Detection of Machine Tool Anomalies from Bayesian Changepoint Recurrence Estimation

Christian Reich*†, Christina Nicolaou*†, Ahmad Mansour*, Kristof Van Laerhoven†

*Corporate Sector Research, Robert Bosch GmbH, Stuttgart, Germany.
Email: {christian.reich,ahmad.mansour,christina.nicolaou2}@de.bosch.com
†Dept. of Electrical Engineering and Computer Science, University of Siegen, Germany. Email: kvl@eti.uni-siegen.de

***Abstract*—In this study, we consider the problem of detecting process-related anomalies for machine tools. The similar shape of successive sensor signals, which arises due to the same process step sequence applied to each workpiece, suggests extracting shape-related features. In recent years, shapelets dominated the field of shape-related features. Unfortunately, they involve a high computational burden due to hyperparameter optimization.**

We introduce alternative shape-related features relying on abrupt signal changes (*changepoints*) reflecting the changes of process steps. During normal operation, changepoints follow a highly recurrent pattern, i.e., appear at similar locations. Thus, being able to distinguish regular, recurrent from abnormal, non-recurrent changepoints allows detecting process anomalies.

For changepoint recurrence estimation, we extend the Bayesian Online Changepoint Detection (BOCPD) method. The extension allows distinguishing normal and abnormal changepoints relying on empirical estimates of the changepoint recurrence distribution. Subsequently, changepoint-related features are introduced and compared to shapelets and wavelet-based features in a case study comprising real-world machine tool data.

Qualitative results verify changepoint locations being comparable to shapelet locations found by the FLAG shapelet approach. Furthermore, quantitative results suggest superior classification performance both to shapelets and wavelet-based features.

***Index Terms*—Bayesian methods, changepoint estimation, anomaly detection, machine monitoring**

I. Introduction

In this study, we focus on detection of process-related anomalies in machine tools. For this purpose, a vibration sensor is attached close to the chipping tool. Due to the same processing steps applied to each workpiece, successive signals recorded during the machining of each workpiece show a similar shape. This suggests the use of shape-related features for detecting process-related anomalies. In recent years, shapelets [1] established themselves among the most promising methodological directions. Shapelets are defined as most informative time series subsequences to represent discriminative differences of signal classes. Unfortunately, shapelet methods often involve optimizing hyperparameters in order to find the optimal length, position and smoothness of shapelets. This results in a high computational burden.

We introduce changepoint-related features as an alternative to shapelets for machine tool monitoring. Changepoints are defined as variations in the parameters of the generative signal model [2]. These deterministic features come without the necessity to optimize hyperparameters. In order to verify the competitiveness of changepoint features, we compare to shapelets and popular wavelet-based features in a case study with real-world machine tool data.

We study for which of these data sets changepoint locations bear discriminative information and summarize assumptions for data characteristics that allow for a changepoint-related detection of process anomalies. Finally, we outline the benefits of being able to distinguish normal recurrent changepoints (which arise due to the repetitive processing step sequence) from non-recurrent (i.e., abnormal) changepoints. In order to estimate recurrence of changepoints, we introduce an empirical *changepoint recurrence distribution (CPRD)*. Extracting the proposed features separately from recurrent and non-recurrent changepoints outperforms extracting them from standard BOCPD changepoints and results in a superior performance compared to shapelets and wavelet features.

II. Related Work

A. Changepoint Estimation

Changepoints can be estimated via piecewise linear approximation [3], clustering models [4], [5], Hidden Markov Models [6], [7] or methods involving penalized likelihood functions [8]–[10]. The methods in this work built on BOCPD [2].

In [11], Maslov et al. proposed an approach to model recurrence of BOCPD changepoints. Recurrence was defined by quasi-periodicity, i.e., by assuming periodic changepoints while allowing small deviations of individual changepoints from this periodic behavior. This method does not allow modeling generic recurrences. For example, changepoints related to machine tool processing steps appear in a recurrent manner but not periodically (i.e., with similar inter-changepoint distances).

Wilson et al. proposed hierarchical BOCPD extensions [12]. Although this approach allows inferring adaptive estimates of the typical frequency (hazard rate) of changepoints, it does not allow to model a recurrent changepoint prior distribution or distinguish recurrent from non-recurrent changepoints.

B. Signal Form Sensitive Features: Shapelets

In [1], Ye et al. proposed shapelets for time series feature learning. Shapelets are sensitive to ordering of time series samples and thus signal form, which allows using them as locally expressive primitives for anomaly detection.

978-1-7281-2927-3/19/$31.00 ©2019 IEEE

Whereas [1] found shapelets parallel to constructing a decision tree, authors in [13] extended the approach to off-the-shelf classifiers by suggesting distance measures between shapelets and signals to produce a transformed data set. As neither length nor position of most informative signal subsequences can be constrained a priori in shapelet selection, finding shapelets resulted in a combinatorial problem. In [14], authors proposed learning shapelets jointly with a linear Support Vector Machine (SVM) rather than finding most discriminative subsequences. This resulted in a highly nonlinear optimization problem, which involved a similar computational effort as [1].

In addition to optimizing length and position of shapelets, learning shapelets in [14] involves optimizing shapelet smoothness. While overly smooth shapelets suppress shape details, non-smooth shapelets yield overfitting to shape details of training signals and thus bad generalization to future test data.

C. Frequency-related Features: Wavelets

In [15], Teti et al. gave an overview over common machine monitoring features. For rotating machinery, wavelet features were listed among most common choices. We extract features mean, variance and peak-to-peak-distance in wavelet subbands for a comparison with changepoint-related features.

III. Theoretical Background

In recent literature, BOCPD manifested among the most powerful changepoint estimators [2]. BOCPD assumes that a predictive distribution $p\big(x_{t+1}|\mathbf{x}_{1:t}\big)$ at time step t can be computed from observations $\mathbf{x}_{1:t}$ (i.e., measurement data) and a latent *run length* variable r_t [2]. This run length r_t is defined as the distance to the last changepoint having occurred in the data. By integrating over the posterior distribution $p\big(r_t|\mathbf{x}_{1:t}\big)$ on the run length r_t, this marginal predictive distribution can be calculated by

$$p\big(x_{t+1}|\mathbf{x}_{1:t}\big) = \sum_{r_t} p\big(x_{t+1}|r_t, \mathbf{x}_t^{(r)}\big) p\big(r_t|\mathbf{x}_{1:t}\big) \tag{1}$$

Here, $\mathbf{x}_t^{(r)}$ are observations associated with the current run r_t, i.e., the last r_t observations of $\mathbf{x}_{1:t}$ [16]. In changepoint detection, however, the focus of interest is not on predicting future observations x_{t+1}, but in finding estimates of the current run length r_t via the conditional posterior distribution

$$p\big(r_t|\mathbf{x}_{1:t}\big) = \frac{p\big(r_t, \mathbf{x}_{1:t}\big)}{p\big(\mathbf{x}_{1:t}\big)}. \tag{2}$$

Henceforth, $p(r_t|\mathbf{x}_{1:t})$ is referred to as *run length distribution*. As its probability mass is highly concentrated at a few peaks, pruning of run lengths with a probability below a threshold (e.g., $\epsilon = 10^{-4}$) can be applied. This reduces run time from $\mathcal{O}\big(T^2\big)$ to $\mathcal{O}\big(T/\epsilon\big)$ as outlined in [16].

The distribution $p\big(r_t, \mathbf{x}_{1:t}\big)$ can be found recursively [2]:

$$p\big(r_t, \mathbf{x}_{1:t}\big) = \sum_{r_{t-1}} p\big(r_t|r_{t-1}\big) p\big(x_t|r_{t-1}, \mathbf{x}_t^{(r)}\big) p\big(r_{t-1}, \mathbf{x}_{1:t-1}\big) \tag{3}$$

The right-hand side of Eq. 3 consists of three terms:

1) The predictive distribution $p\big(x_t|r_{t-1}, \mathbf{x}_{1:t}\big)$ collapses to $p\big(x_t|r_{t-1}, \mathbf{x}_t^{(r)}\big)$, thus depending only on recent $\mathbf{x_t^{(r)}}$.
2) A joint distribution $p\big(r_{t-1}, \mathbf{x}_{1:t-1}\big)$ from time step $t-1$.
3) A conditional prior distribution $p\big(r_t|r_{t-1}\big)$ on changepoints (i.e., $r_t = 0$). Adams et al. proposed to define it as follows for efficient computation (nonzero probability mass only for outcomes $r_t = 0$ and $r_t = r_{t-1} + 1$) [2]:

$$p\big(r_t|r_{t-1}\big) = \begin{cases} H\big(r_{t-1}+1\big) & \text{if } r_t = 0 \\ 1 - H\big(r_{t-1}+1\big) & \text{if } r_t = r_{t-1}+1 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

In the simplest case, the *hazard function* $H(\tau)$ is assumed constant $H(\tau) = 1/\lambda$, resulting in changepoint estimates $p\big(r_t = 0|r_{t-1}\big)$ independent of r_{t-1} [2]. The constant scale parameter λ has to be defined in advance or treated as a further model hyperparameter which has to be optimized [16], [17].

IV. Proposed Methods

A. Changepoint Recurrence Distribution (CPRD)

Due to a typical concentration of probability mass of $p\big(r_t|\mathbf{x}_{1:t}\big)$ at a dominant peak, the most probable run length estimate $\hat{r_t}$ can be approximated sensibly at the maximum a posteriori (MAP) estimate of the run length distribution, i.e.,

$$\hat{r_t} = \arg\max_{r_t} p\big(r_t|\mathbf{x}_{1:t}\big) \tag{5}$$

According to [2], changepoints can be assigned at $\hat{r_t} = 0$. However, for machine tool data with potentially smooth transitions between signal segments, changepoints at these segment borders do not necessarily lead to $\hat{r_t} = 0$, but to a major drop in this most probable run length estimate $\hat{r_t}$. Drops in $\hat{r_t}$ (i.e., where $\hat{r_t}$ does not increase by one) can then be interpreted as changepoints with a non-zero changepoint probability

$$p\big(c_t|\mathbf{x}_{1:t}\big) \triangleq p\big(\hat{r_t}|\mathbf{x}_{1:t}\big)\Big|_{\frac{\partial \hat{r_t}}{\partial t} \neq 1} \tag{6}$$

Changepoints c_t occur not necessarily due to recurrent changes of process steps, but can also be due to signal fluctuations or anomalies. This motivates the necessity to filter recurrent changepoints from the set of all changepoints. In order to achieve this filtering, we propose the following approach.

Changepoint probability vectors $p\big(c_t^{(n)}|\mathbf{x}_{1:t}\big)$ of N training signals are summed up across time steps $t = 1 \dots T$ (Fig. 1, bottom). For each training signal $n = 1 \dots N$, the cumulative probability mass $\sum_{n=1}^{N} p\big(c_t^{(n)}|\mathbf{x}_{1:t}\big)$ increases at locations t of changepoints $c_t^{(n)}$ (i.e., locations t with non-zero probabilities $p\big(c_t^{(n)}|\mathbf{x}_{1:t}\big)$) while staying the same at other time steps t where $p\big(c_t^{(n)}|\mathbf{x}_{1:t}\big) = 0$. Normalizing $\sum_{n=1}^{N} p\big(c_t^{(n)}|\mathbf{x}_{1:t}\big)$ allows interpretation as an empirical probability distribution over recurrence of changepoint positions [18]. We name this distribution *changepoint recurrence distribution (CPRD)*.

$$p\big(c_t^{(1:N)}|c_t^{(n)}, \mathbf{x}_{1:t}\big) \triangleq \frac{\sum_{n=1}^{N} p\big(c_t^{(n)}|\mathbf{x}_{1:t}\big)}{\sum_{n=1}^{N}\sum_{t=1}^{T} p\big(c_t^{(n)}|\mathbf{x}_{1:t}\big)} \tag{7}$$

Recurrence of changepoints $c_t^{(n)}$ at locations t across signals $n = 1 \dots N$ is denoted by the term $c_t^{(1:N)}$. $p\big(c_t^{(1:N)}|c_t^{(n)}, \mathbf{x}_{1:t}\big)$ thus gives an empirical estimate how likely changepoints $c_t^{(n)}$ at locations t were present in all former N signals. This approach thus yields a nonparametric maximum likelihood estimate of recurrent changepoint probabilities [19].

The CPRD can be used to replace the uninformative hazard function $H(\tau) = 1/\lambda$. This allows incorporating empirical information about the recurrence of changepoints directly into the changepoint prior $p\big(r_t|r_{t-1}\big)$. Instead, an alternative approach by estimating all changepoints via BOCPD and using the CPRD to subsequently filter recurrent changepoints from the set of all changepoints is chosen: By multiplying initial BOCPD changepoint estimates $p\big(c_t^{(n)}|\mathbf{x}_{1:t}\big)$ with the empirical CPRD probabilities $p\big(c_t^{(1:N)}|c_t^{(n)}, \mathbf{x}_{1:t}\big)$, a filtering of changepoint estimates regarding their probability of being recurrent is obtained. In the Bayesian framework, this can be interpreted as applying Bayes' Theorem:

$$p\big(c_t^{(n)}|c_t^{(1:N)}, \mathbf{x}_{1:t}\big) = \frac{p\big(c_t^{(1:N)}|c_t^{(n)}, \mathbf{x}_{1:t}\big)p\big(c_t^{(n)}|\mathbf{x}_{1:t}\big)}{p\big(c_t^{(1:N)}|\mathbf{x}_{1:t}\big)} \quad (8)$$

As discussed earlier, the CPRD $p\big(c_t^{(1:N)}|c_t^{(n)}, \mathbf{x}_{1:t}\big)$ acts as a nonparametric estimate of the likelihood of changepoint recurrence. The initial BOCPD changepoint probabilities $p\big(c_t^{(n)}|\mathbf{x}_{1:t}\big)$ for signal n are interpreted as prior estimates of recurrent changepoints. As the goal of the presented approach is finding non-zero probabilities of $p\big(c_t^{(n)}|c_t^{(1:N)}, \mathbf{x}_{1:t}\big)$, normalization to the prior distribution on changepoint recurrence $p\big(c_t^{(1:N)}|\mathbf{x}_{1:t}\big)$ does not have to be considered:

$$p\big(c_t^{(n)}|c_t^{(1:N)}, \mathbf{x}_{1:t}\big) \propto p\big(c_t^{(1:N)}|c_t^{(n)}, \mathbf{x}_{1:t}\big)p\big(c_t^{(n)}|\mathbf{x}_{1:t}\big) \quad (9)$$

Non-zero probabilities $p\big(c_t^{(n)}|c_t^{(1:N)}, \mathbf{x}_{1:t}\big)$ indicate recurrent changepoints. Non-recurrent changepoints are then found as symmetric set difference between BOCPD changepoints $p\big(c_t^{(n)}|\mathbf{x}_{1:t}\big)$ and recurrent changepoints. Thus, subsequent filtering of BOCPD changepoints outputs both recurrent and non-recurrent changepoints. Instead, using the CPRD as hazard function suppresses non-recurrent changepoints. The CPRD hazard function only yields more robust estimates of recurrent changepoints. Additional to predicting future recurrent changepoints, this allows extracting features in comparable segments of each signal, which proved beneficial in tool condition monitoring [20].

B. Run Length Primitives

Due to the challenges involved in finding optimal position, length and smoothness of shapelets, we propose alternative features relying on the deterministic information given by changepoint locations and distances. Both locations and distances of changepoints are encoded in MAP run length estimates $\hat{r_t}$ (Fig. 1, middle, red line), that come as a byproduct of the BOCPD algorithm. In the following, we will refer to them as *run length primitives*. Run length primitives are expressive of the data-generating process (e.g., processing step sequence) in machine tool applications in that they are influenced by positions and distances of changepoints but not by signal variations between them. They can thus be interpreted as features incorporating the information of signal event locations (i.e., regular changes of processing steps and anomalies) while sequences between these events are approximated by non-informative linearly increasing slopes. This yields an automatic, deterministic solution for the trade-off mentioned above regarding choice of smoothness, positions and lengths of representative signal primitives.

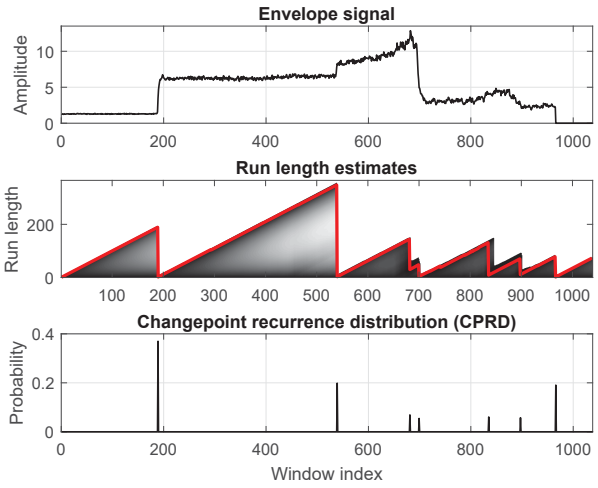


Fig. 1: CPRD estimation. Top: Envelope signal of data set 1 (DS1). Middle: BOCPD run length log probabilities $\log\big(p\big(r_t|\mathbf{x}_{1:t}\big)\big)$ are depicted as gray values, MAP run length estimates $\hat{r_t}$ as red line. Bottom: Changepoint probabilities $p\big(c_t^n|\mathbf{x}_{1:t}\big)$ for the single signal envelope from top figure. Elements of $p\big(c_t^n|\mathbf{x}_{1:t}\big)$ for all N training signals are accumulated across window indices t and normalized in a final step.

V. Case Study: Machine Tool Monitoring

A. Motivating Changepoint-Related Anomaly Detection

When a sensor is placed sufficiently close to the cutting tool, similar segments can be observed in successive signals related to the same sequence of processing steps applied to each workpiece. Changes in processing steps reflect in additional non-recurrent changepoints or missing recurrent changepoints and thus allow for detection of process-related anomalies.

B. Sensor Specifications and Data Sets

The data sets of this case study were recorded at different cutting machine tools using microelectromechanical system (MEMS) vibration sensors. The vibration sensors have a single degree of freedom and sample at a rate of 62.5 kHz.

Data sets comprise sensor signals for six machine tool monitoring applications (cf. Table I). Data set 1 (DS1) contains signals for different degrees of severity of a grinding wheel anomaly. Data set 2 (DS2) comprises signals containing abnormal machine part contacts. Data set 3 (DS3) contains

data for grinding of a workpiece with complex structure, where a gradual decrease in workpiece quality occurs when re-sharpening of the grinding wheel is not applied timely. Thus, DS3 data allows finding out whether changepoint shifts coincide with continuous drifts in signal shape reflecting the decrease of workpiece quality. Data set 4 (DS4) contains data for a turning machine. This data set is used to find out whether breakage of the turning tool can be related to changepoint patterns. Data set 5 and 6 anomalies occurred due to imprecise positioning of the workpieces on the workpiece rest and resulting damages in the grinding wheel.

TABLE I: Data sets and parameter variations.

Data set	Records	Normal	Abnormal	Machine type
DS1	430	312	118	Grinding
DS2	499	400	99	Grinding
DS3	91	59	32	Grinding
DS4	69	66	3	Turning
DS5	3301	3008	293	Grinding
DS6	3692	3670	22	Grinding

C. Experimental Setup

As processing step information is encoded mostly in energy of the signal envelopes, signal envelope representations are computed via $\frac{1}{M}\sum_M|\mathbf{x}_t|$ in each successive signal window of size $M = 1024$ observations $\mathbf{x}_t$. These signal envelope representations are then used as input for the BOCPD algorithm.

In order to use run length primitives with off-the-shelf classifiers, we construct Euclidean similarity measures between shapelets and signals like introduced in [13]. This allows comparison with shapelet approaches like Fused Lasso Generalized Eigenvector (FLAG) [21]. FLAG is chosen as it yields locally sparse, blocky shapelets by regularizing with a fused lasso [22]. The positions of these shapelets are marked by the FLAG indicator vector. To study the relevance of changepoint information in machine tool applications qualitatively, we compare changepoint locations to the FLAG indicator vector.

In accordance with FLAG, we classify with a linear SVM trained with the LIBSVM framework [23]. FLAG hyperparameters α_1 and α_2 are set by cross-validation.

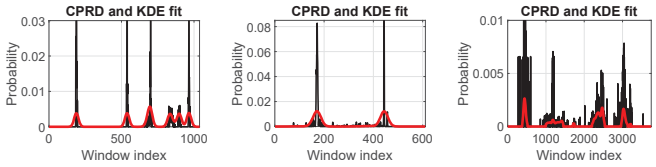
D. Time Complexity

According to [21], FLAG training complexity is $\mathcal{O}\big(T^3 + MT^2\big)$. Here, T is the length of time series and M the number of iterations of the ADMM solver. For optimization of FLAG hyperparameters α_1 and α_2, grid search cross-validation results in a factor $|\alpha_1 \otimes \alpha_2|$, where $\otimes$ is the Kronecker operator and $|\cdot|$ the cardinality operator. Overall training complexity including grid search hyperparameter optimization thus yields a complexity $\mathcal{O}\big(|\alpha_1 \otimes \alpha_2|\big(T^3 + MT^2\big)\big)$ cubic in T.

Training complexity for our changepoint-related approach can be divided into $\mathcal{O}\big(T/\epsilon\big)$ for run length estimation (cf. section III) and $\mathcal{O}\big(\big(N+1\big)T\big)$ both for computation of the CPRD and run length primitives. In summary, this results in $\mathcal{O}\big(\big(\frac{1}{\epsilon}+N+1\big)T\big)$, i.e., linear training complexity.

E. Results

1) CPRD Estimation: CPRDs for DS1 to DS3 (black lines) are depicted in Fig. 2. The broad distribution of probability mass for DS3 results from the non-stationarity, i.e. less recurrent normal changepoint pattern (cf. Fig. 3c). For the other data sets with stationary normal changepoints, CPRD probability mass is concentrated at indices of normal changepoints.



(a) DS1 (b) DS2 (c) DS3

Fig. 2: CPRDs (black lines) and KDE fits (red lines).

Each subfigure additionally shows a Gaussian kernel density estimator (KDE) (red lines) fitted to the CPRDs. KDE post-processing yields similar smoothing effects like estimating the CPRD with a larger number of changepoints and thus a more robust estimator of changepoint recurrence, although individual changepoints of normal signals are marked abnormal (cf. Fig.s 3a and 3b). Changepoint filtering results in the following section were obtained with KDE fits, not the original CPRDs.

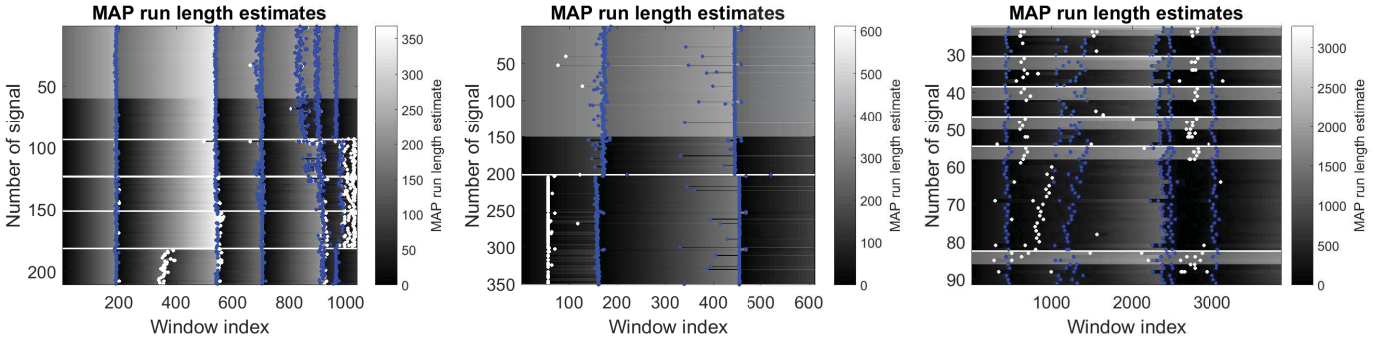
2) Filtering BOCPD Changepoints via CPRD: Fig. 3a depicts changepoints for DS1. Recurrent changepoints likely under the CPRD KDE fit (Fig. 2a, red line) are shown as blue dots, non-recurrent changepoint estimates as white dots.

MAP run length estimates $\hat{r}_t$ are plotted gray-coded in horizontal direction for all signals. Thus, each row depicts the bird's-eye view of the MAP run length estimate for one envelope signal like depicted in the middle plot of Fig. 1.

DS1 normal data consists of signals nr. 1 to 91. Signals nr. 1 to 60 (white overlay) are used for CPRD estimation (Fig. 2a, black line) and subsequent KDE fitting (Fig. 2a, red line). Below, run length estimates for different degrees of grinding wheel anomalies (separated by white lines) are plotted. For all degrees of grinding wheel anomalies, changepoints between indices 700 and 970 vanish. Depending on the degrees of the anomaly, changepoints from index 970 onward vanish or abnormal changepoints between indices 300 and 400 occur. These changepoints persist during abnormal behavior.

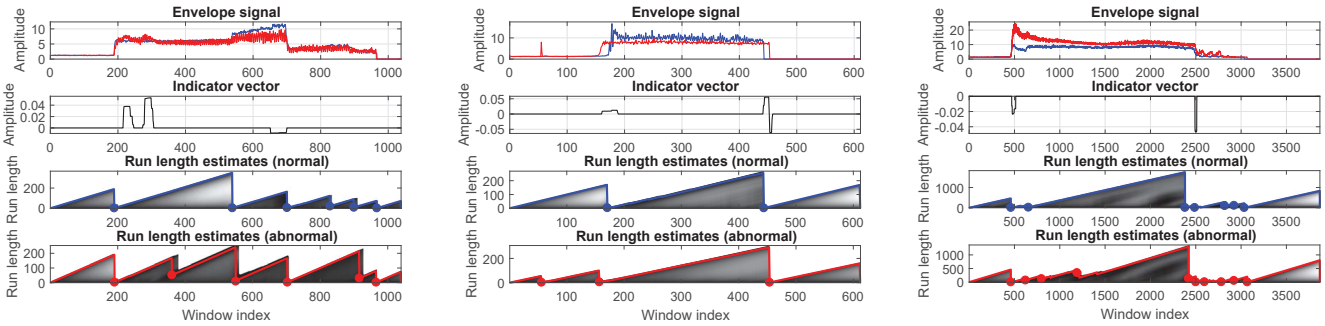
DS2 shows a similarly discrete change between changepoint patterns for normal data (workpiece nr. 1 to 201) and abnormal data, where an additional changepoint occurs at index 55 due to impulse-like artifacts in each abnormal signal (cf. Fig. 3b).

For DS3, each four signals after re-sharpening the grinding wheel (white lines) were assumed normal and used for CPRD estimation. This assumption was made as a decrease in workpiece quality appeared after four workpieces according to domain experts. Other than for DS1 and DS2, only normal changepoints around indices 500, 2500 and 3000 expose a stationary, recurrent pattern (cf. Fig. 3c). This results in a broadly distributed CPRD probability mass (cf. section V-E1) and the normal changepoints not assembled around indices



(a) DS1 (b) DS2 (c) DS3

Fig. 3: MAP run length estimates $\hat{r_t}$ (gray values), recurrent changepoints (blue dots) and non-recurrent changepoints (white dots) filtered from initial BOCPD changepoint estimates. White overlays mark training examples selected for CPRD calculation.



(a) DS1 (b) DS2 (c) DS3

Fig. 4: Run length primitives for DS1 to DS3: Comparison of standard BOCPD changepoint locations and FLAG shapelet locations. Figures 4a to 4c show signal envelope examples for classes normal (blue) and abnormal (red) (first subfigure), FLAG indicator vectors (second subfigure) and run length log probabilities $\log(p(r_t|\mathbf{x}_{1:t}))$ (gray values) with MAP estimates $\hat{r_t}$ for normal data (third subfigure, blue line) and abnormal data (fourth subfigure, red line). Changepoints are depicted as dots.

500, 2500 and 3000 being labeled abnormal. Thus, assuming normal changepoints to be recurrent is not justified for DS3.

3) Qualitative Evaluation: Figure 4a illustrates locations of FLAG shapelets and BOCPD changepoints for normal (blue) and abnormal (red) signal envelopes of DS1. The FLAG indicator vector assigns a shapelet at window index 300 (Fig. 4a, second plot). The corresponding changepoint for the abnormal grinding wheel is assigned at index 350 (bottom plot). The other shapelet locations for DS1 at indices 200 and 700 (Fig. 4a) are found at changepoint locations.

For DS2, anomalies result in an additional changepoint at index 55. This location is not found discriminative by FLAG, while other changepoint locations coincide with regions of interest illustrated by the FLAG indicator vector in Fig. 4b.

Figure 4c depicts shapelet and changepoint locations for DS3. The FLAG indicator vector assigns shapelets at two changepoint locations, others are deemed non-discriminative.

These results confirm the assumptions stated in the CPRD results section V-E1 that changepoint-related features are only useful for data with discrete, persisting shifts in the changepoint pattern during anomalies (DS1 and DS2, not DS3).

4) Quantitative Evaluation: Table II summarizes quantitative results for classification with FLAG shapelets and run length primitives (RLPs). Each data set DS1 to DS6 is splitted into half to create training and test sets. F1 scores of FLAG (row 2) are compared to run length primitives (RLPs) (rows 3-6) and wavelet-based features (rows 7-10). We use db4 wavelet filters and decompose into 8 sub-bands. For each data set, highest F1 scores are highlighted in blue, lowest in red.

TABLE II: Quantitative comparison of FLAG, RLPs and db4 wavelet features via F1 scores for test data (in %).

	DS1	DS2	DS3	DS4	DS5	DS6	**Avg.**
FLAG	95.5	97.4	66.2	46.1	83.7	96.5	**80.9**
RLP (BOCPD)	87.6	92.4	41.2	48.7	99.3	89.1	**76.4**
RLP (recurrent)	84.7	97.0	65.2	20.7	97.8	86.3	**75.3**
RLP (non-rec.)	72.3	39.8	69.8	45.2	47.6	49.8	**54.1**
RLP (both)	84.7	92.5	71.6	100.0	99.1	94.0	**90.3**
Wav_mean	70.9	38.3	57.3	36.1	47.8	49.8	**50.0**
Wav_var	95.2	99.3	43.3	48.9	99.6	95.8	**80.4**
Wav_P2P	62.8	94.3	35.5	46.2	66.1	95.8	**66.8**
All Wav	97.1	99.3	40.0	39.5	98.4	95.8	**78.4**

Classifying with wavelet variance features (row 8) or all

wavelet features combined (last row) yields a powerful baseline for data sets DS1, DS2, DS5 and DS6 and comparable F1 scores to FLAG (row 2). BOCPD run length primitives (row 3) show a performance inferior to FLAG. Run length primitives created from recurrent changepoints (row 4) yield comparable results to BOCPD. Run length primitives for non-recurrent changepoints (row 5) result in poor F1 scores for all data sets except DS1, where anomalies come with a persistent abnormal changepoint pattern (cf. Fig. 4a). Constructing run length primitives from both recurrent and non-recurrent changepoints separately (row 6) yields the best results, suggesting that both missing recurrent changepoints and additional non-recurrent changepoints hold valuable information for anomaly detection.

In summary, anomaly detection with run length primitives performs competitive under following assumptions:

1) Stationarity of changepoint patterns during normal states (as we empirically approximate the recurrent normal changepoint behavior with a static estimator).
2) Stationary abnormal changepoints: Persistent shift between normal and abnormal changepoint pattern.
3) Evolving changepoint patterns that are observable throughout all anomalous signals.

As observable in Fig. 3a and Fig. 3b, the first two assumptions are met for DS1 and DS2 but not for DS3. For DS2, not all signals labeled abnormal (from signal index nr. 202 on) have an additional, abnormal changepoint (cf. Fig. 3b). Thus, the third assumption is not met for DS2, which results in poor F1 scores for non-recurrent run length primitives (cf. Table II).

VI. Conclusions

The goal of this research is to examine the relevance of changepoints for machine tool anomaly detection tasks. Changepoints are estimated with the BOCPD algorithm. Information about changepoint locations enables definition of run length primitives, which are introduced here as an alternative to shapelets with lower computational effort. Their competitiveness is verified qualitatively and quantitatively by comparison with the FLAG shapelet approach [21].

An extension of the BOCPD algorithm is introduced to distinguish recurrent changepoints (evolving due to regular, repetitive processing steps) from non-recurrent (i.e., abnormal) changepoints. Recurrent changepoints allow for a description of the normal process behavior when behaving stationary. Additionally, they can be used for identification of repetitive signal segments and extraction of comparable feature scores [20]. Non-recurrent changepoints yield additional information about process anomalies when behaving similarly stationary.

We verify the relevance of changepoint information in machine tool anomaly detection in a case study involving six real-world data sets. Qualitative results reveal an overlap between FLAG shapelet locations and changepoint locations for data sets with stationary changepoint patterns (e.g., DS1 and DS2). Quantitative results verify superior classification quality (F1 scores) for run length primitives compared to FLAG shapelets and wavelet features when extracted separately from recurrent and non-recurrent changepoints. This confirms the benefit of changepoint recurrence estimation via CPRD.

References

[1] L. Ye and E. Keogh, "Time series shapelets: A new primitive for data mining," in *Proc. of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '09*, Paris, France, Jun. 2009, pp. 947–956.

[2] R. P. Adams and D. J. C. MacKay, "Bayesian online changepoint detection," University of Cambridge, Cambridge, UK, Tech. Rep., Oct. 2007.

[3] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm for segmenting time series," in *Proc. of the International Conference on Data Mining ICDM '01*, San Jose, USA, Nov. 2001, pp. 289–296.

[4] T. W. Liao, "Clustering of time series data - a survey," *Pattern Recognition*, vol. 38, pp. 1857–1874, Nov. 2005.

[5] A. Samé, F. Chamroukhi, G. Govaert, and P. Aknin, "Model-based clustering and segmentation of time series with changes in regime," *Advances in Data Analysis and Classification*, vol. 5, pp. 301–321, Dec. 2013.

[6] T.-C. Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence*, vol. 24, pp. 164–181, Feb. 2011.

[7] S.-Z. Yu, "Hidden semi-markov models," *Artificial Intelligence*, vol. 174, pp. 215–243, Feb. 2010.

[8] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, Jun. 1954.

[9] S. Liu, M. Yamada, N. Collier, and M. Sugiyama, "Change-point detection in time-series data by relative density-ratio estimation," *Neural Networks*, vol. 43, pp. 72–83, Jul. 2013.

[10] P. Fearnhead and G. Rigaill, "Changepoint detection in the presence of outliers," *Journal of the American Statistical Association*, vol. 113, pp. 1–15, Jun. 2018.

[11] A. Maslov, M. Pechenizkiy, Y. Pel, I. Žliobaitė, A. Shklyaev, T. Karkkäinen, and J. Hollmén, "Blpa: Bayesian learn-predict-adjust method for online detection of recurrent changepoints," in *Neural Networks (IJCNN), 2017 International Joint Conference on*, Anchorage, USA, May 2017, pp. 1916–1923.

[12] R. C. Wilson, M. R. Nassar, and J. I. Gold, "Bayesian online learning of the hazard rate in change-point problems," *Neural Computation*, vol. 22, pp. 2452–2476, Sep. 2010.

[13] J. Hills, J. Lines, E. Baranauskas, J. Mapp, and A. Bagnall, "Classification of time series by shapelet transformation," *Data Mining and Knowledge Discovery*, vol. 28, pp. 851–881, Jul. 2014.

[14] J. Grabocka, N. Schilling, M. Wistuba, and L. Schmidt-Thieme, "Learning time-series shapelets," in *Proc. of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '14*, New York, USA, Aug. 2014, pp. 392–401.

[15] R. Teti, K. Jemielniak, G. O'Donnell, and D. Dornfeld, "Advanced monitoring of machining operations," *CIRP Annals-Manufacturing Technology*, vol. 59, pp. 717–739, Aug. 2010.

[16] Y. Saatçi, R. D. Turner, and C. E. Rasmussen, "Gaussian process change point models," in *Proc. of the 27th International Conference on Machine Learning (ICML'10)*, Haifa, Israel, Jun. 2010, pp. 927–934.

[17] R. Turner, Y. Saatçi, and C. E. Rasmussen, "Adaptive sequential bayesian change point detection," University of Cambridge, Cambridge, UK, Tech. Rep., May 2009.

[18] Z. Ghahramani. (2012) Bayesian modelling. [Online]. Available: http://mlg.eng.cam.ac.uk/zoubin/talks/lect1bayes.pdf

[19] A. B. Owen, *Empirical Likelihood*. London, UK: Chapman and Hall, 2001.

[20] C. Reich, A. Mansour, and K. V. Laerhoven, "Embedding intelligent features for vibration-based machine condition monitoring," in *Proc. of European Signal Processing Conference (EUSIPCO'18)*, Rome, Italy, Sep. 2018, pp. 371–375.

[21] L. Hou, J. T. Kwok, and J. M. Zurada, "Efficient learning of timeseries shapelets," in *Proc. of the 30th AAAI Conference on Artificial Intelligence AAAI '16*, Phoenix, Arizona, Feb. 2016, pp. 1209–1215.

[22] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society (Series B)*, vol. 67, pp. 91–108, Feb. 2005.

[23] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, Apr. 2011.