CIRP Journal of Manufacturing Science and Technology 35 (2021) 795–802

Contents lists available at ScienceDirect

CIRP Journal of Manufacturing Science and Technology

journal homepage: www.elsevier.com/locate/cirpj

Data-based ensemble approach for semi-supervised anomaly detection in machine tool condition monitoring

B. Denkena[a], M.-A. Dittrich[a], H. Noske[a,*], D. Stoppel[a], D. Lange[b]

[a] Institute of Production Engineering and Machine Tools, An der Universität 2, 30823 Garbsen, Germany
[b] Marposs Monitoring Solutions GmbH, Buchenring 40, 21272 Egestorf, Germany

ARTICLE INFO

Article history:
Available online 5 October 2021

Keywords:
Condition monitoring
Machine learning
Failure
Ball screw
Maintenance

ABSTRACT

Data-based methods are capable to monitor machine components. Approaches for semi-supervised anomaly detection are trained using sensor data that describe the normal state of machine components. Thus, such approaches are interesting for industrial practice, since sensor data do not have to be labeled in a time-consuming and costly way. In this work, an ensemble approach for semi-supervised anomaly detection is used to detect anomalies. It is shown that the ensemble approach is suitable for condition monitoring of ball screws. For the evaluation of the approach, a data set of a regular test cycle of a ball screw from automotive industry is used.

Introduction

The trend of decreasing lot sizes leads to frequently changing production programs and, thus, to the use of highly automated, flexible usable production equipment. Due to high cost rates of these machines, it is desirable to maximize availability. To ensure high availability, condition monitoring systems are used. These systems enable the detection of machine condition changes and to plan maintenance and servicing as well as to predict machine damage. Because of a wide variety of damage types and operating conditions of machine tools, the exact service life of ball screws cannot be predicted a priori [1]. In addition, the ball screw service life is influenced by the selected type of lubricant [2]. As wear progresses, the precision of ball screws decreases. As a result, tolerances and surface properties of components required by the customer may no longer be met [3]. However, a ball screw can be repaired up to 3 to 4 times, saving 30% to 50% of the cost of installing a new ball screw if its wear level is below 80% [4]. For this reason, reliable monitoring techniques are needed to avoid additional costs for industrial companies [3].

In the past, numerous physical and data-based methods have been developed for condition monitoring [5–10]. A disadvantage of physical models is that they are often not transferable to other components without model adaptions. Data-based approaches, on the other hand, do not require any system knowledge. One challenge in condition monitoring is the availability of information about anomalies. Therefore, it is difficult to derive a performance index that assesses the health state of components based on the monitored variables [11].

In case information about anomalies is not available, methods for semi-supervised anomaly detection can be used [12]. A variety of semi-supervised approaches to component health monitoring have been described in the literature. Qiu et al. present a method for monitoring of bearings using vibration signals. A wavelet filter method is used to suppress noise from the signals. Self-organizing maps are applied to calculate the so-called minimum quantization error (MQE), which is used to detect condition changes of bearings. However, it remains unclear how decision boundaries are calculated [13]. Liao and Lee present an approach for bearing monitoring of a chiller system. In this approach, a test cycle with different workloads is set up and the data of the transition periods is used. To evaluate the condition of the bearing, the data from a test cycle under normal condition is compared with other test cycles that are performed periodically. Using wavelet packet decomposition, features of the vibration sensors are extracted and a principal component analysis is used for feature extraction. Gaussian Mixture Models are used to determine a performance index, which is used to detect incipient anomalies of a bearing [11]. Ruiz-Carcel and Starr develop a method for condition monitoring of ball screws considering position and current data. Using a test rig, various defects are artificially generated (lack of lubrication, spalling, backlash). Decision boundaries are calculated using the T^2 and Q-statistic. The authors show that the Q-statistic achieves better results in terms of monitoring quality in

* Corresponding author.
E-mail address: noske@ifw.uni-hannover.de (H. Noske).

https://doi.org/10.1016/j.cirpj.2021.09.003
1755-5817/

contrast to the T^2-statistic [14]. Zhao et al. use the Mahalanobis-distance for ball screw condition monitoring considering current and velocity signals. Laplacian eigenmaps are used for dimension reduction. The authors show that the derived health index is viable to detect the wear of the ball screw due to insufficient lubrication [15]. Other authors use deep learning approaches for semi-supervised anomaly detection. Zhai et al. apply a variational autoencoder to derive an operation-specific health indicator to quantify machine health conditions [16]. The advantage of deep learning approaches is that features do not have to be extracted manually. However, long training times and the need for large scale data sets might be an obstacle for the successful implementation [17]. In addition, many hyperparameters need to be set in advance which is a challenge using small data sets.

Some of the proposed methods use a risk factor in order to compute decision boundaries. It is observed that often the risk factor is not systematically varied to study the effect on monitoring performance. In addition, the proposed ball screw monitoring methods are statistical and assume a known distribution of signal features. If the assumed distribution of the features deviates strongly from the actual distribution, the monitoring quality is reduced. Although research in the field of condition monitoring has been going on for a long time, it is slowly finding its way into industrial practice. Reasons for this are that no experiments are carried out in science under conditions close to production and production data are lacking to investigate such monitoring systems for industrial use.

Therefore, this work investigates the monitoring quality of an ensemble approach for robust anomaly detection. An ensemble is a set of methods whose predictions are combined in order to produce robust predictions [18]. The proposed ensemble approach combines different methods (e.g. local outlier factor LOF, k-nearest-neighbor KNN, angle-based outlier detection ABOD) and makes no assumptions about the underlying distribution of the data. The dataset used contains test cycles of one ball screw used in the automotive industry which is recorded during an 8 months period.

The remainder of this paper is organized as follows. In chapter 2, the dataset used is described and visualized. Chapter 3 presents the signal features which are used for condition monitoring. Additionally, the ensemble approach is proposed. The results and metrics used to assess monitoring quality are presented and discussed in chapter 4.

Dataset and process analysis

The data set is taken from a MAG Specht 600 5-axis machine tool. 139 test cycles (57 fault cycles) of the ball screw drive of the Z-axis are performed over a period of 8 months and various sensor data are recorded at a frequency of 100 Hz. After 105 test cycles, the ball screw is replaced and another 34 test cycles are performed with a new ball screw containing no anomalies. Each test cycle is performed following the completion of a lot. The torque of the ball screw is acquired via the machine control. In addition, signals from a 3-axis accelerometer (VA-3D) provided by the Marposs Monitoring Solutions GmbH (Artis) is recorded, which is attached to the machine bed. The machine is used for high-volume production in the automotive industry. The reason for replacing the ball screw is the high stress on the Z-axis during machining and the associated wear on the ball raceways. For further investigation, only the segmented signals of the time window [SB, SE] are used which contain sensor data while the ball screw is in motion in forward direction with constant feed. Thereby, SB and SE represent the start and the end of the segmentation window, respectively.

The segmented signals of the entire test cycle of the Z-axis torque at different points in time are shown in Fig. 1.

After 48 test cycles, higher frequencies appear in the torque signal indicating the degradation of the ball screw (start degradation). Finally, after 106 test cycles, the component is replaced. In addition to torque, the accelerometer signals in three spatial directions are depicted in Fig. 2. At the beginning of the test cycle, a peak in time domain appears in the acceleration signals after 48 test cycles. In addition, a new signal plateau forms in further test cycles.

Fig. 3 shows the frequency spectrum for the torque M_{BSD} and the accelerometer Acc_2. For this purpose, the signals are transformed using a Fast Fourier Transform (FFT). In the case of the torque, it can be seen that peaks appear at the start of the degradation (at 18 Hz, 30 Hz, 38 Hz). The amplitude of these peaks shifts and increases with the degree of degradation. For the accelerometer, the start of degradation is not directly visible in the frequency domain. However, significant peaks are formed with further test cycles (at 15 Hz).

In the following chapter, the ensemble approach is presented, which is used to detect state changes. The presented monitoring method requires only data describing the normal state of the ball screw.

Ensemble approach

First, various signal features are extracted using the segmented signals. The extracted features are then used in the second step to perform the state assessment using the ensembles.

Feature generation

The sensor data are available as discrete time series $x_c(i) \in \mathbb{R}$ for $c \in \{1, ..., n_p\}$ test cycles and $i \in \{1, ..., I\}$ time steps. Four feature groups are extracted, and the monitoring quality of each group is examined. The same features are extracted for each signal. A complete overview of the extracted features is provided in Table 1.

The first group consists of the general-purpose features in time domain:

$$\text{Mean,} \quad \bar{x} = \frac{1}{I}\sum_{i=1}^{I} x_i \tag{1}$$

$$\text{Standard deviation,} \quad x_{std} = \sqrt{\frac{1}{I}\sum_{i=1}^{I}(x_i - \bar{x})^2} \tag{2}$$

$$\text{Root mean square,} \quad x_{rms} = \sqrt{\frac{1}{I}\sum_{i=1}^{I} x_i{}^2} \tag{3}$$

$$\text{Kurtosis,} \quad x_{kurt} = \frac{1}{I}\sum_{i=1}^{I}\left(\frac{x_i}{x_{std}}\right)^4 \tag{4}$$

$$\text{Skewness,} \quad x_{skew} = \frac{1}{I}\sum_{i=1}^{I}\left(\frac{x_i - \bar{x}}{x_{std}}\right)^3 \tag{5}$$

$$\text{Peak2peak,} \quad x_{p2p} = |x_{i,max} - x_{i,min}| \tag{6}$$

$$\text{Signal} - \text{to} - \text{noise ratio,} \quad x_{snr} = 10\,log\left(\frac{x^2}{x_{std}{}^2}\right) \tag{7}$$

$$\text{Shape factor,} \quad x_{sf} = \frac{x_{rms}}{\bar{x}} \tag{8}$$

$$\text{Crest factor,} \quad x_{cf} = \frac{|x|_{max}}{x_{rms}} \tag{9}$$

$$\text{Integral,} \quad x_{int} = \sum_{i=2}^{I}\frac{x_{i-1} + x_i}{2}\left(\frac{SE - SB}{I}\right) \tag{10}$$

$$\text{Interquartile range,} \quad x_{ir} = x_{0.75} - x_{0.25} \tag{11}$$

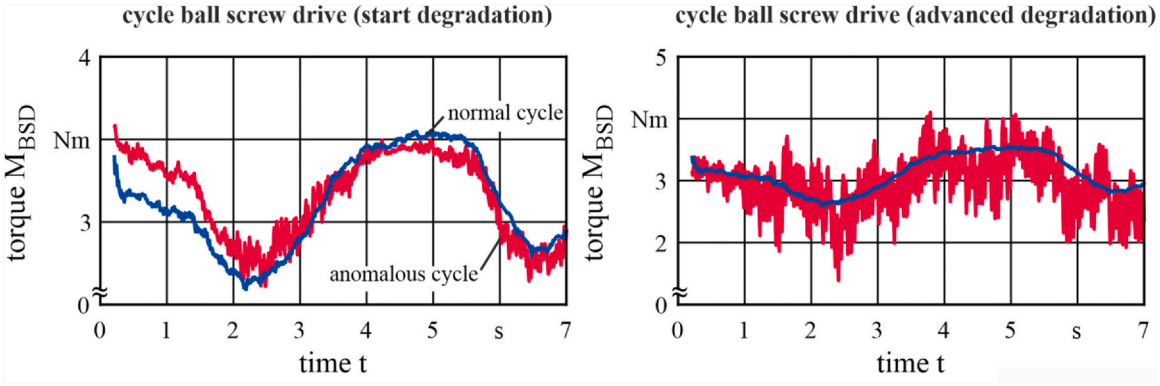

Fig. 1. Visualization of the segmented signal trajectories (torque).

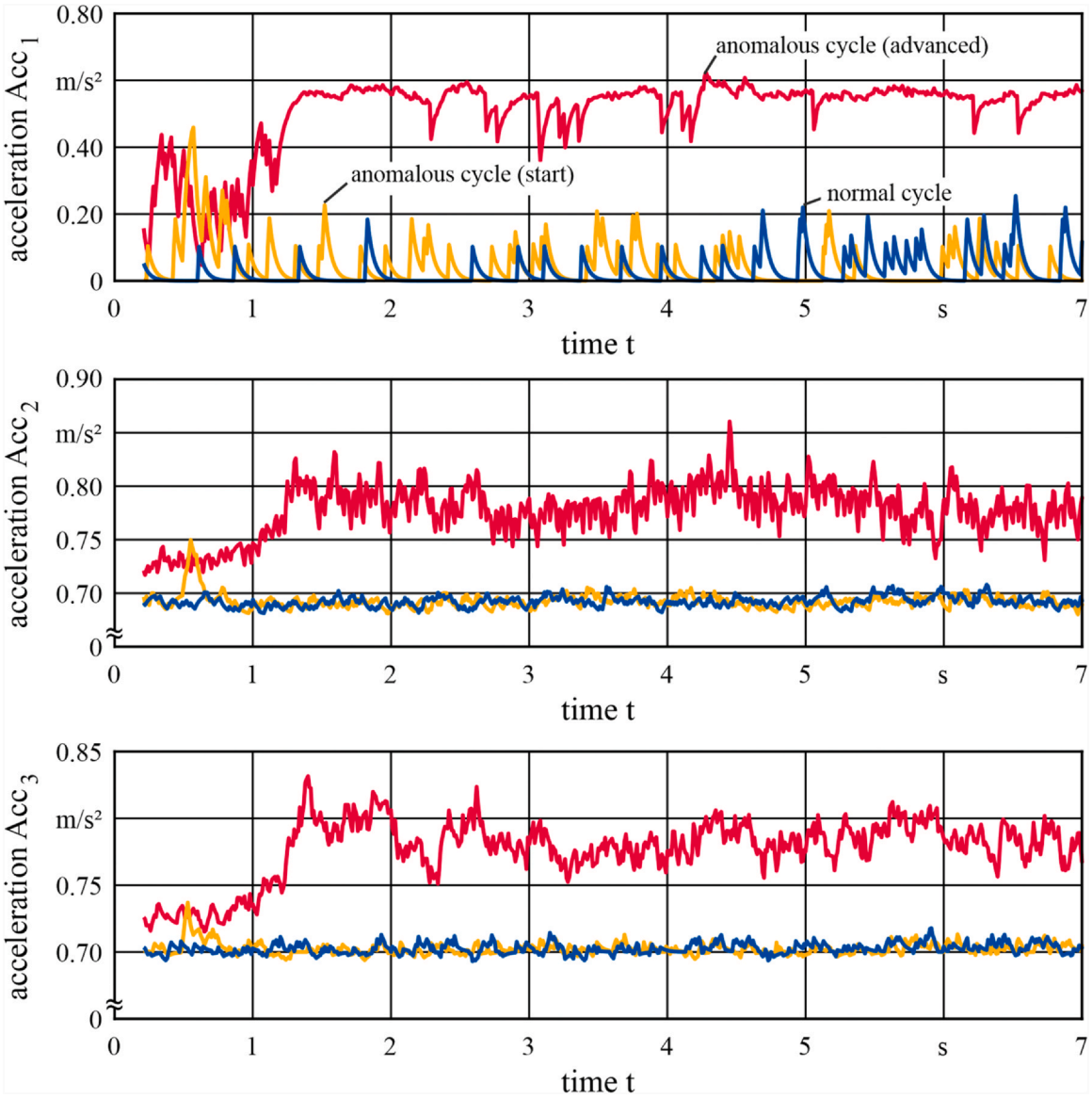

Fig. 2. Visualization of the different segmented signal trajectories (3-axis accelerometer sensor).

Another feature group uses information of the sample autocovariance. The autocovariance indicates how similar a time series x_{i-l} shifted by l discrete time steps is to the original time series x_i. According to Eq. (12), the sample autocovariance is calculated as follows [19]:

$$\hat{\gamma}(l) = \frac{1}{I-1}\sum_{i=1}^{I-l} x_i x_{i-l} \tag{12}$$

The sample autocovariance is calculated for $l \in \{0,...,9\}$. Features are also extracted from the frequency domain by transforming the raw data

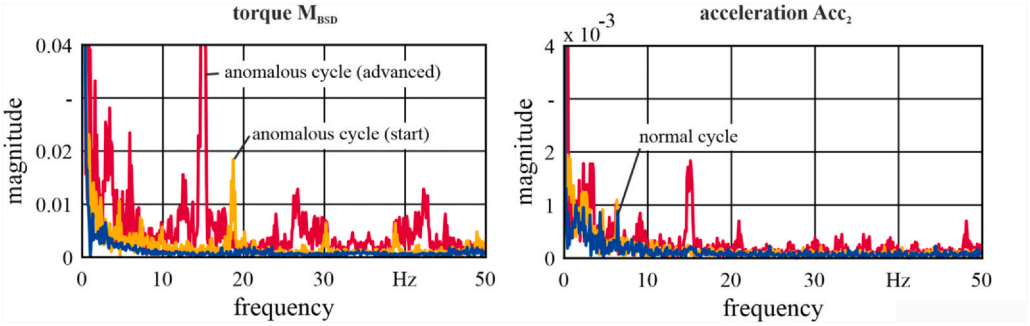

Fig. 3. Visualization of the frequency spectrum for different ball screw conditions.

of all signals using an FFT. The amplitude and frequency of the five most dominant peaks between 10 Hz and 50 Hz are used as another set of features. In addition, the power spectral density (psd) is determined and the features described in the formulas (1)–(12) are extracted to form another feature group. The sciPy library is used to calculate the features from the time domain [20]. Additionally, the statsmodels library is applied to compute the sample autocovariance [19].

Fig. 4 depicts the distribution of selected scaled features for the different sensors. For a better overview, the test cycles containing anomalies are divided into two groups. The group "first level degradation" contains the first 10 cycles after degradation starts. It can be seen that the Peak2Peak (ptp) feature increases with the degradation of the ball screw. This feature is suitable to separate the condition classes especially for the acceleration sensor Acc_2. In the case of torque M_{BSD}, the signal-to-noise ratio (snr) of the power spectral density is appropriate for detecting the degradation of the ball screw.

Some of the extracted features show strong deviations from a normal distribution which is often assumed in statistical anomaly detection approaches. The Shapiro-Wilk test [21] is used to check whether the assumption of a normal distribution (null hypothesis) of the features (features of 48 test cycles without anomalies), can be accepted. Thereby, the null hypothesis is rejected at a p-value of less than 5%. For instance, large deviations from a normal distribution are determined for the following features of the torque M_{BSD}: crest factor ($p = 3.37 \cdot 10^{-6}$), frequency and amplitude of the most dominant peak of the frequency spectrum ($p = 9.73 \cdot 10^{-8}$, $p = 2.13 \cdot 10^{-4}$).

In summary, some features deviate strongly from a normal distribution. At the same time, no features can be preselected without knowledge of fault classes. Assuming a normal distribution can consequently reduce the monitoring quality. For this reason, the next chapter presents an approach that makes no assumptions about the underlying feature distribution.

Ensemble building process and decision-making procedure

A variety of methods exist for unsupervised anomaly detection that do not necessitate assumptions about the underlying distribution of the data. In unsupervised anomaly detection it is assumed that an unlabeled dataset is available which contains normal and faulty data [12]. The group of methods being used for this task includes k-nearest neighbors (KNN), local outlier factor (LOF) or angle-based outlier detection (ABOD) [22–24]. In the industrial environment, the type of fault and the effect of the anomaly on the signal features are usually unknown. As a consequence, the suitability of a particular method for anomaly detection is also unknown a priori. For monitoring purposes, an ensemble for semi-supervised anomaly detection is used. An ensemble is a set of methods whose predictions are combined in order to produce robust predictions [18]. The methods are selected based on the following reasons: The used approaches have a small number of hyperparameters and make no assumptions about the underlying distribution of the data. For anomaly detection, these methods provide a score, which is used to evaluate the test cycles. For an optimal description of the signals, features are extracted after segmenting the time series. The features are used to infer the state of the Z-axis. Since the range and meaning of the output scores varies depending on the method, the scores need to be scaled before an ensemble is formed [25]. Each of the methods in the ensemble computes an anomaly score $S(o)$ for an observation $o \in O$. In this work, an observation o is the standardized feature vector extracted using a test cycle. The approach by Kriegel et al. [25] is adapted to use unsupervised anomaly detection methods in a semi-supervised manner. First, the outlier scores are scaled. An anomaly score must be regular and normal for scores to be combined. Thereby, an anomaly score S is said to be regular if $S(o) \gg 0$ holds for a new observation o in the case of an anomaly. If o is not an anomaly, then $S(o) \approx 0$. To regularize the anomaly scores for KNN and LOF, the following formula (13) is used:

$$Reg_S^{base_S}(o) = max\{0, S(o) - base_S\}. \tag{13}$$

In the case of LOF, $base_{LOF} = 1$ is chosen. For the KNN method, the minimum distance of an observation o to its N_k nearest neighbor ($base_{KNN} = min(dist_{o,i}) \ \forall \ o \epsilon O^{train}, i \epsilon N_{k,o}$) in the training dataset O^{train} is used. For ABOD, formula (14) uses the logarithm function for regularization:

$$Reg_S^{loginv_S}(o) = -\log\left(S(o)/S_{max}^{train}\right). \tag{14}$$

In contrast to KNN and LOF, ABOD produces low scores for anomalies. In the second step, the anomaly scores are normalized, so that the resulting score follows the interval [0,1]. Various options exist for normalizing scores, such as Gaussian scaling. Gaussian scaling is used since it is robust to overfitting due to the small number of parameters (mean and standard deviation). For new

Table 1
Generated signal features for condition monitoring.

Time Domain	General-purpose features (Mean, Standard deviation, Root mean square, Kurtosis, Skewness, Signal-to-Noise-Ratio, Peak2peak, Shape factor, Crest factor, Interquartile range, Integral) Autocovariance (First ten lag coefficients)
Frequency Domain	Amplitude and frequency of the five most dominant peaks between 10 Hz and 50 Hz Power spectral density (Mean, Standard deviation, Root mean square, Kurtosis, Skewness, Signal-to-Noise-Ratio, Peak2peak, Shape factor, Crest factor, Interquartile range, Integral)

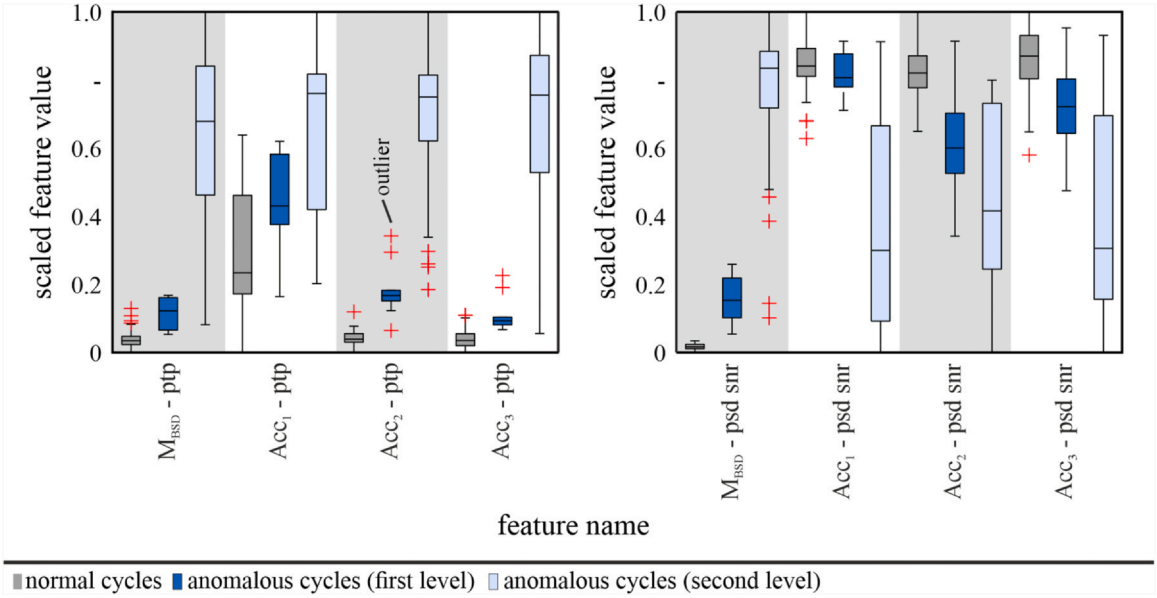

Fig. 4. Visualization of different scaled features.

observations $o \in O^{test}$, a normalized anomaly score is calculated according to formula (15). Before normalization, the mean $\mu_S^{Reg^{train}}$ and the standard deviation $\sigma_S^{Reg^{train}}$ of the regularized anomaly scores of the training data have to be determined. In addition, the Gaussian error function (erf) is used:

$$Norm_S^{gauss}(o) = max\left\{0,\ erf\left(\frac{Reg_S(o) - \mu_S^{Reg^{train}}}{\sigma_S^{Reg^{train}}\sqrt{2}}\right)\right\}, \quad \forall\ o \in O^{test}. \tag{15}$$

It is worth noting that identical anomaly detection methods with different hyperparameters can also be combined to form an ensemble. In the next step, an ensemble is formed in which the average of all scores of $OD_j \in OD$ anomaly detection methods is calculated:

$$P(o) = \frac{1}{|OD|}\sum_{OD_j \in OD} Norm_j(o), \quad \forall\ o \in O^{test}. \tag{16}$$

Considering a risk factor β, formula (17) is used to decide whether or not to produce an alarm when a new observation $o \in O^{test}$ is evaluated:

$$S^{final}(o) = \begin{cases} 1, & if\ P(o) > (1-\beta),\ \forall\ o \in O^{test} \\ 0, & otherwise \end{cases} \tag{17}$$

An alarm is produced in case $S^{final}(o) = 1$. The proposed monitoring system issues an alarm in case of signal/feature changes based on already seen time series. The extent to which changes are allowed depends on the size of the safety factor β. For this reason, the sensitivity of the monitoring system can be adjusted using the risk factor β. With larger values of β the sensitivity of the monitoring system increases and vice versa. Chapter 4 examines different values for β and its effect on the monitoring quality based on the proposed feature groups. In addition, the metrics for the evaluation of the monitoring quality are presented.

Assessment of the monitoring quality

First the correlation of the anomaly score with the degradation of the ball screw is investigated for each method (KNN, LOF and ABOD). An overview of regularized scores based on general purpose features is given in Fig. 5.

These scores serve as health indicators since they increase in value after the degradation starts. It can be noticed that using the general-purpose features, the scores change ten cycles after the anomaly starts. After replacing the ball screw, the scores assume a lower level as expected. The number of nearest neighbors k is identical for all methods. The Minkowski-metric is used to determine the distances. For ABOD, only the k nearest neighbors ($k = 5$) rather than all data points are used to compute the anomaly score (fast ABOD). The raw outlier scores are calculated using the PyOD library [26]. A total of 139 labeled test cycles are available for evaluation containing 57 anomalous test cycles. The ensemble approach is evaluated using two performance indicators. The sensitivity of the monitoring system is determined using the detection rate (DR), which is given by the ratio between detected test cycles with anomalies and the total number of test cycles with anomalies. In addition to the detection rate, the robustness of the monitoring system is determined using the false alarm rate (FR). The FR is calculated by the ratio of cycles without anomalies that are incorrectly classified as faulty to the total number of cycles without anomalies. In case the robustness to false alarms is low, the acceptance for the use of such a system in industrial practice is limited. The performance indicators are calculated as follows:

$$DR = \frac{detected\ faulty\ cycles}{number\ of\ faulty\ cycles}, \tag{18}$$

$$FR = \frac{misclassified\ normal\ cycles}{number\ of\ normal\ cycles}. \tag{19}$$

In a first evaluation step, the FR is determined. The initial training database consists of 10 test cycles without anomalies. For all other test cycles without anomalies, it is determined in each iteration whether a false alarm is issued. After each iteration, the data of a test cycle is added to the training database. The order of the tested cycles is the same for all feature groups. After the FR is determined, the DR is calculated based on the faulty test cycles.

For the ensemble method, the risk factor β is varied and the monitoring quality is examined. Four feature groups are used to detect the degradation, which are described in Table 1. In addition, the monitoring quality of the torque and acceleration sensors is

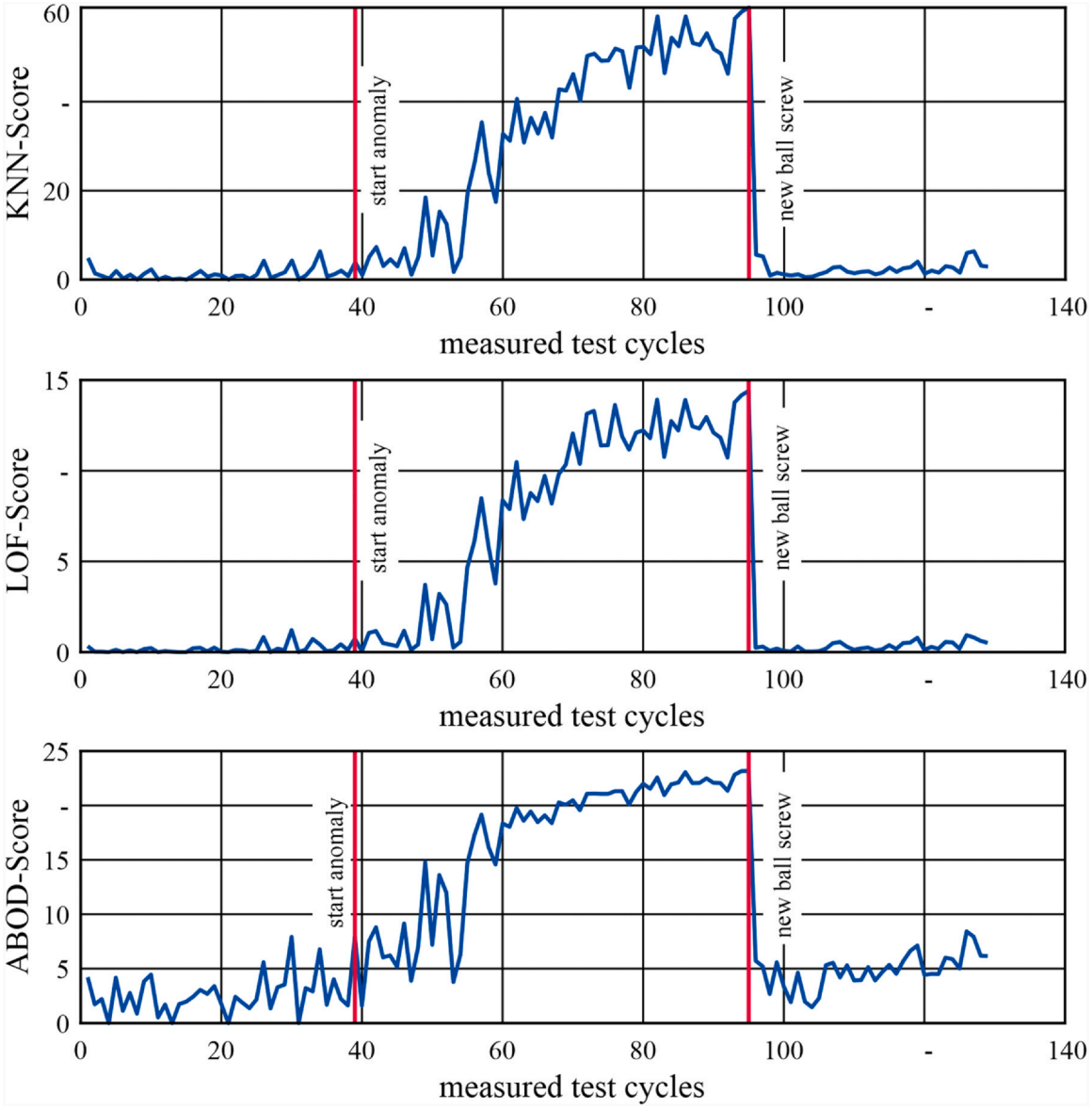

Fig. 5. Regularized scores using the torque M_{BSD} and the general-purpose features.

examined. The results of the evaluation are shown in Table 2. The values of the triaxial accelerometer are aggregated.

First, the monitoring quality of the torque M_{BSD} is examined. Using general-purpose features, no false alarms are issued for different values of beta ($\beta = 10^{-3}$, $\beta = 10^{-4}$, $\beta = 10^{-5}$). For $\beta = 10^{-3}$, no false alarms and a DR of 80.70% is obtained. The autocovariance features lead to smaller false alarm rates. For example, a FR of at most 5.26% is obtained for the selected values for β. At the same time, a lower DR compared to the general-purpose features are achieved. For example, a $\beta = 10^{-3}$ is leading to a DR of 75.44% and a FR of 0%. Most anomalies are detected using the spectral-density features compared to the previous feature groups. A DR of 100.0% and a FR of 0.0% is achieved using a $\beta = 10^{-4}$. Similar to the spectral density features, a high number of anomalies is detected considering the peaks of the frequency spectrum. In summary, it is seen that the ensemble method using the power spectral density features and the peaks of the FFT for the torque M_{BSD} achieves better results in terms of detected anomalies and false alarms than the other feature groups.

In the next step, accelerometer signals ACC_{1-3} are considered for monitoring purposes. High detections rates are realized using the general-purpose features, the autocovariance features as well as the spectral density features. However, high false alarm rates are obtained especially for spectral density features. For example, a $\beta = 10^{-2}$ leads to a FR of 26.32%. One reason for the increased false alarm rate is that signal changes are observed for the acceleration sensor ACC_1 at the beginning of data acquisition. Furthermore, the study shows that the autocovariance features are very robust to false alarms and sensitive to anomalies. No false alarms are issued for different values of β ($10^{-2} - 10^{-5}$).

The investigations highlight that the monitoring quality depends on the hyperparameters such as the safety factor β. However, these cannot be set optimally without knowledge of the fault classes. For industrial applications an initial safety factor $\beta = 10^{-5}$ should be set to avoid false alarms. With a larger amount of available test cycles it is also conceivable to specifically adjust the safety factor via a k-fold-cross validation. In this case, the safety factor is set in such a way that just no false alarms are issued for all folds.

Table 2
Performance indicators overview.

Method: Ensemble approach (KNN [22], LOF [23], ABOD [24]) using the general-purpose features										
Sensor	M_{BSD}					ACC_{1-3}				
β	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
DR(%)	89.47	82.46	80.70	77.19	77.19	98.25	98.25	98.25	98.25	98.25
FR(%)	7.89	5.26	0.00	0.00	0.00	18.42	10.53	7.89	5.26	5.26
Method: Ensemble approach (KNN [22], LOF [23], ABOD [24]) using the autocovariance features										
Sensor	M_{BSD}					ACC_{1-3}				
β	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
DR(%)	80.70	78.95	75.44	73.68	71.92	98.25	96.49	96.49	96.49	96.49
FR(%)	5.26	2.63	0.00	0.00	0.00	2.63	0.00	0.00	0.00	0.00
Method: Ensemble approach (KNN [22], LOF [23], ABOD [24]) using the spectral density features										
Sensor	M_{BSD}					ACC_{1-3}				
β	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
DR(%)	100.0	100.0	100.0	100.0	98.25	98.25	98.25	96.49	94.74	94.74
FR(%)	10.53	2.63	2.63	0.00	0.00	26.32	15.79	5.26	2.63	2.63
Method: Ensemble approach (KNN [22], LOF [23], ABOD [24]) using peaks of the frequency spectrum										
Sensor	M_{BSD}					ACC_{1-3}				
β	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
DR(%)	100	100.0	98.25	98.25	96.49	94.74	78.95	77.19	75.44	71.93
FR(%)	10.53	2.63	0.00	0.00	0.00	31.58	13.16	10.53	5.26	5.26

Conclusion

In this work, a robust ensemble approach was developed for ball screw condition monitoring. The approach was evaluated using the spindle torque and the information from an acceleration sensor. The sensor data was recorded during a regular test cycle. Different feature groups were generated for anomaly detection and their ability to detect the degradation of the ball screw were investigated. It is shown that component degradation has different effects on the feature groups and the used sensor. Without the knowledge of fault classes, the quality of certain features cannot be determined. As a result, the features should not be preselected. To produce robust predictions, an ensemble of different methods (KNN, LOF, ABOD) was created, which does not make any assumptions about the distribution of the sensor data.

The detection rate and the false alarm rate were used to evaluate the monitoring method and the quality of feature groups. The ensemble approach achieves high quality results for the spindle torque considering the features of the power spectral density and the peaks of the Fast Fourier Transformation. For the acceleration sensor, robust results are obtained considering the features of autocovariance.

In future work, the applicability of the presented methods to other fault types and components (e.g. spindles) has to be evaluated. Since there are also features that change with the replacement of a ball screw, the running-in behavior of ball screws should be examined more closely.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Funded by the Lower Saxony Ministry of Science and Culture under grant number ZN3489 within the Lower Saxony “Vorab” of the Volkswagen Foundation, Germany; supported by the Center for Digital Innovations (ZDIN), Germany. We also thank the Marposs Monitoring Solutions GmbH (Artis), Germany, for their support.

References

1 Imiela, J., 2006, Verfügbarkeitssicherung von Werkzeugmaschinen mit Kugelgewindetrieb durch modellbasierte Verschleißüberwachung (Dr.-Ing. Dissertation). Leibniz Universität Hannover.
2 Brecher, C., Witt, S., Yagmur, T., 2009, Influences of Oil Additives on the Wear Behavior of Ball Screws. Production Engineering, 3/3: 323–327. https://doi.org/10.1007/s11740-009-0168-y.
[3] Denkena, B., et al., 2006, Life-cycle oriented development of machine tools, in: Proceedings of 13th CIRP International Conference on Life Cycle Engineering, LCE 2006, Towards a closed loop economy, 05/31/2006 – 06/02/2006, Leuven, Belgium, pp. 693-698.
[4] Machelski, E.J., 1996, Get More Life out of That Ball Screw. Power Transmission Design, 38:53–56.
5 Jardine, A.K.S., Lin, D., Banjevic, D., 2006, A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance. Mechanical Systems and Signal Processing, 20/7: 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012.
6 Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., Siegel, D., 2014, Prognostics and Health Management Design for Rotary Machinery Systems—Reviews, Methodology and Applications. Mechanical Systems and Signal Processing, 42/1–2: 314–334. https://doi.org/10.1016/j.ymssp.2013.06.004.
[7] Lee, W.J., Wu, H., Yun, H., Kim, H., Jun, M.B.G., Sutherland, J.W., 2019, Predictive Maintenance of Machine Tool Systems Using Artificial Intelligence Techniques Applied to Machine Condition Data. Procedia CIRP, 80:506–511. https://doi.org/10.1016/j.procir.2018.12.019.
8 Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., Nandi, A.K., 2020, Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap. Mechanical Systems and Signal Processing, 138:106587https://doi.org/10.1016/j.ymssp.2019.106587.
9 Liu, R., Yang, B., Zio, E., Chen, X., 2018, Artificial Intelligence for Fault Diagnosis of Rotating Machinery: A Review. Mechanical Systems and Signal Processing, 108:33–47. https://doi.org/10.1016/j.ymssp.2018.02.016.
10 Peng, Y., Dong, M., Zuo, M.J., 2010, Current Status of Machine Prognostics in Condition-Based Maintenance: A Review. The International Journal of Advanced

Manufacturing Technology, 50/1–4: 297–313. https://doi.org/10.1007/s00170-009-2482-0.
11 Liao, L., Lee, J., 2009, A Novel Method for Machine Performance Degradation Assessment Based on Fixed Cycle Features Test. Journal of Sound and Vibration, 326/3–5: 894–908. https://doi.org/10.1016/j.jsv.2009.05.005.
12 Chandola, V., Banerjee, A., Kumar, V., 2009, Anomaly Detection. ACM Computing Surveys, 41/3: 1–58. https://doi.org/10.1145/1541880.1541882.
13 Qiu, H., Lee, J., Lin, J., Yu, G., 2003, Robust Performance Degradation Assessment Methods for Enhanced Rolling Element Bearing Prognostics. Advanced Engineering Informatics, 17/3–4: 127–140. https://doi.org/10.1016/j.aei.2004.08.001.
[14] Ruiz-Carcel, C., Starr, A., 2018, Data-Based Detection and Diagnosis of Faults in Linear Actuators. IEEE Transactions on Instrumentation and Measurement, 67/9: 2035–2047. https://doi.org/10.1109/TIM.2018.2814067.
[15] Zhao, S., et al., 2016. A Modified Mahalanobis-Taguchi System Analysis for Monitoring of Ball Screw Health Assessment, in: 2016 IEEE International Conference on Prognostics and Health Management (ICPHM), Ottawa, ON, Canada. 06/20/2016 – 06/22/2016. IEEE, pp. 1–7.
16 Zhai, S., Gehring, B., Reinhart, G., 2021, Enabling Predictive Maintenance Integrated Production Scheduling by Operation-Specific Health Prognostics with Generative Deep Learning. Journal of Manufacturing Systems. https://doi.org/10.1016/j.jmsy.2021.02.006.
17 Liu, R., Yang, B., Zio, E., Chen, X., 2018, Artificial Intelligence for Fault Diagnosis of Rotating Machinery: A Review. Mechanical Systems and Signal Processing, 108:33–47. https://doi.org/10.1016/j.ymssp.2018.02.016.
[18] Caruana, R., Niculescu-Mizil, A., Crew, G., Ksikes, A., 2004. Ensemble Selection from Libraries of Models, in: Twenty-first international conference on Machine learning - ICML '04. Twenty-first International Conference, Banff, Alberta, Canada. 04.07.2004 - 08.07.2004. ACM Press, New York, New York, USA, p. 18.
[19] Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and Statistical Modeling with Python, 92–96. doi:10.25080/Majora-92bf1922-011.
20 Virtanen, et al., 2020, SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17/3: 261–272. https://doi.org/10.1038/s41592-019-0686-2.
[21] Shapiro, S.S., Wilk, M.B., 1965, An Analysis of Variance Test for Normality (Complete Samples). Biometrika, 52/3/4: 591. https://doi.org/10.2307/2333709.
[22] Ramaswamy, S., Rastogi, R., Shim, K., 2000. Efficient Algorithms for Mining Outliers from Large Data Sets, in: Proceedings of the 2000 ACM SIGMOD International Conference on Management of data - SIGMOD '00. the 2000 ACM SIGMOD international conference, Dallas, Texas, United States. 15.05.2000 - 18.05.2000. ACM Press, New York, New York, USA, pp. 427–438.
[23] Breunig, M.M., Kriegel, H.-P., Ng, R.T., Sander, J., 2000. LOF, in: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data - SIGMOD '00. the 2000 ACM SIGMOD International Conference, Dallas, Texas, United States. 05/15/2000 – 05/18/2000. ACM Press, New York, New York, USA, pp. 93–104.
24 Kriegel, H.-P., Schubert, M., Zimek, A., 2008, Angle-Based Outlier Detection in High-Dimensional data. in: Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 08. the 14th ACM SIGKDD International Conference, Las Vegas. 08/24/2008ACM Press, Nevada, USA NewYork, New York, USA: 444.
[25] Kriegel, H.-P., Kroger, P., Schubert, E., Zimek, A., 2011. Interpreting and Unifying Outlier Scores, in: Proceedings of the 2011 SIAM International Conference on Data Mining. Proceedings of the 2011 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, Philadelphia, PA, pp. 13–24.
[26] Zhao, Y., Nasrullah, Z., Li, Z., 2019, PyOD: A Python Toolbox for Scalable Outlier Detection. Journal of Machine Learning Research (JMLR).