IISE Transactions

Taylor & Francis
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uiie21

Distribution-free image monitoring with application to battery coating process

Tingnan Gong, Di Liu, Heeseon Kim, Seong-Hee Kim, Taeheung Kim, Dongki Lee & Yao Xie

To cite this article: Tingnan Gong, Di Liu, Heeseon Kim, Seong-Hee Kim, Taeheung Kim, Dongki Lee & Yao Xie (29 Feb 2024): Distribution-free image monitoring with application to battery coating process, IISE Transactions, DOI: 10.1080/24725854.2024.2308542

To link to this article: https://doi.org/10.1080/24725854.2024.2308542

IISE TRANSACTIONS
https://doi.org/10.1080/24725854.2024.2308542

Distribution-free image monitoring with application to battery coating process

Tingnan Gong[a], Di Liu[a], Heeseon Kim[b], Seong-Hee Kim[a], Taeheung Kim[b], Dongki Lee[b], and Yao Xie[a]

[a]H. Milton School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA; [b]LG Electronics, Seoul, South Korea

ABSTRACT

This article presents a distribution-free image monitoring procedure for a manufacturing process, where a series of images are converted into a vector of two feature values extracted from singular value decomposition. Traditional image-based monitoring methods often make specific assumptions about marginal distributions and spatio-temporal dependence structures, which are often violated in real-world scenarios such as battery coating processes. To overcome this issue, we propose a distribution-free image monitoring procedure that detects a shift in the mean matrix of monitored images. Our method involves performing singular value decomposition of each image matrix in two ways to obtain two values, which are then combined into a bivariate vector. The bivariate vectors are monitored using a distribution-free multivariate CUSUM procedure, for which we determine control limits analytically, enabling convenient and easy implementation of the monitoring procedure. We demonstrate the effectiveness of our proposed procedure, as measured by average run lengths, using various simulated data and a real-data example from a battery coating process.

ARTICLE HISTORY
Received 8 March 2023
Accepted 5 January 2024

KEYWORDS
Statistical process control; CUSUM; image monitoring; distribution free; nonparametric statistical methods

1. Introduction

The monitoring of streaming image data has emerged as a crucial aspect of quality control in modern manufacturing. Recent advancements in sensor technology have made it possible to obtain real-time and high-resolution image data, but the complexity of the data has created challenges in monitoring. Firstly, the high resolution of images results in a high-dimensional problem with a massive number of pixels. Secondly, pixels and image frames exhibit a complex spatio-temporal correlation. Thirdly, the image data may not follow a multivariate normal distribution. To address these challenges, we propose a novel distribution-free monitoring procedure for high-dimensional images, using low-rank models to account for the complex correlation structure in the data while being robust to non-normality.

Our study is motivated by a battery coating process that is prevalent in the battery manufacturing industry. In this process, the electrode slurry is uniformly sprayed onto a long foil within a specific area, as illustrated in Figure 1, during in-control production. The foil moves forward and perpendicular to the sensor's movement, and thickness measurements are collected by the sensor at each trip from one side of the foil to the other. The sensor captures approximately 298 measurements during each trip, which occurs roughly every 7 seconds. Figure 2 presents a heatmap of the thickness measurements of the electrode slurry on two distinct sections of the foil: Figure 2(a) exhibits a uniform in-control pattern, whereas Figure 2(b) demonstrates missing spraying. These observations highlight the importance of monitoring the battery coating process to ensure the uniformity of the sprayed slurry, which is critical for the battery's performance. Our research aims to address this need by proposing a novel monitoring procedure that is capable of detecting any anomalies in the coating process. The battery coating process presents all the challenges mentioned above that make the monitoring task difficult. Figure 3 illustrates the complex characteristics of the measurements in the battery coating process. Figure 3(a) depicts the auto-correlation of the first measurements in the thickness measurement vectors obtained at each sensor trip, while Figure 3(b) shows cross-correlation among a few selected measurements. Furthermore, Figure 3(c) demonstrates the non-normality of the first measurements obtained at each sensor trip. The high-dimensionality issue is also present as each sensor trip collects a vector of 298 measurements, let alone an image consisting of a series of such vectors.

Different monitoring procedures can be used depending on the interpretation of image data. For the thickness measurements in battery coating processes, one approach involves treating the measurements collected in each trip or over a specific part of the foil as a vectorized profile and applying advanced profile monitoring techniques. Alternatively, an image of the electrode slurry thickness over a fixed portion of the foil can be formed, and image monitoring procedures can be applied. In practical settings, the target mean matrix for an image is often rank-one. For example, every row of a target mean matrix is identical in the battery coating process or liquid crystal display screens.

CONTACT Seong-Hee Kim skim@isye.gatech.edu
Supplemental data for this article is available online at https://doi.org/10.1080/24725854.2024.2308542

If one decides to convert images into vectorized data, profile monitoring can be utilized, as discussed in Megahed *et al.* (2011). A profile refers to a functional relationship between explanatory variables and quality attributes of the variables. Expository discussions on applying statistical process control techniques to profile monitoring are available in Woodall *et al.* (2004), and Woodall (2007). However, most profile monitoring techniques, such as those reported by Kazemzadeh *et al.* (2009), Zhu and Lin (2009), and Colosimo and Pacella (2010), assume that measurements are not cross-correlated, consecutive profiles are not auto-correlated, and the measurements are normally distributed. Applying restricted profile monitoring techniques to image data can significantly increase the detection delay or the false alarm rate. For instance, Alwan (1992) and Mastrangelo *et al.* (1996) demonstrate that applying a procedure designed for temporally independent measurements to auto-correlated data can increase the false alarm rate. Noorossana *et al.* (2008) also shows that the violation of the normality assumption can magnify the false alarm rate.

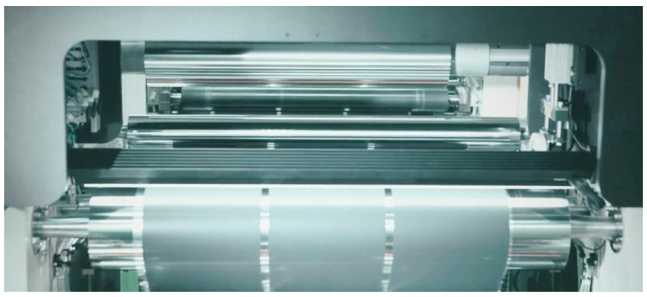


Figure 1. A machine performing battery coating.

Recent developments in profile monitoring have addressed the issue of cross- and auto-correlation to some extent. Some methods, such as combining Principal Component Analysis (PCA) with linear regression (Niaki *et al.*, 2015), or using wavelet-based distribution-free procedures (Lee *et al.*, 2012; Wang *et al.*, 2015) or sparse multi-channel functional PCA (Zhang *et al.*, 2018), have tackled cross-correlation. In contrast, others, such as by Khedmati and Niaki (2016), Wang and Lai (2019), or a multivariate spatio-temporal autoregressive model (Wang *et al.*, 2021), have dealt with auto-correlation. However, these methods have limitations on either the auto-correlation structure or the dimensionality. Zhang *et al.* (2020) tackle the high-dimensional profile through random projection and enhance computation efficiency by ensembling and fusing. However, it requires temporally independent profiles. Although much effort has been made, it is difficult to address cross- and auto-correlation, non-normality, and high dimension simultaneously.

In image monitoring, many techniques can handle cross-correlation but ignore auto-correlation, such as those reported by Wang and Tsung (2005), Megahed *et al.* (2012), and Amirkhani and Amiri (2020). Lu and Tsai (2005) propose a low-rank projection approach without discussing correlation, whereas Alaeddini *et al.* (2018) use novel feature extraction but impose data independence. Yan *et al.* (2014) model images with tensors, which is general but time-consuming. A recent spatio-temporal decomposition methodology, such as

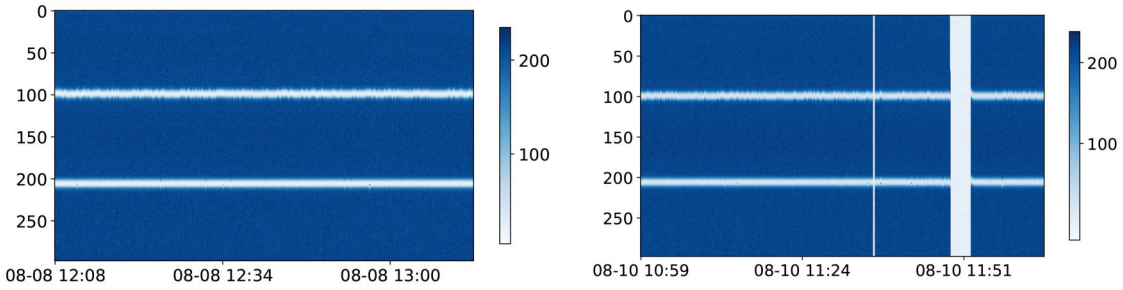


Figure 2. Heatmaps of thickness measurements of the electrode slurry on two different portions of a foil.

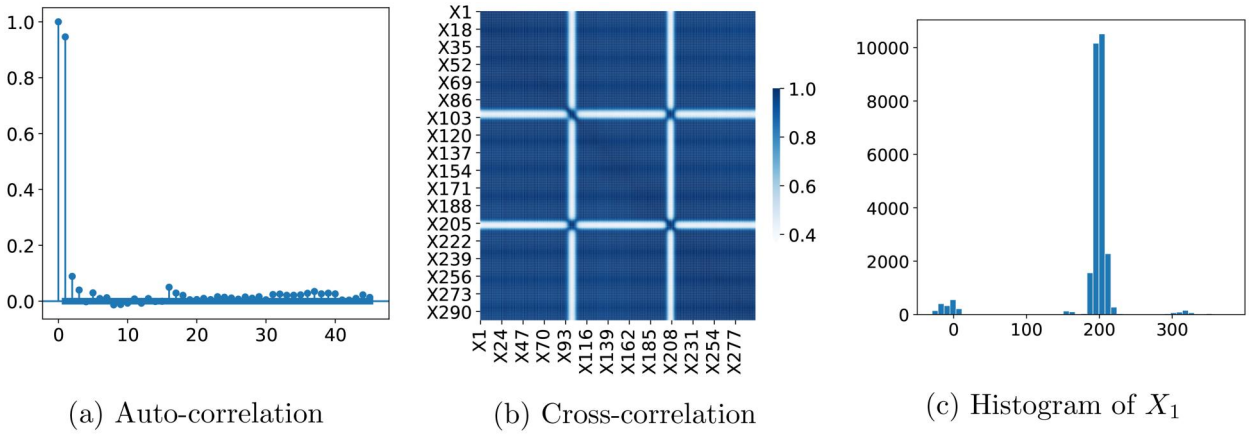


Figure 3. Characteristics of battery coating data.

Yan *et al.* (2017, 2018, 2022), can handle spatio-temporal correlation and high-dimensionality, with focus on hot-spot detection in metal additive manufacturing. Colosimo and Grasso (2018) propose a weighted PCA for the same problem, but its specificity limits general application. Eslami *et al.* (2021) extend the wavelet-based methods of Koosha *et al.* (2017) to handle potential spatio-temporal correlation but still assume normality.

One currently thriving thread of literature employs machine learning or deep learning techniques to monitor high-dimensional data. Deep learning techniques such as Recurrent Neural Networks (Ebrahimzadeh *et al.*, 2019), Long Short-Term Memory autoencoder (Atashgahi *et al.*, 2023), and Generative Adversarial Networks (Li *et al.*, 2019) have been used to monitor data with complex spatio-temporal dependence. Wu *et al.* (2020) and Gupta *et al.* (2022) introduce end-to-end adaptive methods for anomaly detection, focusing on data pre-processing and feature extraction with deep learning techniques. Recently, some studies have explored online training of neural networks to construct control charts and integrate them with conventional statistical process control. Lee *et al.* (2023) perform online training of neural networks to build a CUSUM procedure, whereas Hushchyn *et al.* (2020) develop online neural network classification and online neural network regression for monitoring. It is noteworthy that the aforementioned methods are originally developed for multivariate time series, but can be applied to image monitoring when images are vectorized. Garcia *et al.* (2022) consider multivariate time series but transform them into images to facilitate the training of deep learning models. Thus, if their method is used for image monitoring, it can handle images directly. However, their primary focus is comparing different encoding methods rather than monitoring signals or images with complex dynamics. Although these neural network-based monitoring methods have the potential to handle complex data, the computation time required for online training of neural networks, especially for high-dimensional data, is substantial.

In this article, we propose a distribution-free procedure for effectively monitoring an image process with a rank-one target mean image. Although images can be monitored using profile monitoring with vectorized data, direct monitoring of image data is better for capturing spatial relationships and providing interpretable visualizations. However, cross- and auto-correlations and the inherently high dimensionality make image monitoring more challenging. Our approach is distribution-free in that it does not assume any specific marginal distributions or correlation structures. It also allows for high dimensionality of image data. The proposed procedure utilizes Singular Value Decomposition (SVD) of the image data to reduce dimensionality and then captures any cross-correlation in bivariate vectors constructed from singular values through T^2-type statistics. Additionally, possible auto-correlation is considered by incorporating an asymptotic variance constant, known to capture the variability of a stationary, but dependent, process effectively. Moreover, the control limits are analytically determined based on a pre-specified performance metric, the in-control average run length (ARL_0). The use of analytically determined control limits makes implementation easy and convenient, particularly in cases where only one sample path is available, or the monitoring needs to start immediately without sufficient in-control data or time for calibration. Some preliminary results of this article are included in the second author's doctoral dissertation (Liu, 2022).

The rest of this article is organized as follows: Section 2 defines the problem and notation. Section 3 provides the proposed monitoring statistics and procedure. In Section 4, we present numerical comparisons between the proposed procedure and three baseline procedures. Section 5 illustrates the performance of our procedure in a real-world application involving the battery coating process, and concluding remarks are given in Section 6.

2. Problem and background

In this section, we define the detection problem and provide a preliminary discussion on a relevant distribution-free monitoring procedure for multivariate time-series data.

2.1. Problem definition

Let $\boldsymbol{Y}_n$ represent a $w \times p$ matrix representing the nth image matrix obtained from a monitored process. An image matrix $\boldsymbol{Y}_n$ can comprise pixel values from each image. Or, it can represent measurements captured by a sensor over a specific region, such as thickness measurements on a piece of foil in a battery coating process. Image matrix $\boldsymbol{Y}_n$ is modeled as $\boldsymbol{Y}_n = \boldsymbol{m} + \boldsymbol{\epsilon}_n$, where $\boldsymbol{m}$ is a $w \times p$ deterministic matrix, and $\boldsymbol{\epsilon}_n$ is a $w \times p$ random noise matrix with zero mean. Each entry in $\boldsymbol{\epsilon}_n$ can have any general marginal distribution. The noise matrices may exhibit complex correlations, including cross-correlations among their entries and auto-correlations over time.

The monitoring problem can be formulated as the following online hypothesis testing:

$$\begin{aligned} \mathrm{H}_0: &\quad \boldsymbol{Y}_n = \boldsymbol{m}_0 + \boldsymbol{\epsilon}_n, \quad n = 1, 2, \ldots \\ \mathrm{H}_1: &\quad \boldsymbol{Y}_n = \begin{cases} \boldsymbol{m}_0 + \boldsymbol{\epsilon}_n, & n = 1, 2, \ldots, \tau, \\ \boldsymbol{m}_1 + \boldsymbol{\epsilon}_n, & n = \tau + 1, \tau + 2, \ldots, \end{cases} \end{aligned} \tag{1}$$

where τ is the change point for the mean of $\boldsymbol{Y}_n, \boldsymbol{m}_0$ is the in-control mean matrix of $\boldsymbol{Y}_n$, and $\boldsymbol{m}_1 = \boldsymbol{m}_0 + \boldsymbol{\Delta}$ is the out-of-control mean matrix with a $w \times p$ non-zero shift matrix $\boldsymbol{\Delta}$. In practice, $\boldsymbol{m}_0$ is a known target matrix while $\boldsymbol{m}_1$ is an unknown matrix. Note that we consider only the mean shift, while the statistical properties of $\boldsymbol{\epsilon}_n$ remain unchanged, including cross- and auto-correlations. We have the following assumptions on $\boldsymbol{m}_0$:

Assumption 1. *The in-control mean matrix $\boldsymbol{m}_0$ is a $w \times p$ rank-one matrix. The SVD of $\boldsymbol{m}_0$ is $\boldsymbol{m}_0 = \lambda_0 \boldsymbol{u}_0 \boldsymbol{v}_0^\top$, where λ_0 is the only non-zero singular value of $\boldsymbol{m}_0$; and $\boldsymbol{u}_0$ and $\boldsymbol{v}_0$ are the corresponding left and right singular orthonormal vectors, respectively.*

Assumption 1 only constrains the rank of the in-control mean matrix $\boldsymbol{m}_0$ while the rank of the out-of-control mean

matrix $\boldsymbol{m}_1$ is unrestricted. This assumption on $\boldsymbol{m}_0$ is based on the fact that many manufacturing applications have a specific target mean pattern. A rank one matrix can represent a complex target mean pattern with non-stationary and non-cyclic oscillations as shown in Figure S.1 in Supplement S.1.

2.2. Relevant work

This section reviews a distribution-free CUSUM procedure for monitoring multivariate time-series data, which is used in the proposed procedure in Section 3.

Let $\{\boldsymbol{z}_n : n = 1, 2, \ldots\}$ be a multivariate time series following a general marginal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Assumption 2. *The covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{z}_n$ is positive definite. Moreover, when the mean of $\boldsymbol{z}_n$ shifts from in-control mean $\boldsymbol{\mu}_0$ to out-of-control mean $\boldsymbol{\mu}_1$ at an unknown point, the covariance matrix $\boldsymbol{\Sigma}$ does not change.*

In order to detect a shift in the mean of $\boldsymbol{z}_n$, Liu *et al.* (2022) propose to convert $\{\boldsymbol{z}_n : n = 1, 2, \ldots\}$ to a univariate time-series by computing the T^2-type statistics of each $\boldsymbol{z}_n$ as $T_n = (\boldsymbol{z}_n - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_n - \boldsymbol{\mu}_0)$, and monitor $\{T_n : n = 1, 2, \ldots\}$ with a distribution-free CUSUM procedure proposed by Kim *et al.* (2007). Let ν_0 and σ_T represent the in-control mean and standard deviation of T_n. Then, the one-sided CUSUM statistic is computed as:

$$S_n^+ = \begin{cases} 0, & \text{if } n = 0, \\ \max\{0, S_{n-1}^+ + (T_n - \nu_0 - K)\}, & \text{if } n \geq 1, \end{cases} \tag{2}$$

where K is set to $K = k\sigma_T$ for a pre-selected real-valued constant k.

Assumption 3. *Define the standardized time series of the first n observations as*

$$\mathcal{C}_n(t) \equiv \frac{\sum_{i=1}^{\lfloor nt \rfloor} T_i - nt\nu_0}{\Omega\sqrt{n}} \quad \text{for } t \in [0, 1].$$

There exist finite real constants ν_0 and $\Omega > 0$ such that as $n \to \infty$, the sequence of random functions $\{\mathcal{C}_n(\cdot) : n = 1, 2, \ldots\}$ converges in distribution to standard Brownian motion $\mathcal{W}(\cdot)$ in the Skorohod space $D[0, 1]$. Formally, $\mathcal{C}_n(\cdot) \xrightarrow{\mathcal{D}} \mathcal{W}(\cdot)$, as $n \to \infty$ where $\xrightarrow{\mathcal{D}}$ represents convergence in distribution.

Assumption 3 is a Functional Central Limit Theorem (FCLT). The constant Ω^2 is called the asymptotic variance parameter of process $\{T_n : n = 1, 2, \ldots\}$ and is defined as:

$$\Omega^2 \equiv \lim_{N \to \infty} N \cdot \mathrm{Var}\left(\frac{1}{N}\sum_{n=1}^{N} T_n\right).$$

When auto-correlation is present, Ω^2 is a better measure of process variability than a marginal variance. Techniques for estimating Ω^2 in a distribution-free manner are discussed in Alexopoulos *et al.* (2007). A more comprehensive discussion on the conditions under which a FCLT holds is available in Glynn and Iglehart (1985). Chapter 4.4 of Whitt (2011) states that, from a practical perspective, it is usually reasonable to assume that a FCLT is valid if Ω^2 is finite.

Under a FCLT, Kim *et al.* (2007) show that the CUSUM statistic asymptotically behaves like a drifted Brownian motion and that the in-control average run length (ARL_0) for a given control limit H is approximated by:

$$\frac{\Omega^2}{2K^2}\left\{\exp\left[\frac{2K(H + 1.166\Omega)}{\Omega^2}\right] - 1 - \frac{2K(H + 1.166\Omega)}{\Omega^2}\right\} \approx \mathrm{ARL}_0. \tag{3}$$

To determine the threshold H, one can solve (3) for H with a specified target ARL_0. Once H is determined, the procedure raises an out-of-control alarm at the first point n^* where the statistic S_n^+ exceeds the threshold H, that is, $n^* = \min\{n : S_n^+ \geq H\}$.

Finally, we assume the following assumption holds:

Assumption 4. *When the mean of $\boldsymbol{z}_n$ shifts from in-control mean $\boldsymbol{\mu}_0$ to out-of-control mean $\boldsymbol{\mu}_1$ at an unknown point, the variance parameter of process $\{T_n : n = 1, 2, \ldots\}, \Omega^2$, remains unchanged.*

In this article, we convert the monitoring problem of a high-dimensional process $\{\boldsymbol{Y}_n : n = 1, 2, \ldots\}$ into that of a low-dimensional process $\{\boldsymbol{z}_n : n = 1, 2, \ldots\}$, which requires Assumptions 2, 3, and 4 on $\boldsymbol{z}_n$. It is important to note that although there are no specific assumptions made regarding $\boldsymbol{Y}_n$, it is necessary to ensure that the correlation structure of $\boldsymbol{Y}_n$ meets certain conditions to guarantee that a $\boldsymbol{z}_n$ derived from an *in-control* $\boldsymbol{Y}_n$ conforms to Assumption 3 for the existence of the variance parameter Ω^2. On the other hand, Assumptions 2 and 4 are similar to the usual assumption for mean-shift detection, where variance is assumed to remain unchanged when the mean shifts. We discuss the robustness of our proposed procedure to the violation of these assumptions in Section 3.3 and validate them for battery coating data in Section 5.

3. Distribution-free image monitoring procedure

In this section, we formulate two new statistics to monitor $\boldsymbol{Y}_n$, representing images, and propose our image monitoring procedure. Then, we provide a theoretical basis for the effectiveness of the monitoring procedure.

3.1. Monitoring statistics

We propose two monitoring statistics that utilize the SVD technique, building on Assumption 1, which assumes that the SVD representation of $\boldsymbol{m}_0$ is $\boldsymbol{m}_0 = \lambda_0 \boldsymbol{u}_0 \boldsymbol{v}_0^\top$. The first statistic, denoted by λ_n^P, is constructed by projecting $\boldsymbol{Y}_n$ along the direction of $\boldsymbol{u}_0$ and $\boldsymbol{v}_0$, and it is defined as $\lambda_n^P = \boldsymbol{u}_0^\top \boldsymbol{Y}_n \boldsymbol{v}_0$ for $n = 1, 2, \ldots$. The second statistic, denoted by λ_n^R, is constructed by subtracting the in-control mean matrix $\boldsymbol{m}_0$ from $\boldsymbol{Y}_n$ to obtain the residual matrix: $\boldsymbol{R}_n = \boldsymbol{Y}_n - \boldsymbol{m}_0$ for $n = 1, 2, \ldots$, and then calculating the largest singular value of $\boldsymbol{R}_n$. When the process is in control, $\boldsymbol{R}_n = \boldsymbol{\epsilon}_n$ and the entries of $\boldsymbol{R}_n$ are random variables with mean 0. We define $\boldsymbol{z}_n = [\lambda_n^P, \lambda_n^R]^\top$ and monitor the

bivariate time-series $\{\boldsymbol{z}_n : n = 1, 2, \ldots\}$ using the distribution-free CUSUM procedure reviewed in Section 2.2. Algorithm 1 gives the steps of the Distribution-Free Image Monitoring (DFIM) procedure.

During the setup phase before the implementation of the DFIM procedure, various parameters need to be estimated, including $\boldsymbol{\mu}_0$ (in-control mean of $\boldsymbol{z}_n$), $\boldsymbol{\Sigma}$ (covariance matrix of $\boldsymbol{z}_n$), ν_0 (in-control mean of T_n), σ_T (in-control standard deviation of T_n), and Ω^2 (in-control asymptotic variance constant of T_n). Algorithm 2 provides a detailed description of the estimation process in the setup phase. The combined proposed procedure of Algorithms 1 and 2 is summarized in a flowchart shown in Figure S.2 in Supplement S.2.

Algorithm 1 Distribution-free Image Monitoring Procedure

Input: From the target mean image, obtain $\boldsymbol{m}_0, \boldsymbol{u}_0$, and $\boldsymbol{v}_0$. From in-control training data, obtain $\boldsymbol{\mu}_0, \boldsymbol{\Sigma}$, ν_0, σ_T, and Ω^2. Choose target ARL_0 and k between 0.01 and 0.1.

1: Set $K = k\sigma_T$ and calculate control limit H by solving the following equation:

$$\frac{\Omega^2}{2K^2}\left\{\exp\left[\frac{2K(H+1.166\Omega)}{\Omega^2}\right] - 1 - \frac{2K(H+1.166\Omega)}{\Omega^2}\right\} = \mathrm{ARL}_0. \tag{4}$$

Set $n = 1$ and go to Step 2.

2: Compute $\lambda_n^P = \boldsymbol{u}_0^\top \boldsymbol{Y}_n \boldsymbol{v}_0$ and obtain the largest singular value λ_n^R of $\boldsymbol{R}_n = \boldsymbol{Y}_n - \boldsymbol{m}_0$.

3: Set $\boldsymbol{z}_n = \left[\lambda_n^P, \lambda_n^R\right]^\top$ and compute $T_n = (\boldsymbol{z}_n - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}_n - \boldsymbol{\mu}_0)$.

4: Compute CUSUM statistic S_n^+ as in (2).

5: Raise an out-of-control alarm if $S_n^+ \geq H$. Otherwise, set $n = n + 1$ and go to Step 2.

3.2. Theoretical analysis

This subsection analyzes the in-control and out-of-control mean values of λ_n^P and λ_n^R. Then we discuss the validity of assumptions imposed on $\boldsymbol{z}_n$ and investigate if any shift in the mean matrix from $\boldsymbol{m}_0$ to $\boldsymbol{m}_1$ is accurately reflected in the statistics $\boldsymbol{z}_n$ and T_n.

Algorithm 2 Setup Phase for Distribution-free Image Monitoring Procedure

Input: In-control process $\{\boldsymbol{Y}_n^0 : n = 1, 2, \ldots, N\}$ and target mean $\boldsymbol{m}_0$

1: Perform SVD on $\boldsymbol{m}_0$: $\boldsymbol{m}_0 = \lambda_0 \boldsymbol{u}_0 \boldsymbol{v}_0^\top$.

2: For $n = 1, 2, \ldots, N$, compute $\lambda_n^P = \boldsymbol{u}_0^\top \boldsymbol{Y}_n^0 \boldsymbol{v}_0$ and $\boldsymbol{R}_n^0 = \boldsymbol{Y}_n^0 - \boldsymbol{m}_0$, and perform SVD on $\boldsymbol{R}_n^0$ to obtain the largest singular value λ_n^R. Define $\boldsymbol{z}_n = \left[\lambda_n^P, \lambda_n^R\right]^\top$.

3: Compute in-control sample mean vector denoted by $\hat{\boldsymbol{\mu}}_0$ and sample covariance matrix denoted by $\boldsymbol{S}_z$ using $\boldsymbol{z}_n^0$ for $n = 1, 2, \ldots, N$ as follows:

$$\bar{\boldsymbol{z}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{z}_n^0 \quad \text{and} \quad \boldsymbol{S}_z = \frac{1}{N-1}\sum_{n=1}^{N} (\boldsymbol{z}_n^0 - \bar{\boldsymbol{z}})(\boldsymbol{z}_n^0 - \bar{\boldsymbol{z}})^\top.$$

Estimate $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}$ by $\bar{\boldsymbol{z}}$ and $\boldsymbol{S}_z$, respectively, and pass them to the DFIM procedure.

4: For $n = 1, 2, \ldots, N$, compute $T_n^0 \leftarrow (\boldsymbol{z}_n^0 - \bar{\boldsymbol{z}})^\top \boldsymbol{S}_z^{-1} (\boldsymbol{z}_n^0 - \bar{\boldsymbol{z}})$.

5: Compute sample mean and standard deviation of T_n:

$$\bar{T} \leftarrow \frac{1}{N}\sum_{n=1}^{N} T_n^0 \quad \text{and} \quad s_T \leftarrow \sqrt{\frac{1}{N-1}\sum_{n=1}^{N} (T_n^0 - \bar{T})^2}.$$

Estimate ν_0 and σ_T by $\bar{T}$ and s_T and pass these values to the DFIM procedure.

6: Estimate asymptotic variance parameter Ω^2 of T_n^0 using the CvM estimator (Alexopoulos *et al.*, 2007). The variance estimation algorithm is given in Supplement S.3.

3.2.1. Mean of λ_n^P

Let $\mathbb{E}_0[\cdot]$ and $\mathbb{E}_1[\cdot]$ represent expectation under in-control and out-of-control conditions, respectively. Then the in-control and out-of-control mean values of λ_n^P are as follows:

$$\mathbb{E}_0\left[\lambda_n^P\right] = \mathbb{E}_0\left[\boldsymbol{u}_0^\top \boldsymbol{Y}_n \boldsymbol{v}_0\right] = \boldsymbol{u}_0^\top \mathbb{E}_0[\boldsymbol{Y}_n] \boldsymbol{v}_0 = \boldsymbol{u}_0^\top \boldsymbol{m}_0 \boldsymbol{v}_0 = \lambda_0,$$
$$\mathbb{E}_1\left[\lambda_n^P\right] = \mathbb{E}_1\left[\boldsymbol{u}_0^\top \boldsymbol{Y}_n \boldsymbol{v}_0\right] = \boldsymbol{u}_0^\top \mathbb{E}_1[\boldsymbol{Y}_n] \boldsymbol{v}_0 = \boldsymbol{u}_0^\top \boldsymbol{m}_1 \boldsymbol{v}_0 = \lambda_0 + \boldsymbol{u}_0^\top \boldsymbol{\Delta} \boldsymbol{v}_0.$$

The difference in in-control and out-of-control mean values of λ_n^P is

$$\mathbb{E}_1\left[\lambda_n^P\right] - \mathbb{E}_0\left[\lambda_n^P\right] = \boldsymbol{u}_0^\top \boldsymbol{\Delta} \boldsymbol{v}_0. \tag{5}$$

It is important to note that the difference in (5) may be zero even for a non-zero mean shift $\boldsymbol{\Delta}$ in the image process, resulting in the failure to capture the shift. Therefore, any image monitoring procedure that relies solely on λ_n^P could fail to detect a shift in the process.

3.2.2. Mean of λ_n^R

The analysis of the mean value of λ_n^R is complex because it is the largest singular value of a matrix, $\boldsymbol{R}_n$, with random variables as entries. To gain a better understanding of λ_n^R, we consider a special case where all entries of $\boldsymbol{R}_n$ are assumed to be independent and identically distributed (i.i.d.). In the in-control condition, $\boldsymbol{R}_n$ becomes a random matrix with each entry having expectation 0, while the entries in $\boldsymbol{R}_n$ exhibit different expectations in the out-of-control condition. Then, the following lemmas are applicable to random matrices:

Lemma 1 (Yin *et al.*, 1988). *Suppose $\boldsymbol{C}$ is a $w \times p$ matrix whose entries are i.i.d. random variables having mean zero, variance σ^2, a finite fourth moment, and $w/p \to c$ as $p \to \infty$. Then the largest singular value of $\frac{1}{\sqrt{p}}\boldsymbol{C}$ converges a.s. to $\sigma(1 + \sqrt{c})$ as $p \to \infty$.*

Lemma 2 (Bryce and Silverstein, 2020). *Suppose $\boldsymbol{D}$ is a $w \times p$ random matrix in the following form: $\mathbf{D} = \mathbf{B} + \mathbf{C}$, where $\boldsymbol{B}$ is a deterministic matrix with the largest singular value denoted as η and $\boldsymbol{C}$ is a random matrix satisfying the conditions in Lemma 1. Represent the SVD of $\boldsymbol{B}$ as $\mathbf{B} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^\top = \mathbf{U}\mathbf{G}^\top$, where $\mathbf{G} = \mathbf{V}\boldsymbol{\Lambda}$. Define $\tilde{\Gamma} = \mathbb{E}[\mathbf{C}^\top\mathbf{C}]/w$, $\Sigma_* = \mathbf{G}^\top\tilde{\Gamma}\mathbf{G}, \hat{\Gamma} = \mathbb{E}[\mathbf{C}\mathbf{C}^\top]/p$, and $\Sigma_{**} = \mathbf{U}^\top\hat{\Gamma}\mathbf{U}$. Also, define ψ, a deterministic term depending on the variances of the*

entries of $\boldsymbol{C}$*, as*

$$\psi = \frac{1}{2}\left(\frac{\sqrt{c}}{\gamma^3 wp}[\Sigma_*]_{11} + \frac{1}{\sqrt{c}\gamma}[\Sigma_{**}]_{11}\right),$$

where $c = \lim_{p\to\infty}\frac{w}{p}, \gamma = \lim_{p\to\infty}\frac{\eta}{\sqrt{wp}}$*, and* $[\Sigma_*]_{11}$ *and* $[\Sigma_{**}]_{11}$ *are the (1,1)th entries of* Σ_* *and* Σ_{**}*, respectively. Finally, define Z, a random centered fluctuation term with mean 0, as follows:*

$$Z = \frac{1}{\gamma}\left[\frac{1}{\sqrt{wp}}\boldsymbol{G}^\top \boldsymbol{C}^\top \boldsymbol{U}\right]_{11}.$$

Then the largest singular value of D, denoted by λ*, can be represented as a sum of four terms:* $\lambda = \eta + \psi + Z + \varepsilon$*, where* η *and* ψ *are deterministic terms, Z is a random term that depends on random matrix C, and* $\varepsilon \to 0$ *in probability as* $p \to \infty$*.*

When a monitored process $\{\boldsymbol{Y}_n : n = 1, 2, \ldots\}$ is in control, $\boldsymbol{R}_n$ can be represented as a random matrix and λ_n^R becomes the largest singular value of the random matrix. Then, assuming i.i.d. entries of the random matrix, Lemma 1 implies that for large p, $\mathbb{E}_0[\lambda_n^R] \approx \sigma(\sqrt{w} + \sqrt{p})$. In contrast, when the monitored process is out of control and the mean matrix shifts to $\boldsymbol{m}_1$, a residual matrix is $\boldsymbol{R}_n = \boldsymbol{m}_1 - \boldsymbol{m}_0 + \boldsymbol{\epsilon}_n = \boldsymbol{\Delta} + \boldsymbol{\epsilon}_n$, which can be viewed as a random perturbation of a low-rank matrix $\boldsymbol{\Delta}$. Denote the largest singular value of $\boldsymbol{\Delta}$ as λ_L. When $\boldsymbol{\epsilon}_n$ has i.i.d. entries with variance $\sigma^2, \mathbb{E}[\boldsymbol{\epsilon}_n^\top \boldsymbol{\epsilon}_n]/w$ and $\mathbb{E}[\boldsymbol{\epsilon}_n \boldsymbol{\epsilon}_n^\top]/p$ become diagonal matrices, with all diagonal entries being σ^2. Then in Lemma 2, we have $\gamma = \lambda_L, [\Sigma_*]_{11} = \sigma^2\lambda_L^2, [\Sigma_{**}]_{11} = \sigma^2$, and for large p

$$\mathbb{E}_1[\lambda_n^R] \approx \lambda_L + \frac{\sigma^2(w+p)}{2\lambda_L}.$$

The difference between the in-control and out-of-control mean values of λ_n^R is

$$\begin{aligned}\mathbb{E}_1[\lambda_n^R] - \mathbb{E}_0[\lambda_n^R] &\approx \lambda_L + \frac{\sigma^2(w+p)}{2\lambda_L} - \sigma(\sqrt{w}+\sqrt{p}) \\ &\geq \sigma\sqrt{2(w+p)} - \sigma(\sqrt{w}+\sqrt{p}) \\ &= \sigma(\sqrt{2(w+p)} - \sqrt{w} - \sqrt{p}),\end{aligned} \tag{6}$$

where the inequality holds due to $a + b \geq 2\sqrt{ab}$ for $a, b \geq 0$ with the equality holding when $a = b$, i.e., $\lambda_L = \sigma\sqrt{\frac{w+p}{2}}$. The lower bound in (6) is always non-negative since

$$\begin{aligned}\sqrt{2(w+p)} \geq \sqrt{w} + \sqrt{p} &\iff 2(w+p) \geq w + p + 2\sqrt{wp} \\ &\iff w + p - 2\sqrt{wp} \geq 0 \\ &\iff (\sqrt{w} - \sqrt{p})^2 \geq 0.\end{aligned}$$

When $w \neq p$, the lower bound is strictly positive. Thus, when p is large, $w \neq p$, and $\lambda_L \neq \sigma\sqrt{(w+p)/2}$, a non-zero shift Δ in the image process is reflected into λ_n^R.

We have explored the expectations of λ_n^R in both the in-control and out-of-control conditions and expect the results to hold when auto-correlations exist, as the results are derived from marginal analysis. However, it is unclear whether the results will hold when spatial independence is lost. Spatial independence is the minimum requirement for conducting theoretical analysis in the field of random matrix theory and extending the results to cases with spatial correlation in the random matrix is challenging. Therefore, we conduct experiments in Section 4.2 to numerically verify the results in (6).

3.2.3. Theoretical comparison

From the previous derivations, we have $\mathbb{E}_1[\lambda_n^P] - \mathbb{E}_0[\lambda_n^P] = \boldsymbol{u}_0^\top \boldsymbol{\Delta} \boldsymbol{v}_0$ and $\mathbb{E}_1[\lambda_n^R] - \mathbb{E}_0[\lambda_n^R] \gtrsim \sigma(\sqrt{2(w+p)} - \sqrt{w} - \sqrt{p})$ where the first equation holds for general $\boldsymbol{R}_n$ while the second approximation is derived under the assumption of i.i.d. noise entries with variance σ^2 in $\boldsymbol{\epsilon}_n$. We consider two cases to show the advantages and disadvantages of each statistic.

Case 1: Consider a shift matrix $\boldsymbol{\Delta}$ such that $\boldsymbol{\Delta} = \lambda_L \boldsymbol{u}_d \boldsymbol{v}_d^\top$, where $\boldsymbol{u}_0^\top \boldsymbol{u}_d = 0$. Using (5), we can see that the difference in the in-control and out-of-control mean values of λ_n^P is $\boldsymbol{u}_0^\top \boldsymbol{\Delta} \boldsymbol{v}_0 = \lambda_L \boldsymbol{u}_0^\top \boldsymbol{u}_d \boldsymbol{v}_d^\top \boldsymbol{v}_0 = 0$. This result suggests that any shift $\boldsymbol{\Delta}$ has no effect on the mean of λ_n^P. The same conclusion holds for any shift $\boldsymbol{\Delta}$ such that $\boldsymbol{v}_0^\top \boldsymbol{v}_d = 0$. On the other hand, the mean value of λ_n^R is impacted by the shift $\boldsymbol{\Delta}$ as shown in (6), assuming that the equality does not hold (i.e., $\lambda_L \neq \sigma\sqrt{(w+p)/2}$. Thus, if we want to capture a shift in the monitored process, λ_n^R is a better statistic for this case.

Case 2: Consider a case where $\boldsymbol{u}_d = \boldsymbol{u}_0$ and $\boldsymbol{v}_d = \boldsymbol{v}_0$. In this case, the difference in the in-control and out-of-control means of λ_n^P can be further simplified to λ_L because $\boldsymbol{u}_0$ and $\boldsymbol{v}_0$ are unit vectors. Suppose that

$$\frac{\sigma^2(w+p)}{2\lambda_L} - \sigma(\sqrt{w}+\sqrt{p}) < 0 \quad \text{or} \quad \lambda_L > \frac{\sigma(w+p)}{2(\sqrt{w}+\sqrt{p})}.$$

Then, from (6), we can see that the mean difference of λ_n^R is smaller than λ_L. In this case, λ_n^P is a better statistic because it exhibits a larger shift in the mean, making it more responsive to the shift $\boldsymbol{\Delta}$.

Case 1 demonstrates that λ_n^P alone may fail to capture a shift, highlighting the need to incorporate another statistic to ensure robust mean shift detection. Case 2 shows that λ_n^P can be more effective at capturing a shift than λ_n^R, although our analysis in Section 3.2.2 shows λ_n^R consistently captures a shift. To take advantage of both singular values, we combine the two into a bivariate vector $\boldsymbol{z}_n$ and monitor the vectors. Although one can monitor λ_n^P and λ_n^R individually with two CUSUM charts, the two charts become correlated and make it challenging to set control limits that ensure the overall ARL_0 matches the target.

Further, for $\boldsymbol{\epsilon}_n$ with general spatial correlations, we calculate the difference between the in-control and out-of-control means of T_n statistics because the final monitoring statistics for the DFIM procedure is T_n. When the monitored process is in control, $\mathbb{E}_0[T_n] = \mathbb{E}\left[\text{tr}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_n - \boldsymbol{\mu}_0)(\boldsymbol{z}_n - \boldsymbol{\mu}_0)^\top)\right] = 2$. Let $\text{Cov}_1(\boldsymbol{z}_n)$ represent the covariance matrix of $\boldsymbol{z}_n$ in the out-of-control condition. Then

$$\begin{aligned}\mathbb{E}_1[T_n] &= \text{tr}\left(\boldsymbol{\Sigma}^{-1}\text{Cov}_1(\boldsymbol{z}_n)\right) + \left(\mathbb{E}_1[\boldsymbol{z}_n] - \boldsymbol{\mu}_0\right)^\top \boldsymbol{\Sigma}^{-1}\left(\mathbb{E}_1[\boldsymbol{z}_n] - \boldsymbol{\mu}_0\right) \\ &= 2 + \left(\mathbb{E}_1[\boldsymbol{z}_n] - \boldsymbol{\mu}_0\right)^\top \boldsymbol{\Sigma}^{-1}\left(\mathbb{E}_1[\boldsymbol{z}_n] - \boldsymbol{\mu}_0\right),\end{aligned}$$

where the last equality holds due to Assumption 2 that $\boldsymbol{\Sigma}$ remains unchanged in the out-of-control condition. Then,

we can express the shift in the mean of T_n statistics as follows:

$$\mathbb{E}_1[T_n] - \mathbb{E}_0[T_n] = \left(\mathbb{E}_1[\boldsymbol{z}_n] - \boldsymbol{\mu}_0\right)^\top \boldsymbol{\Sigma}^{-1} \left(\mathbb{E}_1[\boldsymbol{z}_n] - \boldsymbol{\mu}_0\right). \quad (7)$$

If a mean shift $\boldsymbol{\Delta}$ is reflected in λ_n^P or λ_n^R, resulting in a shift in the mean of $\boldsymbol{z}_n$ from $\boldsymbol{\mu}_0$ to $\mathbb{E}_1[\boldsymbol{z}_n] \neq \boldsymbol{\mu}_0$, then (7) is always positive and the proposed procedure can detect the shift.

3.3. Robustness

In this section, we discuss the robustness of the proposed procedure when Assumptions 1, 2, and 4 are violated. Note that the assumption of i.i.d. entries in $\boldsymbol{\epsilon}_n$ is only used for the theoretical analysis of λ_n^R in Section 3.2.2, but the assumption is not needed in this section.

Given that $\boldsymbol{m}_0$ represents the target mean image and is a well-defined reference, the validity of Assumption 1 can be verified easily. In cases where this assumption does not hold, but the decision maker wants to apply the proposed procedure, it essentially performs singular value truncation, retaining only the largest one. As long as a mean shift in $\boldsymbol{Y}_n$ is captured by one of the features λ_n^P or λ_n^R, the conclusions drawn from the theoretical analysis in (7) still hold, which means the proposed procedure can detect the mean shift.

For the violation of Assumption 2, we define $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ as the covariance matrices of $\boldsymbol{z}_n$ in the in-control and out-of-control conditions, respectively. The in-control $\boldsymbol{\Sigma}_0$ replaces $\boldsymbol{\Sigma}$ in the calculation of T_n statistics. Then, the in-control mean of T_n is still 2, but we have the mean shift as follows:

$$\mathbb{E}_1[T_n] - \mathbb{E}_0[T_n] = \left(\mathbb{E}_1[\boldsymbol{z}_n] - \boldsymbol{\mu}_0\right)^\top \boldsymbol{\Sigma}_0^{-1} \left(\mathbb{E}_1[\boldsymbol{z}_n] - \boldsymbol{\mu}_0\right) + \operatorname{tr}\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1\right) - 2. \quad (8)$$

The first term in (8) is identical to the shift size in (7), which is the shift size when Assumption 2 is satisfied. Therefore, we focus on the magnitude of $\operatorname{tr}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1) - 2$. If this magnitude is less than zero, the shift detection may be slower. If it is positive, the detection will be faster. Supplement S.4 shows an analysis of the value of $\operatorname{tr}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1)$. The analysis demonstrates that the proposed procedure will function well in most cases, even when Assumption 2 is violated. Additionally, Supplement S.5 investigates the implication of Assumption 4 violation by using ARL_1 approximations and shows that an increase in Ω^2 accelerates the detection of a shift, while a decrease in Ω^2 does not slow down the detection much when it is already large.

3.4. Computation analysis

Image monitoring applications typically involve high-speed data acquisition processes. To effectively monitor these processes in an online setting, it is essential to have a monitoring procedure with low computational complexity during the monitoring phase.

In the proposed procedure, the majority of the computational efforts are focused on acquiring λ_n^P and λ_n^R. During the monitoring phase, only matrix multiplication is required to obtain λ_n^P, as the singular vectors of the in-control mean matrix are obtained during the setup phase. For λ_n^R, a matrix subtraction and SVD of the residual matrix are performed for each image matrix. After computing λ_n^P and λ_n^R, the T_n statistic is calculated using the in-control mean vector and the inverse covariance matrix estimated during the setup phase. All parameters of the DFIM procedure are also pre-determined during the setup phase. As a result, the SVD of a residual matrix is the dominating operation in the monitoring phase. Fortunately, numerous efficient implementations of SVD have been developed, making the proposed procedure suitable for high-speed image monitoring applications.

4. Numerical experiments

This section compares the proposed procedure with three baseline procedures using numerical experiments. We introduce the testing processes, provide a detailed review of the baseline procedures, and discuss the performance of all four procedures in monitoring images under various cross- and auto-correlation settings.

4.1. Simulated data

We consider two ways to simulate images. Type 1 processes generate vector time series and use them to form image matrices, whereas Type 2 processes directly generate image matrices.

4.1.1. Type 1: Forming an image matrix from vectors

In situations similar to the battery coating, a sensor records a $p \times 1$ observation vector, denoted by $\boldsymbol{x}_t$, at time t. Then, for a window size $w \geq 1$ and an inter-window spacing size $s, 1 \leq s \leq w$, we construct an image matrix, $\boldsymbol{Y}_n$, as follows:

$$\boldsymbol{Y}_n = \begin{bmatrix} \boldsymbol{x}_{(n-1)s+1}^\top \\ \vdots \\ \boldsymbol{x}_{(n-1)s+w}^\top \end{bmatrix}, \quad (9)$$

where $n \geq 1$, and $\boldsymbol{a}^\top$ represents the transpose of a vector $\boldsymbol{a}$. That is, we parallel align $\boldsymbol{x}_t^\top$ into a $w \times p$ matrix, $\boldsymbol{Y}_n$, for $(n-1)s + 1 \leq t \leq (n-1)s + w$. Specifically, when $s = 1$, the matrices correspond to overlapping blocks with a sliding window of size w. On the other hand, when $s = w$, the matrices correspond to non-overlapping blocks of size w. We generate a p-variate vector time-series $\{\boldsymbol{x}_t\}$ as $\boldsymbol{x}_t = \boldsymbol{\alpha} + \boldsymbol{e}_t, \boldsymbol{e}_t = \boldsymbol{\Phi}\boldsymbol{e}_{t-1} + \boldsymbol{\xi}_t$, and $\boldsymbol{e}_0 = \boldsymbol{\xi}_0$ for $t = 1, 2, \ldots$ where $\boldsymbol{\alpha}$ is the marginal mean vector of $\boldsymbol{x}_t, \boldsymbol{e}_t$ are the auto-correlated error terms, $\boldsymbol{\Phi}$ is the auto-correlation coefficient matrix, and $\boldsymbol{\xi}_t$ are i.i.d. normal random vectors with mean 0 and covariance matrix $\boldsymbol{\Sigma}_\xi$. The in-control mean vector is denoted by $\boldsymbol{\alpha}_0$, and the out-of-control mean vector is $\boldsymbol{\alpha}_1$. We construct $\boldsymbol{Y}_n$ following (9) with the inter-window spacing size $s = 1$. The in-control mean matrix of $\boldsymbol{Y}_n$ is a rank-one matrix: $\boldsymbol{m}_0 = \mathbb{E}_0[\boldsymbol{Y}_n] = \begin{bmatrix} \boldsymbol{\alpha}_0 & \boldsymbol{\alpha}_0 & \cdots & \boldsymbol{\alpha}_0 \end{bmatrix}^\top$. We set $\boldsymbol{\alpha}_0$ to be a constant vector with a value of five, which represents a target

uniform height. We also set $p = 200$ and $w = 5$. The auto-correlation coefficient matrix, $\boldsymbol{\Phi}$, is chosen to be a diagonal matrix where $[\boldsymbol{\Phi}]_{ij} = \phi$ if $i = j$ and 0 otherwise. The diagonal entry ϕ is set to 0.3 or 0.7 to test the effect of low and high auto-correlation, respectively. We also test two different cross-correlation structures by varying $\boldsymbol{\Sigma}_\xi$ between the following two models with $\rho = 0.3$: (i) tri-diagonal model: $[\boldsymbol{\Sigma}_\xi]_{ij} = 1$ if $i = j$; ρ if $|i-j| = 1$; 0 otherwise; and (ii) exponential model: $[\boldsymbol{\Sigma}_\xi]_{ij} = \rho^{|i-j|}$.

To simulate out-of-control scenarios, we add a 200-dimensional vector $\boldsymbol{\delta} = [\delta_j]$ for $j = 1, 2, \ldots, 200$ to $\boldsymbol{\mu}_0$. the resulting out-of-control mean vector of $\boldsymbol{x}_t$ is $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \boldsymbol{\delta}$. We consider three different shifts of $\boldsymbol{\delta}$ in our experiments:

1. Sparse: for $1 \le j \le 200$,
$$\delta_j = \begin{cases} d, & \text{if } 18 \le j \le 22, \\ 0, & \text{otherwise.} \end{cases}$$
2. Step: for $1 \le j \le 200$,
$$\delta_j = \begin{cases} \frac{d}{3}, & \text{if } 51 \le j \le 100, \\ \frac{2d}{3}, & \text{if } 101 \le j \le 150, \\ d, & \text{if } 151 \le j \le 200, \\ 0, & \text{otherwise.} \end{cases}$$
3. Zigzag: for $0 \le k \le 4, 1 \le \ell \le 20$,
$$\delta_j = \begin{cases} d\left(1 - \frac{j - 40k}{10}\right), & \text{if } j = 40k + \ell, \\ d\left(-1 + \frac{j - 40k - 20}{10}\right), & \text{if } j = 40k + 20 + \ell. \end{cases}$$

The value of d is adjusted to control the magnitude of the shift, which is defined as $||\boldsymbol{\delta}|| = \sqrt{\sum_j \delta_j^2}$ for Type 1 simulated processes. From the image process $\{\boldsymbol{Y}_n, n = 1, 2, \ldots\}$, the in-control and out-of-control mean matrices are defined as follows:

$$\boldsymbol{m}_0 = \begin{bmatrix} \boldsymbol{\alpha}_0^\top \\ \boldsymbol{\alpha}_0^\top \\ \vdots \\ \boldsymbol{\alpha}_0^\top \end{bmatrix} \quad \text{and} \quad \boldsymbol{m}_1 = \begin{bmatrix} \boldsymbol{\alpha}_1^\top \\ \boldsymbol{\alpha}_1^\top \\ \vdots \\ \boldsymbol{\alpha}_1^\top \end{bmatrix} = \boldsymbol{m}_0 + \boldsymbol{\Delta} \text{ where } \boldsymbol{\Delta} = \begin{bmatrix} \boldsymbol{\delta}^\top \\ \boldsymbol{\delta}^\top \\ \vdots \\ \boldsymbol{\delta}^\top \end{bmatrix}.$$

Figure 4 displays the two-dimensional image representations of the shift matrix $\boldsymbol{\Delta}$ for sparse, step, and zigzag shifts when $d = 1$.

4.1.2. Type 2: Generating image matrices

Different from Type 1 data, Type 2 processes directly generate a series of matrices $\{\boldsymbol{Y}_n : n = 1, 2, \ldots\}$ using $\boldsymbol{Y}_n = \boldsymbol{m} + \boldsymbol{\epsilon}_n$, where $\boldsymbol{Y}_n$ is a $w \times p$ matrix, $\boldsymbol{m}$ is the mean matrix of $\boldsymbol{Y}_n$, and $\boldsymbol{\epsilon}_n$ is a matrix of random noises. Thanks to advances in sensor technologies, capturing image data directly, such as monitoring solar flare activity, has become routine.

For our experiments, we set $w = 100$ and $p = 200$. The noise matrices $\boldsymbol{\epsilon}_n$ are generated as i.i.d. random matrices with a covariance structure represented by $\boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$, where $\boldsymbol{\Sigma}_c$ captures covariance among the columns of $\boldsymbol{\epsilon}_n$, $\boldsymbol{\Sigma}_r$ captures covariance among the rows, and $\otimes$ denotes the Kronecker product. Two distributions are considered for noises:

1. Normal noise: Each $\boldsymbol{\epsilon}_n$ is generated from a matrix normal distribution with covariance $\boldsymbol{\Sigma}_c \otimes \boldsymbol{\Sigma}_r$. Each entry $[\boldsymbol{\epsilon}_n]_{ij}$ has mean 0 and variance 1.
2. Non-normal noise: We start by generating a normally distributed matrix $\boldsymbol{\epsilon}'_n$ following the previous normal-noise case. Then each entry $[\boldsymbol{\epsilon}'_n]_{ij}$ is transformed into $[\boldsymbol{\epsilon}_n]_{ij} = -\log(1 - \Psi([\boldsymbol{\epsilon}'_n]_{ij}))$, where $\Psi(\cdot)$ represents the cumulative distribution function of the standard normal random variable. Consequently, each entry $[\boldsymbol{\epsilon}_n]_{ij}$ follows an exponential distribution with mean 1, and entries in $[\boldsymbol{\epsilon}_n]_{ij}$ are correlated.

We do not incorporate auto-correlation in Type 2 simulated processes since it is already covered by Type 1 processes. Instead, Type 2 processes evaluate the capability to handle non-normal data. Similar to $\boldsymbol{\Sigma}_\xi$ in Section 4.1.1, $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\Sigma}_r$ are modeled either by a tri-diagonal model or an exponential model with $\rho = 0.3$. The form of $\boldsymbol{\Sigma}_r$ is set to be the same as $\boldsymbol{\Sigma}_c$, but with different dimensions. The in-control mean matrix $\boldsymbol{m}_0$ is a constant matrix with all entries equal to five.

To simulate various out-of-control scenarios in images, shift matrices $\boldsymbol{\Delta} = [\Delta_{ij}]$ are added to the in-control mean matrix $\boldsymbol{m}_0$ so that $\boldsymbol{m}_1 = \boldsymbol{m}_0 + \boldsymbol{\Delta}$. We consider five shifts:

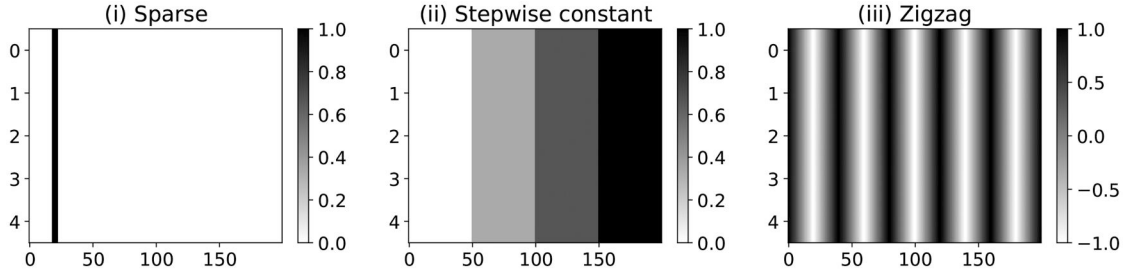


Figure 4. Three shifts of $\boldsymbol{\Delta}$ for Type 1 simulated processes when $d = 1$.

1. Sparse:

$$\Delta_{ij} = \begin{cases} d, & \text{if } 8 \le i \le 13 \text{ and } 18 \le j \le 23, \\ 0, & \text{otherwise.} \end{cases}$$

2. Chessboard: for $0 \le k \le 4, 1 \le \ell \le 10, \quad 0 \le r \le 9, 1 \le s \le 5,$

$$\Delta_{ij} = \begin{cases} d, & \text{if } j = 40k + 10 + \ell \text{ and } i = 10r + s, \\ d, & \text{if } j = 40k + 20 + \ell \text{ and } i = 10r + 5 + s, \\ -d, & \text{if } j = 40k + 30 + \ell \text{ and } i = 10r + s, \\ -d, & \text{if } j = 40k + \ell \text{ and } i = 10r + 5 + s, \\ 0, & \text{otherwise.} \end{cases}$$

3. Ring-wise constant: for $k \ge 0, 0 \le \ell \le 3,$

$$\Delta_{ij} = \begin{cases} d, & \text{if } \lfloor \sqrt{(i-50)^2 + (j-100)^2} \rfloor = 12k + \ell, \\ -d, & \text{if } \lfloor \sqrt{(i-50)^2 + (j-100)^2} \rfloor = 12k + 8 + \ell, \\ 0, & \text{otherwise.} \end{cases}$$

4. Row-wise sine: for $1 \le j \le 200,$

$$\Delta_{ij} = \begin{cases} d \sin(j\pi/10), & \text{if } 1 \le i \le 30, \\ d \sin(2j\pi/10), & \text{if } 31 \le i \le 60, \\ d \sin(3j\pi/10), & \text{if } 61 \le i \le 100. \end{cases}$$

5. Column-wise sine: for $1 \le i \le 100,$

$$\Delta_{ij} = \begin{cases} d \sin(i\pi/5), & \text{if } 1 \le j \le 60, \\ d \sin(2i\pi/5), & \text{if } 61 \le j \le 120, \\ d \sin(3i\pi/5), & \text{if } 121 \le j \le 200. \end{cases}$$

To achieve a different magnitude of a shift for Type 2 simulated processes, the value of d can be adjusted. The magnitude of the shift is defined by $||\boldsymbol{\Delta}|| = \sqrt{\sum_{i,j} \boldsymbol{\Delta}_{ij}^2}$. Figure 5 illustrates these five shifts when $d = 1$.

4.2. Verifying theoretical results for dependent measurements

The analysis in Section 3.2 shows that λ_n^R effectively captures a shift in the mean pattern under the assumption of i.i.d. entries in the random matrix. In this section, we investigate whether the results hold for a matrix with dependent entries by comparing the in-control and out-of-control means of λ_n^R using simulated data. Specifically, we use a Type 1 simulated process with $\phi = 0.3$, the tri-diagonal model with $\rho = 0.3$, and a sparse shift with $||\boldsymbol{\delta}|| = 5$.

In our simulation, we have $p = 200$ and $w = 5$ with $\sigma^2 = 1$. According to (6), the theoretical difference between $\mathbb{E}_0[\lambda_n^R]$ and $\mathbb{E}_1[\lambda_n^R]$ is approximately $\sqrt{2(5+200)} - \sqrt{5} - \sqrt{200} = 3.87$. Using simulated data with dependent measurements, we obtain $\mathbb{E}_0[\lambda_n^R] \approx 19.31$ and $\mathbb{E}_1[\lambda_n^R] \approx 22.14$, yielding the difference ≈ 2.83. We also use simulated data with independent measurements, which have matching means and marginal variances. From the simulated data with independent measurements, we get $\mathbb{E}_0[\lambda_n^R] \approx 15.74$ and $\mathbb{E}_1[\lambda_n^R] \approx 18.30$. The difference between the means is 2.56, which is very close to the difference we obtained for the simulated data with dependent measurements.

Although this experiment does not provide a definitive conclusion, it offers evidence that λ_n^R is capable of accurately capturing a shift in the mean pattern, even when the assumption of i.i.d. measurements is violated.

4.3. Baseline procedures and experimental results

As baseline procedures, we use one profile monitoring and two image monitoring procedures:

1. Multivariate Exponentially Weighted Moving Average (MEWMA) (Wang and Lai, 2019): This procedure fits a regression model to monitor auto-correlated profile vectors. For Type 1 simulated processes, each observation $\boldsymbol{x}_t$ is treated as a profile vector, and the MEWMA procedure is directly applied. However, for Type 2 simulated processes, we vectorize each matrix observation $\boldsymbol{Y}_n$ into a profile of length wp and apply the MEWMA procedure. Specifically, a $w \times p$ matrix $\boldsymbol{Y}_n$ is equivalent to a profile vector $\text{vec}(\boldsymbol{Y}) = (Y_{11}, Y_{12}, \ldots, Y_{1p}, Y_{21}, Y_{22}, \ldots, Y_{2p}, \ldots, Y_{w1}, Y_{w2}, \ldots, Y_{wp})$. The control limit of the procedure is adjusted by trial-and-error simulation for a given target ARL_0.

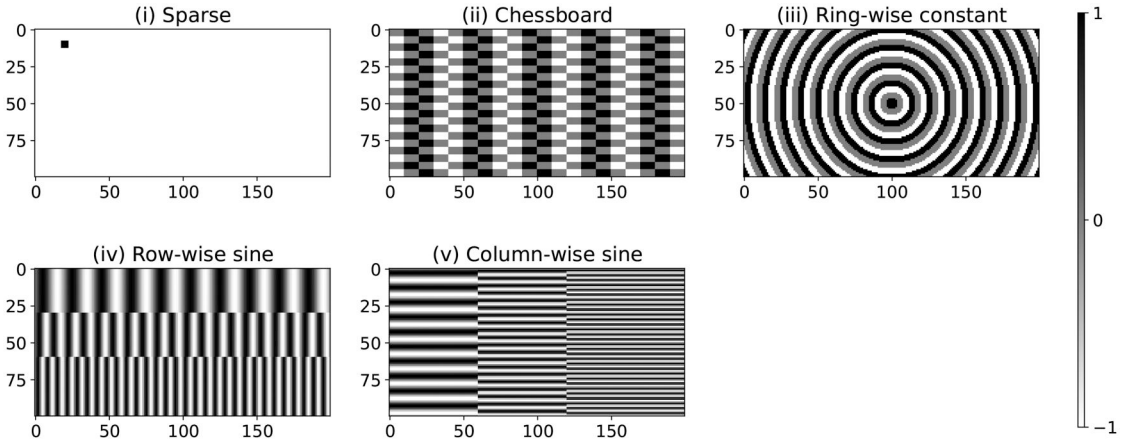


Figure 5. Five shifts of $\boldsymbol{\Delta}$ for Type 2 simulated processes when $d = 1$.

2. Multivariate Generalized Likelihood Ratio (MGLR) (He *et al.*, 2016): The procedure is designed to monitor products with a specific in-control pattern, such as the uniform pattern in LED screens. The MGLR procedure reduces the dimensionality of the original image by dividing it into q equal-sized rectangular areas and focuses on monitoring the summarized quantities from these areas. For Type 1 processes, a rectangular area of 1 by 10 matrix is used with $q = 5 \times 20 = 100$. For Type 2 processes, a rectangular area of 10 by 10 matrix is used with $q = 10 \times 20 = 200$. The control limit of the procedure is determined by trial-and-error simulation for a given target ARL_0.
3. Spatio-Temporal Smooth Sparse Decomposition Model (ST-SSD) (Yan *et al.*, 2018): This procedure can handle streaming data with high dimensionality, high velocity, and complex spatio-temporal structure. The ST-SSD procedure decomposes steaming signals into dynamic background signals and anomalies using regression. The locally estimated anomalies form testing statistics and allow for efficient image monitoring.

In the in-control simulated experiments, we assess the accuracy of the analytically determined control limits by verifying whether they yield actual ARL_0 values close to the target. In the out-of-control simulated experiments, we demonstrate the effectiveness of the proposed procedure in accurately detecting anomalies.

4.3.1. In-control performance

We generate an in-control sample path in each setting to train the three baseline procedures and the DFIM procedure. The length of the in-control sample path for Type 1 and Type 2 settings are 50,000 and 30,000, respectively. For all settings, a target ARL_0 is set to 1000. The control limit of the DFIM procedure is calculated analytically in (4) using estimated parameters and $k = 0.01$. For the ST-SSD procedure, the control limit is calculated as a 0.1% $(= 1/ARL_0)$ sample quantile from a series of monitoring statistics. However, the control limits of the other two baseline procedures need calibration using trial-and-error simulation. We generate 1000 in-control independent sample paths to calibrate their control limits to achieve a target $ARL_0 = 1000$. We use the same set of 1000 in-control paths to verify if the control limits of the ST-SSD and DFIM procedure, whether a quantile estimate or an analytically determined value, produce actual ARL_0 close to the target 1000.

We report the sample standard deviations of the ARLs throughout the simulated experiments, from which the standard error can be calculated as the standard deviation divided by $\sqrt{1000}$. Table 1 shows the control limits of all procedures and their corresponding estimated ARL_0 for each setting of simulated processes. Except for the ST-SSD procedure, the MEWMA, MGLR, and DFIM procedures produce accurate ARL_0 in each setting. The MEWMA and MGLR procedures rely on trial-and-error simulation to ensure the accuracy of their ARL_0. In contrast, the analytical control limit of the DFIM procedure achieves the same accuracy with significantly reduced time and cost. The ST-SSD procedure does not require trial-and-error simulation for control limit calibration, but it does not produce actual ARL_0 close to the target when the covariance matrix is tri-diagonal. This is expected as the control limit is essentially an extreme quantile estimate, which is difficult to estimate accurately with 50,000 or 30,000 data points, especially for auto-correlated data.

4.3.2. Out-of-control performance

In the out-of-control scenario, we independently generate 1000 sample paths and evaluate the performance of the baseline and proposed procedures. Tables 2, 3, and S.2 (in Supplement S.6) compare ARL_1 values of the four procedures for Type 1 and Type 2 simulated processes. The smallest ARL_1 value in each setting is indicated in bold.

Table 2 shows that increasing the auto-correlation parameter ϕ leads to longer detection delays. We do not observe any one procedure dominating the monitoring behavior across different covariance patterns. The DFIM procedure has the smallest ARL_1 values for the sparse and zigzag shifts. The three baseline procedures fail to detect the zigzag shift according to their ARL_1 values, which are close to or larger than the target ARL_0 of 1000. When the shift is stepwise, the DFIM procedure has larger ARL_1 values than the MEWMA procedure, but smaller than the MGLR procedure. The difference in performance between the MEWMA and DFIM procedures is not significant. The ST-SSD procedure fails to detect any shift in Type 1 simulated processes because the decomposition model does not capture the shape information of the background. The designed anomalies in Type 1 data are regular enough to blend into the background signal, as shown in Figure 4.

Table 3 shows the DFIM procedure consistently outperforms the MEWMA and MGLR procedures for all cases for normal noises. The ST-SSD procedure almost fails in the

Table 1. Control limit (H) and actual ARL_0 for various settings of simulated processes with target $ARL_0 = 1000$ (standard deviation in parentheses).

		MEWMA		*MGLR*		*ST-SSD*		*DFIM*	
Process	*Setting*	H	ARL_0	H	ARL_0	H	ARL_0	H	ARL_0
Type 1	$\phi = 0.3$, Tri-diagonal	14.12	999.3(949.82)	62.63	1004.0(1005.44)	6.76	1448.4(1318.88)	107.15	1000.7(838.79)
	$\phi = 0.7$, Tri-diagonal	12.77	1003.4(1006.92)	80.52	1000.7(960.77)	6.36	1212.4(1165.92)	125.45	1000.4(822.89)
	$\phi = 0.3$, Exponential	16.58	999.9(1018.11)	62.45	1002.6(982.43)	6.34	935.7(942.25)	107.43	1005.5(846.57)
	$\phi = 0.7$, Exponential	15.1	998.3(980.56)	81.08	1000.6(1027.23)	6.25	1004.1(986.81)	129.54	998.5(881.62)
Type 2	Normal, Tri-diagonal	23.74	1004.0(955.21)	79.39	1001.9(1041.45)	3.46	749.3(723.02)	59.25	1004.8(871.27)
	Normal, Exponential	31.85	999.1(959.62)	79.55	1003.7(1050.18)	3.72	1035.1(993.67)	54.84	1007.6(897.15)
	Non-normal, Tri-diagonal	21.73	1004.7(1005.09)	79.88	995.4(1019.76)	3.59	1254.5(1170.00)	55.85	1000.7(830.95)
	Non-normal, Exponential	28.4	1002.8(1001.63)	80.05	996.0(1013.58)	3.63	956.4(933.61)	54.24	998.2(850.07)

Table 2. ARL_1 for out-of-control Type 1 processes (standard deviation in parentheses).

Setting	*Shape of* $\boldsymbol{\delta}$	$\|\|\boldsymbol{\delta}\|\|$	*MEWMA*	*MGLR*	*ST-SSD*	*DFIM*
	Sparse	2	291.7 (291.72)	447.9 (688.02)	1406.2 (1316.85)	**210.7 (144.20)**
		5	47.9 (41.00)	24.6 (20.19)	1384.9 (1263.24)	**12.8 (5.47)**
$\phi = 0.3$	Step	1	**48.1 (39.90)**	154.3 (143.42)	1437.7 (1338.98)	99.3 (53.03)
Tri-diagonal		2	**11.0 (7.64)**	18.4 (13.91)	1421.3 (1269.78)	25.3 (11.51)
	Zigzag	2	960.0 (939.82)	1015.7 (1029.34)	1355.2 (1218.11)	**274.3 (176.25)**
		5	937.7 (911.54)	1031.1 (1028.94)	1423.0 (1254.30)	**14.4 (6.54)**
	Sparse	2	688.9 (697.18)	721.5 (715.70)	1175.5 (1104.74)	**679.8 (577.89)**
		5	225.5 (212.05)	227.6 (220.13)	1254.9 (1178.36)	**108.0 (62.34)**
$\phi = 0.7$	Step	1	**218.2** (205.62)	557.4 (542.06)	1121.9 (1083.92)	330.3 (241.16)
Tri-diagonal		2	**55.1** (48.65)	177.4 (175.92)	1142.6 (1107.68)	103.8 (59.99)
	Zigzag	2	964.7 (950.86)	1031.8 (1049.91)	1224.3 (1150.79)	**758.3 (652.91)**
		5	989.2 (1011.31)	998.2 (966.88)	1221.5 (1143.93)	**134.6 (80.48)**
	Sparse	2	330.0 (334.51)	315.6 (320.62)	974.9 (925.04)	**224.2 (147.39)**
		5	58.2 (53.52)	30.5 (26.58)	967.6 (927.13)	**13.4 (6.06)**
$\phi = 0.3$	Step	1	**56.7 (51.52)**	167.4 (164.19)	957.0 (925.70)	113.4 (63.63)
Exponential		2	**12.9 (8.58)**	22.3 (17.69)	925.0 (885.52)	28.9 (13.41)
	Zigzag	2	1024.7 (1056.13)	971.4 (944.37)	1037.2 (960.14)	**285.1 (192.45)**
		5	1029.3 (1049.97)	939.8 (928.26)	984.3 (951.41)	**14.5 (6.79)**
	Sparse	2	732.6 (721.64)	736.6 (740.58)	956.4 (945.17)	**636.7 (525.47)**
		5	266.7 (263.71)	314.9 (447.98)	1009.3 (938.59)	**114.6 (67.55)**
$\phi = 0.7$	Step	1	**271.6 (278.13)**	666.2 (763.75)	991.4 (948.35)	404.7 (414.44)
Exponential		2	**68.1 (63.77)**	233.4 (325.82)	981.6 (934.88)	115.7 (68.84)
	Zigzag	2	1052.6 (1084.03)	848.2 (812.43)	993.5 (925.58)	**714.9 (585.80)**
		5	1013.2 (1081.01)	807.5 (719.61)	1001.6 (961.35)	**138.4 (99.18)**

Table 3. ARL_1 for out-of-control Type 2 processes with normal noises (standard deviation in parenthesis).

Setting	*Shape of* $\boldsymbol{\Delta}$	$\|\|\boldsymbol{\Delta}\|\|$	*MEWMA*	*MGLR*	*ST-SSD*	*DFIM*
	Sparse	10	191.5 (175.72)	180.5 (173.13)	750.3 (724.58)	**134.4 (71.8)**
		20	47.2 (39.09)	13.1 (7.67)	735.9 (722.49)	**1.9 (0.77)**
	Chessboard	10	1038.8 (996.77)	983.9 (944.25)	**238.8 (237.99)**	271.9 (171.43)
		20	1008.5 (1022.65)	986.5 (949.56)	19.4 (19.16)	**7.2 (2.66)**
Normal	Ring	20	218.9 (213.86)	56.1 (52.89)	**9.2 (8.40)**	35.2 (13.31)
Tri-diagonal		30	100.1 (90.13)	11.9 (6.62)	**1.2 (0.54)**	5.2 (1.72)
	Sine (row)	20	976.3 (1008.57)	984.5 (963.12)	49.6 (48.50)	**12.7 (4.45)**
		30	990.2 (1037.35)	1039.9 (1089.72)	4.9 (4.54)	**1.8 (0.61)**
	Sine (column)	20	1003.7 (1057.11)	982.6 (1021.21)	**8.6 (8.11)**	36.8 (13.89)
		30	982.4 (1001.09)	998.7 (969.84)	**1.2 (0.52)**	3.0 (0.97)
	Sparse	10	257.6 (259.43)	270.7 (253.01)	997.9 (1005.54)	**131.4 (73.09)**
		20	64.7 (54.17)	21.0 (15.12)	1052.6 (1035.37)	**2.2 (0.88)**
	Chessboard	10	1031.4 (1001.81)	1091.2 (1055.10)	381.0 (417.71)	**218.5 (135.70)**
		20	958.4 (906.65)	1050.2 (1052.08)	32.2 (34.53)	**7.0 (2.63)**
Normal	Ring	20	284.1 (289.60)	101.8 (101.10)	**14.9 (14.68)**	32.6 (12.67)
Exponential		30	139.8 (129.21)	19.2 (14.44)	**1.5 (0.85)**	5.2 (1.79)
	Sine (row)	20	958.6 (934.14)	1043.6 (1006.06)	79.6 (78.04)	**12.0 (4.26)**
		30	962.0 (973.97)	973.6 (950.01)	7.5 (6.66)	**1.9 (0.63)**
	Sine (column)	20	1042.0 (1044.64)	1073.1 (1079.59)	**14.2 (13.63)**	52.6 (22.52)
		30	997.3 (985.75)	1015.8 (1006.50)	**1.5 (0.84)**	4.2 (1.36)

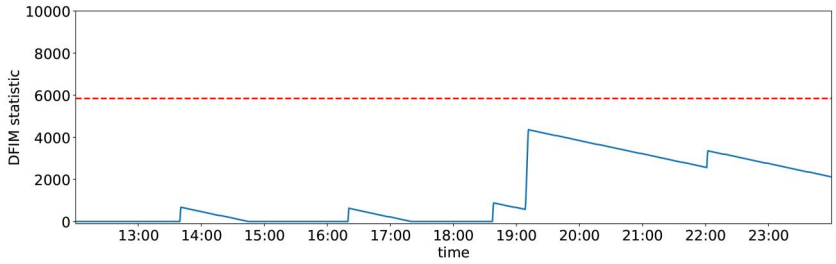
sparse setting, while the ST-SSD and DFIM procedures perform comparably in the remaining settings. When the shift matrix is normally distributed, the DFIM procedure performs slightly better than ST-SSD. For non-normal noises, the DFIM procedure gains an edge and has the smallest ARL_1 except for sparse and ring shifts as shown in Table S.2 in Supplement S.6. The DFIM procedure never fails to detect shifts in any tested setting, which verifies its robustness. The MEWMA and MGLR procedures fail under the chessboard, row-wise, and column-wise sine shifts, and the ST-SSD procedure fails under sparse shifts, while the DFIM procedure produces relatively quick detection.

Overall, the DFIM procedure is efficient, requiring less time and data than the baseline procedures to determine control limits for a given target ARL_0. In most cases tested in this section, it detects various out-of-control shifts with the smallest ARL_1 compared with the baseline procedures, demonstrating competitive performance and its ability to detect even difficult shifts that the baseline procedures fail to detect.

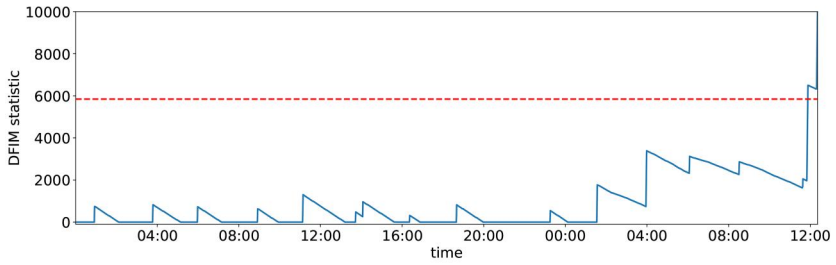
5. Battery coating process

In this section, we demonstrate the performance of the DFIM procedure in a practical application using the battery coating process described in Section 1.

We analyze a time-series dataset containing measurements of battery coating thickness from a real production line. The dataset covers the period from August 8 to August 11, 2018, and includes a total of 27, 311 sensor trips. Maintenance was carried out between August 10 at 13:00 and August 11 at 0:00. The production manager reviewed other indicators and believed that a shift occurred around August 10 at 12:00. To set up the DFIM procedure, we use



(a) Training data from August/8 12:00 to August/9 0:00



(b) Testing data from August/9 0:00 to August/10 13:00

Figure 6. Performance of the DFIM procedure on the battery coating process.

the battery coating data between August 8 at 12:00 and August 9 at 0:00 as training data. Each sensor trip captures a vector of 298 thickness measurements ($p = 298$), and we use a sliding window size of $w = 5$ with an inter-window spacing size of $s = 1$.

Analysis in Supplement S.7 verifies that the battery coating data satisfies Assumptions 1, 2, and 4. To perform the DFIM procedure, the training dataset consists of 5558 image matrices, resulting in 5558 T_n^0 training statistics. We set the target $\text{ARL}_0 = 22,248$, which is four times the training size to ensure a low false alarm rate. Due to the limited availability of sample paths, trial-and-error simulation is not possible, and as a result, the MEWMA and MGLR procedures are not applied. Furthermore, the ST-SSD procedure requires an extreme 0.005% sample quantile, which cannot be accurately estimated with only 5558 samples. Consequently, we do not employ the ST-SSD procedure.

From the training data, we estimate parameters $\boldsymbol{\mu}_0, \nu_0, \sigma_T$, and Ω^2, and the control limit is calculated with the target $\text{ARL}_0 = 22,248$ and $k = 0.01$ using (4). After estimating the in-control covariance matrix of $\boldsymbol{z}_n$ as explained in Supplement S.7, the DFIM procedure is applied separately to the training data and the remaining data. The results of these applications are shown in Figures 6(a) and 6(b), respectively. The dashed lines in the figures represent the control limits, and the solid lines represent the monitoring statistics of the DFIM procedure. During the training period, the DFIM procedure does not raise any false alarms while an alarm is triggered just before August 10 at 12:00, which highlights the effective detection power of our procedure.

6. Conclusion

We present a novel distribution-free procedure that employs SVD to monitor image processes with general marginals, cross-correlations, and auto-correlations. The DFIM procedure offers a significant advantage as it eliminates the need for trial-and-error simulation to determine control limits, thereby allowing for convenient and rapid implementation in practical settings. Our numerical experiments, conducted on both simulated and real processes, demonstrate the effectiveness and robustness of our procedure in anomaly detection.

While the DFIM procedure was originally motivated by monitoring battery coating processes, it can be easily extended to other applications that require image monitoring provided that the in-control mean image is rank-one. Our ongoing research is focused on extending the procedure to accommodate low-rank mean images, which would further expand its applicability to a broader range of applications.

Notes on contributors

Tingnan Gong is a PhD student in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Tech. He received a BS degree in statistics from University of Science and Technology of China (USTC), Hefei, China in 2020, and is currently the 3rd year PhD student.

Di Liu is a data scientist at Google. He received a BS degree in information and computing science from Zhejiang University, Hangzhou, China, in 2011, and a PhD degree in operations research from Georgia Tech, Atlanta, USA, in 2022.

Heeseon Kim works as a senior researcher at LG Electronics in Korea. She has 15 years of working career in the area of business innovation utilizing strategic decision-making based on data. She is currently focusing on manufacturing solutions that can establish smart factories. She has her Master's degree in industrial engineering from the Korea Advanced Institute of Science and Technology in the Republic of Korea in 2006.

Seong-Hee Kim is a professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Tech. She received her

PhD in industrial engineering and management sciences from Northwestern University in 2001. Her research interests include data-based decision-making, simulation optimization, spatiotemporal monitoring, and applications to environmental management and manufacturing.

Taeheung Kim received his PhD degree in industrial engineering from Sungkyunkwan University, Korea, where he now works as a research professor. Before rejoining Sungkyunkwan University, he worked with the Production Research Institute, LG Electronics, as a member of the AI solution team. His main research interests include the design of learning and optimization algorithms to improve real manufacturing processes

Dongki Lee received his BS degree from Hanyang University, Seoul, Republic of Korea, in 2013 and a MS degree from Seoul National University, Seoul, Republic of Korea, in 2018. He is currently working for LG Electronics. His current research topics include prognostics and health management for electric machines.

Yao Xie is the Harold R. and Mary Anne Nash Early Career Professor and Associate Professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Tech. Her research interests are in sequential statistical methods, statistical signal processing, big data analysis, compressed sensing, optimization, and has been involved in applications to wireless communications, sensor networks, medical and astronomical imaging. Dr. Xie previously served as Research Scientist in the Electrical and Computer Engineering Department at Duke University after receiving her PhD in electrical engineering (minor in mathematics) from Stanford University in 2011.

ORCID

Tingnan Gong http://orcid.org/0000-0002-1542-5787
Seong-Hee Kim http://orcid.org/0000-0001-9536-5755
Yao Xie http://orcid.org/0000-0001-6777-2951

Data availability statement

The authors of this study from LG Electronics did not agree that their data from a battery coating process be shared publicly due to security issues, so supporting data obtained from the battery coating process is unavailable.

References

Alaeddini, A., Motasemi, A. and Faruqui, S.H.A. (2018) A spatiotemporal outlier detection method based on partial least squares discriminant analysis and area Delaunay triangulation for image-based process monitoring. *IISE Transactions*, **50**(2), 74–87.

Alexopoulos, C., Argon, N.T., Goldsman, D., Tokol, G. and Wilson, J.R. (2007) Overlapping variance estimators for simulation. *Operations Research*, **55**(6), 1090–1103.

Alwan, L.C. (1992) Effects of autocorrelation on control chart performance. *Communications in Statistics – Theory and Methods*, **21**(4), 1025–1049.

Amirkhani, F. and Amiri, A. (2020) A novel framework for spatiotemporal monitoring and post-signal diagnosis of processes with image data. *Quality and Reliability Engineering International*, **36**(2), 705–735.

Atashgahi, Z., Mocanu, D.C., Veldhuis, R. and Pechenizkiy, M. (2023) Memory-free online change-point detection: A novel neural network approach. *arXiv preprint arXiv:2207.03932*.

Bryc, W. and Silverstein, J.W. (2020) Singular values of large non-central random matrices. *Random Matrices: Theory and Applications*, **9**(04), 2050012.

Colosimo, B.M. and Grasso, M. (2018) Spatially weighted PCA for monitoring video image data with application to additive manufacturing. *Journal of Quality Technology*, **50**(4), 391–417.

Colosimo, B.M. and Pacella, M. (2010) A comparison study of control charts for statistical monitoring of functional data. *International Journal of Production Research*, **48**(6), 1575–1601.

Ebrahimzadeh, Z., Zheng, M., Karakas, S. and Kleinberg, S. (2019) Deep learning for multi-scale changepoint detection in multivariate time series. *arXiv preprint arXiv:1905.06913*.

Eslami, D., Izadbakhsh, H., Ahmadi, O. and Zarinbal, M. (2021) Spatial-nonparametric regression: An approach for monitoring image data. *Communications in Statistics – Theory and Methods*, **52**, 1–24.

Garcia, G.R., Michau, G., Ducoffe, M., Gupta, J.S. and Fink, O. (2022) Temporal signals to images: Monitoring the condition of industrial assets with deep learning image processing algorithms. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, **236**(4), 617–627.

Glynn, P.W. and Iglehart, D.L. (1985) Large-sample theory for standardized time series: An overview, in *Proceedings of the 17th Conference on Winter Simulation*, Association for Computing Machinery, New York, NY, pp. 129–134.

Gupta, M., Wadhvani, R. and Rasool, A. (2022) Real-time change-point detection: A deep neural network-based adaptive approach for detecting changes in multivariate time series data. *Expert Systems with Applications*, **209**, 118260.

He, Z., Zuo, L., Zhang, M. and Megahed, F.M. (2016) An image-based multivariate generalized likelihood ratio control chart for detecting and diagnosing multiple faults in manufactured products. *International Journal of Production Research*, **54**(6), 1771–1784.

Hushchyn, M., Arzymatov, K. and Derkach, D. (2020) Online neural networks for change-point detection. *arXiv preprint arXiv: 2010.01388*.

Kazemzadeh, R., Noorossana, R. and Amiri, A. (2009) Monitoring polynomial profiles in quality control applications. *The International Journal of Advanced Manufacturing Technology*, **42**(7), 703–712.

Khedmati, M. and Niaki, S.T.A. (2016) Phase II monitoring of general linear profiles in the presence of between-profile autocorrelation. *Quality and Reliability Engineering International*, **32**(2), 443–452.

Kim, S.-H., Alexopoulos, C., Tsui, K.-L. and Wilson, J.R. (2007) A distribution-free tabular CUSUM chart for autocorrelated data. *IIE Transactions*, **39**(3), 317–330.

Koosha, M., Noorossana, R. and Megahed, F. (2017) Statistical process monitoring via image data using wavelets. *Quality and Reliability Engineering International*, **33**(8), 2059–2073.

Lee, J., Hur, Y., Kim, S.-H. and Wilson, J.R. (2012) Monitoring nonlinear profiles using a wavelet-based distribution-free CUSUM chart. *International Journal of Production Research*, **50**(22), 6574–6594.

Lee, J., Xie, Y. and Cheng, X. (2023) Training neural networks for sequential change-point detection, in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE Press, Piscataway, NJ, pp. 1–5.

Liu, D. (2022) Distribution-free statistical process control and Bayesian feasibility determination. Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA.

Li, D., Chen, D., Goh, J. and Ng S.-K. (2019) Anomaly detection with generative adversarial networks for multivariate time series. *arXiv preprint arXiv:1809.04758*.

Liu, D., Kim, H., Kim, S.-H., Kim, T., Lee, D. and Xie, Y. (2022) Distribution-free multivariate time-series monitoring with analytically determined control limits. *International Journal of Production Research*, **61**, 1–18.

Lu, C.-J. and Tsai, D.-M. (2005) Automatic defect inspection for LCDs using singular value decomposition. *The International Journal of Advanced Manufacturing Technology*, **25**(1), 53–61.

Mastrangelo, C.M., Runger, G.C. and Montgomery, D.C. (1996) Statistical process monitoring with principal components. *Quality and Reliability Engineering International*, **12**(3), 203–210.

Megahed, F.M., Wells, L.J., Camelio, J.A. and Woodall, W.H. (2012) A spatiotemporal method for the monitoring of image data. *Quality and Reliability Engineering International*, **28**(8), 967–980.

Megahed, F.M., Woodall, W.H. and Camelio, J.A. (2011) A review and perspective on control charting with image data. *Journal of Quality Technology*, **43**(2), 83–98.

Niaki, S.T.A., Khedmati, M. and Soleymanian, M.E. (2015) Statistical monitoring of autocorrelated simple linear profiles based on principal components analysis. *Communications in Statistics – Theory and Methods*, **44**(21), 4454–4475.

Noorossana, R., Vaghefi, A. and Dorri, M. (2008) The effect of non-normality on performance of linear profile monitoring, in *2008 IEEE International Conference on Industrial Engineering and Engineering Management*, IEEE Press, Piscataway, NJ, pp. 262–266.

Wang, D., Li, F. and Liu, K. (2021) Modeling and monitoring of a multivariate spatio-temporal network system. *IISE Transactions*, **55**, 1–17.

Wang, H., Kim, S.-H., Huo, X., Hur, Y. and Wilson, J.R. (2015) Monitoring nonlinear profiles adaptively with a wavelet-based distribution-free CUSUM chart. *International Journal of Production Research*, **53**(15), 4648–4667.

Wang, K. and Tsung, F. (2005) Using profile monitoring techniques for a data-rich environment with huge sample size. *Quality and Reliability Engineering International*, **21**(7), 677–688.

Wang, Y.-H.T. and Lai, Y. (2019) Monitoring of autocorrelated general linear profiles. *Journal of Statistical Computation and Simulation*, **89**(3), 519–535.

Whitt, W. (2011) *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer New York.

Woodall, W.H. (2007). Current research on profile monitoring. *Production*, **17**, 420–425.

Woodall, W.H., Spitzner, D.J., Montgomery, D.C. and Gupta, S. (2004) Using control charts to monitor process and product quality profiles. *Journal of Quality Technology*, **36**(3), 309–320.

Wu, J., Zhao, Z., Sun, C., Yan, R. and Chen, X. (2020) Fault-attention generative probabilistic adversarial autoencoder for machine anomaly detection. *IEEE Transactions on Industrial Informatics*, **16**(12), 7479–7488.

Yan, H., Grasso, M., Paynabar, K. and Colosimo, B.M. (2022) Real-time detection of clustered events in video-imaging data with applications to additive manufacturing. *IISE Transactions*, **54**(5), 464–480.

Yan, H., Paynabar, K. and Shi, J. (2014) Image-based process monitoring using low-rank tensor decomposition. *IEEE Transactions on Automation Science and Engineering*, **12**(1), 216–227.

Yan, H., Paynabar, K. and Shi, J. (2017) Anomaly detection in images with smooth background via smooth-sparse decomposition. *Technometrics*, **59**(1), 102–114.

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

Yin, Y.-Q., Bai, Z.-D. and Krishnaiah, P.R. (1988) On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields*, **78**(4), 509–521.

Zhang, C., Chen, N. and Wu, J. (2020) Spatial rank-based high-dimensional monitoring through random projection. *Journal of Quality Technology*, **52**(2), 111–127.

Zhang, C., Yan, H., Lee, S. and Shi, J. (2018) Weakly correlated profile monitoring based on sparse multi-channel functional principal component analysis. *IISE Transactions*, **50**(10), 878–891.

Zhu, J. and Lin, D.K. (2009) Monitoring the slopes of linear profiles. *Quality Engineering*, **22**(1), 1–12.