Wear detection for a cutting tool based on feature extraction and multivariate regression

Kurt Pichler[1], Mario Huemer[2], Gerhard Kaineder[1], Robert Schlosser[3], Bettina Dorfner[3], and Christian Kastl[1]

[1]*Linz Center of Mechatronics GmbH, Altenberger Str. 69, 4040 Linz, Austria*
[2]*Johannes Kepler University Linz, Altenberger Str. 69, 4040 Linz, Austria*
[3]*Leitz Gmbh & Co. KG, Leitzstraße 80, 4752 Riedau, Austria*

***Abstract*—In this paper, a method for detecting the wear of the cutting tool in laminate production is proposed. First, principal component analysis (PCA) for dimensionality reduction and clustering are used to determine from the measurement data whether a data set was recorded during production or during idling. Then, using only the data sets from actual production, a model for the wear is trained in a feature-based approach. The most relevant features for detecting wear are selected using a filter feature selection approach. Afterwards, an estimator for the wear is determined from the selected features by multivariate regression. A comparison of the results of two different sensor systems shows, that the sensor data already available for process monitoring can be reused for this purpose and that no additional sensor system needs to be installed.**

I. Introduction

In the dynamic landscape of industrial production processes, ensuring optimal performance and minimizing downtime is of paramount importance. Industries across the spectrum, from manufacturing to energy, are increasingly recognizing the value of proactive maintenance strategies to maximize operational efficiency. Among these strategies, condition monitoring and predictive maintenance have emerged as crucial tools that revolutionize maintenance practices and enable organizations to stay one step ahead of potential failures.

Traditionally, maintenance activities were often reactive, relying on routine inspections or repairs after equipment breakdowns occurred. However, this approach has inherent limitations, including unplanned downtime, production losses, and increased costs associated with emergency repairs.

Condition monitoring entails the continuous and real-time assessment of the health and performance of industrial assets. By employing a variety of sensors, advanced analytics, and machine learning techniques, organizations can gather vast amounts of operational data, monitor critical parameters, and identify early warning signs of potential failures. This approach empowers maintenance teams to make data-driven decisions and take corrective actions before equipment malfunctions occur, thus minimizing disruptions and optimizing productivity.

In laminate flooring production, an important step is cutting an angled edge from the laminate flooring. This is done by means of a rotating cutting tool. Of course, it is crucial that the cutting tool is sharp enough to produce a clean cutting edge. The sharpness of the cutting tool must therefore be constantly monitored so that it can be replaced if it becomes too worn.

Many publications report the extraction of features and subsequent classification in the feature space for the monitoring of rotating machines or components. For instance, in [1], a review of features for fault diagnosis for cutting tools is given. Different features are proposed for this use case. Subsequently, the features are used to determine the state of the cutting tools by means of known classification methods such as support vector machines (SVMs), decision trees or k-nearest-neighbor. The review paper [2] provides an overview of features used for fault diagnosis of gearboxes. An assignment of the features to specific fault types is also carried out. Paper [3] proposes a method that performs condition monitoring for a slow speed gearbox. After calculating the transmission error from the raw data, various features are calculated. Using recursive feature elimination, the significant features are selected from a large number of features. Finally, the gearbox condition is determined by random forests. Further approaches to feature extraction in different domains (time, frequency, wavelet,...) and subsequent classification of the state can be found in [4]–[8], just to name a few examples. The review papers [9], [10] on condition monitoring for bearings and gears also suggest many possible diagnostic features for rotating machinery.

In this application-oriented paper, we use well-known features and machine learning approaches to estimate the state of a cutting tool from measured signals. After extracting a set of features from the raw data, we first use an unsupervised approach to determine whether the tool is actually cutting or is in idle state. Then, only from the active measurements, wear is estimated in a supervised approach by selecting the most important features and using them as input for a multivariate support vector regression. A sensor system mounted specifically for the purpose of condition monitoring is compared with a sensor system already in use for process monitoring. The results show that the existing sensor data provide more accurate results and can therefore also be reused for condition monitoring.

The paper is organized as follows: in Section II, the problem statement is given and the test setup is explained. Section III deals with the actual wear estimation. Finally, Section IV provides test results, and in Section V conclusions are drawn.

2835-5776/24/$31.00 ©2024 IEEE
DOI 10.1109/IRI62200.2024.00030

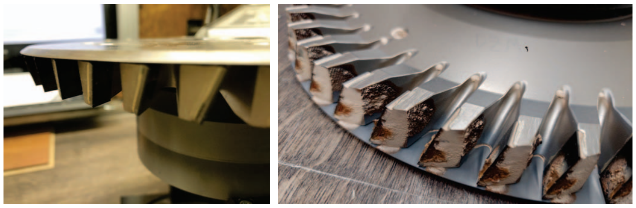

Fig. 1. New (left) and heavily worn (right) cutting tool

II. Problem Statement and Test Setup

As mentioned in the introduction, this paper aims to determine the wear of a cutting tool in laminate production from measurement data of the cutting process. A photo of such a cutting tool in new and worn condition is shown in Fig. 1. Up to now, the personnel of the laminate manufacturer inspect the cutting edge at regular intervals in order to assess the condition of the cutting tool on the basis of the cutting pattern and, if necessary, to change the tool. This regular inspection ties up resources in the plant that could be used more urgently elsewhere. If this inspection is replaced by continuous condition monitoring of the cutting tool, this will firstly relieve the workload of the employees and secondly enable the company to work more efficiently thanks to the resources that are made available as a result.

Test data were acquired during ongoing laminate production operations. In the production process, a co-rotating sensor system, called spike® [11], from pro-micron GmbH was installed at the cutting tool. This co-rotating sensor system (abbreviated as MS1 in the following), that was installed especially for the purpose of condition monitoring, provides the following measurements:

- Tension
- Torsion
- Bending moment in x-direction
- Bending moment in y-direction

The data recording is not completely regular. Data sets of 10 to 13 seconds each are recorded at intervals of between 5 and 10 minutes at a sampling rate of 2500 Hz, regardless of whether laminate is being processed or not.

In addition, data from a second sensor system permanently installed on the machine is also acquired. This sensor system (abbreviated as MS2 in the following) is installed for process monitoring anyway and is always available. It measures the following signals:

- Motor current
- Microphone in the exhaust hood
- Microphone in the cabin
- Infrared thermometer at the cutting edge before cutting
- Infrared thermometer at the cutting edge after cutting

This machine side measuring system records data at a sampling rate of 125 kHz for one minute every 15 minutes.

The two sensor systems record data at different times, they operate independently of each other and are not synchronized. While the motor current can be used as a very clear criterion for cutting activity detection in MS2, there is no such obvious criterion for MS1. Therefore, a method is to be developed for MS1 that detects the actual activity of the cutting tool.

During the test runs, measurement data was recorded during operation over a period of 5 weeks. A new cutting tool was installed at the beginning of each week. Due to the wear of the tools, the product quality is already decreasing towards the end of each week, but is still within the acceptable range. However, there is no continuous information about the ground truth of the wear of the cutting tool. Likewise, there is no information about the different types of laminate that were processed and about different machine settings. There were also some failures of the measuring systems in the course of these 5 weeks, so that no data is available in some cases.

For reasons of confidentiality, the data set cannot be made publicly available.

III. Wear Estimation

To estimate the wear of the cutting tool, in Section III-A a set of features is first extracted from the measured raw data. This feature set is then used in Section III-B to determine whether a data set originates from actual production or was recorded during idling. Subsequently, only the data sets from actual production are used in Section III-C to estimate wear by selecting the features of interest and using them as input for a multivariate support vector regression.

A. Feature Extraction

Each measurement data set from MS1 of 10 to 13 seconds duration consists of the 4 signals tension, torsion, as well as bending moment in x- and y-direction. At a sampling rate of 2500 Hz, this corresponds to a number of at least 100000 data samples per data set. For MS2, this amount of data is even significantly higher, since each data set contains 5 signals of duration 60 seconds at a sampling rate of 125 kHz. In order to significantly reduce the amount of data and thus simplify further evaluation from a computational and memory point of view, 30 features are extracted from each signal. This reduces the number of data samples to 120 per data set for MS1 and 150 for MS2. The data reduction, of course, results in a loss of information. However, the diverse features developed specifically for such purposes usually preserve enough information to be able to continue working with it in a meaningful way. The features include simple statistical measures such as mean value ($\bar{x}$ (1)), standard deviation (std (2)) and skewness (sk (3)) as well as other features recommended in the literature such as short time energy (ste (4)) or many more.The definition of the mentioned features for an exemplary signal represented by its data samples $x_1, \ldots, x_n$ is as follows:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

$$std = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{2}$$

$$sk = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^3} \tag{3}$$

$$ste = \frac{\sum_{i=1}^{n} x_i^2}{n} \tag{4}$$

For more detailed information about these features and a more extensive list of possible features, the reader is referred to [1]–[8].

The result of the feature extraction process is therefore an $m \times 30$ feature matrix for each signal, where $m \in \mathbb{N}$ denotes the total number of datasets/observations.

B. Cutting Activity Detection

As mentioned before, there is a simple way to detect cutting activity for MS2 by applying a threshold to the motor current. However, for MS1, there is no easily interpretable criterion to infer cutting activity. This subsection is therefore only relevant for the data of MS1 to identify the measurements where laminate was actually cut. Since there is also no ground truth available for this, it must be done in an unsupervised way.

In order to further reduce the dimensionality of the feature space, a principal component analysis (PCA) is first carried out. The main idea of dimensionality reduction is to map a high dimensional data set into a lower dimensional space. The purpose behind this can be, for example, to visualize the data or to remove unnecessary information or noise.

There are many methods for dimensionality reduction [12]. These include linear methods like PCA or multidimensional scaling (MDS) as well as nonlinear methods like t-distributed stochastic neighbor embedding (t-SNE), uniform manifold approximation and projection (UMAP) and self-organizing maps (SOM). We use PCA for our application because it can be also applied to new data (provided that the parameters of PCA have been saved). In principal, PCA makes a transformation of the coordinate system such that the data has largest variance along the new first coordinate, second largest variance along the new second coordinate, and so on. By construction, the consecutive principal components are non-correlated with each other. Additionally, the ordering of the principal components allows to eliminate the coordinates with the lowest information content and thus to reduce the dimensionality of the data. Since the PCA also provides the proportion of the explained variance of each principal component, the principal components can be selected that together explain a certain proportion of the total variance.

Now that the data is in the reduced principal component space, it can be clustered into different classes. Since we want to distinguish between two cases (cutting and non-cutting) in this application, we want to divide the data into two clusters. For this purpose, different methods are proposed in the literature [12], for example k-means clustering, k-medoids clustering, Gaussian mixture model clustering or hierarchical clustering. Here, k-means clustering is used, because on the one hand it achieved very good results in our tests, and on the other hand it can be easily applied to new data. Of course, the result is not guaranteed to yield a distinction between cutting and non-cutting. However, after looking at the data, process experts have confirmed that the clusters available here are highly likely to allow this distinction to be made.

When clustering is done, only those observations containing actual laminate cutting are kept, all other observations are discarded. We then assume that we only work with data sets that actually contain cutting. Of course, in the training phase of clustering, process experts must manually check whether the division into cutting or non-cutting makes sense. Once this has been done, this classification can subsequently be automated.

C. Multivariate Regression

Since we can now assume that there are only valid data sets from actual production left, we can now estimate the wear. We are assuming a supervised problem here, since the wear has of course increased continuously over the course of each test week, and when the tool was changed at the end of each week, clear wear was detected by the operators, which would suggest a tool change. However, no real ground truth is available since the actual tool status is not constantly monitored and logged. In a first step, we therefore assume linear wear over the course of each week. That is, we define a linear virtual ground truth that starts at 0 at the beginning of each week and reaches 1 at the end of each week. The value 0 corresponds to the new condition of the tool, while 1 represents the maximum permissible wear at which the tool should be changed. The linearity of wear is of course only a simplifying assumption. In experiments, where actual wear is monitored, this assumption must be confirmed or replaced by a more realistic one (e.g. parabolic). The linear assumption used here was made for this first study in accordance with process experts.

In a first step, the features relevant for the regression must be selected. This is done using a feature selection filter approach. For this purpose, the data is labeled in 3 classes, namely "*beginning of the week (new)*", "*middle of the week (medium wear)*", and "*end of the week (heavy wear)*". Then the features that best distinguish these 3 classes are identified. Since the condition of the cutting tool can be roughly divided into the 3 classes using the selected features, the features are assumed to be suitable candidates for estimating the wear using regression. The actual feature selection is carried out in a forward filter approach [13]. A feature set is initialized empty, and then single features are added sequentially to it. This is repeated, until a certain number of features or a certain information content in the feature space is reached. In each step of the forward feature selection process, Dy-Brodley measure [14] is used as a selection criterion. For a k-dimensional feature space containing data from $C \in \mathbb{N}$ classes, the feature values of each class can be written as a matrix $\mathbf{X}_c \in \mathbb{R}^{n_c \times k}$ with $c \in \{1, \ldots, C\}$ and $n_c \in \mathbb{N}$ denoting the number of observations in class c. When $\boldsymbol{\mu}_c \in \mathbb{R}^k$ and $\Sigma_c \in \mathbb{R}^{k \times k}$ denote the mean values and covariance matrices of each class c, and $\boldsymbol{\mu} \in \mathbb{R}^k$ denotes the mean value of all samples (regardless the class membership), the within scatter $\mathbf{S}_{\mathrm{W}} \in \mathbb{R}^{k \times k}$ is defined as

$$\mathbf{S}_{\mathrm{W}} = \sum_{c=1}^{C} \frac{n_{\mathrm{c}}}{n} \cdot \Sigma_{\mathrm{c}} \tag{5}$$

and the between scatter $\mathbf{S}_{\mathrm{B}} \in \mathbb{R}^{k \times k}$ as

$$\mathbf{S}_{\mathrm{B}} = \sum_{c=1}^{C} \frac{n_c}{n} \cdot \left(\boldsymbol{\mu}_{\boldsymbol{c}} - \boldsymbol{\mu}\right)^{\mathsf{T}} \cdot \left(\boldsymbol{\mu}_{\boldsymbol{c}} - \boldsymbol{\mu}\right), \tag{6}$$

where $n \in \mathbb{N}$ is the overall number of observations, i.e. $n = \sum_{c=1}^{C} n_c$. Then, the Dy-Brodley measure J is defined as

$$J = \operatorname{tr}\left({\mathbf{S}_{\mathrm{W}}}^{-1} \cdot \mathbf{S}_{\mathrm{B}}\right). \tag{7}$$

where $\operatorname{tr}(\cdot)$ denotes the trace operation. By maximizing J in each step when a feature is added, the distance between different classes is maximized while the distance within the same class is kept minimal. The relative change of the Dy-Brodley measure can be used as stopping criterion. If the addition of another feature does not lead to a significant improvement, the feature selection is stopped.

Assuming that k features are selected, these features are used as an input to a support vector regression model with the virtual ground truth as target value. We chose support vector regression because, unlike linear regression, it can also map both linear and nonlinear relationships. Since nonlinearities cannot be excluded a priori in the present application, they must be taken into account in the model. Assume that the finally selected feature values are stored in a matrix $\mathbf{X} \in \mathbb{R}^{m \times k}$ with k features and m observations, where each row $\mathbf{x}_i \in \mathbb{R}^k, i = 1, \ldots, m$ of $\mathbf{X}$ denotes one observation with k features. Furthermore, the target values, in this case the virtual ground truth, are stored in the vector $\mathbf{y} \in \mathbb{R}^m$, where each value y_i denotes the target value of the i-th observation. Then, nonlinear support vector regression finds the coefficients that minimize the Lagrange function

$$\begin{aligned} L =& \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right) G\left(x_i, x_j\right) \\ &+ \epsilon \sum_{i=1}^{m} \left(\alpha_i + \alpha_i^*\right) - \sum_{i=1}^{m} y_i \left(\alpha_i - \alpha_i^*\right) \end{aligned} \tag{8}$$

subject to

$$\begin{aligned} &\sum_{i=1}^{m} \left(\alpha_i - \alpha_i^*\right) = 0 \\ &\forall i : 0 \leq \alpha_i \leq C \\ &\forall i : 0 \leq \alpha_i^* \leq C. \end{aligned} \tag{9}$$

For predicting the target value for a new observation $\mathbf{x} \in \mathbb{R}^k$, the function

$$f(\mathbf{x}) = \sum_{i=1}^{m} \left(\alpha_i - \alpha_i^*\right) G\left(\mathbf{x}_i, \mathbf{x}\right) + b \tag{10}$$

is used. In the previous equations, the constant C is a positive numeric value that controls the penalty imposed on observations that lie outside the ϵ-margin. It is used to prevent overfitting and serves therefore as a regularization parameter. It determines the trade-off between the flatness of $f(\mathbf{x})$ and the amount up to which deviations larger than ϵ are tolerated. For more details about support vector regression, the reader is referred for instance to [15].

To assess the quality of the regression model, the root mean square error (RMSE) between the virtual ground truth and the prediction is calculated. If $\mathbf{y} \in \mathbb{R}^m$ with elements $y_i, i = 1, \ldots, m$ denotes the virtual ground truth and $\hat{\mathbf{y}} \in \mathbb{R}^m$ with elements $\hat{y}_i, i = 1, \ldots, m$ the prediction, the RMSE is defined as

$$rmse(\mathbf{y}, \hat{\mathbf{y}}) = \sqrt{\frac{\sum_{i=1}^{m} \left(\hat{y}_i - y_i\right)^2}{m}}. \tag{11}$$

The smaller the RMSE value, the better the approximation by the model.

IV. Results

As already briefly described earlier, five weeks of data were recorded during operation. In the first step, the 30 features (and additionally a few manually engineered features) for each signal and all measurements of both sensor systems were extracted. Then, cutting activity detection was performed. While for MS2, measurements were simply discarded if the motor current does not exceed a certain threshold, the method described in Section III-A was used for MS1. In PCA, the first 8 principal components were selected, since together they describe more than 95 % of the total variance . In the principal component space, k-means clustering was performed with $k = 2$ classes. Fig. 2 shows the clustering result in the space of the first two principal components. It can already be seen here that two quite clearly defined clusters are formed. If we apply this clustering result to the course of a single well-suited feature, we can see that it is apparently possible to distinguish well between cutting and non-cutting. This was also confirmed by process experts. Fig. 3 shows such a well-suited feature, namely the short-time energy for the bending moment in x-direction, for each of the 5 recorded weeks. The measurements that are also used for further evaluation are shown in blue. In red are those measurements that are discarded because they were not acquired during the actual cutting. One can also clearly see here the missing data due to the failure of the measurement system, especially in week 3 and week 4 of the test runs many data are missing. In total, the following number of valid samples were available in each of the 5 weeks: 203 samples in week 1, 515 samples in week 2, 45 samples in week 3, 124 samples in week 4, 454 samples in week 5.

Subsequently, feature selection and regression were performed using only the remaining measurements after cutting activity detection. Since the two sensor systems recorded at different times, the data must first be brought to common time stamps for joint processing. For this purpose, the

Fig. 2. Clustering result in the first two principal components to distinguish between cutting and non-cutting for MS1.

feature curves were interpolated with common time stamps. In addition, the feature curves were smoothed using the Whittaker smoother [16] to reduce noise effects. In feature selection, the following 4 features were selected for MS1:

- Short-time energy of bending moment in x-direction
- Root mean square of bending moment in x-direction
- Short-time energy of bending moment in y-direction
- Root mean square of bending moment in y-direction

From the features of MS2, the following 3 features were selected:

- Median value of the temperature difference before/after cutting
- Spectral spread of the current
- Median value of the current

All of these features were then used as inputs for the support vector regression. For regression, leave-one-week-out cross validation was applied. That is, one of the 5 recorded weeks was always used for testing, while training was done with the other 4 weeks. The result of the cross validation is shown in Fig. 4. In blue, the virtual ground truth is shown, in red the output of the support vector regression model. It can be seen that the general trend is already very well reproduced. However, the actual performance cannot yet be assessed with the data currently available. On the one hand, the data of 4 of the 5 measurement weeks stop before the tool was changed, because the measurement system failed. On the other hand, the ground truth is only an assumption, which, however, has not been substantiated by actual inspections of the cutting tool.

Especially for week 3 and 4, no significant statement about the quality is possible, the data set here is very limited. In week 1, the trend is well represented, especially at the beginning of the week, but towards the end of the week the model flattens out considerably. The reason could also be, for instance, a different type of laminate or different machine settings. However, since we do not have this information in the present test, this can neither be verified nor denied. For weeks 2 and 5, the model looks very similar to the virtual ground truth. The RMSE for each week can be seen in the last column Table I. The mean value over all weeks is 0.1035. Since for weeks 3 and 4 the measurement systems failed very early, these weeks are not so representative. Therefore, we also calculated the weighted mean, where the duration of the measurements in each week are the weights. However, with a value of 0.1031 this does not change the mean RMSE significantly.

To determine the contribution of the two sensor systems to the result, the same regression analysis was also performed using only the features from MS1 and MS2. The resulting RMSE values as well as the mean and weighted mean values can be seen in Table I in columns 2 and 3, respectively. It can be seen that MS2 gives almost as good results as the combination of both systems, while MS1 alone gives significantly worse results. Therefore, it can be concluded that the data from MS2 can not only be used for process monitoring, but can also be reused for condition monitoring.

TABLE I
CROSS VALIDATION RMSE FOR EACH TEST WEEK

Class	MS1	MS2	Combined
Week 1	0.2609	0.1484	0.1425
Week 2	0.1557	0.0440	0.0372
Week 3	0.0558	0.0143	0.0187
Week 4	0.2488	0.2208	0.2133
Week 5	0.1675	0.0944	0.1060
Mean value	0.1778	0.1044	0.1035
Weighted mean	0.2002	0.1039	0.1031

V. CONCLUSIONS

In this paper we have shown that the principle of data driven wear detection for a cutting tool works. Both cutting activity detection and regression based on extracted features provide plausible results. The most important result is the reusability of MS2 also for the purpose of condition monitoring. This means that no additional sensor system is required, which is a great advantage in terms of cost and economy. However, there is still some development work to be done before the system is ready for actual use.

Most importantly, significantly more test data need to be collected, and the sensor systems have to work more reliably. In addition, a valid ground truth about wear should be available, for instance through regular inspection of the tool. However, this will be difficult to realize in ongoing production operations, as it conflicts with the interests of the laminate producer. Also different laminate types and machine settings should be considered for the regression model. These must be measured or logged in a further test run. For the regression, other models such as kernel regression or neural networks could be used. Also a transformation of the data for regression (for instance Box-Cox transformation [17]) could bring an improvement of the model. In addition, the applicability of the method to other cutting machines and tools should be tested.

ACKNOWLEDGMENT

This work has been supported by the COMET-K2 Center of the Linz Center of Mechatronics (LCM) funded by the

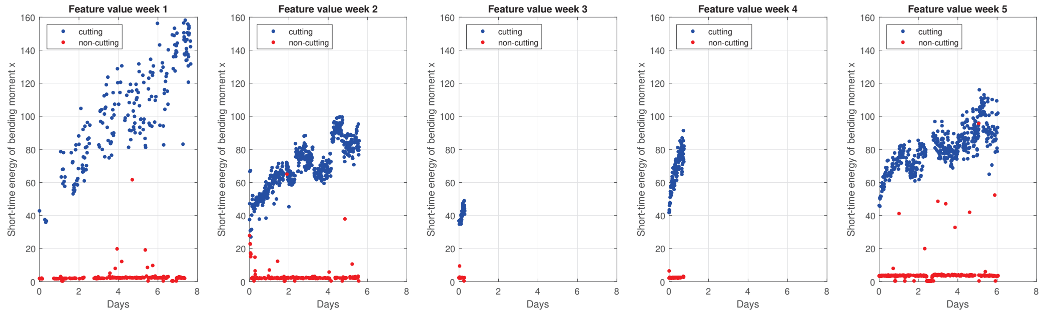

Fig. 3. Feature value short-time energy of the signal bending moment in x-direction for the five test weeks.

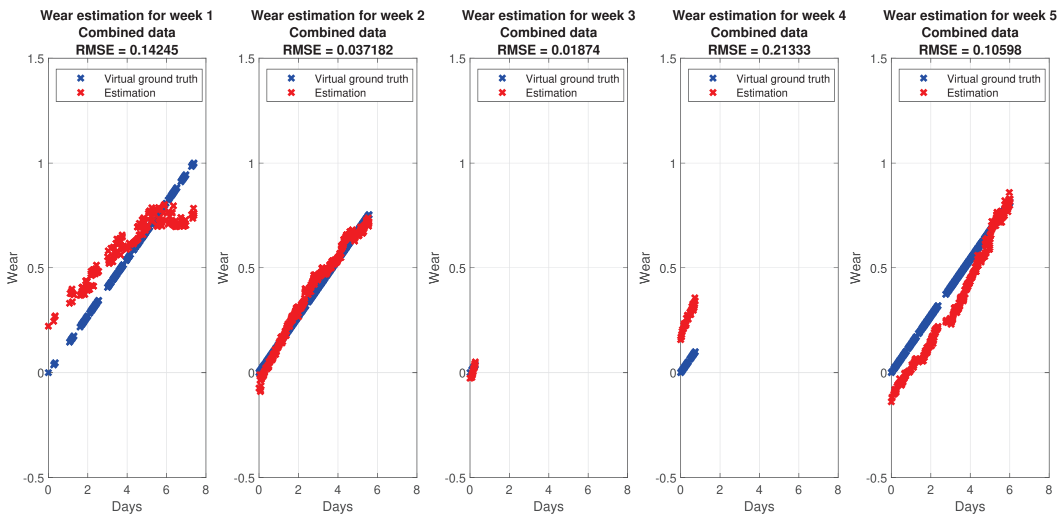

Fig. 4. Estimated wear and virtual ground truth for the five test weeks.

Austrian federal government and the federal state of Upper Austria.

References

[1] N. R. Tambake, B. B. Deshmukh, and A. D. Patange, "Data driven cutting tool fault diagnosis system using machine learning approach: A review," *Journal of Physics: Conference Series*, vol. 1969, no. 1, 2021.

[2] V. Sharma and A. Parey, "A review of gear fault diagnosis using various condition indicators," *Procedia Engineering*, vol. 144, pp. 253–263, 2016.

[3] S. Sendlbeck, A. Fimpel, B. Siewerin, M. Otto, and K. Stahl, "Condition monitoring of slow-speed gear wear using a transmission error-based approach with automated feature selection," *International Journal of Prognostics and Health Management*, 2021.

[4] L. Barbini, A. Ompusunggu, A. Hillis, J. du Bois, and A. Bartic, "Phase editing as a signal pre-processing step for automated bearing fault detection," *Mechanical Systems and Signal Processing*, vol. 91, pp. 407–421, jul 2017.

[5] H. Wang and P. Chen, "Fault diagnosis of centrifugal pump using symptom parameters in frequency domain," *Agricultural Engineering International: the CIGR Ejournal*, vol. IX, 2007.

[6] R. Bajric, N. Zuber, G. Skrimpas, and N. Mijatovic, "Feature extraction using discrete wavelet transform for gear fault diagnosis of wind turbine gearbox," *Shock and Vibration*, vol. 2016, 2016.

[7] A. Hu, L. Xiang, S. Xu, and J. Lin, "Frequency loss and recovery in rolling bearing fault detection." *Chinese Journal of Mechanical Engineering*, vol. 32, no. 35, 2019.

[8] M. Jalil, F. A. Butt, and A. Malik, "Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals," in *The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering*, 2013, pp. 208–212.

[9] D. Wang, K. Tsui, and Q. Mia, "Prognostics and health management: A review of vibration based bearing and gear health indicators," *IEEE Access*, vol. 6, pp. 665–676, 2017.

[10] S. Singh and M. Vishwakarma, "A review of vibration analysis techniques for rotating machines," *International Journal of Engineering Research & Technology*, vol. 4, no. 03, pp. 757–761, 2015.

[11] pro-micron GmbH, "spike®," 2023. [Online]. Available: https://www.pro-micron.de/spike

[12] S. Marsland, *Machine Learning - An Algorithmic Perspective*. Chapman & Hall/CRC, 2009.

[13] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.

[14] J. Dy and C. Brodley, "Feature selection for unsupervised learning," *Journal of Machine Learning Research (JMLR)*, vol. 5, 12 2004.

[15] V. N. Vapnik, *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.

[16] P. H. C. Eilers, "A perfect smoother," *Analytical Chemistry*, vol. 75, no. 14, pp. 3631–3636, 2003, pMID: 14570219. [Online]. Available: https://doi.org/10.1021/ac034173t

[17] G. E. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society Series*, vol. 26, no. 2, pp. 211–252, 1964.