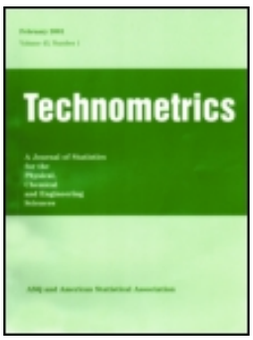

Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: http://www.tandfonline.com/loi/utch20

An Adaptive Sampling Strategy for Online High-Dimensional Process Monitoring

Kaibo Liu, Yajun Mei & Jianjun Shi

To cite this article: Kaibo Liu, Yajun Mei & Jianjun Shi (2015) An Adaptive Sampling Strategy for Online High-Dimensional Process Monitoring, Technometrics, 57:3, 305-319, DOI: 10.1080/00401706.2014.947005

To link to this article: http://dx.doi.org/10.1080/00401706.2014.947005

Accepted author version posted online: 08 Aug 2014.
Published online: 08 Aug 2014.

An Adaptive Sampling Strategy for Online High-Dimensional Process Monitoring

Kaibo Liu

Department of Industrial and Systems Engineering
University of Wisconsin-Madison
Madison, WI 53706
(*kliu8@wisc.edu*)

Yajun Mei and Jianjun Shi

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332
(*yajun.mei@isye.gatech.edu*; *jianjun.shi@isye.gatech.edu*)

Temporally and spatially dense data-rich environments provide unprecedented opportunities and challenges for effective process control. In this article, we propose a systematic and scalable adaptive sampling strategy for online high-dimensional process monitoring in the context of limited resources with only partial information available at each acquisition time. The proposed adaptive sampling strategy includes a broad range of applications: (1) when only a limited number of sensors is available; (2) when only a limited number of sensors can be in "ON" state in a fully deployed sensor network; and (3) when only partial data streams can be analyzed at the fusion center due to limited transmission and processing capabilities even though the full data streams have been acquired remotely. A monitoring scheme of using the sum of top-r local CUSUM statistics is developed and named as "TRAS" (top-r based adaptive sampling), which is scalable and robust in detecting a wide range of possible mean shifts in all directions, when each data stream follows a univariate normal distribution. Two properties of this proposed method are also investigated. Case studies are performed on a hot-forming process and a real solar flare process to illustrate and evaluate the performance of the proposed method.

KEY WORDS: Multiple data streams; Partial information over the spatial domain; Sensor redeployment; Shift detection; Sum of the top-r local statistics.

1. INTRODUCTION

In modern quality control, online monitoring of high-dimensional data streams in real time has become increasingly important due to the pervasive use of multiple sensors to observe a system. Examples of these cases include distributed sensing networks (DSNs) and image sensing devices. Due to the complexity of a system, the amount of data needs to be collected for monitoring becomes extremely large volume and high dimensional. However, with the resource constraints (e.g., limited number of sensors, limited bandwidth or energy constraint, limited transmission and processing time), it often requires us to make a decision based on partial observations during online monitoring. To address this issue, this paper focuses on developing a systematic and scalable adaptive sampling strategy that enables us to actively select the partial "observable" data to maximize the change detection capability of the whole system subject to the resource constraints.

A high-level description of our problem is as follows. Suppose that there are m variables $\mathcal{M} = \{1, \ldots m\}$ of interest. At each acquisition time t, $\mathbf{X}_t = (X_{1,t}, \ldots X_{m,t})'$ denotes the measurement values of these variables. We are interested in detecting a possible change on certain characteristics, say, means or variances, of some components of $\boldsymbol{X}_t$. However, due to resource constraints, only q ($q \le m$) out of m data streams can be observed at each acquisition time, in which the sampling rate is also constrained by the system capacity. Mathematically, let $\delta_{k,t}$ be a binary variable for the observation $X_{k,t}$, such that $\delta_{k,t} = 1$ if and only if $X_{k,t}$ is observed at time t. Then, the resource constraints imply $\sum_{k=1}^{m} \delta_{k,t} = q$ for all time t. The question is how to adaptively choose $\delta_{k,t}$'s at each time and how to use the observed partial data to quickly detect the system change while maintaining a prespecified in-control average run length (ARL) requirement.

To highlight our main ideas, below we will focus on detecting a change on the means of some components of $\mathbf{X}_t$. Specifically, the following assumptions are made: (1) The sampling strategy can be timely implemented without any cost within each sampling interval. (2) When the process is in-control, $\mathbf{X}_t$ is independently and identically distributed (iid) across different time t with a joint distribution function $F(\mathbf{X}_t)$, in which each variable k has a normal probability density function $f(X_{k,t})$. (3) Without losing generality, the in-control mean and the in-control variance of each variable k are assumed to be known as in (Hawkins 1993). In practice, these parameters can be estimated offline from a sufficiently large sample of measurements or known from engineering specifications when designing the process. In addition, a preliminary transformation has been applied to the data, so that each variable has a mean of 0 and a standard deviation of 1. (4) At some finite time γ, there are changes in the means of data streams made at a subset $\mathcal{M}_0 \subset \mathcal{M}$. The distribution of X_t after process changes is denoted as $G(\mathbf{X}_t)$, where each variable k has a normal probability density function $g_k(X_{k,t})$ with mean u_k and variance 1. The change point γ, the subset $\mathcal{M}_0$ and its cardinality, and the magnitude and the di-

TECHNOMETRICS, AUGUST 2015, VOL. 57, NO. 3
DOI: 10.1080/00401706.2014.947005
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/tech*.

rection of the post-change mean u_k are generally unknown. In addition, it is worth mentioning that the problem defined here is called *unstructured* in the sense that we do not assume a spatial structure that relates the data streams to one another (Xie and Siegmund 2013).

The primary contribution of this article is to develop an adaptive sampling strategy over the spatial domain for online high-dimensional process monitoring in the field of statistical process control (SPC). In the SPC literature, adaptive sampling strategies with variable sample sizes and/or sampling intervals are widely used in quality inspection to eliminate the defective units (Montgomery 2009; Li and Qiu 2013). However, these existing adaptive sampling strategies focus on actively selecting partial information over the temporal domain (i.e., decide the data acquisition frequency). In the context of field estimation in the offline spatial statistics, the idea of sequentially adding (reducing) the amount of acquisitions in the spatial domain is applied to explore the correlation between measured samples (Fiorelli et al. 2006; Rahimi et al. 2005). Our application of adaptive sampling strategy over the spatial domain to SPC is new, as we face the need of actively selecting which data streams should be observed in the spatial domain.

From the technical point of view, our proposed adaptive sampling method includes two novel ideas. One is introducing a compensation coefficient to the local data streams for not taking observations, and the other is relocating resources to closely monitor those data streams that most likely involve a change based on past information. These two ideas allow our proposed sampling strategy to behave like a random sampling when the system is in-control and like a greedy sampling strategy when the system is out-of-control. The proposed method is named as top-r based adaptive sampling (TRAS). Our approach is inspired by the monitoring scheme in (Mei 2011), which is based on investigating the sum of top-r local cumulated sum (CUSUM) statistics. However, our approach is different from the one in (Mei 2011) because the monitoring scheme proposed by Mei (2011) assumes that all variables are measurable and independent. Therefore, it cannot be directly employed to the adaptive sampling problem discussed in this article.

The proposed adaptive sampling strategy includes a broad range of applications: (1) when only a limited number of sensors is available during online monitoring (e.g. see Limongelli 2003); (2) when only a limited number of sensors can be in "ON" state in a fully deployed DSN at any given time for data acquisition and transmission purposes (e.g., see Tan et al. 2012); and (3) when only partial data streams can be analyzed at the fusion center due to limited transmission and processing capabilities even though the full data streams have been acquired remotely.

For simplicity, in the remaining of this article, we consider only the scenario that the available data streams are collected by q sensors and we focus on adaptively changing sensor layout for quick change detection. This article is organized as follows: In Section 2, we first review the topics related to our defined problem and then focus on the cumulated sum (CUSUM)-based methodologies for process monitoring. Section 3 proposes an adaptive sampling strategy for online process monitoring and further investigates two properties of the developed method. Parameter settings involved in the algorithm will also be discussed. Section 4 conducts two case studies based on a hot-forming process and also a real solar flare process to test and validate our proposed method in practice. Finally, Section 5 draws a conclusion and discusses the future work.

2. LITERATURE REVIEW

This section contains two subsections. In Section 2.1, we will review a couple of topics that are closely related to our defined problem from the application point of view. Then, in Section 2.2, we will do a quick review of the CUSUM-based methodologies from the methodology point of review, as we will use the CUSUM control chart specifically as a demonstration to monitor each local data stream and then combine them together to produce an adaptive sampling strategy for online high-dimensional process monitoring.

2.1 Review of Related Topics

Our research is related to a couple of topics in the literature from the application point of view. The first one is the optimal design of a sensor system in a DSN. Comprehensive reviews of the state-of-the-art developments in DSNs for quality and productivity improvements can be found in (Ding et al. 2006; Mandroli et al. 2006). The conventional sensor system design often assumes that the sensor layout is fixed during online monitoring. However, numerous studies have shown that variables without sensor deployment will have a much lower detectability and diagnosability (Li and Jin 2010; Liu and Shi 2013). To address this issue, adaptive sensor allocation algorithms have been developed to improve the monitoring and diagnostic capabilities. One of the most recent research efforts has been done by Liu et al. (2013), who integrated the multivariate T^2 control chart and causal structures in a Bayesian network (BN) to develop an adaptive sensor allocation strategy. Their approach can greatly reduce the detection delay and increase the diagnostic accuracy compared with the fixed strategy in (Li and Jin 2010; Liu and Shi 2013). Unfortunately, there are three main limitations in that approach: (1) a BN must be known; (2) it only focuses on single mean shift detection; and (3) the computational complexity is exponential in the number of variables.

Another closely related topic is the theory of searching and tracking targets (Lim et al. 2006; Zoghi and Kahaei 2010; Kagan and Ben-Gal 2013), in which the problem is to study how to most effectively employ limited resources when trying to find an object whose location is not precisely known (Frost and Stone 2001). Although many research efforts have been made on this topic, most of them assume that (1) a single target already exists in the searching space but at some unknown location; and (2) the searching result is binary (Katenka et al. 2008), which can be used to update the location probability distribution of the target over the space by a Bayesian statistics approach. In the case that there are multiple search agents and multiple targets, additional questions arise regarding collective decision making, sharing information, and online communication between the search agents. On the contrary, this article aims at developing an online monitoring scheme, in which the data streams are continuous flows of data. In addition, our method does not require any assumption on the potential failure pat-

terns (i.e., does not assume a prior probability distribution of the change locations), nor information sharing (e.g. considering the current spatial relationships between multiple search agents).

2.2 Review of CUSUM-Based Methodologies for Process Monitoring

From the methodology point of review, our proposed methodologies rely heavily on efficient univariate control charts for monitoring univariate local data streams. To demonstrate our main ideas, we will choose the CUSUM procedure in this article as our baseline univariate control chart, though our ideas can be extended to other efficient univariate control charts such as exponentially weighted moving average (EWMA) and Shewhart charts. Below, we will do a quick review of the CUSUM-based methodologies for univariate data streams and its extension to the multivariate case.

CUSUM-based methodologies are a set of sequential procedures to calculate cumulative sums based on likelihood ratios, which can be used for detecting a shift in a process. The CUSUM procedure was first developed by Page (1954), who proposed to monitor a univariate variable with CUSUM statistic $W_{1,t}$ at time $t(t = 1, 2, \ldots)$:

$$W_{1,t} = \max\left(W_{1,t-1} + \log\frac{g_1(X_{1,t})}{f\left(X_{1,t}\right)}, 0\right) \quad \text{and} \quad W_{1,0} = 0. \tag{1}$$

In the case that $f\left(X_{1,t}\right)$ and $g_1(X_{1,t})$ are normal distribution density functions with common variances 1, and means 0 and u_1, respectively, Equation (1) is reduced to

$$W_{1,t} = \max\left(W_{1,t-1} + u_1 X_{1,t} - \frac{u_1^2}{2}, 0\right) \quad \text{and} \quad W_{1,0} = 0. \tag{2}$$

When the postchange mean u_1 is unknown, the standard approach is to estimate it by the maximum likelihood method (Lorden 1971; Tsui et al. 2012) or replace it with a constant parameter, which represents the interested-smallest magnitude of mean shift to be detected, $u_{\min}$. When $u_{\min} \geq 0$, the CUSUM control chart is only able to detect a positive mean shift. In order to detect the mean shift in both positive and negative directions, the two-sided CUSUM procedures are suggested (Page 1954). In addition, several efforts have been made to combine the two-sided procedures into a single control chart (see Crosier 1986; Cheng and Thaga 2005a).

Since the univariate CUSUM control chart was proposed, numerous contributions have been made to study the CUSUM control chart for the multivariate case. One approach is to directly consider the density function or other properties of multivariate vector observations. Healy (1987) viewed the CUSUM procedures as a series of sequential probability ratio tests and proposed a method for detecting a shift in both the mean vector and the covariance matrix based on a linear combination of individual measurements. Crosier (1988) suggested two multivariate CUSUM procedures, in which the first one is based on the square root of T^2 statistics and the second one is based on a vector-valued CUSUM scheme. Pignatiello and Runger (1990) proposed two similar multivariate CUSUM schemes, which are simpler but have a better ARL performance than the ones in (Crosier 1988). Hawkins (1993) introduced CUSUM procedures based on a set of regression-adjusted variables. Ngai and Zhang (2001) considered a two-sided CUSUM chart for monitoring process mean via projection pursuit method. Chan and Zhang (2001) further extended the method for monitoring the covariance matrix. Qiu and Hawkins (2001, 2003) proposed a rank-based multivariate CUSUM procedure, which is distribution free. Cheng and Thaga (2005b) developed a multivariate max-CUSUM chart to simultaneously monitor the shifts in both mean vector and covariance structure with different directions. Qiu (2008) further proposed a distribution-free multivariate CUSUM procedure based on log-linear modeling, which transfers the multivariate vector observations into a vector with only binary elements. Zou and Qiu (2009) integrated the SPC methods with LASSO-based variable selection algorithms, which can provide a balanced protection against various shifts. A comprehensive introduction to this topic can be found in a recent book by Qiu (2013). However, this research line relies on the estimated joint density function or other properties of multivariate vector observations, and the computational cost is often high, which limits their application to high-dimensional processes monitoring.

A different approach of using CUSUM for the multivariate case is to combine local univariate procedure together into a single global scheme. One intuitive idea is to monitor CUSUMs of individual measurements simultaneously and raise an out-of-control alarm whenever any of the univariate CUSUM charts indicates an out-of-control (i.e., based on the maximum of local CUSUM statistics) (see Woodall and Ncube 1985; Tartakovsky et al. 2006). Specifically, the alarm will be raised at the first time t such that $\max_k W_{k,t} \geq b$. Mathematically, this procedure is defined by the integer-valued stopping time:

$$T_{\max}(b) = \inf\{t \geq 1 : \max_k W_{k,t} \geq b\}, \tag{3}$$

where "inf" is used to denote that $T_{\max}(b) = \infty$ if such t does not exist; and b is a constant number that determines the operating characteristics of the procedure (i.e., maintain a prespecified in-control ARL requirement). The approach of using the maximum of local CUSUM statistics is preferable when the number of root cause variable is one. Another monitoring scheme was proposed by Mei (2010), which is based on the sum of all local CUSUM statistics. The proposed stopping time is

$$T_{\text{sum}}(c) = \inf\{t \geq 1 : \sum_{k=1}^{m} W_{k,t} \geq c\}, \tag{4}$$

where c is a suitable constant number like b in Equation (3). This procedure is most effective when multiple variables simultaneously shift. Indeed, $T_{\text{sum}}(c)$ is a maximum likelihood ratio-based procedure when changes are allowed to occur at different times for different data streams.

Recently, Mei (2011) further extended $T_{\text{sum}}(c)$ to propose another global monitoring scheme based on the sum of the largest r local CUSUM statistics. This is a maximum likelihood ratio-based procedure under the additional assumptions that different data streams have different change-points and at most r data streams are affected. Note that $T_{\max}(b)$ in Equation (3) and

$T_{\text{sum}}(c)$ in Equation (4) can be regarded as the special cases of using the sum of top-r local CUSUM statistics when $r = 1$ and $r = m$, respectively.

3. ADAPTIVE SAMPLING STRATEGY

When dealing with a process monitoring problem in the context of limited resources, there are two intuitive ways of choosing the sample strategy. The first one is to (randomly or purposely) fix q out of m data streams to monitor the process. However, the approach relies on that the set of q data streams and the set $\mathcal{M}_0$ (the set of variables with shift) intersect; otherwise, it will completely miss to detect the change. Considering an example with only $m = 1000$ variables, 5 shift variables, and $q = 5$ sensors randomly distributed, the probability that these two sets interact is only 0.0248. As the dimension parameter m becomes larger, this probability will become even smaller, and thus this approach (pre-fix q out of m data streams) is ineffective to monitor the process change in high-dimensional settings when the potential failure locations are unknown. The second approach is called random sampling, which will randomly choose q of m data streams to monitor at each time. Although this idea is straightforward, it is surprising that both the application and the performance of this approach (i.e. randomly choose q out of m data streams to monitor the process) to high-dimensional settings has not been fully studied yet in the SPC literature. By using the random sampling algorithm, the probability that these two sets never interact to each other will decrease as time goes. This provides us a confidence that the shift variables will be monitored at certain time. Unfortunately, an obvious limitation of this random sampling approach is that if the monitored variable with change is not significantly large to trigger an out-of-control alarm immediately, then it will likely switch to other variables to monitor at next time step, which deteriorates the detection performance. Thus, one interesting question here is whether we can develop a more intelligent sampling strategy that can integrate the advantages of these two intuitive sampling approaches in an automatic way. If so, the new method would be able to quickly detect a wide range of possible changes with no prior knowledge of the potential failure patterns by adaptively and automatically adjusting to and then sticking to sampling the shift variables.

To tackle this challenging question, we propose to extend the idea of top-r local CUSUM statistics in Mei (2011) for effective monitoring of the high-dimensional process in the context of limited resources. In the proposed approach, we first monitor each data stream locally through some classical computationally simple but efficient univariate control charts, and then combine all local univariate control charts together to produce a single global monitoring statistic. There are two novel ideas in the proposed strategy: (1) introducing a compensation coefficient Δ to the local statistics for not taking observations, and (2) focusing on only the r most significant local univariate control charts and considering the sum of the corresponding local statistics (in the log likelihood scale) for some suitable choice of r.

The reason of introducing the compensation coefficient Δ is to make sure that no data streams will be left unobserved for a long period when the system is in-control and thus the proposed method will behave similarly to the random sampling method. On the other hand, the sum of top-r local statistics will smooth out the noise when at most r data streams are affected and thus our proposed algorithm will become a greedy algorithm and eventually stick to sampling those out-of-control data streams. In other words, our proposed method can be regarded as a combination of the two intuitive sampling approaches, random sampling when the system is in-control and fixed (greedy) sampling at the shift variables when the system is out-of-control. The novelty of our method is that it can integrate these two intuitive sampling approaches automatically.

We mentioned before that the proposed TRAS algorithm will be an effective tool for monitoring of high-dimensional process. Here, we would like to further clarify the meaning of the word "effective" in our context. First, "effective" should be evaluated by comparing with a baseline model, which is the random sampling method, rather than the SPC/CUSUM method that knows exactly where the changes will occur. This is because the core challenge here is "how to manage the limited monitoring resource" rather than a traditional SPC problem which is "how to monitor a specified target." Thus, the effectiveness of an algorithm or control chart should be compared with a baseline method (random sampling) in term of how well they can manage the limited monitoring resource for SPC purpose. In this regards, the "absolute" performance of our proposed TRAS algorithm on some extreme conditions might not be numerically appealing, but it can still be promising if the baseline (random sampling) method performs much worse. For instance, let us consider an extreme example in which we want to detect a mean shift in $m = 10^5$ variables but only with one resource $q = 1$. The detection delays of all reasonable SPC methods will be much larger as compared to those SPC/CUSUM methods that have a precise knowledge about which variable will be changed, and thus such approach will lead to an unfair comparison. Second, in this paper, from an asymptotic point of view (see Section 3.2 for more details), we will demonstrate that the proposed TRAS algorithm will lead to an "effective" detection of the change variable in the long run. However, since the asymptotic results do not always translate into empirical properties due to slow convergence rate, we acknowledge that it would be an interesting question to investigate the practical "effectiveness" of the proposed method, which will be studied as a future study.

Below, we will present the details of the proposed method in Section 3.1 and further discuss its properties in Section 3.2. Parameter settings of the proposed method are discussed in Section 3.3.

3.1 Methodology Development

There are three essential components in our proposed adaptive sampling strategy: (1) how to construct local statistics with dynamic partial observations; (2) when to indicate process is out-of-control; and (3) how to redistribute sensor layout. Below we will explain each of them in detail and then provide an overview of our proposed method.

3.1.1 Local Statistics. In this article, we are interested in detecting both positive and negative mean shifts in any variable. Thus, a two-sided local statistic for each variable k at time t is defined as (see Cheng and Thaga 2005a, b):

$$W_{k,t} = \max\left(W_{k,t}^{(1)}, W_{k,t}^{(2)}\right), \tag{5}$$

where $W_{k,t}^{(1)}$ and $W_{k,t}^{(2)}$ are the local statistics for detecting positive and negative mean shifts, respectively.

Due to limited resources, only q data streams are available at each acquisition time. Therefore, depending on whether a sensor is deployed on the variable k, the adaptive sampling strategy focuses on defining how to construct the local statistics, $W_{k,t}^{(1)}$ and $W_{k,t}^{(2)}$. In the scenario that a sensor is deployed on variable k at time t, we use the following equation to calculate the local statistics based on Equation (2) (for $t = 1, 2, \ldots$):

$$W_{k,t}^{(1)} = \max\left(W_{k,t-1}^{(1)} + u_{\min}X_{k,t} - \frac{u_{\min}^2}{2}, 0\right)$$

and

$$W_{k,t}^{(2)} = \max\left(W_{k,t-1}^{(2)} - u_{\min}X_{k,t} - \frac{u_{\min}^2}{2}, 0\right), \qquad (6)$$

where $W_{k,0}^{(1)} = W_{k,0}^{(2)} = 0$. Since the postchange mean u_k in Equation (2) is usually unknown, it has been replaced by a constant positive parameter $u_{\min}$, the interested-smallest magnitude of mean shift to be detected, as introduced in Section 2 (see the guidelines in Section 3.3 on how to determine the value for $u_{\min}$).

On the other hand, when a sensor is not deployed on variable k at time t, we construct the local statistics by introducing a compensation coefficient $\Delta \geq 0$:

$$W_{k,t}^{(1)} = W_{k,t-1}^{(1)} + \Delta \quad \text{and} \quad W_{k,t}^{(2)} = W_{k,t-1}^{(2)} + \Delta, \qquad (7)$$

where Δ is a constant tuning parameter for not taking observations when updating the local statistics (see the guidelines in Section 3.3 on how to determine the value for Δ).

It is worth mentioning that the two-sided local statistic $W_{k,t}$ is always nonnegative for all variables k and all time t based on Equations (5)–(7), and it will be large no matter whether the process has a positive or negative mean shift in variable k. In the following study, we will assume that the constructed local statistics are all two-sided, unless otherwise specified.

3.1.2 Stopping Time. Let $W_{(1),t} \geq \cdots \geq W_{(k),t} \geq \cdots \geq W_{(m),t}$ denote the decreasing order statistics of $\{W_{k,t}, k = 1, 2, \ldots, m\}$. Inspired by Mei (2011), the stopping time of the monitoring procedure can be determined as

$$T_{\text{top},r}(d) = \inf\left\{t \geq 1 : \sum_{k=1}^{r} W_{(k),t} \geq d\right\}, \qquad (8)$$

where $W_{(k),t}$ denotes the kth largest local statistics and d is a suitable constant number like b and c in Equations (3) and (4). Empirical results have shown that using the sum of top-r local CUSUM statistics is sensitive and robust to detect process change with suitable choices of r (Mei 2011) (see the guidelines in Section 3.3 on how to determine the values for r and d).

3.1.3 Redistribute Sensor Layout. We let $l_{(k),t}$ denote the variable index of the decreasing order statistic $W_{(k),t}$. Since a larger local statistic indicates that the variable is more likely to have a mean shift, we will redistribute sensor resources onto the variables with the first q largest local statistics after checking the stopping rule at each time t:

$$S = \{l_{(1),t}, \ldots, l_{(q),t}\}, \qquad (9)$$

where S denotes the new sensor layout. In the case that there are more than one variable achieving the same statistic value as $W_{(q),t}$, which results in a nonunique solution for S, we will randomly choose one of the solutions to update S.

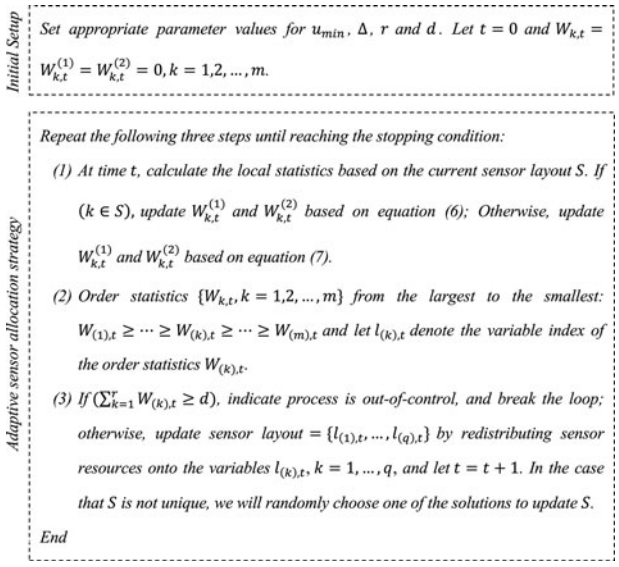

Figure 1. The overall flow chart of the TRAS algorithm.

3.1.4 Overview of the TRAS Algorithm. The initial sensor layout is not critical since the adaptive sampling strategy will be updated at each acquisition time according to the online measurements. An overview of the TRAS algorithm is illustrated in Figure 1.

Recall that no matter whether the process has a positive or negative mean shift in variable k, the corresponding local statistic $W_{k,t}$ will become large. Thus, our proposed method is *robust* in the sense that it is able to detect a wide range of possible shifts in all directions without making any assumption on the spatial structure of variables or requiring any prior knowledge on the potential failure pattern. On the other hand, since only m local statistics are recursively calculated at each time epoch, the computational cost is linear in the number of variables. Therefore, the proposed method is *scalable* for online monitoring of a large number of data streams. It is interesting to compare our proposed approach with the existing multivariate SPC methods from the spatial-temporal viewpoint. The existing multivariate SPC methods (e.g., T^2 control chart) often first look at the spatial domain and then consider the temporal domain. Our proposed approach switches the spatial and temporal domains and thus leads to a scalable scheme with significantly reduced computational complexity. Furthermore, since the proposed method has a recursive formula, in which the next sampling strategy only depends on the current information collected at the fusion center (i.e., the algorithm satisfies Markov property), it can save much cost and time for storing large amount of data.

3.2 Properties of TRAS Algorithm

In this section, we provide two properties associated with the TRAS algorithm. These two properties ensure that the TRAS algorithm can not only quickly detect a wide range of possible shifts in all directions, but also identify the shift variables under certain constraints. In the properties, the following variable plays an important role, which characterizes the difference

between the true postchange mean u_k and our preassigned parameters $u_{\min}$ and Δ:

$$\rho_k = u_{\min} u_k - \frac{u_{\min}^2}{2} - \Delta. \tag{10}$$

Recall that $\mathcal{M} = \{1, \ldots, m\}$. We first investigate the property when there is no mean shift or only small mean shifts occurring in the system (i.e. $\rho_k \leq 0$ for all variables k from the set $\mathcal{M}$).

Property 1. Assume $\rho_k \leq 0$ for all variables k from the set $\mathcal{M}$. Let set U denote those $k \in \mathcal{M}$ such that a sensor will never be redistributed to the variable k after some finite time t_0. Then as $d \to \infty$, $P(U = \emptyset) \to 1$, where $\emptyset$ represents the empty set (proof is in Appendix A).

The first property discusses the randomness of sensor layout when either the system is in-control or only small mean shifts occur. When $\rho_k \leq 0$, a sensor will not stick to the variable k, but will be redistributed to other variables with infinite number of times. In other words, sensors will not stick to a certain layout and any variable will not be always left unattended though only a limited number of variables can be observable at any given time.

Let card$(*)$ denote the cardinality of the set $*$. Next, we investigate the property when modest mean shifts occur in the system (i.e. there exists some variable k such that $\rho_k > 0$).

Property 2. Let $B = \{k \in \mathcal{M} : \rho_k > 0\}$. Then, for any finite time t, once a sensor is deployed on the variable $k \in B$ at time t, there is a nonzero probability such that the sensor will never be redistributed to other variables, as $d \to \infty$ (proof is in Appendix B).

The second property indicates that sensor resources will eventually stick to the set of variables with modest mean shifts. According to Property 1, we have already shown that sensor resources will not stick to the variables in $\mathcal{M} \backslash B$, and thus they must be redistributed to the variables in B at some time. Since once a sensor is deployed on the variable in B, there is always a nonzero probability such that the sensor will never be redistributed to other variables. Therefore, sensor resources will eventually stick to the variables that belong to B. It worth mentioning that when the number of sensors $q \geq$ card(B), all variables in B will eventually have sensor deployed and the remaining sensors are redistributed among the variables in $\mathcal{M} \backslash B$. As a result, this property also indicates that the TRAS algorithm is able to localize the process changes. However, since the TRAS algorithm is developed on the basis of individual local information, it can only identify the shift variables instead of the root causes.

From the above two properties, one may ask how our proposed TRAS algorithm jumps from the state of random sampling in property 1 to the greedy sampling in property 2? In other words, what opportunities the proposed TRAS algorithm actually capitalizes on while the random sampling strategy misses? When the process is out-of-control and our proposed algorithm has not observed any of the change variables yet (i.e. behave as the random sampling), the variables that are not currently monitored will gain higher probability to be monitored in the next iteration by the introduction of the parameter Δ. This implies that our proposed algorithm is indeed better than the random sampling method since it encourages the method to move towards the change variables. On the other hand, when our proposed algorithm has observed some change variables, the change variables will have higher probability to still be monitored in the next iteration since the local statistics of change variables will be aggregated in a larger magnitude than the others not being monitored by using the CUSUM approach. In summary, as compared to the random sampling method, our TRAS algorithm is good at the following two things: (i) when change occurs, the method is intended to monitor the change variables rather than randomly walking over them; (ii) when the method happens to capture the change variables, it has the capability to stick to these variables.

3.3 Parameter Settings

In this section, we will discuss how to set the values of four parameters, $u_{\min}$,Δ, r, and d in the TRAS algorithm.

1. The selection of $u_{\min}$: As discussed in Section 3.1.1, the parameter $u_{\min}$ represents the interested-smallest magnitude of mean shift to be detected. In reality, the practitioner can determine the value of $u_{\min}$ based on the engineering domain knowledge to ensure production yield. For example, in a manufacturing process, the tolerance specifications for each variable specified in product/process design can be used to determine the interested-smallest magnitude of mean shift, $u_{\min}$ (Liu and Shi 2013).
2. The selection of Δ: $\Delta \geq 0$ is a tuning parameter, which is associated with how frequently the sensor resources will be reallocated. In practice, Δ cannot be either too large or too small. If Δ is too large, then the reallocation scheme will be dominated by Δ. As a result, a sensor will be redistributed onto the variable that currently does not have a sensor deployed after each time no matter whether this variable has a shift or not. By property 2, if we want to detect and localize the shift from variable k, then $\rho_k > 0$ or equivalently, $u_{min}|u_k| - \frac{u_{\min}^2}{2} > \Delta$. On the other hand, according to property 1, if Δ is too small, then the monitoring system may take a longer time to reallocate sensors from the in-control variables onto the shift variables.
3. The selection of r: At first, r should be smaller than q, the total number of sensors. Otherwise, the sum of top-r local statistics will include unobserved variables when the system is out-of-control, which will degrade the performance of the control chart. Ideally, r should be chosen approximately equal to the total number of root cause variables. On the other hand, Mei (2011) showed that when the total number of root cause variables is unknown, the monitoring scheme with relatively small r value is more robust to detect a wide range of possible shifts.
4. The selection of d: d is the threshold to stop the monitoring procedure. The practitioner can determine the optimal d value from sufficiently large in-control measurements or via Monte Carlo simulation and bootstrap technique (Efron and Tibshirani 1993; Chatterjee and Qiu 2009). The value of d is related to the prescribed in-control ARL of the monitoring scheme, ARL_U, when no change occurs in the system.

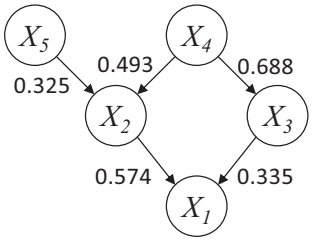

Figure 2. BN structure of a hot-forming process (Li and Jin 2010).

4. CASE STUDIES

4.1 Hot-Forming Process

In this section, we will evaluate the performance of the proposed TRAS algorithm and compare it with the existing one in Liu et al. (2013) under different shift scenarios based on a hot-forming process. A Bayesian network (BN) for the hot-forming process was identified by Li and Jin (2010) and is shown in Figure 2. There are four process variables [4: temperature; 3: material flow stress; 2: tension in workpiece; and 5: blank holding force (BHF)] and one quality variable (1: final dimension of workpiece). Figure 3 illustrates a two-dimensional (2-D) physical illustration of the hot-forming process. All variables are assumed to follow standard normal distribution when the system is under normal operation condition. In addition, the linear Gaussian parameterization of a BN is assumed to be known (Li and Jin 2010):

$$X_i = \sum_{k=1}^{\mathrm{card}(\mathrm{PA}(i))} p\left(\mathrm{PA}_k(i), i\right) X_{\mathrm{PA}_k(i)} + V_i,$$

where $\mathrm{PA}(i)$ denotes the parents of i; $p\left(\mathrm{PA}_k(i), i\right)$ is called the *path coefficient* (the number annotated on the arc of the BN), which refers to the conditional probability $p(X_i|X_{\mathrm{PA}_k(i)})$, and also the causal influence from the k th parent of i to i; $V_i \sim N(0, \sigma_i^2)$ represents the random noise that cannot be described by the linear part of the model, and is assumed to be independent to $X_{\mathrm{PA}_k(i)}$ and $V_j (j \neq i)$.

As mentioned in Section 1, the method in Liu et al. (2013) is based on integrating the multivariate T^2 control chart and the causal structures [in the following pages, we denote this algorithm as causation-based adaptive (CBA) algorithm] when only partial information is available. However, it has four limitations when it is used in practice. First, it assumes that a priori knowledge about the BN for the system is available. Second, it assumes that only a single mean shift may occur in the system until it is detected by the control chart. Third, the computational cost is exponential in the total number of variables m, which makes it difficult to be implemented online, especially for large-volume and high-dimensional datasets. Fourth, it only takes the spatial relationship of measurements into consideration, but ignores the temporal relationship when detecting process changes. Hence, it still results in a relatively long detection delay. To demonstrate the effectiveness of our proposed method, in the following case studies, we assume that the BN is unknown when implementing the TRAS algorithm, whereas the BN is available when implementing the CBA algorithm in phase II monitoring and diagnosis.

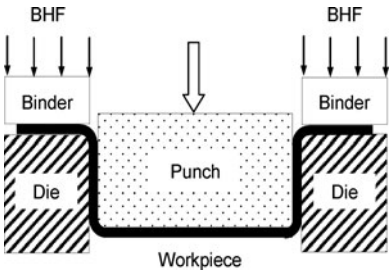

Figure 3. Two-dimensional illustration of the hot-forming process (Li and Jin 2010).

4.1.1 Single Mean Shift Case. In this case study, we focus on comparing the performance of the proposed TRAS algorithm with the CBA algorithm when only a single mean shift may occur at any variable in the hot-forming process. Specifically, the following discussions will consider five potential single mean shift scenarios with different shift magnitudes, $\tau = 1.5, 2, 2.5, 3$ and different number of available sensors, $q = 2, 3, 4$. In addition, the in-control ARL, ARL_U, is chosen to be 100. Since the TRAS algorithm can only identify the shift variables, it is not suitable for root cause diagnosis, especially when the causal structure is unknown. However, since only a single mean shift may occur in the system and the shift will propagate and dilute along the BN (e.g. a mean shift in X_2 with $u_2 = 1$ will result in a mean shift in X_1 with $u_1 = 0.574$), we will use a simple rule of root cause diagnosis for the TRAS algorithm: the variable with the largest CUSUM statistic is the root cause variable. This diagnostic approach can demonstrate how many times the out-of-control alarm is triggered with the largest contribution from the root cause variable. To evaluate the monitoring and the diagnostic capabilities, we specifically focus on the following metrics when each time the same amount of mean shift occurs at each variable: (1) the maximum out-of-control ARL, $\mathrm{RL}_{\max}$; (2) the average out-of-control ARL, $\overline{\overline{\mathrm{RL}}}$; (3) the minimum rate of *uniquely correct diagnosis*, $\mathrm{UF}_{\min}$; and (4) the average rate of *uniquely correct diagnosis*, $\overline{\overline{\mathrm{UF}}}$. A diagnosis result is called *uniquely correct* if the true mean shift variable is uniquely identified. According to the definitions of these metrics (see Appendix C for details), $\mathrm{RL}_{\max}$ and $\overline{\overline{\mathrm{RL}}}$ are desired to be as small as possible, whereas $\mathrm{UF}_{\min}$ and $\overline{\overline{\mathrm{UF}}}$ are desired to be as large as possible. Parameter settings (e.g. values for τ, q, and α) and performance evaluation metrics (e.g. $\mathrm{RL}_{\max}$, $\overline{\overline{\mathrm{RL}}}$, $\mathrm{UF}_{\min}$ and $\overline{\overline{\mathrm{UF}}}$) are selected to be consistent with (Liu et al. 2013).

The TRAS algorithm is implemented during online monitoring no matter when the system is in-control or out-of-control. Thus, the sensor layout is not unique at the moment when a mean shift occurs, and it can affect the monitoring performance. In order to address this issue, we consider the *worst initial sensor deployment (WISD)*, which will provide the least favorable sensor layout for detection when a mean shift occurs in the system. Define $\{\rho_{(k)}\}$ as the increasing order statistics of $\{\rho_k\}$. In other words, when a mean shift occurs at variable k, we assume that sensors are deployed right on the variables associated with $\{\rho_{(1)}, \ldots, \rho_{(q)}\}$, no matter how sensor placement has evolved be-

fore. In this way, we can get the most conservative performance evaluation of the TRAS algorithm.

Table 1 summarizes the performance comparisons between the TRAS algorithm and the CBA algorithm under different combinations of mean shift magnitude τ, compensation coefficient Δ, and number of available sensors q. In this study, we choose $r = 2$ and $u_{\min} = 1.5$ according to the characteristics of the actual system. The detailed simulation steps can be found in Appendix C. The numbers in the "CBA" column can be referred to (Liu et al. 2013). The column "Impv." shows the percentage of improvement (i.e., positive value) or deterioration (i.e., negative value) in each performance metric by implementing the TRAS algorithm ($\Delta = 0.1$) over the CBA algorithm.

According to Table 1, we can draw the following conclusions:

1. The TRAS algorithm outperforms the CBA algorithm in both monitoring and diagnostic capabilities. In most of the scenarios, the TRAS algorithm can significantly improve the detection delay (which is related to $\mathrm{RL}_{\max}$ and $\overline{\overline{\mathrm{RL}}}$ metrics) and also the diagnostic accuracy (which is related to $\mathrm{UF}_{\min}$ and $\overline{\overline{\mathrm{UF}}}$ metrics) compared with the CBA algorithm, although the causal information is unknown when implementing the TRAS algorithm. On the other hand, the differences between the $\mathrm{RL}_{\max}$ and the $\overline{\overline{\mathrm{RL}}}$ metrics are much smaller for the TRAS algorithm under different mean shift scenarios. This characteristic ensures that a mean shift can be detected in a timely manner no matter where the root cause is. In this way, a more robust monitoring scheme is obtained, and thus less defective products will be produced due to quick detection of process changes.
2. The performance of the TRAS algorithm is relatively stable when Δ changes within a certain range. However, when Δ is set to be a very large number (e.g. $\Delta = 1.5$: in this way, $\rho_k < 0$ for $u_k = 1.5$), the performance of the TRAS algorithm deteriorates very fast, especially when the number of available sensors is small (due to page limits, the result is omitted). On the other hand, the system associated with different number of available sensors need different optimal Δ values. For example, when $q = 2$, $\Delta = 0.1$ provides the best performance compared with $\Delta = 0.01$ or 0.5. On the contrary, when $q = 3$ or 4, $\Delta = 0.5$ provides the best performance compared with $\Delta = 0.01$ or 0.1. According to the empirical study, as the number of available sensors increases, a larger Δ value is preferred.
3. As τ increases, the detection delay of both methods decreases, which is consistent with our intuitions; however, the amount of improvement in the "Impv." columns becomes smaller. It is known that the CUSUM-type chart is more sensitive to detect small mean shifts than the Shewhart chart (Montgomery 2009). However, in this example, even with large mean shifts, the detection delay of the TRAS algorithm (which is based on the CUSUM statistics) is still comparable to the CBA algorithm (which is based on T^2 statistics). On the other hand, the diagnostic accuracy of both methods becomes smaller as τ increases. This is because as the mean shift becomes significant, the abnormality can be more easily noticed by the system even with a poorly deployed sensor layout.
4. As q increases, less information is lost in the system, and thus the detection delay of both methods decreases. In addition, the amount of improvements in both the detection delay and the diagnostic accuracy of the TRAS algorithm become less pronounced. As a result, the TRAS algorithm is more sensitive and robust than the CBA algorithm for process change detection, especially when the number of available sensors is small.

4.1.2 Multiple Mean Shifts Case. The CBA algorithm is only able to detect single mean shift, but the TRAS algorithm can detect a wide range of possible shifts in all directions. Consequently, in this study, we focus on evaluating the performance of the TRAS algorithm when multiple mean shifts occur in the network. In addition, we are interested in studying the effect of sensor layout on the detection delay of the TRAS algorithm when the mean shifts right occur in the system. Specifically, two approaches to the initial sensor deployment are considered in the following analysis. The first one is called the WISD as introduced in 4.1.1, and the second approach is called *random initial sensor deployment (RISD)*, which will randomly distribute q sensors to the variables when mean shifts right occur in the system. By comparing these two approaches, we can get a thorough study about the robustness of the TRAS algorithm to the initial sensor layout. Similarly to the single mean shift case, different mean shift magnitudes τ and different number of available sensors q are considered. In this study, Δ is chosen to be 0.1. The evaluation process is similar to the single mean shift case in Section 4.1.1, and thus it is omitted here.

Tables 2 and 3 elaborate the performance of the TRAS algorithm when two variables simultaneously have a mean shift with identical magnitude but in the same and the opposite directions, respectively. According to the results in Tables 2 and 3, we can get the following conclusions:

1. As q and τ increase, less information is lost during online monitoring and the mean shift becomes more significant, and thus detection delay decreases. In addition, since mean shifts with different directions will mitigate each other when propagating to downstream variables, the detection delay in Table 3 is larger than that in Table 2.
2. Compared with the RISD, the WISD only has a little larger detection delay in both $\mathrm{RL}_{\max}$ and $\overline{\overline{\mathrm{RL}}}$ metrics (the difference is within one ARL in almost all scenarios). Thus, this study shows that the sensor layout at the moment when a mean shift occurs in the system has little effect on the monitoring performance. In other words, the TRAS algorithm is able to timely update the sensor layout for detecting process changes.

4.2 Solar Flare Detection

In this section, we conduct a case study based on a real dataset collected from the solar data observatory, which illustrates the occurrences of solar flares. A solar flare is defined as a sudden, transient, and intense variation in brightness, which is usually observed over the Sun's surface (Augusto et al. 2011). A solar

Table 1. Performance comparisons between the TRAS algorithm and the CBA algorithm under different combinations of Δ, τ, and q values for single mean shift case

	CBA	TRAS ($\Delta = 0.01$)	TRAS ($\Delta = 0.1$)	TRAS ($\Delta = 0.5$)	Impv.	CBA	TRAS ($\Delta = 0.01$)	TRAS ($\Delta = 0.1$)	TRAS ($\Delta = 0.5$)	Impv.
	$q = 2$ and $\tau = 1.5$					$q = 2$ and $\tau = 2$				
RL_{max}	27.24	9.03	8.77	10.36	67.80%	15.15	6.45	6.19	7.25	59.14%
$\overline{\overline{RL}}$	17.66	8.40	8.17	9.93	53.74%	9.06	5.98	5.79	6.95	36.09%
UF_{min}	0	0.56	0.60	0.75	–	0	0.49	0.54	0.73	-
$\overline{\overline{UF}}$	0.40	0.83	0.84	0.92	110%	0.45	0.82	0.84	0.92	86.67%
	$q = 2$ and $\tau = 2.5$					$q = 2$ and $\tau = 3$				
RL_{max}	9.16	5.25	5.05	5.83	44.87%	5.69	4.67	4.41	5.01	22.50%
$\overline{\overline{RL}}$	5.38	4.90	4.74	5.58	11.90%	3.66	4.32	4.16	4.81	−13.66%
UF_{min}	0	0.41	0.45	0.66	-	0	0.33	0.36	0.54	-
$\overline{\overline{UF}}$	0.47	0.80	0.82	0.91	74.47%	0.48	0.77	0.79	0.88	64.58%
	$q = 3$ and $\tau = 1.5$					$q = 3$ and $\tau = 2$				
RL_{max}	17.75	7.88	7.72	7.65	56.51%	8.67	5.51	5.40	5.33	37.72%
$\overline{\overline{RL}}$	13.03	7.27	7.16	7.10	45.05%	6.26	5.10	5.01	4.97	19.97%
UF_{min}	0.20	0.64	0.67	0.73	235%	0.23	0.58	0.62	0.74	169.57%
$\overline{\overline{UF}}$	0.61	0.87	0.88	0.90	44.26%	0.69	0.87	0.88	0.92	27.54%
	$q = 3$ and $\tau = 2.5$					$q = 3$ and $\tau = 3$				
RL_{max}	5.06	4.43	4.36	4.28	13.83%	3.50	3.53	3.76	3.69	−7.43%
$\overline{\overline{RL}}$	3.76	4.10	4.03	3.99	−7.18%	2.68	3.86	3.46	3.42	−29.1%
UF_{min}	0.22	0.52	0.55	0.69	150%	0.18	0.44	0.46	0.61	155.56%
$\overline{\overline{UF}}$	0.68	0.86	0.87	0.92	27.94%	0.64	0.84	0.86	0.91	34.38%
	$q = 4$ and $\tau = 1.5$					$q = 4$ and $\tau = 2$				
RL_{max}	17.44	7.51	7.36	7.14	57.8%	8.35	5.22	5.16	5.03	38.2%
$\overline{\overline{RL}}$	12.08	6.85	6.75	6.55	44.12%	5.86	4.74	4.69	4.57	19.97%
UF_{min}	0.56	0.63	0.65	0.71	16.07%	0.42	0.56	0.59	0.67	40.48%
$\overline{\overline{UF}}$	0.72	0.86	0.87	0.89	20.83%	0.77	0.85	0.86	0.89	11.69%
	$q = 4$ and $\tau = 2.5$					$q = 4$ and $\tau = 3$				
RL_{max}	4.70	4.22	4.14	4.03	11.91%	3.27	3.59	3.57	3.47	−9.17%
$\overline{\overline{RL}}$	3.46	3.76	3.72	3.64	−7.51%	2.47	3.17	3.15	3.10	−27.53%
UF_{min}	0.20	0.46	0.49	0.59	145%	0.05	0.36	0.39	0.48	680%
$\overline{\overline{UF}}$	0.75	0.82	0.83	0.87	10.67%	0.72	0.79	0.80	0.84	11.11%

flare can emit large energetic charged particles, which may disturb the Earth's ionosphere and radio communications. Thus, there is a pressing need to detect the solar flare as soon as possible. At each second, tons of solar flare images will be captured and generated by the satellite. However, due to the transient characteristic of the solar flare process and large amount of dataset, traditional methodologies for image change detection (Radke et al. 2005; Qiu 2005) by analyzing the full data stream can usually exceed the transmission and processing capabilities during online monitoring, and thus are incapable of detecting solar flares in real time.

Table 2. Performance evaluations of the TRAS algorithm under different combinations of initial sensor layouts, τ and q values for multiple (two) mean shifts with same direction

	TRAS (WISD)	TRAS (RISD)	TRAS (WISD)	TRAS (RISD)	TRAS (WISD)	TRAS (RISD)	TRAS (WISD)	TRAS (RISD)
$q = 2$	$\tau = 1.5$		$\tau = 2$		$\tau = 2.5$		$\tau = 3$	
RL_{max}	5.98	5.20	4.51	3.61	3.86	2.88	3.45	2.41
$\overline{\overline{RL}}$	5.05	4.24	3.94	3.05	3.39	2.46	3.05	2.10
$q = 3$	$\tau = 1.5$		$\tau = 2$		$\tau = 2.5$		$\tau = 3$	
RL_{max}	5.11	4.43	3.75	3.06	3.13	2.38	2.77	2.01
$\overline{\overline{RL}}$	4.24	3.63	3.20	2.58	2.68	2.05	2.36	1.74
$q = 4$	$\tau = 1.5$		$\tau = 2$		$\tau = 2.5$		$\tau = 3$	
RL_{max}	4.37	4.12	3.04	2.84	2.43	2.21	2.09	1.85
$\overline{\overline{RL}}$	3.75	3.39	2.73	2.39	2.21	1.89	1.90	1.58

Table 3. Performance evaluations of the TRAS algorithm under different combinations of initial sensor layouts, τ and q values for multiple (two) mean shifts with different directions

	TRAS (WISD)	TRAS (RISD)	TRAS (WISD)	TRAS (RISD)	TRAS (WISD)	TRAS (RISD)	TRAS (WISD)	TRAS (RISD)
$q = 2$	$\tau = 1.5$		$\tau = 2$		$\tau = 2.5$		$\tau = 3$	
RL_{max}	8.25	7.59	5.83	5.07	4.78	3.97	4.26	3.43
$\overline{\overline{RL}}$	7.00	6.31	5.08	4.26	4.22	3.34	3.76	2.83
$q = 3$	$\tau = 1.5$		$\tau = 2$		$\tau = 2.5$		$\tau = 3$	
RL_{max}	7.13	6.49	4.94	4.27	3.98	3.24	3.45	2.71
$\overline{\overline{RL}}$	5.96	5.33	4.23	3.56	3.44	2.75	3.01	2.30
$q = 4$	$\tau = 1.5$		$\tau = 2$		$\tau = 2.5$		$\tau = 3$	
RL_{max}	6.65	6.01	4.59	3.94	3.61	2.99	3.10	2.44
$\overline{\overline{RL}}$	5.30	4.90	3.68	3.27	2.91	2.52	2.46	2.08

The dataset is recorded in a video format and is publicly available online at *http://nislab.ee.duke.edu/MOUSSE/index.html*. There are in total 300 frames in the video, each of which contains a size of $232 \times 292 = 67744$ dimensional online data. According to the video, there are at least two obvious transient flares which occur at frames $t = 187 \sim 202$ and $t = 216 \sim 268$, respectively. For monitoring purposes, the background information has been already removed and the remaining data is approximately normally distributed as mentioned in (Xie et al. 2013).

In this study, the parameters are selected as $u_{\min} = 3$, $\Delta = 0.1$, $r = 40$ (we have also tried other combinations of the parameters, and achieved a similar result). In addition, we assume that only 2000 out of 67,744 pixels are available (i.e. $q = 2000$) at each data frame, and can be sent back to the fusion center for analysis due to limited transmission and processing capabilities.

Figure 4(a) and Figure 5(a) show the snapshots of the video at frames $t = 198$ and $t = 230$, when the first and second solar flares are brightest, respectively. Figure 4(b) and Figure 5(b) illustrate the snapshots at frames $t = 202$ and $t = 268$, when the first and second solar flares are nearly over, respectively. Figure 4(c) is the snapshot at frame $t = 186$, which is the moment right before the first solar flare occurs, and Figure 5(c) shows the snapshot at frame $t = 300$, which is the last frame of this video. Figures 4(d)–(f) and Figures 5(d)–(f) demonstrate the locations of sampled data streams which are marked by white dots at the corresponding frames $t = 198, 202, 186$, and $t = 230, 268, 300$, respectively.

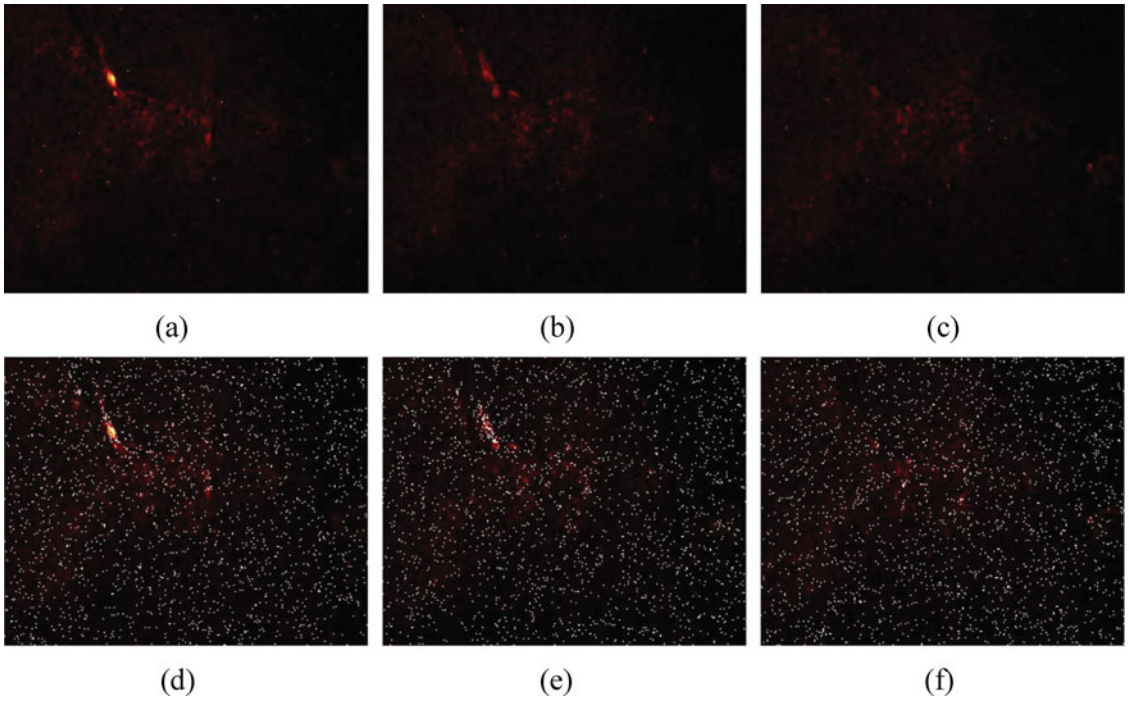

Figure 4. Detection of the first solar flare: snapshots of the video (a) at frame $t = 198$, when the first solar flare is brightest; (b) at frame $t = 202$, when the first solar flare is nearly over; and (c) at frame $t = 186$, the moment right before the first solar flare starts. The locations of 2000 sampled data streams are (d) at frame $t = 198$; (e) at frame $t = 202$; and (f) at frame $t = 186$.

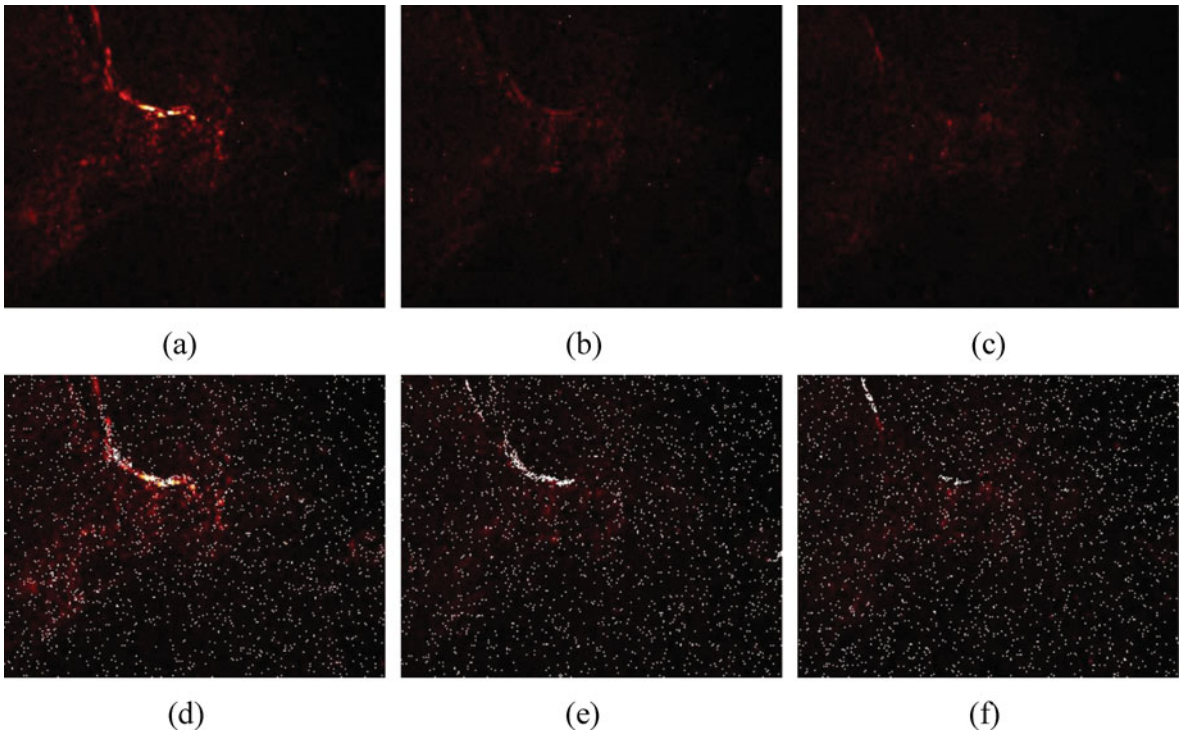

Figure 5. Detection of the second solar flare: snapshot of the video (a) at frame $t = 230$, when the second solar flare is brightest; (b) at frame $t = 268$, when the second solar flare is nearly over; and (c) at frame $t = 300$, the last frame of this video. The locations of 2000 sampled data streams are (d) at frame $t = 230$; (e) at frame $t = 268$; and (f) at frame $t = 300$.

Figures 4 and 5 show that our method cannot only detect the occurrences of solar flares, but also localize the flares. The locations of the sampled data streams clearly show the patterns of the solar flare in Figures 4(e) and 5(e). On the other hand, when there is no solar flare occurring, locations of sampled data streams are nearly random and do not show any obvious pattern as illustrated in Figure 4(f). In addition, comparing Figure 5(e) and (f), when the solar enters a normal state after the flare disappears, the adaptive sampling strategy does not stick to the locations of the previous solar flare and tends to be random again. Therefore, these two phenomena further validate property 1 of the TRAS algorithm. Due to this characteristic, the TRAS algorithm has also the potential to detect multiple shifts that occur at different times.

Although methodologies have already been investigated to estimate the threshold d, accurately determining its value given a prescribed in-control ARL is still a challenging problem, especially when the training sample is small. Chatterjee and Qiu (2009) proposed a bootstrap procedure that repeatedly draws observations with replacement from the in-control data to estimate the control limit. Qiu (2008) and Qiu and Li (2011) applied Monte Carlo simulation and used a bisection search algorithm to calculate the threshold value. In this study, we estimate the value of threshold d based on the data in the first 100 frames with bootstrap technique (the detailed steps can be found in Appendix D) to demonstrate the effectiveness of our algorithm. Phaladiganon et al. (2011, 2013) conducted a comprehensive study between the number of in-control data size and the control limit estimation when using bootstrap-based control charts. They showed that the variability of control limit estimation is greater when a small number of bootstrap samples are involved, but stabilizes as the number of bootstrap samples increases. In addition, they concluded that the determination of the appropriate number of bootstrap samples to use is not obvious. Zou and Tsung (2009) further showed that how large the number of in-control samples needed to accurately estimate the threshold d depends on the dimension m. To limit the scope of this study, we decide not to investigate the sensitivity of our algorithm in this article but leave it as a future work.

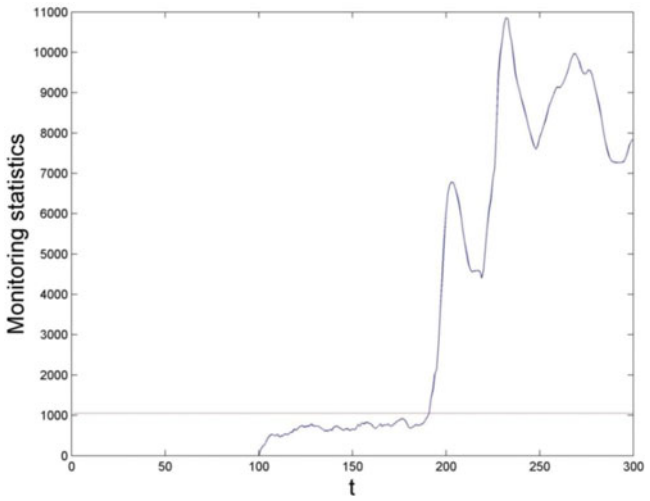

Figure 6. The monitoring statistics (i.e. the sum of top-r local statistics) over different acquisition time by implementing the TRAS algorithm.

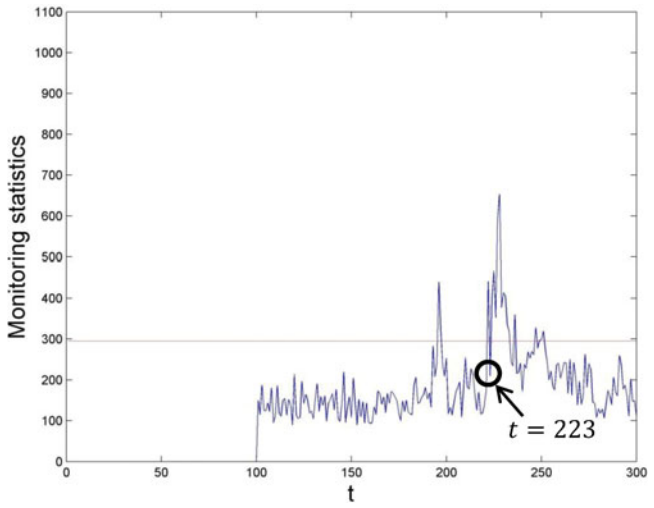

Figure 7. The monitoring statistics (i.e. the sum of top-r local statistics) over different acquisition time by implementing the random sampling algorithm.

Figure 6 plots the values of the monitoring statistics by implementing the TRAS algorithm at frames $t = 101 \sim 300$. The horizontal line represents $d = 1050$, which corresponds to the in-control ARL of 2500. Although 2000 pixels are available, which accounts for only 2.95 % of the total information, our method can still quickly detect the two solar flares at frames $t = 190$ and $t = 221$ as shown in Figure 6. This is comparable to the results in (Xie et al. 2013), where the two solar flares are detected at frames $t = 191$ and $t = 217$. However, the method in (Xie et al. 2013) is based on the generalized likelihood ratio procedure (Siegmund and Venkatraman 1995) and assumes that all 67,744 pixels in each data frame are available. In addition, there is no recursive formula in (Xie et al. 2013) and thus their method is not suitable for online monitoring applications.

To further demonstrate the effectiveness of the TRAS algorithm, we have also considered the random sampling approach with the same parameter settings and the detection result is shown in Figure 7. The random sampling approach detects the two solar flares at frames $t = 196$ and $t = 222$, which is slower than the TRAS algorithm. Moreover, it sends lots of misleading messages (e.g., the in-control sample at $t = 223$) which can create much trouble in practice as the users need to determine if the out-of-control sample at $t = 222$ is truly out-of-control or just a false-positive error. As a contrast, the difference between the out-of-control and the in-control samples by implementing the TRAS algorithm is much clearer in Figure 6, and thus the TRAS algorithm provides a better detection performance.

5. CONCLUSION

Quick process change detection is an important and challenging topic in many industrial and civilian applications. This article develops a novel adaptive sampling monitoring scheme by introducing a compensation coefficient to the local data streams for not taking observations and by using the sum of top-r local statistics for online process change detection. The use of this framework has several advantages over other approaches, which include significant decreases in computational cost (the complexity is only linear in the number of variables), and extensive savings for physical sensors, data acquisition, transmission, and processing time. The proposed adaptive sampling algorithm, which is named as TRAS, has two properties: (1) quickly detect a wide range of possible changes with no prior knowledge of the potential failure patterns; and (2) adaptively select/sample the data streams to be observed from the whole streaming data to maximize the sensitivity for process change detection when there are only q ($q \leq m$) out of m data streams observable at each acquisition time. Two properties of this algorithm are also given with proofs in this article. The methodology has been tested and validated on a hot-forming process and a real solar flare example. Both studies have demonstrated the capabilities of the TRAS algorithm to quickly detect and also localize the process changes.

This research establishes a new direction in online process monitoring by developing an adaptive sampling strategy to determine which data streams should be observed in the spatial domain when only a limited number of resources are available. Due to the simplicity and effectiveness of the TRAS algorithm, it can also be extended to other applications such as syndromic surveillance in epidemiology, network traffic control, intrusion detection, and surveillance video. There are several important topics for future research that are related to this work. First, the current compensation coefficient Δ is a constant number. Further studies can be done to investigate the adaptive value of Δ based on online measurements to further minimize the detection delay. Second, the theoretical basis of the threshold selection and the sensitivity analysis of the in-control data size on the threshold estimation are challenging but important topics that need further study. Moreover, it is an interesting question to derive a relationship function between the out-of-control ARL and the parameters of the TRAS algorithm, which not only depends the parameters q and $|\mathcal{M}_0|$, but also m, Δ and the specific failure pattern (magnitude and location of the shift variables). This can provide us a guideline that ensures the TRAS algorithm is able to detect the interested failure pattern within a user specified ARL requirement under certain conditions.

Moreover, it will be interesting to extend or modify the TRAS algorithm to non-Gaussian distributions. Here the local CUSUM control chart has to be replaced by another efficient local univariate control chart that is capable to monitor each data stream locally with nonnormal distributions (via either parametric or nonparametric approaches). The good news is that it is still applicable to apply our two main ideas of introducing the compensation coefficient Δ for not taking observation and using the sum of top-r local statistics (in a log-likelihood scale) to monitor the change in the system. As a concrete example, we may consider using a likelihood-ratio-based control chart to replace the CUSUM control chart when local data streams are Bernoulli or Poisson distributed. Then we can still introduce the compensation coefficient Δ to each unobserved control chart and use the sum of top-r local statistics (in a log-likelihood scale) to monitor the change in the system as we did before. We expect that the robustness of our proposed method on non-Gaussian data depends on the robustness of the univariate local control charts on non-Gaussian data. Similarly, it is also possible to revise our algorithm to detect shifts in both local means and local variances (see Healy 1987), though we feel that it will not be

effective to detect a shift on the correlations between different data streams, as the TRAS algorithm ignores the correlation structures by summing up the local statistics. We will investigate the aforementioned extensions of our method and also the performances in the future work.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the National Science Foundation under grant NSF CMMI-1362529, CMMI-1362876, and DMS-0954704. The authors would also like to thank the Associate Editor and three anonymous referees for their constructive comments and generous help.

APPENDIX A: PROOF OF PROPERTY 1

This appendix proves the first property of the TRAS algorithm. Denote $\mathcal{M}\backslash U = \{k \in \mathcal{M} | k \notin U\}$. First of all, let us consider the following lemma:

Lemma 1. $W_{k',t} \geq W_{k,t}$, for all $t > t_0$, all $k' \in \mathcal{M}\backslash U$, and all $k \in U$.

We can prove lemma 1 via contradiction. Assume there exists some $t > t_0$, $k' \in \mathcal{M}\backslash U$, and $k \in U$, such that $W_{k,t} > W_{k',t}$. Let S_t denote the sensor layout at time t. If $k' \in S_t$, then the sensor on variable k' will be redistributed to other variables in $\mathcal{M}\backslash U$ at time $t+1$. Since the incremental part of the local statistics for both variables k' and k is always Δ, a sensor cannot be redistributed back to variable k' without first deployed on variable k. In this way, $k' \in U$, which contradicts to the assumption. On the other hand, if $k' \notin S_t$, then it follows the same logic that a sensor cannot be redistributed to variable k' without first deployed on variable k, which also contradicts to the assumption. Therefore, we have proved lemma 1 and now we will use it to prove the first property of the TRAS algorithm.

Since sensor resources will never be redistributed to the variables in U after time t_0, all variables $k' \in \mathcal{M}\backslash U$ must have sensor deployed with infinite number of times. Now, let us first consider the case when $\rho_{k'} < 0$. Without loss of generality, we assume $E(X_{k',t}) = u_{k'} \geq 0$. The incremental part of the local statistics for any variable without sensor deployed is always Δ, while the incremental part for any variable k' with sensor deployed is either $u_{\min}X_{k',t} - \frac{u_{\min}^2}{2}$ or $-u_{\min}X_{k',t} - \frac{u_{\min}^2}{2}$ or 0. Since only one term (either $u_{\min}X_{k',t} - \frac{u_{\min}^2}{2}$ or $-u_{\min}X_{k',t} - \frac{u_{\min}^2}{2}$) can be positive at any given time and $X_{k',t}$ is an iid normal random variable, there must be a series of $X_{k',t}$ such that either $\sum_{t=1}^{\infty}(u_{\min}X_{k',t} - \frac{u_{\min}^2}{2}) \geq t$ or $\sum_{t=1}^{\infty}(-u_{\min}X_{k',t} - \frac{u_{\min}^2}{2}) \geq t$. Since $\rho_{k'} < 0$, $P(\sum_{t=1}^{\infty}(u_{\min}X_{k',t}; -\frac{u_{\min}^2}{2}) \geq t) + P(\sum_{t=1}^{\infty}(-u_{\min}X_{k',t'} - \frac{u_{\min}^2}{2}) \geq t) = \lim_{t\to\infty}(1 - \varphi(\frac{\sqrt{t}}{u_{\min}} + \frac{\sqrt{t}u_{\min}}{2} - u_{k'}\sqrt{t})) + \lim_{t\to\infty}(1 - \varphi(\frac{\sqrt{t}}{u_{\min}} + \frac{\sqrt{t}u_{\min}}{2} + u_{k'}\sqrt{t})) = 0$. Thus, there is no such series of $X_{k',t}$. Second, let us consider the case when $\rho_{k'} = 0$. Define $Y_{k',n} = u_{\min}X_{k',t_0+n} - \frac{u_{\min}^2}{2} - \Delta$ and $H_{k',n} = \sum_{i=1}^{n} Y_{k',i}$. Then, $Y_{k',n}$ is an iid normal random variable with mean 0 and variance $u_{\min}^2$. Consequently, $\{H_{k',n} : n \geq 1\}$ refers to the oscillating random walk process. Denote $D = \min\{H_{k',n} : n \geq 1\}$. Thus, $D \xrightarrow{\text{a.s.}} -\infty$ as $n \to \infty$ (Gut 1988). In other words, there must exist a time t, such that $W_{k,t} > W_{k',t}$. Therefore, it is impossible that sensor resources are always reallocated among the variables that only belong to $\mathcal{M}\backslash U$. In other words, U must be an empty set. In this way, we have finished the proof for property 1.

APPENDIX B: PROOF OF PROPERTY 2

This appendix proves the second property of the TRAS algorithm. Let S_t denote the sensor layout at time t. Considering the variable k, where $k \in S_t$ and $k \in B$, then $W_{k,t} \geq W_{k',t}$ for all $k' \notin S_t$. Without loss of generality, let $E(X_{k,t}) = u_k \geq 0$. Define $Y_{k,n} = X_{k,t+n} - \frac{u_{\min}}{2} - \frac{\Delta}{u_{\min}}$ and $H_{k,n} = \sum_{i=1}^{n} Y_{k,i}$. Then, $Y_{k,n}$ is an iid normal random variable with mean $u_k - \frac{u_{\min}}{2} - \frac{\Delta}{u_{\min}}$ and variance 1. Consequently, $\{H_{k,n} : n \geq 0\}$ refers to the Guassian random walk process, where $H_{k,0} = 0$. Denote $D = \min\{H_{k,n} : n \geq 0\}$ We are interested to show that $P(D = 0)$ is nonzero as the stopping criteria $d \to \infty$. Since $\mathrm{P(D=0) = P(H_{k,n} \geq 0, \text{for all} n \geq 0)}$, it is equivalent to show that there is a nonzero probability that once a sensor is deployed on the variable $k \in B$ at time t, it will never be reallocated to other variables (i.e. $W_{k,t+n} \geq W_{k',t+n}$ for all $k' \notin S_t$, and $n \geq 0$). According to (Chang and Peres 1997; Janssen and Van Leeuwaarden 2007), the probability that the minimum of the Gaussian random walk is zero satisfies:

$$P(D=0) = \sqrt{2}\delta_k \exp\left\{\frac{\delta_k}{\sqrt{2\pi}} \sum_{r=0}^{\infty} \frac{\zeta\left(\frac{1}{2} - r\right)}{r!\,(2r+1)} \left(-\frac{\delta_k^2}{2}\right)^r\right\} \quad \text{for} \quad 0 < \delta_k < 2\sqrt{\pi}, \tag{B.1}$$

where $\delta_k = u_k - \frac{u_{\min}}{2} - \frac{\Delta}{u_{\min}}$ and $\zeta()$ is the Riemann zeta function. It is worth mentioning that $P(D = 0)$ is an increasing function as δ_k gets larger (even when $\delta_k \geq 2\sqrt{\pi}$), which means the random walk is less likely to go back to 0 as δ_k increases. In this way, we have finished the proof for property 2.

APPENDIX C: DETAILS OF THE SIMULATION STEPS

This appendix elaborates the details of the simulation steps when comparing the performance between the TRAS algorithm and the CBA algorithm under different combinations of Δ ($\Delta = 0.01, 0.1, 0.5$), τ ($\tau = 1.5, 2, 2.5, 3$), and q ($q = 2, 3, 4$) values for single mean shift case.

1. Given each value of the number of available sensors q, the mean shift magnitude τ, and the incremental parameter Δ, the following substeps are performed for each single mean shift fault scenario i ($i = 1, \ldots 5$) (i.e. mean shift occurs at variable i):
   i. A dataset of $\mathbf{X}_t = \{X_{1,t}, X_{2,t}, X_{3,t}, X_{4,t}, X_{5,t}\}$ with M(= 5000) samples is generated, in which a mean shift τ is introduced at $X_{i,1}$, the first sample of X_i.
   ii. Implement the WISD. For each incoming sample, calculate the local statistics and update sensor layout based on the TRAS algorithm in Figure 1.
   iii. Index of the first out-of-control sample is recorded as RL_i. In addition, set UF_i equal to one if the variable associated with the largest local statistic is the root cause variable, i.
   iv. Repeat Steps 1–3 for N(= 10,000) times. The average of RL_i, $\overline{\mathrm{RL}_i}$ and the average of UF_i, $\overline{\mathrm{UF}_i}$ are computed and recorded.
2. Calculate the maximum out-of-control ARL, $\mathrm{RL}_{\max} = \max_i(\overline{\mathrm{RL}_i})$ and the average out-of-control ARL, $\overline{\overline{\mathrm{RL}}} = \sum_i \overline{\mathrm{RL}_i}/5$. Similarly, calculate the minimum uniquely correct diagnosis rate, $\mathrm{UF}_{\min} = \min_i(\overline{\mathrm{UF}_i})$ and the average of uniquely correct diagnosis rate, $\overline{\overline{\mathrm{UF}}} = \sum_i \overline{\mathrm{UF}_i}/5$.
3. Repeat Steps 1 and 2 for different combinations of Δ ($\Delta = 0.01, 0.1, 0.5$), τ ($\tau = 1.5, 2, 2.5, 3$), and q ($q = 2, 3, 4$) values, and present the results in Table 1.

APPENDIX D: DETAILS OF FINDING THE THRESHOLD VALUE d

This appendix describes the detailed steps to estimate the threshold value d given any prescribed in-control ARL, ARL_U. Specially, we conduct the following evaluation processes:

1. Set $d_{\min}$ and $d_{\max}$ a small and a large values as the initial lower and upper bounds of d, respectively. Let $d = \frac{d_{\min}+d_{\max}}{2}$.
2. Generate a bootstrap dataset with $M(= 5000)$ samples by randomly drawing the data with replacement from the first 100 frames.
3. Implement the TRAS algorithm and record the index of the first out-of-control sample, RL.
4. Repeat Steps 2–3 for $N(= 10{,}000)$ times, which produce N bootstrap datasets. Calculate the average of RL, $\overline{\text{RL}}$.
5. If the $\overline{\text{RL}}$ is larger than ARL_U, let $d_{\max} = d$; Otherwise, let $d_{\min} = d$. Then update $d = \frac{d_{min}+d_{max}}{2}$.
6. Repeat Steps 2–5 until there is only small difference between $\overline{\text{RL}}$ and ARL_U.

[*Received August 2013. Revised July 2014.*]

REFERENCES

Augusto, C. R. A., Fauth, A. C., Navia, C. E., Shigeouka, H., and Tsui, K. H. (2011), "Connection Among Spacecrafts and Ground Level Observations of Small Solar Transient Events," *Experimental Astronomy*, 34, 177–197. [312]

Chang, J. T., and Peres, Y. (1997), "Ladder Heights, Gaussian Random Walks, and the Riemann Zeta Function," *The Annals of Probability*, 25, 787–802. [317]

Chan, L. K., and Zhang, J. (2001), "Cumulative Sum Control Charts for the Covariance Matrix," *Statistica Sinica*, 11, 767–790. [307]

Chatterjee, S., and Qiu, P. (2009), "Distribution-Free Cumulative Sum Control Charts Using Bootstrap Based Control Limits," *Annals of Applied Statistics*, 3, 349–369. [310,315]

Cheng, S. W., and Thaga, K. (2005a), "Max-CUSUM Chart for Autocorrelated Processes," *Statistica Sinica*, 15, 527–546. [307,308]

——— (2005b), "Multivariate max-CUSUM Chart," *Quality Technology and Quantitative Management*, 2, 221–235. [307,308]

Crosier, R. B. (1986), "A New Two-Sided Cumulative Sum Quality Control Scheme," *Technometrics*, 28, 187–194. [307]

——— (1988), "Multivariate Generalizations of Cumulative Sum Quality-Control Schemes," *Technometrics*, 30, 291–303. [307]

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Boca Raton, FL: Chapman and Hall. [310]

Fiorelli, E., Leonard, N. E., Bhatta, P., Paley, D., Bachmayer, R., and Fratantoni, D. M. (2006), "Multi-AUV Control and Adaptive Sampling in Monterey Bay," *IEEE Journal of Oceanic Engineering*, 31, 935–948. [306]

Frost, J. R., and Stone, L. D. (2001), *Review of Search Theory: Advances and Applications to Search and Rescue Decision Support*, U.S. Coast Guard Research and Development Center, Groton. [306]

Gut, A. (1988), *Stopped Random Walks, Limit Theory and Applications*, Berlin: Springer. [317]

Hawkins, D. M. (1993), "Multivariate Quality Control Based on Regression Adjusted Variables," *Technometrics*, 33, 61–75. [305,307]

Healy, J. D. (1987), "A Note on Multivariate CUSUM Procedures," *Technometrics*, 29, 409–412. [307,316]

Janssen, A. J. E. M., and Van Leeuwaarden, J. S. H. (2007), "On Lerch's Transcendent and the Gaussian Random Walk," *The Annals of Applied Probability*, 17, 421–439. [317]

Kagan, E., and Ben-Gal, I. (2013), *Probabilistic Search for Tracking Targets: Theory and Modern Applications*, London: Wiley. [306]

Katenka, N., Levina, E., and Michailidis, G. (2008), "Robust Target Localization From Binary Decisions in Wireless Sensor Networks," *Technometrics*, 50, 448–461. [306]

Lorden, G. (1971), "Procedures for Reacting to a Change in Distribution," *Annals of Mathematical Statistics*, 42, 1897–1908. [307]

Li, J., and Jin, J. (2010), "Optimal Sensor Allocation by Integrating Causal Models and Set-Covering Algorithms," *IIE Transaction*, 42, 564–576. [306,311]

Li, Z., and Qiu, P. (2013), "Statistical Process Control Using Dynamic Sampling," *Technometrics*, in press, available at *http://users.phhp.ufl.edu/pqiu/research/DS-SPC.pdf*. [306]

Lim, H. B., Lam, V. T., Foo, M. C., and Zeng, Y. (2006), "Adaptive Distributed Resource Allocation in Wireless Sensor Networks," in *Proceedings of the 2nd International Conference on Mobile Ad-hoc and Sensor Networks*, Hong Kong, China. [306]

Limongelli, M. P. (2003), "Optimal Location of Sensors for Reconstruction of Seismic Response Through Spline Function Interpolation," *Earthquake Engineering and Structural Dynamics*, 32, 1055–1074. [306]

Liu, K., and Shi, J. (2013), "Objectives-Oriented Optimal Sensor Allocation Strategy for Process Monitoring and Diagnosis by Multivariate Analysis in a Bayesian Network," *IIE Transactions*, 45, 630–643. [306,310]

Liu, K., Zhang, X., and Shi, J. (2013), "Adaptive Sensor Allocation Strategy for Process Monitoring and Diagnosis in a Bayesian Network," *IEEE Transactions on Automation Science and Engineering*, in press available at *http://www2.isye.gatech.edu/~jshi33/Publications/JournalPapers/p82.pdf*. [306,311,312]

Mandroli, S. S., Shrivastava, A., and Ding, Y. (2006), "A Survey of Inspection Strategy and Sensor Distribution Studies in Discrete-Part Manufacturing Processes," *IIE Transactions*, 38, 309–328. [306]

Montgomery, D. C. (2009), *Introduction to Statistical Quality Control*, New York: Wiley. [306,312]

Mei, Y. J. (2010), "Efficient Scalable Schemes for Monitoring a Large Number of Data Streams," *Biometrika*, 97, 419–433. [307]

——— (2011), "Quickest Detection in Censoring Sensor Networks," in *Proceedings of IEEE International Symposium on Information Theory*, 2148–2152. [306,307,308,309,310]

Ngai, H. M., and Zhang, J. (2001), "Multivariate Cumulative Sum Control Charts Based on Projection Pursuit," *Statistica Sinica*, 11, 747–766. [307]

Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika*, 41, 100–114. [307]

Phaladiganon, P., Kim, S. B., Chen, V. C. P., Baek, J. G., and Park, S. K. (2011), "Bootstrap-Based Multivariate Control Charts," *Communications in Statistics: Simulation and Computation*, 40, 645–662. [315]

Phaladiganon, P., Kim, S. B., Chen, V. C., and Jiang, W. (2013), "Principal Component Analysis-Based Control Charts for Multivariate Nonnormal Distributions," *Expert Systems with Applications*, 40, 3044–3054. [315]

Pignatiello, J. J., and Runger, G. C. (1990), "Comparisons of Multivariate CUSUM Charts," *Journal of Quality Technology*, 22, 173–186. [307]

Qiu, P. (2005), *Image Processing and Jump Regression Analysis*, New York: Wiley. [313]

——— (2008), "Distribution-Free Multivariate Process Control Based on Log-Linear Modeling," *IIE Transactions*, 40, 664–677. [307,315]

——— (2013), *Introduction to Statistical Process Control*, Boca Raton, FL: Chapman & Hall/CRC. [307]

Qiu, P., and Hawkins, D. M. (2001), "A Rank-Based Multivariate CUSUM Procedure," *Technometrics*, 43, 120–132. [307]

——— (2003), "A Nonparametric Multivariate Cumulative Sum Procedure for Detecting Shifts in all Directions," *Journal of the Royal Statistical Society,* Series D, 52, 151–164. [307]

Qiu, P., and Li, Z. (2011), "On Nonparametric Statistical Process Control of Univariate Processes," *Technometrics*, 53, 390–405. [315]

Radke, R. J., Andra, S., Al-Kofahi, O., and Roysam, B. (2005), "Image Change Detection Algorithms: A Systematic Survey," *IEEE Transactions on Image Processing*, 14, 294–307. [313]

Rahimi, M. H., Kaiser, W., Sukhatme, G. S., and Estrin, D. (2005), "Adaptive Sampling for Environmental Field Estimation Using Robotic Sensors," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3692–3698. [306]

Siegmund, D., and Venkatraman, E. S. (1995), "Using the Generalized Likelihood Ratio Statistic for Sequential Detection of a Change-Point," *The Annals of Statistics*, 23, 255–271. [316]

Tan, R., Xing G., Chen, J., Song W., and Huang, R., (2012), "Fusion-Based Volcanic Earthquake Detection and Timing in Wireless Sensor Networks," *ACM Transaction on Sensor Networks (ACM TOSN)*, 9, 2, 17:1–25. [306]

Tartakovsky, A. G., Rozovskii, B. L., Blazek, R. B., and Kim, H. (2006), "Detection of Intrusions in Information Systems by Sequential Change-Point Methods" (with discussion), *Statistical Methodology*, 3, 252–340. [307]

Tsui, K-L, Han, S. W., Jiang, W., and Woodall, W. H. (2012), "A Review and Comparison of Likelihood Based Charting Methods," *IIE Transactions*, 44, 724–743. [307]

Woodall, W. H., and Ncube, M. M. (1985), "Multivariate CUSUM Quality Control Procedure," *Technometrics*, 27, 285–292. [307]

Xie, Y., Huang, J., and Willett, R. (2013), "Changepoint Detection for High-Dimensional Time Series with Missing Data," *IEEE Journal of Selected Topics in Signal Processing*, 7, 12–27. [314,316]

Xie, Y., and Siegmund, D. (2013), "Sequential Multi-Sensor Change-Point Detection," *The Annals of Statistics*, 41, 670–692. [306]

Zoghi, M., and Kahaei, M. H. (2010), "Adaptive Sensor Selection in Wireless Sensor Networks for Target Tracking," *IET Signal Processing*, 4, 530–536. [306]

Zou, C., and Qiu, P. (2009), "Multivariate Statistical Process Control Using LASSO," *Journal of the American Statistical Association*, 104, 1586–1596. [307,315]