A change-point–based control chart for detecting sparse mean changes in high-dimensional heteroscedastic data

Zezhong Wang and Inez Maria Zwetsloot
Department of Systems Engineering, City University of Hong Kong, Kowloon, Hong Kong

ABSTRACT
Because of the “curse of dimensionality,” high-dimensional processes present challenges to traditional multivariate statistical process monitoring (SPM) techniques. In addition, the unknown underlying distribution of and complicated dependency among variables such as heteroscedasticity increase the uncertainty of estimated parameters and decrease the effectiveness of control charts. In addition, the requirement of sufficient reference samples limits the application of traditional charts in high-dimension, low-sample-size scenarios (small n, large p). More difficulties appear when detecting and diagnosing abnormal behaviors caused by a small set of variables (i.e., sparse changes). In this article, we propose two change-point–based control charts to detect sparse shifts in the mean vector of high-dimensional heteroscedastic processes. Our proposed methods can start monitoring when the number of observations is a lot smaller than the dimensionality. The simulation results show that the proposed methods are robust to nonnormality and heteroscedasticity. Two real data examples are used to illustrate the effectiveness of the proposed control charts in high-dimensional applications. The R codes are provided online.

1. Introduction

In modern manufacturing systems, widely used sensors and internet technologies make it possible to collect hundreds of measurements related to the final product and its production over time. These measurements can be used to evaluate process conditions and the necessity of corrective actions. Statistical process monitoring (SPM) tools for such applications in high-dimensional processes are becoming prevalent. Detecting abnormality is a challenging task for traditional multivariate control charts, such as the Hotelling’s $T^2$ control chart, because the “curse of dimensionality” affects the parameter estimates and signal detection ability of multivariate control charts significantly (Hastie, Tibshirani, and Friedman 2009).

The underlying distribution and parameters of a target process are crucial factors for setting up control charts. These are usually unknown and need to be estimated based on in-control (IC) data in a Phase I analysis. The uncertainty of estimation grows rapidly with dimensionality, especially of the estimated covariance matrix (Zhang, Chen, and Wu 2020). One overlooked characteristic, which can also affect the accuracy of parameter estimates in SPM solutions for high-dimensional data, is heteroscedasticity (Hong, Balzano, and Fessler 2018). We define heteroscedasticity as the inequality of error variance over time. In some applications, heteroscedasticity is regarded as undesirable and needs to be detected. On the contrary, we focus on situations in which heteroscedasticity is an intrinsic part of the stable process. Therefore, our aim is to monitor for changes in the mean, given the heteroscedastic nature of the process.

For example, heteroscedasticity is an intrinsic aspect of financial data in financial surveillance. It is explicitly modeled by autoregressive conditional heteroscedastic (ARCH) models (Engle 1982) or generalized autoregressive conditional heterokedastic (GARCH) models (Bollerslev 1986). Schipper and Schmid (2001) and Bodnar (2009) used univariate and multivariate GARCH models to develop dispersion control charts for monitoring shifts in the variance. It is challenging to apply the same strategy to high-
dimensional scenarios because of the computational complexity (Frisén 2008).

Heteroscedasticity is also an inherent property of data whenever data collection procedures are affected by environmental or dynamic external factors that are difficult to remove. For example, astronomical measurements of stars taken at various times are affected by the changing atmospheric effects, which cause the amount of noise to vary across time (Tamuz, Mazeh, and Zucker 2005). Another example of intrinsic heteroscedasticity is the residual of dengue-related data with unequal variance in different months, displayed in Figure 1(a) (Wang and Zwetsloot 2019). Figure 1(b) shows the heteroscedasticity of a vibration index collected by sensors when the system functions normally (Wang, Goedhart, and Zwetsloot 2023).

In this article, we consider heteroscedasticity an inherent characteristic of the process and wish to detect changes in the location parameter. We therefore treat the different variances over time as a common cause variability in the processes. All variables share the same heteroscedasticity. We focus on robust methods to maintain consistent performance under dynamic covariance matrices for monitoring high-dimensional heteroscedastic data.

Another challenge is detecting sparse shifts in high-dimensional data. Dimension reduction is one solution for monitoring high-dimensional processes. Principal components analysis (PCA) is widely used in monitoring high-dimensional time series as an efficient dimension reduction tool. The PCA-based control charts first extract the latent variables from the original data and then monitor the derived statistics. An excellent overview is given by De Ketelaere, Hubert, and Schmitt (2015). It is rare that all variables change simultaneously in a process. More often, in high-dimensional processes, the shift is sparse; that is, a small subset of variables is likely to cause the abnormality. Because of the curse of dimensionality, the out-of-control (OC) behaviors caused by sparse changes are more likely to be buried in the noise (Shu and Fan 2018).

The capability of PCA-based methods to detect sparse shifts is questionable because the extracted latent variables may overlook important data features. After successfully detecting an alarm, it may also be difficult to locate and identify the assignable causes because the interpretation of the derivative statistic may be meaningless (Wang and Jiang 2009). Ebrahimi, Ranjan, and Paynabar (2022) proposed an adaptive PCA-based SPC framework, adaptive PC selection (APC), for high-dimensional data streams. This method is valid for monitoring sparse shifts, and a root-cause diagnostics stage is proposed to determine the shifted variables. However, its performance decreases for nonnormal data.

Combining variable selection algorithms and control charts is an alternative for detecting sparse changes. Zou and Qiu (2009) proposed a LEWMA chart by combining a least absolute shrinkage and selection operator (LASSO) test statistic with a MEWMA chart. Independently, Wang and Jiang (2009) proposed a VS-MSPC control chart, which first combined a forward variables selection (FVS) algorithm with a multivariate Shewhart chart. Jiang, Wang, and Tsung (2012) and Abdella et al. (2017) extended the FVS algorithm to the VS-EWMA chart and the VS-CUSUM chart, respectively. One bonus associated with these variable selection–based methods is that the tasks of detection and diagnosis are naturally integrated and conveniently and simultaneously resolved because all the selected variables are assumed to be responsible for the OC signals (Jiang, Wang, and Tsung 2012).

The multivariate normality assumption limits the application of the aforementioned variable selection control charts, and heteroscedasticity is a potential factor that affects their performance. Robust control charts are preferable, and nonparametric methods are
appropriate when the underlying process distribution is nonnormal and/or challenging to verify (Ebadi et al. 2022). Zou, Wang, and Tsung (2012), Holland and Hawkins (2014), Huang, Kong, and Huang (2014), Chen, Zi, and Zou (2016), Shu and Fan (2018), and Zhang, Chen, and Wu (2020) developed different distribution-free methods based on the change-point model. These methods are also appropriate for monitoring the location shifts when the number of variables is much greater than the number of individual observations $(p > n)$. This is also referred to as a high-dimension, low-sample-size (HDLSS) scenario. The change-point control chart proposed by Li et al. (2014) can detect mean changes at the beginning of a process and identify corresponding assignable causes. It is robust to nonnormality, as shown in their simulation results. Change-point–based methods can support practitioners in identifying root causes. Huang, Kong, and Huang (2014) discussed the change-point estimate algorithm in the post-signal diagnosis phase. For a systematic review of change-point estimation in SPC, refer to Amiri and Allahyari (2012).

To overcome the three challenges of monitoring high-dimensional processes—(1) heteroscedasticity, (2) sparsity of the change, and (3) parameter estimation when $p > n$—we propose change-point–based control charts for detecting sparse mean shifts in high-dimensional heteroscedastic data. We consider abrupt and sustained shifts. The change-point methods allow us to deal naturally with the third challenge. We apply supremum-based tests to deal with the first and second challenges. The proposed methods, called the $RH$-$CP$ chart and the $RH$-$SCP$ chart, where $RH$ stands for robust to heteroscedasticity, are valid in HDLSS scenarios. Another attractive feature is their robustness to nonnormality. A data-driven method is used to obtain control limits. We propose a post-signal diagnosis method to estimate the change-point and assignable causes.

The structure of this article is as follows. In Section 2, we propose our novel change-point–based control charts, the $RH$-$CP$ chart and the $RH$-$SCP$ chart, for detecting sparse shifts in the mean vector of possibly heteroscedastic nonnormal data. Section 3 shows the performance of proposed methods and a comparison with another method. In Section 4, we propose a post-signal diagnosis procedure. In Section 5, two examples are used to showcase the proposed methods. Section 6 concludes the article.

### 2. Change-point–based control chart

#### 2.1. Change-point–based monitoring scheme

Given a series of independent and individual $p$-dimensional observations $X_i = (X_{i1}, X_{i2}, ..., X_{ip})'$, $i = 1, 2, ..., n$, the change-point model is

$$H_0 : X_i \sim F_p(\mu_0, \Sigma_i) \text{ and } H_1 : X_i \sim \left\{ \begin{array}{ll} F_p(\mu_0, \Sigma_i) & , \quad i \leq \tau, \\ F_p(\mu_1, \Sigma_i) & , \quad i > \tau, \end{array} \right. $$

where $F_p^0$ and $F_p^1$ are the IC and OC distribution functions, $\mu_0$ and $\mu_1$ are the corresponding unknown IC and OC mean vectors, and $\Sigma_i$ is the IC covariance matrix, which changes over time. The change-point $\tau$, which indicates the process changes after $X_\tau$, is also unknown and needs to be estimated.

Various test statistics for mean vectors have been adapted to high-dimensional process monitoring. Zou et al. (2015) developed a new control chart based on the goodness-of-fit test proposed by Zhang (2002). Multivariate distribution-free tests proposed by Randles (2000), Choi and Marden (1997), and Bickel (1969) have been adapted to monitoring schemes by Zou and Tsung (2011), Holland and Hawkins (2014), and Chen, Zi, and Zou (2016), respectively. Li et al. (2014) extended the high-dimensional two-sample test proposed by Chen and Qin (2010) to a self-starting control chart. Shu and Fan (2018) developed a nonparametric control chart based on the high-dimensional one-sample test proposed by Saha, Sarkar, and Ghosh (2017).

High-dimensional test statistics can be categorized as sum-of-squares–based test statistics and supremum-based test statistics. The former is powerful when there are many small differences between $\mu_0$ and $\mu_1$—in other words, when the signals are dense but weak (Chen and Qin 2010). However, this type of method could be ineffective under the sparsity assumption, as the accumulation of all differences will not be significantly influenced by a few large differences, especially in a big data stream (Gregory et al. 2015). For the case in which a few variables have large shifts and the signals are sparse but strong, a supremum-based test statistic may perform better. To detect the sparse changes effectively, we proposed a control chart based on the supremum-based test statistics modified from Chang et al. (2017).

Chang et al. (2017) proposed two tests for location parameters—a nonstudentized test and a studentized test—both of which are valid in the HDLSS scenario. A fully data-driven Monte Carlo simulation is proposed to compute the critical values. According to
their simulation results, both tests are robust against nonnormality. They loosen the assumption of the covariance matrix, so the two test samples can be drawn from distributions with different covariance matrices. Both tests are potentially robust for testing high-dimensional means under heteroscedasticity. When the sample size is small, the nonstudentized test can maintain nominal significance; hence, it is recommended for practical application. Standardization is necessary to reduce the impact of unequal variances before applying the nonstudentized test. The studentized method involves sample variance estimation and is recommended for a larger sample size because its performance depends on accurate and precise variance estimates.

According to Chang et al. (2017), their tests based on very large dimensions might result in large critical values. They added a preliminary screening step for dimension reduction to enhance its power before conducting the tests. The screening procedure requires static sample sizes, which is violated in process monitoring and the change-point model. Consequently, we take the nonstudentized and studentized tests and modify them to be applicable in an online monitoring setting.

To test for \( H_0 \) vs. \( H_1 \) in Eq. [1], we partition the \( n \) observations into two sets, \( \{X_1, \ldots, X_k\} \) and \( \{X_{k+1}, \ldots, X_n\} \) at a split point \( k(3 \leq k \leq n - 3) \). Each subset contains at least three observations in order to compute the sample mean and variance. Hence, this method can start monitoring with at least six observations, making it appropriate for monitoring short-run processes. As in Chang et al. (2017), the nonstudentized statistics with \( n \) observations and split point \( k \) are defined as follows:

\[
T^{(1)}_{n,k} = \max_{1 \leq r \leq p} \frac{\sqrt{k(n-k)}|X_{k,r} - \overline{X}_{n-k,r}|}{\sqrt{k\hat{\sigma}^2_{k,r} + (n-k)\hat{\sigma}^2_{n-k,r}}}, \quad [2]
\]

where \( \overline{X}_{k,r} = \frac{1}{k} \sum_{i=1}^{k} X_{i,r} \) and \( \overline{X}_{n-k,r} = \frac{1}{n-k} \sum_{i=k+1}^{n} X_{i,r} \).

The monitoring statistic at time point \( n \), referred to as \( U^{(1)}_n \), is the maximum value of \( T^{(1)}_{n,k} \) over all split points \( k \):

\[
U^{(1)}_n = \max_{3 \leq k \leq n-3} T^{(1)}_{n,k}. \quad [3]
\]

We plot \( U^{(1)}_n \) on a chart with a limit \( h^{(1)}_p \) and signal when \( U^{(1)}_n > h^{(1)}_p \). The change-point \( \tau \) can be estimated directly by the corresponding \( k \) when \( U^{(1)}_n \) exceeds the control limit \( h^{(1)}_p \):

\[
\hat{\tau}^{(1)}_n = \arg \max_{3 \leq k \leq n-3} (T^{(1)}_{n,k} > h^{(1)}_p). \quad [4]
\]

By standardizing Eq. [2] with sample variances, we get the monitoring statistics and change-point estimation based on studentized test statistics.

\[
T^{(2)}_{n,k} = \max_{1 \leq r \leq p} \frac{\sqrt{k(n-k)}|X_{k,r} - \overline{X}_{n-k,r}|}{\sqrt{k\hat{\sigma}^2_{k,r} + (n-k)\hat{\sigma}^2_{n-k,r}}}, \quad [5]
\]

where \( \overline{X}_{k,r} \) and \( \overline{X}_{n-k,r} \) are the same as in Eq. [2].

\[
\hat{\sigma}^2_{k,r} = \frac{1}{k-1} \sum_{i=1}^{k} (X_{i,r} - \overline{X}_{k,r})^2 \quad \text{and} \quad \hat{\sigma}^2_{n-k,r} = \frac{1}{n-k-1} \sum_{i=k+1}^{n} (X_{i,r} - \overline{X}_{n-k,r})^2
\]

\((n - k)\) are the pre-shift sample variance and the post-shift sample variance of variable \( r \). Furthermore, the monitoring statistic equals

\[
U^{(2)}_n = \max_{3 \leq k \leq n-3} T^{(2)}_{n,k}. \quad [6]
\]

The change-point \( \tau \) can be estimated directly by the corresponding \( k \) when \( U^{(2)}_n \) exceeds the control limit:

\[
\hat{\tau}^{(2)}_n = \arg \max_{3 \leq k \leq n-3} (T^{(2)}_{n,k} > h^{(2)}_p). \quad [7]
\]

### 2.2. Moving window

Note that for progressive online monitoring, our statistics will be based on larger samples (i.e., \( n \) increases with time). The efficiency of detecting sparse shifts decreases with the sample size (\( n \)) because the signals are more likely to be buried by noise from the data. Another practical problem with the method is that the time and complexity of data processing multiply with incoming observations; \( n - 5 \) iterations are needed to compute the statistic from \( n \) observations. A third issue is that the sample size influences the statistics; therefore, the control limits are dynamic. To shorten the detection delay and promote the sensitivity of the proposed monitoring schemes, we have added a moving window to the proposed methods.

For a fixed window size \( W \), the monitoring procedure starts with the first \( W \) observations (\( X_1, \ldots, X_W \)). After collecting \( s \) new observations (\( s \) is the step size), the oldest \( s \) observations are excluded and the procedure moves to the second window (\( X_{1+s}, \ldots, X_{W+s} \)). An appropriate step size can speed up the computation and reduce the serial correlation, making the monitoring statistics approximately independent. The observation matrix at time point \( n(n > W) \) is

\[
X_n = (X_{n-W+1}, X_{n-W+2}, \ldots, X_n).
\]

The window-based nonstudentized statistics are

\[
T^{(1)}_{n,W,k} = \max_{1 \leq r \leq p} \frac{\sqrt{k(W-k)(X_{W-k,r} - \overline{X}_{n-k,r})}}{\sqrt{W}}, \quad [8]
\]
where \( k^*(3 \leq k^* \leq W - 3) \) is the split point inside the current window and \( \bar{X}_{t-r,r} \) is the mean of the first \( k^* \) observations in window \( W_n \). The corresponding charting statistic based on this window is

\[
U_{n,W}^{(1)} = \max_{3 \leq k^* \leq W-3} T_{n,W,k^*}^{(1)}.
\]  

Note that we observe \( U_{n,W}^{(1)} \) at each \( s \) time point only. Estimating the change-point is also straightforward. If the control chart signals at time point \( n \), the corresponding change-point estimate is

\[
\hat{\tau}_{n,W}^{(1)} = n - W + \arg\max_{3 \leq k^* \leq W-3} \left( T_{n,W,k^*}^{(1)} | R_{n,W}^{(1)} > \hat{h}_{p,W}^{(1)} \right).
\]  

Similarly, the window-based studentized statistic is

\[
T_{n,W,k^*}^{(2)} = \max_{3 \leq k^* \leq W-3} \frac{\sqrt{k^*(W-k^*)} | \bar{X}_{t-r,r} - \bar{X}_{W-k^*-r}^{(1)} |}{\sqrt{k^* \sigma^2_{W-k^*}^{(1)} + (W-k^*) \sigma^2_{W-k^*-r}^{(1)}}},
\]  

where \( \sigma^2_{W-k^*}^{(1)} \) is the sample variance of first \( k^* \) observations in window \( W_n \) and \( \sigma^2_{W-k^*}^{(1)} \) is the sample variance of the rest of the \( W-k^* \) observations. The corresponding charting statistic based on this window is

\[
U_{n,W}^{(2)} = \max_{3 \leq k^* \leq W-3} T_{n,W,k^*}^{(2)}.
\]  

And the corresponding change-point estimate is

\[
\hat{\tau}_{n,W}^{(2)} = \arg\max_{3 \leq k^* \leq W-3} \left( T_{n,W,k^*}^{(2)} | R_{n,W}^{(2)} > \hat{h}_{p,W}^{(2)} \right) + n - W
\]  

The magnitude of studentized and nonstudentized monitoring statistics depends on the window size, which is a prespecified fixed value. Consequently, the control limits are constant values and need to be simulated for different dimensionality and window sizes. The RH-CP chart signals when \( U_{n,W}^{(1)} > \hat{h}_{p,W}^{(1)} \), where \( \hat{h}_{p,W}^{(1)} \) is the control limit. And the RH-SCP chart signals when \( U_{n,W}^{(2)} > \hat{h}_{p,W}^{(2)} \), where \( \hat{h}_{p,W}^{(2)} \) is the control limit.

The sample mean and variance are part of the statistics in Eqs. (8) and (11). Determining the window size is a tradeoff between estimation accuracy (large windows) and a quick start of the monitoring scheme (small windows). If the practitioner prefers sensitive methods, a large window is recommended.

### 2.3. Control limits

Chang et al. (2017) used Monte Carlo simulations to compute the critical values for the proposed maximum-type statistics. The empirical distribution of a set of i.i.d. Gaussian random vectors from \( N(0, \Sigma_{1,2}) \), where \( \Sigma_{1,2} = \frac{n}{m+n} \Sigma_1 + \frac{m}{m+n} \Sigma_2 \), is used to estimate the critical value. Estimation requires a known and fixed sample size, which is violated in our proposed methods. The distributions of \( U_{n,W}^{(1)} \) and \( U_{n,W}^{(2)} \) are complicated and unknown, so we use a data-driven bootstrap method to determine the control limits based on their empirical distributions. One bonus of the bootstrap method is that the effect from underlying heteroscedasticity is already incorporated.

The control limits are determined by setting a prespecified false alarm probability (FAP) instead of the ARL because the ARL is affected by the window size and step size directly. Moreover, simulation for ARLs is more time-consuming. FAP = \( \text{Pr}(RL \leq n|H_0) \) is the probability of at least one false alarm in the process from time 1 to \( n \) (Chakraborti, Human, and Graham 2008). Based on this, we used the bootstrap method in Algorithm 1 to compute the control limits.

**Algorithm 1: Bootstrap method for control limits**

1. Import \( B \), \( n \), \( W \), \( s \), and \( \mathbb{X}_p \), where \( B \) is the number of bootstrap samples, \( n \) is the monitoring interval, \( W \) is the window size, and \( s \) is the step size.
2. Draw \( B \) bootstrap samples of size \( W \) with replacement from \( \mathbb{X}_p \).
3. Use Eqs. (8) and (9) or Eqs. (11) and (12) to calculate the charting statistics \( U_{W,b}^{(1)} \), \( b = 1, 2, ..., B, l = 1, 2 \), for every bootstrap sample. An empirical distribution of \( U_{W,b}^{(1)} \) can be built as \( \hat{F}_W(t) = B^{-1} \sum_{b=1}^{B} I\{U_{W,b}^{(1)} \leq t\} \).
4. Given that FAP = \( \alpha \), the control limits can be estimated as \( \hat{h}_{p,W}^{(1)} = \inf \{t \in \mathbb{R} : \hat{F}_W(t) \geq \sqrt{1 - \frac{\alpha}{s}}\} \), where \( Q = \max \{z \in \mathbb{Z} | z \leq \left( \frac{n-W}{s} \right) \} + 1 \).

Step 4 in Algorithm 1 is valid when an appropriate step size, \( s \), can make \( \text{Pr}(U_{n,W} > \hat{h}_{p,W} | \text{no signal before } n, H_0) \leq 1 - \sqrt{1 - \frac{\alpha}{s}} \) in other words, the sequential \( U_{n,W}^{(1)} \) and \( U_{n,W}^{(2)} \) are approximately independent. A simple simulation is used to verify this statement. In each simulation run, 100 observations are generated from \( N_{50}(0, \Sigma_r) \). A sequence of \( U_{n,W}^{(1)} \) and their autocorrelations are calculated with \( s = \)}
1, 3, 5 and $W = 20, 30, 40$. Finally, the autocorrelations (from lag 1 to lag 10) are averaged over 100 replications. As shown in Figure 2, when $s = 5$, the autocorrelations are weak for all lags. The results are consistent with different dimensionality, and the same conclusion holds for $U_{n,W}^{(2)}$. Hence, we recommend using $s \geq 5$ for computational convenience and accuracy of the control limits. We also explore the relationship between the sample size of $X_p$ and the estimated control limits. For different scenarios with various $p$ and $W$, we recommend using at least 150 observations to achieve reliable control limits.

3. Performance study

In this section, we evaluate the performance of the RH-CP and RH-SCP charts and compare them with another method. We use the empirical FAP to evaluate IC performances and propose two new metrics to assess OC performances.

According to Frisén (2009), the choice of performance metrics should rely on the characteristics of surveillance methods. The probability of detection and the expected delay are valid for the change-point model. Here we use the proportion of runs that are stopped by a signal out of runs to compute the detecting power of the proposed methods. This metric is called the detection rate (DR), where

$$DR = \frac{\sum_{i=1}^{n} I(U_{n,W}^{(i)} \geq \beta_{n,W}(H_1))}{n}$$

for $l = 1, 2$. The proposed method is sensitive to changes when $DR$ is close to 1.

Timely detection is also important for a monitoring scheme. We define the conditionally expected detection delay (CED), which is the time between the average stop point and the change-point, as

$$CED = \frac{\sum_{i=1}^{n} \min(\arg_{t}(U_{n,W}^{(i)} \geq \beta_{n,W}(H_1)))}{n} - \tau.$$  

CED is influenced by $W$, $s$, and $\tau$ directly. When $W \leq \tau$, the smaller CED reflects better performance. When $W > \tau$ and $CED = W - \tau$, it means that the control chart can detect a signal as early as the first window. For more uses and discussion of DR and CED in SPM, see Knoth (2015), Liu et al. (2023), Yu et al. (2022), Alghuried and Moghaddass (2022), and Lee et al. (2022). We discuss the highlights of the simulation results below.

3.1. Comparison

As discussed in Section 1, Chen, Zi, and Zou (2016) proposed a distribution-free EWMA control chart (DFEWMA) based on the change-point model. It is efficient in detecting sparse shifts in location parameters under HDLSS scenarios. We select the DFEWMA control chart because its MATLAB code is available online. We realize that other methods could also be good comparisons, but either there is no code available (e.g., LEWMA chart and VS-CUSUM by Zou and Qiu (2009) and Abdella et al. (2017)) or the implementation is so slow that it is impractical to run the method (e.g., the method proposed by Li et al. (2014)). Hence, we focus on the DFEWMA chart because it is the only method that is comparable, has code available, and is fast to run.

One property of the DFEWMA method is that its IC run length always follows the geometric distribution. According to Theorem 1 in Chen, Zi, and Zou (2016), the FAP satisfies the following equation:

$$FAP = P(\text{RL} \leq n) = 1 - (1 - \beta)^n.$$  

$\beta$ is used to determine the control limits, which makes it comparable with our proposed methods. They also provided an equation for change-point estimation without further evaluation.
Robustness under heteroscedasticity is of interest to us, without a loss of generality, so we use $N_p(0, I_p)$ as a baseline model, with $p = 20, 50, 100$. To imitate and manipulate the heteroscedastic model $N_p(0, \Sigma_i)$, we multiply the constant covariance matrix by a time-dependent parameter $\lambda_i$. The heteroscedastic model is $N_p(0, \lambda_i I_p)$, where $\lambda_i$ is a set equal to \{0.5, 0.6, ..., 1.4, 1.5, 1.4, ..., 0.6\} and repeated until the end of the process. As a result, each variable has a unit variance on average within a prolonged time period. We set the window size as $W = 30$ for the three methods. The smoothing parameter in the DFEWMA chart is 0.1 according to the guidelines in Chen, Zi, and Zou (2016). The reference sample for DFEWMA consists of $m_0 = 100$ observations. For the proposed RH-CP chart and RH-SCP chart, we set $s = 5$ and use 500 observations to compute the control limits by $B = 10,000$ replications. The first $\nu = 25\%$ components in the location parameter change by $\delta = 1.5$ or 2 after $\tau = 25$ observations. The absolute shift sizes are fixed with dynamic variance. Therefore, when $\lambda_i = 1$, the shifts are $1.5\sigma_i$ and $2\sigma_i$. When $\lambda_i < 1$, the real shifts are relatively larger than $1.5\sigma_i$ and $2\sigma_i$. Symmetrically, the relevant shifts are smaller than $1.5\sigma_i$ and $2\sigma_i$ when $\lambda_i > 1$. All performance results in this section are obtained from 2,000 simulation runs unless indicated otherwise.

All methods are designed to achieve a $FAP = 0.01$ with $n = 100$ observations. The Empirical column in Table 1 shows the empirical $FAP$ values. Surprisingly, the empirical $FAP$ values of the DFEWMA chart are about five times larger than the designed values when the data follow a standard normal distribution. Chen, Zi, and Zou (2016) claimed that with a given $\beta$, the simulated $FAP = Pr(\text{RL} \leq n)$ of the DFEWMA chart is close to the theoretical one. One possible explanation of the inconsistency is that when $FAP = 0.01$, the corresponding $\beta$ is much smaller. The estimation of such a long-tail percentile is challenging. More variables and observations include more uncertainty, which may also cause the inflation of $FAP$.

Both the RH-CP chart and the RH-SCP chart can guarantee low $FAP$ values under baseline models. Under heteroscedasticity, the empirical $FAP$ of all three methods increases significantly. The simulated $FAP$ of the RH-SCP chart is about twice the predefined value, which is the most robust.

The incomparable $FAP$ values make the OC performance analysis meaningless. Therefore, we adjust the control limits of our proposed methods to get a similar empirical $FAP$ as the DFEWMA chart. The complementary results show that $DR$ positively correlates with $FAP$. With a large $FAP$, the $DR$ of the proposed methods is approximately equal to 1. The RH-CP chart has a larger $DR$ in all OC scenarios, and it is more sensitive than the RH-SCP chart. Heteroscedasticity reduces the $DR$ of the proposed methods, but the RH-CP chart is less affected.

The DFEWMA method shows a smaller $CED$ than the proposed methods in all scenarios, since it uses an increasing window size and moves with every new observation. When $\tau = 25$ and $W = 30$, the proposed methods have a theoretical minimum detection delay of five, and they can achieve the ideal $CED$ with a large $FAP$ and shift size. Overall, the maximum detection delay is about half the window size, which is reasonable for a change-point model. Heteroscedasticity reduces the $CED$ of the DFEWMA chart by 1; this is related to the large $FAP$ and increases the $CED$ of the proposed methods by a few observations. Another universal conclusion is that all three methods can achieve good $DR$ and fast $CED$ for all considered $p$.

Figure 3 shows the $DR$ and $CED$ curves with varying $\delta$ in heteroscedastic processes. Overall, with increasing $\delta$, the $DR$ of all methods converges to 1 and achieves a much smaller detection delay. Figure 3(a) shows the $DR$ with $\nu = 10\%$ under various

| Model       | Methods | Design | Empirical | DR | CED | DR | CED |
|-------------|---------|--------|-----------|----|-----|----|-----|
| $N_{50}(0, I_p)$ | DFEWMA | 0.01   | 0.042     | 1.00 | 6.08 | 1.00 | 4.95 |
|             | RH-CP   | 0.01   | 0.006     | 0.91 | 11.99 | 1.00 | 8.29 |
|             | RH-SCP  | 0.01   | 0.005     | 0.56 | 12.57 | 0.97 | 9.09 |
| $N_{50}(0, I_p)$ | RH-CP   | 0.05   | 0.031     | 0.99 | 8.44  | 1.00 | 5.41 |
|             | RH-SCP  | 0.05   | 0.046     | 0.87 | 11.14 | 1.00 | 7.06 |
| $N_{100}(0, I_p)$ | DFEWMA | 0.01   | 0.047     | 1.00 | 4.38  | 1.00 | 3.64 |
|             | RH-CP   | 0.01   | 0.009     | 0.99 | 10.46 | 1.00 | 6.39 |
|             | RH-SCP  | 0.01   | 0.008     | 0.75 | 11.86 | 1.00 | 7.86 |
| $N_{100}(0, I_p)$ | RH-CP   | 0.05   | 0.045     | 1.00 | 6.78  | 1.00 | 5.02 |
|             | RH-SCP  | 0.05   | 0.050     | 0.96 | 9.69  | 1.00 | 5.97 |

When $m_0 = 100$, $\nu = 25\%$, $\tau = 25$, $W = 30$. |
dimensional models for our proposed methods. When \( \delta > 2 \), the proposed RH-CP and RH-SCP charts can achieve \( DR = 1 \) approximately. In higher-dimensional scenarios such as \( p = 100 \), the detection power grows faster because of more shifted variables. The difference between Figure 3(a) and (b) is sparsity, and our
proposed methods converge faster with more shifted variables caused by lower sparsity levels. Figure 3(c) and (d) illustrate similar conclusions; with more shifted variables, the CED values for both methods decrease faster and converge to the minimum detection delay of five when \( \delta \) is large.

In Table 1, the DFEWMA chart outperforms the proposed methods. One reasonable explanation for this result is the reference sample, as the DFEWMA chart compares new observations with IC data points. To make a fair comparison and to illustrate the improvement of a large window, we use \( W = 100 \) for our methods to detect changes at \( \tau = 125 \) in heteroscedastic models with \( p = 50 \) and 100. For the DFEWMA chart, we use \( W = 30 \) and \( \tau = 25 \), since it has a reference sample with size \( m_0 = 100 \). Figure 3(e) and (f) plot the DR curves for the RH-CP chart, RH-SCP chart, and DFEWMA chart with \( \nu = 10\% \) and \( \nu = 25\% \), respectively. Our proposed methods can converge to \( DR = 1 \) as soon as \( \delta = 1 \) with \( W = 100 \). The improvement is significant compared with Figure 3(a) and (b) when \( W = 30 \). The lower sparsity can improve the performance of all methods, and the positive effect is more significant on the DFEWMA chart. The CED of our proposed methods with \( W = 100 \) decreases faster than with \( W = 30 \). The trends of CED are not monotonous when \( \delta \) is small. Figure 3(g) and (h) show that the CED increases for small \( \delta \) before decreasing again for larger \( \delta \). This phenomenon is related to the increasing DR as well as the large window size. As the DR increases with the increase of \( \delta \), we will have more detected signals, and because the shift is still relatively small, some of these signals will only be observed far into the window, resulting in a larger CED. When \( \delta \) is large, our proposed methods can achieve the minimum detection delay of seven. Hence, when the change-point is not at the beginning of the process, our proposed methods can achieve a similarly good performance as the DFEWMA control chart based on a large window.

An obvious drawback of the DFEWMA is the excessive empirical FAP. If there is no heteroscedasticity and the process practitioner does not mind many false alarms, we recommend the DFEWMA method. We believe practitioners should use our proposed control charts for all other scenarios. The reason is that, although its detection power is a little smaller, monitoring can start quickly, and false alarms will be close to the nominal level. In addition, our methods exhibit consistent performance under heteroscedastic data, whereas the performance of the DFEWMA chart is severely affected (which is why we do not consider it in the following simulations for robustness).

Another drawback of the DFEWMA method is the extra computation time in processing high-dimensional data. When \( p = 100 \) and \( W = 30 \), it takes 2 minutes to process 200 IC observations, but the RH-CP chart requires only 1 second. One convincing explanation is that individually estimating the control limits for each time point is time-consuming in the DFEWMA chart. Further, Chen, Zi, and Zou (2016) considered 0.5 the maximum correlation in their models. Wang, Goedhart, and Zwetsloot (2023) applied the DFEWMA method for monitoring highly dependent variables, with a strong correlation of 0.9. Hence, when the underlying dependency structure is strong and complicated, we recommend our proposed methods, which showed robustness to correlation in the simulation results.

### 3.2. Robustness

To comprehensively analyze the robustness and efficiency of our proposed methods, we designed two other covariance matrices as replacements for the baseline model. First, we consider \( D_p = \left( I_{p_1} \quad 0 \right) \left( 0 \quad 2I_{p-p_1} \right) \), where \( p_1 = 0.2p \). This model can test the influence of unequal variances and standardization for the RH-CP chart. Second, the covariance matrix \( \Sigma_p = (\sigma_{i,m})_{1 \leq i, m \leq p} \), where \( \sigma_{i,m} = 0.95^{\left|i-m\right|} \), is used to explore the impact of correlation. Third, a multivariate \( t \) distribution with 30 degrees of freedom, denoted as \( t_{p,30}(\mathbf{0}, \Sigma_p) \), is used to explore the robustness to nonnormality. We also investigate the influence of dimensionality, window size, sparsity, and heteroscedasticity. Again, the FAP is used to compare the IC performances. The DR, CED, and change-point estimate (CPE) are used to assess the OC performances.

| Model \( \mathbf{0}, \Sigma_p \) | Methods | \( W = 30 \) | \( W = 40 \) | \( W = 30 \) | \( W = 40 \) | \( W = 30 \) | \( W = 40 \) |
|-----------------------------|---------|-------------|-------------|-------------|-------------|-------------|-------------|
| \( N_p(0, D_p) \) | RH-CP   | 0.014       | 0.008       | 0.008       | 0.008       | 0.009       | 0.013       |
| \( N_p(0, \Sigma_p) \) | RH-SCP  | 0.008       | 0.007       | 0.010       | 0.008       | 0.014       | 0.011       |
| \( t_{p,30}(0, \Sigma_p) \) | RH-CP   | 0.015       | 0.008       | 0.005       | 0.014       | 0.004       | 0.007       |
| \( t_{p,30}(0, \Sigma_p) \) | RH-SCP  | 0.009       | 0.007       | 0.008       | 0.010       | 0.012       | 0.011       |
| \( N_p(0, \mathbf{I}_p) \) | RH-CP   | 0.004       | 0.004       | 0.002       | 0.014       | 0.010       | 0.012       |
| \( N_p(0, \mathbf{I}_p) \) | RH-SCP  | 0.004       | 0.005       | 0.005       | 0.012       | 0.008       | 0.012       |
| \( N_p(0, \mathbf{I}_p) \) | RH-CP   | 0.050       | 0.030       | 0.045       | 0.029       | 0.062       | 0.036       |
| \( N_p(0, \mathbf{I}_p) \) | RH-SCP  | 0.033       | 0.049       | 0.026       | 0.045       | 0.036       | 0.074       |
| \( N_p(0, \mathbf{I}_p) \) | RH-CP   | 0.037       | 0.019       | 0.027       | 0.018       | 0.029       | 0.028       |
| \( N_p(0, \mathbf{I}_p) \) | RH-SCP  | 0.011       | 0.027       | 0.012       | 0.053       | 0.023       | 0.028       |
| \( t_{p,30}(0, \mathbf{I}_p) \) | RH-CP   | 0.013       | 0.017       | 0.022       | 0.016       | 0.038       | 0.030       |
| \( t_{p,30}(0, \mathbf{I}_p) \) | RH-SCP  | 0.020       | 0.019       | 0.020       | 0.028       | 0.022       | 0.035       |

When the designed FAP = 0.01.
can keep the charts. Both the change point shifts from high-dimensional processes. Therefore, the RH-SCP average, but the simulated scenarios. The window size has no effect on the heteroscedasticity. 

sponding to such a small scale, the shifts occur in variables with smaller variances; other variables with large variances disturb the RH-SCP chart and cause a small DR. The CED and CPE corresponding to such a small DR are meaningless.

Table 3. DR, CED, and CPE of the RH-CP and RH-SCP charts, under various models and $W_0$.

| Model | Methods | $W = 30$ | $W = 40$ | $W = 30$ | $W = 40$ |
|-------|---------|----------|----------|----------|----------|
| $N_{100}(0,D_{100})$ | RH-CP   | 0.02     | 0.14     | 31.38    | 20.44    |
|       | RH-SCP  | 0.09     | 1.00     | 9.06     | 15.01    |
| $N_{100}(0,\Sigma_{100})$ | RH-CP   | 0.93     | 0.99     | 8.93     | 15.15    |
|       | RH-SCP  | 0.69     | 0.94     | 11.15    | 15.74    |
| $\Gamma_{100,30}(0,\Sigma_{100})$ | RH-CP   | 0.91     | 0.99     | 9.10     | 15.22    |
|       | RH-SCP  | 0.61     | 0.91     | 11.74    | 15.96    |
| $N_{100}(0,\lambda D_{100})$ | RH-CP   | 0.06     | 0.05     | 34.67    | 36.45    |
|       | RH-SCP  | 0.94     | 1.00     | 10.67    | 15.08    |
| $N_{100}(0,\lambda \Sigma_{100})$ | RH-CP   | 0.83     | 0.96     | 10.87    | 15.78    |
|       | RH-SCP  | 0.63     | 0.86     | 11.77    | 17.22    |
| $\Gamma_{100,30}(0,\lambda \Sigma_{100})$ | RH-CP   | 0.80     | 0.91     | 11.31    | 15.88    |
|       | RH-SCP  | 0.58     | 0.84     | 12.16    | 17.48    |

When $p = 100$, $\tau = 25$, $\delta = 2$, nominal FAP = 0.01, $v = 10\%$.

Table 4. DR, CED, and CPE of the RH-CP and RH-SCP charts, under various models and $W_0$.

| Model | Methods | $W = 30$ | $W = 40$ | $W = 30$ | $W = 40$ |
|-------|---------|----------|----------|----------|----------|
| $N_{100}(0,D_{100})$ | RH-CP   | 1.00     | 1.00     | 5.00     | 15.00    |
|       | RH-SCP  | 1.00     | 1.00     | 5.01     | 15.00    |
| $N_{100}(0,\Sigma_{100})$ | RH-CP   | 0.96     | 1.00     | 8.26     | 15.10    |
|       | RH-SCP  | 0.93     | 0.99     | 8.74     | 15.18    |
| $\Gamma_{100,30}(0,\Sigma_{100})$ | RH-CP   | 0.75     | 0.97     | 10.31    | 15.49    |
|       | RH-SCP  | 0.68     | 0.93     | 11.01    | 15.77    |
| $N_{100}(0,\lambda D_{100})$ | RH-CP   | 1.00     | 1.00     | 5.00     | 15.00    |
|       | RH-SCP  | 0.93     | 0.99     | 8.74     | 15.18    |
| $N_{100}(0,\lambda \Sigma_{100})$ | RH-CP   | 0.95     | 0.97     | 10.44    | 15.52    |
|       | RH-SCP  | 0.85     | 0.95     | 11.03    | 15.77    |
| $\Gamma_{100,30}(0,\lambda \Sigma_{100})$ | RH-CP   | 0.70     | 0.89     | 11.26    | 16.89    |
|       | RH-SCP  | 0.84     | 0.95     | 11.03    | 15.77    |
| $\Gamma_{100,30}(0,\lambda \Sigma_{100})$ | RH-CP   | 0.64     | 0.90     | 11.82    | 17.04    |
|       | RH-SCP  | 0.58     | 0.84     | 12.16    | 17.48    |

When $p = 100$, $\tau = 25$, $\delta = 2$, nominal FAP = 0.01, $v = 25\%$.

Table 2 shows the FAP of the proposed control charts. Both the RH-CP chart and the RH-SCP chart can keep the FAP at a low level for all considered scenarios. The window size has no effect on the FAP. Heteroscedasticity can multiply the FAP threefold on average, but the simulated FAP values are still acceptable. Thus, the proposed methods are robust to heteroscedasticity.

Table 1 showed that both the RH-CP chart and the RH-SCP chart are more efficient in detecting large shifts from high-dimensional processes. Therefore, the following comparisons only include the scenarios under $p = 100$ and $\delta = 2$. Sparsity level $v$ is used to control the number of changed variables; after the change point $\tau = 25$, the location parameter becomes $\mu = (\delta, 0_{p(1-v)})$. Table 3 shows results for $v = 10\%$ and Table 4 for $v = 25\%$.

One evident difference between Table 3 and Table 4 is the performance of the RH-CP chart under $N_p(0, D_p)$ and $N_p(0, \lambda D_p)$ distributions. When $v = 10\%$, the shifts occur in variables with smaller variances; other variables with large variances disturb the RH-CP method and cause a small DR. The CED and CPE corresponding to such a small DR are meaningless.

In applications, it is likely that different variables have different scales. It is recommended that the scales be standardized before monitoring starts. This can be done using application knowledge, if available. Alternatively, an initial data sample can be collected (Phase I) to estimate sample standard deviations for each process. We highly recommend the RH-CP method when standardization techniques are feasible and the RH-SCP method when standardization of scales is not feasible.

The RH-CP chart is more sensitive to sparse changes than the RH-SCP chart. When $v = 10\%$, the DR values of the RH-CP chart are greater than 0.9 with $\Sigma_{100}$. Under heteroscedastic models, it can still keep $DR > 0.8$. As expected, both methods are more efficient when more variables shift ($v = 25\%$). The improvement is significant in model $N_{100}(0, D_{100})$ because the shifts occur in variables with larger variance.

Both methods can achieve higher detection power, smaller detection delay, and more accurate CPE with larger window sizes. The improvement is more noticeable for the RH-SCP method because the estimated standard deviations are more accurate. The CED shows that the proposed methods can signal within 10 observations after a change or in the first window. Determining the optimal window size is a tradeoff between detection delay and sensitivity and should depend on the expected change-point location and application.

Correlation and/or heavy tails can reduce the DR of the RH-SCP chart by about 30 percent, and they have no significant impact on CED and CPE. The RH-CP outperforms the RH-SCP by showing more robustness. The same conclusion can be drawn under heteroscedastic scenarios, although the variability from covariance matrices can cause an approximate 10 percent overall reduction of DR.

Overall, we presented and compared the robustness and sensitivity of the proposed methods. The RH-CP chart and the RH-SCP chart can maintain low FAPs under nonnormal and heteroscedastic distributions. They can also detect large changes in high-dimensional mean vectors with appropriate window sizes under various conditions. The RH-CP method is recommended, with standardization in data preparation.

4. Diagnosis

After detecting a signal, it is of interest to diagnose the root causes. Eqs. [10] and [13] give the change-point estimate for the RH-CP and RH-SCP charts. One limitation of the proposed supremum-based methods is that the signal is directly caused by one
variable, and the abnormal behavior in other variables is ignored. We propose the following diagnostic procedure for the RH-CP chart to locate all suspicious variables:

$$\hat{V}_n^{(1)} = \arg \left( \frac{\sqrt{k'(W - k')}|\bar{X}_{k',r} - \bar{X}_{W-k',r}|}{\sqrt{W}} > h_{p,W}^{(1)}|U_{n,W}^{(1)} > h_{p,W}^{(1)}, H_{1}\right), \tag{14}$$

where $\hat{V}_n^{(1)}$ is the set of variables that yield a signal in the RH-CP method. To evaluate the accuracy of the diagnostic procedure, we compared $\hat{V}_n^{(1)}$ with the pre-defined variable set $V_n = (X_1, ..., X_{sp})$. The detection rate of variables (DRV) is defined as

$$DRV^{(1)} = P\left(\frac{\left|\hat{V}_n^{(1)} \cap V_n\right|}{|V_n|} \geq \frac{\sum_{i \in \hat{V}_n^{(1)}} I(i \leq p \times v)}{p \times v}\right), \tag{15}$$

where $i$ is an element in $\hat{V}_n^{(1)}$ (the set of suspicious variables). $DRV^{(2)}$ for the RH-SCP chart can be derived similarly.

Table 5 shows the post-signal diagnostic performances of the proposed methods with $p = 100$, $\tau = 25$, $W = 40$, and $\delta = 2$ under different models. The RH-CP method performs better in finding the suspicious variables from various models. Without heteroscedasticity, both can correctly detect approximately 90 percent of the change variables under correlated models. Heteroscedasticity has a limited effect on the performance of both methods. One possible way to locate assignable causes is to add a variable selection algorithm to the proposed method.

5. Case study

In this section, we illustrate the proposed methods on two real data sets.

### 5.1. Semiconductor data

The first data set is from a semiconductor manufacturing process under constant surveillance via monitoring signals/variables collected from sensors and/or measurement points. This data set is available online. It consists of 590 variables, each with 1,567 records in chronological order. A classification label (±1) is given to indicate whether the product passes or fails, where −1 represents pass and 1 is fail. Among these, 1,463 observations are from the IC sample, and the remaining 104 observations compose the OC sample.

Preprocessing of constant values, null values, and potential outliers in the data set is necessary. We remove the constant variables from both samples, leaving 416 variables. The outliers in the IC sample are identified by Tukey’s fences (Tukey 1977) and replaced by the variable median. The variable median replaces the missing values in both samples. After that, all the observations are standardized by the sample mean and standard deviation from the IC observations so that all variables have similar ranges. The proposed methods are applied to monitor the standard scores. Figure 4 shows the heteroscedastic standard scores of some variables when they are in control.

The control limits are simulated by the proposed bootstrap method in Algorithm 1, based on the IC sample with $FAP = 0.05$ and $n = 100$. Based on the assumption that the process starts from an IC status, the changes occur at time $\tau$, where $\tau = 25, 50$. As shown in Table 6, the RH-CP chart is efficient in monitoring real high-dimensional processes. The $DR$ is equal to 1 in all scenarios. It can achieve timely detection in the first window when $W > \tau$ or with an acceptable delay when $W < \tau$. But the RH-SCP chart performs poorly. One explanation is that both the mean vector and the dispersion parameter change after $\tau$, and the standardization step in the RH-SCP chart shrinks the location changes. These results confirm that the RH-CP chart is robust to unknown distributions and heteroscedasticity.

### 5.2. Monitoring the health condition of escalators

This case uses vibration data from escalators. The data are part of a project to develop a comprehensive health condition model for escalators using mathematical analysis of the related parameters. Sensors were installed to collect vibration data for four important escalator components: gearbox, main drive, motor, and tension carriage. Each component had two vibration sensors that recorded three vibration profiles per day from January 4 to October 20, 2021. Participants
in this project did not give written consent for their data to be shared publicly, so supporting data are not available.

Figure 5(a) shows the vibration profiles collected by two sensors in one day. The x-axis is frequency, and the y-axis is amplitude. The shapes are different and are complicated to analyze. Therefore, we summarize the profiles using indexes proposed by Li et al. (2016). Each profile is described by 13 indexes; in other words, 13 features are used to describe one component, yielding 104 variables for one escalator.

We use the records from January to July 2021 as Phase I. According to the fault log, we removed the observations related to faults. Finally, we have 450 IC data points (three measures per day for six months). The other 140 observations from August to October are used for Phase II monitoring. All data are standardized by the IC sample mean and variance. Two standardized health indicators of one gearbox sensor in Phase I are shown in Figure 5(b).

We set $W = 20$ and $FAP = 0.01$ to simulate the control limits from the IC data and apply them to Phase II monitoring. Figure 6 plots the control charts. The RH-SCP chart signals at 30, 35, 40, but no reported faults are identified during this period. Both charts can detect changes at 80, 85, 90, and 95, and the
estimated change-point is at 75 on September 7 (see the vertical line). After applying the post-signal diagnosis procedure from Section 4, the gearbox and motor are identified as root causes. We also observe significant shape changes in the spectrum before and after that date, and investigations show a link with maintenance activity. These two cases support the claim that the RH-CP method is more sensitive and more robust to unknown distributions, which is recommended for real applications.

6. Conclusion

The curse of dimensionality increases the unreliability of parameter estimates, especially when numerous IC observations are unavailable. Sparsity is another practical aspect when analyzing control chart performance, which challenges signal detection and diagnosis methods. The major contribution of this article is that it considers heteroscedasticity as a common cause of variability in processes and explores the effect of heteroscedasticity in process monitoring. We proposed two change-point–based control charts for monitoring large, sparse changes in a high-dimensional mean vector under heteroscedasticity, specifically for HDLSS scenarios. Our method is useful for detecting changes in the location parameter while ignoring variance changes, but we do not recommend it for detecting changes in covariance structure.

A moving window is added to speed up computation and increase sensitivity. Results of our experiments show that the proposed RH-CP and RH-SCP charts can effectively detect large, sparse shifts. In addition, they can accurately estimate the change-point and the potential OC variables. The RH-CP chart is more robust to correlation, nonnormality, and heteroscedasticity. A reliable standardization step is necessary before applying the RH-CP chart. Although all the simulations are based on a finite set of observations, the proposed methods can be applied for online monitoring with continuous and infinite observations. In addition, the experiments provide information about determining the optimal window size, where a tradeoff between sensitivity and false alarms needs to be considered. The real-world case studies illustrate the robustness and practicability of the RH-CP chart in applications. In comparison, our proposed charts are not as sensitive as the spatial-rank–based DFEWMA chart. However, the DFEWMA chart needs a large reference sample, is slow to compute, is not robust, and shows excessive false alarms, none of which are issues in our method. Future directions for research should focus on improving the sensitivity in detecting small and medium changes in high-dimensional heteroscedastic processes. One possibility is to adopt a variable selection algorithm before starting to monitor. An alternative is exploring other statistics that are robust to heteroscedasticity.

Disclosure statement

No potential conflict of interest was reported by the author(s).
About the authors

Zezhong Wang received her doctoral degree from the Department of Systems Engineering, City University of Hong Kong. She is a postdoc researcher at the Hong Kong Industrial Artificial Intelligence and Robotics Centre (FLAIR). Her research interests include statistical process monitoring, data science, and artificial intelligence applications. Her email address is zzwang@hkflair.org.

Inez M. Zwetsloot is an associate professor in the Department of Business Analytics at the University of Amsterdam, The Netherlands. Previously she was assistant professor at the Department of Systems Engineering, City University of Hong Kong. She is the recipient of the Young Statistician Award (ENBIS, 2021) and the Feigenbaum Medal (ASQ, 2021). Her research focuses on using statistics and analytics for solving business challenges using data. This includes work on statistical process monitoring, network analytics, quality engineering, and data science. Her email address is i.m.zwetsloot@uva.nl.

ORCID

Zezhong Wang http://orcid.org/0000-0002-7147-1489
Inez Maria Zwetsloot http://orcid.org/0000-0002-6144-4188

References

Abdella, G. M., K. N. Al-Khalifa, S. Kim, M. K. Jeong, E. A. Elsayed, and A. M. Hamouda. 2017. Variable selection-based multivariate cumulative sum control chart. Quality and Reliability Engineering International 33 (3):565–78. doi: 10.1002/qre.2041.

Alghuried, A., and R. Moghaddass. 2022. Anomaly detection in large-scale networks: A state-space decision process. Journal of Quality Technology 54 (1):65–92. doi: 10.1080/00224065.2020.1805379.

Amiri, A., and S. Allahyari. 2012. Change point estimation methods for control chart postsignal diagnostics: A literature review. Quality and Reliability Engineering International 28 (7):673–85. doi: 10.1002/qre.1266.

Bickel, P. J. 1969. A distribution free version of the Smirnov two sample test in the p-variate case. The Annals of Mathematical Statistics 40 (1):1–23. doi: 10.1214/aoms/1177697800.

Bodnar, O. 2009. Application of the generalized likelihood ratio test for detecting changes in the mean of multivariate GARCH processes. Communications in Statistics—Simulation and Computation 38 (5):919–38. doi: 10.1080/03610910802691861.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31 (3):307–27. doi: 10.1016/0304-4076(86)90063-1.

Chakraborti, S., S. Human, and M. Graham. 2008. Phase I statistical process control charts: An overview and some results. Quality Engineering 21 (1):52–62. doi: 10.1080/0898210820445561.

Chang, J., C. Zheng, W.-X. Zhou, and W. Zhou. 2017. Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. Biometrics 73 (4):1300–10. doi: 10.1111/biom.12695.

Chen, N., X. ZI, and C. Zou. 2016. A distribution-free multivariate control chart. Technometrics 58 (4):448–59. doi: 10.1080/00401706.2015.1049750.

Chen, S. X., and Y.-L. Qin. 2010. A two-sample test for high-dimensional data with applications to gene-set testing. The Annals of Statistics 38 (2):808–35. doi: 10.1214/09-AOS716.

Choi, K., and J. Marden. 1997. An approach to multivariate rank tests in multivariate analysis of variance. Journal of the American Statistical Association 92 (440):1581–90. doi: 10.1080/01621459.1997.10473680.

De Ketelaere, B., M. Hubert, and E. Schmitt. 2015. Overview of PCA-based statistical process-monitoring methods for time-dependent, high-dimensional data. Journal of Quality Technology 47 (4):318–35. doi: 10.1080/00224065.2015.11918137.

Downs, G. W., and D. M. Rocke. 1979. Interpreting heteroscedasticity. American Journal of Political Science 23 (4): 816–28. doi: 10.2307/2110809.

Ebadi, M., S. Chenouri, D. K. Lin, and S. H. Steiner. 2022. Statistical monitoring of the covariance matrix in multivariate processes: A literature review. Journal of Quality Technology 54 (3):269–89. doi: 10.1080/00224065.2021.1889419.

Ebrahimi, S., C. Ranjan, and K. Paynabar. 2022. Monitoring and root-cause diagnostics of high-dimensional data streams. Journal of Quality Technology 54 (1):20–43. doi: 10.1080/00224065.2020.1805377.

Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50 (4):987–1007. doi: 10.2307/1912773.

Frisén, M. 2008. Financial surveillance. Vol. 71. Hoboken, NJ: Wiley Online Library.

Frisén, M. 2009. Optimal sequential surveillance for finance, public health, and other areas. Sequential Analysis 28 (3): 310–37. doi: 10.1080/0747494090341605.

Gregory, K. B., R. J. Carroll, V. Baladandayuthapani, and S. N. Lahiri. 2015. A two-sample test for equality of means in high dimension. Journal of the American Statistical Association 110 (510):837–49. doi: 10.1080/01621459.2014.934826.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. The elements of statistical learning: Data mining, inference, and prediction. New York, NY: Springer Science & Business Media.

Holland, M. D., and D. M. Hawkins. 2014. A control chart based on a nonparametric multivariate change-point model. Journal of Quality Technology 46 (1):63–77. doi: 10.1080/00224065.2014.11917954.

Hong, D., L. Balzano, and J. A. Fessler. 2018. Asymptotic performance of pca for high-dimensional heteroscedastic data. Journal of Multivariate Analysis 167:435–52. doi: 10.1016/j.jmva.2018.06.002.

Huang, S., Z. Kong, and W. Huang. 2014. High-dimensional process monitoring and change point detection using embedding distributions in reproducing kernel Hilbert space. IIE Transactions 46 (10):999–1016. doi: 10.1080/0740817X.2013.855848.
Jiang, W., K. Wang, and F. Tsung. 2012. A variable-selection-based multivariate EWMA chart for process monitoring and diagnosis. *Journal of Quality Technology* 44 (3):209–30. doi: 10.1080/00224065.2012.11917896.

Knoth, S. 2015. Run length quantiles of EWMA control charts monitoring normal mean or/and variance. *International Journal of Production Research* 53 (15): 4629–47. doi: 10.1080/00207543.2015.1005253.

Lee, W., T. H. McCormick, J. Neil, C. Sodja, and Y. Cui. 2022. Anomaly detection in large-scale networks with latent space models. *Technometrics* 64 (2):241–52. doi: 10.1080/00401706.2021.1952900.

Li, C., R.-V. Sánchez, G. Zurita, M. Cerrada, and D. Cabrera. 2016. Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning. *Sensors* 16 (6):895. doi: 10.3390/s16060895.

Li, Y., Y. Liu, C. Zou, and W. Jiang. 2014. A self-starting control chart for high-dimensional short-run processes. *International Journal of Production Research* 52 (2):445–61. doi: 10.1080/00207543.2013.832001.

Liu, L., J. Liu, H. Wang, S. Tan, Q. Guo, and X. Sun. 2023. A KLMS dual control chart based on dynamic nearest neighbor kernel space. *IEEE Transactions on Industrial Informatics* 19 (5):6950–62. doi: 10.1109/TII.2022.3209248.

Randles, R. H. 2000. A simpler, affine-invariant, multivariate, distribution-free sign test. *Journal of the American Statistical Association* 95 (452):1263–8. doi: 10.1111/j.1369-7486.2000.00337.x.

Saha, E., S. Sarkar, and A. K. Ghosh. 2017. Some high-dimensional one-sample tests based on functions of interpoint distances. *Journal of Multivariate Analysis* 161:83–95. doi: 10.1016/j.jmva.2017.07.006.

Schipper, S., and W. Schmid. 2001. Control charts for GARCH processes. *Nonlinear Analysis: Theory, Methods & Applications* 47 (3):2049–60. doi: 10.1016/S0362-546X(01)00332-7.

Shu, L., and J. Fan. 2018. A distribution-free control chart for monitoring high-dimensional processes based on interpoint distances. *Naval Research Logistics (NRL)* 65 (4):317–30. doi: 10.1002/nav.21809.

Tamuz, O., T. Maze, and S. Zucker. 2005. Correcting systematic effects in a large set of photometric light curves. *Monthly Notices of the Royal Astronomical Society* 356 (4):1466–70. doi: 10.1111/j.1365-2966.2004.08585.x.

Tukey, J. W. 1977. *Exploratory data analysis*. Vol. 2. Reading, MA: Creative Education.

Wang, K., and W. Jiang. 2009. High-dimensional process monitoring and fault isolation via variable selection. *Journal of Quality Technology* 41 (3):247–58. doi: 10.1080/00224065.2009.11917780.

Wang, Z., R. Goedhart, and I. M. Zwetsloot. 2023. Monitoring high-dimensional heteroscedastic processes using rank-based EWMA methods. *Computers & Industrial Engineering* 184: 109544. doi: 10.1016/j.cie.2023.109544.

Yu, L., I. M. Zwetsloot, N. T. Stevens, J. D. Wilson, and K. L. Tsui. 2022. Monitoring dynamic networks: A simulation-based strategy for comparing monitoring methods and a comparative study. *Quality and Reliability Engineering International* 38 (3):1226–50. doi: 10.1002/qre.2944.

Zamba, K., and D. M. Hawkins. 2006. A multivariate change-point model for statistical process control. *Technometrics* 48 (4):539–49. doi: 10.1198/004017006000000291.

Zhang, C., N. Chen, and J. Wu. 2020. Spatial rank-based high-dimensional monitoring through random projection. *Journal of Quality Technology* 52 (2):111–27. doi: 10.1080/00224065.2019.1571336.

Zhang, J. 2002. Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 64 (2):281–94. doi: 10.1111/1467-9868.00337.

Zou, C., W. Jiang, and F. Tsung. 2011. A LASSO-based diagnostic framework for multivariate statistical process control. *Technometrics* 53 (3):297–309. doi: 10.1198/TECH.2011.10034.

Zou, C., and P. Qiu. 2009. Multivariate statistical process control using LASSO. *Journal of the American Statistical Association* 104 (488):1586–96. doi: 10.1198/jasa.2009.tm08128.

Zou, C., and F. Tsung. 2011. A multivariate sign EWMA control chart. *Technometrics* 53 (1):84–97. doi: 10.1198/TECH.2010.09095.

Zou, C., Z. Wang, and F. Tsung. 2012. A spatial rank-based multivariate EWMA control chart. *Naval Research Logistics (NRL)* 59 (2):91–110. doi: 10.1002/nav.21475.

Zou, C., Z. Wang, X. Zi, and W. Jiang. 2015. An efficient online monitoring method for high-dimensional data streams. *Technometrics* 57 (3):374–87. doi: 10.1080/00401706.2014.940089.