Environmental and Ecological Statistics (2019) 26:217–238
https://doi.org/10.1007/s10651-019-00425-4

To reduce or not to reduce: a study on spatio-temporal surveillance

Junzhuo Chen[1] · Chuljin Park[2] · Seong-Hee Kim[1] · Yao Xie[1]

Received: 5 February 2019 / Revised: 22 August 2019 / Published online: 13 September 2019

Abstract

The majority of control charts based on scan statistics for spatio-temporal surveillance use full observation vectors. In high-dimensional applications, dimension-reduction techniques are usually applied. Typically, the dimension reduction is conducted as a post-processing step rather than in the data acquisition stage and thus, a full sample covariance matrix is required. When the dimensionality of data is high, (i) the sample covariance matrix tends to be ill-conditioned due to a limited number of samples; (ii) the inversion of such a sample covariance matrix causes numerical issues; and (iii) aggregating information from all variables may lead to high communication costs in sensor networks. In this paper, we propose a set of reduced-dimension (RD) control charts that perform dimension reduction during the data acquisition process by spatial scanning. The proposed methods avoid computational difficulties and possibly high communication costs. We derive a theoretical measure that characterizes the performance difference between the RD approach and the full observation approach. The numerical results show that the RD approach has little performance loss under several commonly used spatial models while enjoying all the benefits of implementation. A case study on water quality monitoring demonstrates the effectiveness of the proposed methods in real applications.

Keywords Reliability · Scan statistics · Spatio-temporal surveillance · Statistical process control (SPC) · Statistical computing

Handling Editor: Pierre Dutilleul.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10651-019-00425-4) contains supplementary material, which is available to authorized users.

✉ Seong-Hee Kim
skim@isye.gatech.edu

[1] H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

[2] Department of Industrial Engineering, Hanyang University, Seoul, South Korea

1 Introduction

Spatio-temporal surveillance has numerous applications, including disease outbreak detection (Lee et al. 2014), river water quality monitoring (Bartram and Ballance 1996), and computer network intrusion detection (Park et al. 2014). The objective of spatio-temporal surveillance is to quickly detect changes that occur in a monitored area based on data streams collected from different locations. Typically, the data streams are observations of a particular quality index that can be modeled as a discrete (e.g., network intrusion counts and mortality) or continuous (e.g., disease incidence rate and contaminant concentration) random variable. Multivariate control charts are commonly used for spatio-temporal surveillance. A change in the monitored area is detected when the monitoring statistic exceeds a control limit that is pre-specified to control the false-alarm rate. We use the in-control average run length (ARL_0) to characterize the false-alarm-rate of a control chart. In order to measure the detection performance, we use the out-of-control average run length (ARL_1) which represents the expected number of samples needed to raise an alarm. Thus, the goal is to achieve a short detection delay (ARL_1) with a targeted ARL_0.

Commonly used multivariate control charts include the T^2 chart (Hotelling 1947), multivariate exponentially weighted moving average (MEWMA) chart (Lowry et al. 1992), and multivariate cumulative sum (MCUSUM) chart. The univariate CUSUM chart (Page 1954) is widely adopted due to its efficiency in detecting small shifts and its recursive implementation that facilitates online monitoring. In the multivariate setting, MCUSUM methods can be classified as those based on the likelihood ratio (LR) statistic and the Hotelling's T^2 statistic. Healy (1987) uses an LR statistic assuming that the vector observations are i.i.d. multivariate normal. Crosier (1988) computes a T^2 statistic for each vector observation and then forms a CUSUM chart based on a sequence of T^2 statistics. A key difference between the LR-based and the T^2-based MCUSUM statistics is that the former assumes known shift direction, while the latter does not make such an assumption.

One challenge in spatio-temporal surveillance is the high dimensionality of the data. For example, Spiegelhalter et al. (2012) consider a problem where 200,000 indicators of excess mortality are monitored simultaneously in the UK health system. In addition, spatial correlation among different locations usually exists and holds valuable information for change detection. In order to tackle these issues, Jiang et al. (2011) propose a spatial-scanning based MCUSUM chart, which constructs an LR statistic for each spatial cluster and scans through all possible clusters to detect a potential disease outbreak. Lee et al. (2014) extend the method by using an analytical formula proposed by Kim et al. (2007) for control limit approximation. Lee et al. (2015) further extend previous work to more general distributions. Other examples of high-dimensional spatio-temporal surveillance include astronomical imaging, where high-resolution video streams are monitored for solar flare detection, as studied in Xie et al. (2013) and Liu et al. (2015). In this problem, each observation is a 67,744-dimensional vector consisting of image pixels and the goal is to detect the emergence of a sparse signal. Sequential detection of a sparse change is an active area of research, see, for example, the work by Xie and Siegmund (2013) and Liu et al. (2016). However, these works usually assume independent data streams without considering spatial correlation.

All the methods reviewed above employ full observation vectors when constructing monitoring statistics. Even for charts with the spatial scanning approach (Jiang et al. 2011; Lee et al. 2014 and Lee et al. 2015), a full-size covariance matrix is involved in computing the detection statistic. For high-dimensional problems in the presence of spatial correlation, several difficulties exist when applying these control charts that require a full-size covariance matrix: (i) the sample covariance matrix tends to be ill-conditioned due to the number of samples being relatively small compared to the dimension of the covariance matrix; (ii) the ill-conditioned sample covariance matrix causes numerical issues due to the matrix inversion involved in the computation of the statistics; and (iii) the communication cost can be high for distributed sensor networks, because all sensors need to exchange information with each other (Guerriero et al. 2009).

In order to tackle the difficulties caused by high-dimensionality, a viable solution is to perform a dimensionality reduction technique. The existing dimension reduction techniques include the principal component analysis (PCA) (Mishin et al. 2014); random linear projections such as Runger (1996), Bodnar and Schmid (2005) and Skubałlska-Rafajlowicz (2013); and wavelet transform methods such as Lee et al. (2012) and Wang et al. (2015a). These methods still require full observation vectors because dimension-reduction is performed as a *post-processing* step rather than *in the data acquisition stage*.

Another method to tackle the high-dimensionality problem is to perform *reduced-dimension* spatial scanning. In this method, the entire monitored area is broken into overlapping local clusters of certain radii and only a subset of sensors or locations is used within the clusters, as in Fig. 1. Then, a control-chart is constructed for each local cluster and a change is detected whenever any of the local clusters fires an alarm. In doing this, each control chart only monitors a small number of data streams that fall within the scanning cluster. In this paper, we consider the reduced-dimension spatial scanning based on data streams *within the local clusters* only, rather than the full observation vectors as done in earlier works such as Jiang et al. (2011). Hence, our method never needs to acquire full-dimension observations, and dimension reduction is performed *during data acquisition*. Our method is suitable for distributed processing required by sensor networks. A recent related work by Xie et al. (2015) based on linear projection may also be used for spatial scanning; however, spatial correlation is not considered in their study. Moreover, due to the recursive nature of the CUSUM statistics, our method is suitable for *in situ* processing, which implies that raw data do not need to be stored, and this is again preferable in sensor networks.

Although the idea of the reduced-dimension spatio-temporal scanning is not new and, in fact, is often used in practice, one important question has not yet been answered in the literature: *How much do we lose by using reduced-dimension observations in the presence of spatial correlation?* We provide a precise answer to the amount of loss by characterizing the difference of reduced-dimension charts and full-dimension charts in terms of their ARL_1 under a fixed ARL_0. Our analysis shows that the RD approach usually utilizes 0–20% more observations in ARL_1 than the full observation approach for a reasonable range of spatial correlation among neighboring locations. Even when we lose over 20%, the absolute differences in ARL_1 are as small as $1-5$ observations in most cases. In addition, we show that it is even possible that the RD approach may

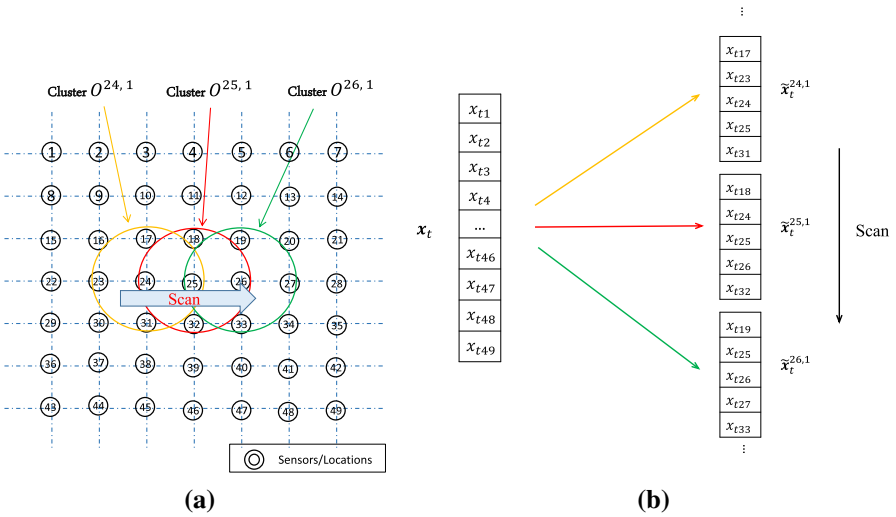

Fig. 1 **a** A monitored area with $p = 49$ locations or sensors and illustration of the spatial scanning using a circular shaped region; **b** Mapping of a full-dimension observation vector into reduced-dimensional vectors corresponding to scanning regions

perform better than the full observation approach (i) when T^2 based charts for unknown shift directions are used or (ii) when certain groups are completely independent of other groups, as in a water quality monitoring example presented in Sect. 6. This is a blessing, since an anomaly usually affects a local region and spatial correlation tends to decay with distance. Thus, restricting to local sensors when forming a monitoring statistic should not significantly degrade the detection performance.

The remainder of the paper is organized as follows. In Sect. 2, we formulate the problem and review existing MCUSUM charts that employ the concept of spatial scanning. In Sect. 3, we present reduced-dimension MCUSUM charts. In Sect. 4, we present theoretical analysis for the performance comparison of the reduced-dimension and full-dimension charts. In Sect. 5, we present the numerical results. In Sect. 6, we present an application in water quality monitoring, and in Sect. 7 we present the concluding remarks .

2 Background

In this section, we define the notations and our problem. Then, a few spatial scanning control charts are presented as representative charts, which adopt the full observation approach.

2.1 Notation and Problem

In spatio-temporal surveillance, observations (e.g., a quality index) are sequentially collected from different locations in the monitored area. Using sequential observations, one desires to detect a possible change in the monitored area as soon as possible.

Suppose there are p locations (sensors). For simplicity, we assume that the monitored area is rectangular and the locations (sensors) sit at a lattice of $p = MN$ points. This rectangular assumption on the shape of the monitored area can be relaxed, as we show in Sect. 6.

Let $q = (m, n)$, where $m = 1, 2, \ldots, M$ and $n = 1, 2, \ldots, N$ denote the two-dimensional spatial coordinate of a location. The location is indexed by c, where $c = (m-1)N + n$. Let $P = \{1, 2, \ldots, p\}$ be the set of monitoring locations. At time t, the observation is a p-dimensional vector $\boldsymbol{x}_t = [x_{t1}, x_{t2}, \ldots, x_{tp}]'$. Assume that different observation vectors are temporally independent but spatially correlated with covariance matrix $\boldsymbol{\Sigma}$. We further assume that the covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{x}_t$ is known or can be estimated from data. The change only affects the mean, and the covariance matrix remains the same. Under the hypothesis of no change, the observations $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ are i.i.d. normally distributed with mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}$. Alternatively, there exists a change-point κ in time, and a subset of neighboring locations are affected by a change that occurs at the change-point. For the locations affected by the change, the means of their observations are shifted, while observations from the unaffected locations remain the same distribution. This corresponds to a shift in the mean vector from $\boldsymbol{\mu}_0$ to another vector $\boldsymbol{\mu}_1$. Without loss of generality, assume that the observation vectors are standardized so that $\boldsymbol{\mu}_0 = \mathbf{0}$ and $[\boldsymbol{\Sigma}]_{i,i} = 1, \forall i = 1, \ldots, p$, where $[\cdot]_{i,j}$ denotes the (i, j)th element of a matrix.

Due to spatial correlation, an anomaly often affects a cluster of neighboring locations. We assume that the shift cluster is circular to facilitate the notation (such an assumption can be relaxed as demonstrated in Sect. 6). A cluster is a set of locations $O^{c,r} = \{j \,|\, \|q_j - q_c\| \le r,\ j \in P\}$, where c is the center of the cluster, r is the radius of the cluster and $\|\cdot\|$ denotes the ℓ_2-norm of a vector. In our setting (sensors are placed over a grid), the radius r is usually selected from a discrete set of values. Let $R \subseteq \{1,\ \sqrt{2},\ 2,\ 2\sqrt{2}, \cdots\}$ be the set of possible values of r. Define a p-dimensional vector $[\boldsymbol{\mu}_{c,r}]_j = \delta_j$ for all $j \in O^{c,r}$, and 0 otherwise. Here, $[\cdot]_j$ denotes the jth element of a vector and δ_j denotes the shift magnitude of the jth location. Hence, if an anomaly affects the cluster $O^{c,r}$, then $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_{c,r}$.

Our objective is to detect anomalies that affect a cluster $O^{c,r}$ ($c \in P$ and $r \in R$) by testing whether the mean of the observations has shifted from a nominal vector $\boldsymbol{\mu}_0$ to a different vector $\boldsymbol{\mu}_1$.

2.2 The LR-F-MCUSUM chart

Jiang et al. (2011) consider a spatial scanning control chart based on LR statistics. The method is based on full observation vectors and requires a full covariance matrix. Spatial scanning is achieved by zeroing out the portion of the mean vector that falls out of the scanning region. We refer to their method as the LR-F-MCUSUM chart hereafter. For a hypothetical shift cluster with center c and radius r, the sequence of monitoring statistics in the LR-F-MCUSUM chart is given by

$$S_t^{c,r} = \max\left(0,\ \max_{1 \le \tau \le t} \sum_{i=\tau}^{t} \ell_i^{c,r}\right), \quad t = 1, 2, \ldots, \tag{1}$$

where $\ell_i^{c,r}$ is the log LR statistic when the post-change mean vector is $\boldsymbol{\mu}_{c,r}$:

$$\ell_i^{c,r} = \boldsymbol{\mu}_{c,r}^{'} \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{x}_i - \frac{\boldsymbol{\mu}_{c,r}}{2} \right), \quad i = 1, 2, \ldots, t. \tag{2}$$

When constructing the monitoring statistic in (1), we need to search over the unknown change-point by maximizing with respect to τ, i.e., the variable that represents a putative change-point. Note that the statistic for each cluster $S_t^{c,r}$ in (1) can be computed recursively

$$S_t^{c,r} = \max\{0, S_{t-1}^{c,r} + \ell_t^{c,r}\}, \quad t = 1, 2, \ldots, \quad \text{and} \quad S_0^{c,r} = 0. \tag{3}$$

If the true shift center and radius (c, r) are known, we detect a change as soon as $S_t^{c,r}$ exceeds a pre-specified control limit.

In practice, usually neither the shift center c nor the radius r is known a priori. In this case, one has to scan over all possible values of c and r, calculate the corresponding MCUSUM statistic, and form a global detection statistic by taking the maximum: $S_t^{**} = \max_{c \in P, r \in R} S_t^{c,r}$ for $t = 1, 2, \ldots$. A change is detected whenever S_t^{**} exceeds a pre-specified control-limit $h_\ell^{**} > 0$, which is specified according to the requirement for the ARL_0.

2.3 The T^2-F-MCUSUM chart

When there is no prior information on the magnitude and direction of the mean shift, an MCUSUM chart based on T^2-statistic is more appropriate. In order to make a reasonable comparison with reduced dimension charts, we introduce a T^2-F-MCUSUM chart based on full-dimension observation vectors. It performs spatial scanning while using the full covariance matrix, as analogous to Jiang et al. (2011). Given a cluster $O^{c,r}$, we modify the full observation vector $\boldsymbol{x}_t$ by replacing all elements that are not in the cluster with zeros $[\boldsymbol{x}_t^{c,r}]_j = [\boldsymbol{x}_t]_j$ for $j \in O^{c,r}$, and 0 otherwise.

For each modified observation $\boldsymbol{x}_t^{c,r}$, we compute a T^2 statistic

$$a_t^{c,r} = \boldsymbol{x}_t^{c,r\,'} \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_t^{c,r} - \mu_T - k\sigma_T, \tag{4}$$

where k is a positive real-valued constant, $\mu_T = \text{E}[\boldsymbol{x}_t^{c,r\,'} \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_t^{c,r}]$ and $\sigma_T^2 = \text{Var}[\boldsymbol{x}_t^{c,r\,'} \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_t^{c,r}]$, which are the *in-control* mean and standard deviation of the T^2 statistic, respectively. The calculations for μ_T and σ_T^2 are discussed in Sect. 4. The monitoring statistic $T_t^{c,r}$ is computed recursively over time

$$T_t^{c,r} = \max\{0, T_{t-1}^{c,r} + a_t^{c,r}\}, \quad t = 1, 2, \ldots, \quad \text{and} \quad T_0^{c,r} = 0. \tag{5}$$

With an unknown shift center and radius, we again search over all possible clusters and sizes to construct the global detection statistic $T_t^{**} = \max_{c \in P, r \in R} T_t^{c,r}$ for $t = 1, 2, \ldots$ and detection is performed by comparing T_t^{**} with a control limit h_a^{**}.

3 RD-MCUSUM charts

In this section, we present the reduced-dimension (RD) approach. For each scan cluster, a control chart is constructed for RD observation vectors and only utilizes local covariance. We develop two versions of MCUSUM charts, based on the LR statistic and the T^2 statistic, respectively.

3.1 The LR-RD-MCUSUM chart

We begin by considering an LR-based chart, which is referred as the LR-RD-MCUSUM chart hereafter. For each scan cluster $O^{c,r}$, we truncate the original data vector $\boldsymbol{x}_t$ into a lower dimensional vector $\tilde{\boldsymbol{x}}_t^{c,r}$, where $[\boldsymbol{x}_t]_j$ is positioned in $\tilde{\boldsymbol{x}}_t^{c,r}$ if $j \in O^{c,r}$, and is eliminated otherwise, as illustrated in Fig. 1b. The monitoring statistic for that particular cluster is computed over vectors of dimension $|O^{c,r}|$, where $|\cdot|$ denotes the cardinality of a set. At each time, the LR statistic for $O^{c,r}$ is computed in the following manner

$$\tilde{\ell}_i^{c,r} = \tilde{\boldsymbol{\mu}}_{c,r}^{'} \boldsymbol{\Sigma}_{c,r}^{-1} \left(\tilde{\boldsymbol{x}}_i^{c,r} - \frac{\tilde{\boldsymbol{\mu}}_{c,r}}{2} \right), \quad i = 1, 2, \ldots, t. \tag{6}$$

Here, $\tilde{\boldsymbol{\mu}}_{c,r}$ and $\boldsymbol{\Sigma}_{c,r}$ are the sub-vector and sub-matrix of $\boldsymbol{\mu}_{c,r}$ and $\boldsymbol{\Sigma}$, respectively. Then, based on (6), the detection statistic $\tilde{S}_t^{c,r}$ is computed recursively for each cluster, similar to (3). Finally, a global monitoring statistic is formed by taking the maximum over all clusters and sizes

$$\tilde{S}_t^{**} = \max_{c \in P, r \in R} \tilde{S}_t^{c,r}, \quad t = 1, 2, \ldots. \tag{7}$$

An alarm is signaled whenever $\tilde{S}_t^{**}$ exceeds a pre-specified control limit $\tilde{h}_\ell^{**}$.

3.2 The T^2-RD-MCUSUM chart

We now construct the reduced-dimension T^2-based chart, which is referred to as the T^2-RD-MCUSUM chart hereafter. Given a shift cluster, we compute a T^2 statistic at each time step as follows,

$$\tilde{a}_t^{c,r} = \tilde{\boldsymbol{x}}_t^{c,r'} \boldsymbol{\Sigma}_{c,r}^{-1} \tilde{\boldsymbol{x}}_t^{c,r} - \tilde{p} - k\sqrt{2\tilde{p}}, \quad i = 1, 2, \ldots, t, \tag{8}$$

where $\tilde{p} = |O^{c,r}|$. Note that when the process is in-control, the mean and variance of the T^2 statistic is given by

$$\mathrm{E}[\tilde{\boldsymbol{x}}_t^{c,r\,'} \boldsymbol{\Sigma}_{c,r}^{-1} \tilde{\boldsymbol{x}}_t^{c,r}] = \tilde{p}, \quad \text{and} \quad \mathrm{Var}[\tilde{\boldsymbol{x}}_t^{c,r\,'} \boldsymbol{\Sigma}_{c,r}^{-1} \tilde{\boldsymbol{x}}_t^{c,r}] = 2\tilde{p}.$$

Based on $\tilde{a}_t^{c,r}$, the monitoring statistic for each cluster is computed recursively over time, similar to (5). The global detection statistic is constructed by maximizing over

Table 1 Summary of charts with the full and RD approaches

	F-MCUSUM	RD-MCUSUM
LR based	$S_t^{**} = \max_{c,r} S_t^{c,r} \geq h_\ell^{**}$	$\tilde{S}_t^{**} = \max_{c,r} \tilde{S}_t^{c,r} \geq \tilde{h}_\ell^{**}$
T^2 based	$T_t^{**} = \max_{c,r} T_t^{c,r} \geq h_a^{**}$	$\tilde{T}_t^{**} = \max_{c,r} \tilde{T}_t^{c,r} \geq \tilde{h}_a^{**}$

all clusters: $\tilde{T}_t^{**} = \max_{c\in P, r\in R} \tilde{T}_t^{c,r}$ for $t = 1, 2, \ldots$. An alarm is signaled when $\tilde{T}_t^{**}$ exceeds a control limit $\tilde{h}_a^{**}$.

Hereafter, the charts based on full observation vectors are referred to as the F-MCUSUM charts, including the LR-F-MCUSUM and T^2-F-MCUSUM charts, while the reduced dimension charts are referred to as the RD-MCUSUM charts, including the LR-RD-MCUSUM and T^2-RD-MCUSUM charts. Table 1 summarizes our methods and terminology.

4 Theoretical analysis for effects of spatial correlation

In this section, we analytically compare the F-MCUSUM and RD-MCUSUM charts in terms of detection performance. We use ARL_1 given a target ARL_0 as our performance metric.

Both the F-MCUSUM and RD-MCUSUM charts use scan statistics to search for the true shift cluster from the set of possible shift clusters $\{O^{c,r} | c \in P,\ r \in R\}$ at each time step. In practice, the center and the radius of an anomaly are unknown and, thus, the entire monitored area needs to be scanned. In this section, we perform a theoretical analysis of the F-MCUSUM and RD-MCUSUM charts in a simplified setting, that is, when the actual shift cluster is known. This simplified situation provides a few insights into understanding the impact of dimensionality reduction on the performance of an MCUSUM chart with spatial scanning.

Suppose a shift affects a cluster with center c and radius r, i.e., $O^{c,r}$ is the actual shift cluster. Suppose the cluster $O^{c,r}$ contains $\tilde{p}$ locations. Without loss of generality, assume that the affected locations correspond to the first $\tilde{p}$ entries in the observation vectors, for instance, through reindexing. Hence, the post-change mean vector is $\boldsymbol{\mu}_1 = [\mu_1, \ldots, \mu_{\tilde{p}}, 0, \ldots, 0]' = [\tilde{\boldsymbol{\mu}}_{c,r}', 0, \ldots, 0]'$, where $\tilde{\boldsymbol{\mu}}_{c,r}' = [\mu_1, \ldots, \mu_{\tilde{p}}]' \neq \mathbf{0}$. We can relate the original full observation vector $\boldsymbol{x}_t$ to the reduced-dimension observation vector $\tilde{\boldsymbol{x}}_t^{c,r}$ in (6) and (8) and the modified full observation vector $\boldsymbol{x}_t^{c,r}$ in (4) as $\boldsymbol{x}_t = [\tilde{\boldsymbol{x}}_t^{c,r'}, 0, \ldots, 0]' + [0, \ldots, 0, \hat{\boldsymbol{x}}_t^{c,r'}]' = \boldsymbol{x}_t^{c,r} + [0, \ldots, 0, \hat{\boldsymbol{x}}_t^{c,r'}]'$, where $\tilde{\boldsymbol{x}}_t^{c,r} = [x_{t1}, \ldots, x_{t\tilde{p}}]'$ and $\hat{\boldsymbol{x}}_t^{c,r'} = [x_{t(\tilde{p}+1)}, \ldots, x_{tp}]'$. Furthermore, partition the $p \times p$ dimensional covariance matrix for the full observation vectors accordingly $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$, where $\boldsymbol{\Sigma}_{11} \in R_{\tilde{p}\times\tilde{p}}$, $\boldsymbol{\Sigma}_{12} \in R_{\tilde{p}\times(p-\tilde{p})}$, $\boldsymbol{\Sigma}_{21} \in R_{(p-\tilde{p})\times\tilde{p}}$ and $\boldsymbol{\Sigma}_{22} \in R_{(p-\tilde{p})\times(p-\tilde{p})}$. Using the Schur complement (Zhang 2006) of $\boldsymbol{\Sigma}_{11}$, we write the inverse of $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{A^*}^{-1} & \boldsymbol{\Sigma}_{B}^{-1} \\ \boldsymbol{\Sigma}_{C}^{-1} & \boldsymbol{\Sigma}_{D}^{-1} \end{bmatrix},$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_{A^*}^{-1} &= \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}; \\ \boldsymbol{\Sigma}_{B}^{-1} &= -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}; \\ \boldsymbol{\Sigma}_{C}^{-1} &= -(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}; \quad \text{and} \\ \boldsymbol{\Sigma}_{D}^{-1} &= (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}. \end{aligned} \tag{9}$$

4.1 Relations between statistics in full- and reduced-dimension charts

As a basis for the subsequent analysis, we derive relations for the LR and T^2 statistics used in the F- and the RD-MCUSUM charts. Using Eqs. (2), (6) and (9), the original LR statistic in (2) can be written as

$$\begin{aligned} \ell_t^{c,r} &= \tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{11}^{-1}\tilde{\boldsymbol{x}}_t^{c,r} - 0.5\tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{11}^{-1}\tilde{\boldsymbol{\mu}}_{c,r} + \tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{A^*}^{-1}\tilde{\boldsymbol{x}}_t^{c,r} \\ &\quad + \tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{B}^{-1}\hat{\boldsymbol{x}}_t^{c,r} - 0.5\tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{A^*}^{-1}\tilde{\boldsymbol{\mu}}_{c,r} \\ &= \tilde{\ell}_t^{c,r} + \tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{A^*}^{-1}\tilde{\boldsymbol{x}}_t^{c,r} + \tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{B}^{-1}\hat{\boldsymbol{x}}_t^{c,r} - 0.5\tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{A^*}^{-1}\tilde{\boldsymbol{\mu}}_{c,r}. \end{aligned} \tag{10}$$

From (10), it is evident that the dimension reduction for the LR-based chart is equivalent to eliminating the last three terms on the right-hand side of (10) from the full LR statistic. It is worth mentioning that $\tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{B}^{-1}\hat{\boldsymbol{x}}_t^{c,r}$ contains only noise information, since no mean shift occurs in $\hat{\boldsymbol{x}}_t^{c,r}$.

Similarly, we derive a relation for the T^2 statistics used in the F- and the RD-MCUSUM charts. Using equation (9), we obtain the in-control mean and variance of the full version T^2 statistic as follows:

$$\begin{aligned} \mathrm{E}[\boldsymbol{x}_t^{c,r\,'}\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_t^{c,r}] &= \tilde{p} + \mathrm{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\} \quad \text{and} \\ \mathrm{Var}[\boldsymbol{x}_t^{c,r\,'}\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_t^{c,r}] &= 2\mathrm{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\} + 4\mathrm{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\} + 2\tilde{p}. \end{aligned}$$

Hence, the statistics $a_t^{c,r}$ defined in (4) and $\tilde{a}_t^{c,r}$ defined in (8) are related to each other via

$$\begin{aligned} a_t^{c,r} &= \tilde{\boldsymbol{x}}_t^{c,r'}(\boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{A^*}^{-1})\tilde{\boldsymbol{x}}_t^{c,r} - (\tilde{p} + \mathrm{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\}) \\ &\quad - k[2\mathrm{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\} + 4\mathrm{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\} + 2\tilde{p}]^{1/2} \\ &= \tilde{a}_t^{c,r} + \tilde{\boldsymbol{x}}_t^{c,r'}\boldsymbol{\Sigma}_{A^*}^{-1}\tilde{\boldsymbol{x}}_t^{c,r} - \mathrm{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\} \\ &\quad - k([2\mathrm{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\} + 4\mathrm{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\} + 2\tilde{p}]^{1/2} - (2\tilde{p})^{1/2}). \end{aligned}$$

Table 2 Drift and variance parameters of the LR based charts

	LR-F-MCUSUM	LR-RD-MCUSUM
d_0	$-\frac{1}{2}\boldsymbol{\mu}_{c,r}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{c,r}$	$-\frac{1}{2}\tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{11}^{-1}\tilde{\boldsymbol{\mu}}_{c,r}$
$d_{c,r}$	$\frac{1}{2}\boldsymbol{\mu}_{c,r}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{c,r}$	$\frac{1}{2}\tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{11}^{-1}\tilde{\boldsymbol{\mu}}_{c,r}$
$\Omega_0^2(\Omega_{c,r}^2)$	$\boldsymbol{\mu}_{c,r}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{c,r}$	$\tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{11}^{-1}\tilde{\boldsymbol{\mu}}_{c,r}$

Table 3 Drift and variance parameters of the T^2-statistic based charts

	T^2-F-MCUSUM	T^2-RD-MCUSUM
d_0	$-k[2\text{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\}+4\text{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\}+2\tilde{p}]^{1/2}$	$-k(2\tilde{p})^{1/2}$
$d_{c,r}$	$\tilde{\boldsymbol{\mu}}_{c,r}^{'}(\boldsymbol{\Sigma}_{11}^{-1}+\boldsymbol{\Sigma}_{A^*}^{-1})\tilde{\boldsymbol{\mu}}_{c,r}-k[2\text{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\}+4\text{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\}+2\tilde{p}]^{1/2}$	$\tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{11}^{-1}\tilde{\boldsymbol{\mu}}_{c,r}-k(2\tilde{p})^{1/2}$
Ω_0^2	$2\text{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\}+4\text{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\}+2\tilde{p}$	$2\tilde{p}$
$\Omega_{c,r}^2$	$\Omega_0^2+4\tilde{\boldsymbol{\mu}}_{c,r}^{'}(\boldsymbol{\Sigma}_{11}^{-1}+\boldsymbol{\Sigma}_{A^*}^{-1})\boldsymbol{\Sigma}_{11}(\boldsymbol{\Sigma}_{11}^{-1}+\boldsymbol{\Sigma}_{A^*}^{-1})\tilde{\boldsymbol{\mu}}_{c,r}$	$2\tilde{p}+4\tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{11}^{-1}\tilde{\boldsymbol{\mu}}_{c,r}$

4.2 Performance metric: ARL_1 measure

Our objective is to compare the detection performance of the F- and the RD-MCUSUM charts in terms of ARL_1 for a fixed ARL_0. We define a performance metric called the ARL_1 *measure*. For a fixed ARL_0, a smaller ARL_1 measure implies shorter detection delay.

Kim et al. (2007) derive a formula to approximate ARL of a single CUSUM chart for both in-control and out-of-control processes. The formula is given by

$$\text{ARL} \approx \begin{cases} \dfrac{\Omega^2}{2d^2}\left\{\exp\left[-\dfrac{2d(H+1.166\Omega)}{\Omega^2}\right]-1+\dfrac{2d(H+1.166\Omega)}{\Omega^2}\right\}, & \text{if } d\neq 0; \\ \left(\dfrac{H+1.166\Omega}{\Omega}\right)^2, & \text{if } d=0; \end{cases} \tag{11}$$

where d is the drift parameter, Ω^2 is the variance parameter, and H is the control limit.

In our settings, specifically, if observations are temporally independent, we have $d=\text{E}[\ell_t^{c,r}]$, $\Omega^2=\text{Var}[\ell_t^{c,r}]$ for the LR-F-MCUSUM chart, and $d=\text{E}[\tilde{\ell}_t^{c,r}]$, $\Omega^2=\text{Var}[\tilde{\ell}_t^{c,r}]$ for the LR-RD-MCUSUM chart. In addition, we have $d=\text{E}[a_t^{c,r}]$, $\Omega^2=\text{Var}[a_t^{c,r}]$ for the T^2-F-MCUSUM chart and $d=\text{E}[\tilde{a}_t^{c,r}]$, $\Omega^2=\text{Var}[\tilde{a}_t^{c,r}]$ for the T^2-RD-MCUSUM chart. In the following, we denote the in-control drift and variance parameters as d_0 and Ω_0^2, respectively. The out-of-control parameters are defined similarly. If the shift center is c and radius is r, we denote the out-of-control drift and variance as $d_{c,r}$ and $\Omega_{c,r}^2$, respectively. Tables 2 and 3 summarize these parameters.

The in-control drift d_0 is always negative but the out-of-control drift $d_{c,r}$ can either be negative (if the shift magnitude is too small) or positive. In this section, we assume that the shift magnitude is sufficiently large so that $d_{c,r}$ is positive.

Consider a special function called the Lambert W function. Let $W_0(\cdot)$ be the principal branch of the Lambert W function (Corless et al. 1996). For a fixed target ARL_0, using (11), we derive an approximation for ARL_1 as follows:

$$\text{ARL}_1 \approx -\frac{\Omega_0^2}{2d_0 d_{c,r}}\left[\epsilon_{\eta_0} + \frac{\Omega_{c,r}^2/d_{c,r}}{\Omega_0^2/d_0}\right] - 1.166\frac{1}{d_{c,r}}(\Omega_0 - \Omega_{c,r}), \tag{12}$$

where

$$\epsilon_{\eta_0} = -W_{-1}(-e^{-\eta_0}) - \eta_0, \quad \text{and} \quad \eta_0 = \frac{2d_0^2}{\Omega_0^2}\text{ARL}_0 + 1.$$

The derivation is presented in Section A1 of the online supplement.

Note that the ϵ_{η_0} function tends to be very flat. Moreover, with a fixed ARL_0, the values of η_0 for the F- and RD-MCUSUM charts are very close. The term $\frac{\Omega_{c,r}^2/d_{c,r}}{\Omega_0^2/d_0}$ is always a constant, -1, for the LR-based charts (based on values in Table 2) and is expected to have similar values for T^2-F-MCUSUM and T^2-RD-MCUSUM charts for a spatial model, where spatial covariance decays as the distance increases (see Section A1 of the online supplement for a detailed discussion). In addition, for the LR-based charts, since $\Omega_0 = \Omega_{c,r}$ due to the assumption in Sect. 2.1, the second term in the right-hand side of (12) is equal to zero, and for the T^2-based charts, it is small compared to the first term. Thus, when we fix ARL_0, to compare ARL_1s of the F- and RD-MCUSUM charts, we may compare the values of $\left|\Omega_0^2/(d_0 d_{c,r})\right|$, which we define as the ARL_1 measure.

In the following, we denote the ARL_1 measure for LR-F-MCUSUM and LR-RD-MCUSUM charts as m_{LR} and $\tilde{m}_{\text{LR}}$, respectively, and for T^2-F-MCUSUM and T^2-RD-MCUSUM charts as m_{T^2} and $\tilde{m}_{T^2}$, respectively. For LR-based charts, we obtain,

$$m_{\text{LR}} = \frac{4}{\boldsymbol{\mu}_{c,r}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{c,r}} = \frac{4}{\tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{11}^{-1}\tilde{\boldsymbol{\mu}}_{c,r} + \tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{A^*}^{-1}\tilde{\boldsymbol{\mu}}_{c,r}} \le \frac{4}{\tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{11}^{-1}\tilde{\boldsymbol{\mu}}_{c,r}} = \tilde{m}_{\text{LR}}. \tag{13}$$

Equation (13) indicates that m_{LR} is always smaller than or equal to $\tilde{m}_{\text{LR}}$, which, in turn, implies that the ARL_1 of the method using full observation vectors is always smaller than that of the chart with reduced dimension vectors. Since a smaller ARL_1 measure implies a smaller ARL_1, we expect that the LR-F-MCUSUM chart generally detects a shift faster than the LR-RD-MCUSUM chart.

For T^2-based charts, we have

$$m_{T^2} = \left[k^2 - k\frac{\tilde{\boldsymbol{\mu}}_{c,r}^{'}(\boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{A^*}^{-1})\tilde{\boldsymbol{\mu}}_{c,r}}{[2\text{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\} + 4\text{tr}\{\boldsymbol{\Sigma}_{A^*}^{-1}\boldsymbol{\Sigma}_{11}\} + 2\tilde{p}]^{1/2}}\right]^{-1}$$

and

$$\tilde{m}_{T^2} = \left[k^2 - k\frac{\tilde{\boldsymbol{\mu}}_{c,r}^{'}\boldsymbol{\Sigma}_{11}^{-1}\tilde{\boldsymbol{\mu}}_{c,r}}{\sqrt{2\tilde{p}}}\right]^{-1}.$$

Note that the theoretical performance measure and the above analysis are applicable to a spatial covariance matrix, $\mathbf{\Sigma}$, with a general structure. Several commonly used spatial covariance structures are listed below. We use d to denote the distance between two sensors, $C(d|\rho)$ to denote the correlation function between two sensors, which is a function of d and other parameters and $\mathbb{1}_{\{\cdot\}}$ to denote an indicator function.

1. Four-value model: $C(d|\rho) = 1\mathbb{1}_{\{d=0\}} + \rho\mathbb{1}_{\{d=1\}} + \frac{\rho}{2}\mathbb{1}_{\{d=\sqrt{2}\}}$ for $\rho \in [0, 1]$.
2. Polynomial model: $C(d|\rho) = 1\mathbb{1}_{\{d=0\}} + \rho^d\mathbb{1}_{\{d>0\}}$ for $\rho \in [0, 1]$.
3. Matérn model: $C(d|\theta) = 1\mathbb{1}_{\{d=0\}} + \frac{1}{2^{v-1}\Gamma(v)}(\sqrt{2}v^{1/2}d/\theta)^v K_v(\sqrt{2}v^{1/2}d/\theta)\mathbb{1}_{\{d>0\}}$ for $\theta > 0$, where K_v is the modified Bessel function of order v (Ripley 2005).

Based on the correlation function, the entries of the covariance matrix $[\mathbf{\Sigma}]_{i,j}$ are determined as $C(d(q_i, q_j)|\theta)$, where q_i and q_j are the coordinates of sensors i and j, respectively.

4.2.1 An illustrative example

Using a simple illustrative example, we calculate how much we lose or gain in terms of ARL_1 when reduced dimension vectors are used. We use a tridiagonal spatial covariance matrix as an example. The correlation between two sensors is ρ if they are neighboring, and 0 otherwise. Such a covariance matrix is denoted by $\mathbf{\Sigma}_1(\rho) \in R_{p\times p}$, with $[\mathbf{\Sigma}_1(\rho)]_{i,j} = 1$ if $i = j$, $[\mathbf{\Sigma}_1(\rho)]_{i,j} = \rho$ if $|i - j| = 1$ and $[\mathbf{\Sigma}_1(\rho)]_{i,j} = 0$ otherwise.

We use $p = 5$, $\tilde{p} = 2$ and $\mu_{c,r} = [1, 1, 0, 0, 0]'$ in this example. The ratio $m_{\text{LR}}/\tilde{m}_{\text{LR}}$ is calculated as a function of the spatial correlation ρ. Since a spatial correlation is unlikely to be very large in practice, we consider ρ in the range of $0 \le \rho \le 0.3$. A ratio smaller than one implies that the charts with full observation vectors have a smaller ARL_1 than the charts with reduced observation vectors, and vice versa.

For LR charts, the ratio of ARL_1 measure for the full- and reduced-dimension methods can be simplified to

$$\frac{m_{\text{LR}}}{\tilde{m}_{\text{LR}}} = \left[1 + \frac{\tilde{\boldsymbol{\mu}}_{c,r}^{'}\mathbf{\Sigma}_{A^*}^{-1}\tilde{\boldsymbol{\mu}}_{c,r}}{\tilde{\boldsymbol{\mu}}_{c,r}^{'}\mathbf{\Sigma}_{11}^{-1}\tilde{\boldsymbol{\mu}}_{c,r}}\right]^{-1} = \frac{6\rho^3 - 6\rho^2 - 2\rho + 2}{\rho^4 + 4\rho^3 - 5\rho^2 - 2\rho + 2}.$$

We plot this ratio as a function of ρ in Fig. 2a. For LR-based charts, the ratio is always smaller than 1, as expected and it decreases as the spatial correlation ρ increases. This indicates that for LR charts, methods based on full observation vectors are always better. However, the performance loss of using reduced-dimensional vectors is small (less than 7%, as shown in the plot). Thus, we expect that when the spatial correlation decays reasonably fast, in the case of known center and radius, the reduced-dimension charts do not lose much detection power compared to its full version.

For T^2-based charts, there is no simple analytic expression for the ratio of ARL_1 measures. Hence, we calculate the ratio numerically. Figure 2b depicts the plot of $m_{T^2}/\tilde{m}_{T^2}$ as a function of ρ. Interestingly, the ratio is slightly greater than 1, thereby indicating that the RD method may perform slightly better than the full-dimension counterparts for the particular covariance structure that we consider in this example.

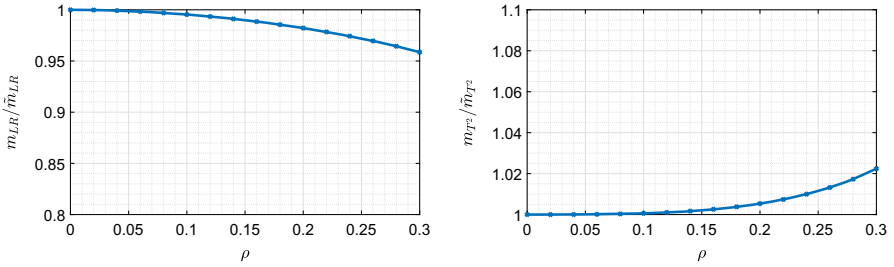

Fig. 2 Simple example with $p = 5$, $\tilde{p} = 2$, and $\mu_{c,r} = [1, 1, 0, 0, 0]'$: **a** $m_{\mathrm{LR}}/\tilde{m}_{\mathrm{LR}}$ as a function of ρ; and **b** $m_{T^2}/\tilde{m}_{T^2}$ as a function of ρ

However, in general, the ratio is smaller than 1 depending on the covariance structure, which we present in Sect. 5.

We also conduct simulation experiments assuming that the shift cluster is known. Due to the brevity of the paper, the results are presented in Figure S1 of the online supplement. In summary, the simulation results match the ARL_1 measures very well. In addition, the RD-MCUSUM charts use only 3–4% more ARL_1 than the F-MCUSUM charts for a reasonable range of spatial correlation, when only a single shift cluster is considered. For T^2 charts, there is almost no performance loss for the settings that we considered.

5 Experiments

In this section, we conduct various numerical experiments under more realistic settings, where the shift center and/or the shift radius are unknown.

5.1 Experimental setup

We first consider the case in which the shift center is unknown, and the radius is known to be $r = 1$. In this case, we scan over p possible shift clusters with different centers at each time step. Second, we consider the case in which both the shift center c and radius r are unknown. We scan over a set R of possible radii at every possible shift center. In the experiments, we use $R = \{1, \sqrt{2}\}$. In a bell-shaped signal case, which is discussed toward the end of Sect. 5.2, we consider $R = \{0, 1, \sqrt{2}, 2, 2\sqrt{2}, 3, 3\sqrt{2}\}$.

We run the control charts on three commonly used spatial models (the four-value model, polynomial model and Matérn model, as introduced in Sect. 4.2). In applications like environmental monitoring, sensors tend to be placed within a reasonable distance and the spatial correlation between two locations is usually not high. Thus, we test $\rho \in \{0,\ 0.02,\ 0.04, \ldots,\ 0.3\}$. In the Matérn model, we test $\theta \in \{0,\ 0.054,\ 0.108, \ldots,\ 0.81\}$ and use order $v = \frac{1}{2}$. Note that $\theta = 0.8$ for the Matérn model corresponds to the spatial correlation among neighboring locations, $\rho \approx 0.3$.

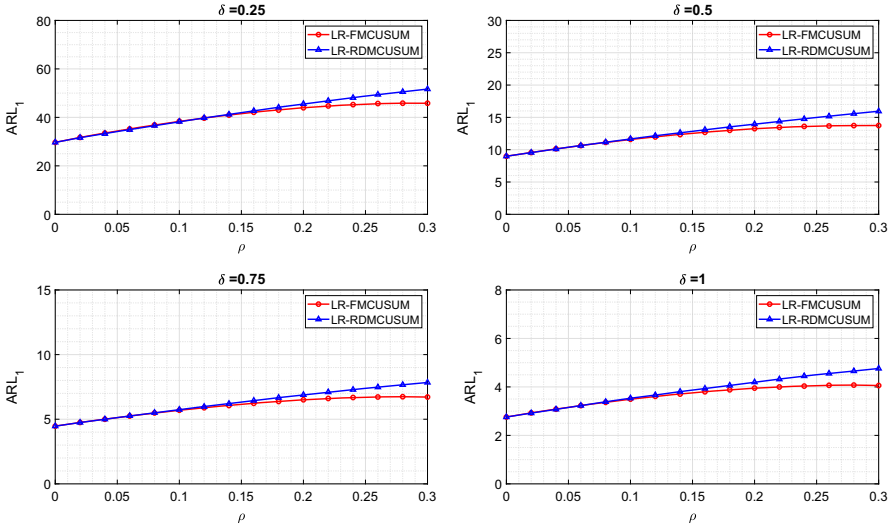

Fig. 3 Simulated ARL_1 of LR-F-MCUSUM and LR-RD-MCUSUM charts on the four-value model with $r_{out} = \sqrt{2}$ in the unknown center and unknown radius case

The monitored area in the simulation experiments has a dimensionality of $p = 7 \times 7$. For the out-of-control state, homogeneous shifts (shifts of all affected locations in the cluster have the same magnitude) with magnitudes $\delta = 0.25,\ 0.5,\ 0.75$ and 1 are tested. The targeted ARL_0 is fixed to 1000 in all the cases. Moreoever, all simulated ARL values are obtained based on 10,000 simulation replications.

5.2 Results

For the sake of brevity, the results with an unknown center but known radius are included in Section A3.1 of the online supplement and we present the results with the unknown center and unknown radius in this subsection.

In the unknown center and unknown radius case, we denote the actual radius as r_{out} and consider the possible radius $R \in \{1, \sqrt{2}\}$. At each time step, the control chart scans over $2p$ possible shift clusters. Figures 3 and 4 compare ARL_1 of LR-F-MCUSUM and LR-RD-MCUSUM charts on two different spatial correlation structures. Figures 5 and 6 present results for T^2-F-MCUSUM and T^2-RD-MCUSUM charts. The results for the Matérn model with $r_{out} = \sqrt{2}$ are illustrated in Figures S23 and S24 in the online supplement. The remaining results, including plots of simulated ARL_0 and ARL_1 when $r_{out} = 1$, and the percentage difference of simulated ARL_1 are presented in Section A3.2 of the online supplement.

Based on the numerical results, we conclude that in general, the reduced-dimension charts do not severely sacrifice the ARL_1 performance in the range of correlation tested in the paper. Therefore, the RD-MCUSUM charts can be a powerful and easy-to-implement alternative of the F-MCUSUM charts, particularly when the dimension of the monitored area is high and the full covariance matrix is ill-conditioned.

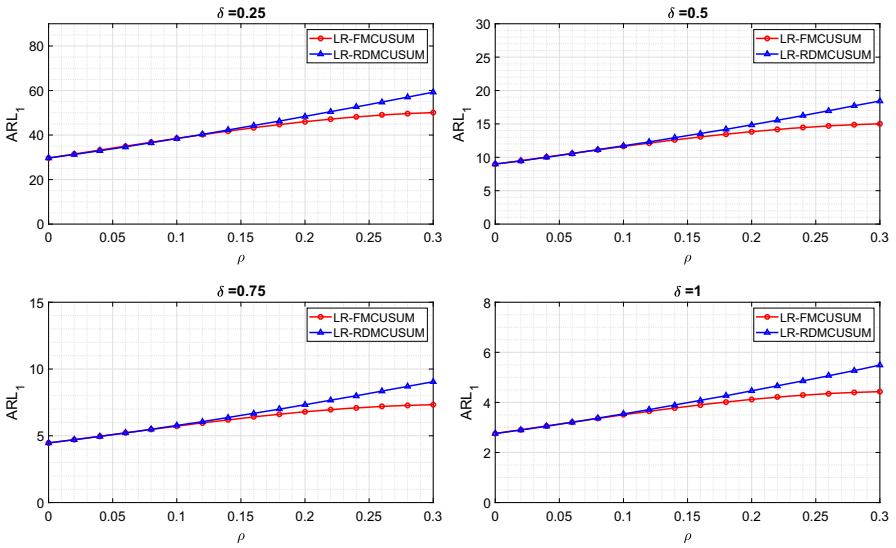

Fig. 4 Simulated ARL_1 of LR-F-MCUSUM and LR-RD-MCUSUM charts on the polynomial model with $r_{out} = \sqrt{2}$ in the unknown center and unknown radius case

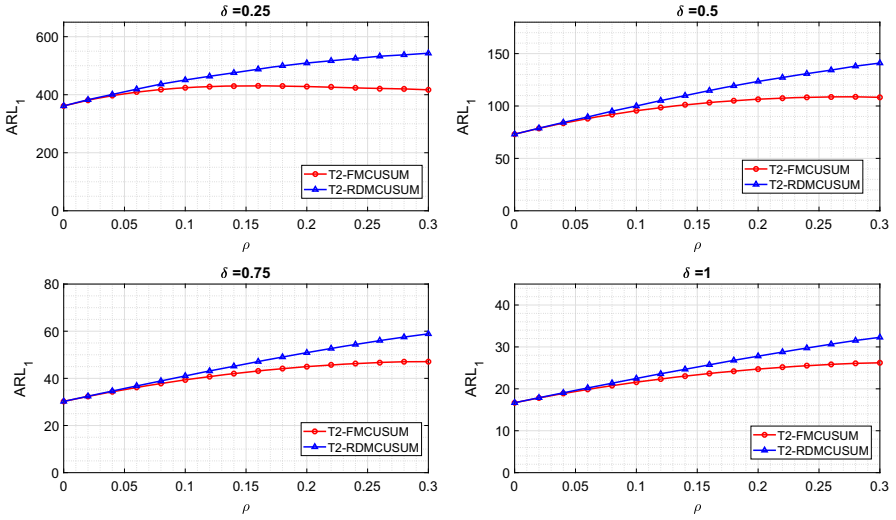

Fig. 5 Simulated ARL_1 of T^2-F-MCUSUM and T^2-RD-MCUSUM charts on the four-value model with $r_{out} = \sqrt{2}$ in the unknown center and unknown radius case

Finally, a "bell-shaped" signal is considered rather than a "box-shaped" one. Thus far, we assume that the shift signal is box-shape, which implies that an anomaly only affects a cluster of neighboring sensors but not all of them. Moreover, we only consider a homogeneous shift, which means that the shift magnitude at all affected sensors are equal. In reality, there exist other forms of shift signals, such as a bell-shaped signal. Such a shift signal occurs at center $c \in P$, affecting all sensors, and the shift magnitude

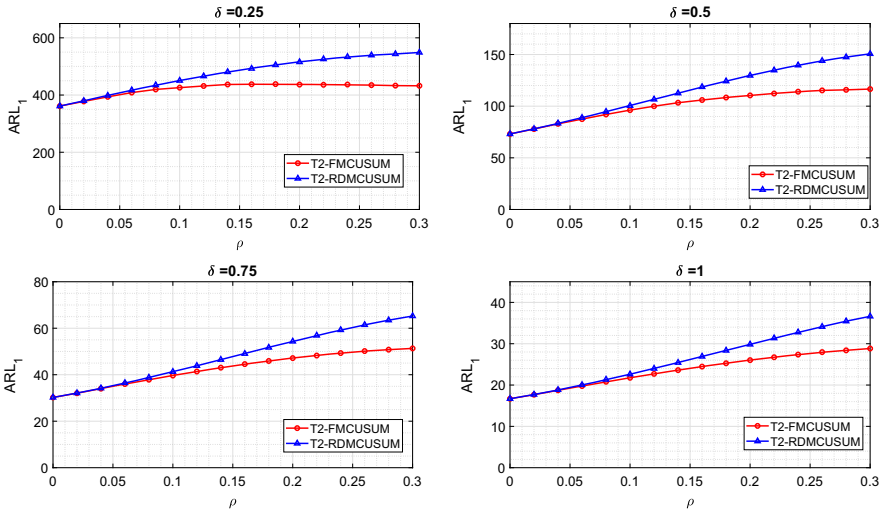

Fig. 6 Simulated ARL_1 of T^2-F-MCUSUM and T^2-RD-MCUSUM charts on the polynomial model with $r_{\text{out}} = \sqrt{2}$ in the unknown center and unknown radius case

decays with the distances from c. In this case, reduced-dimension charts are destined to lose performance, since each reduced-dimension observation can only capture a portion of the signal energy. The smaller the scan radius, the more dimensionality reduction we achieve, however, a greater loss of performance incurs. In this example, we consider a larger set of possible shift radii, $R = \{0, 1, \sqrt{2}, 2, 2\sqrt{2}, 3, 3\sqrt{2}\}$. Section A3.3 of the online supplement includes the performances of the proposed charts and demonstrates how our ARL_1 measure can be used in selecting a scan radius in the reduced dimension charts.

5.3 The ill-conditioned covariance matrix

We demonstrate the benefit of RD-MCUSUM charts when the full covariance matrix is ill-conditioned. We consider a monitored area with dimensionality $p = 49$. The spatial covariance matrix of the first 48 sensors is constructed using a polynomial model with $\rho = 0.3$. We assume that the 49_{th} sensor and the first sensor have a high correlation, 0.99, and consequently, the full spatial covariance matrix of the monitored area is ill-conditioned (with its determinant ≈ 0). In this case, F-MCUSUM charts are not applicable, as the inversion of an ill-conditioned matrix is unstable. Figure 7 shows the LR-RD-MCUSUM charts based on simulated data when homogeneous shifts with different magnitudes δ are added at time $t = 50$. The control limits are adjusted so that $\text{ARL}_0 = 1000$. As we can observe from Fig. 7, LR-RD-MCUSUM charts successfully detect the mean shifts in all cases.

6 Application: water quality monitoring

In this section, we apply the proposed methods to real-time water quality monitoring for a river network. The objective is to detect contaminant spills that cause water

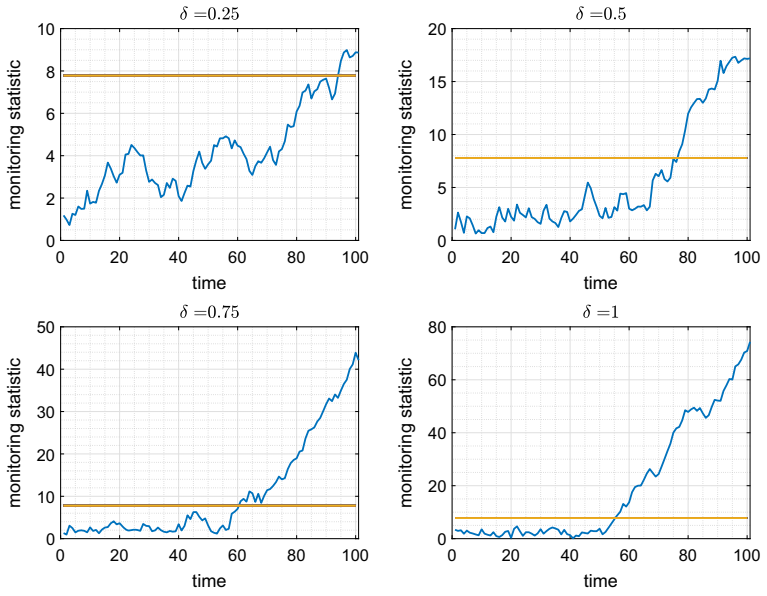

Fig. 7 LR-RD-MCUSUM chart detects a mean shift when the full covariance matrix is ill-conditioned. Mean shifts occur at $t = 50$

pollution in the river. In this application, the shape of the monitored region is not rectangular and scan clusters are non-circular.

6.1 Data

We study a river network for the Altamaha River in Georgia, United States. The shape of the river is shown in Fig. 8a. The nodes in the river network represent a possible spill location and potential monitoring locations where concentration data are collected.

The contaminant concentration data for such a river network is simulated by the Storm Water Management Model (SWMM) developed by the United States Environmental Protection Agency. SWMM requires geologic, geometric and fundamental hydrodynamics data to construct a river network. Given rainfall information, the location, intensity and duration of a contaminant spill, SWMM simulates the contaminant transport process through the river over a period of time. In the simulation, rain events and spill events bring randomness to the contaminant transport process. The same dataset used in Telci and Aral (2011) is adopted in our case study to generate rain events in ten sub-catchments shown in Fig. 8b.

A spatial correlation exists among the concentration data collected at different locations in the river network due to the nature of hydrodynamics. We construct a spatial model for the river network based on Ver Hoef and Peterson (2010). See Section A4 of the online supplement for the detailed spatial model.

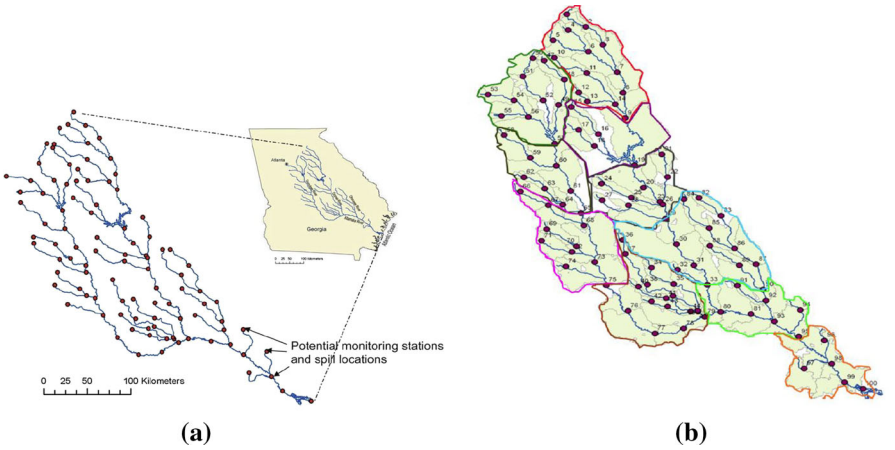

Fig. 8 **a** Shape and monitoring locations and **b** ten sub-catchments of the Altamaha River (Telci and Aral 2011)

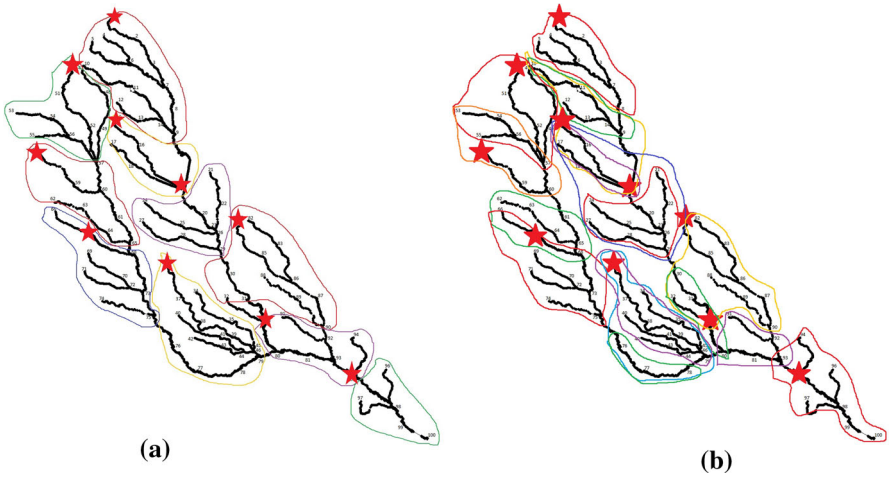

Fig. 9 Two sets of scan clusters for spatial scanning: **a** non-overlapped clusters (consistent with the 10 sub-catchements of the Altamaha river network); **b** overlapped clusters. Red stars represent possible spill locations

6.2 Results of the water quality monitoring application

Here we present results of the application of LR-F-MCUSUM and LR-RD-MCUSUM charts to online detection of contaminant spills in the Altamaha river network. Among the 100 nodes on the river network, 10 of them are used as potential contaminant spill locations, which are marked as red stars in Fig. 9, and the remaining 90 nodes are used for collecting measurements every 15 min. In each replication, we run SWMM to simulate the river network during a 10-day period. A single instantaneous spill with a spill location randomly selected from one of the ten possible locations is generated.

Table 4 Detection performance (ARL_1) obtained using the non-overlapped (NOV) and overlapped (OV) sets of scan clusters. Numbers in parentheses are standard error of ARLs

Intensity	ARL		False alarm		Fail to detect	
	F	RD	F (%)	RD (%)	F (%)	RD (%)
NOV						
Low	40.68 (4.11)	46.91 (4.60)	7	7	8	9
Medium	30.11 (2.31)	33.45 (2.90)	7	7	2	9
High	23.79 (1.39)	25.65 (1.68)	9	7	0	0
OV						
Low	36.45 (3.73)	39.16 (4.22)	8	9	4	8
Medium	33.09 (2.54)	30.44 (2.26)	4	3	1	3
High	27.30 (2.76)	27.68 (2.43)	4	3	1	3

The spill starting time is uniformly distributed between 15 and 20 h. The intensity of the contaminant spills follows a uniform distribution, and we consider three different levels: U (10, 100) g/l (low), U (100, 250) g/l (medium), and U (250, 500) g/l (high).

Since the Altamaha river network does not have a regular shape and the monitoring stations are not located on the uniform grid, we do not use the circle shape clusters for spatial scanning. Instead, we construct scan clusters based on locations of the sensors and topology of the river network. Two sets of clusters are considered: (i) ten clusters that are consistent with the ten sub-catchments of the Altamaha River and have no overlap (Fig. 9a), and (ii) 18 clusters, some of which are partially overlapping (Fig. 9b).

For the LR-based MCUSUM charts, we set a minimum size of change that we aim to detect to be 0.05 g/l and use it to construct the post-change mean vectors, $\boldsymbol{\mu}_{c,r}$ and $\tilde{\boldsymbol{\mu}}_{c,r}$. In order to conduct a fair comparison, the thresholds for both LR-F-MCUSUM and LR-RD-MCUSUM charts are adjusted so that the in-control average run lengths are 10 days (960 samples).

We generate 300 simulated contaminant spills (100 spills in each level of intensity). Detection performances of the two methods using the two sets of scan clusters are summarized in Table 4. From Table 4, we can observe that the ARL performances of the two methods are consistent with our analysis in the previous sections: the chart using full observations achieves a slightly smaller ARL. A further detailed comparison is presented in Table S2 from Section A5 of the online supplement. The table shows that, almost under all settings, the percentage of cases where the LR-RD-MCUSUM chart performs no worse than (performs better than or similarly with) the LR-F-MCUSUM chart is higher than 60%. We conclude that the RD-based charts show very competitive performances in practice. Given that the RD-based charts enjoy distributed computing with a little loss in ARL performances, they should be considered as a good alternative or even a better option for large-scale sensor networks, especially when the full covariance matrix is ill-conditioned.

7 Conclusion

We study the reduced-dimension control charts for spatio-temporal surveillance in the presence of various spatial covariance structures. The reduced-dimension charts perform spatial scanning by breaking the entire monitoring area into overlapping clusters as well as computing the detection statistics locally while incorporating local covariance. In the presence of high-dimensional data streams, the reduced-dimension approach enjoys lower communication complexity and better numerical stability. More specifically, provided that a set of possible shift radii is fixed (because the set depends on a shift type rather than the size of a monitored area), the dimension-reduced charts have a computation time that is proportional to p, while the computation time of the full-observation charts increases proportional to p^3. We also study the loss and gain in quantified performance on account of ignoring full spatial covariance through systematic theoretical and numerical studies. Considering the benefits of dimensionality reduction, the RD charts can be a powerful and cheaper alternative to the full-observation charts in high-dimensional problems, especially when the inversion of the full-size matrix is a computational bottleneck or when a full-size covariance matrix is ill-conditioned. Finally, it should be noted that our results hold under a small spatial correlation (up to 0.3). Moreover, observation vectors are assumed to be temporally independent, which would be a reasonable assumption when observation vectors are sufficiently spaced out in time.

Acknowledgements This material is based upon work supported by NSF under Grants CMMI-1538746.

References

Bartram J, Ballance R (1996) Water quality monitoring: a practical guide to the design and implementation of freshwater quality studies and monitoring programmes. CRC Press, New York

Bodnar O, Schmid W (2005) Multivariate control charts based on a projection approach. Allgemeines Statistisches Archiv 89:75–93

Brynjarsdóttir J, Berliner LM (2014) Dimension-reduced modeling of spatio-temporal processes. J Am Stat Assoc 109:1647–1659

Corless RM, Gonnet GH, Hare DE, Jeffrey DJ, Knuth DE (1996) On the LambertW function. Adv Comput Math 5:329–359

Crosier RB (1988) Multivariate generalizations of cumulative sum quality-control schemes. Technometrics 30:291–303

Gaetan C, Guyon X (2010) Spatial statistics and modeling. Springer, New York

Guerriero M, Willett P, Glaz J (2009) Distributed target detection in sensor networks using scan statistics. In: IEEE transactions on signal processing, vol. 57, pp 2629–2639

He Q, Zhou S (2014) Discriminant locality preserving projection chart for statistical monitoring of manufacturing processes. Int J Prod Res 52:5286–5300

Healy JD (1987) A note on multivariate CUSUM procedures. Technometrics 29:409–412

Hotelling, H (1947) "Multivariate quality control," Techniques of statistical analysis

Jiang W, Han SW, Tsui K-L, Woodall WH (2011) Spatiotemporal surveillance methods in the presence of spatial correlation. Stat Med 30:569–583

Kim S-H, Alexopoulos C, Tsui K-L, Wilson JR (2007) A distribution-free tabular CUSUM chart for autocorrelated data. IIE Trans 39:317–330

Lee J, Hur Y, Kim S-H, Wilson JR (2012) Monitoring nonlinear profiles using a wavelet-based distribution-free CUSUM chart. Int J Prod Res 50:6574–6594

Lee ML, Goldsman D, Kim S-H (2015) Robust distribution-free multivariate CUSUM charts for spatiotemporal biosurveillance in the presence of spatial correlation. IIE Trans Healthc Syst Eng 5:74–88
Lee ML, Goldsman D, Kim S-H, Tsui K-L (2014) Spatiotemporal biosurveillance with spatial clusters: control limit approximation and impact of spatial correlation. IIE Trans 46:813–827
Liu K, Mei Y, Shi J (2015) An adaptive sampling strategy for online high-dimensional process monitoring. Technometrics 57:305–319
Liu K, Zhang R, Mei Y (2016) Scalable SUM-Shrinkage Schemes for Distributed Monitoring Large-Scale Data Streams. arXiv:1603.08652
Lowry CA, Woodall WH, Champ CW, Rigdon SE (1992) A multivariate exponentially weighted moving average control chart. Technometrics 34:46–53
Mishin D, Brantner-Magee K, Czako F, Szalay AS (2014) "Real time change point detection by incremental PCA in large scale sensor data," In: High performance extreme computing conference (HPEC), 2014 IEEE. IEEE pp 1–6
Page ES (1954) Continuous inspection schemes. Biometrika 41:100–115
Park Y, Baek SH, Kim S-H, Tsui K-L (2014) Statistical process control-based intrusion detection and monitoring. Qual Reliab Eng Int 30:257–273
Ripley BD (2005) Spatial statistics, vol 575. Wiley, New York
Rogerson PA (1997) Surveillance systems for monitoring the development of spatial patterns. Stat Med 16:2081–2093
Rogerson PA, Yamada I (2004) Approaches to syndromic surveillance when data consist of small regional counts. Morbidity and Mortality Weekly Report pp 79–85
Runger GC (1996) Projections and the U-squared multivariate control chart. J Qual Technol 28:313–319
Skubałlska-Rafajlowicz E (2013) Random projections and Hotelling's T2 statistics for change detection in high-dimensional data analysis. Int J Appl Math Comput Sci 23:447–461
Sonesson C (2007) A CUSUM framework for detection of space–time disease clusters using scan statistics. Stat Med 26:4770–4789
Spiegelhalter D, Sherlaw-Johnson C, Bardsley M, Blunt I, Wood C, Grigg O (2012) Statistical methods for healthcare regulation: rating, screening and surveillance. J R Stat Soc 175:1–47
Tango T (1995) A class of tests for detecting 'general'and 'focused'clustering of rare diseases. Stat Med 14:2323–2334
Telci IT, Aral MM (2011) Contaminant source location identification in river networks using water quality monitoring systems for exposure analysis. Water Qual Expos Health 2:205–218
Tsui K-L, Chiu W, Gierlich P, Goldsman D, Liu X, Maschek T (2008) A review of healthcare, public health, and syndromic surveillance. Qual Eng 20:435–450
Ver Hoef JM, Peterson EE (2010) A moving average approach for spatial statistical models of stream networks. J Am Stat Assoc 105:6–18
Wang H, Kim S-H, Huo X, Hur Y, Wilson JR (2015a) Monitoring nonlinear profiles adaptively with a wavelet-based distribution-free CUSUM chart. Int J Prod Res 53:4648–4667
Wang K, Jiang W, Li B (2015b) A spatial variable selection method for monitoring product surface. Int J Prod Res 54:4161–4181
Xie Y, Huang J, Willett R (2013) Change-point detection for high-dimensional time series with missing data. IEEE J Select Top Sign Process. 7:12–27
Xie Y, Siegmund D (2013) Sequential multi-sensor change-point detection. Annal Stat 41:670–692
Xie, Y, Wang M, Thompson A (2015) Sketching for sequential change-point detection. In: 2015 IEEE global conference on signal and information processing (GlobalSIP), IEEE, pp 78–82
Yamamoto M, Hwang H (2017) Dimension-reduced clustering of functional data via subspace separation. J Classif 34:294–326
Zhang F (2006) The Schur complement and its applications, vol 4. Springer, New York

Junzhuo Chen is a Ph.D. candidate in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. He received a B.S. degree in Industrial Engineering from the Shanghai Jiao Tong University, Shanghai, China, in 2014. His research interests include environmental monitoring and simulation optimization.

Chuljin Park is an assistant professor in the Department of Industrial Engineering at Hanyang University. He completed his Ph.D. in Operations Research at Georgia Institute of Technology in 2013. He

received the CHADONGWAN Young Researcher Award, KORMS, in 2018. His research interests include simulation optimization, simulation-based machine learning and environmental and healthcare system management.

Seong-Hee Kim is a professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. She received a B.S. in Industrial Management from Korea Advanced Institute of Science and Technology (KAIST) and an M.S. & Ph.D. in Industrial Engineering and Management Sciences from Northwestern University. Her research interest centers on ranking and selection procedures for stochastic simulation, optimization via simulation, statistical output analysis, quality control, and applications of simulation methods to environmental management.

Yao Xie is the Harold R. and Mary Anne Nash Early Career Professor and Assistant Professor in the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology. She previously served as a research scientist in the Electrical and Computer Engineering Department at Duke University after receiving her Ph.D. in Electrical Engineering (minor in Mathematics) from Stanford University in 2011. Her research interests are in sequential statistical methods, statistical signal processing, big data analysis, compressed sensing, optimization, with applications to wireless communications, sensor networks, medical and astronomical imaging.