Article

Kernel-Based Multivariate Nonparametric CUSUM Multi-Chart for Detection of Abrupt Changes

Lei Qiao and Bing Wang *

School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China; ql0130@lixin.edu.cn
* Correspondence: bingwang@lixin.edu.cn

Abstract: In many cases, it is difficult to obtain precise distributional information on multivariate sequences. Therefore, there is a need to propose nonparametric methods for monitoring multivariate sequences. This article discusses the multivariate change detection problem and utilizes the kernel function as the statistic to construct the nonparametric Multivariate Cumulative Sum multi-chart, under the assumption that there is prior information about the abrupt changes. Through theoretical and numerical analysis, we show that the proposed control chart is more effective compared to other existing control charts. The good monitoring effect of this method demonstrates a strong potential for application.

1. Introduction

Statistical Process Control (SPC) is an important problem in numerous applications. The development of science and the increase in data storage have led to an increasing need for Multivariate Statistical Process Control (MSPC) [1,2]. The purpose of change detection is to raise an alarm as soon as possible when a change occurs. In order to evaluate the detectability of the control chart, we usually use two kinds of average run lengths (*ARLs*): one is in-control (IC) *ARL*, denoted by ARL_0, and the other is out-of-control (OC) *ARL*, denoted by ARL_1. With the same ARL_0, a control chart with a smaller ARL_1 is better and can detect changes more effectively.

In the literature, many researchers focus on this area and propose various methods to detect multivariate variables. Multivariate Statistical Process Control techniques were established by Hotelling in his pioneering paper in 1947 [3], which monitored the mean vector of multiple quality variables following a multivariate normal distribution. Subsequently, the field has attracted many experts to conduct related research and many different multivariate control charts have been proposed. In the literature, multivariate control charts can be classified into two categories according to the presence or absence of the assumption of data distribution. The first category with a distribution assumption is parametric. These include the Multivariate Exponentially Weighted Moving Average control chart (MEWMA [4,5]), Multivariate Cumulative Sum control chart (MCUSUM [6,7]) and other control charts, such as [8–10], which use different strategies to construct control charts to detect multivariate variables. These charts typically require the distribution of the observed data to obey a parametric (usually normal) form. However, in most cases, we lack the distributional information, or we do not have enough samples to estimate the exact distribution. The second category is nonparametric without distribution assumption, i.e., distribution-free control charts. Unfortunately, unlike many univariate nonparametric charts, it is challenging to design a distribution-free MSPC scheme based on conventional construction [11]. There are many multivariate nonparametric control charts proposed.

Refs. [12,13] discussed multivariate process control in the context of multivariate data. Ref. [14] proposed general and multiplicative nonparametric ratio models for data envelopment analysis problems with interval data. Ref. [15] provided a variable sampling interval (VSI) Shewhart $\bar{X}$ chart. Ref. [2] proposed a multivariate sign EWMA chart without distribution assumption, which requires some in-control observations. Other nonparametric multivariate control charts such as [16–18] also tried to use different methods to detect changes without parametric assumption.

In this paper, we develop a nonparametric or distribution-free control chart for monitoring multivariate data. Our approach is different from the existing nonparametric SPC methods in the sense that we apply the kernel function to construct the CUSUM multi-chart, that is, we construct a CUSUM-type control chart that does not require the distributional information of the multivariate data. Meanwhile, the method proposed in this paper solves three challenges in multivariate change detection:

- The first challenge is as follows: It is known that the traditional CUSUM control chart needs both pre-change and post-change information [7]. However, in reality, for online detection, we usually do not know the post-change information, we only have some pre-change observations. In order to solve this challenge, we construct the nonparametric multivariate CUSUM multi-chart. The idea of the multi-chart approach originates from [19], which shows that the multi-chart approach can detect different types of changes.
- The second challenge is how to capture important features when constructing a control chart for multivariate data. It is known that for multivariate data, such as when the dimension is $p = 30$, which is a relatively large number, statistics like Hotelling's T^2 may contain errors and trigger false alarms in online change detection [3]. In order to overcome this challenge, we utilize the kernel function to capture information for the multivariate data.
- The last challenge is the amount of historical pre-change observations; we may not have too many historical pre-change observations, that is, the amount of historical pre-change observations required for the method proposed in this paper does not need to be very large, making this method promising for application. For nonparametric control charts, historical pre-change observations are required. However, if the number of observations needed is excessively large, the loss outweighs the gain, and in many application scenarios, it is not achievable [20].

From the above statements, we can conclude that the kernel-based multivariate nonparametric CUSUM multi-chart proposed in this paper can overcome the aforementioned three challenges. We also present the theoretical results and simulation results of the control chart proposed in this paper.

The remainder of this paper is organized as follows: In Section 2, we first present a brief review of the reproducing kernel Hilbert space (RKHS) and then introduce the kernel-based multivariate nonparametric CUSUM multi-chart proposed in this paper, along with some theoretical properties. Section 3 uses the simulated data to present that the proposed kernel-based multivariate nonparametric CUSUM multi-chart has a better performance compared with other existing control charts. The conclusion and recommendations are given in Section 4. The proofs of the theoretical results are listed in Appendix A.

2. Kernel-Based Multivariate Nonparametric CUSUM Multi-Chart

In this section, we propose the kernel-based CUSUM multi-chart for the multivariate observation sequences. To illustrate this model in detail, we first briefly review the reproducing kernel Hilbert space (RKHS). This involves providing the definition of the kernel function and the statistics used for change detection. With the kernel function, we define the nonparametric CUSUM multi-chart and obtain the theoretical properties to ensure the performance of this model.

It is known that for a traditional univariate CUSUM control chart, we need to know the pre-change and post-change distributions to define the statistics used in the control

chart. However, the post-change information is usually unknown in reality, as is the distribution information. So, in this section, the kernel-based CUSUM multi-chart can solve the above problem when both the post-change information and the distribution information are unknown.

2.1. Examples of Kernels

In this section, we construct the CUSUM multi-chart based on the kernel function. Here, we give some examples of kernels that are widely used, as follows:

- When $\mathcal{X} = \mathcal{R}^d$, $k^{lin}(x, y) = <x, y>_{\mathcal{R}^d}$ defines the linear kernel. When $d = 1$, the KCP (kernel change-point) then coincides with the algorithm proposed in [21].
- When $\mathcal{X} = \mathcal{R}^d$, $k_h^G(x, y) = \exp[-||x - y||^2/(2h^2)]$ defines the Gaussian kernel with bandwidth $h > 0$, which is used in the experiments in Section 3.
- When $\mathcal{X} = \mathcal{R}^d$, $k_h^L(x, y) = \exp[-||x - y||/h]$ defines the Laplace kernel with bandwidth $h > 0$, which is used in the experiments in Section 3.
- When $\mathcal{X} = \mathcal{R}^d$, $k_h^e(x, y) = \exp(<x, y>_{\mathcal{R}^d}/h)$ defines the exponential kernel with bandwidth $h > 0$. Note that, unlike Gaussian and Laplace kernels, the exponential kernel is not translation-invariant.

Classical kernels can be found in the books by [22], with the advantageous properties of the Gaussian kernel and Laplace kernel; we use these two kernels in the simulations and compare the results with other online detection methods. Kernel change-point methods can be used in offline monitoring problems and online monitoring problems. Here, we consider online change detection. So, we have pre-change information through historical pre-change observations, but the exact post-change information is usually unknown.

2.2. Kernel Change-Point Online Detection Method

Let $X_{-m+1}, X_{-m+2}, \dots, X_0, X_1, X_2, X_3, \dots, X_N$ be the multivariate observation sequence. And, $X_i, -m + 1 \leq i \leq N \in \mathcal{R}^p$ are all p-dimensional vectors. We assume that X_{-m+1}, X_{-m+2}, ..., X_0 are historical pre-change observations, and the change-point is defined as τ $(1 \leq \tau \leq N)$. The observation sequence changes from an in-control to an out-of-control state after the change-point τ. From the m historical pre-change observations, we can estimate the pre-change mean vector denoted as $\hat{\mu}_0 \in \mathcal{R}^p$. While the post-change mean vector, denoted by $\mu_1 \in \mathcal{R}^p$, usually is not known in online change detection, we can obtain the possible boundary of the post-change mean vector, denoted by D, based on empirical knowledge or prior information. With the possible boundary D, we can define the c reference post-change mean vectors $\mu_1^k, 1 \leq k \leq c$ to construct multiple control charts to detect different types of changes.

With the kernel function, estimated pre-change mean vector $\hat{\mu}_0$, and possible post-change mean vectors μ_1^k, where $1 \leq k \leq c$, we can construct the kernel-based CUSUM multi-chart. The kernel-based CUSUM multi-chart is defined as follows: For fixed $c \geq 2$,

$$T_{MC} = \min_{1 \leq k \leq c} T_k,$$

$$T_k = \min \left\{ 1 \leq n \leq N : \max_{1 \leq j \leq n} \sum_{i=j}^{n} \ln \frac{k(X_i, \mu_1^k)}{k(X_i, \hat{\mu}_0)} > d_n^k \right\},$$

$$(1)$$

where $d_n^k \in \mathcal{R}$ is called the control limit. If an abrupt change occurs, we warn as soon as the statistic exceeds the control limit. So, it is important to study the statistical properties of T_{MC}. For p-dimensional observations, possible changes may occur in each dimension. Suppose that in each dimension we give l pre-specified known reference values, then $c = l^p$. However, in reality, the changes usually only occur in several dimensions, and in the simulation, we consider the changes only occur in selected $[p/5]$ dimensions.

The traditional probability likelihood ratio function is replaced by the kernel ratio function in (1). Under a nonparametric assumption, the CUSUM-type control chart can not use the probability likelihood ratio function. Therefore, we aim to find an alternative

function to replace the probability function. The reasons why we consider the kernel function are listed as follows:

- The kernel function contains a lot of information, especially in high dimensional cases, and is commonly used in classification problems [23]. While change detection is a very special "classification" problem, in change detection, we need to classify the data as either in-control or out-of-control data. Similarly, can we apply the kernel function in change detection?
- Actually, the kernel function is widely used in the literature in online and offline change detection problems, as demonstrated in [24,25]. Can we naturally apply this function in a CUSUM-type control chart?

For the two reasons mentioned above, we apply the kernel function to replace the probability function and construct the CUSUM multi-chart to solve the problem that the shift is unknown before detection.

Similarly, for the kernel-based CUSUM multi-chart defined in this section, in order to evaluate the detection power of any kernel-based test T with $ARL_0(T) = \mathbf{E}_0(T) = \gamma$, we also propose an index called *KCPI* (Kernel Control Chart Performance Index), as in [19], which is as follows:

$$KCPI(T) = \exp\{-\int_D w(\mu_1)\frac{ARL_1(T) - ARL_1^*}{ARL_1^*}d\mu_1\}, \tag{2}$$

where

$$ARL_1(T) = \mathbf{E}_1(T).$$

The definition of *ARL* is the average value of run length T. Since the state is divided into in-control and out-of-control states, there are two *ARLs*. $\mathbf{E}_0$ means the mathematical expectation of the in-control state, and the in-control *ARL* is denoted by $ARL_0 = \mathbf{E}_0(T)$. Similarly, $\mathbf{E}_1$ stands for the mathematical expectation of the out-of-control state; the out-of-control *ARL* is denoted by $ARL_1 = \mathbf{E}_1(T)$. ARL_1^* is the lower bound and $w(\mu_1)$ is the weight of the unknown post-change mean vector μ_1, that is, we do not know the exact value of μ_1, but from engineering knowledge, we know the possible boundary of μ_1 denoted by D and the weight function denoted by $w(\mu_1)$. If we have no additional information about the unknown μ_1, the weight function is $w(\mu_1) = 1/\int_D d\mu_1$.

Since the purpose of change detection is to provide an alert as soon as possible in the case of an abrupt change, a smaller value of ARL_1 represents an earlier alarm. For comparison purposes, we usually need a benchmark, i.e., a common ARL_0 for different control charts. So, in general, in order to compare the performance of several control charts, we ensure they have a common ARL_0 and then compare the ARL_1 of the control charts. The smaller the value of ARL_1, the better the performance of the control chart, which is the essence of the definition of the *KCPI*. If $KCPI(T_{MC})$ is larger, the performance of the kernel-based CUSUM multi-chart is better. So, finding the optimal design of the kernel-based CUSUM multi-chart involves finding the maximum value of $KCPI(T_{MC})$. Under the multivariate observation cases, if finding the maximum value of $KCPI(T_{MC})$ is hard, we can choose stochastic reference values of post-change $\mu_1^k, 1 \leq k \leq c$, as long as the corresponding *KCPI* is large.

2.3. Theoretical Properties of Kernel-Based CUSUM Multi-Chart

In this section, we discuss the theoretical properties of the proposed kernel-based CUSUM multi-chart to ensure the performance of change detection.

Similarly, as in Theorem 3 of [26,27], for a large γ, we can obtain the lower bound $ARL_1^* = [\mathbf{E}_1(T)]^*$ with $ARL_0(T) = \mathbf{E}_0(T) = \gamma$, as follows:

$$[\mathbf{E}_1(T)]^* \sim \frac{\ln\gamma}{KI(\mu_1, \hat{\mu}_0)}, \tag{3}$$

where

$$KI(\mu_1, \hat{\mu}_0) = \mathbf{E}_1\left[\ln \frac{k(X, \mu_1)}{k(X, \hat{\mu}_0)}\right]. \tag{4}$$

For a large N, γ, in order to obtain the asymptotic ARL_1 of the kernel-based CUSUM multi-chart, we first divide the region D into several disjointed subsets by using the reference post-change $\mu_1^k, 1 \leq k \leq c$, according to the kernel-based Kullback–Leibler information distance in (4).

Let

$$D_k = \left\{\mu_1 \in D : KI(\mu_1, \mu_1^k) \leq \min_{l \neq k} KI(\mu_1, \mu_1^l)\right\}$$

for $1 \leq k \leq c$. Thus, we have a disjointed division of the region D, with $D_k, 1 \leq k \leq c$, and then,

$$KCPI(T_{MC}) = \exp\{-\sum_{k=1}^c \int_{D_k} w(\mu_1)\frac{ARL_1(T_{MC}) - ARL_1^*}{ARL_1^*}d\mu_1\}. \tag{5}$$

The definition of the kernel-based inner Kullback–Leibler information distance is as follows:

$$\begin{aligned}
KI(\mu_1, \mu_1^k, \hat{\mu}_0) &= \mathbf{E}_1\left[\ln \frac{k(X, \mu_1^k)}{k(X, \hat{\mu}_0)}\right] \\
&= \mathbf{E}_1\left[\ln \frac{k(X, \mu_1)}{k(X, \hat{\mu}_0)}\right] - \mathbf{E}_1\left[\ln \frac{k(X, \mu_1)}{k(X, \mu_1^k)}\right] \\
&= KI(\mu_1, \hat{\mu}_0) - KI(\mu_1, \mu_1^k).
\end{aligned} \tag{6}$$

The last equation holds with the definition of $KI(\mu_1, \hat{\mu}_0)$ and $KI(\mu_1, \mu_1^k)$ in (4). Here, we assume that the reference $\mu_1^k, 1 \leq k \leq c$ satisfies the following conditions:

(1) $KI(\mu_1, \mu_1^k, \hat{\mu}_0) > 0$ if $\mu_1 \in D_k$ for $1 \leq k \leq c$.
(2) $KI(\mu_1, \hat{\mu}_0) \geq 0$ for $\mu_1 \in D$.

The first condition means that if $\mu_1 \in D_k$, then $KI(\mu_1, \hat{\mu}_0) > KI(\mu_1, \mu_1^k)$, which is consistent with the intuitive reality. We can interpret that the "distance" from μ_1 to μ^k is smaller than the distance from μ_1 to $\hat{\mu}_0$. The reference post-change μ_1^k should be designed to be closer to the real but unknown post-change μ_1. The second condition ensures that the kernel-based Kullback–Leibler information distance is non-negative, which imposes constraints on the selection of different kernel functions. So, in the simulation, we choose the Gaussian kernel and Laplace kernel; these two kernel functions all satisfy the above two conditions. In order to give a clear form of (5), we set

$$d_{lim}^k = \lim_{n \to \infty} d_n^k.$$

By using a similar method as in [28], we can prove the following theorems. Theorem 1 shows the asymptotic results regarding ARL_1 for a large N, γ.

Theorem 1. *Let $T_k, 1 \leq k \leq c$ have a common $ARL_0 = \mathbf{E}_0(T_k) = \gamma$. If $\mu_1 \in D_k$, then, a.s.,*

$$\begin{aligned}
\frac{T_k}{T_l} &\leq \frac{\max(0, KI(\mu_1, \mu_1^l, \hat{\mu}_0)}{KI(\mu_1, \mu_1^k, \hat{\mu}_0)}, \\
T_k &< T_l, \\
\frac{T_{MC}}{d_{lim}^k} &= \frac{T_k}{d_{lim}^k} + o(1) \to \frac{1}{KI(\mu_1, \mu_1^k, \hat{\mu}_0)}
\end{aligned} \tag{7}$$

for $\tau = 1, k \neq l$, and

$$\mathbf{E}_1\left[\frac{T_{MC}}{d_{lim}^k}\right] = \mathbf{E}_1\left[\frac{T_k}{d_{lim}^k}\right] + o(1) = \frac{1}{KI(\mu_1, \mu_1^k, \hat{\mu}_0)}(1 + o(1)) \tag{8}$$

for large N, γ.

By introducing the kernel-based Kullback–Leibler information distance in (4) and the kernel-based inner Kullback–Leibler information distance in (6), we can obtain the statistical properties of T_{MC}. For the kernel-based CUSUM multi-chart, in order to compare their ARL_1s, there are two common ARL_0s, γ' and γ, respectively, for T_k and T'_{MC}. The dynamic control limits $d_n'^k, 1 \leq k \leq c$ are for γ', and $d_n^k, 1 \leq k \leq c$ are for γ, and the inequality $d_{lim}'^k > d_{lim}^k$ holds. Writing $T'_k = T(d_n'^k)$ for short, then

$$\mathbf{E}_0\left[T'_k\right] = \gamma' > \mathbf{E}_0\left[T_k\right] = \gamma = \mathbf{E}_0(T'_{MC}) \tag{9}$$

for $1 \leq k \leq c$, where $T'_{MC} = \min_{1 \leq k \leq c}\{T'_k\}$.

So, from Theorem 1, we have

$$ARL_1(T'_{MC}) \to \frac{d_{lim}'^k}{KI(\mu_1, \mu_1^k, \hat{\mu}_0)} \quad a.s.$$

And, from (3),

$$ARL_1^* \to \frac{\ln \gamma}{KI(\mu_1, \hat{\mu}_0)} \quad a.s.$$

Then, $KCPI(T_{MC})$ can have a simple form for a large N and γ:

$$
\begin{aligned}
KCPI(T_{MC}) = KCPI(\mu_1^k, 1 \leq k \leq c) &= \exp\left\{-\int_D w(\mu_1)\frac{ARL_1(T) - ARL_1^*}{ARL_1^*}d\mu_1\right\} \\
&= \exp\left\{1 - \int_D w(\mu_1)\frac{ARL_1(T)}{ARL_1^*}d\mu_1\right\} \\
&= \exp\left\{1 - \sum_k \int_{D_k} w(\mu_1)\frac{KI(\mu_1, \hat{\mu}_0)}{KI(\mu_1, \hat{\mu}_0) - KI(\mu_1, \mu_1^k)}d\mu_1\right\}.
\end{aligned}
\tag{10}
$$

Then, we give a theorem about the relationship between the kernel-based CUSUM multi-chart and its constituent chart. Indeed, Theorem 2 shows that the performance of the kernel-based CUSUM multi-chart is better than that of its constituent charts in detecting an unknown abrupt change. That is to say, it makes sense to use the kernel-based CUSUM multi-chart; the kernel-based CUSUM multi-chart has better performance than a single kernel-based CUSUM chart in detecting unknown abrupt changes.

Theorem 2. *$p_l, 1 \leq l \leq c$ are positive numbers and satisfy*

$$\sum_{1 \leq l \leq c} p_l = 1.$$

If condition (7) holds for T'_{MC} and for $\mu_1 \in D_k$, there is μ_1^l, $k \neq l$ such that $KI(\mu_1, \mu_1^l, \hat{\mu}_0) < KI(\mu_1, \mu_1^k, \hat{\mu}_0)$, and then for every $\mu_1 \in D$,

$$\sum_{1 \leq l \leq c} p_l ARL_1(T_l) > ARL_1(T'_{MC})$$

for large N, γ.

The theorems above illustrate the beneficial properties of our proposed kernel-based CUSUM multi-chart.

3. Comparison and Analysis of Simulation Results

In this section, we present simulation results to demonstrate the performance of the proposed kernel-based CUSUM multi-chart. And, we choose the commonly used multivariate normal distribution, multivariate t distribution, and multivariate Gamma distribution as the simulation examples. For comparison, we use different control charts to compute the out-of-control average run length ARL_1, under the condition that the in-control average run length ARL_0 is the same. For detection problems, we expect to detect changes promptly as soon as changes occur. Hence, with the same in-control average run length (ARL_0), the smaller the value of out-of-control average run length (ARL_1), the better the performance of the detection.

In the simulation, in order to evaluate the performance of change detection for different shifts, the criterion is as follows:

$$
\begin{aligned}
KCPI(T) &= \exp\{-\int_D w(\mu_1)\frac{ARL_1(T) - ARL_1^*}{ARL_1^*}d\mu_1\} \\
&\propto -\int_D w(\mu_1)\frac{ARL_1(T) - ARL_1^*}{ARL_1^*}d\mu_1 \\
&\propto -\int_D \frac{ARL_1(T) - ARL_1^*}{ARL_1^*}d\mu_1 \quad \text{(if no prior information)} \\
&\propto -\int_D ARL_1(T)d\mu_1.
\end{aligned}
$$

We choose a number of possible values of μ_1 as a representation in the numerical simulations and compute the out-of-control ARL_1. In this view, we can also conclude that the performance of the control charts depends on the value of the out-of-control average run length (ARL_1); the smaller the value, the better the performance.

Remark 1. *In this section, we compare the performance with different control charts, and the performance is measured by the ARL. It is well known that the bandwidth has an impact on the kernel function estimation. The optimal bandwidth h in regression problems has been studied in [29,30]. However, in the monitoring problem discussed in this paper, the stopping rule is that if an abrupt change occurs, we issue a warning whenever the statistic exceeds the control limit. In the case of the same ARL_0, the control limit was also influenced by h, so if an abrupt change occurs, the bandwidth h in Gaussian and Laplace kernel functions has a joint effect on the statistic and the control limit, and their effects theoretically cancel each other out.*

In this section, we compare the Gaussian kernel-based CUSUM multi-chart, Laplace kernel-based CUSUM multi-chart, and other existing control charts. The performance (i.e., detectability) is measured by the ARL. Since the out-of-control state is unknown before change detection, we choose several representative out-of-control states in the simulation.

3.1. The Case for Multivariate Normal Distribution

In this section, we present simulation results to demonstrate the performance of the Gaussian kernel-based CUSUM multi-chart (KMulti-Gaussian), the Laplace kernel-based CUSUM multi-chart (KMulti-Laplace), the distribution-free multivariate goodness-of-fit chart (DFMGoF), the self-starting EWMAC (SSEWMAC) chart [31], the chart based on the change-point and generalized likelihood ratio test (short for ChangePt [32]), and the RTC chart [16]. In the simulation, we consider the Gaussian and Laplace kernel functions-based CUSUM multi-charts with DFMGoF, RTC, ChangePt, and SSEWMAC. We compare the values of ARL_1 under the case that $ARL_0 \approx 200$ and $m = 100$. The bandwidth in the Laplace and Gaussian kernels is set to $h = 1$.

Without loss of generality, the mean vector μ_0 is set to $\mathbf{0}$, and the covariance matrix $\Sigma_0 = I$ is chosen to be diagonal. We consider $p = 10, 30$, representing the low-dimensional

and high-dimensional cases, respectively. We compare the out-of-control performance of the six competing charts.

Similar to other MSPC studies, it is impossible to enumerate all the change patterns to allow a full-scale study of the charts' performance. Following similar studies in the literature [2,20,31,32], we consider scenarios as examples: shifts in the process mean vector in the first $[p/5]$ components of size δ, i.e., $\mu_1 = \mu_0 + \delta e$ with $e = (1, \cdots, 1, 0, \cdots, 0)^T$. We compare their performance in detecting mean shifts of magnitude $\delta = 0.25, 0.5, 1, 2, 4$, respectively.

The upper part of Table 1 represents the value of the ARL_1 for different control charts for $p = 10$ with the same value of $ARL_0 \approx 200$. As we have elaborated before, the smaller the value of ARL_1, the better the detection performance, given the same ARL_0. Thus, from Table 1, we can conclude that KMulti-Gaussian outperforms the other five charts when $p = 10$. The lower part of Table 1 represents the value of the ARL_1 for different control charts for $p = 30$ with the same value of $ARL_0 \approx 200$; the conclusion is entirely similar.

Table 1. The comparison of the ARL_1 in detecting location shifts when $m = 100$ for a multivariate normal distribution.

p	δ	KMulti-Gaussian	KMulti-Laplace	DFMGoF	RTC	ChangePt	SSEWMAC
	0.25	42.63	45.24	93.9	132	186	122
	0.5	15.77	16.97	36.1	51.8	134	70.3
10	1	5.32	5.67	12.4	12.1	36.4	11.6
	2	1.84	1.85	5.83	6.23	12.4	4.04
	4	1.01	1.01	3.93	4.72	4.81	1.59
	0.25	23.22	25.92	83.1	100	186	87.8
	0.5	7.71	8.33	20.6	24.7	148	22.9
30	1	2.46	2.36	8.4	7.62	52.9	9.09
	2	1.03	1.03	4.08	4.82	23.9	4.6
	4	1.01	1.01	2.74	3.9	12.2	2.37

In our simulation setting, we observed that the charts generally have a smaller ARL_1 when $p = 30$ than that when $p = 10$ given the same δ. This is because the detectability is largely determined by the Mahalanobis distance of the shifted mean vector from the IC mean vector, $\Delta = (\mu_1 - \mu_0)^T \Sigma_0^{-1} (\mu_1 - \mu_0)$ (see [31]). Given Σ_0 and the change pattern in our study, we have $\Delta_{p=30} > \Delta_{p=10}$ given the same δ. This can partially explain the better performance when $p = 30$.

Our proposed method has a priori information compared to other methods and extracts the information in the form of kernel functions, such that the simulation results are not unexpected.

3.2. The Case for Multivariate t Distribution

In this section, we consider the multivariate *t* distribution. We use $t_{p,\epsilon}$, which represents multivariate *t* distribution with ϵ degrees of freedom. Let pre-change observations belong to multivariate *t* distribution with 5 degrees of freedom, denoted by $t_{p,5}$. We also consider the scenario as follows: shifts in the process mean vector in the first $[p/5]$ components of size δ, i.e., $\mu_1 = \mu_0 + \delta e$ with $e = (1, \cdots, 1, 0, \cdots, 0)^T$. We compare their performance in detecting mean shifts of magnitude $\delta = 0.25, 0.5, 1, 2, 4$, respectively. Table 2 also compares our proposed methods with other existing methods.

The upper part of Table 2 represents the value of the ARL_1 for different control charts for $p = 10$ with the same value of $ARL_0 \approx 200$, and the lower part of Table 2 represents the value of the ARL_1 for different control charts for $p = 30$ with the same value of $ARL_0 \approx 200$. As we have elaborated before, the smaller the value of the ARL_1, the better the detection performance, given the same ARL_0. Thus, from Table 2, we can conclude that KMulti-Laplace outperforms the other five charts when $p = 10$ and $p = 30$.

Table 2. The comparison of the ARL_1 in detecting location shifts when $m_0 = 100$ for a multivariate t distribution.

p	δ	KMulti-Gaussian	KMulti-Laplace	DFMGoF	RTC	ChangePt	SSEWMAC
	0.25	60.75	51.60	140	151.6	188	162
	0.5	25.54	20.75	58.5	68.3	155	88.9
10	1	9.04	7.40	16.4	15.1	59.9	21.4
	2	3.58	2.89	7.37	6.74	19.9	8.1
	4	1.45	2.01	4.42	5.21	8.21	4.11
	0.25	63.83	43.05	127	136	184	140
	0.5	18.81	11.89	37.7	45.6	154	55.3
30	1	6.70	4.14	11.5	9.38	64.4	14.2
	2	2.76	1.55	5.46	5.77	31.5	6.43
	4	1.01	1.01	3.31	4.58	15.8	3.6

3.3. The Case for Multivariate Gamma Distribution

In this section, we consider the multivariate Gamma distribution. We use $Gamma_{p,\epsilon}$, which represents the multivariate Gamma distribution with the shape parameter ϵ and scale parameter of one. Let the pre-change observations belong to multivariate Gamma with the shape parameter of three and scale parameter of one, denoted by $Gamma_{p,3}$. We also consider the scenario as follows: shifts in the process mean vector in the first $[p/5]$ components of size δ, i.e., $\mu_1 = \mu_0 + \delta e$ with $e = (1, \cdots, 1, 0, \cdots, 0)^T$. We compare their performance in detecting mean shifts of magnitude $\delta = 0.25, 0.5, 1, 2, 4$, respectively.

The upper part of Table 3 represents the value of the ARL_1 for different control charts for $p = 10$ with the same value of $ARL_0 \approx 200$, and the lower part of Table 3 represents the value of the ARL_1 for different control charts for $p = 30$ with the same value of $ARL_0 \approx 200$. As we have elaborated before, the smaller the value of the ARL_1, the better the detection performance, given the same ARL_0.

Table 3. The comparison of the ARL_1 in detecting location shifts when $m_0 = 100$ for a multivariate Gamma distribution.

p	δ	KMulti-Gaussian	KMulti-Laplace	DFMGoF	RTC	ChangePt	SSEWMAC
	0.25	72.73	73.94	88.3	147	188	137
	0.5	35.76	35.03	26.8	59.3	152	67.2
10	1	14.41	14.33	13.2	10.5	52.9	17.4
	2	5.69	5.62	7.27	6.43	17.2	7.39
	4	2.30	2.57	4.57	5.05	6.87	3.4
	0.25	57.78	64.62	43.4	126	183	109
	0.5	19.84	24.56	17.2	28.4	159	35.7
30	1	7.26	7.04	9.2	7.43	62.5	11.9
	2	2.62	2.35	5.16	5.25	27.8	5.72
	4	1.14	1.09	3.3	4.13	14.7	2.85

From Table 3, we can see that DFMGoF works better in some specific situations ($p = 30$ and $\delta = 0.25, 0.5$), but note that from all the drift cases combined, KMulti-Gaussian is still better. We assume that all drift cases are equally likely. For example, the ARL_1 of DFMGoF is 14.38 less than the ARL_1 of KMulti-Gaussian when $\delta = 0.25$ and $p = 30$, but we set a weight of $0.25/4$ when measuring performance in aggregate. Combining all the out-of-control states considered, KMulti-Gaussian and KMulti-Laplace are better than the other control charts.

From the three tables of the simulation results, we can conclude that the performance of the kernel-based multivariate nonparametric CUSUM multi-chart (KMulti-Gaussian and

KMulti-Laplace) is better than the other four existing control charts, as the ARL_1 is smaller when the ARL_0 is the same.

Remark 2. *In this section, we compared the performance of different multivariate control charts. The performance of the methods proposed in this paper may not always be the best because the KCPI in (5) is an integrated form for all possible post-change mean vectors μ_1. Combining all the out-of-control states considered, the value of the ARL_1 for the method proposed in this paper is better compared with other existing control charts.*

Remark 3. *The average run length (ARL) is a very important measurement in change detection. The in-control average run length, denoted by ARL_0, is, in the case of the observations, a pre-change observation. The out-of-control average run length, denoted by ARL_1, is, in the case of the observations, a post-change observation. The purpose of change detection is to alarm as quickly as possible if an abrupt change occurs. Hence, the smaller the value of the ARL_1, the better the performance of detection when the ARL_0 is the same.*

4. Concluding Remarks

4.1. The Novelty of the Kernel-Based Multivariate CUSUM Multi-Chart

Traditional CUSUM multi-chart [19,28] is designed for one-dimensional data with a known distribution. Generalizing to multivariate data is not as straightforward. On the one hand, it requires some computation. On the other hand, we would like to design a multivariate CUSUM multi-chart that does not depend on the distribution. This paper proposes a kernel-based multivariate nonparametric CUSUM multi-chart to solve the following three challenges:

- In reality, we usually lack post-change information in online detection.
- Effectively capturing important features is essential when the dimension is $p = 30$, a slightly larger number. The statistics, such as Hotelling's T^2, may obtain the wrong alarm in online detection.
- In some cases, the amount of historical pre-change observations is not large.

4.2. Future Research Perspectives

We initially attempted to construct a control chart using the kernel function, which in turn was found to have good theoretical properties and simulation performance. We defined the kernel function-based CUSUM multi-chart and *KCPI*, which further gives the theoretical properties. In the simulation, we compared our proposed method with other existing methods and showed the advantages of our proposed method in terms of the average run length, a crucial index.

At the same time, it has some shortcomings. We assumed that multivariate data have prior knowledge. As the number of dimensions increases, acquiring prior knowledge becomes more and more difficult. For example, in the actual production process, there are many steps involved in production. We can monitor the sensor data from these steps as multivariate data, but acquiring more precise prior knowledge may necessitate a substantial investment of time and money.

We can conduct further research on the correlation between multivariate data internally and investigate the relationship between multidimensional variables when constructing control charts. This is a valuable research area to consider in the future. We can also construct parallel control charts using different kernel functions to monitor changes. In the simulation results above, the Gaussian kernel function and Laplace kernel function each have their own advantages and disadvantages. As to which kernel function works better, it depends on the data and how they change. In some data, this kernel function works well. In other data, another kernel function works well. The subsequent research intends to combine different kernel functions to construct parallel control charts.

Abbreviations

The following abbreviations are used in this manuscript:

CUSUM	Cumulative Sum
SPC	Statistical Process Control
MSPC	Multivariate Statistical Process Control
IC	in-control
OC	out-of-control
ARL	average run length
ARL_0	in-control average run length
ARL_1	out-of-control average run length
MEWMA	Multivariate Exponentially Weighted Moving Average
MCUSUM	Multivariate Cumulative Sum
VSI	variable sampling interval
RKHS	reproducing kernel Hilbert space
KCP	kernel change-point
KCPI	Kernel Control Chart Performance Index
KI	Kullback–Leibler information distance
KMutl-Gaussian	Gaussian kernel-based CUSUM multi-chart
KMutl-Laplace	Laplace kernel-based CUSUM multi-chart
DFMGoF	distribution-free multivariate goodness-of-fit chart
RTC	Real-Time Constrasts
ChangePt	change-point and generalized likelihood ratio test
SSEWMAC	self-starting EWMA chart

Appendix A

Proof of Theorem 1. Based on the proofs provided in [19,33], we can easily prove Theorem 1. The only difference is that the measure function is a kernel function. □

Proof of Theorem 2. Based on the proofs provided in [19,33], we can easily prove Theorem 2. The only difference is that the measure function is a kernel function. □

References

1. Woodall, W.H.; Montgomery, D.C. Some Current Directions in the Theory and Application of Statistical Process Monitoring. *J. Qual. Technol.* **2014**, *46*, 78–94. [CrossRef]
2. Zou, C.; Tsung, F. A Multivariate Sign EWMA Control Chart. *Technometrics* **2011**, *53*, 84–97. [CrossRef]
3. Hotelling, H. Multivariate quality control—Illustrated by the air testing of sample bombsights. In *Techniques of Statistical Analysis*; Eisenhart, C., Hastay, M.W., Wallis, W.A., Eds.; McGraw-Hill: New York, NY, USA, 1947; pp. 111–184.
4. Lowry, C.A.; Woodall, W.H.; Champ, C.W.; Rigdon, S.E. Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics* **1992**, *34*, 46–53. [CrossRef]
5. Runger, G.C.; Prabhu, S.S. A Markov Chain Model for the Multivariate Exponentially Weighted Moving Averages Control Chart. *J. Am. Stat. Assoc.* **1996**, *91*, 1701–1706. [CrossRef]
6. Crosier, R.B. Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics* **1988**, *30*, 243–251. [CrossRef]
7. Woodall, W.H.; Ncube, M.M. Multivariate CUSUM Quality Control Procedures. *Technometrics* **1985**, *27*, 285–292. [CrossRef]

8. Hawkins, D.M. Multivariate Quality Control Based on Regression-Adjusted Variables. *Technometrics* **1991**, *33*, 61–75. [CrossRef]
9. Wang, K.; Jiang, W. High-Dimensional Process Monitoring and Fault Isolation via Variable Selection. *J. Qual. Technol.* **2009**, *41*, 247–258. [CrossRef]
10. Zou, C.; Qiu, P. Multivariate Statistical Process Control Using LASSO. *J. Am. Stat. Assoc.* **2009**, *40*, 1586–1596. [CrossRef]
11. Zhang, C.; Chen, N.; Zou, C. Robust Multivariate Control Chart Based on Goodness-of-fit Test. *J. Qual. Technol.* **2016**, *48*, 139–161. [CrossRef]
12. Alt, F. Multivariate quality control: State of the art. In *ASQC Quality Congress Transactions*; American Society for Quality Control: Milwaukee, WI, USA, 1982; pp. 886–893.
13. Alt, F.; Smith, N. Multivariate process control. *Handb. Stat.* **1988**, *7*, 333–351.
14. Emrouznejad, A.; RostamyMalkhalifeh, M.; HatamiMarbini, A.; Tavana, M. General and multiplicative non-parametric corporate performance models with interval ratio data. *Appl. Math. Model.* **2012**, *36*, 5506–5514. [CrossRef]
15. Reynolds, M.R., Jr.; Amin, R.W.; Arnold, J.C.; Nachlas, J.A. X charts with variable sampling intervals. *Technometrics* **1988**, *30*, 181–192. [CrossRef]
16. Deng, H.; Runger, G.; Tuv, E. System Monitoring with Real-Time Contrasts. *J. Qual. Technol.* **2012**, *44*, 9–27 . [CrossRef]
17. Hwang, W.; Runger, G.; Tuv, E. Multivariate Statistical Process Control with Artificial Contrasts. *IIE Trans.* **2007**, *39*, 659–669. [CrossRef]
18. Holland, M.D.; Hawkins, D. A Control Chart Based on A Nonparametric Multivariate Change-Point Model. *J. Qual. Technol.* **2014**, *46*, 63–77. [CrossRef]
19. Han D.; Tsung, F. Detection and diagnosis of unknown abrupt changes using CUSUM multi-chart schemes. *Sequantial Anal.* **2007**, *26*, 225–249. [CrossRef]
20. Zou, C.; Wang, Z.; Tsung, F. A Spatial Rank-Based Multivariate EWMA Control Chart. *Nav. Res. Logist. (NRL)* **2012**, *59*, 91–110. [CrossRef]
21. Lebarbier, E. Detecting multiple change-points in the mean of a Gaussian process by model selection. *Signal Process.* **2005**, *85*, 717–736. [CrossRef]
22. Scholkopf, B.; Smola, A. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2001.
23. Evers, L.; Messow, C.-M. Sparse kernel methods for high-dimensional survival data. *Bioinformatics* **2008**, *24*, 1632–1638. [CrossRef] [PubMed]
24. Arlot, S.; Celisse, A.; Harchaoui, Z. A kernel multiple change-point algorithm via model selection. *J. Mach. Learn. Res.* **2019**, *162*, 1–56.
25. Li, S.; Xie, Y.; Dai, H.; Song, L. Scan b -statistic for kernel change-point detection. *Seq. Anal.* **2019**, *38*, 503–544. [CrossRef]
26. Lorden, G. Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **1971**, *42*, 520–527. [CrossRef]
27. Siegmund, D. *Sequential Analysis: Tests and Confidence Intervals*; Springer: New York, NY, USA, 1985.
28. Han, D.; Tsung, F.; Hu, X.; Wang, K. CUSUM and EWMA multi-charts for detection a range of mean shifts. *Stat. Sin.* **2007**, *17*, 1139–1164.
29. Amini, M.; Roozbeh, M. Optimal partial ridge estimation in restricted semiparametric regression models. *J. Multivar. Anal.* **2015**, *136*, 26–40. [CrossRef]
30. Roozbeh, M. Optimal QR-based estimation in partially linear regression models with correlated errors using GCV criterion. *Comput. Stat. Data Anal.* **2018**, *117*, 45–61. [CrossRef]
31. Maboudou-Tchao, E.M.; Hawkins, D.M. Self-Starting Multivariate Control Charts for Location and Scale. *J. Qual. Technol.* **2011**, *43*, 113–126. [CrossRef]
32. Zamba, K.; Hawkins, D.M. A Multivariate Change-Point Model for Change in Mean Vector and/or Covariance Structure. *J. Qual. Technol.* **2009**, *41*, 285–303. [CrossRef]
33. Qiao, L.; Han, D. CUSUM multi-chart for detecting unknown abrupt changes under finite measure space for network observation sequences. *Statistics* **2021**, *55*, 489–513. [CrossRef]