety in a system.

Overall, the proposed approach shows good slow-speed wear estimation results throughout the whole experiment range. Even at the beginning of lifetime for small quantities of wear, which intuitively are harder to estimate, the results are acceptable. This does especially apply when using a GBRT model. The basis for calculations are three experiments with one of them used for training. The obtained results are even more promising considering the rather low number of experiments used for training and testing. However, this also raises the question of the overall validity of the results when it comes to general gear shapes and setups. All experiments use the same experiment setup

with one spur gear stage. Because the transmission error and the vibrational behavior is sensitive to gear shapes and gear system design, additional experiments are nearly imperative. The same accounts for variations in operating conditions such as speed or load. Beyond the physical influences, higher rotational speeds also reduce the angle measurement quality due to a limited sample rate and hence reduce the transmission error determination accuracy.

## 6. CONCLUSION

Monitoring of slow-speed gear wear is an important contribution to realizing a predictive maintenance strategy to avoid critical failures of gear systems.

With this paper, we show that an approach based on the dynamic transmission error that contains an automated feature selection process is a suitable way for slow-speed wear estimation. The approach includes extracting a large set of features from the preprocessed transmission error in order to find and select the most relevant features. Therefore, the feature selection and elimination process contain a combined filter and embedded feature selection method including recursive feature elimination. By doing so, only a small amount of domain knowledge is necessary. Nevertheless, due to the explicit nature of calculation, the resulting features still allow for interpretability if required. Comparing both RFR and GBRT models used for subsequent training, GBRT slightly outperforms RFR, but at a high overall performance level.

Considering everything, the presented approach offers good estimations for slow-speed wear in the course of condition monitoring. It even seems to be promising for higher speeds in combination with vibrational analysis, which, however, needs to be addressed in further research.

## NOMENCLATURE

<i>FDR</i>	False discovery rate
<i>FZG</i>	Gear Research Centre, TUM
<i>GBRT</i>	Gradient boosted regression trees
<i>IQR</i>	Interquartile range
<i>LSTM</i>	Long-short term memory
$Q_1$	First quartile
$Q_3$	Third quartile
$R^2$	Coefficient of determination
<i>RFR</i>	Random forest regression
<i>RMSE</i>	Root mean square error
$THOLD_k$	Threshold at rank k
<i>cv</i>	Cross-validated
$f_{1,2}$	Shaft frequency (pinion, wheel), in Hz
$f_{gm}$	Gear mesh frequency, in Hz
$f_{ht}$	Hunting tooth frequency, in Hz
<i>k</i>	Rank of p-value
<i>m</i>	Number of hypothesis tests
$\tilde{x}$	Overall signal median

$x_t$	Signal value at timestep $t$
$z_{1,2}$	Number of teeth (pinion, wheel)
$z_{gcd}$	Greatest common divisor of pinion and wheel number of teeth
$z_t^r$	Robust z-score at timestep $t$
$\Delta\varphi(t)$	Transmission error, in $^\circ$
$\Delta\varphi_{1,2}(t)$	Rotation angle (pinion, wheel), in $^\circ$
$\sigma$	Standard deviation

## REFERENCES

- Aggarwal, C. C. (2017). *Outlier Analysis* (2nd edition). Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-47578-3>
- Albon, C. (2018). *Machine learning with Python cookbook: Practical solutions from preprocessing to deep learning* (1st edition). Sebastopol, CA: O'Reilly Media.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29, 1165–1188.
- Brecher, C., Gorgels, C., Hesse, J., & Hellmann, M. (2011). Dynamic transmission error measurements of a drive train. *Production Engineering*, 5(3), 321–327. <https://doi.org/10.1007/s11740-011-0310-5>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd International Conference on Machine Learning*, 161–168. <https://doi.org/10.1145/1143844.1143865>
- Cerrada, M., Zurita, G., Cabrera, D., Sánchez, R.-V., Artés, M., & Li, C. (2016). Fault diagnosis in spur gears based on genetic algorithm and random forest. *Mechanical Systems and Signal Processing*, 70-71, 87–103. <https://doi.org/10.1016/j.ymssp.2015.08.030>
- Chin, Z. Y., Smith, W. A., Borghesani, P., Randall, R. B., & Peng, Z. (2021). Absolute transmission error: A simple new tool for assessing gear wear. *Mechanical Systems and Signal Processing*, 146, 107070. <https://doi.org/10.1016/j.ymssp.2020.107070>
- Choy, F. K., Polyshchuk, V., Zakrajsek, J. J., Handschuh, R. F., & Townsend, D. P. (1996). Analysis of the effects of surface pitting and wear on the vibration of a gear transmission system. *Tribology International*, 29(1), 77–83. [https://doi.org/10.1016/0301-679X\(95\)00037-5](https://doi.org/10.1016/0301-679X(95)00037-5)
- Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing*, 307, 72–77. <https://doi.org/10.1016/j.neucom.2018.03.067>
- Christ, M., Kempa-Liehr, A. W., & Feindt, M. (2017). Distributed and parallel time series feature extraction for industrial big data applications. *ArXiv E-Prints*.
- DIN ISO 14635-1 (2006). *DIN ISO 14635-1:2006-05: Zahnräder - FZG-Prüfverfahren - Teil 1: FZG-Prüfverfahren A/8,3 /90 zur Bestimmung der relativen Fresstragfähigkeit von Schmierölen (ISO 14635-1:2000), Gears - FZG test procedures - Part 1: FZG test method A/8,3/90 for relative scuffing load-carrying capacity of oils*.
- Ding, H., & Kahraman, A. (2007). Interactions between nonlinear spur gear dynamics and surface wear. *Journal of Sound and Vibration*, 307(3-5), 662–679. <https://doi.org/10.1016/j.jsv.2007.06.030>
- Feng, Z., & Zuo, M. J. (2012). Vibration signal models for fault diagnosis of planetary gearboxes. *Journal of Sound and Vibration*, 331(22), 4919–4939. <https://doi.org/10.1016/j.jsv.2012.05.039>
- Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do We Need Hundreds of Classifiers to Solve Real World Classification Problems? *J. Mach. Learn. Res.*, 15(1), 3133–3181.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Fromberger, M., Weinberger, U., Kohn, B., Utakapan, T., Otto, M., & Stahl, K. (2016). Condition Monitoring by Position Encoders. In W. Kropp (Ed.), *Proceedings of the Inter-Noise 2016: 45th International Congress and Exposition on Noise Control Engineering : Towards a quieter future : August 21-24, 2016, Hamburg*. Berlin: Deutsche Gesellschaft für Akustik e.V.
- Fulcher, B. D. (2017). Feature-based time-series analysis. *ArXiv E-Prints*.
- Fulcher, B. D., & Jones, N. S. (2014). Highly Comparative Feature-Based Time-Series Classification. *IEEE Transactions on Knowledge and Data Engineering*, 26(12), 3026–3037. <https://doi.org/10.1109/TKDE.2014.2316504>
- Guyon, I., & Elisseeff, A. (2000). 10.1162/153244303322753616. *CrossRef Listing of Deleted DOIs*, 1, 1157–1182. <https://doi.org/10.1162/153244303322753616>
- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, 46(1/3), 389–422. <https://doi.org/10.1023/A:1012487302797>
- Han, T., Jiang, D., Zhao, Q., Wang, L., & Yin, K. (2018). Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery. *Transactions of the Institute of*

- Measurement and Control*, 40(8), 2681–2693.  
<https://doi.org/10.1177/0142331217708242>
- Hu, C., Smith, W. A., Randall, R. B., & Peng, Z. (2016). Development of a gear vibration indicator and its application in gear wear monitoring. *Mechanical Systems and Signal Processing*, 76-77, 319–336.  
<https://doi.org/10.1016/j.ymssp.2016.01.018>
- Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufer, M., Verstockt, S., . . . van Hoescke, S. (2016). Convolutional Neural Network Based Fault Detection for Rotating Machinery. *Journal of Sound and Vibration*, 377, 331–345.  
<https://doi.org/10.1016/j.jsv.2016.05.027>
- Kaggle (2019). Kaggle's State of Data Science and Machine Learning 2019: Enterprise Executive Summary.
- Kuang, J. H., & Lin, A. D. (2001). The Effect of Tooth Wear on the Vibration Spectrum of a Spur Gear Pair. *Journal of Vibration and Acoustics*, 123(3), 311–317.  
<https://doi.org/10.1115/1.1379371>
- Lee, K., Kim, J.-K. K., Kim, J., Hur, K., & Kim, H. (2018). Stacked Convolutional Bidirectional LSTM Recurrent Neural Network for Bearing Anomaly Detection in Rotating Machinery Diagnostics. *1st International Conference on Knowledge Innovation and Invention*, 98–101.
- Lei, Y., Jia, F., Lin, J., Xing, S., & Ding, S. X. (2016). An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data. *IEEE Transactions on Industrial Electronics*, 63(5), 3137–3147. <https://doi.org/10.1109/TIE.2016.2519325>
- Louppe, G. (2014). *Understanding Random Forests: From Theory to Practice. Dissertation*: Université de Liège.
- Mark, W. D. (2015). Time-synchronous-averaging of gear-meshing-vibration transducer responses for elimination of harmonic contributions from the mating gear and the gear pair. *Mechanical Systems and Signal Processing*, 62-63, 21–29.  
<https://doi.org/10.1016/j.ymssp.2015.03.006>
- McFadden, P. D. (1986). Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Vibration. *Journal of Vibration and Acoustics*, 108(2), 165–170. <https://doi.org/10.1115/1.3269317>
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press. Retrieved from <https://ebookcentral-proquest-com.eaccess.ub.tum.de/lib/munchentech/detail.action?docID=3339490>
- Niemann, G., & Winter, H. (2003). *Getriebe allgemein, Zahnradgetriebe - Grundlagen, Stirnradgetriebe: Band 2: Getriebe allgemein, Zahnradgetriebe - Grundlagen, Stirnradgetriebe* (2., völlig neubearb. Aufl., 2. berichtigter Nachdr., korrigierter Nachdr. *Maschinenelemente: / G. Niemann; H. Winter ; Bd. 2.* Berlin: Springer. <https://doi.org/10.1007/978-3-662-11873-3>
- Olson, R. S., La Cava, W., Mustahsan, Z., Varik, A., & Moore, J. H. (2018). Data-driven advice for applying machine learning to bioinformatics problems. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 23, 192–203.
- Pacheco, F., Valente de Oliveira, J., Sánchez, R.-V., Cerrada, M., Cabrera, D., Li, C., . . . Artés, M. (2016). A statistical comparison of neuroclassifiers and feature selection methods for gearbox fault diagnosis under realistic conditions. *Neurocomputing*, 194, 192–206.  
<https://doi.org/10.1016/j.neucom.2016.02.028>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2020, October 14). sklearn.preprocessing.robust\_scale — scikit-learn 0.23.2 documentation. Retrieved from [https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.robust\\_scale.html](https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.robust_scale.html)
- Randall, R. B. (2011). *Vibration-based condition monitoring: Industrial, aerospace, and automotive applications*. Chichester, UK: John Wiley & Sons. Retrieved from <http://lib.myliblibrary.com?id=310092>  
<https://doi.org/10.1002/9780470977668>
- Samanta, B. (2004). Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mechanical Systems and Signal Processing*, 18(3), 625–644.  
[https://doi.org/10.1016/S0888-3270\(03\)00020-7](https://doi.org/10.1016/S0888-3270(03)00020-7)
- Samuel, P. D., & Pines, D. J. (2005). A review of vibration-based techniques for helicopter transmission diagnostics. *Journal of Sound and Vibration*, 282(1-2), 475–508.  
<https://doi.org/10.1016/j.jsv.2004.02.058>
- Scheffer, C., & Girdhar, P. (2004). *Practical machinery vibration analysis and predictive maintenance*. Oxford, UK: Newnes. Retrieved from <http://ebookcentral.proquest.com/lib/munchentech/detail.action?docID=293533>
- Schultheiss, H., Tobie, T., Michaelis, K., Höhn, B.-R., & Stahl, K. (2014). The Slow-Speed Wear Behavior of Case-Carburized Gears Lubricated with NLGI 00 Grease under Boundary Lubrication Conditions. *Tribology Transactions*, 57(3), 524–532.  
<https://doi.org/10.1080/10402004.2014.883005>
- Sharma, V., & Parey, A. (2016). A Review of Gear Fault Diagnosis Using Various Condition Indicators. *Procedia Engineering*, 144, 253–263.  
<https://doi.org/10.1016/j.proeng.2016.05.131>

- Shen, C., Wang, D., Kong, F., & Tse, P. W. (2013). Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier. *Measurement*, *46*(4), 1551–1564.  
<https://doi.org/10.1016/j.measurement.2012.12.011>
- Siewerin, B. J., Dobler, A., Tobie, T., & Stahl, K. (2020). Applicability of an Oil Based Calculation Approach for Wear Risk and Wear Lifetime to Grease Lubricated Gear Pairings. In ASME (Ed.), *Print proceedings of the international design engineering technical conferences & computers and information in engineering conference*. AMER SOC OF MECH ENGINEER.  
<https://doi.org/10.1115/DETC2019-97439>
- Soualhi, A., Hawwari, Y., Medjaher, K., Guy, C., Razik, H., & Guillet, F. (2018). PHM survey: Implementation of signal processing methods for monitoring bearings and gearboxes. *International Journal of Prognostics and Health Management*, *9*.
- Wu, S., Zuo, M. J., & Parey, A. (2008). Simulation of spur gear dynamics and estimation of fault growth. *Journal of Sound and Vibration*, *317*(3-5), 608–624.  
<https://doi.org/10.1016/j.jsv.2008.03.038>
- Yan, R., Gao, R. X., & Chen, X. (2014). Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Processing*, *96*, 1–15.  
<https://doi.org/10.1016/j.sigpro.2013.04.015>
- Yang, R., Huang, M., Lu, Q., & Zhong, M. (2018). Rotating Machinery Fault Diagnosis Using Long-short-term Memory Recurrent Neural Network. *IFAC-PapersOnLine*, *51*(24), 228–232.  
<https://doi.org/10.1016/j.ifacol.2018.09.582>
- Zhao, R., Yan, R., Wang, J., & Mao, K. (2017). Learning to Monitor Machine Health with Convolutional Bi-Directional LSTM Networks. *Sensors*, *17*(2).  
<https://doi.org/10.3390/s17020273>
- Zhu, J., Nostrand, T., Spiegel, C., & Morton, B. (2014). Survey of condition indicators for condition monitoring systems. *PHM 2014 - Proceedings of the Annual Conference of the Prognostics and Health Management Society 2014*, 635–647.