Bayesian Estimation of Recurrent Changepoints for Signal Segmentation and Anomaly Detection

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
†Department of Electrical Engineering and Computer Science, University of Siegen, Germany. Email: kvl@eti.uni-siegen.de

***Abstract*—Signal segmentation is a generic task in many time series applications. We propose approaching it via Bayesian changepoint algorithms, i.e., by assigning segments between changepoints. When successive signals show a recurrent changepoint pattern, estimating changepoint recurrence is beneficial for two reasons: While recurrent changepoints yield more robust signal segment estimates, non-recurrent changepoints bear valuable information for unsupervised anomaly detection.**

This study introduces the *changepoint recurrence distribution (CPRD)* as an empirical estimate of the recurrent behavior of observed changepoints. Two generic methods for incorporating the estimated CPRD into the process of assessing recurrence of future changepoints are suggested. The knowledge of non-recurrent changepoints arising from one of these methods allows additional unsupervised anomaly detection.

The quality both of changepoint recurrence estimation via CPRD and of changepoint-related signal segmentation and unsupervised anomaly detection are verified in a proof-of-concept study for two exemplary machine tool monitoring tasks.

***Index Terms*—Bayesian methods, online learning, signal segmentation, anomaly detection**

I. Introduction

For many time series applications, both finding anomalies and a sensible segmentation of signals can be interpreted as two sides of one coin when signals show a recurrent structure. For example, electrocardiogram (ECG) signals behave highly recurrent for normal cardiac behavior and thus come with repetitive signal segments. Cardiac anomalies can then be found both by extraction of suitable features in recurrent signal segments and by abnormalities in this recurrent behavior itself, i.e., changes of the signal segmentation structure [1].

Another application field with an interest in estimating recurrent signal behavior is machine tool monitoring. When a sensor is placed sufficiently close to the cutting tool, similar signal segments can be observed for each processed workpiece. This is due to the same sequence of processing steps applied to each workpiece. Again, tracking deviations from this recurrent behavior allows for detection of (sudden) abnormal process alterations, while extracting features in comparable, recurrent signal segments allows for detection of drifting anomalies (cf., e.g., [2] for tool condition monitoring applications).

In this study, we assign signal segments between *changepoints*. Changepoints are defined as variations in the generative, statistical parameters of signal models [3]. Generic changepoint estimation can be approached with the Bayesian Online Changepoint Detection (BOCPD) algorithm [3].

We extend the BOCPD approach by introducing a *changepoint recurrence distribution (CPRD)*. The CPRD allows estimating recurrent behavior of observed changepoints and can then be used to improve robustness of signal segmentation by assigning segments between recurrent changepoints only. Additionally, we show that non-recurrent changepoints yield expressive features for an unsupervised anomaly detection.

The approach is illustrated for two real-world machine tool data sets in a proof-of-concept study on changepoint-related signal segmentation and unsupervised anomaly detection.

II. Related Work

Popular signal segmentation approaches comprise piecewise linear approximation methods [4], cluster-based methods [5], [6], Hidden Markov Models [7]–[9] and algorithms involving a penalized likelihood function of the data [10]–[12].

The BOCPD algorithm introduced in [3] allows dividing signals into non-overlapping segments of stationary generative data distributions between changepoints. Different work extending BOCPD to model data-generating distributions more flexibly [13], [14] or to use changepoint information for the sake of robust time series predictions [15] emerged quickly.

In [16], an approach explicitly dedicated to modeling recurrence of changepoints was proposed. Here, recurrence was defined by quasi-periodicity, i.e., by assuming periodic recurrence of changepoints while allowing small deviations of individual changepoints from this periodic behavior. This assumption does not allow to model a generic recurrent (but non-periodic) structure of the data.

Wilson et al. proposed a hierarchical extension of the BOCPD approach in [17]. Although this approach allows inferring an adaptive estimate 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 as desired in this work.

III. Theoretical Background

A. Bayesian Online Changepoint Detection (BOCPD)

BOCPD assumes that a predictive distribution $p(x_{t+1}|\mathbf{x}_{1:t})$ 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 [3]. r_t is

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

defined as the distance to the last changepoint having occurred in the data. We obtain this predictive distribution by integrating over the run length posterior distribution $p(r_t|\mathbf{x}_{1:t})$:

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

Here, $\mathbf{x}_t^{(r)}$ is the set of observations associated with the current run r_t, i.e., the last r_t observations of $\mathbf{x}_{1:t}$ [13]. In changepoint detection, the focus of interest is not on predicting the most probable future observation x_{t+1}, but in finding the most probable estimate of the current run length r_t via the conditional posterior distribution

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

Henceforth, this conditional posterior distribution is referred to as *run length distribution*. As probability mass of the run length distribution 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 $\mathcal{O}(t^2)$ to $\mathcal{O}(t/\epsilon)$ as outlined in [13].

The distribution $p(r_t, \mathbf{x}_{1:t})$ can be found recursively [3]:

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

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

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

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

The function $H(\tau)$ is named *hazard function* [18]. In the simplest case, an uninformative constant hazard function $H(\tau) = 1/\lambda$ can be chosen as discussed in [3]. This results in making changepoint estimates $p(r_t = 0|r_{t-1})$ independent of r_{t-1}. Here, λ is a constant scale parameter which has to be defined in advance or can be treated as a further model hyperparameter which has to be optimized [13], [14].

For sensor data it is common to assume independent and identically distributed (iid) normal observations x_t and a Normal-Inverse-Gamma parameter prior $p(\mu, \sigma^2|\mu_0, \kappa, \alpha, \beta)$:

$$x_t \sim \mathcal{N}(\mu, \sigma^2), \quad (5)$$
$$\mu \sim \mathcal{N}(\mu_0, \sigma^2/\kappa), \quad (6)$$
$$\sigma^{-2} \sim \text{Gamma}(\alpha, \beta). \quad (7)$$

Here, α and β are the shape parameter and rate parameter of the Gamma distribution and κ acts as a scaling factor for the variance σ^2. As prior $p(\mu, \sigma^2|\mu_0, \kappa, \alpha, \beta)$ and posterior $p(\mu, \sigma^2|\mathbf{x}_{1:t})$ form a conjugate pair for the assumptions made above, updates of parameters $\{\mu_0, \kappa, \alpha, \beta\}$ yield a computationally convenient closed form solution [19].

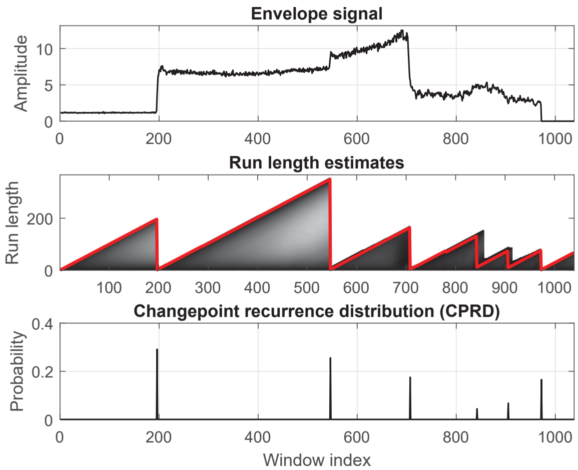


Fig. 1: CPRD estimation. Top: DS1 envelope signal. Middle: Pruned BOCPD solution. Run length log probabilities $\log(p(r_t|\mathbf{x}_{1:t}))$ depicted in gray, MAP estimates $\hat{r}_t$ as bold red line. Bottom: First step of CPRD for signal from top.

IV. Proposed Methods

A. Changepoint Recurrence Distribution (CPRD)

Due to the typical concentration of probability mass of $p(r_t|\mathbf{x}_{1:t})$ 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.

$$\hat{r}_t = \arg\max_{r_t} p(r_t|\mathbf{x}_{1:t}) \quad (8)$$

According to [3], 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

$$p(c_t|\mathbf{x}_{1:t}) \triangleq p(\hat{r}_t|\mathbf{x}_{1:t})\Big|_{\frac{\partial \hat{r}_t}{\partial t} \neq 1}, \quad (9)$$

where $\frac{\partial}{\partial t}$ denotes a derivate with respect to t. Changepoints c_t occur not only due to recurrent changes of process steps, but also due to signal fluctuations or anomalies. This motivates the necessity to filter recurrent changepoints from the set of all changepoints. We propose the following approach for filtering.

Changepoint probability vectors $p(c_t^{(n)}|\mathbf{x}_{1:t})$ 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 $\sum_{n=1}^{N} p(c_t^{(n)}|\mathbf{x}_{1:t})$ increases at locations t of changepoints $c_t^{(n)}$ (i.e., locations t with non-zero probabilities

$p(c_t^{(n)}|\mathbf{x}_{1:t})$ while staying the same at other time steps t where $p(c_t^{(n)}|\mathbf{x}_{1:t}) = 0$. Normalizing $\sum_{n=1}^{N} p(c_t^{(n)}|\mathbf{x}_{1:t})$ allows interpretation as an empirical probability distribution over recurrence of changepoint positions [20]. We name this distribution *changepoint recurrence distribution (CPRD)*.

$$p(c_t^{(1:N)}|c_t^{(n)}, \mathbf{x}_{1:t}) \triangleq \frac{\sum_{n=1}^{N} p(c_t^{(n)}|\mathbf{x}_{1:t})}{\sum_{n=1}^{N}\sum_{t=1}^{T} p(c_t^{(n)}|\mathbf{x}_{1:t})} \tag{10}$$

Recurrence of changepoints $c_t^{(n)}$ at locations t across signals $n = 1 \ldots N$ is denoted by the term $c_t^{(1:N)}$. $p(c_t^{(1:N)}|c_t^{(n)}, \mathbf{x}_{1:t})$ 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 non-parametric maximum likelihood estimate of recurrent changepoint probabilities [21].

The CPRD allows incorporating further prior information. For instance, if times of processing step changes are available, this deterministic prior knowledge can be utilized to complement the empirical information of observed changepoints.

B. Estimation of Future Recurrent Changepoints via CPRD

1) CPRD as informative hazard function: The CPRD can be used to replace the uninformative hazard function $H(\tau) = 1/\lambda$ introduced in [3]. This allows incorporating empirical information about the recurrence of observed changepoints directly into the changepoint prior $p(r_t|r_{t-1})$ (refer to [3] for a detailed discussion). Hence, this approach yields more robust estimates of recurrent signal segments in future signals by suppressing non-recurrent changepoints.

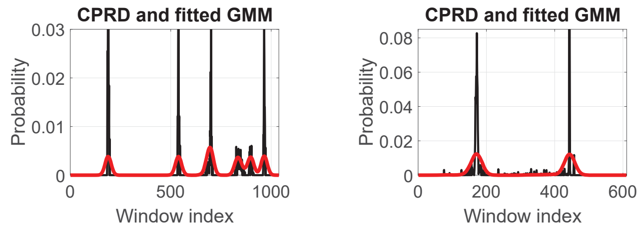
2) CPRD for filtering of BOCPD changepoints: An alternative approach is estimating all changepoints via BOCPD and using the CPRD to separate recurrent from non-recurrent changepoints in a subsequent step: By multiplying initial BOCPD changepoint estimates $p(c_t^{(n)}|\mathbf{x}_{1:t})$ with the empirical CPRD probabilities $p(c_t^{(1:N)}|c_t^{(n)}, \mathbf{x}_{1:t})$, a filtering of changepoint estimates regarding their probability of being recurrent is obtained. This can be interpreted as applying Bayes' Theorem:

$$p(c_t^{(n)}|c_t^{(1:N)}, \mathbf{x}_{1:t}) = \frac{p(c_t^{(1:N)}|c_t^{(n)}, \mathbf{x}_{1:t}) p(c_t^{(n)}|\mathbf{x}_{1:t})}{p(c_t^{(1:N)}|\mathbf{x}_{1:t})} \tag{11}$$

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

$$p(c_t^{(n)}|c_t^{(1:N)}, \mathbf{x}_{1:t}) \propto p(c_t^{(1:N)}|c_t^{(n)}, \mathbf{x}_{1:t}) p(c_t^{(n)}|\mathbf{x}_{1:t}) \tag{12}$$

Non-zero probabilities $p(c_t^{(n)}|c_t^{(1:N)}, \mathbf{x}_{1:t})$ indicate recurrent changepoints. Non-recurrent changepoints are then found as symmetric set difference between BOCPD changepoints $p(c_t^{(n)}|\mathbf{x}_{1:t})$ and recurrent changepoints.



(a) DS1 (b) DS2

Fig. 2: CPRDs and GMM fits for data sets DS1 and DS2.

For stationary behavior of normal changepoints, estimating the CPRD with a large number of normal training signals results in a smooth distribution. For a smaller number of training signals, a post-processing of the CPRD by fitting a kernel density estimator or Gaussian Mixture Model (GMM) can similarly increase smoothness of the CPRD and thus yield more robust changepoint filtering results. We smooth by fitting a GMM, as this yields meaningful features (distance of changepoints to cluster centers, cluster membership probabilities, etc.) for a changepoint-related anomaly detection.

V. Proof-of-Concept Study

The benefits of changepoint recurrence estimation for signal segmentation and unsupervised anomaly detection are studied here for two machine tool monitoring tasks. First, we estimate CPRDs for both data sets (Section V-B1). Then, we compare using these CPRDs as informative hazard function (Section V-B2) or for changepoint filtering (Section V-B3). Finally, we illustrate how to use non-recurrent changepoints and CPRD estimates for unsupervised anomaly detection (Section V-B4).

A. Data Sets

Data set 1 (DS1) consists of 312 normal sensor signals (recorded for a grinding wheel with normal behavior) and 118 signals for different degrees of severity of a grinding wheel anomaly. Data set 2 (DS2) comprises 400 normal signals and 99 signals with machine part collisions. The collisions result in a single impulse-like artifact and thus one additional abnormal changepoint for abnormal DS2 signals.

As changes of processing steps are best observable in signal envelope energy, signal envelopes computed via $\frac{1}{M}\sum_{t=1}^{M}|\mathbf{x}_t|$ in each successive signal window of size $M = 1024$ observations $\mathbf{x}_t$ are used as input for BOCPD changepoint estimation.

B. Results

1) CPRD Estimation: Fig. 2 shows the CPRD estimates (black line) and their GMM fits (bold red line) for DS1 and DS2. Fitting a GMM to the CPRD results in a smoother probability distribution as discussed in Section IV-B.

2) CPRD as informative hazard function: Results of utilizing this GMM as an informative hazard function are depicted in Fig. 3a for an exemplary, abnormal DS1 signal (top). The abnormal changepoint at window index 355 which is assigned when using a constant hazard function (middle) is suppressed by the informative hazard function (i.e., GMM fitted to CPRD) in the bottom figure. A similar behavior is illustrated in Fig. 3b for the abnormal changepoint at index 55 of an abnormal DS2 signal. This confirms the validity of an informative CPRD hazard function for robust signal segmentation.

3) CPRD for filtering of BOCPD changepoints: Results of filtering BOCPD changepoints with a GMM fitted to the DS1 CPRD are depicted in Fig. 4a. 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 DS1 envelope signal. Normal data consist of signals nr. 1 to 91. Signals nr. 1 to 60 (white overlay) are used for estimation of the CPRD (Fig. 2a, black line). Below, run length estimates for different degrees of grinding wheel anomalies (separated by white lines) are plotted.

Recurrent changepoints likely under the GMM fitted to the CPRD are depicted as blue dots. They allow dividing signals into recurrent segments more robustly than via initial BOCPD changepoints (both blue and red dots) similar to the approach in Section V-B2. For all degrees of grinding wheel anomalies, additional non-recurrent changepoints (red dots) occur between indices 300 and 400 or indices 970 and 1040.

For DS2, abnormal machining (signals nr. 201 to 350) frequently resulted in machine part collisions, which resulted in additional changepoints at index 55 (red dots) (Fig. 4b).

Such variations in changepoint patterns are not detectable by the CPRD hazard approach in Section V-B2 and support the benefit of non-recurrent changepoints for anomaly detection.

4) Unsupervised Anomaly Detection: The discrimination of BOCPD changepoints into recurrent and non-recurrent obtained by the approach in Section V-B3 can be used for an unsupervised detection of process anomalies. We consider the following features to be useful for anomaly detection:

- N_{NC}: Number of non-recurrent changepoints in a signal.
- D_C: Average distance of BOCPD changepoints in a signal to closest cluster centers of the GMM CPRD fit.
- MPA: Probability of having at least one abnormal changepoint in a signal (calculated as maximum of the probabilities of all BOCPD changepoints not to be generated by the GMM CPRD fit).

Results are summarized in Table I. Feature scores (columns 3-7) are stated as medians of normal class (N) and abnormal (AN) classes. In DS1, different degrees of grinding wheel anomalies yield multiple AN classes. F1 scores for anomaly detection with each feature are stated in the last column. We classify an anomaly for feature scores more than two normal class standard deviations distant from the normal class median.

Scores for each feature show clear differences between normal and abnormal classes and result in a decent predictive quality as confirmed by the F1 scores. When we consider the full feature set (i.e., a three-dimensional feature space), F1 scores improve to 99.0 % (DS1) and 97.6 % (DS2).

TABLE I: F1 scores for changepoint-related features.

Data	Feature	N	AN1	AN2	AN3	AN4	F1 score
DS1	N_{NC}	0	3	3	3	2	87.5 %
	D_C	13.1	29.4	30.0	33.4	36.9	97.1 %
	MPA	80.9	99.0	99.5	100	100	96.1 %
	All						**99.0 %**
DS2	N_{NC}	0	1				85.9 %
	D_C	3.8	30.6				84.8 %
	MPA	87.9	100.0				83.9 %
	All						**97.6 %**

VI. Conclusions

This work introduces CPRD, a method to assess changepoints in time series data regarding their likelihood to be recurrent. The CPRD can be used either as an informative hazard function in the BOCPD algorithm or as an empirical estimate of the changepoint recurrence likelihood in a separate changepoint partitioning step. Both approaches allow dividing signals into recurrent segments for subsequent extraction of comparable feature scores.

Non-recurrent changepoints, which come as a byproduct of the latter approach, yield information for unsupervised anomaly detection. This is verified for three exemplary anomaly detection features suggested in this work.

Although our experiments focus on machine tool monitoring, the proposed methods can be extended to other applications with a recurrent changepoint structure (e.g., ECG analysis).

References

[1] A. Jovic and N. Bogunovic, "Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features," *Artificial Intelligence in Medicine*, vol. 51, pp. 175–186, March 2011.

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

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

[4] 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, November 2001, pp. 289–296.

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

[6] 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, December 2013.

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

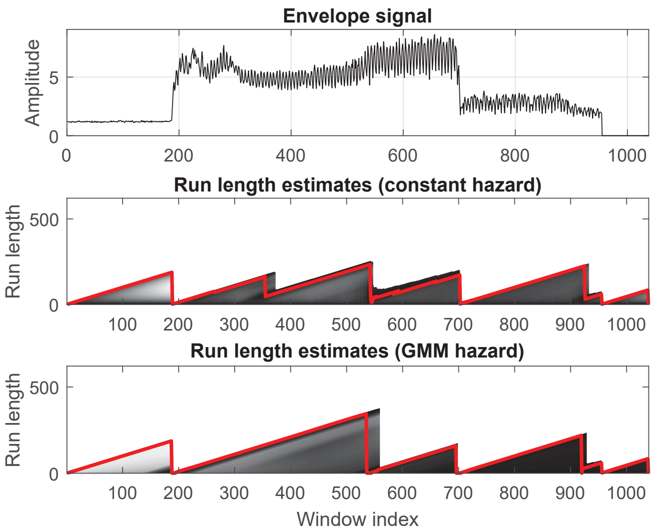
[8] S.-Z. Yu, "Hidden semi-markov models," *Artificial Intelligence*, vol. 174, pp. 215–243, February 2010.

[9] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, pp. 82–97, November 2012.

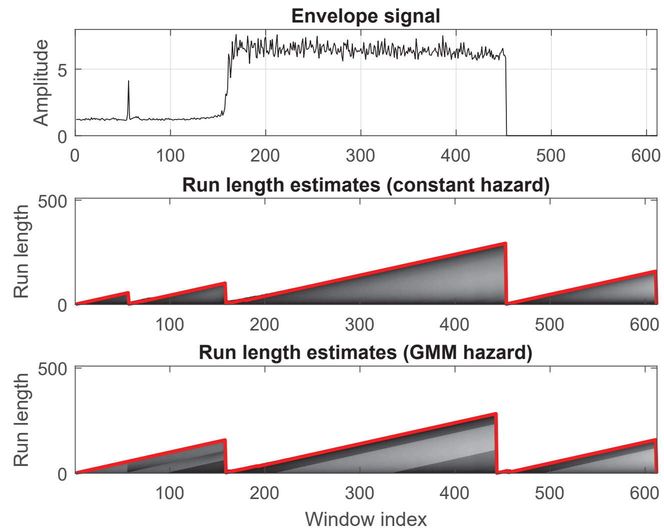
[10] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, June 1954.

[11] 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, July 2013.

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

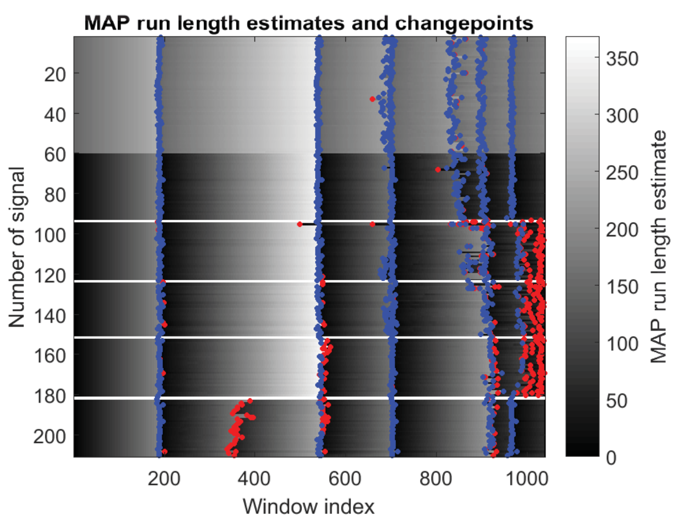


(a) DS1

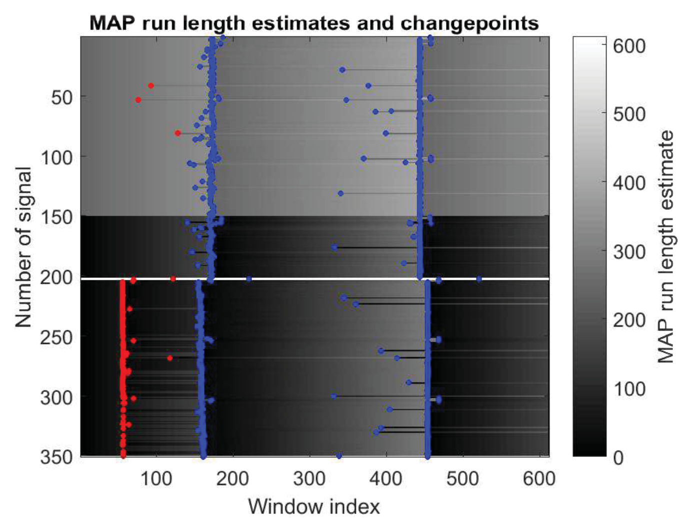


(b) DS2

Fig. 3: Recurrent changepoint estimation with CPRD-based hazard functions. Top: Abnormal envelope signal. Middle/bottom: Run length log probabilities (gray) and MAP estimates $\hat{r_t}$ (red line) for constant (middle) and GMM hazard functions (bottom).



(a) DS1



(b) DS2

Fig. 4: CPRD-based changepoint filtering. Figures show MAP run length estimates obtained via BOCPD, recurrent (blue dots) and non-recurrent (red dots) changepoints for subsets of DS1 and DS2. A white overlay marks signals for CPRD estimation.

[13] 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, June 2010, pp. 927–934.

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

[15] R. Garnett, M. A. Osborne, and S. J. Roberts, "Sequential bayesian prediction in the presence of changepoints," in *Proc. of the 26th International Conference on Machine Learning (ICML'09)*, Montreal, Canada, June 2009, pp. 345–352.

[16] 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.

[17] 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, September 2010.

[18] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*, John Wiley & Sons, Hoboken, USA, 2011.

[19] K. P. Murphy, *Conjugate Bayesian Analysis of the Gaussian Distribution*, University of British Columbia, Vancouver, Canada, 2007.

[20] Z. Ghahramani, *Bayesian Modelling*, University of Cambridge, Cambridge, UK, 2012.

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