An Adaptive Approach for Online Monitoring of Large Scale Data Streams

Shuchen Cao [a] and Ruizhi Zhang [b] *
[a] Department of Statistics, University of Nebraska-Lincoln, Lincoln, USA
[b] Department of Statistics, University of Georgia, Athens, USA

October 8, 2024

Abstract

In this paper, we propose an adaptive top-r method to monitor large-scale data streams where the change may affect a set of unknown data streams at some unknown time. Motivated by parallel and distributed computing, we propose to develop global monitoring schemes by parallel running local detection procedures and then use the Benjamin-Hochberg (BH) false discovery rate (FDR) control procedure to estimate the number of changed data streams adaptively. Our approach is illustrated in two concrete examples: one is a homogeneous case when all data streams are i.i.d with the same known pre-change and post-change distributions. The other is when all data are normally distributed, and the mean shifts are unknown and can be positive or negative. Theoretically, we show that when the pre-change and post-change distributions are completely specified, our proposed method can estimate the number of changed data streams for both the pre-change and post-change status. Moreover, we perform simulations and two case studies to show its detection efficiency.

Keywords: False discovery rate, CUSUM, quickest change detection, process control

1 Introduction

Process monitoring and change-point detection of high-dimensional streaming data has many important applications, such as network security (Polunchenko et al., 2012; Tartakovsky et al., 2013), medical diagnostics (Nika et al., 2015), or intrusion detection on video surveillance (Oberti et al., 2002; Xiong and Lee, 1998). In many cases, the high-dimensional data streams may have complicated spatial and temporal correlation structures, which may cause difficulty in detecting the true change quickly. To remove the effect of correlation in advance, some decorrelation techniques methods are commonly used. For

*CONTACT:ruizhi.zhang@uga.edu

example, Apley and Tsung (2002); Xie et al. (2012); Yan et al. (2018); Qiu et al. (2020) develop some spatial-temporal models. Then, the model residuals are used for monitoring and change detection. Moreover, Jin and Shi (1999); Paynabar et al. (2016) propose to monitor some informative features such as PCA coefficients or Wavelets from the original high-dimensional data. See Woodall and Montgomery (2014); Qiu (2020) for a comprehensive review. For both types of methods, the (standardized) residuals or feature coefficients can often be assumed to be independent and identically distributed. Thus, in the paper, we mainly focus on the problem of monitoring large-scale independent data streams.

However, it is nontrivial to develop efficient monitoring procedures for large-scale independent data streams due to two challenges. The first is about computational complexity because of the need for real-time monitoring in some applications. The second is about the unknown sparsity of the number of affected data streams. To address the computation issue, many computationally efficient algorithms are proposed by combing efficient local detection statistics, say CUSUM (Page, 1954), in different ways. Tartakovsky et al. (2006) proposed a "MAX" scheme by taking the maximum of local CUSUM statistics to get a global monitoring statistic. Mei (2010) developed a "SUM" scheme by combining the summation of local CUSUM statistics to get a global monitoring statistic. A "Top-r" scheme is proposed by Mei (2011) and Liu et al. (2015). It takes the summation of the top r largest CUSUM statistics to obtain a global statistic. In Liu et al. (2019) and Zhang and Mei (2018), the general "SUM-shrinkage" framework is proposed. It applies shrinkage transformation on local CUSUM statistics and takes the summation of these shrunk statistics to obtain a global monitoring statistic. Benefiting from the recursive format of the CUSUM, all of these methods can be computed recursively and thus can be implemented in real time. However, it turns out the "MAX" scheme is efficient in detecting the change only if the number of affected data streams is very small, while the "SUM" scheme is efficient only if a lot of data streams are changed. Moreover, as shown in Liu et al. (2019) and Zhang and Mei (2018), for some popular thresholding functions, such as soft-thresholding, hard-thresholding, and order-thresholding, the optimal choices of thresholding parameters in the corresponding "SUM-shrinkage" procedure also depend on the number of affected data streams. To address the challenge of an unknown but sparse number of affected data streams, some generalized-likelihood-ratio-based methods are proposed in Xie and Siegmund (2013) and Fellouris and Sokolov (2016), which have been shown to be second-order optimal for the Gaussian data streams in Fellouris and Sokolov (2016) and Chan (2017). However, these methods are computationally infeasible for online monitoring large-scale data streams over a long time period. Moreover, Liu et al. (2015); Xian et al. (2021); Zhang and Mei (2023) studied the problem of monitoring multiple data streams when only partial information is available each time. They have proposed some adaptive sampling and change detection methods under the sampling constraint, which is different from our problem when the information of all data streams is always available.

In this article, we propose an adaptive top-r scheme that does not rely on the information on the number of affected data streams and can be implemented in real-time. The key idea of our proposed approach is based on a false discovery rate control procedure, called the Benjamini and Hochberg (BH) procedure (Benjamini and Liu, 1999), which is originally aimed to select rejected hypotheses in multiple testing problems. Using such BH procedure, at each time, we will select some data streams and obtain a global monitoring statistic by taking the summation of their CUSUM statistics. Therefore, instead of using a fixed r in the "Top-r" scheme, our method will select different numbers of data streams adaptively to different post-change scenarios. We should also mention that in literature, false discovery rate is also used as an error metric for online monitoring of multiple data streams. Some procedures have been developed to yield smaller detection delays while controlling the false alarm rate (Li and Tsung (2009); Gandy and Lau (2013); Chen et al. (2020)). However, although our proposed method is motivated by the false discovery rate control, our method is still designed to control the in-control average run length instead of the false discovery rate.

Our research makes three main contributions. First, our proposed adaptive top-r method provides a natural connection between the false discovery rate control in multiple testing problems and the change-point detection of large-scale data streams. Second, our proposed method does not rely on the prior knowledge of the number of changed data streams and thus is adaptive to various post-change scenarios. Numerical experiments are conducted to illustrate the detection efficiency of our proposed method. Third, theoretically, we show that for our proposed method, under mild conditions, on average, the number of selected data streams is close to one before the change and is close to the true number of affected data streams after the change. Such consistent results imply the nice detection efficiency of our method.

This paper is organized as follows: In Section 2, we give some preliminaries about false discovery rate (FDR) control and then introduce schemes for change-point detection of multiple independent data streams. In Section 3, we introduce our proposed adaptive top-r scheme when both pre-change and post-change distributions are fully specified and show their theoretical properties. We then propose to combine the adaptive CUSUM procedure with our adaptive top-r method to develop a monitoring procedure for the normal distribution with an unknown mean shift in Section 4. The simulation results are provided in Section 5 to illustrate the performance of the scheme. Finally, in Section 6, we conduct two case studies using our proposed method. The proofs of theorems are presented in the Appendix.

2 Preliminaries

In this section, we first provide some background information for the False Discovery Rate (FDR) control in multiple hypothesis testing problems. Then, we provide some preliminaries about process monitoring and change-point detection of multiple data streams.

2.1 *False discovery rate control*

Suppose we have k hypotheses $H_1, H_2, ..., H_k$ to be tested. For a given data set and some tests or decision rules, we can get a decision of "reject" or "accept" for each of the k hypotheses. Suppose there are a total of R rejected hypotheses (discoveries), and V of them are true null hypotheses (false discoveries). Then, the False Discovery Rate (FDR) is defined as the expected value of false discoveries, i.e.,

$$\mathrm{FDR} = \mathbf{E}(\frac{V}{R}). \tag{1}$$

The FDR concept was formally described by Benjamini and Hochberg (Benjamini and Hochberg, 1995), where the so-called Benjamini–Hochberg (BH) procedure was proposed to control the FDR under certain conditions. Specifically, for a given FDR level $q \in (0, 1)$, suppose the test produces a p-value p_i for each hypothesis. The order statistics of these p-values are denoted by $p_{(1)} \leq p_{(2)} \leq ... \leq p_{(k)}$. Let i^* be the smallest i so that $p_{(i)} \geq \frac{i}{k}q$. Then the BH procedure will reject i^* hypotheses corresponding to the p-values $p_{(1)} \leq p_{(2)} \leq ... \leq p_{(i^*)}$. This procedure is referred to as the step-down BH procedure. There is another variant of the step-up BH procedure defined in a similar way. In this article, we will focus on this step-down procedure since it often allows more discoveries than the step-up counterpart while still controlling the FDR at a rate q (Gavrilov et al., 2009).

2.2 *Change-point detection of large-scale data streams*

Suppose we are monitoring K data streams $X_{k,n}, k = 1, 2, ..., K$. Assume that the data $X_{k,n}$'s are initially independent and identically distributed (i.i.d.) with probability density function (pdf) $f_0(x)$. At some unknown time $\nu \geq 1$, an undesired event occurs and changes the distributions of an unknown but sparse set $\mathcal{S}$ of data streams so that the affected data streams $X_{k,n}$'s have another distribution $f_1(x)$ when $n \geq \nu$. The goal is to develop an efficient monitoring procedure to detect the change as soon as possible.

When $K = 1$ or when monitoring a single local data stream, say, the kth data stream, the problem has been well studied in the literature of sequential change-point detection; see Page (1954), Shiryaev (1963), Lorden (1971), Pollak (1985), Moustakides (1986), Lai (1995). One of the most efficient detection procedure is Page's CUSUM procedure, which raises an alarm at the first time n when the local CUSUM statistic $W_{k,n}$ exceeds some prespecified threshold, where $W_{k,n}$ has a recursive form:

$$W_{k,n} = \max\Big(W_{k,n-1} + \log \frac{f_1(X_{k,n})}{f_0(X_{k,n})},\ 0\Big). \tag{2}$$

When the number of data streams K is large, many efficient global monitoring procedures are developed by combining local CUSUM statistics to a global monitoring statistic in different ways.

The "SUM" schemes, introduced in Mei (2010), will raise an alarm the first time when the summation of local CUSUM statistics exceeds a prespecified global threshold:

$$N_{\text{sum}}(b) = \inf\Big\{n \geq 1 : \sum_{k=1}^{K} W_{k,n} \geq b\Big\}. \tag{3}$$

The "MAX" scheme introduced in Tartakovsky et al. (2006) will raise an alarm at the first time when the largest CUSUM statistic exceeds a predetermined threshold:

$$N_{\max}(b) = \inf\Big\{n \geq 1 : \max_{1\leq k\leq K} W_{k,n} \geq b\Big\}. \tag{4}$$

Since the "SUM" scheme uses information from all data streams, including the unchanged ones, it will work well only if a lot of data streams are changed. Meanwhile, the "MAX" scheme only uses the information of one data stream with the largest CUSUM, and it will have a better detection performance only if one or very few data streams are changed.

The "Top-r" scheme in Zhang and Mei (2018), which uses the summation of the top r largest CUSUM statistics to get the global monitoring statistic:

$$N_r(b) = \inf\Big\{n \geq 1 : \sum_{i=1}^{r} W_{(K-i+1),n} \geq b\Big\} \tag{5}$$

where $W_{(1),n} \leq W_{(2),n} \leq \ldots \leq W_{(K),n}$ are the order statistics of the K CUSUM statistics $W_{1,n}, \ldots, W_{K,n}$. We can see the "SUM" scheme and the "MAX" scheme are the special case of the "Top-r" scheme when $r = K$ and $r = 1$ respectively. Extensive numerical results imply the "Top-r" scheme will have better performance if the true number of changed data streams is around r (Liu et al., 2015; Zhang and Mei, 2018).

The "SUM-shrinkage" scheme (Liu et al., 2019; Zhang and Mei, 2018) uses the shrinkage function on each local CUSUM statistic and sums them together to obtain the global monitoring statistic:

$$N_{\mathrm{G}}(b) = \inf\Big\{n \geq 1 : \sum_{k=1}^{K} h_k(W_{k,n}) \geq b\Big\}, \tag{6}$$

where h_k is some shrinkage transformation functions. For instance, the "hard-thresholding" transformation function takes the form of $h_k(W_{k,n}) = W_{k,n}\mathbb{1}(W_{k,n} \geq b_k)$, and the "soft-thresholding" transformation function takes the form of $h_k(W_{k,n}) = \max(W_{k,n} - b_k, 0)$, where b_k is the local threshold for the kth data stream. We can see that when $b_k = 0$, the "SUM-shrinkage" scheme with the above two transformation functions will become the "SUM" scheme in (3). As shown in Zhang and Mei (2018), the optimal choice for the values of b_k may depend on the true number of affected data streams.

Although these schemes are computationally fast because of the recursive form of the CUSUM, their detection efficiencies are affected by the true number of affected data streams, which may be unknown in practice.

To address the unknown number of affected data streams, Zou et al. (2015) proposed an efficient monitoring procedure by combining the ideas of Tukey's higher criticism statistics and goodness-of-fit statistics:

$$N_{gof}(b) = \inf\left\{ n \geq 1 : \sum_{i=1}^{K} \left\{ \log\left[\frac{U_{(i),n}^{-1} - 1}{(K-1/2)(i-3/4) - 1} \right] \right\}^2 \times \mathbb{1}(U_{(i),n} > (i-3/4)/K) \right\}, \tag{7}$$

where $U_{i,n} = F_n(W_{i,n}), F_n(.)$ denotes the cdf of the CUSUM statistic $W_{i,n}$ under the pre-change distribution. $U_{(1),n} \leq U_{(2),n} \leq \cdots \leq U_{(K),n}$ are the order statistics of $(U_{1,n}, \cdots, U_{K,n})$. Since there is no closed form for the cdf of CUSUM, an approximation to the null steady-state distribution of the CUSUM statistic developed in Grigg and Spiegelhalter (2008) is used. However, such a method only works for detecting the mean shift of the normal distribution.

Thus, in this paper, we will propose an adaptive "Top-r" monitoring scheme based on the false discovery rate control. As we will show later, our proposed method does not rely on a pre-knowledge about the true number of affected data streams, can be adaptive to different post-change scenarios, and also works for many distributions other than normal distributions.

3 Our proposed monitoring procedure for known post-change distributions

In this section, we propose an adaptive "Top-r" scheme to monitor a large number of data streams when the pre-change and post-change distributions are completely specified. Then we will investigate some theoretical properties of our proposed method.

We consider the problem of change-point detection of K independent data streams $X_{k,n}$ as we described in Section 2.2. At the high level, at each time n,, we borrow the idea of the BH procedure in the FDR control and consider testing K hypotheses for these K data streams. Let R_n be the number of rejected hypotheses from the BH procedure. Then, we calculate the summation of the largest R_n CUSUM statistics as the global monitoring statistic and raise the alarm if it exceeds a certain pre-defined threshold. The key question is how to get p-values for each data stream in order to use the BH procedure for FDR control. Since the CUSUM statistic is a good indicator of the potential change for each data stream, here we will calculate p-values using the CUSUM statistics. Specifically, let $W_{k,n}$ be the CUSUM statistic for k^{th} data stream at time n, and $w_{k,n}$ be the actually observed value of the CUSUM statistic. Then, we consider the corresponding p-value $p^*_{k,n} = \mathbf{P}_0(W_{k,n} \geq w_{k,n})$, where $\mathbf{P}_0$ denotes the probability measure when data follow the pre-change distribution f_0. However, due to the complexity of the sampling distribution of the CUSUM statistic, there is no closed-form expression for the exact p-value $p^*_{k,n}$. Thus, we propose to use

$$p_{k,n} = \exp(-W_{k,n}) \tag{8}$$

to conduct the BH FDR control and construct the global monitoring statistics. By the property of CUSUM (see appendix A in Siegmund (2013)): $\mathbf{P}_0(W_{k,n} \geq x) \leq e^{-x}$ for any $x \geq 0$, we have $p^*_{k,n} \leq p_{k,n}$. Thus, $p_{k,n}$ is an upper bound of the exact p-value $p^*_{k,n}$.

We then summarize our proposed adaptive Top-r monitoring procedure below:

Let R_n be the number of rejected hypotheses using the $p_{k,n}$ and the BH procedure. Then we have

$$R_n = \begin{cases} K, \quad \text{if } p_{(r),n} < \frac{r}{K}\alpha \text{ for all } r = 1, 2, \ldots, K \\ \min\left\{1 \leq r \leq K : p_{(r),n} \geq \frac{r}{K}\alpha\right\}, \quad \text{otherwise,} \end{cases} \tag{9}$$

where $p_{(1),n} \leq p_{(2),n} \leq \ldots \leq p_{(K),n}$ are the order statistics of the K statistics $p_{1,n}, \ldots, p_{K,n}$. That means at each time n, we will reject top R_n of hypotheses and use the corresponding CUSUM statistics to construct the global monitoring procedure. In other words, given a preset global threshold b, we will raise an alarm the first time

$$T_\alpha(b) \quad = \quad \inf\Big\{n \geq 1 : \sum_{r=1}^{R_n} W_{(K+1-r),n} \geq b\Big\}, \tag{10}$$

where $W_{(1),n} \leq W_{(2),n} \leq \ldots \leq W_{(K),n}$ are the order statistics of the K CUSUM statistics $W_{1,n}, \ldots, W_{K,n}$. We should emphasize that although we did not use the exact p-value to conduct the FDR control and select the rejected hypotheses, as we will show in the later simulation section, our proposed global monitoring procedure (10) using the simple upper bound $p_{k,n}$ in (8) can detect various change scenarios quickly. Moreover, similar to the

BH procedure for the FDR control in multiple testing problems, the tuning parameter α in our proposed method may also control the false discovery rate when no change occurs. Intuitively, the choice of α will also affect the detection performance of our proposed monitoring procedure. Based on our simulation and case study results, we can set α as a small number (for example, $\alpha = 0.1$) to detect the change quickly.

Next, we will present the theoretical properties of our proposed adaptive top-r monitoring procedure. We first introduce some notations. We use $\mathbf{E}_0$ to denote the expectation when there are no changes in the data streams, i.e., data $X_{k,n}$ are i.i.d with pdf f_0. We use $\mathbf{E}_m$ to denote the expectation when the data stream changes at the first time $\nu = 1$ and the pdf of m out of K data streams are changed to f_1. Let R_n be the number of rejected data streams as defined in (9), then we have the following theorem:

Theorem 1. *For* $0 < \alpha \leq \frac{1}{2}$, *we have*

$$1 \leq \mathbf{E}_0(R_n) < 1 + \frac{\alpha}{(1+\alpha)^2}[\log(K-1) - \alpha + 1]. \tag{11}$$

Moreover, if $e^{n\mu} > (m+1)(K-m)$, *and* $\frac{K}{e^{n\mu}} < \alpha < \frac{K}{(m+1)(K-m)}$, *we have*

$$\left(1 - \frac{\sigma^2}{n[\frac{1}{n}\log(\frac{K}{\alpha}) - \mu]^2}\right)(m+1) \leq \mathbf{E}_m(R_n) < m + 1 + \frac{(m+K)(K-m-1)}{2}\beta, \tag{12}$$

where $\mu = \int f_1(x)\log\frac{f_1(x)}{f_0(x)}dx$, $\sigma^2 = \int f_1(x)(\log\frac{f_1(x)}{f_0(x)})^2 dx - \mu^2$, *and* $\beta = \frac{(K-m)\alpha/K}{[1-(m+1)(\frac{K-m}{K})\alpha]^2}$.

The proof of Theorem 1 is postponed to the Appendix. Based on Theorem 1, we can see when there is no change, as $\alpha \to 0$, we have $\mathbf{E}_0(R_n) \to 1$. This implies on average we will only use the information of one data stream to monitor whole data streams when there is no change. This will reduce much noise information. Moreover, as $n \to \infty$, $\frac{\sigma^2}{n[\frac{1}{n}\log(\frac{K}{\alpha})-\mu]^2}$ goes to 0. As $\alpha \to 0$, β goes to 0. Therefore, for a very small α satisfying the condition $\frac{K}{e^{n\mu}} < \alpha < \frac{K}{(m+1)(K-m)}$, we have $\mathbf{E}_m(R_n) \to m+1$. This result means if m data streams change, on average, approximately $m+1$ data streams will be rejected and used to monitor the process. Therefore, our proposed method does not include more additional noise information when monitoring these data streams and can be adaptive to different change scenarios.

4 Our proposed monitoring procedure for unknown post-change means

Suppose we are monitoring K data streams $X_{k,n}$'s. Initially, the data $X_{k,n}$'s are iid $N(0,1)$. At some unknown time ν, the distribution of the k-th local data stream might change to $N(\mu,1)$ if affected. Here, we assume the post-change mean μ is unknown, but the minimum magnitude of the shift $\rho > 0$ is known. That is, assume $|\mu| \geq \rho$. We then propose to use the adaptive Top-r method and the adaptive CUSUM statistics (Lorden and Pollak, 2008; Liu et al., 2019) to construct a global monitoring procedure.

Specifically, since we are interested in detecting both positive and negative local mean shifts for affected data streams, we will use the two-sided adaptive CUSUM statistic in Liu et al. (2019) for each local data stream at time n :

$$W_{k,n} = \max(W_{k,n}^{(1)}, W_{k,n}^{(2)}), \tag{13}$$

where $W_{k,n}^{(1)}$ and $W_{k,n}^{(2)}$ are the local one-side adaptive CUSUM detection statistics of Lorden and Pollak (2008) for detecting positive and negative mean shifts, respectively. Specifically,

$$\begin{aligned} W_{k,n}^{(1)} &= \max\left(W_{k,n-1}^{(1)} + \hat{\mu}_{k,n}^{(1)} X_{k,n} - \frac{1}{2}(\hat{\mu}_{k,n}^{(1)})^2, 0\right), \\ W_{k,n}^{(2)} &= \max\left(W_{k,n-1}^{(2)} + \hat{\mu}_{k,n}^{(2)} X_{k,n} - \frac{1}{2}(\hat{\mu}_{k,n}^{(2)})^2, 0\right), \end{aligned} \tag{14}$$

where

$$\hat{\mu}_{k,n}^{(1)} = \max\Big(\rho, \frac{s + S_{k,n}^{(1)}}{t + T_{k,n}^{(1)}}\Big) > 0, \qquad \hat{\mu}_{k,n}^{(2)} = \min\Big(-\rho, \frac{-s + S_{k,n}^{(2)}}{t + T_{k,n}^{(2)}}\Big) < 0, \tag{15}$$

and for $j = 1, 2$ and for any k, the sequences $(S_{k,n}^{(j)}, T_{k,n}^{(j)})$ are defined recursively

$$\begin{pmatrix} S_{k,n}^{(j)} \\ T_{k,n}^{(j)} \end{pmatrix} = \begin{cases} \begin{pmatrix} S_{k,n-1}^{(j)} + X_{k,n-1} \\ T_{k,n-1}^{(j)} + 1 \end{pmatrix} & \text{if } W_{k,n-1}^{(j)} > 0 \\ \begin{pmatrix} 0 \\ 0 \end{pmatrix} & \text{if } W_{k,n-1}^{(j)} = 0 \end{cases} \tag{16}$$

Note that $\hat{\mu}_{k,n}^{(1)}$ and $\hat{\mu}_{k,n}^{(2)}$ in (15) are the estimates of the post-change mean when restricted to the positive and negative values, respectively, under the assumption that $|\mu| \geq \rho$. Clearly, $W_{k,n}^{(1)}$ is designed to detect positive local mean shift, whereas $W_{k,n}^{(2)}$ is to detect negative local mean shifts. Also, the two-sided local detection statistic $W_{k,n}$ in (13) is always nonnegative for any k at any time step n, and it will become large when there is a local mean shift no matter whether such mean shift is positive or negative. Moreover, $t > 0, s > 0$ are pre-specified constants, and s/t can be considered a prior estimate of the post-change mean. We will use $s = 1, t = 4$ as the default setting in the simulation study.

Then, we will raise an alarm the first time

$$T^u_\alpha(b) \quad = \quad \inf\Big\{n \geq 1 : \sum_{r=1}^{R_n} W_{(K+1-r),n} \geq b\Big\}, \tag{17}$$

where $W_{(1),n} \leq W_{(2),n} \leq \ldots \leq W_{(K),n}$ are the order statistics of the K adaptive CUSUM statistics $W_{1,n}, \ldots, W_{K,n}$ in (13). R_n is the number of rejected hypothesis defined in (9) using $p_{k,n} = \exp(-W_{k,n})$.

5 Simulation

In this section, we evaluate the performance of our proposed adaptive top-r monitoring procedures, compare them with existing methods, and validate our theoretical results under different scenarios. For all these methods, we first find appropriate values of the global stopping threshold b so that the in-control average run length of these methods is equal to some designed value (within the range of sampling error). Then, using the obtained stopping threshold b, we simulate the detection delays of these methods when the change happens at $n = 1$ under different sparse cases of the number of affected data streams. Note since all of the monitoring statistics of these methods are nonnegative, the average of these simulated detection delays when change happens at time $n = 1$ is an estimate of the worst-case detection delay (Lorden, 1971), which is often used to quantify the theoretical detection performance of the monitoring procedure. All Monte Carlo simulations are based on 2500 repetitions.

5.1 Mean shift of normal distribution

In the first experiment, we focus on detecting a mean shift in the normal distribution. We set the number of data streams $K = 100$, and the pre-change distribution is the standard normal distribution, i.e., $f_0 \sim \mathcal{N}(0,1)$. If the data stream is affected after the change point, the post-change distribution $f_1 \sim \mathcal{N}(1,1)$.

For comparison, we consider our proposed method $T_\alpha(b)$ in (10) with two $\alpha = 0.1, 0.2$, the "Top-r" procedure $N_r(b)$ in (5) with $r = 5, 10$, the "MAX" scheme $N_{\max}(b)$ in (4), the "SUM" scheme $N_{\text{sum}}(b)$ in (3), the "SUM-shrinkage" scheme $N_G(b, b_k)$ in (6) with $b_k = -\log(0.1) = 2.3026$, and the goodness-of-fit test $N_{gof}(b)$ in (7). Moreover, note for our proposed methods, we use an upper bound of the p-value of the CUSUM statistic $p_{k,n} = \exp(-W_{k,n})$ to conduct BH FDR control and find the number of rejected data streams R_n as in (9). To see the robustness and efficiency of such an approximation of p-values, we also consider similar procedures by using the actual p-value $p^*_{k,n} = \mathbf{P}_0(W_{k,n} \geq w_{k,n})$. However, as we mentioned before, there is no closed form of the p-values of the CUSUM statistics except for the normal distribution with the mean shift (Grigg and Spiegelhalter, 2008). Thus, at each time n, we simulate the empirical distribution of the CUSUM statistics and obtain the estimated $p^*_{k,n}$ by computing the percentage of the simulated CUSUM statistics that are greater than the observed CUSUM statistic $w_{k,n}$. The corresponding procedure is denoted by $N^*_\alpha(b)$. To make a pair comparison, we also consider two choices of $\alpha = 0.1$ and $\alpha = 0.2$.

Note all of the above methods assume the pre-change and post-change distributions to be known. Then we also consider our proposed method $T^u_\alpha(b)$ in (17) with $\alpha = 0.1, 0.2$ and the method $N_{XS}(b, p_0 = 0.1)$ in Xie and Siegmund (2013) based on generalized likelihood ratio (GLR):

$$N_{XS}(b, p_0) = \inf\{n \geq 1 : \max_{0 \leq i < n} \sum_{k=1}^{K} \log(1 - p_0 + p_0 \exp[\frac{(U^+_{k,n,i})^2}{2}] \geq b\},$$

where for all $1 \leq k \leq K, 0 \leq i < n$,

$$U^+_{k,n,i} = \max\big(0, \frac{1}{\sqrt{n-i}} \sum_{j=i+1}^{n} X_{k,j}\big).$$

These two methods do not use the information of the post-change mean.

For all these methods, we find the threshold b so that the in-control average run length is 5000. Table 1 summarizes the resulting detection delays. First, we can see the "MAX" scheme has the smallest detection delay when the number of the changed data stream is one. However, when the number of data streams becomes larger, even if the number of

Table 1: A comparison of the detection delays for the mean shift of the normal distribution with in-control average run length equal to 5000.

	# affected data streams					
	1	3	5	8	10	20
Methods use the post-change mean information						
$T_{\alpha=0.1}(b=12.43)$	24.5(0.15)	10.4(0.04)	9(0.01)	8(0)	8(0)	7(0)
$T^*_{\alpha=0.1}(b=17.2)$	25.2(0.18)	14.2(0.01)	11.6(0.04)	10.1(0.03)	9.5(0.03)	6.0(0)
$T_{\alpha=0.2}(b=17.8)$	32.2(0.16)	13.7(0.06)	8.3(0.01)	8(0.01)	8(0)	6(0)
$T^*_{\alpha=0.2}(b=21.9)$	26.6(0.23)	14.6(0.02)	11.8(0.04)	10.1(0.01)	9.9(0.01)	5.0(0.01)
$N_{\text{sum}}(b=88.7)$	52.1(0.35)	21.8(0.12)	14.7(0.07)	10.3(0.04)	8.7(0.03)	5.3(0.02)
$N_{\max}(b=11.3)$	23.2(0.18)	16.2(0.09)	14.3(0.07)	12.9(0.06)	12.4(0.05)	11(0.04)
$N_{r=5}(b=29.55)$	29.6(0.21)	14.2(0.07)	10.7(0.05)	8.7(0.03)	8(0.03)	6.3(0.02)
$N_{r=10}(b=44.08)$	34.3(0.24)	15.4(0.08)	11.1(0.05)	8.5(0.03)	7.5(0.03)	5.5(0.02)
$N_{\text{G}}(b=21.5, b_k=2.3)$	33.9(0.23)	15.4(0.08)	11.1(0.05)	8.5(0.03)	7.5(0.03)	5.3(0.02)
$N_{gof}(b=227.6)$	34.9(0.28)	15.6(0.05)	11.3(0.05)	8.7(0.03)	7.7(0.03)	5.2(0.02)
Methods do not use the post-change mean information						
$T^u_{\alpha=0.1}(b=12.6)$	24.3(0.13)	12.9(0.06)	10.9(0.05)	9.1(0.02)	8.8(0.02)	7.6(0.02)
$T^u_{\alpha=0.2}(b=14.7)$	26.3(0.17)	14.8(0.06)	12.1(0.04)	9.4(0.02)	9.1(0.02)	7.8(0.01)
$N_{XS}(b=19.5, p_0=0.1)$	31.7(0.24)	13.4(0.09)	9.4(0.06)	6.7(0.04)	5.7(0.03)	3.4(0.02)

affected data streams is three, its performance becomes much worse than other methods. Moreover, the "SUM" has a larger detection delay when the number of changed data streams is less than 10. Furthermore, although the methods $N_{XS}(b, p_0 = 0.1)$, $N_{\text{G}}(b, b_k = 2.3)$, and $N_{r=10}(b)$ that designed for detecting change of 10 data streams have smaller detection delay when the number of changed data streams is around 10, their detection delays are larger than our proposed procedures $T_{\alpha=0.1}(b)$ when the number of changed data streams is less than or equal to 5.

It is interesting to see that although the method $N_{r=5}(b)$ is designed to detect a change of five data streams, our proposed methods $T_{\alpha=0.1}(b)$ and $T_{\alpha=0.2}(b)$ are still better than it when the number of changed data streams is around five. Theoretically, it has been shown that the top-r method is optimal when the number of affected data streams is r. However, it is an asymptotic result when the in-control average run length goes to infinity.

Therefore, the top-r method may not perform best for the finite in-control average run length. One explanation is that for the top-r when r=5, under the in-control status, we still need to compute the summation of the largest five CUSUM statistics to obtain the global monitoring statistic. Therefore, we should choose a larger threshold b to obtain the designed in-control average run length. However, for our adaptive top-r method, the number of rejected data streams R_n may be smaller than five under the in-control status. Thus, we just need to use a smaller threshold b to obtain the designed in-control average run length. On the other hand, under the out-of-change status, when five data streams are affected, R_n may be close to five. Thus, our adaptive top-r method may perform better than the top-r method in some finite sample cases.

Moreover, note when the post-change mean is unknown, compared to the GLR-based method $N_{XS}(b)$, our adaptive method $T_\alpha^u(b)$ still has small detection delays when the number of affected data streams is small. It is also interesting to see that our proposed adaptive top-r methods $T_\alpha(b)$ have competitive performance compared with the $T_\alpha^*(b)$, which is constructed using the simulated p-values of the CUSUM statistics. This finding implies the robustness and efficiency of our proposed methods by using a simple upper bound of the p-values.

All of these results imply that although our proposed method does not rely on the information on the number of changed data streams, it can detect a wide range of sparse change scenarios quickly.

5.2 Variance change of normal distribution

In the second experiment, we focus on detecting a change of variance in the normal distribution. We set the number of data streams $K = 100$, and the pre-change distribution is the standard normal distribution, i.e., $f_0 \sim \mathcal{N}(0, 1)$. If the data stream is affected after the change point, the post-change distribution $f_1 \sim \mathcal{N}(0, 2)$.

Since the $N_{gof}(b), N_{XS}(b)$ and our proposed method $T_\alpha^u(b)$ are only designed for detecting the mean shift of normal distribution, in this study, we will only compare the performance of our proposed method $T_\alpha(b)$ in (10) with two $\alpha = 0.1, 0.2$, the "Top-r" procedure $N_r(b)$ in (5) with $r = 5, 10$, the "MAX" scheme $N_{\max}(b)$ in (4), the "SUM" scheme $N_{\text{sum}}(b)$ in (3), the "SUM-shrinkage" scheme $N_G(b, b_k)$ in (6) with $b_k = -\log(0.1) = 2.3026$, and the adaptive top-r method using simulated p-values $T^*_\alpha(b)$ with two $\alpha = 0.1, 0.2$.

We find the threshold b for all these methods, so the in-control average run length is 1000. Table 2 summarizes the resulting detection delays. First, we can see the "MAX" scheme has the smallest detection delay when the number of the changed data stream is one. However, when the number of data streams becomes larger, even if the number of affected data streams is three, its performance becomes much worse than our proposed methods. Moreover, the "SUM" has a larger detection delay when the number of changed data streams is less than 10. Furthermore, although our proposed methods T_α have a similar performance with $N_{r=5}$ when the number of affected data streams is five and have a similar performance with $N_{r=10}$ when the number of affected data streams is ten, overall, our proposed adaptive top-r methods have smaller detection delay for a wide range of the number of affected data streams. Moreover, it is also interesting to see that our proposed adaptive top-r methods $T_\alpha(b)$ have competitive performance compared with the $T^*_\alpha(b)$, which is constructed using the simulated p-values of the CUSUM statistics. This finding implies the robustness and efficiency of our proposed methods by using a simple upper bound of the p-values.

Table 2: A comparison of the detection delays for the variance change of the normal distribution with in-control average run length equal to 1000.

	# affected data streams					
	1	3	5	8	10	20
$T_{\alpha=0.1}(b = 10.18)$	13.1(0.18)	6.7(0.08)	5.8(0.06)	4.5(0.03)	3.9(0.02)	2.5(0.02)
$T^*_{\alpha=0.1}(b = 12.4)$	13.6(0.16)	6.8(0.06)	5.3(0.03)	4.3(0.03)	3.3(0.01)	3.0(0.01)
$T_{\alpha=0.2}(b = 10.97)$	13.9(0.2)	6.6(0.07)	5.6(0.06)	4.6(0.02)	3.2(0.02)	2.6(0.02)
$T^*_{\alpha=0.2}(b = 13.7)$	13.4(0.15)	6.5(0.05)	4.6(0.03)	3.9(0.02)	3.4(0.01)	3.0(0.01)
$N_{\text{sum}}(b = 55.7)$	26.5(0.38)	11.3(0.13)	7.7(0.08)	5.4(0.05)	4.6(0.04)	2.7(0.02)
$N_{\max}(b = 8.94)$	13.0(0.26)	7.1(0.13)	5.5(0.09)	4.3(0.07)	3.8(0.06)	2.8(0.04)
$N_{r=5}(b = 23.79)$	16.6(0.29)	7.6(0.11)	5.4(0.07)	4.0(0.05)	3.4(0.04)	2.2(0.03)
$N_{r=10}(b = 34.67)$	19.3(0.31)	8.6(0.12)	6.0(0.07)	4.3(0.05)	3.7(0.04)	2.3(0.02)
$N_{\text{G}}(b = 12.9, b_k = 2.3)$	16.7(0.29)	7.6(0.11)	5.4(0.07)	4.0(0.05)	3.4(0.04)	2.2(0.03)

5.3 Change of Poisson distribution

In the third experiment, we focus on detecting a change in the parameter in the Poisson distribution. We set the number of data streams $K = 100$, and the pre-change distribution is the Poisson distribution with parameter $\lambda = 1$, i.e., $f_0 \sim \text{Poisson}(1)$. If the data stream is affected after the change point, the post-change distribution $f_1 \sim \text{Poisson}(2)$.

For the same reason as we stated in the previous subsection, in this study, we will only compare the performance of our proposed method $T_\alpha(b)$ in (10) with two $\alpha = 0.1, 0.2$, the "Top-r" procedure $N_r(b)$ in (5) with $r = 5, 10$, the "MAX" scheme $N_{\max}(b)$ in (4), the "SUM" scheme $N_{\text{sum}}(b)$ in (3), the "SUM-shrinkage" scheme $N_G(b, b_k)$ in (6) with $b_k = -\log(0.1) = 2.3026$, and the adaptive top-r method using simulated p-values $T^*_\alpha(b)$ with two $\alpha = 0.1, 0.2$.

We find the threshold b for all these methods, so the in-control average run length is 1000. Table 3 summarizes the resulting detection delays. First, the "MAX" scheme has the smallest detection delay when the number of the changed data stream is one. However, when the number of data streams becomes larger, even if the number of affected data streams is three, its performance becomes much worse than our proposed methods. Moreover, the "SUM" has a larger detection delay when the number of changed data streams is less than 10. Furthermore, our proposed method $T_{\alpha=0.2}$ has a smaller detection delay than $N_{r=5}$ and $N_{r=10}$ when the number of affected data streams is not greater than 10, which implies our adaptive top-r methods have better performance for a wide range of the number of affected data streams. Moreover, similar to the previous simulation study, our proposed adaptive top-r methods $T_\alpha(b)$ have competitive performance compared with the $T^*_\alpha(b)$, which is constructed using the simulated p-values of the CUSUM statistics. This finding implies the robustness and efficiency of our proposed methods by using a simple upper bound of the p-values.

5.4 Change of exponential distribution

In the fourth experiment, we focus on detecting a change in the parameter in the exponential distribution. We set the number of data streams $K = 100$, and the pre-change distribution is the exponential distribution with parameter $\lambda = 1$, i.e., $f_0 \sim \text{Exp}(1)$. If the data stream

Table 3: A comparison of the detection delays for the change of the Poisson distribution with in-control average run length equal to 1000.

	# affected data streams					
	1	3	5	8	10	20
$T_{\alpha=0.1}(b = 12.71)$	44.0(0.44)	23.5(0.09)	21.5(0.04)	16.1(0.03)	18.9(0.02)	12.0(0.04)
$T^*_{\alpha=0.1}(b = 16.2)$	46.0(0.58)	26.1(0.11)	16.5(0.11)	14.9(0.02)	17.8(0.03)	16.3(0.04)
$T_{\alpha=0.2}(b = 13.44)$	45.1(0.45)	23.9(0.1)	18.8(0.1)	14.9(0.04)	12.7(0.02)	11.8(0.04)
$T^*_{\alpha=0.2}(b = 19.6)$	54.7(0.67)	26.6(0.2)	16.0(0.02)	12.6(0.09)	15.6(0.02)	13.8(0.05)
$N_{\text{sum}}(b = 91.86)$	81.5(1.07)	38.3(0.39)	26.5(0.22)	19.3(0.14)	16.4(0.11)	10.1(0.06)
$N_{\max}(b = 9.1)$	40.9(0.62)	27(0.28)	22.9(0.21)	20.4(0.17)	19.3(0.16)	16.6(0.12)
$N_{r=5}(b = 26.6)$	49.6(0.67)	26.0(0.25)	19.7(0.16)	15.9(0.11)	14.3(0.1)	11.1(0.06)
$N_{r=10}(b = 41.06)$	56.9(0.75)	28.2(0.27)	20.7(0.17)	16.0(0.11)	14.1(0.09)	10.1(0.05)
$N_{\text{G}}(b = 18.4, b_k = 2.3)$	56.7(0.74)	28.2(0.26)	20.7(0.17)	16.0(0.11)	14.1(0.09)	9.9(0.05)

is affected after the change point, the post-change distribution $f_1 \sim \text{Exp}(2)$.

We still compare the performance of our proposed method $T_\alpha(b)$ in (10) with two $\alpha = 0.1, 0.2$, the "Top-r" procedure $N_r(b)$ in (5) with $r = 5, 10$, the "MAX" scheme $N_{\max}(b)$ in (4), the "SUM" scheme $N_{\text{sum}}(b)$ in (3), the "SUM-shrinkage" scheme $N_G(b, b_k)$ in (6) with $b_k = -\log(0.1) = 2.3026$, and the adaptive top-r method using simulated p-values $T^*_\alpha(b)$ with two $\alpha = 0.1, 0.2$.

We find the threshold b for all these methods, so the in-control average run length is 1000. Table 4 summarizes the resulting detection delays. From the table, we can observe similar phenomena as previous simulation studies, which implies our proposed adaptive top-r methods can detect a wider range of possible changes efficiently.

5.5 Study of the number of rejected data streams

In the last experiment, we simulate the expectation of rejected data streams by our proposed method, i.e., R_n in (9) to validate the results in Theorem 1. In this experiment, we will run our proposed method $T_\alpha(b)$ for a long time without stopping to see the limit performance of R_n. Specifically, we still set $K = 100$, $f_0 \sim \mathcal{N}(0, 1)$, $f_1 \sim N(1, 1)$, and set $n = 200$. Then, we simulate the R_n under different scenarios of numbers of affected data streams by Monte

Table 4: A comparison of the detection delays of for the change of the exponential distribution with in-control average run length equal to 1000.

	# affected data streams					
	1	3	5	8	10	20
$T_{\alpha=0.1}(b=11.22)$	30.3(0.22)	15.9(0.08)	13.5(0.04)	10.9(0.03)	10.7(0.03)	7.1(0.01)
$T'_{\alpha=0.1}(b=13.9)$	30.2(0.23)	17.0(0.06)	15.6(0.04)	12.4(0.05)	11.9(0.05)	5.8(0.02)
$T_{\alpha=0.2}(b=13.12)$	33.6(0.27)	15.7(0.07)	13.5(0.04)	10.9(0.01)	10.9(0.01)	7.4(0.01)
$T'_{\alpha=0.2}(b=16.3)$	31.3(0.29)	17.2(0.07)	14.6(0.03)	12.1(0.04)	11.7(0.04)	5.9(0.01)
$N_{\text{sum}}(b=76.6)$	62.0(0.53)	27.9(0.18)	19.1(0.11)	13.5(0.07)	11.4(0.06)	6.8(0.03)
$N_{\max}(b=8.94)$	29.9(0.03)	17.8(0.15)	14.6(0.11)	12.3(0.09)	11.3(0.08)	9.0(0.06)
$N_{r=5}(b=25.48)$	37.5(0.36)	18.1(0.13)	13.3(0.09)	10.3(0.06)	9.1(0.05)	6.4(0.03)
$N_{r=10}(b=38.72)$	43.5(0.39)	20.1(0.14)	14.4(0.09)	10.7(0.06)	9.3(0.05)	6.2(0.03)
$N_{\text{G}}(b=16.04, b_k=2.3)$	41.5(0.38)	19.4(0.13)	14.0(0.09)	10.6(0.06)	9.2(0.05)	6.2(0.03)

Carlo simulations with 2500 repetitions. The resulting average value and the standard deviation of R_n are reported in Table 5. The table shows that our proposed method rejects one data stream on average when there is no change. Also, after the change occurs, our proposed method rejects approximately the number of truly affected data streams plus one. We further plot the average number of rejected data streams R_n with respect to the time n for our proposed adaptive top-r method with $\alpha = 0.1$ and $\alpha = 0.2$ in Figure 1 and Figure 2 respectively. From these plots, we can see for different numbers of changed data streams, the average number of rejected data streams $\mathbf{E}(R_n)$ by our method will converge closely to the true number of data streams quickly. These results are consistent with our Theorem 1 and imply that our proposed method does not rely on the information of the number of affected data streams and can be adaptive to detect the change point under different post-change scenarios.

6 Case study

In this section, we apply our proposed adaptive top-r method on two real datasets: one is for solar flare detection, and the other is for fault detection in a progressive forming process. Both case studies show that our proposed method is efficient to detect the change quickly

Table 5: Average rejected data streams of our proposed method at $n = 200$ with $\alpha = 0.1$ and $\alpha = 0.2$.

	# affected data streams						
	0	1	3	5	10	20	100
$\alpha = 0.1$	1.1(0.25)	2.1(0.37)	4.2(0.51)	6.3(0.61)	11.6(0.81)	21.9(1.01)	100(0)
$\alpha = 0.2$	1.1(0.41)	2.3(0.56)	4.5(0.77)	6.7(0.94)	12.2(1.21)	23.0(1, 55)	100(0)

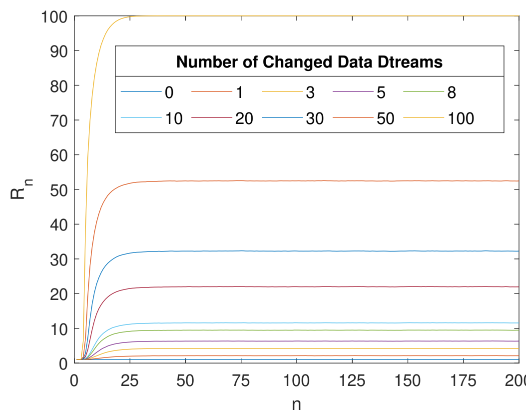

Figure 1: $\mathbf{E}(R_n)$ when $\alpha = 0.1$

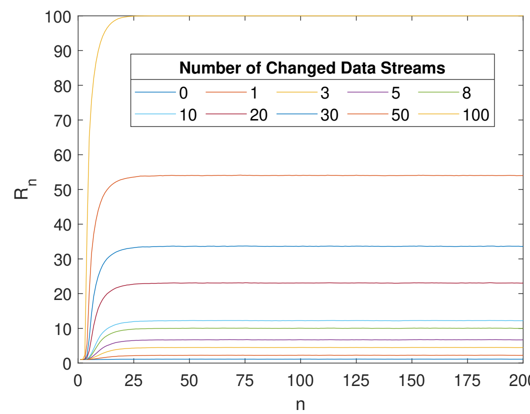

Figure 2: $\mathbf{E}(R_n)$ when $\alpha = 0.2$

without knowing the number of truly changed data streams.

6.1 Solar flare detection

In this section, we use our proposed method to detect the solar flare occurrence. A solar flare is an intense release of energy, usually observed near the sun's surface as a sudden increased bright area. With intense radio emissions, solar flares can create radiation hazards, disable satellites, and interfere the radio communication. Thus, it is desirable to detect the solar flare as soon as possible.

The dataset is a sequence of 300 solar images recorded in video form, which is available at "`https://voices.uchicago.edu/willett/research/software/mousse/`". The size of each image is $232 \times 292 = 67744$ pixels. Since the original image data are correlated, we use the method in Xie et al. (2012) to remove the background and then monitor the residues for solar flare detection. The video shows that a solar flare occurs from time $n = 187$ to $n = 202$ and reaches the brightest scene at approximately $n = 202$. Another one occurs from $n = 216$ to $n = 268$ and is brightest at approximately $n = 233$. Figure 3 shows the residual image at $n = 268$ when the second solar flare is happening.

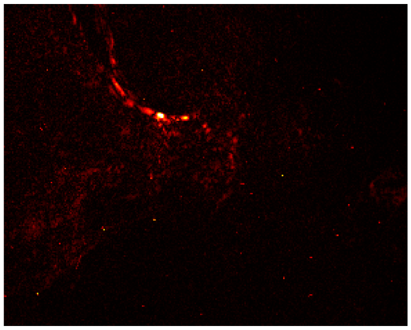

Figure 3: Solar image at n=268

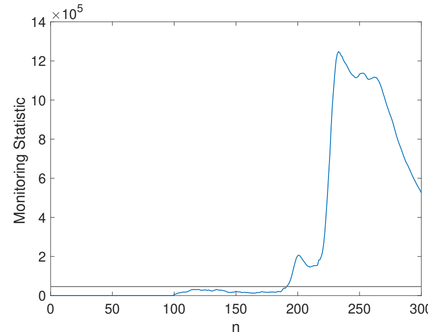

Figure 4: Our proposed global monitoring statistics

We then apply our proposed adaptive top-r procedure $T_\alpha(b)$ in (10) with $\alpha = 0.2$ on the residual images. The threshold $b = 46000$ is selected based on sampling with replacement of the first 100 in-control training data so that the average run length (ARL) is 2500. The Monte Carlo simulation is conducted with 100 repetitions. Figure 4 shows the global monitoring statistic of our proposed method, i.e., the sum of the top R_n CUSUM statistics over time. The horizontal line in the plot shows the threshold $b = 46000$. Based on figure 4, we will raise an alarm and detect the first solar flare at the time around $n = 192$. This is comparable to the results in Xie et al. (2012) and Liu et al. (2015).

6.2 Profile monitoring and fault detection

In this section, we apply our proposed adaptive top-r method to a real dataset that includes functional profile data from a progressive forming process (Lei et al., 2010). The data contains 307 profiles data from the normal process and 69 profiles data under five different faults of the process. Each profile data include $2^{11} = 2028$ measurement points. Figure 5 shows one profile and five different faulty profile data. We follow the literature (Zhou et al., 2006; Lei et al., 2010; Zhang et al., 2018) and use Haar Wavelets to exact features from the original nonlinear profile data. Then we will compare the detection performance of our proposed method with $T_{\alpha=0.1}(b)$ in (10), $N_{sum}(b)$ in (3), N_{max} in (4), and N_r with $r = 8$ and $r = 10$ to monitor the first 512 standardized Haar coefficients.

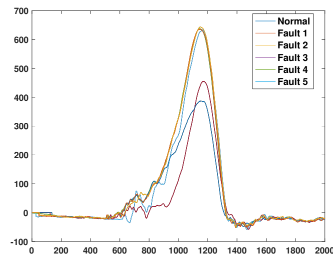

Figure 5: sample profiles selected from the normal profile data and the 5 groups of fault data

We find the threshold b for each method by sampling from the normal profiles so that the average run length (ARL) is 1000. With the determined threshold b, we apply each method to detect changes in anomalous profiles for each faulty group separately by conducting 100 simulations with sampling from their 69 fault profiles with replacement.

Table 6 shows the average detection delays with standard deviations inside the brackets for these different methods. We can see that the SUM scheme has the longest detection delay for all faults. Besides the SUM scheme, all other methods can detect the change in just one sample for fault 2 and fault 3. For Fault 1,4, and 5, our method can detect the change with the fewest samples.

Table 6: Detection delay for faults in each of the 5 operations

Method	Fault 1	Fault 2	Fault 3	Fault 4	Fault 5
$T_{\alpha=0.1}(b = 78.1)$	2(0)	1(0)	1(0)	5.01(0.001)	3(0)
$N_{sum}(b = 612.9)$	2.09(0)	1.24(0.004)	2(0)	6.12(0.008)	4.18(0.004)
$N_{max}(b = 15.6)$	2.3(0)	1(0)	1(0)	6.03(0.006)	3(0.002)
$N_{r=10}(b = 103.5)$	2(0)	1(0)	1(0)	5.59(0.006)	3.04(0.002)
$N_{r=8}(b = 88.55)$	2(0)	1(0)	1(0)	5.57(0.006)	3.07(0.003)

7 Conclusion

In this paper, we propose an adaptive scheme for monitoring large-scale data streams. The scheme estimates the changed data streams by using the idea of the BH false discovery rate control procedure, which is originally used to determine the rejected hypotheses in multiple hypothesis testing problems. We theoretically prove that this method is adaptive to the unknown number of changed data streams. We further extend this adaptive top-r method to address the problem of detecting an unknown mean shift. Simulation and case studies show that our proposed method is efficient in detecting change and can be easily implemented in real-time. Further study can focus on investigating the false alarm rate and the worst-case detection delay of our proposed adaptive top-r monitoring.

Acknowledgments

The authors greatly thank the editor, the associate editor, and anonymous referees for many constructive comments and suggestions, which improved the quality of the paper greatly. Dr. Zhang's research is supported in part by NSF grant ECCS-2236565.

Notes on contributors

Shuchen Cao received her PhD degree in Statistics from the University of Nebraska-Lincoln in 2023, and her bachelor's degree of science in mathematics from the University of Illinois at Urbana-Champaign. She is currently working as a statistical scientist at Bayer. Her research interests include sequential hypothesis testing, quickest change detection, experiment design, and deep learning models.

Ruizhi Zhang received his B.S. degree in Mathematics from the University of Science and Technology of China in 2014, and his Ph.D. degree in industrial and systems engineering from Georgia Institute of Technology in 2019. He is currently an assistant professor in the Department of Statistics at the University of Georgia, Athens, USA. His research interests include change-point detection, sequential analysis, robust statistics, and high-dimensional statistical inference.

Appendix

Proof of Theorem 1. First we prove the result of $\mathbf{E}_0(R_n)$ in (11). Note

$$\mathbf{E}_0(R_n) = \mathbf{P}_0(R_n \geq 1) + \sum_{i=2}^{K} \mathbf{P}_0(R_n \geq i). \tag{18}$$

By the definition of R_n in (9), we have $\mathbf{P}_0(R_n \geq 1) = 1$. Thus, we just need to prove that $\sum_{i=2}^{K} \mathbf{P}_0(R_n \geq i) < \frac{\alpha}{(1-\alpha)^2}[\log(K-1) - \alpha + 1]$.

Note for fixed $2 \leq i \leq K$, we have

$$\mathbf{P}_0(R_n \geq i) = \mathbf{P}_0(\bigcap_{r=1}^{i-1}\{p_{(r),n} < \frac{r}{K}\alpha\}) \leq \mathbf{P}_0(p_{(i-1),n} < \frac{i-1}{K}\alpha). \tag{19}$$

Let Y be the number of elements in $\{p_{j,n} : p_{j,n} < \frac{i-1}{K}\alpha, j = 1, .., K\}$, then $Y \sim \text{Bin}(K, q)$, where $q = \mathbf{P}_0\{p_{j,n} < \frac{i-1}{K}\alpha\}$. Then by (19), we have

$$\mathbf{P}_0(R_n \geq i) \leq \mathbf{P}_0(p_{(i-1),n} < \frac{i-1}{K}\alpha) = \mathbf{P}(Y \geq i-1). \tag{20}$$

By the property of CUSUM: $\mathbf{P}_0(W_{k,n} \geq x) \leq e^{-x}$ for any $x \geq 0$, we have

$$q = \mathbf{P}_0\big(p_{j,n} < \frac{i-1}{K}\alpha\big) = \mathbf{P}_0\big(W_{j,n} > \log\frac{K}{(i-1)\alpha}\big) \leq \frac{(i-1)}{K}\alpha. \tag{21}$$

Note when $\alpha \leq \frac{1}{2}$, $\frac{Kq(1-q)}{(i-1-Kq)^2}$ is an increasing function of q. Thus, from (21) and Chebyshev's inequality we have

$$\begin{aligned}
\mathbf{P}(Y \geq i-1) &\leq \mathbf{P}(|Y - Kq| \geq i-1-Kq) \leq \frac{Kq(1-q)}{(i-1-Kq)^2} \leq \frac{(i-1)\alpha[K-(i-1)\alpha]}{K[i-1-(i-1)\alpha]^2} \\
&= \frac{\alpha[K-(i-1)\alpha]}{K(i-1)(1-\alpha)^2} = \frac{\alpha}{(1-\alpha)^2}\frac{1}{i-1} - \frac{\alpha^2}{K(1-\alpha)^2}.
\end{aligned} \tag{22}$$

Therefore,

$$\begin{aligned}
\sum_{i=2}^{K}\mathbf{P}(R_n \geq i) &< \sum_{i=2}^{K}\frac{\alpha}{(1-\alpha)^2}\frac{1}{i-1} - \frac{\alpha^2}{K(1-\alpha)^2} < \frac{\alpha}{(1-\alpha)^2}[\log(K-1)+1] - \frac{\alpha^2}{(1-\alpha)^2} \\
&< \frac{\alpha}{(1-\alpha)^2}[\log(K-1) - \alpha + 1],
\end{aligned} \tag{23}$$

which completes the proof of (11).

Next, we prove the results of $\mathbf{E}_m(R_n)$ in (12). Note that

$$\mathbf{E}_m(R_n) = \sum_{i=1}^{m+1}\mathbf{P}_m(R_n \geq i) + \sum_{i=m+2}^{K}\mathbf{P}_m(R_n \geq i). \tag{24}$$

Similar to (19), when $2 \le i \le m+1$, we have

$$\mathbf{P}_m(R_n \ge i) = \mathbf{P}_m(\bigcap_{r=1}^{i-1}\{p_{(r),n} < \frac{r}{K}\alpha\}) \ge \mathbf{P}_m(p_{(i-1),n} < \frac{1}{K}\alpha). \tag{25}$$

Without loss of generality, assume the first m data streams are changed. Clearly, we have

$$\mathbf{P}_m(p_{(i-1),n} < \frac{1}{K}\alpha) \ge \prod_{j=1}^{m}\mathbf{P}_m(p_{j,n} < \frac{1}{K}\alpha). \tag{26}$$

Let $Z_{j,t} = \log\frac{f_1(X_{j,t})}{f_0(X_{j,t})}$ be the log-likelihood ratio of data $X_{j,t}$ with mean $\mu = \int f_1(x)\log\frac{f_1(x)}{f_0(x)}dx$ and variance $\sigma^2 = \int f_1(x)(\log\frac{f_1(x)}{f_0(x)})^2dx - \mu^2$. For the first m data streams, i.e., $1 \le j \le m$, let $\bar{Z}_{j,t} = \frac{1}{t}(\sum_{i=1}^{t} Z_{j,t})$. Then by Chebyshev's inequality, when $\alpha > \frac{K}{e^{n\mu}}$, we have

$$\begin{aligned}\mathbf{P}_m(p_{j,n} < \frac{1}{K}\alpha) = \mathbf{P}_m(W_{j,n} > \log(\frac{K}{\alpha})) \ge \mathbf{P}_m(\sum_{t=1}^{n} Z_{j,t} > \log(\frac{K}{\alpha})) = \mathbf{P}(\bar{Z}_{j,t} > \frac{1}{n}\log(\frac{K}{\alpha}))\\ > 1 - \mathbf{P}(|\bar{Z}_{j,t} - \mu| \ge \mu - \frac{1}{n}\log(\frac{K}{\alpha})) > 1 - \frac{\sigma^2}{n[\mu - \frac{1}{n}\log(\frac{K}{\alpha})]^2}\end{aligned} \tag{27}$$

Thus, we have

$$\begin{aligned}\left(1 - \frac{\sigma^2}{n[\frac{1}{n}\log(\frac{K}{\alpha}) - \mu]^2}\right)(m+1) &\le \sum_{i=1}^{m+1}\mathbf{P}_m(R_n \ge i)\\ &\le m+1.\end{aligned} \tag{28}$$

Moreover, for $m+2 \le i \le K$, by (19), we have

$$\mathbf{P}_m(R_n \ge i) \le \mathbf{P}_m(p_{(i-1),n} < \frac{i-1}{K}\alpha). \tag{29}$$

Note by (27), for $1 \le j \le m$, as $n \to \infty$, $\mathbf{P}_m(p_{j,n} < \frac{1}{K}\alpha) \to 1$. Let Y be the number of elements in $\{p_{j,n} : p_{j,n} < \frac{i-1}{K}\alpha, j = m+1, .., K\}$. Then, we have

$$\mathbf{P}_m(p_{(i-1),n} < \frac{i-1}{K}\alpha) = \mathbf{P}(Y \ge i-m-1). \tag{30}$$

Since $Y \sim \text{Bin}(K-m, q)$, where $q = \mathbf{P}_0\{p_{j,n} < \frac{i-1}{K}\alpha\}$, using the Chebyshev's inequality and the similar approach as in (22), for $m+2 \le i \le K$, we have

$$\begin{aligned}
&\mathbf{P}(Y \ge i-m-1) = \mathbf{P}(Y-(K-m)q \ge i-m-1-(K-m)q) \\
\le&\frac{(K-m)q(1-q)}{[i-m-1-(K-m)q]^2} = \frac{\frac{K-m}{K}(i-1)\alpha(1-\frac{i-1}{K}\alpha)}{[i-m-1-\frac{K-m}{K}(i-1)\alpha]^2} \\
<&\frac{\frac{K-m}{K}(i-1)\alpha}{[i-m-1-\frac{K-m}{K}(i-1)\alpha]^2} < \frac{\frac{K-m}{K}(i-1)\alpha}{[(i-m-1)(1-(m+1)(\frac{K-m}{K}\alpha)]^2} \\
=&\frac{i-1}{(i-m-1)^2}\beta < (i-1)\beta,
\end{aligned} \tag{31}$$

where

$$\beta = \frac{(K-m)\alpha/K}{[1-(m+1)(\frac{K-m}{K})\alpha]^2} > 0.$$

Therefore,

$$\sum_{i=m+2}^{K} \mathbf{P}_m(R_n \ge i) \le \sum_{i=m+2}^{K} (i-1)\beta \le \frac{(m+K)(K-m-1)}{2}\beta. \tag{32}$$

Combining (24),(28), and (32), we have

$$\left(1 - \frac{\sigma^2}{n[\frac{1}{n}\log(\frac{K}{\alpha}) - \mu]^2}\right)(m+1) \le \mathbf{E}_m(R_n) < m+1+\frac{(m+K)(K-m-1)}{2}\beta, \tag{33}$$

which completes the proof of (12). □

References

Apley, D. W. and F. Tsung (2002). The autoregressive t 2 chart for monitoring univariate autocorrelated processes. *Journal of Quality Technology 34*(1), 80–96.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological) 57*(1), 289–300.

Benjamini, Y. and W. Liu (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *Journal of Statistical Planning and Inference 82*(1), 163–170.

Chan, H. P. (2017). Optimal sequential detection in multi-stream data. *The Annals of Statistics 45*(6), 2736–2763.

Chen, J., W. Zhang, and H. V. Poor (2020). A false discovery rate oriented approach to parallel sequential change detection problems. *IEEE Transactions on Signal Processing 68*, 1823–1836.

Fellouris, G. and G. Sokolov (2016). Second-order asymptotic optimality in multisensor sequential change detection. *IEEE Transactions on Information Theory 62*(6), 3662–3675.

Gandy, A. and F. D.-H. Lau (2013). Non-restarting cumulative sum charts and control of the false discovery rate. *Biometrika 100*(1), 261–268.

Gavrilov, Y., Y. Benjamini, and S. K. Sarkar (2009). An adaptive step-down procedure with proven fdr control under independence. *The Annals of Statistics 37*(2), 619–629.

Grigg, O. A. and D. J. Spiegelhalter (2008). An empirical approximation to the null unbounded steady-state distribution of the cumulative sum statistic. *Technometrics 50*(4), 501–511.

Jin, J. and J. Shi (1999). Feature-preserving data compression of stamping tonnage information using wavelets. *Technometrics 41*(4), 327–339.

Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society: Series B (Methodological) 57*(4), 613–644.

Lei, Y., Z. Zhang, and J. Jin (2010). Automatic tonnage monitoring for missing part detection in multi-operation forging processes. *Journal of manufacturing science and engineering 132*(5).

Li, Y. and F. Tsung (2009). False discovery rate-adjusted charting schemes for multistage process monitoring and fault identification. *Technometrics 51*(2), 186–205.

Liu, K., Y. Mei, and J. Shi (2015). An adaptive sampling strategy for online high-dimensional process monitoring. *Technometrics 57*(3), 305–319.

Liu, K., R. Zhang, and Y. Mei (2019). Scalable sum-shrinkage schemes for distributed monitoring large-scale data streams. *Statistica Sinica 29*(1), 1–22.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics 42*(6), 1897–1908.

Lorden, G. and M. Pollak (2008). Sequential change-point detection procedures that are nearly optimal and computationally simple. *Sequential Analysis 27*(4), 476–512.

Mei, Y. (2010). Efficient scalable schemes for monitoring a large number of data streams. *Biometrika 97*(2), 419–433.

Mei, Y. (2011). Quickest detection in censoring sensor networks. In *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 2148–2152. IEEE.

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions. *The Annals of Statistics 14*(4), 1379–1387.

Nika, V., P. Babyn, and H. Zhu (2015). Change detection of medical images for three dimensional volumetric data. *Journal of Theoretical and Computational Science 2*, 135.

Oberti, F., L. Marcenaro, and C. S. Regazzoni (2002). Real-time change detection methods for video-surveillance systems with mobile camera. In *2002 11th European Signal Processing Conference*, pp. 1–4.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika 41*(1/2), 100–115.

Paynabar, K., C. Zou, and P. Qiu (2016). A change-point approach for phase-i analysis in multivariate profile monitoring and diagnosis. *Technometrics 58*(2), 191–204.

Pollak, M. (1985). Optimal detection of a change in distribution. *The Annals of Statistics 13*(1), 206–227.

Polunchenko, A. S., A. G. Tartakovsky, and N. Mukhopadhyay (2012). Nearly optimal change-point detection with an application to cybersecurity. *Sequential Analysis 31*(3), 409–435.

Qiu, P. (2020). Big data? statistical process control can help! *The American Statistician 74*(4), 329–344.

Qiu, P., W. Li, and J. Li (2020). A new process control chart for monitoring short-range serially correlated data. *Technometrics 62*(1), 71–83.

Shiryaev, A. N. (1963). On optimum methods in quickest detection problems. *Theory of Probability & Its Applications 8*(1), 22–46.

Siegmund, D. (2013). *Sequential analysis: tests and confidence intervals*. Springer Science & Business Media.

Tartakovsky, A. G., A. S. Polunchenko, and G. Sokolov (2013). Efficient computer network anomaly detection by changepoint detection methods. *IEEE Journal of Selected Topics in Signal Processing 7*(1), 4–11.

Tartakovsky, A. G., B. L. Rozovskii, R. B. Blažek, and H. Kim (2006). Detection of intrusions in information systems by sequential change-point methods. *Statistical Methodology 3*(3), 252–293.

Woodall, W. H. and D. C. Montgomery (2014). Some current directions in the theory and application of statistical process monitoring. *Journal of Quality Technology 46*(1), 78–94.

Xian, X., C. Zhang, S. Bonk, and K. Liu (2021). Online monitoring of big data streams: A rank-based sampling algorithm by data augmentation. *Journal of Quality Technology 53*(2), 135–153.

Xie, Y., J. Huang, and R. Willett (2012). Change-point detection for high-dimensional time series with missing data. *IEEE Journal of Selected Topics in Signal Processing 7*(1), 12–27.

Xie, Y. and D. Siegmund (2013). Sequential multi-sensor change-point detection. *The Annals of Statistics 41*(2), 670–692.

Xiong, W. and J. C.-M. Lee (1998). Efficient scene change detection and camera motion annotation for video classification. *Computer Vision and Image Understanding 71*(2), 166–181.

Yan, H., K. Paynabar, and J. Shi (2018). Real-time monitoring of high-dimensional functional data streams via spatio-temporal smooth sparse decomposition. *Technometrics 60*(2), 181–197.

Zhang, R. and Y. Mei (2018). Asymptotic statistical properties of communication-efficient quickest detection schemes in sensor networks. *Sequential Analysis 37*(3), 375–396.

Zhang, R., Y. Mei, and J. Shi (2018). *Wavelet-Based Profile Monitoring Using Order-Thresholding Recursive CUSUM Schemes*, pp. 141–159. Cham: Springer International Publishing.

Zhang, W. and Y. Mei (2023). Bandit change-point detection for real-time monitoring high-dimensional data under sampling control. *Technometrics 65*(1), 33–43.

Zhou, S., B. Sun, and J. Shi (2006). An spc monitoring system for cycle-based waveform signals using haar transform. *IEEE Transactions on Automation Science and Engineering 3*(1), 60–72.

Zou, C., Z. Wang, X. Zi, and W. Jiang (2015). An efficient online monitoring method for high-dimensional data streams. *Technometrics 57*(3), 374–387.