Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 88 (2020) 312–317

www.elsevier.com/locate/procedia

13th CIRP Conference on Intelligent Computation in Manufacturing Engineering, CIRP ICME '19

Wear curve based online feature assessment for tool condition monitoring

Berend Denkena[a], Benjamin Bergmann[a], Tobias H. Stiehl[a]*

[a]*Leibniz Universität Hannover, Institute of Production Engineering and Machine Tools, An der Universität 2, 30823 Garbsen, Germany*

* Corresponding author. Tel.: +49-511-762-18003; fax: +49-511-762-5115. *E-mail address:* stiehl@ifw.uni-hannover.de

Abstract

The performance of a process monitoring system is determined by the information available to it. Existing methods for selecting relevant process information (features) work offline with data of faulty processes that is often unavailable or neglect random disturbances. This increases the risk of choosing non-sensitive features. Hence, this paper investigates whether a non-sensitive feature is detectable online in an initial selection of features presumed to be sensitive. A method for quantifying and assessing trends in features online is described. In the validation with turning and drilling processes, a single non-sensitive feature was detected successfully in seven out of eight test cases.

Peer review under the responsibility of the scientific committee of the 13th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 17-19 July 2019, Gulf of Naples, Italy.

Keywords: Tool condition monitoring; Feature selection; Online

1. Introduction

Tool condition monitoring systems can increase machine availability and reduce the risk of damaging workpieces. The systems extract information (features) from sensor signals, analyze them and thereby detect tool wear, breakage or other anomalies during machining. The performance of these systems depends on the sensitivity of the evaluated features to the aforementioned failures. The certain selection of such features, however, requires machine- and process-specific data of deficient machining, the very instance to be prevented. Common practice, however, includes analyzing the signal in hindsight without knowing the true wear or from experience [1, 2]. Inherent to this process is the risk of selecting features that are insufficiently sensitive to failures. Online inspection of an initial feature selection by assessing the long-term behavior of features might reduce the risk of inadequate feature selection.

Nomenclature

α	confidence level
I	count of features assessed
N	count of process repetitions / segments
R_i^2	coefficient of determination relating to feature i
R_{thres}^2	threshold for features to be accepted as trending
Q	test value of Dixon's Q test
Q_{crit}	critical value for Dixon's Q test
x_i	sequence of means for feature i
y_i	low pass filtered sequence of means for feature i
z_i	normed change for feature i

The importance of feature selection for system sensitivity has been stressed extensively, e.g. in [1] or [3]. Combined, both papers review more than 50 publications covering twelve different signal sources and sensors, over ten basic principles for feature extraction, as well as feature selection methods. As these can be combined in various ways, an abundance of features is available, most of which are distorted and indifferent

2212-8271 © 2020 The Authors. Published by Elsevier B.V.

Peer review under the responsibility of the scientific committee of the 13th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 17-19 July 2019, Gulf of Naples, Italy.
10.1016/j.procir.2020.05.054

to process conditions [1]. The methods either select failure sensitive features based on data of faulty processes, if available, or otherwise features assumed to be failure sensitive based on experience, e.g. [4]. In [5], [6], [7] and [8], for example, the trend of features was correlated to measured tool wear or surface roughness of the machined component, to identify features suited for wear prediction. In addition, [9] correlated features to an ideal wear trend, represented by a straight line with a slope of 40°, as a selection criterion.

All of the cited works assessed the sensitivity of features offline after at least one tool life was consumed. The resulting feature selection is specific to the examined characteristics of the machining process. The selected features are then used to monitor the very process previously examined assuming that the determined sensitivity to wear persists. However, even well-correlated features can be randomly disturbed [1], for example, due to a dirty or malfunctioning sensor. The issue is emphasized by Jemielniak [10], who introduces the repeatability of sensitive features as a selection criterion.

Established practice in industry extends to monitoring a process with features determined to be sensitive for comparable process characteristics. This further increases the risk of selecting unsuited features, as monitoring tasks are process-oriented problems closely related to the characteristics of the specific machining process [11].

Unsuited initial selections and random disturbances are generally addressed by streaming or online feature selection methods. They are designed to continuously select features parallel to a process or task performed. Concepts used to determine the relevance of a feature rely on, for instance, Chi-Squared statistics, probabilistic significance or entropy-based values [12]. Selected online feature selection methods were examined for metal cutting applications, e.g. integrated with the decision-making method Parsimonious Ensemble+ [13]. For tool wear detection in turning or drilling, however, methods for online feature selection have received little attention.

Existing methods for feature selection in tool condition monitoring work predominantly in hindsight of the monitored process. This allows using high-quality information such as real failure data. However, even a flawless selection of features might lead to impaired monitoring performance, as offline selection cannot overcome risks arising from random disturbances. While online feature selection methods are generally capable of detecting such disturbances, existing methods are neither specifically designed nor evaluated for tool wear monitoring. A reliable method, however, is required for autonomous operation and parameterization of a monitoring system. This work examines how trends in features are detectable and utilizable during tool wear monitoring to exclude an irrelevant feature from an initial selection online.

2. Trend-Based Online Feature Assessment

As many potent feature selection methods exist, an initial selection mostly composed of sensitive features is assumed. Therefore, a method is designed to identify and exclude an individual non-sensitive feature from the initial selection. In this work, the focus is on failures or other anomalies developing over a prolonged period of time, such as wear. Trends are analysed based on repetitive machining operations, represented here by reoccurring process segments of identical turning or drilling operations. The method provides an assessment after every new repetition that is completed.

Firstly, the method quantifies trends in the features monitored. For this, the offset of a signal in every repetition is removed to compensate drifts in sensor signals that result from sources other than the process. A mean for each signal i is then calculated per repetition n resulting in a sequence of means $x_i[n]$. The resulting sequence of means is normed to process fluctuations represented by noise in the sequence. This is done by calculating the dispersion in the sequence (1) in reference to a moving average filtered (5 samples) version $y_i[n]$ of the original sequence $x_i[n]$. The dispersion measure s_i of a sequence $x_i[n]$ is determined and the normed change $z_i[n]$ computed:

$$s_i = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x_i[n]-y_i[n]\right)} \tag{1}$$

$$z_i[n] = \frac{x_i[n]-x_i[1]}{s_i} \tag{2}$$

Next, a regression function is fitted in the normed change $z_i[n]$ of each feature using the method of least squares. For this, a linear function is used as the method is intended to identify non-sensitive features in the stage of uniform wear rates. This yields a regression statistic that quantifies the trend in features (Figure 1).

Secondly, outlier detection is performed to identify features that do not trend with the majority of features. For this, the coefficients of determination (R_i^2) of the linear regression for the individual normed changes are compared. Two different methods are subsequently compared for this task:

- Thresholds and majority approach
  The coefficient of determination quantifies the proportion of variance in the sequence of means that is predictable with the fitted regression function. A feature i is considered to show a trend if the majority of the variance in the sequence can be accounted to the trend. This translates to the coefficient of determination with the following condition:

  $$R_i^2[n] > R_{thres} \quad \text{with } R_{thres} = 0.5 \tag{3}$$

  where n is the repetition after which the calculation is performed. A feature is considered non sensitive if it is the only feature not showing a trend.

- Dixon's Q-Test
  This outlier test is suitable for small sample sizes of 3 to 7. According to [14], it is robust against a variety of non-normal distributions for small sample sizes of 3 to 5. To assess whether the smallest value is an outlier, the Q value is calculated in a first step according to [15]:

$$Q[n] = \frac{R_2^2[n] - R_1^2[n]}{R_I^2[n] - R_1^2[n]} \quad (4)$$

where the evaluated sample is composed of the previously calculated coefficients of determination sorted in an ascending order $R_1^2 \le R_2^2 \le \cdots \le R_I^2$ for any repetition n. The smallest value $R_1^2[n]$ is considered an outlier if:

$$Q[n] \ge Q_{crit} \quad (5)$$

where Q_{crit} is the critical value. It is a function of the confidence level α and the number of elements in the samples I. Here, a one-tailed test is performed with critical values Q_{crit} as determined by [15]. If $R_1^2[n]$ is classified as an outlier, then the associated feature is considered to be non-sensitive.

Features determined to be non-sensitive by the method are not considered in the decision making part of the monitoring system. This state persists until the next process repetition is completed and features are reassessed. If the feature is again considered to be non-sensitive, it is again excluded from the decision-making process (Figure 2).

3. Experiments Conducted

Indexable inserts and drills were worn in series of repetitive face turning and drilling operations. Established features for monitoring were extracted, processed and their behavior with increasing machining time was examined.

3.1. Experimental Setup and Process Segmentation

Experiments with face turnings comprised two series employing a previously unused indexable insert (CNMG 120408). Machining in a series continued until the operator observed clear indicators of tool wear (diminished workpiece surface quality, emitted sound, chip structure). Machining was then stopped and the wear of the indexable insert was examined with a microscope (width of flank wear land VB, general condition). Workpieces had a diameter of 100 mm, a length of 200 mm and consisted of 42CroMo4+QT (DIN EN 10025). The employed machining center (DMG MORI NTX 1000) had reached operating temperature before experiments started and no cooling lubricant was used.

Experiments for drilling comprised two series of holes (depth 30 mm, diameter 6.8 mm) employing previously unused high-speed steel drills. Drilling was continued until the drill broke or the operator observed a red glowing tip. Up to 118 holes were drilled continuously in a single disk consisting of S355JR (DIN EN10025-2) before it was replaced by a new workpiece to continue the series if necessary. Idle time of up to 15 minutes occurred before drilling continued on the new workpiece. Within a single disk, holes closer to the rotational axis of the disk were drilled before the more distant ones. The automated program was based on a drilling cycle with chip breaking (every 5 mm) as provided by the employed universal lathe (Gildemeister CTX420 linear).

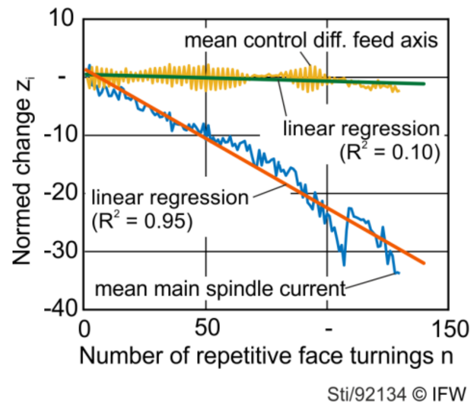

Fig. 1. Exemplary calculation of regression functions (experimental series 1)

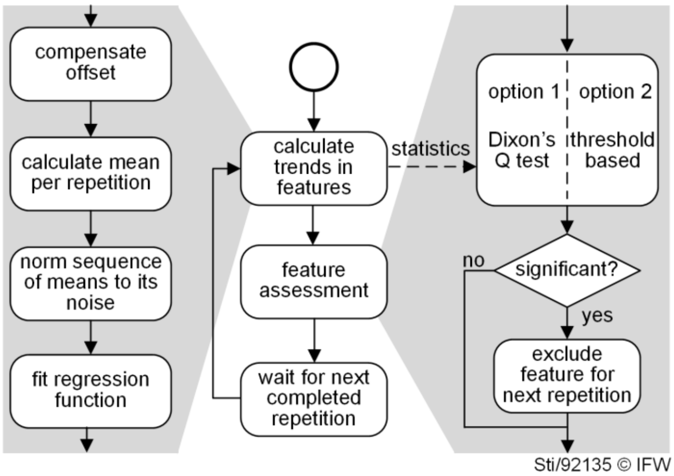

Fig. 2. Flow chart for online feature assessment

A total of four series were conducted, thereby consuming two indexable inserts and two drills. While for turning cutting speed was altered between series, feed rate and depth of the cut remained constant with 0.2 mm and 0.75 mm respectively. For drilling, the feed rate was fixed to 0.1 mm while the cutting speed was altered (Table 1).

On both machines, process data was recorded using a Beckhoff IPC with TwinCat (Figure 3 is exemplary for face turning operation). For turning, six signals were accessed as provided by the control of the machine tool: currents of the non-feed axis x and y, the feed axis z, torque and current of the main spindle drive, and the interpolation type. For drilling, three different signals were recorded: torque of the drive rotating the drill, the torque of the feed axis z, and the control difference of the axis z.

Table 1. Experiments conducted

series	process	cutting speed (m/min)	number of workpieces	repetitions
1	turning	300	1	130
2	turning	350	1	142
3	drilling	75	1	88
4	drilling	60	3	334

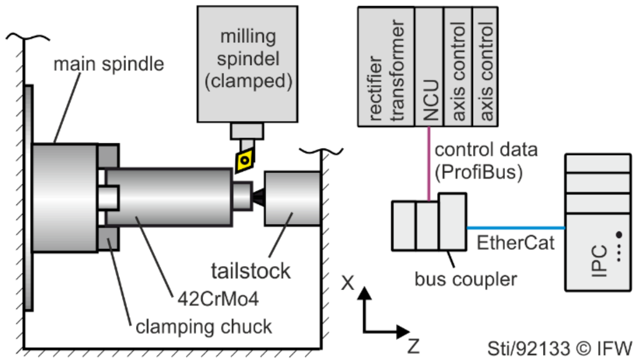

Fig. 3. Setup for experiments and sensor signal acquisition in face turning

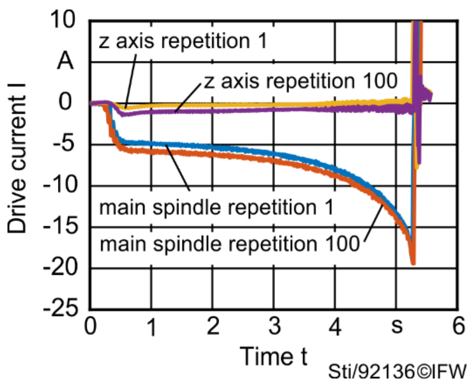

Fig. 4. Sensor signals recorded for repetition 1 and 100 in the sequence

Once recorded, the process data was divided into segments of identical face turning or drilling operations representing a single repetition. This process is based on the interpolation type and is suitable for online operation. Figure 4 shows such a repetition from the beginning and from the end of a series of face turnings.

3.2. Signal Feature Sensitivity and Test Samples

To determine the sensitivity of features to wear, the data generated from the experiments is processed analogous to the trend assessment part of the method yielding the normed change $z_i[n]$. The latter expresses the observed change to the first repetition in ratio to process fluctuations. By analogy to statistical process monitoring, the value $z_i[n]$ is interpreted as an approximate for the sensitivity of the underlying feature to wear. A feature is considered sensitive if the last five values of its normed change $z_i[n]$ are all greater than 5. Presented below are normed changes for face turnings (series 2, Fig. 5) and drillings (series 4, Fig.6). Features considered non-sensitive are the mean control difference of the feed axis for drilling (series 2) and the mean current driving the y-axis for turning (series 4).
The recorded data is structured into test sets labeled with ground truth to validate the method described in section 2. Test sets comprise all or a subset of the features recorded and are named A to F (Table 2). They include between 3 and 5 features in total and 1 or 0 non-sensitive features. Because data for drilling experiments comprises only three signals, it is limited to subset F. For turning, different subsets are composed from the features available. Table 2 specifies features that are part of a subset.

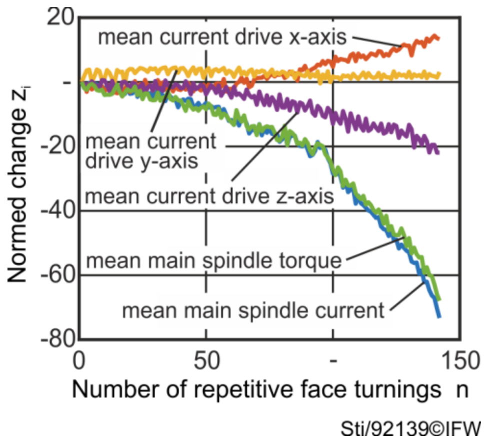

Fig. 5. Mean of signals over repetitive face turnings

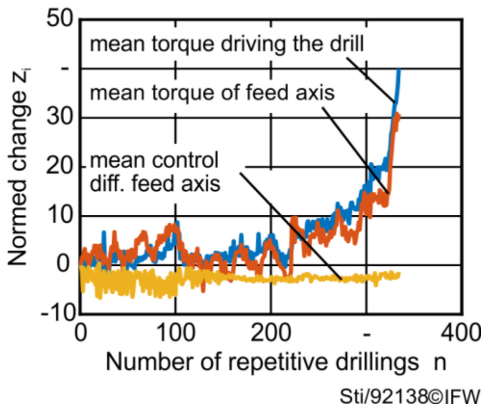

Fig. 6. Mean of signals over repetitive drilling operations

4. Results

The performance of the previously described methods is evaluated using the defined test sets. Figure 7 exemplifies the behavior of the coefficients of determination for test set 2A (series 2, feature subset A), a turning process. The main spindle current and torque show a trend first. The current driving the y-axis and the z-axis is considered trending after repetition 89 and 75, respectively ($R^2_{thres} = 0.5$). The current driving the x-axis shows no trend as the threshold is not reached at any point of time. Dixon's Q test is significant from repetition 90 onwards classifying the signal of the current driving the y-axis as non-sensitive.

Table 2. Composition of feature subsets

Subset	Torque of tool drive	x-axis	main spindle	Drive currents of axis main spindle	x	y	z	Control diff. z-axis
A			+	+	+	+	+	
B			+	+	+		+	
C				+	+	+	+	
D			+	+	+			
E				+		+	+	
F	+	+						+

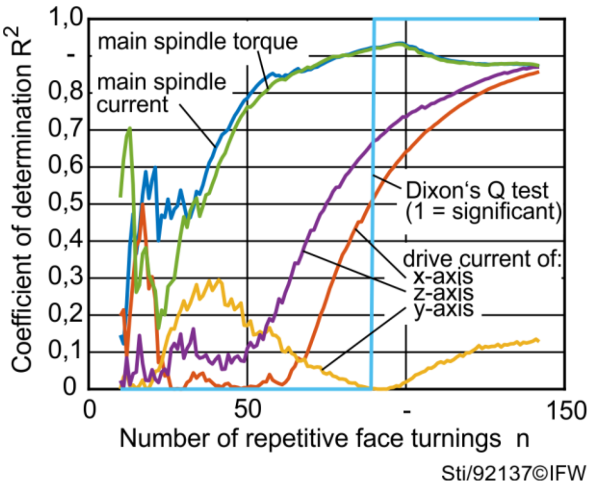

Fig. 7. Result of the trend evaluation for test set 2A (turning)

The method is assessed by evaluating its result after the last repetition of the sequence is completed. Table 3 summarizes the results under the use of Dixon's Q test as well as the threshold and majority based approach. If a test set contains a non-sensitive feature and it is detected, the assessment is correct. When all features in a test set are sensitive, the assessment is correct if no non-sensitive feature is detected. The assessment is incorrect in any other case. The confidence level α and the threshold R^2_{thres} were varied to gauge their influence on the result. Using Dixon's Q test, 18 out of 36 test sets were correctly evaluated. Using threshold-based decisions, 33 out of 36 test sets were correctly evaluated.

Dixon's method, for example, failed the test set 2B (series 2, feature subset B) with $\alpha = 0.05$. Three out of the four features in the test set are closely grouped, whereas the fourth is not. This causes the Dixon Q test to detect an outlier. However, all features in the test set are trending and eventually reach a value of $R^2 > 0.8$. The same mechanism results in failing test case 1D that contains three sensitive features, two of which hold virtually redundant information (torque and current of the main

Table 3. Results of validation. Feature assessment by the method matched ground truth (+), did not match ground truth (-)

Test case (series and feature set)	Irrelevant feature present?	Correctly evaluated with Dixon's Q test, $\alpha =$ 0.01	0.05	0.1	Thresholds, $R^2_{thres} =$ 0.3	0.4	0.5
1A	yes	-	-	+	+	+	+
1B	no	+	+	+	+	+	+
1C	yes	-	-	-	+	+	+
1D	no	+	-	-	+	+	+
1E	yes	-	-	-	+	+	+
2A	yes	+	+	+	+	+	+
2B	no	+	-	-	+	+	+
2C	yes	+	+	+	+	+	+
2D	no	+	+	+	+	+	+
2E	yes	+	+	+	+	+	+
3F	yes	-	-	-	-	+	+
4F	yes	-	-	-	+	-	-

spindle). Due to the high correlation of these two signals, the third signal is assessed as an outlier, passing the test case with $\alpha = 0.01$ only. While passing this test case, the threshold-based decision making failed test case 4B. This is due to $R^2 < 0.4$ for the mean of the torque of the feed axis. The feature is sensitive to wear, but the contained noise and seasonality account for the majority of the variance.

In 9 out of 12 test cases, a change in the confidence level did not cause the outcome of the Dixon's Q test to change. Results of the threshold-based decision-making remained the same in 10 out of 12 test cases despite varying the threshold.

5. Conclusion

Various feature selection methods exist, which mostly perform feature selection in hindsight of the monitoring task utilizing real failure data. However, a risk remains as even features found to be well correlating with the failure might be randomly disturbed. Additionally, common practice for parameterization of process monitoring systems in industry settings increases that risk.

To address this matter a method was described that assesses an initial selection of features online aiming to detect a single non-sensitive feature if present. For this, long term trends in repetitive machining operations are quantified. A linear function is fitted for each feature separately. The resulting statics are then either evaluated by the Dixon's Q outlier test or a threshold and majority based approach.

The method was characterized and validated using a total of four series of repetitive milling and drilling operations. Twelve different test sets with different feature combinations were created from the data and assessed manually. The performance of the method was evaluated by comparing its assessment to the manually determined ground truth. Results show that a single non-sensitive feature is detectable in an initial selection of otherwise sensitive features. The displayed threshold-based decision-making mechanism evaluated 33 out of 36 test cases correctly, failing to detect an irrelevant feature in all three cases. A sensitive feature was never falsely classified as irrelevant. This detection characteristic is suitable to complement existing monitoring systems as falsely excluding a sensitive feature is a highly unfavorable situation. The approach based on Dixon's Q test evaluated 18 out of 36 test sets correctly, performing inferior to the threshold-based approach.

While it was evaluated whether a test set was eventually correctly assessed or not, it was not determined when non-sensitive features were first detected. Additionally, in sporadic test sets, the assessment of the method altered multiple times during the course of the sequence.

The data used to validate the method is unambiguous, featuring only one non-sensitive feature per series. Thus, the performance of the method in more challenging data might substantially vary.

Future work will address the method's tendency towards false classification when high portions of variance result from noise or seasonality. For this, additional statistics of the regression might be used. Further, test sets with more diverse sensitive and non-sensitive features should be evaluated. The method should also be adapted to handle more than one non-sensitive feature.

Acknowledgment

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 313912117.

References

[1] Teti R, Jemielniak K, O'Donnell G, Dornfeld D. Advanced monitoring of machining operations. CIRP Annals – Manufacturing Technology 2010;59:717-739.

[2] Sick B. On-line and Indirect Tool Wear Monitoring in Turning with Artificial Neural Networks: A Review of More Than a Decade of Research. Mechanical Systems and Signal Processing 2002,16:487–546.

[3] Jantunen E. A summary of methods applied to tool condition monitoring in drilling. International Journal of Machine Tools & Manufacture 2002;42:997-1010.

[4] Denkena B, Dahlmann D, Damm J. Self-adjusting Process Monitoring System in Series Production. Procedia CIRP 2015;33:233-238.

[5] Jemielniak K, Urbański T, Kossakowska J, Bombiński S. Tool condition monitoring based on numerous signal features. International Journal of Advanced Manufacturing Technology 2012;59:73-81.

[6] Jemielniak K, Arrazola PJ. Application of AE and cutting force signals in tool condition monitoring in micro-milling. CIRP Journal of Manufacturing Science and Technology 2008;1:97-102.

[7] Scheffer C, Heyns PS. An industrial tool wear monitoring system for interrupted turning. Machanicla Systems and Signal Processing 2004;18:1219-1242.

[8] Da ZJ, Sadler JP, Jawahir IS. Predicting optimum cutting conditions for turning operations at varying tool-wear states. Transactions of the North American Manufacturing Research Institution of SME 1997;25:75-80.

[9] Scheffer C, Heyns PS. Wear monitoring in turning operations using vibration and strain measurements. Mechanical Systems and Signal Processing 2001;15:1182-1202.

[10] Jemielniak K. Tool Condition Monitoring in Metal Cutting. In: Grzegorzewski P, Kochanski A, Kacprzyk J, editors. Soft Modeling in Industrial Manufacturing. Studies in Systems, Decision and Control. Cham: Springer; 2019. p. 45-74.

[11] Abellan-Nebot JV, Subirón FR. A review of machining monitoring systems based on artificial intelligence process models. The International Journal of Advanced Manufacturing Technology 2010;47:237-257.

[12] Duangsoithong R, Windeatt T. Relevance and redundancy analysis for ensemble classifiers. In: Perner P, editor. Machine Learning and Data Mining in Pattern Recognition. Lecture Notes in Computer Science. Berlin: Springer; 2009. pp. 206-220.

[13] Pratama M, Dimla E, Tjahjowidodo T, Pedrycz W, Lughofer E. Online Tool Condition Monitoring Based on Parsimonious Ensemble+. IEEE Transactions on Cybernetics. 2018. Article in Press. DOI: 10.1109/TCYB.2018.2871120

[14] Chernick MR. A Note on the Robustness of Dixon's Ratio Test in Small Samples. The American Statistician 1982;36:140.

[15] Böhrer A. One-sided and Two-sided Critical Values for Dixon's Outlier Test for Sample Sizes up to n=30. Economic Quality Control 2008;23:5-13.