

Adaptive Change Detection for Long-Term Machinery Monitoring Using Incremental Sliding-Window

Teng Wang¹ · Guo-Liang Lu¹ · Jie Liu² · Peng Yan¹

Received: 10 June 2017/Revised: 1 September 2017/Accepted: 29 September 2017/Published online: 25 October 2017
© The Author(s) 2017. This article is an open access publication

Abstract Detection of structural changes from an operational process is a major goal in machine condition monitoring. Existing methods for this purpose are mainly based on retrospective analysis, resulting in a large detection delay that limits their usages in real applications. This paper presents a new adaptive real-time change detection algorithm, an extension of the recent research by combining with an incremental sliding-window strategy, to handle the multi-change detection in long-term monitoring of machine operations. In particular, in the framework, Hilbert space embedding of distribution is used to map the original data into the Re-producing Kernel Hilbert Space (RKHS) for change detection; then, a new adaptive threshold strategy can be developed when making change decision, in which a global factor (used to control the coarse-to-fine level of detection) is introduced to replace the fixed value of threshold. Through experiments on a range of real testing data which was collected from an experimental rotating machinery system, the excellent detection performances of the algorithm for engineering applications were demonstrated. Compared with state-of-the-art methods, the proposed algorithm can be more

suitable for long-term machinery condition monitoring without any manual re-calibration, thus is promising in modern industries.

Keywords Machine monitoring · Change detection · Long-term monitoring · Adaptive threshold

1 Introduction

Detection of structural changes from an operational process is a major goal in machinery monitoring which enables to solve many practical problems ranging from early fault detection, safety protection as well as other process control problems. Existing works are mainly based on a retrospective analysis of a data stream composed of numerical condition monitoring (CM) variables, such as vibration, sound, power consumption. The basic idea of a standard retrospective change detection mainly relies on estimating the logarithm of the likelihood ratio between two distributions [1, 2]. This kind of strategy argues that the detection of a change can be converted into the detection of the parameter difference between the two distributions before and after this change point. As a consequence, the retrospective change detection aims to estimate this parameter difference of distributions based upon likelihood ratio statistics. Change decision can be made by performing a null hypothesis testing with a threshold. Many effective tools for this goal such as the cumulative sum metric (CUSUM) [3–6], geometric moving average (GMA) [7, 8] and the generalized likelihood ratio test (GLRT) [9–12] have been widely used. For example, Willersrud et al. [12] developed the GLRT to make efficient downhole drilling washout detection with the multivariate t -distribution; Reñones et al. [6] used the CUSUM analysis for multi-

Supported by National Natural Science Foundation of China (Grant Nos. 61403232, 61327003), Shandong Provincial Natural Science Foundation of China (Grant No. ZR2014FQ025), and Young Scholars Program of Shandong University, China (YSPSDU, 2015WLJH30).

✉ Guo-Liang Lu
luguoliang@sdu.edu.cn

¹ Key Laboratory of High-Efficiency and Clean Mechanical Manufacture of MOE, School of Mechanical Engineering, Shandong University, Jinan 250061, China

² Department of Mechanical and Aerospace Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada

tooth machine tool fault detection. Although these methods have been experimentally demonstrated the effectiveness in various fields, due to the requirement of data after change point, a large detection delay is an essential limitation of these methods for real applications [13]. On the other hand, real-time change detection aims to detect changes as soon as possible when a change occurs, this requirement is crucial in many real-life scenarios such as security monitoring [14, 15], health care [16, 17], automated factory [18, 19] as well as machine operation monitoring studied in this paper. In operation of real-time change detection, at each time when a datum is input, it evaluates what extent the input datum is likely to be a change-point by a certain type of measuring score [20] which does not need any input data after change time. The real-time approaches have succeeded in solving many practical applications (e.g., wind turbine condition monitoring [21], driver vigilance monitoring [22]), and thus are promising [23].

The goal of this paper is to further advance this research line of real-time detection methods. More specifically, our main contributions in this paper are summarized in two folds. The first contribution is to apply a martingale-based framework proposed in our recent article [24] to long-term machine monitoring by combining an incremental sliding-window strategy. The basic idea of the original martingale is to directly learn a statistical regularity from already observed data, and then detect possible change(s) by investigating how much each data is deviated from the regularity using martingale by testing exchangeability. That framework, however, only works for at-most-one-point change detection, thus unsuitable for the cases containing multiple changes in long-term monitoring applications. In this paper, we introduce an incremental sliding-window strategy for solving this problem.

Recall that, the threshold value for change decision making is a key factor of detection accuracy. Potential weakness of the majority of exiting algorithms, e.g., Refs. [5–11], is that they need either human-made instruction/intervention or an off-line cross-validation to confirm the value of threshold before operation, and thus make them largely limited in real applications. Another contribution of this paper is to develop a new adaptive threshold when performing change decision making. In particular, we introduce an alternative factor: a fixed-global parameter used to control the coarse-to-fine level of detection, instead of the fixed value of threshold, for change detection (see Section 3.2 for details). By using this factor, at each step of change decision, the threshold value can be adaptively computed from the already observed data.

Besides the methodological extensions of the proposed method, we also conducted validations on an experimental setup to investigate effectiveness/priority of the method for

change detection with large datasets. For more details, please see Section 4.

The rest of the paper is structured as follows. Section 2 presents the outline of martingale framework for machine monitoring. Section 3 formulates problems addressed in this paper and provides our proposed methods, followed by experimental results in Section 4. Section 5 concludes this paper and shows the future work.

2 Martingale Based Change Detection for Machinery Monitoring

Assuming that we have collected a data stream composed of numerical CM variables from an operational process of a machine, i.e., $X = \{x_1, \dots, x_i, \dots, x_n\}$ where $x_i, i \in \{1, 2, \dots, n\}$ is the variable value at time i , three points are provided in the following to support the use of martingale for change detection:

- (1) The changes are detected by testing the null hypothesis that all n (strangeness) values (which corresponds to x_1, x_2, \dots, x_n , respectively) are exchangeable in the index, through the corresponding exchangeability martingale M_1, M_2, \dots, M_n , where M_n is a measurable function of s_1, s_2, \dots, s_n , satisfying

$$M_n = E(M_{n+1} | M_1, M_2, \dots, M_n). \quad (1)$$

- (2) The following Doob's inequality [25] can be used for rejecting this null hypothesis for a large value of M_n :

$$P(\exists n | M_n \geq \lambda) \leq \frac{1}{\lambda}. \quad (2)$$

- (3) This (exchangeability) martingale is constructed from a p value, the probability of obtaining a test statistic at least as extreme as the one that was actually observed, and the p -value is obtained by a strangeness value appropriately determined in each specific application.

On the basis of the above three points, the outline of performing martingale for change detection is described as follows (see Ref. [25] for more details):

Step 1: The randomized power martingale (RPM) [26] is constructed from the computed s_1, s_2, \dots, s_n by

$$M_t = \prod_{i=1}^t (\varepsilon \hat{p}_i^{\varepsilon-1}), t \in \{1, 2, \dots, n\}, \quad (3)$$

where $\varepsilon \in (0, 1)$ and \hat{p}_i 's are the value computed from \hat{p} –value functions:

$$\hat{p}_i(\{s_1, \dots, s_{i-1}\}, s_i) = \frac{\#\{j : s(j) > s(i)\} + \theta_i \#\{j : s(j) = s(i)\}}{i}, \quad (4)$$

where $\#\{\bullet\}$ is a counting function and θ_i is a random value from a uniformly distribution of $[0, 1]$, $j \in \{1, 2, \dots, i-1\}$.

Step 2: The following inequality is then tested based on the Doob's inequality for any $t \in \{1, 2, \dots, n\}$ to test the hypothesis as below:

$$\begin{aligned} H_0 : \text{no change} : 0 < M_t < \lambda, \\ H_A : \text{change occurs} : M_t \geq \lambda. \end{aligned} \quad (5)$$

That is, if the martingale value M_t is greater than a predefined threshold λ , H_A in Eq. (5) is satisfied, i.e., a change occurs on the time t . Otherwise, the martingale test satisfying H_0 continues to operate as long as $0 < M_t < \lambda$.

3 Problem Formulation and Proposed Scheme

In Section 2, we have provided the outline of martingale-test for change detection. Two problems have to be further considered for long-term machinery monitoring:

- (1) How to deal with multi-change detection in long-term monitoring?
- (2) Is it possible to adaptively compute the threshold value when making change decision?

In the following, we will discuss these two problems in details and provide our proposed schemes.

3.1 Change Detection Using Incremental Sliding-Window

Problem 1. As shown in Eq. (3) and Eq. (5), M_t can be sequentially processed with a fixed-length L sliding-window over the given data stream, and all possible change candidates $t \in \{1, 2, \dots, n\}$ are tested. This process however may be unsuitable for long-term monitoring applications. A key feature of real machine operations is temporal variations, i.e., one operation can last for a long time or only a few seconds. Hence, it is difficult to use a fixed-length L sliding-window to capture transitions (i.e., changes from an operational state to another) in long-term monitoring. More specifically, a small length of L causes over-change-detection and a large length of L causes a large delay. To overcome this problem, we combine the martingale with an incremental sliding-window strategy [27] to design a real-time change detection algorithm for Eq. (5).

Proposed scheme: By virtue of incremental sliding-window, the length L can be automatically updated depending on whether a change is detected or not at time t :

$$\begin{aligned} \text{If } t \text{ is no change} : n_{t+1} &= n_t, \quad L_{t+1} = L_t + \Delta L, \\ \text{If } t \text{ is a change} : n_{t+1} &= t, \quad L_{t+1} = L_1, \end{aligned} \quad (6)$$

where n_t is the starting time when computing the current martingale and L_t is the length of corresponding sliding window at time t . The process starts with $n_1 = 1$ and $L_1 = 1$, and ends with $n_t + L_t > n$ where n is the length of a given data stream in off-line applications or ends at an pre-defined stopping time in on-line applications. Here, it is worth mentioning that ΔL is an increasing step to update the sliding window and was set as 1 in the following experiments.

3.2 Adaptive Threshold for Change Detection

Problem 2. When making change decision by testing the null hypothesis shown in Eq. (5), the threshold of λ is essential as it balances the detection *precision* and *recall* (their definitions will be given in Section 4.3). In general, the value of λ is pre-defined empirically or confirmed by a prior estimation for change detection. It is, however, often difficult to confirm the optimal value in real-world applications. To address this problem, unlike existing works that directly used the original monitored variables for change detection (e.g., Refs. [6–12, 25]), we utilize the Hilbert space embedding of distribution (HED, also called the kernel mean or mean map) to map the original data $\{x_i\}$, $i \in \{1, 2, \dots, n\}$ into the Re-producing Kernel Hilbert Space (RKHS) (see Figure 1). Without going into details, the idea of using HED for change detection is straightforward. By this, the probability distribution is now represented as an element of a RKHS, and the change can be thus detected by using a well behaved smoothing kernel function, whose values are small on the data belonging to the same pattern and large on the data from different patterns.

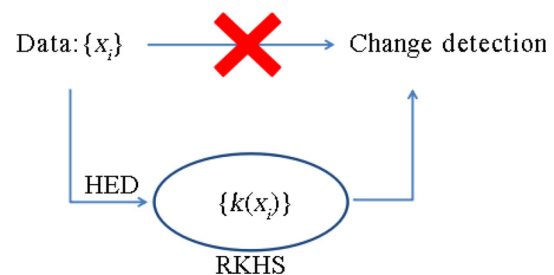


Figure 1 Use of HED to map original data into RKHS for change detection

Proposed scheme: Inspired by Ref. [28], probabilistic distributions can be embedded in RKHS. The center of the HED are the mean mapping functions:

$$\begin{aligned}\mu(p_x) &= E(k(\{x_i\})), \\ \mu(\{x_i\}) &= \frac{1}{t} \sum_{i=1}^t k(x_i),\end{aligned}\quad (7)$$

where $\{x_i\}$ $i = 1, 2, \dots, t$ are assumed to be I.I.D sampled from the distribution P_x . Under mild conditions, $\mu(P_x)$ (same for $\mu(\{x_i\})$) is an element of the Hilbert space. Mapping $\mu(P_x)$ is attractive because each data point x_i has a one-to-one correspondence with mapping $\mu(P_x)$. Thus, we can use the function norm $s(\mu(P_x), k(x_t))$ (instead of $s(P\{x_1, x_2, \dots, x_{t-1}\}, x_t)$) used in Ref. [1]) to quantify the strangeness value s_t for x_t . We do not need to access the actual distributions but rather finite samples to calculate/estimate the distribution ρ .

Lemma 1. As long as the rademacher average [29], which measures the “size” of a class of real-valued functions with respect to a probability distribution, is well behaved, finite sample yield error converges to zero, thus they empirically approximate $\mu(P_x)$ (see Ref. [28] for more details).

The success of kernel methods largely depends on the choice of the kernel function k which is chosen according to the domain knowledge or universal kernels. In this paper, we employ the widely-used Gaussian radial basis function (RBF) kernel by

$$k(x_i) = \exp\left(-\frac{1}{2\sigma^2} \|x_i - \bar{x}\|^2\right), \quad (8)$$

where \bar{x} and σ are the sample mean and standard deviation of the data stream $\{x_1, x_2, \dots, x_t\}$. We next construct s_t to measure the strangeness of x_t to the past data stream up to time $t-1$, i.e., $\{x_1, x_2, \dots, x_{t-1}\}$ in RKHS, as

$$s_t = s(\mu(p_x), k(x_t)) = |k(x_t) - k_{t-1}^c|, \quad (9)$$

where k_{t-1}^c is the kernel center of the data stream, and $|\bullet|$ is distance metric. Here, it is worth mentioning that in real engineering scenarios, the CM variables are often composing of multidimensional values measured from multiple sensors at each time instance, we thus use the Mahalanobis distance [30] to compute the strangeness s_t , by considering correlations between variables such that different patterns in each dimension can be identified and analyzed [30], computed by

$$s_t = \sqrt{(k(x_t) - k_{t-1}^c)' \sum^{-1} (k(x_t) - k_{t-1}^c)} \quad (10)$$

where \sum is the covariance matrix.

Since we used RBF as the kernel function as given in Eq. (8), an isolated data point can be certified if $s_t \geq \alpha \cdot \sigma$,

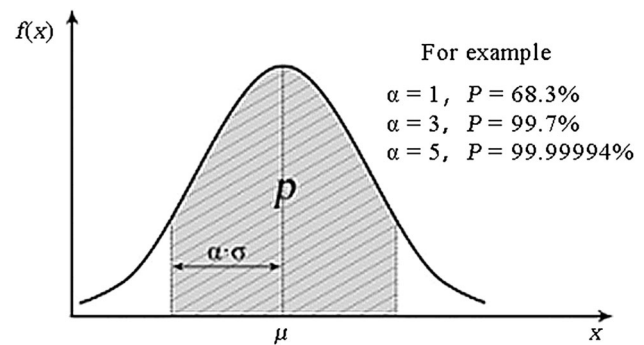


Figure 2 Different α and corresponding confidence level P in a Gaussian distribution

where α is a fixed-global factor controlling the confidence level of detection and σ is the standard deviation computed from existed data (that is, an adaptive threshold).

Based on this fact, kernelized change decision can be made by re-writing Eq. (5) as follows,

$$\begin{aligned}\text{If } 0 < M_t < \alpha \cdot K \cdot \sigma_{t-1} &: \text{ no change,} \\ \text{If } M_t \geq \alpha \cdot K \cdot \sigma_{t-1} &: \text{ change occurs,}\end{aligned}\quad (11)$$

where K is a projection coefficient of data from RKHS to martingale space and σ_{t-1} can be computed adaptively from the data stream up to time $t-1$. In real implementation, the employed Gaussian function is often standardized as a normal Gaussian function (i.e., $\mu = 0$ and $\sigma = 1$). Figure 2 gives typical confidence levels corresponding to different α in Gaussian distribution. Thus K can be fixed as $K \approx 2.17$ by an off-line estimation. The estimation is made as follows: (a) given a set of data streams containing changes, we first make a definition of detection accuracy as $q = N/P$ where N is the number of correctly detected changes and P is the number of total detected changes, and set up a threshold value λ^* to guarantee a perfect accuracy, i.e., $q = 100\%$; (b) then, decrease the value of λ^* gradually to make sure q not decrease; (c) once q decreases, K is computed by

$$K = \frac{\lambda^*}{5 \cdot \sigma}$$

since five times of σ can guarantee all changes can be detected.

4 Experimental Verification

In this section, we aim to demonstrate the effectiveness and priority of the proposed adaptive change-detection algorithm for long-term machine monitoring. Thus, Experiment I and Experiment II are conducted to answer the following two questions:

- (1) Will the proposed incremental sliding-window be more suitable than the fixed-length sliding-window for long-term machine monitoring?
- (2) Can the adaptive detection algorithm be effective for change detection and how does it perform with large datasets?

4.1 Experimental Setup

Figure 3 shows the experimental setup where various motor speed inputs were applied to the setup to simulate changes in real machinery operations. We used the sound signal as the tested CM signal for testing due to its relatively large applicability in implementation. The relevant CM sound signals were acquired by a microphone mounted on the gearbox, and then sent to equipment for a PC.

Figure 4 shows a data stream of collected CM sound signals where three changes indicating transitions from one operational state to another are provided by a human instructor.

4.2 Experiment I: Performance of Incremental Sliding-Window

In our method, we propose to use incremental sliding-window, instead of length-fixed sliding window, for long-

term machine monitoring. To evaluate the effectiveness/priority of this strategy for change detection, a set of testing threshold values from 1 to 20 at an interval of 1 were checked. Particularly, Figure 5 shows three detection cases of using incremental sliding-window for the testing data stream (shown in Figure 4). It is apparent that the performance differs greatly with different values of λ , and the best performance has been achieved when $\lambda = 7$. Specifically, the smaller value of brings higher false alarms but larger value takes a larger detection delay. Here, it is demonstrated again the important roles of threshold value λ for change detection. In the following, since the value of $\lambda = 7$ has been shown the success of change detection for the testing data, we compare this performance with that of using a fixed-length L sliding window detection. From Figure 6 that shows the results of three cases of $L = \{3000, 5000, 7000\}$, we can observe that the small length of $L = 3000$ causes over-detection (i.e., more false alarms) and the large length of $L = 7000$ cause a large detection delay; it tends to achieve a good detection performance when $L = 5000$, while one notes that it is often difficult to fix the length of L of the sliding-window considering the great temporal variations of collected CM signal from real machine operations.

In overall, taking the results shown in Figure 5 and Figure 6 together, the followings were interestingly found:

- The length-fixed sliding-window martingale requires more parameters, i.e., λ and L , for performing change detection, which requires a more complicated prior estimation of them before usage;
- By the incremental sliding-window martingale, only one parameter: λ is required, which inspires an extension of change detection by adaptive threshold.

Both of them inspire the adaptive threshold given in Section 3.2, which will be evaluated in the following.

4.3 Experiment II: Performance of Adaptive Threshold

In this section, we will evaluate the proposed adaptive threshold for machine monitoring. Here, it is noted that since in the Section 4.2, we have demonstrated the priority of using incremental sliding-window for long-term machine monitoring, in this section, we only test the performance of adaptive threshold with incremental sliding-window.

Figure 7 shows the results where all changes in the testing CM data (the same data shown in Figure 4) have been successfully detected without any false alarms when setting α as 3.0 but when λ decreases, more false alarms are brought out. In addition, it is also observed that on the basis of considering the projection coefficient K from RKHS to

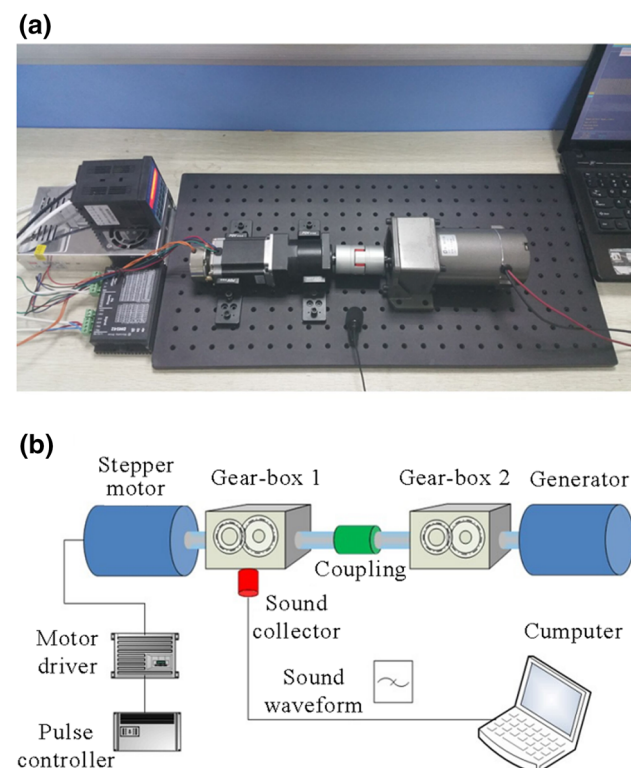


Figure 3 Experimental setup for collection of testing data: (a) Front view; (b) Schematics

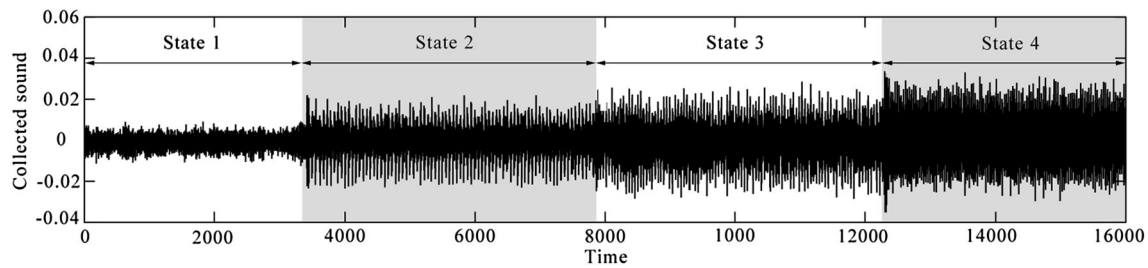


Figure 4 A collected CM sound data stream which contains three changes indicating transitions of machine operations

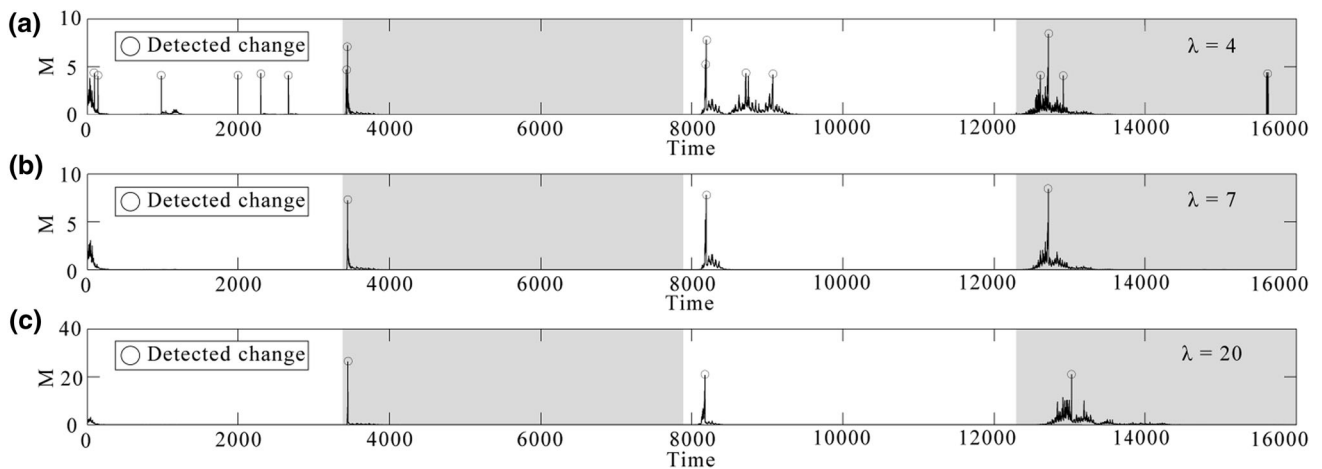


Figure 5 Results of using incremental sliding-window martingale: (a) detection with $\lambda = 4$, mean change delay is 164 with 13 false alarms; (b) detection with $\lambda = 7$, mean change delay is 263 with no false alarm; (c) detection with $\lambda = 20$, mean change delay is 320 with no false alarm

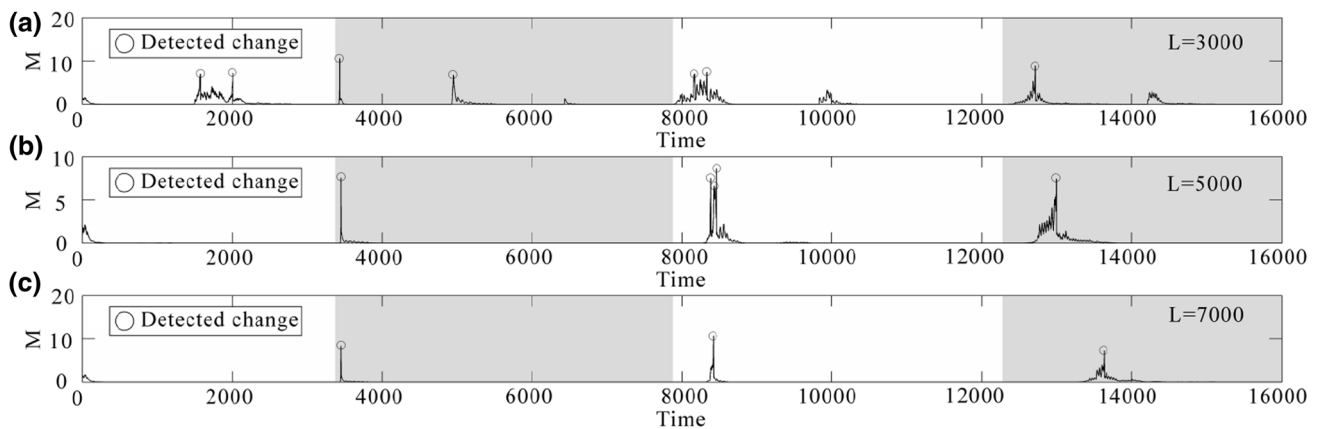


Figure 6 Experimental results of using length-fixed sliding-window martingale for $\lambda = 7$: (a) $L = 3000$, (b) $L = 5000$, (c) $L = 7000$

martingale space and the global fixed confidence level α , the threshold value can be computed/adjusted adaptively according to the standard deviation computed from the past data at each step of change decision making as mentioned in Eq. (11), in other words, the threshold value is not fixed in the whole process, that is different from many existing works [7–11, 25]. All of them are consistent with the analysis previously made in Section 3.2. Moreover, to verify the effectiveness of the adaptive threshold with large

datasets, we collected streams totally containing 90 changes for evaluation. The performance evaluation is based on two retrieval performance indicators: *precision* and *recall*, which are defined respectively as

$$\text{Precision} = \frac{\text{Number of Correct Detectins of Changes}}{\text{Number of Detections of Changes}}$$

$$\text{Recall} = \frac{\text{Number of Correct Detectins of Changes}}{\text{Number of Ture Changes}}.$$

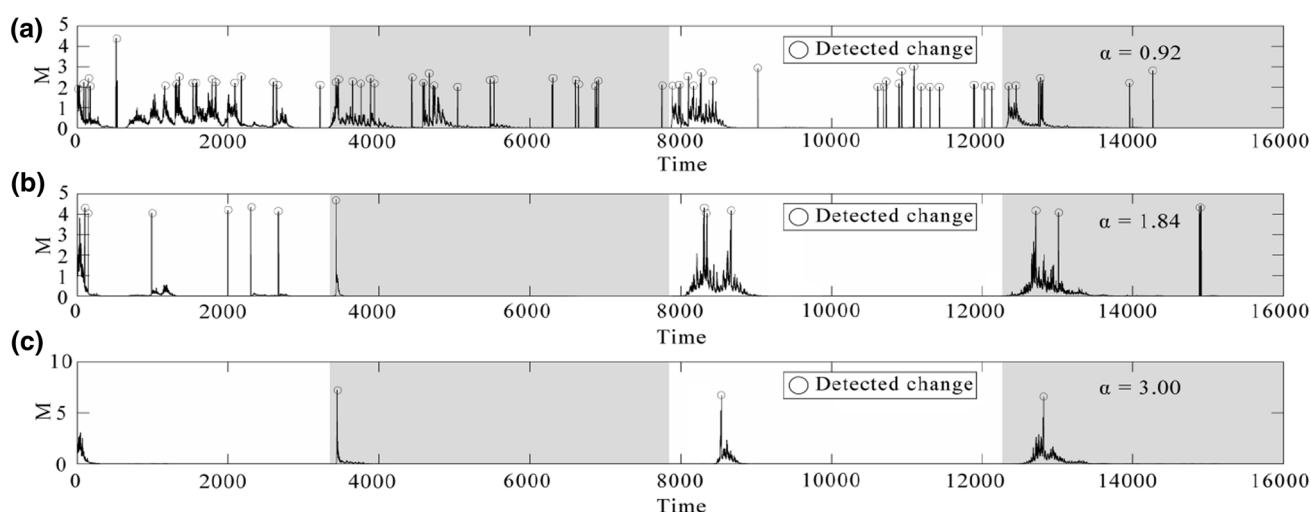


Figure 7 Experimental results by adaptive threshold when setting α as (a) 0.92, (b) 1.84 and (c) 3.00

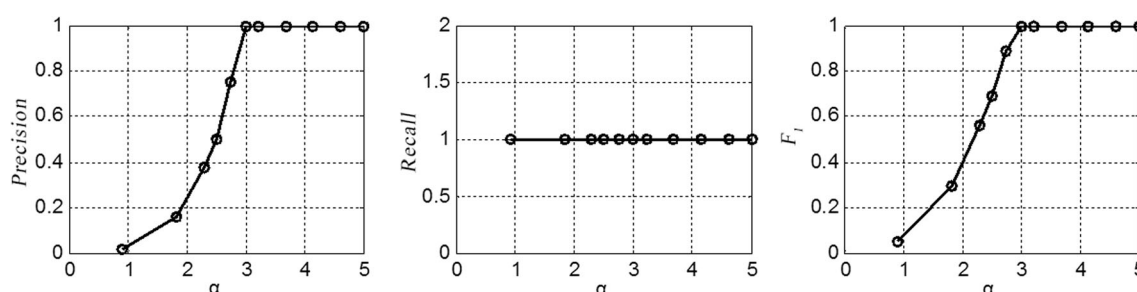


Figure 8 Detection performances on large datasets: (a) *precision*, (b) *recall*, (c) F_1

Precision is the probability that a detection is actually correct, i.e., a true change. *Recall* is the probability that the detection recognizes a true change.

In addition, we also use a single performance indicator F_1 defined as

$$F_1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}.$$

Apparently, F_1 is a harmonic mean between *precision* and *recall*, and a high value of F_1 ensures reasonably a high balance between *precision* and *recall*.

Figure 8 shows the detection performances for different values of $\alpha \in \{0.92, 1.84, 2.30, 2.50, 2.75, 3.00, 3.22, 3.69, 4.15, 4.61, 5.00\}$. Specifically, it can be found in Figure 8(a) that with an increasing value of α , the detection *precision* increases and achieves the best performance when $\alpha > 3$. On the other hand, for the *recall* shown in Figure 8(b), our proposed method can always obtain a perfect performance, that is 100% which means the all true changes can be successfully detected for tested values of α . These results can be more clearly observed in Figure 8(c) where F_1 can achieve the best performance when $\alpha > 3$. All of these results are not surprising because three times of

σ guarantees approximate 99.8% (as shown in Figure 2) data points have been contained in a Gaussian distribution. Considering a smaller α brings a smaller detection delay as shown in Figure 7, it is thus recommended $\alpha = 3$ when using the proposed method in real applications.

5 Conclusions

In this paper, we have extended our recent work [25] to long-term machine monitoring where two schemes are proposed: 1) using the incremental sliding-window to solve the problem of multi-change detection; and 2) developing an adaptive threshold when making change decision. Experimental results on an experimental setup demonstrated great successes of the proposed method in multi-change detection in long-term monitoring. With this, it can be concluded that the improved algorithm is feasible for a new generation of long-term machine monitoring systems. In view of this, further work will be done to continue verifying the capability of the improved algorithm for detecting a wider range of changes when operating a machine to make it ready for commercial exploitation.

In addition, considering that the detection delay is one of essential aspects to be considered when design a detection method, another future work is to extract informative features to represent the raw collected data for modeling in order to further decrease the delay of our method when detecting changes

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Fokianos Konstantinos, E Gombay, A Hussein. Retrospective change detection for binary time series models. *Journal of Statistic Planning and Inference*, 2014, 145(2): 102-112.
2. N Peach. Detection of Abrupt Changes: Theory and Applications. M Basseville; I V Nikiforov, ed. *Journal of the Royal Statistical Society*, 2015, 15(158): 326-327.
3. A G Tartakovsky, A S Polunchenko. Quickest change point detection in distributed multisensor systems under unknown parameters. *International Conference on Information Fusion, IEEE*, Istanbul, Turkey, July 9-12, 2013: 1-8.
4. Hafiz Zafar Nazir, Muhammad Riaz, Ronald J M M Does, et al. Robust CUSUM control charting. *Quality Engineering*, 2013, 25(3): 211-224.
5. H Cho. Change-point detection in panel data via double CUSUM statistic. *Electronic Journal of Statistics*, 2016, 10(2): 2000-2038.
6. A Reñones, L J D Miguel, J R Perán. Experimental analysis of change detection algorithms for multitooth machine tool fault detection. *Mechanical Systems and Signal Processing*, 2009, 23(7): 2320-2335.
7. K Atashgar. Identification of the change point: an overview. *International Journal of Advanced Manufacturing Technology*, 2013, 64(9-12): 1663-1683.
8. M A Graham, S Chakraborti, S W Human. A nonparametric exponentially weighted moving average signed-rank chart for monitoring location. *Computational Statistics and Data Analysis*, 2013, 55(8): 2490-2503.
9. A Willersrud, M Blanke, L Lmsland. Incident detection and isolation in drilling using analytical redundancy relations. *Control Engineering Practice*, 2015, 41: 1-12.
10. A S Willsky, H L Joned. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *Automatic Control IEEE Transactions on*, 1976, 21(1): 108-112.
11. A Willersrud, M Blanke, L Lmsland, et al. Drillstring Washout Diagnosis Using Friction Estimation and Statistical Change Detection. *IEEE Transactions on Control Systems Technology*, 2015, 23(5): 1886-1900.
12. A Willersrud, M Blanke, L Lmsland, et al. Fault diagnosis of downhole drilling incidents using adaptive observers and statistical change detection. *Journal of Process Control*, 2015, 30: 90-103.
13. C Aldrich, L Auret. *Unsupervised process monitoring and fault diagnosis with machine learning methods*. London: Springer Publishing Company, Incorporated, 2013: 1-15.
14. B Yildiz. *Development of a hybrid intelligent system for on-line real-time monitoring of nuclear power plant operations*. Boston: Massachusetts Institute of Technology, 2003.
15. A J Tickle. Morphological scene change detection for night time security. *Proceedings of SPIE - The International Society for Optical Engineering*, 2012, 8540(24): 2288-2291.
16. H Wang, C Zhang, T Shi, et al. Real-time EEG-based detection of fatigue driving danger for accident prediction. *International Journal of Neural Systems*, 2015, 25(2): 1550002.
17. K C Huang, T Y Huang, C H Chuang, et al. An EEG-based fatigue detection and mitigation system. *International Journal of Neural Systems*, 2016, 26(4): 1650018.
18. M Karapirli, E Kandemir, S Akyol, et al. Failure mode and effect analysis in asset maintenance: a multiple case study in the process industry. *International Journal of Production Research*, 2013, 51(4): 1055-1071.
19. J Lee, F Wu, W Zhao, M Ghaffari, et al. Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications. *Mechanical Systems and Signal Processing*, 2014, 42(1-2): 314-334.
20. Y Urabe, K Yamanishi, R Tomioka, et al. Real-time change-point detection using sequentially discounting normalized maximum likelihood coding. *Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, Shenzhen, China, Springer-Verlag, May 24-27, 2011: 185-197.
21. W Yang, P J Tavner, W Tian. Wind turbine condition monitoring based on an improved spline-kernelled Chirplet transform. *IEEE Transactions on Industrial Electronics*, 2015, 62(10): 6565-6574.
22. C T Lin, C H Chuang, H C S Huang, et al. Wireless and wearable EEG system for evaluating driver vigilance. *IEEE Transactions on Biomedical Circuits and Systems*, 2014, 8(2): 165-176.
23. A Pouliezios, G S Stavrakakis. *Real time fault monitoring of industrial processes*. Netherlands: Kluwer Academic Publishers, 2013: 463-519.
24. G L Lu, Y Q Zhou, C H Lu, et al. A novel framework of change-point detection for machine monitoring. *Mechanical Systems and Signal Processing*, 2017, 83: 533-548.
25. J L Doob. Boundary properties of functions with finite Dirichlet integrals. *Annales-Institut institute Fourier*, 1962, 12: 573-621.
26. V Vovk, I Nouredinov, A Gammernan. Testing exchangeability on-line. *Machine Learning, Proceedings of the Twentieth International Conference DBLP*, Washington, August 21-24, 2003: 768-775.
27. E Tsamoura, A Gounaris, Y Manolopoulos. Incorporating change detection in the monitoring phase of adaptive query processing. *Journal of Internet Services and Applications*, 2016, 7(1): 1-18.
28. A Smola, A Gretton, L Song, et al. A Hilbert space embedding for distributions. *International Conference on Algorithmic Learning Theory*, Sendai, Japan, Springer Berlin Heidelberg, October 1-4, 2007: 13-31.
29. P L Bartlett, S Mendelson. Rademacher and Gaussian Complexities: Risk Bounds and Structural Results. *Journal of Machine Learning Research*, 2002, 3: 463-482.
30. P C Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences*, 1936, 2: 49-55.

Teng Wang, received his bachelor degree from Hefei University of Technology in 2016, and now is a master candidate at Shandong University. His research interest includes machine monitoring, prognostics and health management.

Guo-Liang Lu, received the bachelor degree and master degree from Shandong University, Jinan, China, in 2006 and 2009, respectively. He received his PhD degree from the Graduate School of Information Science and Technology of Hokkaido University, Sapporo, Japan, in March 2013. He is currently an associate professor in Shandong University. His research interests mainly include signal processing, machine monitoring, machine vision and visual servo control.

Jie Liu, is currently an Associate Professor in the Dept. of Mechanical & Aerospace Engineering at Carleton University, Ottawa, Canada. He obtained his bachelor degree in Electronics and Precision Engineering from Tianjin University (China) in 1998, his master degree in Control Engineering from Lakehead University (Canada) in 2005, and his PhD degree in Mechanical Engineering from the University of Waterloo (Canada) in 2008. Before joining Carleton, he worked as an Postdoctoral Fellow in the Dept. of Mechanical Engineering at UC Berkeley for one and half years. Dr. Liu has been engaged in interdisciplinary research in the areas of prognostics and health management, intelligent mechatronic systems, and battery management systems for more than thirteen years. His research results have been disseminated through over 35 journal publications and 20 conference papers. He is a Steering Committee Member of Annual IEEE PHM Conferences, an Associate Editor of IEEE

Transactions on Reliability, an IEEE Senior Member, and a registered Professional Engineer in Ontario, Canada.

Peng Yan, received the bachelor degree and master degree from Southeast University in 1997 and 1999 respectively, and the PhD degree from the Ohio State University, Columbus, OH, in 2003, all in Electrical Engineering. From 2004 to 2005, he worked as postdoctorate researcher at the Ohio State University. From 2005 to 2011, he held various industry positions including a Senior Staff Engineer at Seagate Technology at Shakopee MN and a Staff Scientist at United Technology Research Center (UTRC) at Easthartford CT. Currently he is a full professor at the school of Mechanical Engineering of Shandong University, China. His research interests include robust control and control of high precision mechatronics.