Taylor & Francis
Taylor & Francis Group

Quality Technology & Quantitative Management

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ttqm20

Comprehensive review of high-dimensional monitoring methods: trends, insights, and interconnections

Fuhad Ahmed, Tahir Mahmood, Muhammad Riaz & Nasir Abbas

To cite this article: Fuhad Ahmed, Tahir Mahmood, Muhammad Riaz & Nasir Abbas (26 Aug 2024): Comprehensive review of high-dimensional monitoring methods: trends, insights, and interconnections, Quality Technology & Quantitative Management, DOI: 10.1080/16843703.2024.2395745

To link to this article: https://doi.org/10.1080/16843703.2024.2395745

Published online: 26 Aug 2024.

QUALITY TECHNOLOGY & QUANTITATIVE MANAGEMENT
https://doi.org/10.1080/16843703.2024.2395745

Comprehensive review of high-dimensional monitoring methods: trends, insights, and interconnections

Fuhad Ahmed[a], Tahir Mahmood[b], Muhammad Riaz[a] and Nasir Abbas[a]

[a]Department of Mathematics, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia; [b]School of Computing, Engineering and Physical Sciences, University of the West of Scotland, Paisley, UK

ABSTRACT

High-dimensional data refers to a dataset that contains many variables or features, typically with many more features (p) than observations (n) (i.e. $n < p$). With technological advancements in sensors, high-dimensional data are becoming increasingly common in process-monitoring applications. Therefore, this study presents a comprehensive overview of high-dimensional monitoring methods, in which 82 articles published from 2004 to 2023 were found to be relevant. The literature on high-dimensional monitoring can be divided into three approaches: control charts based on dimension reduction, variable selection, and high-dimensional techniques. Furthermore, the literature on each approach is divided in terms of control chart structures such as memory-less (Hotelling's T^2), memory type (multivariate exponentially weighted moving average (MEWMA) and multivariate cumulative sum (MCUSUM)), and others. Real-life datasets from different fields, such as industry, medical science, chemical engineering, and image processing, which have frequently been used in high-dimensional monitoring, are also listed. It is noted that the literature on high-dimensional monitoring increased after 2016, and most studies were designed using high-dimensional techniques. Moreover, most studies proposed memory types and other structures for monitoring high-dimensional data. This review article offers a comprehensive summary of the current state-of-the-art high-dimensional monitoring research and identifies potential areas for future research.

ARTICLE HISTORY
Received 19 October 2023
Accepted 19 August 2024

KEYWORDS
Dimension reduction techniques; High-dimensional data; multivariate control charts; statistical process control; variable selection methods

1. Introduction

In recent years, collecting and analyzing high-dimensional data has become integral to many fields, including bioinformatics, finance, physics, social science, and machine learning. High-dimensional data refers to a dataset that contains many variables or features, typically with many more features (p) than observations (n) (i.e. $n < p$). The simultaneous collection of thousands to millions of features for each object or person is a key element of modern data. These data are referred to as high-dimensional (HD) data. For instance, in high-throughput DNA sequencing, the number of features recorded for each patient can range from tens to thousands, while the number of patients is modest. Similarly, the dataset may include many features but a few labelled examples in speech or image recognition. An image can be considered high-dimensional data, with each pixel representing a separate feature. High-dimensional data can be found in almost every field, including industry (Kim et al., 2019; Li et al., 2014), medical science (Fan et al., 2021; Kim et al., 2020; Maboudou-Tchao et al., 2023), chemical engineering (Sun et al., 2013), and image processing (He et al., 2018).

CONTACT Tahir Mahmood tahir.mahmood@uws.ac.uk School of Computing, Engineering and Physical Sciences, University of the West of Scotland, Paisley PA1 2BE, UK

In the current era, analyzing and interpreting high-dimensional data is challenging because it can be difficult to visualize and comprehend data with many dimensions. This can also lead to overfitting and spurious correlations if not adequately handled. Thus, high-dimensional data analysis poses unique challenges that traditional statistical methods cannot capture. The primary challenge associated with high-dimensional data relates to the well-known 'curse of dimensionality' phenomenon. Moreover, increasing the number of features increases the risk of overfitting, leading to poor predictive accuracy of the model. Traditional statistical methods can only handle a small number of features and have become inefficient when handling high-dimensional data. Therefore, the need for methods that can handle high-dimensional data has led to a new area of research known as 'high-dimensional data analysis'. High-dimensional data analysis aims to develop statistical methods to extract meaningful patterns or structures from high-dimensional data while overcoming the challenges posed by the 'curse of dimensionality'. These methods must be computationally efficient, statistically robust, and scalable. Recent trends have seen significant advancements in high-dimensional monitoring methods. Theoretical developments have enhanced our understanding of how to address the challenges posed by high-dimensional data. For example, new techniques such as memory-based control charts and kernel-based methods have been introduced to better handle complex data structures and interactions. These advancements have led to improved practical applications across various domains, including quality control and process monitoring. Despite these advancements, limitations remain; some methods, while offering improved detection of subtle changes, may require extensive computational resources and may not be suitable for all types of data. Understanding the strengths and weaknesses of these techniques is crucial for their effective implementation.

High-Dimensional data analysis is often divided into two main groups: high-dimensional approaches and analysis based on dimension reduction methods. Figure 1 displays the high-dimensional analytics taxonomy. Dimension reduction techniques and variable selection strategies are two more categories under which the dimension reduction methods are subdivided. With the goal of retaining as much crucial information as possible, dimension reduction strategies reduce a dataset's characteristics, dimensions, or variables to a lower-dimensional space. Nonetheless, a subset of variables (sometimes referred to as features or predictors) is often chosen from a large number of variables in a dataset using variable selection techniques. High-Dimensional data may be reduced to a lower-dimensional space using both dimension reduction and variable selection strategies. However, these methods can provide inadequate data when the variables have a high degree of correlation (multicollinearity). Furthermore, both methods are computationally intensive and time-consuming. These drawbacks have pushed researchers to use so-called high-dimensional methods traits that eschew variable selection or dimension reduction.

In the era of Industry 4.0, the revolution in sensor technology has provided remarkable features for storing high-dimensional data in temporal frames. Therefore, most industries are interested in monitoring high-dimensional data to improve the quality of their production and service sectors. Over the last few decades, Statistical Process Control (SPC) methodologies have been used to monitor and control production processes to ensure efficient, effective, and consistent operations. This methodology involves collecting and analyzing data in real-time to identify and respond to variations in the production process. Control charts are one of seven magnificent SPC tools that aim to detect and prevent abnormalities or deviations from the expected process performance. However, with the advent of high-dimensional data, traditional control charts may not be sufficient because they fail to detect complex interactions between variables, leading to poor process monitoring. Therefore, many studies have proposed control charts based on high-dimensional data over the last few decades. Hence, this research was designed as follows:

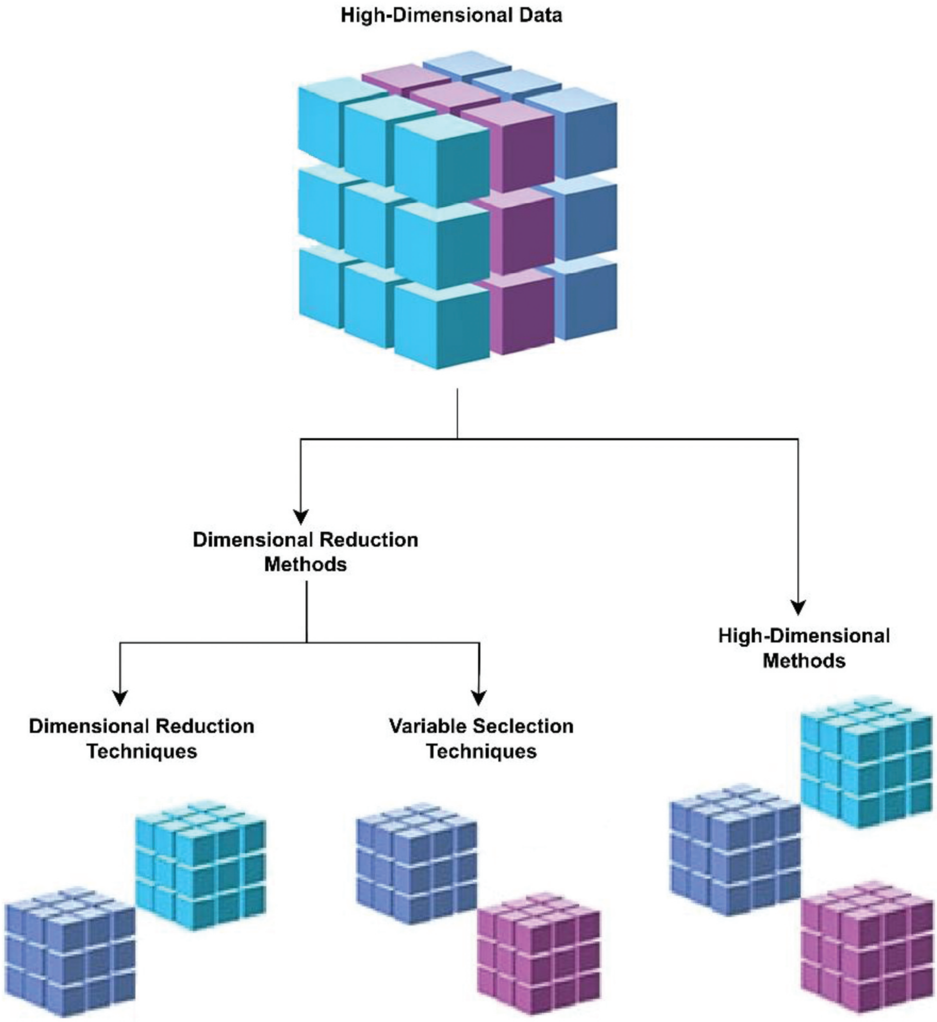

Figure 1. Taxonomy of high-dimensional analytics.

(i) to review the existing studies related to control charts for high-dimensional data,
(ii) to analyze the high-dimensional control chart studies,
(iii) to list the high-dimensional datasets used in monitoring studies,
(iv) to classify the existing literature on high-dimensional control charts,
(v) to highlight the research gaps in high-dimensional monitoring studies.

The rest of the article is organized as follows: Section 2 consists of an analysis of existing control chart studies to monitor high-dimensional data while also listing the high-dimensional datasets used in the monitoring studies; Section 3 provides a review of basic multivariate control charting setups; Section 4 discusses the classes of high-dimensional monitoring studies; and, Section 5 presents the concluding remarks and future directions for high-dimensional monitoring studies.

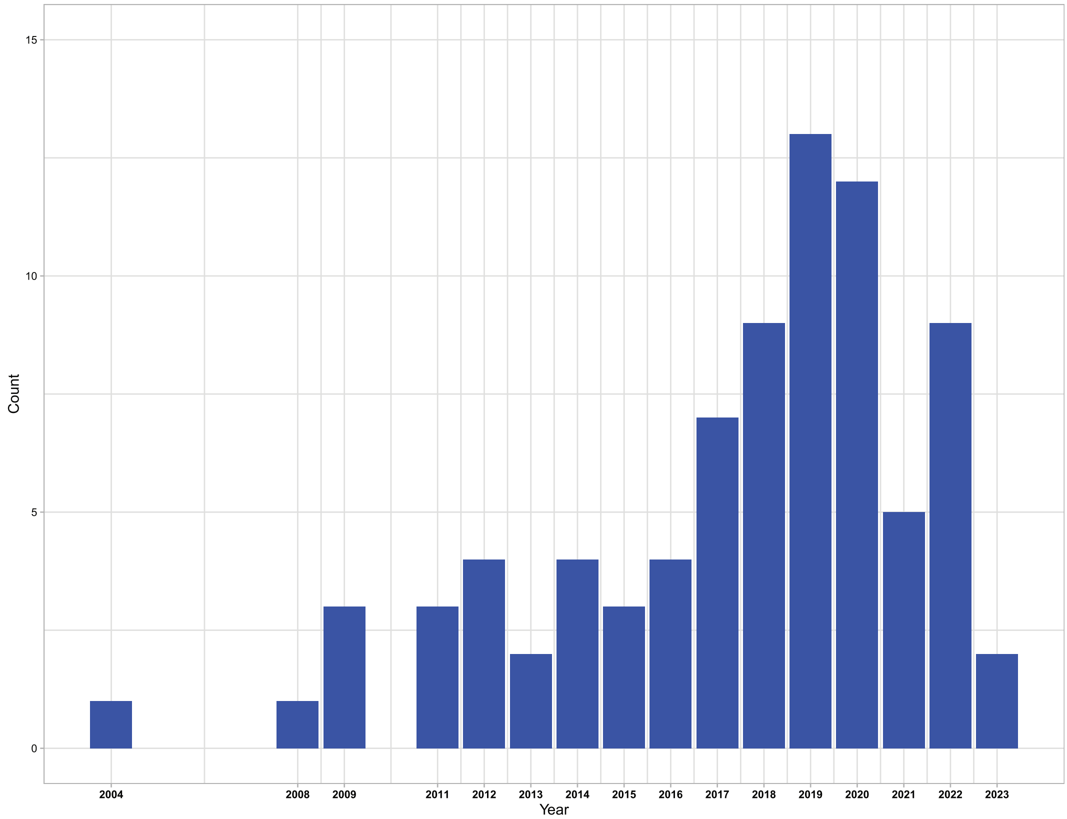

Figure 2. Number of high-dimensional control chart studies over the years.

2. Data analysis of high-dimensional control chart studies

In this study, a total of 82 articles between 2004 and 2023 were reviewed. As illustrated in Figure 2, a few studies were initially conducted between 2004 and 2016, and 25 articles were published on high-dimensional control charts. After 2016, there was an immense increase in literature, with 57 studies being published on high-dimensional monitoring schemes. This trend demonstrates the importance of designing review studies for the surveillance of high-dimensional data.

From this point on, the literature on high-dimensional control charts is separated into three categories: dimension reduction, variable selection, and high-dimension techniques. Additionally, there are three types of charting structures: memory-less (like Hotelling's) and memory-type (like Multivariate Exponentially Weighted Moving Average (MEWMA) or Multivariate Cumulative Sum (MCUSUM) charts and other structures). As shown in Figure 3, among the 82 studies on high-dimensional process monitoring, 23 studies used dimension-reduction techniques to reduce the data dimensions before implementing traditional multivariate control charts. Of these, 12 studies applied HT control chart, four studies used MEWMA or MCUSUM charts, and the rest used other charts. Ten studies utilized the variable selection approach to reduce the data before implementing traditional multivariate control charts. Among them, only one article proposed HT scheme, five suggested MEWMA or MCUSUM control charts, and four applied charts that were different from HT, MEWMA, and MCUSUM. A large number of studies (49 articles) considered neither dimension reduction nor variable selection methods but were purely based on high-dimensional data techniques. Among these studies, only three articles applied memory-less monitoring structures; 20 were based on memory type, and 26 were based on other structures.

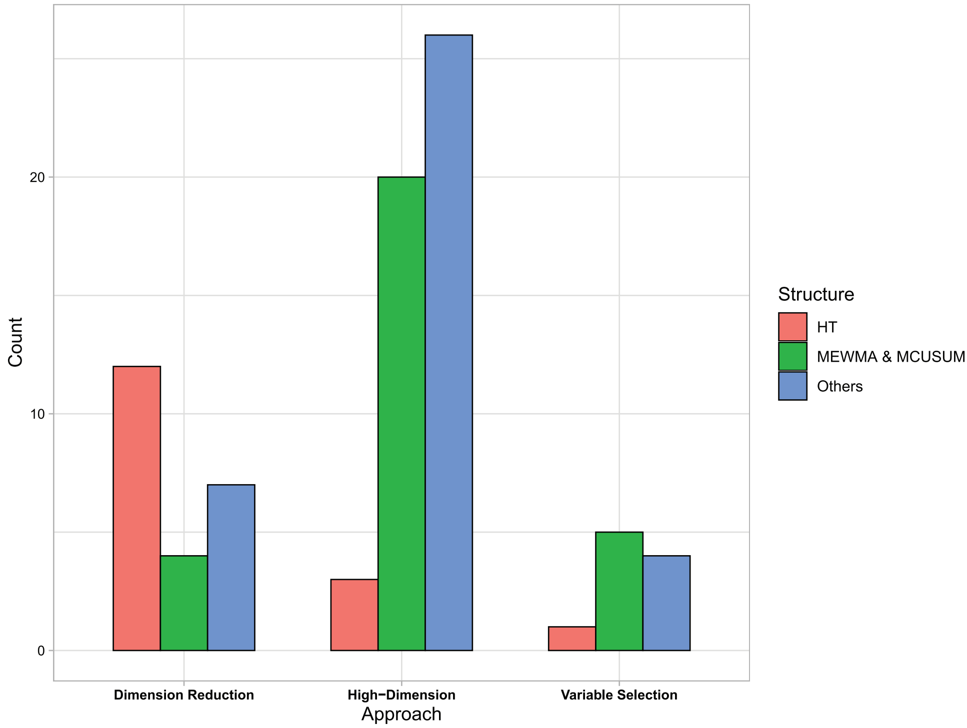

Figure 3. Count plot for high-dimensional control chart studies with respect to approaches and structures.

Figure 4 shows the co-occurrences of keywords used in studies focusing on high-dimensional process monitoring. It is evident that the keywords 'dimensionality reduction' and 'pca' occurred together frequently, whereas 'cusum', 'robustness', and 'performance' are the three keywords that appeared together in the high-dimensional monitoring studies. Similarly, other co-occurring keywords were 'control charts', 'statistical process control', 'multivariate', 'regression', 'profiles', and 'mean vector'.

In the SPC literature, various high-dimensional data types are utilized, including those from the industrial, chemical, imaging, solar, and medical sciences (cf. Table 1). We found that the utilization of industrial data was most prevalent in this review. A total of 21 datasets based on the industry sector were found among the numerous high-dimensional dataset types (cf. Table 1). Meanwhile, the numbers of high-dimensional datasets used in the chemical and medical areas are 7 and 5, respectively.

3. General structure of multivariate control charts

Assume $X_1, X_2, \ldots, X_p$ are random variables that follow a multivariate normal distribution with mean vector $\boldsymbol{\mu}_0$ and covariance matrix Σ_0. Let the i^{th} observation vector of the process be denoted by X_{ij}, where $i = 1, 2, \ldots, n; j = 1, 2, \ldots, p$. The Hotelling's T^2 statistic is defined as:

$$T^2 = \left(X_{ij} - \bar{X}_j\right)^T S^{-1} \left(X_{ij} - \bar{X}_j\right),$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$, and $S = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)\left(x_{ij} - \overline{x_j}\right)^T$..

The lower control limit of Hotelling's T^2 chart is set to be zero while the upper control limit under phase I is obtained by $UCL = \frac{(n-1)^2}{n}\beta_{\alpha,p/2,(n-p-1)/2}$, where $\beta_{\alpha,p/2,(n-p-1)/2}$ is upper α^{th} percentile of beta distribution with parameters $p/2$ and $(n-p-1)/2$ (Bersimis et al., 2007; Tracy

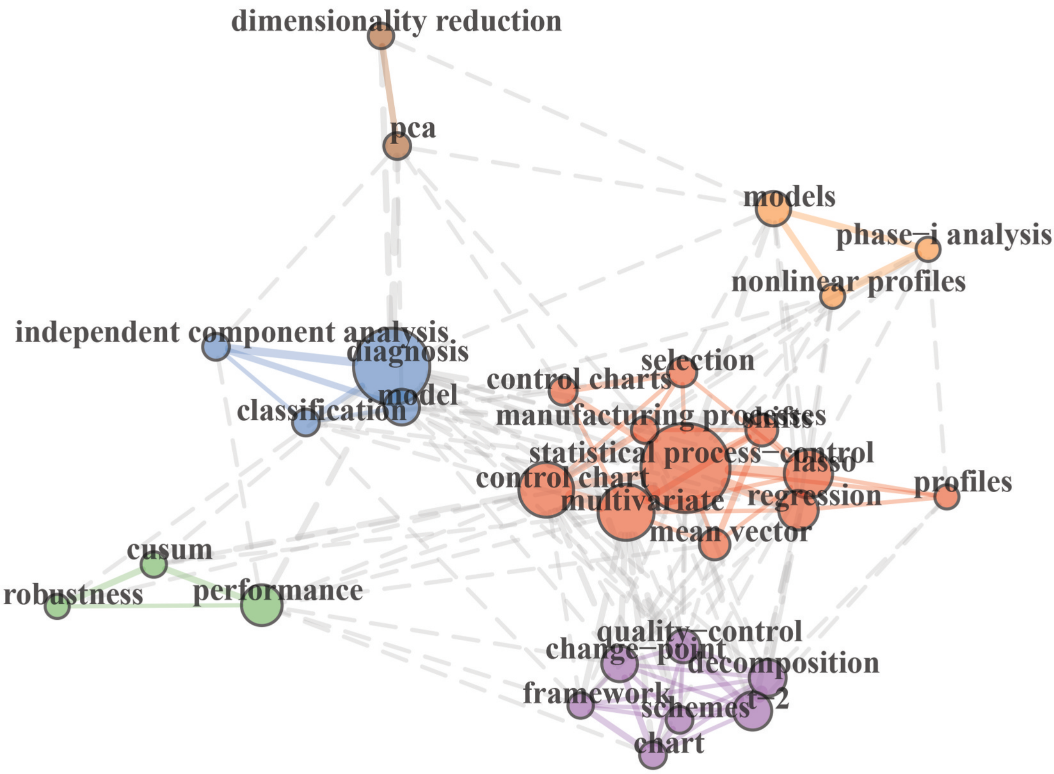

Figure 4. Keyword co-occurrences of high-dimensional control chart studies.

et al., 1992). Similarly, the upper control limit for phase II is obtained by $UCL = \frac{p(n+1)(n-1)}{n^2-np} F_{\alpha,p,(n-p)}$, where $F_{\alpha,p,(n-p)}$ is the upper α^{th} percentile of the F distribution with parameters p and $(n-p)$.

Note that Hotelling's T^2 chart is effective only for detecting large shifts in the parameter of interest (Mahmood et al., 2019). However, the MEWMA chart is preferred for detecting small to moderate shifts in the parameters of interest (Mahmood & Erem, 2023). Lowry et al. (1992) introduced the MEWMA chart, where the MEWMA statistic is defined as:

$$z_{ij} = R\left(X_{ij} - \boldsymbol{\mu}_0\right) + (1-R)z_{i;j}$$

Where $\boldsymbol{R} = diag\left(r_1, r_2, \ldots, r_p\right)$ is user-chosen smoothing parameters $0 \leq r_j \leq 1$ for $j = 1, 2, \ldots, p$, $z_{0j} = 0$, and $\mathbf{I}$ is the identity matrix. If the charted statistic $E_i^2 = z^T \sum_z^{-1} z > UCL$, the MEWMA chart gives an out-of-control signal. Here, Σ_z is the variance-covariance matrix of z_{ij}, and UCL is the upper control limit of the MEWMA chart.

The MCUSUM chart is an alternative for monitoring small to moderate shifts in the parameter of interest (Erem & Mahmood, 2023). Crosier (1988) presented two procedures by extending the univariate CUSUM chart to a multivariate case and named it the MCUSUM control chart. The MCUSUM statistic can be expressed as:

$$C_i = \left[\left(S_{i-1} + \bar{x}_j - u_0\right)^T \left(\frac{\Sigma}{n}\right)^{-1} \left(S_{i-1} + \bar{x}_j - u_0\right)\right]^{1/2},$$

where the covariance matrix is Σ, and the cumulative sums S_i are defined as

Table 1. High-dimensional datasets used in the selected studies.

Sector	Dataset	References	Dataset	References
Industrial Datasets	Industrial penicillin production plant	Hu and Yuan (2008) Hu and Yuan (2009) R. C. Wang et al. (2017)	PCB packaging process	Z. Wang et al. (2017)
	Spur gear production	Abdella et al. (2020)	High-speed milling process	S. Kim et al. (2020)
	A hexagonal blot manufacturing	Abdella et al. (2017)	Chocolate conching batch process	Peres et al. (2019)
	Semiconductor manufacturing process	Zou et al. (2014) Zou et al. (2015), Shu and Fan (2018), Abdella et al. (2019), J. Li (2019), W. Li et al. (2020) C. Zhang et al. (2020), Mukherjee and Marozzi (2022) Xiang et al. (2022) Ye et al. (2023)	Tennessee Eastman process (TEP)	J. Yu (2012a) Samuel and Cao (2014) J. Yu and Zhang (2020) S. Kim and Turkoz (2022) Ye et al. (2023) R. C. Wang et al. (2017)
	Paper product manufacturing process	Alaeddini et al. (2018)	White wine production process	Zou et al. (2011), Z. Li and Tsung (2019), Y. Li et al. (2020) Merlo et al. (2022)
	Bolt monitoring	J. Kim et al. (2016) Turkoz et al. (2019)	Forging process	Y. Wang et al. (2018)
	Aluminium smelter company	Zhao et al. (2021)	Bearing	J.-B. Yu (2011)
	Aircraft maintenance	Qi et al. (2016)	Surface contact quality of the radio frequency MEMS switch	Huang et al. (2014)
	Steel rolling process, and a stamping process	H. Yan et al. (2018)	Soil temperature	Gómez et al. (2022)
	Continuous stirred tank heater	S. Kim et al. (2022)	Hard disk drives monitoring system	Du and Zou (2018)
	Logistic company	Zou, Ning, et al. (2012)		
Medical Science Data	Breast Cancer	J.-M. Kim et al. (2020) Fan et al. (2021) E. Maboudou-Tchao et al. (2023)	Biomanufacturing Process & Ultra-precision Machining	Kan and Yang (2015) Kan and Yang (2017)
	Lung cancer	E. M. Maboudou-Tchao (2021)	Metabolomics	Ullah et al. (2017)
	E. coli	D. Yan et al. (2019)		
Chemical Data	LNG process	Sangahn (2019)		
	Wastewater treatment process	J. Yu (2012b) Odom et al. (2018)	Heavy oil-fired combustion test facility	Sun et al. (2013)
	Water quality monitoring	J. Chen et al. (2019)	Water vapor flux divergence (WVFD) dataset	Fang et al. (2022)
	Pressure swing adsorption (PSA)	R. C. Wang et al. (2017)	Thin-film-transistor liquid-crystal display process	S. Lee et al. (2019)
Image Data	Cherenkov gamma-ray telescope	Jang et al. (2017)	Camera performance measurement	He et al. (2018)
Solar Data	Solar data	Gómez et al. (2022)	Solar flare detection dataset	J. Guo et al. (2022)
	Solar data observatory	H. Yan et al. (2018)		

(*Continued*)

Table 1. (Continued).

Sector	Dataset	References	Dataset	References
Miscellaneous	Monthly rates of vehicle robberies	Moala et al. (2019)	Traffic monitoring	Feng et al. (2020)
	National Football League (NFL)	Martinez et al. (2020)	Sonar	E. M. Maboudou-Tchao (2021)
	KDDCUP99/NSL_KDD/UNSW-NB15/ISCX-URL2016/TON_IoT	Shaohui et al. (2022)	Vertical density profiles (VDP)	K. Wang and Jiang (2009), H. Zhang and Albin (2009) S. Chen and Nembhard (2011)

$$S_i = \begin{cases} 0 & ifC_i \leq k \\ \left(S_{i-1} + \bar{X}_j - \boldsymbol{\mu}_0\right)\left(1 - \frac{k}{C_i}\right) & ifC_i > k \end{cases},$$

If the charting statistic $T_i^2 = \left[S^T\left(\frac{\Sigma}{n}\right)^{-1}S\right] > h$, the MCUSUM chart gives an out-of-control signal. It is noteworthy that h is the upper control limit.

4. High-dimensional control chart studies

In this review, we found that the literature on high-dimensional monitoring methods can be classified into three approaches:

(i) Dimension-reduction-based control charts: First, the dimension-reduction technique is used to reduce the dimensions of the data, and then traditional multivariate control charts are implemented.
(ii) Variable-selection-based control charts: First, the variable selection method is used to reduce the data, and then traditional multivariate control charts are implemented.
(iii) High-dimensional technique based control charts adopt high-dimensional control charts without dimension reduction or variable selection.

Furthermore, the literature on each approach is categorized with respect to different structures, such as memory-less (Hotelling's T^2), memory type (i.e. MEWMA and MCUSUM), and other control charts. The divisions of all studies with respect to approaches and structures are listed in Table 2. The details of each approach for the structures are provided in the preceding subsections.

4.1. High-dimensional control charts using dimension reduction techniques

Dimension reduction is the process of reducing a dataset's features, dimensions, or variables to a lower-dimensional space while preserving as much essential ;information ;as possible. Dimension reduction aims to reduce the complexity of data for better understanding, visualization, and analysis while avoiding issues that arise with high-dimensional data, such as the dimensionality curse. This review classifies dimension reduction methods into component analysis, projection-based, and other. Currently, component analysis is widely used for dimension reduction in SPCs. Commonly used component analysis methods are Principal Component Analysis (PCA), Local and Global PCA (LGPCA), Multiway PCA (MPCA), Kernel PCA (KPCA), Multistate Adaptive-Dynamic PCA (MSAD-PCA), Adaptive-Dynamic PCA (AD-PCA), Functional PCA (FPCA), Multichannel Functional PCA (MFPCA), Uncorrelated Multi-linear PCA (UMPCA), Multi-linear PCA (MPCA), Vectorized PCA (VPCA), and Sparse Multi-channel Functional PCA (SMFPCA).

Table 2. Summary of the high-dimensional control chart literature.

Technique	Control Charts: Hotelling's T^2	MEWMA & MCUSUM	Others
Dimension Reduction	J. Yu (2012a), J.-M. Lee et al. (2004), Hu and Yuan (2008), Skubalska-Rafajłowicz (2013), Sun et al. (2013), Samuel and Cao (2014), Fujiwara and Kano (2017), Odom et al. (2018), S. Lee et al. (2019), J. Yu and Zhang (2020), Shaohui et al. (2022), J. Yu and Liu (2022)	J.-B. Yu (2011), C. Zhang et al. (2020), J. Chen et al. (2019) Gómez et al. (2022)	Ning and Tsung (2012), Bae et al. (2016), Kan and Yang (2017), C. Zhang et al. (2018a), C. Zhang et al. (2018b), Y. Wang et al. (2018), J.-M. Kim et al. (2020)
Variable Selection	J. Kim et al. (2016)	Zou et al. (2011), Zou, Ning, et al. (2012), Abdella et al. (2017), Sangahn (2019), Zhao et al. (2021)	K. Wang and Jiang (2009) Peres et al. (2019), D. Yan et al. (2019), Qin et al. (2020), J. Guo et al. (2022)
High-Dimension	Ullah et al. (2017), M. R. Ahmad and Ahmed (2021), Ahmadi-Javid and Ebadi (2021)	Huang et al. (2014), Liu et al. (2015), Zou et al. (2015), Zou et al. (2015) Qi et al. (2016), Z. Wang et al. (2017), Alaeddini et al. (2018), Du and Zou (2018), J. Li (2019), Z. Li and Tsung (2019), Moala et al. (2019), S. Ahmad et al. (2020), Feng et al. (2020), S. Kim et al. (2020), J. Li (2020), Y. Li et al. (2020), Fang et al. (2022), S. Kim et al. (2022), E. Maboudou-Tchao et al. (2023), Ye et al. (2023)	Hu and Yuan (2009), H. Zhang and Albin (2009), S. Chen and Nembhard (2011), J. Yu (2012b), Y. Li et al. (2014), Zou et al. (2014), Kan and Yang (2015), Rato et al. (2016), Turkoz et al. (2016), Jang et al. (2017), R. C. Wang et al. (2017), He et al. (2018), Shu and Fan (2018), H. Yan et al. (2018), Abdella et al. (2019), X. Guo et al. (2019), J. Kim et al. (2019) Turkoz et al. (2019) Abdella et al. (2020), W. Li et al. (2020), Martinez et al. (2020), Fan et al. (2021), E. M. Maboudou-Tchao (2021), Merlo et al. (2022), Mukherjee and Marozzi (2022), S. Kim and Turkoz (2022), Xiang et al. (2022)

Some projection-based methods used in SPC include Locality Preserving Projections (LPP), Multiway Locality Preserving Projections (MLPP), Normal Random Projections (NRP), Ensemble Random Projections (ERP), and Tensor Locality Preserving Projections (TLPP). Other methods used for reducing dimensionality in SPC include Joint Decorrelation (JD) and Stochastic Proximity Embedding (JD-SPE), One-Dimensional Residual Convolutional Auto-Encoder (1DRCAE), Recursive Tensor Recovery (RTR), Spatial Scanning (SS), Local Outlier Factor (LOF), and Simplicial Depth Statistic (SDS). The details on the SPC methods using the above-stated dimension reduction methods are presented in the following subsections.

4.1.1. The Hotelling's T^2 charts

In recent decades, several studies have utilized dimension reduction techniques to handle high-dimensional process data and employed Hotelling's T^2 chart for process monitoring. J.-

M. Lee et al. (2004) used KPCA to capture nonlinear relationships in variables, proposing Hotelling's T^2 and SPE charts to monitor process variations, demonstrated in a biological wastewater treatment process. Hu and Yuan (2008) applied MLPP for dimension reduction and also proposed Hotelling's T^2 and SPE charts, showing in a penicillin fed-batch case study that MLPP better preserved data structure and captured intrinsic information compared to MPCA. J. Yu (2012a) proposed a technique for projecting observational data into a low-dimensional space using LGPCA, incorporating Hotelling's T^2 and SPE charts. This exceeded PCA and LPP charts in terms of data separation, error detection rate, and sensitivity. Skubalska-Rafajłowicz (2013) proposed a multivariate Hotelling's T^2 chart with random projections for high-dimensional process monitoring. Sun et al. (2013) employed KPCA for dimension reduction and Hotelling's T^2 and Q statistics to find abnormalities in combustion processes. Samuel and Cao (2014) adapted KPCA to kernel density estimation (KDE) for process monitoring, revealing that KPCA with KDE produced greater fault detection rates than KPCA and PCA techniques alone, despite slightly higher false alarm rates that remained below acceptable bounds. Overall, combining KPCA with KDE enhanced nonlinear process monitoring performance.

Fujiwara and Kano (2017) introduced the JD-SPE-T^2 chart, which combines JD and SPE methodologies to enhance data visualization and monitoring. JD is a blind source separation approach, while SPE converts high-dimensional data into two dimensions. This new approach outperforms the typical Hotelling's T^2 chart using MSPC. Odom et al. (2018) employed MSAD-PCA to reduce dimensions of complicated multistate process data, comparing SPE and Hotelling's T^2charts to AD-PCA-based charts. Their research, which included a decentralized wastewater treatment plant instance, demonstrated MSAD-PCA's superiority in monitoring nonlinear, non-stationary, and serially autocorrelated processes. S. Lee et al. (2019) suggested a Variational Auto-Encoder (VAE) based control chart that combines Hotelling's T^2 and SPE statistics for nonlinear and non-normal high-dimensional systems. This technique outperforms previously used latent variable-based charts in both simulated and real-life TFT-LCD datasets. J. Yu and Zhang (2020) presented Manifold Regularized Stacked Autoencoders (MRSAE) for dimension reduction, employing Hotelling's T^2 and SPE charts. They demonstrated improved fault detection rates in TEP and fed-batch fermentation penicillin (FBFP) processes compared to other DNNs, stacked denoising autoencoders (SDAE), PCA, KPCA, and LGPCA. Shaohui et al. (2022) introduced a PCA mix algorithm with bootstrap control limits for Hotelling's T^2 charts, which outperformed traditional approaches in detecting network anomalies. J. Yu and Liu (2022) used a deep neural network model called the 1DRCAE model for dimension reduction, incorporating SPE and Hotelling's T^2 charts, which outperformed other approaches in simulation studies as well as in real-life implementations using TE and FBFP datasets.

4.1.2. The MEWMA charts

In the SPC literature, numerous studies have used MEWMA charts for high-dimensional process monitoring following dimension reduction. J.-B. Yu (2011) employed LPP for dimension reduction and established an EWMA statistic combining Hotelling's T^2 and SPE to quantify bearing performance deterioration. This LPP-EWMA approach accurately identified early bearing deterioration and monitored its performance over time. C. Zhang et al. (2020) employed random ensemble projections to generate subspaces and devised a local monitoring technique with a spatial rank test, resulting in an RPSR chart. This RPSR chart beat MCUSUM (Mei, 2010), AnRank (Qiu & Hawkins, 2003), and the Kernel Hilbert chart named RKHS (Huang et al., 2014) charts for identifying correlations, sparse changes, and managing non-normal distributions, as shown by case studies on handwritten number identification and semiconductor production processes. Further, Gómez et al. (2022) employed a recursive tensor recovery (RTR) model to transform high-dimensional streaming data into a low-dimensional subspace, which was then integrated with an EWMA control chart. To enhance sampling, they created the tensor sequential sampling (TSS) approach. The suggested RTR-EWMA chart outperformed the TRAS (Liu et al., 2015) and spatial-adaptive sampling method (SASAM) (Wang et al., 2018) charts in collecting both geographical and temporal information. This strategy was tested using solar and soil temperature data case studies.

4.1.3. The MCUSUM charts

Chen et al. (2019) introduced reduced-dimension (RD) control charts that use spatial scanning to minimize dimension during data collecting while monitoring high-dimensional process means. The authors created MCUSUM charts using LR and Hotelling's T^2 statistics, comparing full-dimension (F-MCUSUM) and reduced-dimension (RD-MCUSUM) charts with Average Run Length (ARL). The RD-MCUSUM charts were successful and simple to use, particularly with high-dimensional, ill-conditioned covariance matrices. A water quality monitoring case study confirmed the actual efficiency of RD-MCUSUM charts.

4.1.4. Other charts

In addition to established techniques like Hotelling's T^2, MEWMA, or MCUSUM-based control charts, new measure-based approaches have evolved for monitoring high-dimensional data. Ning and Tsung (2012) presented a Local Outlier Factor (LOF) technique for mixed-type data, which demonstrated advantages over simple depth statistics, particularly in clustered datasets. Bae et al. (2016) developed a nonparametric monitoring technique based on data depth, with measurements such as Tukey depth being particularly useful for SPC purposes, especially when data deviates from normality. Kan and Yang (2017) developed a network-generalized likelihood ratio (NGLR) chart for effective process change detection that was validated in ultraprecision machining and biomanufacturing. When compared to Hotelling's T^2, NGLR successfully characterized complicated alterations in time-varying picture data structures. Y. Wang et al. (2018) suggested a unique-threshold multivariate PCA for multichannel profile monitoring, which employs functional PCA followed by soft-thresholding for feature selection. Soft thresholding in PCA improved performance significantly when tested on a forging production dataset.

Zhang et al. (2018a) developed monitoring techniques for multiple sensor profiles utilizing FPCA and MFPCA, which outperformed UMPCA, multilinear PCA (MPCA) and VPCA in ARL and SDRL. A manufacturing process case study had similar outcomes. C. Zhang et al. (2018b) developed SMFPCA for modeling multi-channel profiles, which demonstrated flexibility in defining correlation structures while producing sparse and interpretable PCA scores. When modeling weak inter-profile correlations, SMFPCA outperformed MFPCA, UMPCA, MPCA, and VPCA-based charts. Furthermore, the simulation findings were validated by implementing charts on the manufacturing system data. Moreover, J.-M. Kim et al. (2020) presented regression-based r-control charts for binary asymmetrical data with multicollinearity, which combine PCA and FPCA with binary response regression models. A comparison of efficiency between simulated and actual breast cancer data revealed that the neural network model outperformed PCA-based and FPCA-based binary response regression models using GLM and neural network techniques, demonstrating its usefulness in monitoring both simulated and real binary response data.

4.2. High-dimensional control charts using variable selection techniques

In literature, variable selection methods are often used in high-dimensional process monitoring to select a subset of variables (also known as features or predictors) from many variables in a dataset. Choosing only relevant variables for monitoring in a high-dimensional process is crucial because the number of variables often far exceeds the number of observations. Variable selection aims to reduce the dimensionality of the dataset and increase fault detection accuracy. Various variable selection techniques, such as the step-down procedure, LASSO-based procedure, conditional sum of square-based procedure, adaptive variable selection, adaptive LASSO, forward-selection technique, penalized likelihood, and integrated multiway principal component analysis, are currently used for high-dimensional process monitoring.

4.2.1. The Hotelling's T^2 charts

Historically, few studies have utilized variable selection strategies to reduce high-dimensional process data in conjunction with Hotelling's T^2 chart for parameter monitoring. J. Kim et al. (2016) used the adaptive step-down (ASD) approach and Hotelling's T^2 chart to detect mean vector changes. In comparison to MYT procedures (Mason et al., 1995), step-down procedures (Sullivan et al., 2007), and LASSO-based procedures (Zou et al., 2011), the suggested technique demonstrated enhanced diagnostic capabilities, especially in detecting mean-shifted variables when shifts occurred in a limited number of variables. Simulation findings showed that the suggested technique outperformed MYT and step-down processes for minor shift levels and LASSO-based methods for greater shifts.

4.2.2. The MEWMA charts

The MEWMA chart has been used in various high-dimensional process monitoring studies following variable selection approaches. For example, Zou et al. (2011) suggested a LASSO-based diagnostic method that combined Bayesian information criterion (BIC) with an adaptive LASSO approach, followed by MEWMA chart monitoring of a white wine production dataset, demonstrating its usefulness in multivariate process diagnostics. The popularity of method stems from the piecewise linearity of the LASSO solution route, which reduces computing effort. Zou, Ning, et al. (2012) also developed a LASSO-based EWMA chart (LEWMA) for Phase II analysis, which is a variable-selection-driven multivariate control strategy. Comparative simulations revealed the LEWMA chart's acceptable diagnostic capability across a variety of circumstances, and its efficacy was confirmed in a real-world logistics services case study. Sangahn (2019) used the conditional sum of squares-based variable selection approach to analyze high-dimensional data, creating the CVS-MEWMA chart for monitoring processes such as LNG. The CVS-MEWMA chart outperformed other multivariate SPC charts (Shewhart, Hotelling's T^2, and MEWMA) and contemporary variable selection-based control charts (VSMSPC and VS-MEWMA). Zhao et al. (2021) developed two improved MEWMA control charts, AVS-MEWMA and ALEWMA, to track sparse mean changes. A comparison of anticipated weighted run length (EWRL) and ARL showed that AVS-MEWMA and ALEWMA were more stable and superior than current VS-MEWMA and LEWMA charts. Real-world applications using metallic aluminium data confirmed the efficacy of the suggested adaptive control charts.

4.2.3. The MCUSUM chart

Abdella et al. (2017) presented a variable-selection-based multivariate cumulative sum (VS-MCUSUM) chart for mean monitoring. The VS-MCUSUM chart outperformed standard MCUSUM, Hotelling's T^2 chart, and a variable selection-based chart (VS-EWMA) in identifying tiny process changes using specified parameters. Validation using hexagonal blot production data validated its efficacy.

4.2.4. Other charts

Alternative measure-based charts have arisen to monitor high-dimensional data, giving alternatives to standard forms like Hotelling's T^2, MEWMA, or MCUSUM control charts. K. Wang and Jiang (2009) developed the variable-selection-based Multivariate SPC (VS-MSPC) chart, a variable-selection-based Multivariate SPC method that beat Hotelling's T^2 in both simulated and actual Vertical Density Profiles (VDP) data trials. D. Yan et al. (2019) suggested a VS-based control chart that employs penalized likelihood and a Gaussian Mixture Model (GMM), resulting in strong fault identification in high-dimensional situations. Peres et al. (2019) addressed the constraints of Multiway Principal Component Analysis (MPCA) with the Pareto Variable Selection (PVS)-MPCA approach, demonstrating its effectiveness in a chocolate conching batch process case study. J. Guo et al. (2022) introduced the Bayesian Spike-Slab Composite Decomposition (BSSCD) technique for decomposing high-dimensional signals, with the Thompson Sampling Bayesian Spike-Slab Composite Decomposition change detection (TS-BSSCD(O) outperforming other online anomaly detection methods on a solar flare dataset.

4.3. High-dimensional control charts without the use of dimension reduction and variable selection techniques

Dimension reduction and variable selection are useful techniques for reducing high-dimensional data to a lower-dimensional space. However, these techniques provide insufficient information when the variables are highly correlated (known as multicollinearity). Additionally, both methods can be computationally intensive and time-consuming. These shortcomings have motivated researchers to select techniques that do not involve dimension reduction or variable selection. Over the years, many studies have been designed in SPC literature to address high-dimensional process monitoring without using dimension reduction or variable selection techniques. A comprehensive review of SPC studies is presented in the following subsections.

4.3.1. The Hotelling's T^2 charts

Ahmadi-Javid and Ebadi (2021) proposed a two-stage approach for monitoring mean changes in high-dimensional normally distributed multistream processes, which employs Hotelling and chi-square charts in the first phase and Shewhart charts for each stream in the second. Control limits and performance measurements were calculated correctly for both independent and dependent streams. Their technique has practical advantages, such as requiring only one chart when the process is under control. A comparative numerical study revealed that the technique often beats an existing method mentioned by (Meneces et al., 2008). Ullah et al. (2017) introduced a James-Stein shrinkage estimator that uses leave-one-out resampling to generate separate Hotelling's T^2 statistics. This method beats traditional Hotelling's T^2 in simulation and metabolomics datasets. M. R. Ahmad and Ahmed (2021) introduced a modified Hotelling's T^2 statistic for monitoring high-dimensional data. The version includes unified control limits for individual observations, subgroup cases, and both phases of control charts.

4.3.2. The MEWMA charts

Huang et al. (2014) developed the KMD-EWMA chart, a control chart based on Reproducing Kernel Hilbert Space (RKHS) that can detect a broader range of process variations than typical changes in mean or variance, even in complex high-dimensional processes that do not follow a Gaussian distribution. The KMD-EWMA chart outperformed Hotelling's T^2 and MEWMA charts in practical applications, as shown by ARL comparisons. Qi et al. (2016) introduced a one-sided MEWMA chart for high-dimensional data, which performed better in simulated data and aviation maintenance databases than previous techniques. Z. Wang et al. (2017) developed the CP chart, an EWMA-based chart for monitoring high-dimensional Poisson datasets, which outperformed the MP-CUSUM chart in Printed Circuit Board (PCB) manufacturing process. Du and Zou (2018) proposed a dynamic multiple-testing technique for high-dimensional data streams that control FDR using the EWMA scheme, demonstrating good monitoring performance in simulated experiments and illustrating a hard disk drive monitoring system.

Alaeddini et al. (2018) developed a spatiotemporal outlier identification approach that combines Partial Least Squares Discriminant Analysis (PLSDA) and Area Delaunay Triangulation (ADT) of

SPEs with an EWMA control chart. Their approach, KPLSDA_ADT, was compared against five different outlier identification methods and shown superiority. Z. Li and Tsung (2019) introduced the MVP chart, which combines a strong high-dimensional covariance matrix test with the MEWMA approach for monitoring variability. Comparative comparison with rival charts demonstrated the MVP chart's usefulness, particularly in identifying changes in covariance matrix blocks in higher dimensions. Feng et al. (2020) introduced the High-Dimensional EWMA (HDEWMA) and paired HDEWMA (PHD-EWMA) control charts, which integrates a test statistic into a MEWMA chart for Phase II process monitoring and outperforms its upgraded version, HDEWMA* and PHD-EWMA* charts. S. Kim et al. (2020) proposed a novel approach that uses L2 norm regularization to reduce mean process estimates to zero, expanding Ridge-based Multivariate SPC (RMSPC) to Adaptive RMSPC and RMEWMA charts. Their suggested charts were successful in detecting modest mean changes in high-dimensional processes, outperforming traditional multivariate SPC and current VS-based charts in online monitoring. S. Ahmad et al. (2020) suggested a MEWMA chart based on a Fast Region-based Convolutional Network approach, which has benefits such as eliminating the need for feature extraction for high-dimensional data and considerably increasing fault detection accuracy.

Li et al. (2020)proposed the HAMEWMA control chart, a nonparametric control chart based on the Hamiltonian path, and showed that it outperformed traditional multivariate nonparametric control charts like (Hawkins & Maboudou-Tchao, 2007), SREWMA (Zou, Wang, et al., 2012), and DFEWMA (Chen et al., 2016). Fang et al. (2022) developed the Data-Driven Pairwise EWMA (DPEWMA) chart for monitoring severe weather occurrences utilizing Water Vapor Flux Divergence (WVFD) data, which takes use of pairwise correlation between dimensions to monitor High-Dimensional Binary Data Streams (HBDS) efficiently. S. Kim et al. (2022) proposed the multivariate Bayesian EWMA (BEWMA) chart, which outperformed MEWMA and VSMEWMA in different shift situations with fewer sparse setups, as illustrated in a case study using a constantly stirred tank heater. E. Maboudou-Tchao et al. (2023) introduced LSMEWMA and SMEWMA charts based on Least Squares Support Vector Data Description (LS-SVDD) and Support Vector Data Description (SVDD), which outperformed traditional MEWMA (Lowry et al., 1992), RMEWMA (Kim et al., 2020) and L_0 MEWMA (Jiang et al., 2012) charts.

4.3.3. The MCUSUM chart

Liu et al. (2015) proposed a scalable CUSUM chart based on the sum of TRAS (Top-r-based Adaptive Sampling) statistics for online process change detection, which can identify mean changes in all directions assuming univariate normal distributions. Zou et al. (2015) presented a unique control chart that integrates information from several data streams using a CUSUM method based on a strong goodness-of-fit test, exhibiting superiority in semiconductor production operations. J. Li (2019) proposed a two-stage CUSUM process to overcome the limitations of single-stage solutions. The first stage utilizes IC ARL to set control limits, while the second stage employs Per-Comparison Error Rate (PCER) to reduce false alarms while recognizing OC data streams. This approach outperformed Y. Li and Tsung (2009) and Y. Li and Tsung (2012) methods, retaining control of global PCER in all conditions. Moala et al. (2019) investigated MCUSUM and MEWMA control charts for high-dimensional Spatial-Temporal Autoregressive Moving Average (STARMA) processes and found that the two performed similarly, with MCUSUM marginally better in certain circumstances.

Li (2020) proposed global monitoring statistics (G_t) based on order statistics of local monitoring statistics, and used the Cumulative Sum (CUSUM) statistic to identify anomalous conditions. Comparative simulation studies compared the suggested statistic (G_t) to global monitoring statistics by Zou et al. (2015) $(G||t^z)$ and a thresholding-based technique $(G||t^L)$. The suggested technique is computationally equivalent to G_t^L based monitoring, but substantially faster than G_t^z based methods. Ye et al. (2023) proposed a nonparametric asynchronous monitoring (NAM) framework for online monitoring of high-dimensional, asynchronous, and heterogeneous data streams. This approach supports various sampling intervals and arbitrary distributions, providing more flexibility than earlier synchronous and parametric assumptions. Local CUSUM statistics based on quantile-based ordered intervals were used to calculate a global monitoring statistic, which outperformed compensatory schemes and benchmark approaches such as MRS-KNN. The suggested NAM technique was shown to be successful in identifying fast changes when tested on semiconductor manufacturing process and TEP datasets.

4.3.4. Other charts

In SPC literature, many studies have been conducted using neither the dimension-reduction /variable-selection method nor conventional structures such as Hotelling's T^2, MEWMA, or MCUSUM control charts. For example, H. Zhang and Albin (2009) proposed the χ^2 control chart method, which is effective in detecting outliers while preserving most non-outlier profiles. The authors conducted both a simulation and case study analysis using the VDP dataset to evaluate the performance of the proposed chart, using the average percentage (and standard deviation) as the evaluation measure, and found satisfactory results in detecting faults. Hu and Yuan (2009) proposed a method for monitoring batch processes using tensor factorization and the TLPP. Unlike existing methods, such as multiway principal component analysis (MPCA) and MLPP, which represent batch data as a vector in high-dimensional space, TLPP avoids information loss by not vectorizing the data. The effectiveness of the TLPP was tested using benchmark fed-batch penicillin fermentation process data and two industrial fed-batch processes of penicillin and cephalosporin fermentation and compared with MPCA and MLPP monitoring techniques. The results were shown on SPE charts, and it was found that the TLPP detected faults earlier than the MLPP and MPCA. In addition, the results for different numbers of reference batches in the benchmark fed-batch penicillin fermentation examples demonstrated that TLPP had a much lower false alarm rate than MLPP and MPCA, especially with smaller reference datasets. S. Chen and Nembhard (2011) proposed an HD control chart for monitoring profiles. Their approach relies on the Adaptive Neyman (AN) test for coefficients and the Discrete Fourier Transform (DFT) algorithm. The chart can track profiles without necessitating the fitting of regression models and is effective for monitoring both linear and nonlinear profiles with autocorrelated stationary noise.

Yu (2012b) developed a Nonlinear Kernel Gaussian Mixture Model (NKGMM)-based chart for chemical process defect detection and diagnosis, and demonstrated its efficiency in monitoring simulated wastewater treatment processes with diverse operating modes. Compared to GMM and Independent Component Analysis (ICA)-based charts, the NKGMM technique performed better in identifying and diagnosing problems, especially in step change and drift errors. Y. Li et al. (2014) presented the HC chart for high-dimensional change-point detection, which can detect measurement changes across dimensions early, resulting in a shorter average run time than the ZH control chart (Zamba & Hawkins, 2006). Zou et al. (2014) developed Penalized Profile Outlier Detection (PPOD) methods, which use penalized regression to locate outliers in high-dimensional profile vectors. Simulations and real-world industrial data were used to evaluate PPOD-R's usefulness in spotting outliers across diverse settings, notably in semiconductor production. Kan and Yang (2015) proposed a dynamic network framework for profile monitoring of high-dimensional imaging streams, which includes effective online control charting for complicated picture profiles. Turkoz et al. (2016) presented the Distribution Free Adaptive Step-Down (DFASD) approach for finding fault variables, which addresses difficulties such as variable correlation and computational complexity in high-dimensional data. DFASD outperformed other approaches in both normal and non-normal datasets.

Jang et al. (2017) suggested three real-time contrast-based control charts (FWRF, GWRF, and MWRF) for variance monitoring in high-dimensional data, demonstrating their usefulness across several data types. R. C. Wang et al. (2017) conducted a study of data visualization strategies for chemical industry defect identification, emphasizing the efficacy of time-explicit Kiviat diagrams in various chemical processes. Shu and Fan (2018) developed an Interpoint Distance-based control chart (IPD chart) for mean vector monitoring in high-dimensional settings and demonstrated its effectiveness using simulations of real semiconductor production data. H. Yan et al. (2018) introduced a monitoring approach based on the sequential likelihood ratio of a Spatio-Temporal Smooth Sparse Decomposition (ST-SSD) test, which performed better in real-world case studies than previous methods. He et al. (2018) presented a distance-based multivariate process control chart (D-SVM chart) based on Support Vector Machines (SVM), which showed success in identifying mean and variance changes in a case study, including camera performance metrics from mobile phone manufacturing.

Guo et al. (2019) suggested an RTC-based technique for defect detection in industrial picture data, resulting in reliable defect estimate and identification. Abdella et al. (2019) developed the Adaptive LASSO Thresholding-norm chart (ALT-norm chart), which can identify shifts of varying magnitudes and patterns without depending on shift-pattern data. J. Kim et al. (2019) introduced the RPLR (ridge-penalized likelihood ratio) chart, which is excellent for monitoring changes in the covariance matrix without needing shift pattern information. Turkoz et al. (2019) proposed Bayesian Support Vector Data Description (BSVDD) for fault diagnosis in non-normal data, which outperformed existing approaches in identifying irregular patterns. W. Li et al. (2020) suggested a marginal data-driven wMDR-control procedure (MDW) to diagnose high-dimensional data streams while efficiently reducing the missed discovery rate. Martinez et al. (2020) presented One-Class Peeling (OCP) method to identify outliers in multivariate data, displaying greater performance over existing approaches. Abdella et al. (2020) assessed the T-COV chart's ability to identify changes in the covariance matrix and found that it outperformed the Conditional Entropy (CE) control chart.

Maboudou-Tchao (2021) proposed the Support Tensor Vector Data Description (STVDD) chart for high-dimensional data utilizing second-order tensors, which outperformed SVDD, particularly in distribution-free settings. Fan et al. (2021) proposed the SLED chart for Phase I control, which demonstrated more detection power than current approaches for detecting breast cancer. S. Kim and Turkoz (2022) introduced the Bayesian Sequential Update (BSU) chart, which outperforms MEWMA and VS-MEWMA at spotting tiny movements fast. Merlo et al. (2022) coupled Mahalanobis distances with the Mann-Whitney statistic to generate MW(D) or MW(I) charts, providing a distribution-free solution for high-dimensional processes. Xiang et al. (2022) used a data-driven LIS-based diagnostic approach to identify aberrant mean changes in linked high-dimensional data streams, and their findings were successful. Mukherjee and Marozzi (2022) proposed Shewhart-type nonparametric control charts for combined monitoring of location and scale parameters in multivariate high-dimensional processes, proving performance using simulations and real-world data from semiconductor production.

5. Conclusions and future directions

The term 'high-dimensional data' refers to a dataset comprising many variables or features, usually more than the number of observations. This paper provides a detailed overview of high-dimensional monitoring methods using a sample of 82 relevant articles published between 2004 and 2023. Three categories can be found in the literature on high-dimensional control charts: control charts based on the dimension reduction approach, variable selection method, and high-dimensional techniques. In addition, the literature on each approach is categorized by control chart structures such as memory-less (Hotelling T^2 chart), memory-based (MEWMA and MCUSUM charts), and others. Real-life examples of high-dimensional datasets utilized in various fields, such as industry, medical science, chemical engineering, and image processing, are also presented. This study found that research on high-dimensional control charts increased significantly after 2016, and most studies aimed to develop high-dimensional control charts using high-dimensional techniques. Furthermore, most studies introduced memory-based (MEWMA and MCUSUM) charts and other structures to monitor high-dimensional data.

After a comprehensive review of the studies on high-dimensional monitoring methods, several future research directions were observed, with some significant ones being listed below:

- **Monitoring the covariance matrix of high-dimensional data**: In the above review, very few studies are designed to monitor the covariance matrix of high-dimensional data (Abdella et al., 2020; Fan et al., 2021; Kim et al., 2019; Li & Tsung, 2019). However, monitoring the mean vector of high-dimensional data will be seriously affected if the covariance matrix of the high-dimensional data is unstable. Hence, more effective high-dimensional data-based charts are proposed for monitoring covariance matrices. In addition, one may offer combined structures to monitor the mean and covariance matrices simultaneously.
- **Joint monitoring of mean vector and covariance matrix of high-dimensional data**: As discussed above, monitoring the mean vector of high-dimensional data requires stability in the covariance matrix. Instead of proposing combined structures, joint monitoring methods based on a single plotting statistic of high-dimensional data's mean vector and covariance matrix will be another possible opportunity to look at simultaneous shifts in means and variances in high-dimensional data.
- **Phase I analysis of the high-dimensional control charts**: In practice, most processes do not have known parameters (Case-U); hence, an effective Phase I analysis is typically required to determine the parameters of the monitoring mechanism. The above analysis of high-dimensional control chart studies shows that most were designed under the assumption of known parameters (Case-U). Hence, exploring the high-dimensional control chart under Case-U will be an excellent contribution to the current literature.
- **Effect of estimation errors on high-dimensional control charts**: In the current era of Industry 4.0, efficient sensors are used to record data; however, sometimes, natural estimation errors are recorded due to the sensor's lifetime, settlement, connectivity, synchronization, and human mishandling. Hence, this issue has not been studied, and it is a possible research topic to be highlighted and addressed using the high-dimensional control chart based on estimation error models.
- **Effect of violation of independence and normality assumptions**: A good number of works have been carried out to monitor the shifts in high-dimensional data when the assumptions of independence and normality assumptions are not violated. However, only a few studies have been accomplished when the independence assumption is not satisfied. For example, Moala et al. (2019) used STARMA process to deal with such issues. Similarly, only some studies tried to develop proper control charts for high-dimensional data when the normality assumption is violated. Hence, an interesting research direction would be analyzing the changes in the detection ability of high-dimensional control charts when the abovementioned assumptions are not satisfied.
- **Problems with sensitivity, robustness, and changepoint detection**: The design of high-dimensional control charts should be moderately sensitive. Excessive sensitivity can detect shifts in the process caused by natural or minor variations that will produce false alarms, which is not expected. A lack of sensitivity may lead to late detection of shifts in the process, which will result in an ineffective control chart. Hence, high-dimensional control charts with moderate sensitivity should be considered in future studies. In addition, changepoint or outlier detection for high-dimensional data is not as simple as the low-dimensional cases. Therefore, constructing robust high-dimensional control charts that will effectively detect shifts with moderate sensitivity to extreme values can be a possible direction for future researchers in this area of research.
- **Mean or variance problem of two or more populations**: The above comprehensive review found that most of the high-dimensional control charts are designed for detecting a change in the mean vector or covariance matrix of the high-dimensional data of a single population/group. However, sometimes researchers are interested in real-time detection of change in the

mean vectors of two or more groups or variance matrices of two or more groups. For example, biologists are mostly interested in simultaneously examining mean changes in the genes of male and female patient groups (Chen & Qin, 2010; Thulin, 2014; Zhang & Wang, 2021). To support such cases, extending the high-dimensional control charts for real-time detection of change in the mean vectors of two or more groups or variance matrices of two or more groups will be an exceptional addition to the literature.

- **Count-based high-dimensional control charts**: Most high-dimensional control charts are variable-type charts developed for continuous-type high-dimensional data, such as the multivariate normal distribution, multivariate Gamma distribution, multivariate Cauchy distribution, and multivariate t-distribution. However, many processes generate count-based or discrete-type high-dimensional data that follow a multivariate Poisson distribution, multivariate binomial distribution, multivariate negative binomial distribution, and multivariate Conway-Maxwell Poisson distribution. An example is the number of delamination in the (PCB) packaging process (Wang et al., 2017). Hence, extending high-dimensional control charts to count-based or discrete-type high-dimensional data is a possible future research direction.
- **Regression model based high-dimensional control charts**: Generally, some covariates are recorded along with the main high-dimensional study variables. Therefore, linear or nonlinear profile methods are usually preferred to monitor data by maintaining the relationships among variables. In the current review, a few studies (Chen & Nembhard, 2011; Kim et al., 2020; Zou et al., 2014) have focused on linear and nonlinear profile monitoring under a high-dimensional setup. Hence, extending the existing literature to linear and nonlinear high-dimensional profile monitoring can be another direction to enhance the literature.
- **Machine learning based high-dimensional control charts**: With the fourth revolution in the 21st century, the use of machine learning techniques from artificial intelligence has become a demanding area of research for monitoring high-dimensional data. This study has noticed a few of the research (He et al., 2018; Lee et al., 2019; Maboudou-Tchao, 2021; Martinez et al., 2020; Yu & Liu, 2022; Yu & Zhang, 2020) applied machine learning tools to monitor shifts in the process. More studies are needed to make a solid bridge between machine learning techniques and already proposed methods to address high-dimensional monitoring.
- **Computational libraries or packages for high-dimensional control charts**: A total of 82 high-dimensional control chart studies are discussed in the current review. However, the proposed methods are not available in any computational software. Adding these methods in the form of packages in libraries in R-language and Python may decrease the complexity and allow practitioners to implement them easily and frequently.

Table of Acronyms

Acronym	Definition
1DRCAE	One-Dimension Residual Convolutional Auto-Encoder
ADD	Average Detection Delay
ADT	Area Delaunay Triangulation
ALEWMA	Adaptive LEWMA
ALT	Adaptive LASSO Thresholding
ARL	Average Run Length
ASD	Adaptive Step-Down
AVS-MEWMA	Adaptive VS-MEWMA
BSS	Blind Source Separation
BSSCD	Bayesian Spike-Slab Composite Decomposition
BSU	Bayesian Sequential Update
BSVDD	Bayesian Support Vector Data Description
CE	Conditional Entropy

CR	Correctness Rates
DFASD	Distribution Free Adaptive Step-Down
DL	Deep Learning
DNN	Deep Neural Network
DPCA	Dynamic PCA
DPEWMA	Data-Driven Pairwise EWMA
EER	Expected Error Rates
EQL	Extra Quadratic Loss
EWRL	Expectation Weighted Run Length
FAR	False Alarm Rate
Fast R-CNN	Fast Region based Convolutional Network
FBFP	Fed-Batch Fermentation Penicillin
FCR	Fault Correctness Ratio
FDR	Fault Detection Rate
FPCA	Functional PCA
GLR	Generalized Likelihood Ratio
GMM	Gaussian Mixture Model
HBDS	High-Dimensional Binary Data Streams
HDEWMA	High-Dimensional EWMA
ICA	Independent Component Analysis
IHDEWMA	Individual HDEWMA
JD	Joint Decorrelation
KNN_MSE	K-Nearest Neighbor-Mean Squared Errors
KPCA	Kernel Principal Component Analysis
KPLSDA	Kernel PLSDA
KPLSDA_MAX_E	KPLSDA Maximum Error
KPLSDA_MAX_KE	KPLSDA top K Errors
KPLSDA_MSE	KPLSDA Mean Squared Errors
LAM_MSE	Linear Autoregressive Model-Mean Squared Errors
LEWMA	LASSO-based EWMA
LGPCA	Local and Global Principal Component Analysis
LPP	Locality Preserving Projections
LSMEWMA	Least Squares Support Vector Data Description MEWMA
MEWMV	Multivariate Exponentially Weighted Moving Variance
MFPCA	Multichannel Functional Principal Component Analysis
MLPP	Multiway Locality Preserving Projections
MPCA	Multi-linear PCA
MPCA	Multiway Principal Component Analysis
MRSAE	Manifold Regularized Stacked Auto-Encoders
MEWMA	Multivariate Exponentially Weighted Moving Average
MCUSUM	Multivariate Cumulative Sum
NCSC	Nearest Correlation Spectral Clustering
NFL	National Football League
NKGMM	Nonlinear Kernel Gaussian Mixture Model
OCP	One-Class Peeling
PCA	Principal Component Analysis
PHD-EWMA	Paired HDEWMA
PLR	Penalized Likelihood Ratio
PLSDA	Partial Least Squares Discriminant Analysis
PSA	Pressure Swing Adsorption
PVS – MPCA	Pareto Variable Selection-Multiway Principal Component Analysis
RARL	Relative Average Run Length
RMDP	Robust Minimum Diagonal Product
RMI	Relative Mean Index
RPLR	Ridge Penalized Likelihood Ratio
RTC	Real-Time Contrasts
SDAE	Stacked Denoising Auto-Encoders
SDRL	Standard Deviations of the Run-Length
SMEWMA	Support Vector Data Description MEWMA
SMFPCA	Sparse Multi-channel Functional Principal Component Analysis
SPC	Statistical Process Control

SPCRAE	Sparse Principal Component Regression with Autocorrelated Errors
SPE	Squared Prediction Error
SPE	Stochastic Proximity Embedding
SSMRL	Steady-State Median Run Length
STVDD	Support Tensor Vector Data Description
SVDD	Support Vector Data Description
TEP	Tennessee Eastman Process
TLPP	Tensor Locality Preserving Projections
TRAS	Top-r based Adaptive Sampling
UMPCA	Uncorrelated Multi-linear PCA
VAE	Variational Auto-Encoder
VDP	Vertical Density Profiles
VPCA	Vectorized PCA
VS-MSPC	Variable-Selection-Based Multivariate SPC
WVFD	Water Vapor Flux Divergence

Disclosure statement

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

Notes on contributors

Fuhad Ahmed has completed his MS (Applied Statistics) from Department of Mathematics, College of Computing and Mathematics, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia. Her current research interests include Statistical Process Monitoring, Sampling Techniques and Time Series Modelling.

Tahir Mahmood got his degree of BS (Hons.) in Statistics with distinction (Gold Medalist) from the Department of Statistics, University of Sargodha, Sargodha, Pakistan. He has received an MS (Applied Statistics) degree from the Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia. He received his PhD degree from the Department of Systems Engineering, City University of Hong Kong, Hong Kong. Nowadays, he is a Lecturer in the School of Computing, Engineering and Physical Sciences, University of the West of Scotland, United Kingdom. Before this, he worked as a lecturer in the Hong Kong Metropolitan University, Hong Kong, and as an Assistant Professor in the Industrial and Systems Engineering Department, KFUPM, Saudi Arabia. His current research interests include Statistical Process Monitoring, Data Mining and Generalized Linear Modelling.

Muhammad Riaz obtained PhD degree in Statistics from the Institute for Business and Industrial Statistics, University of Amsterdam, The Netherlands, in 2008. He holds the position of Professor in the Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia. His current research interests include statistical process control, nonparametric techniques, and experimental design.

Nasir Abbas is Associate Professor of Statistics at the Department of Mathematics, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia. His current research interests include mathematical statistics and quality control, particularly control charting methodologies under parametric and nonparametric environments.

ORCID

Tahir Mahmood http://orcid.org/0000-0002-8748-5949
Muhammad Riaz http://orcid.org/0000-0002-7599-6928
Nasir Abbas http://orcid.org/0000-0002-8622-5467

References

Abdella, G. M., Al-Khalifa, K. N., Kim, S., Jeong, M. K., Elsayed, E. A., & Hamouda, A. M. (2017). Variable selection-based multivariate cumulative sum control chart. *Quality and Reliability Engineering International*, *33*(3), 565–578. https://doi.org/10.1002/qre.2041

Abdella, G. M., Kim, J., Kim, S., Al-Khalifa, K. N., Jeong, M. K., Hamouda, A. M., & Elsayed, E. A. (2019). An adaptive thresholding-based process variability monitoring. *Journal of Quality Technology*, *51*(3), 242–256. https://doi.org/10.1080/00224065.2019.1569952

Abdella, G. M., Maleki, M. R., Kim, S., Al-Khalifa, K. N., & Hamouda, A. M. S. (2020). Phase-I monitoring of high-dimensional covariance matrix using an adaptive thresholding LASSO rule. *Computers & Industrial Engineering*, *144*, 106465. https://doi.org/10.1016/j.cie.2020.106465

Ahmad, M. R., & Ahmed, S. E. (2021). On the distribution of the T2 statistic, used in statistical process monitoring, for high-dimensional data. *Statistics & Probability Letters*, *168*, 108919. https://doi.org/10.1016/j.spl.2020.108919

Ahmad, S., Enshaei, N., Naderkhani, F., & Awasthi, A. (2020, June 8–10). *Integrated deep learning and statistical process control for online monitoring of manufacturing processes*. Paper Presented at the 2020 IEEE International Conference on Prognostics and Health Management (ICPHM), Detroit, Michigan, United States of America.

Ahmadi-Javid, A., & Ebadi, M. (2021). A two-step method for monitoring normally distributed multi-stream processes in high dimensions. *Quality Engineering*, *33*(1), 143–155. https://doi.org/10.1080/08982112.2020.1786118

Alaeddini, A., Motasemi, A., & 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. https://doi.org/10.1080/24725854.2017.1386336

Bae, S. J., Do, G., & Kvam, P. (2016). On data depth and the application of nonparametric multivariate statistical process control charts. *Applied Stochastic Models in Business and Industry*, *32*(5), 660–676. https://doi.org/10.1002/asmb.2186

Bersimis, S., Psarakis, S., & Panaretos, J. (2007). Multivariate statistical process control charts: An overview. *Quality and Reliability Engineering International*, *23*(5), 517–543. https://doi.org/10.1002/qre.829

Chen, J., Park, C., Kim, S.-H., & Xie, Y. (2019). To reduce or not to reduce: A study on spatio-temporal surveillance. *Environmental and Ecological Statistics*, *26*(3), 217–238. https://doi.org/10.1007/s10651-019-00425-4

Chen, N., Zi, X., & Zou, C. (2016). A distribution-free multivariate control chart. *Technometrics*, *58*(4), 448–459. https://doi.org/10.1080/00401706.2015.1049750

Chen, S., & Nembhard, H. B. (2011). A high-dimensional control chart for profile monitoring. *Quality and Reliability Engineering International*, *27*(4), 451–464. https://doi.org/10.1002/qre.1140

Chen, S. X., & Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Annals of Statistics*, *38*(2), 808–835, 828. https://doi.org/10.1214/09-AOS716

Crosier, R. B. J. T. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, *30*(3), 291–303. https://doi.org/10.1080/00401706.1988.10488402

Du, L., & Zou, C. (2018). On-line control of false discovery rates for multiple datastreams. *Journal of Statistical Planning and Inference*, *194*, 1–14. https://doi.org/10.1016/j.jspi.2017.10.006

Erem, A., & Mahmood, T. (2023). A bivariate CUSUM control chart based on exceedance statistics. *Quality and Reliability Engineering International*, *39*(4), 1172–1191. https://doi.org/10.1002/qre.3285

Fan, J., Shu, L., Yang, A., & Li, Y. (2021). Phase I analysis of high-dimensional covariance matrices based on sparse leading eigenvalues. *Journal of Quality Technology*, *53*(4), 333–346. https://doi.org/10.1080/00224065.2020.1746212

Fang, Z., Li, W., Liu, X., Pu, X., & Xiang, D. (2022). Online monitoring of high-dimensional binary data streams with application to extreme weather surveillance. *Journal of Applied Statistics*, *49*(16), 4122–4136. https://doi.org/10.1080/02664763.2021.1971633

Feng, L., Ren, H., & Zou, C. (2020). A setwise EWMA scheme for monitoring high-dimensional datastreams. *Random Matrices: Theory and Applications*, *9*(2), 2050004. https://doi.org/10.1142/s2010326320500045

Fujiwara, K., & Kano, M. (2017, July). *Development of correlation-based process characteristics visualization method and its application to fault detection*. Paper Presented at the 2017 13th IEEE International Conference on Control & Automation (ICCA), Ohrid, Macedonia.

Gómez, A. M. E., Li, D., & Paynabar, K. (2022). An adaptive sampling strategy for online monitoring and diagnosis of high-dimensional streaming data. *Technometrics*, *64*(2), 253–269. https://doi.org/10.1080/00401706.2021.1967198

Guo, J., Yan, H., & Zhang, C. (2022). A bayesian partially observable online change detection approach with Thompson Sampling. *Technometrics*, *65*(2), 179–191. https://doi.org/10.1080/00401706.2022.2127914

Guo, X., He, Z., & Chen, H. (2019, April 12–15). *A real-time contrasts method for monitoring image data*. Paper Presented at the 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), Tokyo, Japan.

Hawkins, D. M., & Maboudou-Tchao, E. M. (2007). Self-starting multivariate exponentially weighted moving average control charting. *Technometrics*, *49*(2), 199–209. https://doi.org/10.1198/004017007000000083

He, S., Jiang, W., & Deng, H. (2018). A distance-based control chart for monitoring multivariate processes using support vector machines. *Annals of Operations Research*, *263*(1), 191–207. https://doi.org/10.1007/s10479-016-2186-4

Hu, K., & Yuan, J. (2008). Multivariate statistical process control based on multiway locality preserving projections. *Journal of Process Control*, *18*(7), 797–807. https://doi.org/10.1016/j.jprocont.2007.11.002

Hu, K., & Yuan, J. (2009). Batch process monitoring with tensor factorization. *Journal of Process Control*, *19*(2), 288–296. https://doi.org/10.1016/j.jprocont.2008.03.003

Huang, S., Kong, Z., & Huang, W. (2014). High-dimensional process monitoring and change point detection using embedding distributions in reproducing kernel Hilbert space. *IIE Transactions*, *46*(10), 999–1016. https://doi.org/10.1080/0740817X.2013.855848
Jang, S., Park, S. H., & Baek, J.-G. (2017). Real-time contrasts control chart using random forests with weighted voting. *Expert Systems with Applications*, *71*, 358–369. https://doi.org/10.1016/j.eswa.2016.12.002
Jiang, W., Wang, K., & Tsung, F. (2012). A variable-selection-based multivariate EWMA chart for process monitoring and diagnosis. *Journal of Quality Technology*, *44*(3), 209–230. https://doi.org/10.1080/00224065.2012.11917896
Kan, C., & Yang, H. (2015, August 24–28). *Network models for monitoring high-dimensional image profiles*. Paper Presented at the 2015 IEEE International Conference on Automation Science and Engineering (CASE), Gothenburg, Sweden.
Kan, C., & Yang, H. (2017). Dynamic network monitoring and control of in situ image profiles from ultraprecision machining and biomanufacturing processes. *Quality and Reliability Engineering International*, *33*(8), 2003–2022. https://doi.org/10.1002/qre.2163
Kim, J., Abdella, G. M., Kim, S., Al-Khalifa, K. N., & Hamouda, A. M. (2019). Control charts for variability monitoring in high-dimensional processes. *Computers & Industrial Engineering*, *130*, 309–316. https://doi.org/10.1016/j.cie.2019.02.012
Kim, J., Jeong, M. K., Elsayed, E. A., Al-Khalifa, K. N., & Hamouda, A. M. S. (2016). An adaptive step-down procedure for fault variable identification. *International Journal of Production Research*, *54*(11), 3187–3200. https://doi.org/10.1080/00207543.2015.1076948
Kim, J.-M., Wang, N., Liu, Y., & Park, K. (2020). Residual control chart for binary response with multicollinearity covariates by neural network Model. *Symmetry*, *12*(3), 381. https://doi.org/10.3390/sym12030381
Kim, S., Jeong, M. K., & Elsayed, E. A. (2020). A penalized likelihood-based quality monitoring via L2-norm regularization for high-dimensional processes. *Journal of Quality Technology*, *52*(3), 265–280. https://doi.org/10.1080/00224065.2019.1571348
Kim, S., & Turkoz, M. (2022). Bayesian sequential update for monitoring and control of high-dimensional processes. *Annals of Operations Research*, *317*(2), 693–715. https://doi.org/10.1007/s10479-021-04188-9
Kim, S., Turkoz, M., & Baek, J. W. (2022). High-speed monitoring of multidimensional processes using bayesian updates. *Institute of Electrical and Electronics Engineers Access*, *10*, 97450–97464. https://doi.org/10.1109/ACCESS.2022.3206369
Lee, J.-M., Yoo, C., Choi, S. W., Vanrolleghem, P. A., & Lee, I.-B. (2004). Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, *59*(1), 223–234. https://doi.org/10.1016/j.ces.2003.09.012
Lee, S., Kwak, M., Tsui, K.-L., & Kim, S. B. (2019). Process monitoring using variational autoencoder for high-dimensional nonlinear processes. *Engineering Applications of Artificial Intelligence*, *83*, 13–27. https://doi.org/10.1016/j.engappai.2019.04.013
Li, J. (2019). A two-stage online monitoring procedure for high-dimensional data streams. *Journal of Quality Technology*, *51*(4), 392–406. https://doi.org/10.1080/00224065.2018.1507562
Li, J. (2020). Efficient global monitoring statistics for high-dimensional data. *Quality and Reliability Engineering International*, *36*(1), 18–32. https://doi.org/10.1002/qre.2557
Li, W., Xiang, D., Tsung, F., & Pu, X. (2020). A diagnostic procedure for high-dimensional data streams via missed discovery rate control. *Technometrics*, *62*(1), 84–100. https://doi.org/10.1080/00401706.2019.1575284
Li, Y., Liu, Y., Zou, C., & Jiang, W. (2014). A self-starting control chart for high-dimensional short-run processes. *International Journal of Production Research*, *52*(2), 445–461. https://doi.org/10.1080/00207543.2013.832001
Li, Y., Pei, D., & Wu, Z. (2020). A multivariate non-parametric control chart based on run test. *Computers & Industrial Engineering*, *149*, 106839. https://doi.org/10.1016/j.cie.2020.106839
Li, Y., & Tsung, F. (2009). False discovery rate-adjusted charting schemes for multistage process monitoring and fault identification. *Technometrics*, *51*(2), 186–205. https://doi.org/10.1198/TECH.2009.0019
Li, Y., & Tsung, F. (2012). Multiple attribute control charts with false discovery rate control. *Quality and Reliability Engineering International*, *28*(8), 857–871. https://doi.org/10.1002/qre.1276
Li, Z., & Tsung, F. (2019). A control scheme for monitoring process covariance matrices with more variables than observations. *Quality and Reliability Engineering International*, *35*(1), 351–367. https://doi.org/10.1002/qre.2403
Liu, K., Mei, Y., & Shi, J. (2015). An adaptive sampling strategy for online high-dimensional process monitoring. *Technometrics*, *57*(3), 305–319. https://doi.org/10.1080/00401706.2014.947005
Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, *34*(1), 46–53. https://doi.org/10.2307/1269551
Maboudou-Tchao, E., Harrison, C. W., & Sen, S. (2023). A comparison study of penalized likelihood via regularization and support vector-based control charts. *Quality Technology & Quantitative Management*, *20*(2), 147–167. https://doi.org/10.1080/16843703.2022.2096198
Maboudou-Tchao, E. M. (2021). High-dimensional data monitoring using support machines. *Communications in Statistics - Simulation and Computation*, *50*(7), 1927–1942. https://doi.org/10.1080/03610918.2019.1588312

Mahmood, T., & Erem, A. (2023). A bivariate exponentially weighted moving average control chart based on exceedance statistics. *Computers & Industrial Engineering*, *175*, 108910. https://doi.org/10.1016/j.cie.2022.108910

Mahmood, T., Wittenberg, P., Zwetsloot, I. M., Wang, H., & Tsui, K. L. (2019). Monitoring data quality for telehealth systems in the presence of missing data. *International Journal of Medical Informatics*, *126*, 156–163. https://doi.org/10.1016/j.ijmedinf.2019.03.011

Martinez, W. G., Weese, M. L., & Jones-Farmer, L. A. (2020). A one-class peeling method for multivariate outlier detection with applications in phase I SPC. *Quality and Reliability Engineering International*, *36*(4), 1272–1295. https://doi.org/10.1002/qre.2629

Mason, R. L., Tracy, N. D., & Young, J. C. (1995). Decomposition of T^2 for multivariate control chart interpretation. *Journal of Quality Technology*, *27*(2), 99–108. https://doi.org/10.1080/00224065.1995.11979573

Mei, Y. (2010). Efficient scalable schemes for monitoring a large number of data streams. *Biometrika*, *97*(2), 419–433. https://doi.org/10.1093/biomet/asq010

Meneces, N. S., Olivera, S. A., Saccone, C. D., & Tessore, J. (2008). Statistical control of multiple-stream processes: A shewhart control chart for each stream. *Quality Engineering*, *20*(2), 185–194. https://doi.org/10.1080/08982110701241608

Merlo, J., Cordero-Franco, A. E., & Tercero-Gómez, V. G. (2022). Nonparametric multivariate processes monitoring with guaranteed in-control performance for changes in location. *Computers & Industrial Engineering*, *166*, 107940. https://doi.org/10.1016/j.cie.2022.107940

Moala, A. B., Ho, L. L., & Quinino, R. C. (2019). Multivariate control charts to monitor the monthly frequency of vehicle robberies in São Paulo city. *Spatial Statistics*, *29*, 49–65. https://doi.org/10.1016/j.spasta.2018.09.002

Mukherjee, A., & Marozzi, M. (2022). Nonparametric phase-ii control charts for monitoring high-dimensional processes with unknown parameters. *Journal of Quality Technology*, *54*(1), 44–64. https://doi.org/10.1080/00224065.2020.1805378

Ning, X., & Tsung, F. (2012). A density-based statistical process control scheme for high-dimensional and mixed-type observations. *IIE Transactions*, *44*(4), 301–311. https://doi.org/10.1080/0740817X.2011.587863

Odom, G. J., Newhart, K. B., Cath, T. Y., & Hering, A. S. (2018). Multistate multivariate statistical process control. *Applied Stochastic Models in Business and Industry*, *34*(6), 880–892. https://doi.org/10.1002/asmb.2333

Peres, F. A. P., Peres, T. N., Fogliatto, F. S., & Anzanello, M. J. (2019). Fault detection in batch processes through variable selection integrated to multiway principal component analysis. *Journal of Process Control*, *80*, 223–234. https://doi.org/10.1016/j.jprocont.2019.06.002

Qi, D., Li, Z., & Wang, Z. (2016). On-line monitoring data quality of high-dimensional data streams. *Journal of Statistical Computation and Simulation*, *86*(11), 2204–2216. https://doi.org/10.1080/00949655.2015.1106542

Qin, S. J., Dong, Y., Zhu, Q., Wang, J. & Liu, Q. (2020). Bridging systems theory and data science: A unifying review of dynamic latent variable analytics and process monitoring. *Annual Reviews in Control*, *50*, 29–48. https://doi.org/10.1016/j.arcontrol.2020.09.004

Qiu, P., & Hawkins, D. (2003). A nonparametric multivariate cumulative sum procedure for detecting shifts in all directions. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *52*(2), 151–164. https://doi.org/10.1111/1467-9884.00348

Rato, T., Reis, M., Schmitt, E., Hubert, M., & De Ketelaere, B. (2016). A systematic comparison of PCA-based Statistical Process Monitoring methods for high-dimensional, time-dependent Processes. *AIChE Journal*, *62*(5), 1478–1493. https://doi.org/10.1002/aic.15062

Samuel, R. T., & Cao, Y. (2014, September). *Fault detection in a multivariate process based on kernel PCA and kernel density estimation*. Paper Presented at the 2014 20th International Conference on Automation and Computing, Cranfield, UKnited Kingdom (pp. 12–13).

Sangahn, K. (2019). Variable selection-based SPC procedures for high-dimensional multistage processes. *Journal of Systems Engineering and Electronics*, *30*(1), 144–153. https://doi.org/10.21629/JSEE.2019.01.14

Shaohui, M., Tuerhong, G., Wushouer, M., & Yibulayin, T. (2022). PCA mix-based Hotelling's T^2 multivariate control charts for intrusion detection system. *IET Information Security*, *16*(3), 161–177. https://doi.org/10.1049/ise2.12051

Shu, L., & Fan, J. (2018). A distribution-free control chart for monitoring high-dimensional processes based on interpoint distances. *Naval Research Logistics*, *65*(4), 317–330. https://doi.org/10.1002/nav.21809

Skubalska-Rafajłowicz, E. (2013). Random projections and Hotelling's T2 statistics for change detection in high-dimensional data streams. *International Journal of Applied Mathematics and Computer Science*, *23*(2), 447–461. https://doi.org/10.2478/amcs-2013-0034

Sullivan, J. H., Stoumbos, Z. G., Mason, R. L., & Young, J. C. (2007). Step-down analysis for changes in the covariance matrix and other parameters. *Journal of Quality Technology*, *39*(1), 66–84. https://doi.org/10.1080/00224065.2007.11917674

Sun, D., Lu, G., Zhou, H., & Yan, Y. (2013). Condition monitoring of combustion processes through flame imaging and kernel principal component analysis. *Combustion Science and Technology*, *185*(9), 1400–1413. https://doi.org/10.1080/00102202.2013.798316

Thulin, M. (2014). A high-dimensional two-sample test for the mean using random subspaces. *Computational Statistics & Data Analysis*, *74*, 26–38. https://doi.org/10.1016/j.csda.2013.12.003
Tracy, N. D., Young, J. C., & Mason, R. L. (1992). Multivariate control charts for individual observations. *Journal of Quality Technology*, *24*(2), 88–95. https://doi.org/10.1080/00224065.1992.12015232
Turkoz, M., Kim, S., Jeong, Y.-S., Al-Khalifa, K. N., & Hamouda, A. M. (2016). Distribution-free adaptive step-down procedure for fault identification. *Quality and Reliability Engineering International*, *32*(8), 2701–2716. https://doi.org/10.1002/qre.2096
Turkoz, M., Kim, S., Jeong, Y.-S., Jeong, M. K., Elsayed, E. A., Al-Khalifa, K. N., & Hamouda, A. M. (2019). Bayesian framework for fault variable identification. *Journal of Quality Technology*, *51*(4), 375–391. https://doi.org/10.1080/00224065.2018.1507561
Ullah, I., Pawley, M. D. M., Smith, A. N. H., & Jones, B. (2017). Improving the detection of unusual observations in high-dimensional settings. *Australian & New Zealand Journal of Statistics*, *59*(4), 449–462. https://doi.org/10.1111/anzs.12210
Wang, A., Xian, X., Tsung, F., & Liu, K. (2018). A spatial-adaptive sampling procedure for online monitoring of big data streams. *Journal of Quality Technology*, *50*(4), 329–343. https://doi.org/10.1080/00224065.2018.1507560
Wang, K., & Jiang, W. (2009). High-dimensional process monitoring and fault isolation via variable selection. *Journal of Quality Technology*, *41*(3), 247–258. https://doi.org/10.1080/00224065.2009.11917780
Wang, R. C., Baldea, M., & Edgar, T. F. (2017). Data visualization and visualization-based fault detection for chemical processes. *Processes*, *5*(3), 45. https://doi.org/10.3390/pr5030045
Wang, Y., Mei, Y., & Paynabar, K. (2018). Thresholded multivariate Principal component analysis for phase I multichannel profile monitoring. *Technometrics*, *60*(3), 360–372. https://doi.org/10.1080/00401706.2017.1375993
Wang, Z., Li, Y., & Zhou, X. (2017). A statistical control chart for monitoring high-dimensional poisson data streams. *Quality and Reliability Engineering International*, *33*(2), 307–321. https://doi.org/10.1002/qre.2005
Xiang, D., Qiu, P., Wang, D., & Li, W. (2022). Reliable post-signal fault diagnosis for correlated high-dimensional data streams. *Technometrics*, *64*(3), 323–334. https://doi.org/10.1080/00401706.2021.1979100
Yan, D., Zhang, S., & Jung, U. (2019). A variable-selection control chart via penalized likelihood and Gaussian mixture model for multimodal and high-dimensional processes. *Quality and Reliability Engineering International*, *35*(4), 1263–1275. https://doi.org/10.1002/qre.2458
Yan, H., Paynabar, K., & Shi, J. (2018). Real-time monitoring of high-dimensional functional data streams via spatio-temporal smooth sparse decomposition. *Technometrics*, *60*(2), 181–197. https://doi.org/10.1080/00401706.2017.1346522
Ye, H., Zheng, Z., Cheng, J.-R. C., Hable, B., & Liu, K. (2023). Online monitoring of high-dimensional asynchronous and heterogeneous data streams for shifts in location and scale. *International Journal of Production Research*, *62*(3), 720–736. https://doi.org/10.1080/00207543.2023.2172474
Yu, J. (2012a). Local and global principal component analysis for process monitoring. *Journal of Process Control*, *22*(7), 1358–1373. https://doi.org/10.1016/j.jprocont.2012.06.008
Yu, J. (2012b). A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes. *Chemical Engineering Science*, *68*(1), 506–519. https://doi.org/10.1016/j.ces.2011.10.011
Yu, J.-B. (2011). Bearing performance degradation assessment using locality preserving projections. *Expert Systems with Applications*, *38*(6), 7440–7450. https://doi.org/10.1016/j.eswa.2010.12.079
Yu, J., & Liu, X. (2022). One-dimensional residual convolutional auto-encoder for fault detection in complex industrial processes. *International Journal of Production Research*, *60*(18), 5655–5674. https://doi.org/10.1080/00207543.2021.1968061
Yu, J., & Zhang, C. (2020). Manifold regularized stacked autoencoders-based feature learning for fault detection in industrial processes. *Journal of Process Control*, *92*, 119–136. https://doi.org/10.1016/j.jprocont.2020.06.001
Zamba, K., & Hawkins, D. M. (2006). A multivariate change-point model for statistical process control. *Technometrics*, *48*(4), 539–549. https://doi.org/10.1198/004017006000000291
Zhang, C., Chen, N., & Wu, J. (2020). Spatial rank-based high-dimensional monitoring through random projection. *Journal of Quality Technology*, *52*(2), 111–127. https://doi.org/10.1080/00224065.2019.1571336
Zhang, C., Yan, H., Lee, S., & Shi, J. (2018a). Multiple profiles sensor-based monitoring and anomaly detection. *Journal of Quality Technology*, *50*(4), 344–362. https://doi.org/10.1080/00224065.2018.1508275
Zhang, C., Yan, H., Lee, S., & Shi, J. (2018b). Weakly correlated profile monitoring based on sparse multi-channel functional principal component analysis. *IISE Transactions*, *50*(10), 878–891. https://doi.org/10.1080/24725854.2018.1451012
Zhang, H., & Albin, S. (2009). Detecting outliers in complex profiles using a $\chi 2$ control chart method. *IIE Transactions*, *41*(4), 335–345. https://doi.org/10.1080/07408170802323000
Zhang, H., & Wang, H. (2021). A more powerful test of equality of high-dimensional two-sample means. *Computational Statistics & Data Analysis*, *164*, 107318. https://doi.org/10.1016/j.csda.2021.107318

Zhao, W., Wang, Z., & Wu, C. (2021). Adaptive multivariate EWMA charts for monitoring sparse mean shifts based on parameter optimization design. *Journal of Statistical Computation and Simulation*, *91*(13), 2670–2683. https://doi.org/10.1080/00949655.2021.1904242
Zou, C., Jiang, W., & Tsung, F. (2011). A lasso-based diagnostic framework for multivariate statistical process control. *Technometrics*, *53*(3), 297–309. https://doi.org/10.1198/TECH.2011.10034
Zou, C., Ning, X., & Tsung, F. (2012). LASSO-based multivariate linear profile monitoring. *Annals of Operations Research*, *192*(1), 3–19. https://doi.org/10.1007/s10479-010-0797-8
Zou, C., Tseng, S.-T., & Wang, Z. (2014). Outlier detection in general profiles using penalized regression method. *IIE Transactions*, *46*(2), 106–117. https://doi.org/10.1080/0740817X.2012.762486
Zou, C., Wang, Z., & Tsung, F. (2012). A spatial rank-based multivariate EWMA control chart. *Naval Research Logistics*, *59*(2), 91–110. https://doi.org/10.1002/nav.21475
Zou, C., Wang, Z., Zi, X., & Jiang, W. (2015). An efficient online monitoring method for high-dimensional data streams. *Technometrics*, *57*(3), 374–387. https://doi.org/10.1080/00401706.2014.940089