Adaptive Partially-Observed Sequential Change Detection and Isolation

Xinyu Zhao[1], Jiuyun Hu[1], Yajun Mei[2], Hao Yan[1] *
[1]School of Computing and Augmented Intelligence
Arizona State University,
[2]School of Industrial and Systems Engineering,
Georgia Institute of Technology

Abstract

High-dimensional data has become popular due to the easy accessibility of sensors in modern industrial applications. However, one specific challenge is that it is often not easy to obtain complete measurements due to limited sensing powers and resource constraints. Furthermore, distinct failure patterns may exist in the systems, and it is necessary to identify the true failure pattern. This work focuses on the online adaptive monitoring of high-dimensional data in resource-constrained environments with multiple potential failure modes. To achieve this, we propose to apply the Shiryaev–Roberts procedure on the failure mode level and utilize the multi-arm bandit to balance the exploration and exploitation. We further discuss the theoretical property of the proposed algorithm to show that the proposed method can correctly isolate the failure mode. Finally, extensive simulations and two case studies demonstrate that the change point detection performance and the failure mode isolation accuracy can be greatly improved.

Keywords: Shiryaev–Roberts procedure, multi-arm bandit, sequential change-point detection, adaptive sampling, multiple failure modes

1 Introduction

Nowadays, most industrial applications are instrumented with hundreds or thousands of

sensors due to the advancement in sensing technology. Real-time process monitoring and

fault diagnosis are among the benefits that can be gained from effective modeling and analysis of the produced high-dimensional streaming data. Classical researches for process monitoring of high-dimensional streaming data focus on a fully observable process, which means at each sampling time point, all the variables can be observed for analysis (Yan et al., 2018). However, it is often infeasible to acquire measurements of all these sensing variables in real time due to limited sensing resources, sensing capacity, sensor battery, or other constraints such as system transmission bandwidth, memory, storage space, and processing speed in modern industrial applications (Liu et al., 2015). Furthermore, under change detection and isolation setting, we assume that the engineered systems that are being studied have several distinct failure modes and patterns but do not know which failure mode may occur beforehand. Overall, this paper focuses on change point detection under resource-constrained environments with multiple potential failure modes.

The first motivating example is in the hot forming process (Li and Jin, 2010) as shown in Fig. 1(a). There are five sensing variables in the system: the final dimension of workpiece $\mathbf{X}_1$, the tension in workpiece $\mathbf{X}_2$, material flow stress $\mathbf{X}_3$, temperature $\mathbf{X}_4$, and blank holding force $\mathbf{X}_5$. These five variables can be represented as a Bayesian network, as shown in Fig. 1(a). For example, if we know that the change of $\mathbf{X}_4$ and $\mathbf{X}_5$ are the two major failure sources in the system. If $\mathbf{X}_4$ changes, $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)$ will also change. Furthermore, If $\mathbf{X}_5$ changes, only $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_5)$ will change. Therefore, different failure modes may affect a different subset of sensors differently.

Another example comes from in-situ hot-spots detection in the laser powder bed fusion (LPBF) process in the metal additive manufacturing process. A thermal camera is often

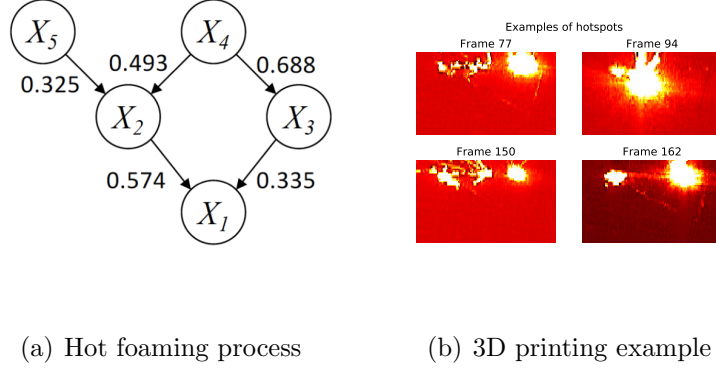(a) Hot foaming process　　　　　(b) 3D printing example

Figure 1: Examples of Complex data in Various Industrial Applications **Left** figure shows an example of a hot foaming process. **Right** figure shows an example of monitoring the thermal images in additive manufacturing.

used to monitor the stability of the process while the product is being produced on a layer-by-layer basis. Here, detecting the hot-spots early is crucial for further product quality control. Fig. 1(b) show an example of such hot-spots from the thermal camera. Given that the anomaly or hot-spots can only occur on the edge/corner of the scanning path, multiple failure modes can be defined.

There are a few challenges of sequential change-point detection under the sampling constraint: 1) From the previous examples, the failure mode distribution can be quite complicated. For example, in the hot foaming process, as shown in Fig. 1(a), we aim to detect the failure mode with the weakly conditional dependency on the graph; In the laser powder bed fusion process, as shown in Fig. 1(b), we aim to detect the spatially clustered hot-spots. 2) Even though we assume that we have prior knowledge of different potential failure modes, we do not know which failure mode may occur in the system. The main challenge is to balance the exploration of all potential failure modes and the
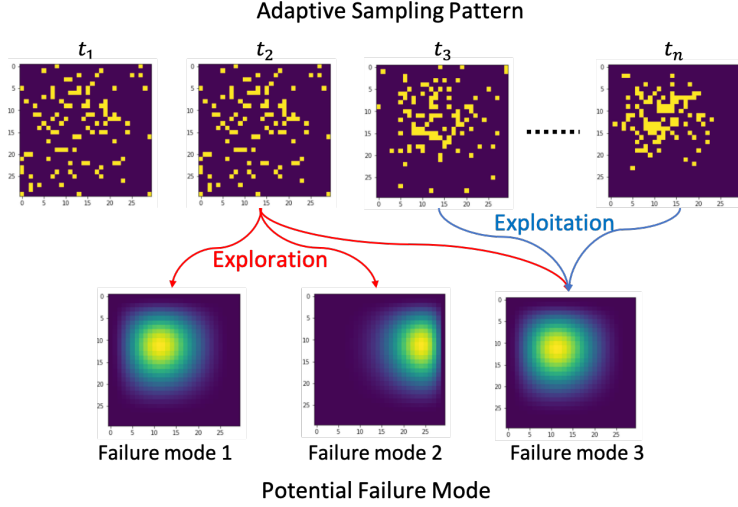
Figure 2: Conceptual Illustration of the Balance of Exploration and Exploitation; The sampling patterns at t_1, t_2 focus on the exploration of all failure modes. The sampling pattern at $t_3, \cdots, t_n$ focuses on the exploitation of the failure mode 3.

exploitation to focus on the most probable failure mode. A conceptual illustration of the proposed algorithm is provided in Fig. 2. The illustration example has shown an example that the sampled points are performed on the 2D spatial domain. The sampling patterns at time t_1 and t_2 focus on exploration for all failure modes and the sampling patterns at $t_3, \cdots, t_n$ focus on exploitation for failure mode 3. In general, it is hard to decide when the algorithms should change to exploitation or which failure mode they should focus on. Finally, given that the multiple failure modes have quite complex shapes and distributions, the exploration, and exploitation among these modes are often quite challenging.

There are also many works focusing on change-point detection under resources constraint. Most of the existing works are proposed based on the "local monitoring and global decision" framework, which focuses on monitoring each data stream independently using

local monitoring statistics and then fusing these local monitoring statistics together via a global decision framework. For example, Liu et al. (2015) proposed scalable and efficient algorithms for adaptive sampling for online monitoring. The method introduced a compensation parameter for the unobserved variables to increase the chance of exploring them. Recently, Zhang and Mei (2020) proposed to combine the powerful tools of the multi-arm bandit problem for efficient real-time monitoring of HD streaming data. However, these works either assume the data stream is independent or cannot take advantage of the failure mode information in some systems, which fails to monitor and identify the correct failure pattern. For a complete literature review of monitoring of high-dimensional streaming data, please see Section 2.

To generalize the sequential change-point detection framework to both detect and identify the correct failure modes, *change detection and isolation literature* has been proposed in the literature, which also inspires this research. The change-point detection and isolation often assume that there are a set of pre-defined post-change distributions. The goal is not only to detect the change with the shortest detection delay but also to identify which change mode occurs in the system. For example, Chen et al. (2020) proposed a Bayesian method to decide on a procedure to identify both the change point as well as the correct change mode. For a complete literature review of change detection and isolation, please see Section 3.3. However, these works typically assume that the data is fully observed, which cannot be applied to partially observed data.

To address the challenge of multiple failure modes and partially observed data, we propose a novel Multiple Thompson Sampling Shiryaev-Roberts-Pollak (MTSSRP) Method

by a modified "local monitoring and global decision framework". As far as the authors know, this is the first work that discusses the adaptive sampling framework for failure mode detection and isolation. Unlike the literature on the monitoring of HD streaming data, where the local monitoring statistics are defined at each individual sensor, we propose to define the local statistics for each individual failure mode. This enables the proposed MTSSRP to take advantage of the failure mode information, which is very important in the high-dimensional space, given that failure mode information can significantly reduce the search space since there are unlimited ways that change may occur in the high-dimensional space. To quantify the uncertainty of unobserved sensing variables for different failure modes, we propose to apply the Shiryaev-Robert (SR) procedure for sequential change point detection on the failure mode level.

Furthermore, to balance the exploration and exploitation, we will borrow the idea from Multi-arm Bandit (MAB). MAB aims to sequentially allocate a limited set of resources between competing "arms" to maximize their expected gain, where the reward function for each arm is not known. MAB provides a principled way to balance exploration and exploitation. To apply MAB for change point detection under the sampling constraint, we propose to use the SR statistics of the selected failure modes as the reward function in the Multi-arm Bandit (MAB) problem (Zhang and Mei, 2020). However, different from (Zhang and Mei, 2020), the selection of arm is on the sensor level, where the SR statistics is defined on the failure-mode level. For high-dimensional data, specifying the joint distribution of high-dimensional data can be very challenging. Therefore, this paper will explore spatial structures for defining the failure mode distributions as shown in 3.5. This

6

paper also discussed that with the independence assumption of the distribution variable, the computational efficiency could be greatly improved.

The paper is organized as follows. In Section 2, we will review the existing literature on change detection and isolation. We will also discuss works on process monitoring with resource constraints. We further introduce our proposed method and then discuss its property in Section 3. All the proofs are available in the outline supplementary materials. Then, we apply the proposed approach to both the simulated data and evaluate its performance and compare the existing methods in Section 4. Furthermore, we apply the proposed method to two real cases in Section 5, respectively. Concluding remarks are given in Section 6.

2 Literature Review

In this section, we will provide a more detailed review of statistical process control or sequential change point detection methods. We will briefly classify the methods for the following four categories: monitoring of independent HD streaming data, monitoring of functional data or profile monitoring, process monitoring with the resource constraint, change-point detection, and isolation.

In the first category, monitoring the HD streaming data has often been treated as monitoring the multiple independent univariate data streams. There are two distinct frameworks for monitoring independent data streams in recent years. First, the "global monitoring" framework focused on directly designing the global monitoring statistics for the process monitoring or change point detection for high-dimensional data (Xie and Siegmund, 2013; Wang and Mei, 2015; Cho and Fryzlewicz, 2015; Chan et al., 2017). However, the global

monitoring framework is typically computationally inefficient for high-dimensional data. Second, the "local monitoring and global decision" framework focus on monitoring each data stream independently using local monitoring statistics and then fusing these monitoring statistics together via the global statistics (Mei, 2010, 2011). The benefit is that these methods are typically computationally efficient and can be scalable to high-dimensional data. However, these methods are often limited to the independent data stream. Finally, this framework is targeted for the case of fully observed data, which may not be applicable under the resource constraint.

In the second category, profile monitoring techniques have been proposed to tackle the complex spatial correlation structures. Dimensionality reduction techniques, such as principal component analysis (PCA), are widely used. Various types of alternatives such as multivariate functional PCA (Paynabar et al., 2016), multi-linear PCA (Grasso et al., 2014), and tensor-based PCA (Yan et al., 2015) are proposed. On the other hand, non-parametric methods based on local kernel regression (Qiu et al., 2010) and splines (Chang and Yadama, 2010) are developed. To monitor the non-smooth waveform signals, a wavelet-based mixed effect model is proposed in (Paynabar and Jin, 2011). However, for both PCA-based methods and non-parametric methods, they typically assume that the change alternative is not known. To utilize the anomaly structures, smooth sparse decomposition methods have been proposed and utilize two sets of basis functions, the background basis and anomaly basis, to represent the spatial structures of the background and anomaly, which have been applied to smooth profiles (Yan et al., 2017, 2018) and waveform profiles (Yue et al., 2017). However, all the profile monitoring techniques assume that the complete measurements are

8

given and cannot be applied for HD data with partial observations.

In the third category, many existing works focus on the change point detection with the sampling constraint. Here, we will briefly classify the existing monitoring methods with the sampling constraint into two categories, monitoring the i.i.d data stream and monitoring the correlated data stream. For monitoring the i.i.d data stream with the sampling constraint, Liu et al. (2015) proposed a top-R-based adaptive sampling strategy as a combination of random sampling in the in-control state and fixed sampling in the out-of-control state. Another work by (Zhang and Mei, 2020) converts the problem into a MAB framework and adaptively selects the sensors with Thompson Sampling. Recent work by (Gopalan et al., 2021) provides an information-theoretic lower bound for the detection delay but focuses on monitoring changes along one dimension for a single failure mode.

However, due to the i.i.d assumption, these methods might not be suitable for data with complex distributions in reality. To deal with this problem, Xian et al. (2018) proposed an adaptive sampling strategy that can handle the correlated data generated from a multinomial distribution. For monitoring correlated data streams with the sampling constraint, these methods can be classified into monitoring data generated from the Bayesian Network and spatial profile. For example, Liu et al. (2013) and Liu and Shi (2013) proposed a sensor allocation strategy according to a Bayesian Network to detect changes with multivariate T^2 control chart. Another work discussed the problem when there is a spatial correlation among sensors and proposed a spatial-adaptive sampling strategy to focus on suspicious spatial clusters(Wang et al., 2018). However, these methods either consider the data stream with spatial correlation (Wang et al., 2018; Ren et al., 2020; Gómez et al., 2022) or modeled

by the Bayesian Network structures (Liu et al., 2013; Liu and Shi, 2013), which fails to apply to the problem with general failure mode distributions as discussed in this paper.

Finally, there is a large amount of work focused on the case when there are multiple failure modes, and it is necessary to identify the true failure while detecting the changes. The problem of sequential change detection with multiple failure modes is usually called change detection and isolation. The goal is to find the best decision procedure that can control the false alarm rate as well as the false isolation probability. The problem is of importance since it is common in different applications like fault diagnosis, process monitoring, and object identification (Nikiforov et al., 1993; Willsky, 1976; Malladi and Speyer, 1999). The major works in change detection and isolation can be categorized into Bayesian and non-Bayesian directions. Nikiforov (1995) proposed a change detection/isolation framework as an extension of Lorden's results (Lorden, 1971) which follows non-Bayesian schema. Another two works formulate the problem into a Bayesian version, which considers the change point as a random variable (Chen et al., 2020; Malladi and Speyer, 1999).

3   Proposed Methodology

In this section, we will first describe the problem formulation of partially observed multi-mode change detection based on high-dimensional (HD) streaming data with sampling control in Section 3.1. We will describe the proposed MTSSRP methodology in Section 3.2. We will prove important properties of the proposed algorithms about the average run length and failure mode isolation guarantee in Section 3.3. We will give the guidelines to select the tuning parameters of the proposed MTSSRP method in Section 3.4. Finally, we

will give a discussion and several guidelines on selecting the failure mode distributions in Section 3.5.

3.1 Problem Formulation and Background

Suppose we are monitoring data stream $X_{j,t}$ for $j = 1, \cdots, p$ and $t = 1, 2, \cdots, T$, where p is the number of dimensionality in the system and T is the monitored time length. We assume that the data streams follow joint distribution f_0 before change as $\mathbf{X}_t \sim f_0$ for $t < \nu$. At some unknown change time $\nu \in \{1, 2, \cdots, \}$, an undesirable event occurs and causes an abrupt change of the data stream into one or few failure modes. For example, the after-change distributions $f_k \in \mathcal{F}$ can be anyone from a family of distributions $\mathcal{F} = \{f_1, \cdots, f_K\}$. In another word, after the change, $\mathbf{X}_t \sim f_k, k = 1, \cdots, K$ for $t > \nu$. In other words, we do not assume that we know which failure mode occurs in the system. Following the change detection and isolation framework, we do assume that a single failure mode f_k may occur after the change. We will discuss the case with multiple failure mode. Here, $f_0, f_1, \cdots f_K$ are the joint distribution for all sensing variables. The sensing variables can also also be correlated. Finally, in practice, we can set the magnitude of the joint distribution as the interested magnitude of the change to be detected.

Furthermore, we assume that given the resource constraint, it is not possible to observe all the data streams. For the partially observed data with sampling control, the set of the observed data is denoted by $\mathbf{y}_t = \{X_{j,t}, j \in C_t\}$. Here, C_t is the set of observed sensor indices at time t, which can be selected online. In other words, we can define $a_{j,t}$ as the binary variable denoting whether the variable j is observed at time t, $C_t = \{j : a_{j,t} = 1\}$.

11

Finally, the sampling constraint is represented by $\sum_{j=1}^{p} a_{j,t} = q$ at each time $t = 1, 2 \cdots$, which means at each time t, only q sensing variables can be observed from all p variables.

The objective of this paper is to design an efficient adaptive sampling algorithm and the change point detection algorithm to automatically distribute sensing resources according to the knowledge of the system failure modes such that the change can be detected quickly as soon as it occurs, and the corresponding failure mode can be identified accurately while maintaining the false alarm constraint.

3.2 Proposed Algorithm

In Section 3.2, we will introduce the proposed methodology with the following major steps, monitoring statistics update, change point detection decision, failure mode isolation, and planning for adaptive sampling. The overall framework is shown in Fig. 3, and the detailed steps are as follows:

1. **Monitoring statistics update**: We first construct the monitoring statistics of partially observed HD streaming data for each failure mode based on the SR procedure. The detailed step is discussed in Section 3.2.1.

2. **Change point detection decision**: According to the updated monitoring statistics for each failure mode, a top-R statistic is used to conduct the global decision. We will then raise a global alarm if the process has gone out of control and decide which failure mode has occurred. The detailed step is discussed in Section 3.2.2.

3. **Planning for adaptive sampling**: If the change is not detected, we will update the
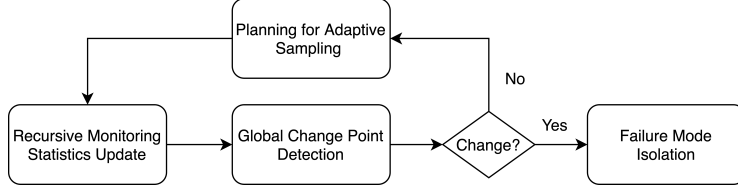
Figure 3: Procedure of the proposed method

sampling layout dynamically according to the historical observations. To achieve this, we propose to borrow the Thompson sampling idea to decide the next sampling layout, where the data is randomly sampled from the identified failure mode distribution. The detailed step is discussed in Section 3.2.3. Furthermore, the optimization algorithm to solve this planning and optimal sampling decision is discussed in Section 3.2.4. The selected sampling patterns will be used to update the monitoring statistics recursively.

4. **Failure mode isolation**: Finally, if the change is detected, we will isolate and identify the true failure mode in the system.

3.2.1 Recursive Monitoring Statistics Update

In this subsection, we will discuss the proposed method of constructing the Shiyaev-Roberts (SR) statistics for each failure mode $k \in \{1, \cdots, K\}$ with missing observations. Here, we denote the local SR statistics at time t as $R_{k,t}$. Here, $\mathbf{y}_t$ is the set of observed data streams. We will follow the same rule of updating the local statistics $r_{k,t}$.

$$R_{k,t} = (R_{k,t-1} + 1) \frac{\tilde{f}_{C_t,k}(\mathbf{y}_t)}{\tilde{f}_{C_t,0}(\mathbf{y}_t)}.$$

13

Here, $\tilde{f}_{C_t,k}$ is the joint distribution of the observed data $\mathbf{y}_t$ at set C_t. For computational efficiency and stability, it is recommended to use the $r_{k,t} = \log R_{k,t}$, updated as

$$r_{k,t} = \log(\exp(r_{k,t-1}) + 1) + \log \frac{\tilde{f}_{C_t,k}(\mathbf{y}_t)}{\tilde{f}_{C_t,0}(\mathbf{y}_t)}.$$

We set $R_{k,0} = 0$ initially and update the statistics accordingly.

3.2.2 Detection Decision and Failure Mode Isolation

We will combine the local statistics for each failure mode to construct a global stopping time. Here, if we know that the system only contains one failure mode, we propose to use the largest of the monitoring statistics $r_{k,t}$ to trigger the alarm. For example, if $r_{k,t}$ is larger than a control limit A, to raise the alarm.

$$T = \inf\{t \geq 1 : \max_k r_{k,t} \geq A\} \tag{1}$$

However, if we know that there are multiple failure modes in the system, the summation of the top $r_{k,t}$ statistics can be used to trigger the alarm.

$$T = \inf\{t \geq 1 : \sum_{k=1}^{K_s} r_{(k),t} \geq A\} \tag{2}$$

Finally, to isolate the most probable failure mode when the change is detected, we propose to use the monitoring statistics with the largest index $\hat{k}$, computed as

$$\hat{k} = \arg\max_k r_{k,T}. \tag{3}$$

3.2.3 Planning for Adaptive Sampling

In this section, we present an efficient method to plan and select the best sampling pattern to observe at the next time point. Suppose that we have observed $\mathbf{y}_1, \cdots, \mathbf{y}_{t-1}$ now and

14

the goal is to determine $\{a_{j,t}\}$ at next time t, which is a binary variable denoting whether variable j is observed or not at time t. Inspired by the MAB, we propose to maximize the reward function, defined by the monitoring statistics of the top few selected failure modes. More specifically, we propose to use the summation of the SR statistics of the top-K_s failure modes as the reward function, where K_s is a pre-defined parameter to balance the exploration and exploitation. In other words, we can define the reward function as $S_t = \sum_{k=1}^{K_s} r_{(k),t}$, where $r_{(k),t}$ is the rank of the statistics such as $r_{(1),t} \geq \cdots \geq r_{(K_s),t}$. One specific challenge is that to compute the reward function S_t for planning, we need to compute $r_{(k),t}$, which requires $\mathbf{x}_t$ to be fully observed. However, given that we are still at time $t-1$ yet and data x_t has not been observed yet, it is impossible to compute and optimize S_t for the planning problem.

To solve this planning problem, we propose to optimize a sampled version of the monitoring statistics S_t, defined as $\tilde{S}_t = \sum_{k=1}^{K_s} a_{k,t} r_{(k),t}$. Borrowing from the Thompson sampling algorithm, we would like to use the sampled version of $\mathbf{x}_t$, denoted as $\tilde{\mathbf{x}}_t^k$ as

$$\tilde{r}_{k,t} = \log(\exp{(r_{k,t-1})} + 1) + \log \frac{f_k(\tilde{\mathbf{x}}_t^k)}{f_0(\tilde{\mathbf{x}}_t^k)}, \tag{4}$$

where $\tilde{\mathbf{x}}_t^k \sim f_k$ is sampled from the kth failure mode. Finally, one can optimize $a_{j,t}$ by maximizing the sampled version of S_t, denoted as $\tilde{S}_t$ as

$$\max_{a_{j,t}} \tilde{S}_t \quad \text{subject to} \quad \sum_j a_{j,t} = q, a_{j,t} = \{0,1\} \tag{5}$$

Finally, we will discuss how to solve the optimization (5) in Section 3.2.4.

3.2.4 Optimization for Planning

In general, optimizing (5) is often challenging. If the problem dimension p and the number of selected sensing variables q are small, we can enumerate all the $\binom{p}{q}$ possible combinations of the sampling layouts. However, given the time complexity is $O\left(\binom{p}{q}\right)$, enumeration of all possible combinations is not feasible for large p, q.

Here, we will first present the closed-form solution for a special case of the proposed algorithm, where the joint failure mode distribution $f_k(\mathbf{x}_t) = \prod_j f_{j,k}(x_{t,j})$ can be approximated by independent but not necessarily identical distributions in each dimension j. Notice that if the possible failure modes in each dimension j is finite, the total number of possible failure modes for all dimensions is finite. Here, we would like to derive the analytical solution to optimize (5) under this setting in Proposition 1.

Proposition 1. *If $f_k(\mathbf{x}_t) = \prod_j f_{j,k}(x_{t,j})$ for $k = 0, \cdots, K$, the set C_t in (5) can be solved by selecting the indices of the largest q of $s_{t,j}$, denoted as $s_{t,(1)}, s_{t,(2)}, \cdots, s_{t,(q)}$. Here $s_{t,j}$ is defined as*

$$s_{t,j} = \sum_{k=1}^{K_s} \log \frac{f_{j,(k)}(\tilde{x}_{j,t}^k)}{f_{j,0}(\tilde{x}_{j,t}^k)} \tag{6}$$

$s_{t,(j)}$ is the order statistics, defined as $s_{t,(1)} \geq s_{t,(2)} \geq \cdots \geq s_{t,(q)} \geq s_{t,(q+1)} \geq \cdots \geq s_{t,(p)}$.

We would like to mention that the computation of (6) in Proposition 1 is actually very efficient. To compute each $s_{t,j}$, it requires the summation of K_s terms, which is of $O(K_s)$ complexity. To compute all p sensing variables at each time t, it requires only $O(pK_s)$ complexity at each time to decide the best sampling layout. Here, the limitation is that $f_{k,j}$ is assumed to be independent over different data dimension j. However, we find that

16

even the distribution of each failure mode is not independent, this approximation can still achieve a pretty reasonable solution.

In this paper, we will only focus on the monitoring of continuous variables and assume that the data follows a normal distribution $f_k \sim N(\boldsymbol{\mu}_k, \Sigma_k)$. However, as derived in the proposed framework, this method can be generalized to other distributions quite easily. Finally, as mentioned in Proposition 1, if we will further assume that Σ_k is diagonal as $\Sigma_k = diag(\sigma_{k,1}^2, \cdots \sigma_{k,p}^2)$, we can derive a simpler formula for $s_{t,j}$ in Proposition 2.

Proposition 2. *Given that* $f_{k,j} = N(\mu_{k,j}, \Sigma_k)$, *where* $\Sigma_k = diag(\sigma_{k,1}^2, \cdots \sigma_{k,p}^2)$. *We can derive* $s_{t,j} = \sum_{k=1}^{K_s} (\frac{1}{\sigma_{k,j}^2}(\tilde{x}_{j,t}^k - \mu_{k,j})^2 - \frac{1}{\sigma_{k,0}^2}(\tilde{x}_{j,t}^k - \mu_{k,j})^2), \ \tilde{x}_t \sim f_k,$

We would like to emphasize that the independence assumption of each failure mode distribution is actually not required for the proposed algorithm. It is only useful to derive the closed-form solution in solving (5). If the spatial dimension is not independent for different failure modes, the proposed planning procedure can still be optimized without the assumptions by approximating the optimal solution. Here, we propose a greedy algorithm to detect the $a_{j,t}$ sequentially. The detailed step is given as follows. First, we can select the first sensing variable j_1 to optimize the sampled version $\tilde{S}_t$ by $j_1 = \arg\max \tilde{S}_t, \sum_j a_{j,t} = 1, a_{j,t} = \{0, 1\}$. After j_1 is decided, we would like to choose the second sensing variable j_2 by $j_2 = arg\max \tilde{S}_t, \sum a_{j,t} = 2, a_{j_1,t} = 1, a_{j,t} = \{0, 1\}$. We will continue the procedure until j_q is selected. In conclusion, the set of the observed sensing index is given as $C_t = \{j_1, \cdots, j_q\}$. Given that we only need to enumerate all p dimensions in each of the q iterations, the time complexity can be reduced to $O(pq)$. Despite the efficiency, the greedy forward selection strategy usually does not produce a global optimum.

Here, we would like to highlight the major difference between the proposed method and the existing literature on monitoring of the i.i.d data stream such as (Zhang and Mei, 2020): 1) the number of failure modes K does not need to be the same as the dimensionality p of the data stream; 2) The normal data distribution f_0 does not need to be i.i.d according to each dimension as $x_j \overset{i.i.d}{\sim} f(x)$, for all j; 3) For each failure mode, it can include overlapping sets of sensing variables.

Finally, we would like to point out a special version of the proposed method and how it links to (Zhang and Mei, 2020), if we are interested in monitoring the i.i.d data stream with the focus of detecting the change of each individual sensing variable.

Proposition 3. *For before change $H_0 : x_j \overset{i.i.d}{\sim} f(x)$, for all j. After change, for j^{th} failure mode, where $j \in \{1, \cdots, p\}$, only 1 distribution changed the distribution to $g(x)$ as $x_j \sim g(x)$, where the rest $x_{j'} \overset{i.i.d}{\sim} f(x), j' \neq j$ still follows the pre-change distribution. The proposed algorithm will result in the sampled updating rule as in (Zhang and Mei, 2020):*

$$R_{j,t} = \begin{cases} \frac{g(\tilde{x}_{j,t})}{f(\tilde{x}_{j,t})}(R_{j,t-1} + 1) & a_{j,t} = 1 \\ \\ R_{j,t-1} + 1 & a_{j,t} = 0 \end{cases} \tag{7}$$

Proposition 3 shows a special case for the proposed algorithm, which assumes the change only affected a few data streams and the algorithms try to identify the change with the resources constraint. Under the current setting, the proposed algorithm will become another sampled version of the TSSRP algorithm. Many previous works, including (Liu et al., 2015; Zhang and Mei, 2020) have studied this setting. However, the proposed algorithm can be generalized into any other joint distributions of different failure modes.

3.3 Properties of the Proposed Algorithm

Here, we will prove two important properties of the algorithms about the bound of the average run length and the failure mode isolation in Theorem 4 and Theorem 5, respectively.

Theorem 4 (Average Run Length). *Let $T = \inf\{t \geq 1 : r_{(1),t} \geq A_2\}$. Then we have that under the null hypothesis where no changes occur, $\mathbb{E}T \geq A_1/K$, $\mathbb{E}T = O(A_1)$, where $A_1 = e^{A_2}$.*

Theorem 4 provides a lower and upper bound for the Average Run Length if no changes occur. Theorem 4 provides us the guidance to select conservative upper and lower bounds of the control limit A. Specifically, $K * ARL$ can serve as the upper bound in the bisection search to speed up the threshold choosing procedure.

Theorem 5 (Failure Mode Isolation). *Assume $\mathbf{X}_1 \ldots \mathbf{X}_\nu \sim f_0$, $\mathbf{X}_{\nu+1} \ldots \sim f_k$. All distributions of failure modes are continuous, and the KL divergence of the distributions of failure mode l and the true failure mode k follows $0 < KL(f_k \| f_l) < \infty$, and $\mathrm{Var}_{x \sim f_k}[\log \frac{f_k}{f_l}] < \infty$, for all $l \neq k$. The probability that $P(r_{k,t} > r_{l,t}) \to 1$ as $t \to \infty$.*

Theorem 5 provides the behavior of the largest SRP statistics when time goes to infinity under the alternative hypothesis (where the failure mode k occurs). It shows that the adaptive sampling algorithm will always be able to isolate the true failure mode k if $t \to \infty$.

In some cases, there might be multiple failure modes happening at different time points after the change point ν, i.e. for some $t_1 > \nu$, $\mathbf{X}_{t_1} \sim f_k$; for some $t_2 > \nu$, $\mathbf{X}_{t_2} \sim f_l$. Corollary 6 is proved.

Corollary 6. *Assume* $\mathbf{X}_1 \ldots \mathbf{X}_\nu \sim f_0$. *Let* $\mathcal{K} = \{k : \mathbf{X}_t \sim f_k \text{ for some } t > \nu\}$ *be the set of true failure modes. If we further assume that the support of different failure modes are non-overlapping, then for any* $k \in \mathcal{K}$ *and* $l \notin \mathcal{K}$, *we have* $\lim_{t \to \infty} P(r_{k,t} > r_{l,t}) = 1$.

Corollary 6 assumes that different failure modes are not overlapping with each other, we can prove in Corollary 6 that SRP statistics of the true failure modes will be larger compared to those of the other potential failure modes. We would like to point out that Corollary 6 is not always true for failure mode distributions that are potentially overlapped with each other. For example, if half of the data after the change follows f_1 and the other half after the change follows f_2, it might be possible that the true failure mode identified would be $f_3 = \frac{1}{2}(f_1 + f_2)$.

3.4   Choice of Parameters

Here, we will present practical guidelines for tuning parameter selection. Given that the number of sensing variables q typically depends on the available sensing resources in the particular applications, we only need to select the following parameters: the number of top-R selected failure modes K_s, the control limit threshold A, and the failure mode distribution f_k and f_0.

Choice of the number of observed failure modes K_s: First, the number of selected failure modes for the monitoring statistics should be smaller than the total number of potential failure modes. Ideally, K_s should be chosen as large as the total number of true failure modes in the system. In practice, we found that increasing K_s to be more than the true number of failure modes in the system would lead the algorithm to explore more

potential failure modes or increase the exploration power. However, if K_s is too large, the algorithm is not able to focus on the actual failure modes, which decreases the exploitation power.

Choice of threshold A**:** The choice of control limit A can be determined by the In-control ARL (or ARL_0). If A is large, ARL_0 will also increase. In practice, we can set an upper bound of the A by utilizing the Theorem 4 and then use the binary search algorithm to find the best A for a fixed ARL_0.

3.5 Selection of failure mode distribution f_k

Finally, the complex joint distributions of different failure modes also bring significant computational complexity, which will be addressed in this paper.

Selecting the failure mode distribution is very important to achieve better change detection and isolation performance. However, it is very challenging to provide accurate failure modes definition for high-dimension data without any domain knowledge. In this work, we mainly focus on detecting mean-shift in high dimension data. We further discuss how to define the failure modes based on our knowledge of the high-dimension data. Overall, there are two strategies for choosing the most appropriate failure mode distribution f_k. 1) If there is prior knowledge about the failure mode distributions, we can set the distribution according to the prior knowledge. For example, if we know that the hot spots are clustered, each failure mode distribution can be assumed as the mean-shift of the IC distribution f_0 with an individual B-spline or Gaussian kernel basis. If we know that the post-change distribution is sparse, a simple way is to set the failure mode distribution as the mean

shift of each individual sensor. 2) If we do not know the failure mode distributions, we can collect some samples for each failure mode and use these samples to estimate the failure mode distribution.

4 Simulation Study

Here, we will evaluate the proposed method in a simulation study. We will start with the simulation setup in Section 4.1 with two different scenarios: the non-overlapping case and the overlapping case. Then, we will evaluate the proposed algorithm in these two scenarios. To test the robustness of the proposed algorithm in the case of multiple failure modes coexist, we also perform the sensitivity analysis to evaluate the performance in Appendix 2.

4.1 Simulation Setup

Here, we will discuss the two scenarios for the simulation setup. We are trying to distinguish multiple failure modes by whether these failure modes have overlapping support. For example, for the first "non-overlapping" case, we assume that different failure modes f_i and f_j have non-overlapping support.

4.1.1 Scenario 1: the non-overlapping case

We will first discuss the non-overlapping case of the proposed method. In the simulation, we let the data dimension $p = 1000$, the number of failure mode $K = 50$, each time the algorithm will select $q = 10$ sensors at each time. Here, we assume the normal data or in

(a) All 50 Non-overlapping po- (b) All 49 overlapping potential
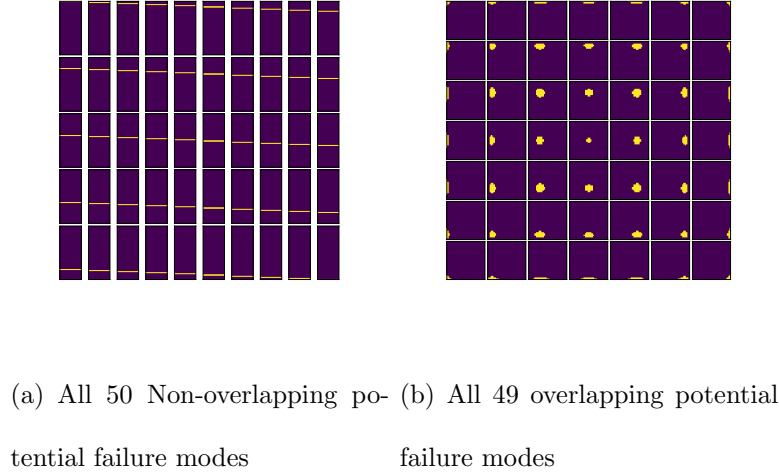tential failure modes   failure modes

Figure 4: Failure Modes for Overlapping and Non-overlapping Cases

control (IC) data follows $f_0 = N(0, I)$. After the change, the out-of-control (OC) data have

K failure modes, where $f_k = N(\boldsymbol{\mu}_k, I)$ and $\boldsymbol{\mu}_k = \sum_{j=sk}^{(s+1)k} e_j$. $e_j = (0, \cdots, 1, \cdots, 0)$, and

only $j^t h$ element is 1. Here, three failure modes have been selected after the change. In

another word, if we organize the 1000 data streams into a 50×20 image, each failure pattern

would be each row of pixels which can be visualized in Fig. 4(a). Therefore, different failure

modes are not overlapped, given they contain different sensing variables.

Finally, we assume that we are only accessible to 10 out of $p = 1000$ data streams to

observe. At each time step, we adaptive select data streams to monitor the whole process.

We want to detect the correct failure mode as soon as possible.

4.1.2   Scenario 2: the overlapping case

We will discuss the second scenario, where the failure patterns are generated as small spatial

clusters. In this case, we set the data as 2-D images with size 30×30 with total dimension

$p = 900$. Here, the failure modes are generated using B-spline basis with 7 knots in both x and y directions. As shown in Fig. 4(b), we end up with $7^2 = 49$ potential failure modes. After the change happens, we randomly select a failure mode as the true failure mode. In this situation, some failure modes might overlap with each other, which will be more challenging for the algorithm to isolate the real changes. Finally, during the monitoring process, we can select 10 out of $p = 900$ data streams adaptively to observe online.

4.2   Simulation Result

Here, we will compare the proposed MTSSRP with the following benchmark methods: 1) TSSRP method (Zhang and Mei, 2020), which is introduced in detail in Appendix. 2) TRAS method Liu et al. (2015), where the local CUSUM statistics is used for each individual data stream and later fused together via the Top-R rule. To show the upper-bound and lower-bound performance, we will also add three simple alternatives: 1) Random, where we randomly select q sensors at each time step with the top-r statistics by monitoring each sensor individually. 2) Oracle, where we not only have access to all the data streams but also the failure mode distribution information using the monitoring statistics as MTSSRP. 3) MRandom, where we apply the same monitoring statistics as MTSSRP, which considers the failure mode distribution information in the monitoring statistics, but we randomly select the sensors at each time step. We evaluate the proposed method with two metrics which are detection delay and failure isolation accuracy.

First, we would like to compare the detection delay of the proposed method and all the benchmark methods. Here, we set the in-control average run length (i.e., denoted as

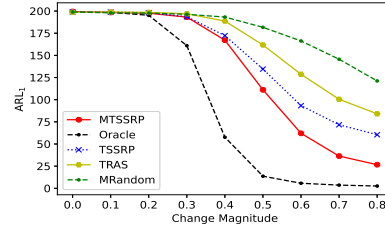Table 1: Average Run Length and Failure Mode Isolation Accuracy for single failure

Case		nonoverlap				overlap			
Change Magnitude		$\delta = 0.5$		$\delta = 0.8$		$\delta = 0.5$		$\delta = 0.8$	
Metrics		ARL_1	Accuracy	ARL_1	Accuracy	ARL_1	Accuracy	ARL_1	Accuracy
Oracle		13.71(11.56)	0.99(0.04)	2.52(1.08)	1.0(0.0)	13.68(12.05)	1.0(0.05)	2.57(1.43)	1.0(0.03)
Competing Methods	MTSSRP	**112.05(66.96)**	**0.75(0.43)**	**26.49(17.09)**	**0.99(0.03)**	**111.16(66.01)**	**0.75(0.43)**	**26.66(17.71)**	**0.98(0.14)**
	TSSRP	124.66(58.2)	0.55(0.50)	61.19(30.67)	0.81(0.39)	134.51(58.38)	0.48(0.5)	60.5(35.92)	0.81(0.39)
	TRAS	164.84(49.07)	0.46(0.50)	75.53(42.51)	0.98(0.13)	161.75(51.48)	0.47(0.5)	84.08(50.73)	0.93(0.25)
	MRandom	178.1(41.49)	0.28(0.45)	104.55(48.71)	0.92(0.28)	181.67(39.91)	0.23(0.42)	121.16(54.09)	0.8(0.4)
	Random	199.97(1.07)	-	199.66(5.08)	-	200.0(0.0)	-	199.95(1.61)	-

ARL_0) for all methods as 200 and compare their out-of-control ARL or average detection delay (i.e., denoted as ARL_1) with 1000 replications. Here, we will compute the ARL_1 for different change magnitudes ($\delta = 0.5, 0.8$) in Table 1. We also compared the ARL_1 and Isolation Accuracy from different magnitude δ (0.1 to 0.8) in Fig. 5. From the results, we can see that the proposed MTSSRP has better ARL_1 compared to other benchmark methods. MTSSRP performs much better than MRandom, which validates the efficiency of the proposed sampling strategy. The advantage of MTSSRP over TSSRP shows that considering the failure mode information can greatly improve the performance. We further compare the isolation accuracy as shown in Fig. 5(c) and Fig. 5(d). It can be seen that the proposed MTSSRP achieves better performance than others. It can also reach pretty high accuracy when δ is greater than 0.6.

To understand how the proposed algorithm balances the exploration and exploitation automatically, we would like to plot both the SR statistics for each failure mode (i.e., in red) and when this particular failure mode has observed sensors (i.e., in black dot) for both the potential failure mode (i.e., failure mode doesn't happen in this run) and the true failure
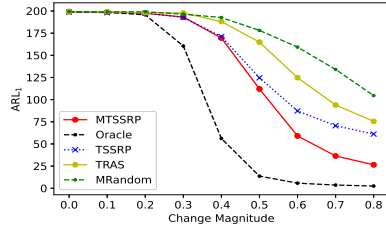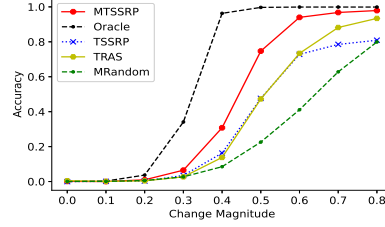
(a) ARL_1 in Non-overlapping Cases (b) ARL_1 in Overlapping Cases
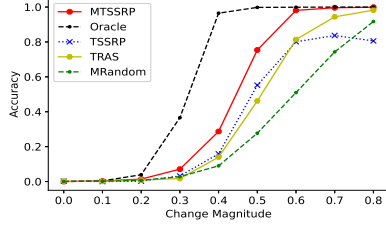
(c) Accuracy in Non-overlapping (d) Accuracy in Overlapping Cases
Cases

Figure 5: Out-of-control Average Run Length (ARL_1) and Failure isolation accuracy for

Different Change Magnitude δ

mode in Fig. 6 together. Here, the failure mode with the observed sensors can be defined as

that there are observed sensing variables located in the non-zero location of the mean-shift

of that particular failure mode. From Fig. 6, it is clear that the statistics for potential is

quite small compared to the statistics for the true failure mode. From Fig. 6(a), we can

also observe that the SR statistics will grow naturally if this particular failure mode is not

observed. This will encourage the sensing variables to be allocated to this particular failure

mode eventually. Furthermore, if the particular failure mode is observed where no change

occurs, the monitoring statistics will drop significantly, indicating that this failure mode

has dropped significantly. On the other hand, from Fig. 6(b), we can clearly see that after

time 100 when the change occurs, the true failure mode statistics will grow significantly as

long as that particular failure mode is observed.

26

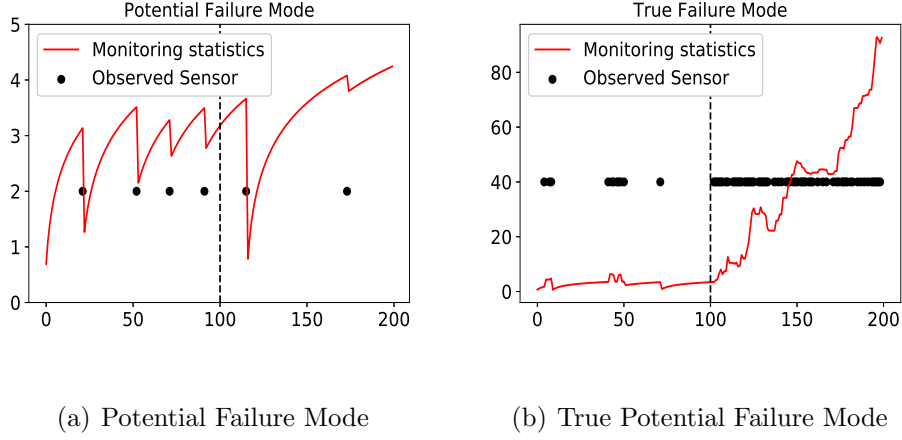(a) Potential Failure Mode   (b) True Potential Failure Mode

Figure 6: SR statistics for the True Failure Mode. The change happens at time $t = 100$. The left figure shows the monitoring statistics for the potential failure mode, where the monitoring statistics are small. The right figure shows the monitoring statistics for the true failure mode, and the monitoring statistics increase dramatically at time $t = 100$.

5   Case Study

In this section, we will evaluate the proposed MTSSRP algorithm in the laser powder bed fusion process monitoring. The code and dataset have been provided in `https://github.com/hyan46/ExtendedTSSRP`. We will evaluate the performance of the MTSSRP and compare it with the state-of-the-art benchmark methods in a case study of hotspot detection in a 3-D printing process. Two additional case studies about the Tonnage signal monitoring and COVID-19 hotspot detection have been provided in the Supplementary Material.

Here, we will implement the proposed algorithm into the hot-spots detection in the process monitoring in the Laser Powder Bed Fusion (LPBF) process. A 300 fps video sequence was acquired during the realization of one layer of the part by using the setup
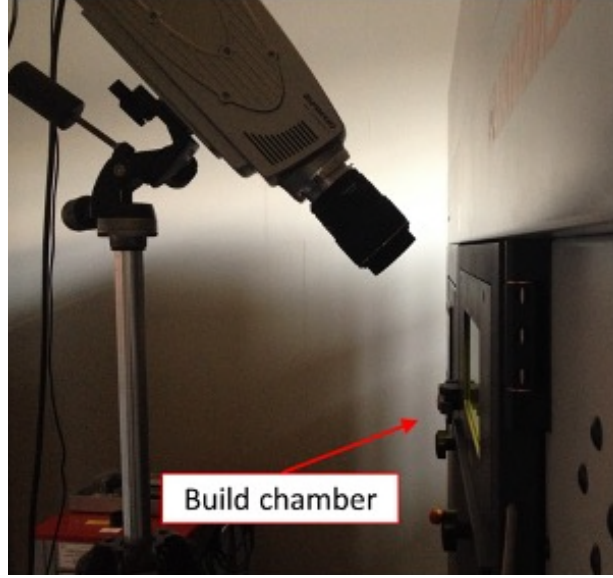
27

Figure 7: Hot-spots Detection in LPBF

shown in Fig. 7, which consists of a thermal camera mounting outside the LPBF chamber monitoring the hot-spots events. The observed image is of size 121×71 pixels. Previous studies showed that the occurrence of local over-heating conditions might yield geometrical distortions (Yan et al., 2020; Colosimo and Grasso, 2018). The hot-spots caused by the formation of solidified balls will cause the local heat accumulation and inflate from one layer to another. Therefore, the overall goal of this study is to detect such hot-spots quickly. For more details about the setup of this experiment and some preliminary works related to this dataset, please refer to (Grasso et al., 2017; Colosimo and Grasso, 2018; Yan et al., 2020). The dataset is also publicly available at `http://doi.org/10.6084/m9.figshare.7092863`.

In this example, it is not easy to obtain the failure mode data beforehand, and therefore, we rely on domain knowledge to define the failure mode distribution. First, we know that the hot-spots must be in the scanning path. Second, we know that the hot-spots must be locally clustered. Therefore, we define the failure modes as each individual B-spline basis

28

overlapped with the printing regions. In this dataset, there are four different events starting from 77, 94, 150, and 162. From Table 2, the proposed algorithm can detect the change at time 79, 95, 152, and 162 with only 200 sensing variables out of 8591 sensing variables. In comparison, TSSRP can detect all four events but with a much larger delay. However, TRAS can only detect Event 4, which fails to detect the first three events. The original image frame, the sampling patterns, and the selected failure modes at the detected time for these four events are shown in Fig. 8. From Fig. 8, we can observe that the algorithm can quickly converge the sampled points to the true hot-spots location at the upper left corner.
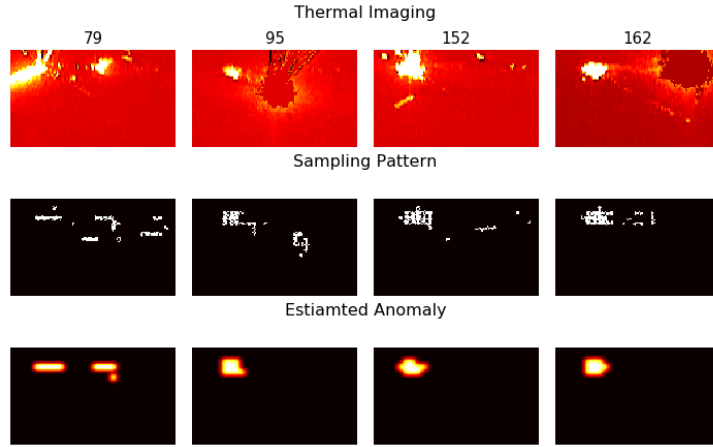

Figure 8: Original Thermal Images, Sampling Patterns, and Detected Failure Modes

6 Conclusion

Online change detection of high-dimensional data under multiple failure modes is an important problem in reality. In this paper, we propose to borrow the concept from Bayesian change point detection and MAB to adaptively sample useful local components, given the

Table 2: Detection Time in the LPBF Process

Event Times		Time of first signal			
		Event 1	Event 2	Event 3	Event 4
Actual Change Time		77	94	150	162
Competing Methods	MTSSRP	**79**	**95**	**152**	**162**
	TSSRP	80	99	156	164
	TRAS	-	-	-	165

distributions of the multiple failure modes. Our proposed algorithm can balance between exploration of all possible system failure modes or exploitation of the most probable system failure mode. Furthermore, we also studied the properties of the proposed methods and showed the proposed algorithm could isolate the correct failure mode. Our simulation and case study show that the proposed algorithm, by considering the failure mode information, can significantly reduce the detection delay.

7 Supplemental Material

The online supplementary materials contain details of the change point detection framework, simulation results, three additional case studies (i.e., tonnage monitoring, COVID-19 monitoring), and technical proofs.

8   Acknowledgement

9   Funding and Disclosure

References

Hock Peng Chan et al. Optimal sequential detection in multi-stream data. *The Annals of Statistics*, 45(6):2736–2763, 2017.

Shing I. Chang and Srikanth Yadama. Statistical process control for monitoring non-linear profiles using wavelet filtering and b-spline approximation. *International Journal of Production Research*, 48(4):1049–1068, 2010.

Jie Chen, Wenyi Zhang, and H Vincent Poor. A bayesian approach to sequential change detection and isolation problems. *IEEE Transactions on Information Theory*, 67(3): 1796–1803, 2020.

Haeran Cho and Piotr Fryzlewicz. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 475–507, 2015.

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

Ana María Estrada Gómez, Dan Li, and Kamran Paynabar. An adaptive sampling strategy for online monitoring and diagnosis of high-dimensional streaming data. *Technometrics*, 64(2):253–269, 2022.

Aditya Gopalan, Braghadeesh Lakshminarayanan, and Venkatesh Saligrama. Bandit quickest changepoint detection. *Advances in Neural Information Processing Systems*, 34: 29064–29073, 2021.

M Grasso, BM Colosimo, and M Pacella. Profile monitoring via sensor fusion: the use of pca methods for multi-channel data. *International Journal of Production Research*, 52 (20):6110–6135, 2014.

Marco Grasso, Vittorio Laguzza, Quirico Semeraro, and Bianca Maria Colosimo. In-process monitoring of selective laser melting: spatial detection of defects via image data analysis. *Journal of Manufacturing Science and Engineering*, 139(5), 2017.

Jing Li and Jionghua Jin. Optimal sensor allocation by integrating causal models and set-covering algorithms. *IIE Transactions*, 42(8):564–576, May 2010.

Kaibo Liu and Jianjun Shi. Objective-oriented optimal sensor allocation strategy for process monitoring and diagnosis by multivariate analysis in a Bayesian network. *IIE Transactions*, 45(6):630–643, June 2013.

Kaibo Liu, Xi Zhang, and Jianjun Shi. Adaptive sensor allocation strategy for process monitoring and diagnosis in a bayesian network. *IEEE Transactions on Automation Science and Engineering*, 11(2):452–462, 2013.

Kaibo Liu, Yajun Mei, and Jianjun Shi. An Adaptive Sampling Strategy for Online High-Dimensional Process Monitoring. *Technometrics*, 57(3):305–319, July 2015.

Gary Lorden. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, pages 1897–1908, 1971.

Durga P Malladi and Jason L Speyer. A generalized shiryayev sequential probability ratio test for change detection and isolation. *IEEE Transactions on Automatic Control*, 44(8): 1522–1534, 1999.

Yajun Mei. Efficient scalable schemes for monitoring a large number of data streams. *Biometrika*, 97(2):419–433, 2010.

Yajun Mei. Quickest detection in censoring sensor networks. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2148–2152. IEEE, 2011.

I Nikiforov, V Varavva, and V Kireichikov. Application of statistical fault detection algorithms to navigation systems monitoring. *Automatica*, 29(5):1275–1290, 1993.

Igor V Nikiforov. A generalized change detection problem. *IEEE Transactions on Information theory*, 41(1):171–187, 1995.

Kamran Paynabar and Jionghua Jin. Characterization of non-linear profiles variations using mixed-effect models and wavelets. *IIE transactions*, 43(4):275–290, 2011.

Kamran Paynabar, Changliang Zou, and Peihua Qiu. A change-point approach for phase-i analysis in multivariate profile monitoring and diagnosis. *Technometrics*, 58(2):191–204, 2016.

Peihua Qiu, Changliang Zou, and Zhaojun Wang. Nonparametric profile monitoring by mixed effects modeling. *Technometrics*, 52(3), 2010.

Haojie Ren, Changliang Zou, Nan Chen, and Runze Li. Large-scale datastreams surveillance via pattern-oriented-sampling. *Journal of the American Statistical Association*, pages 1–15, 2020.

Andi Wang, Xiaochen Xian, Fugee Tsung, and Kaibo Liu. A spatial-adaptive sampling procedure for online monitoring of big data streams. *Journal of Quality Technology*, 50 (4):329–343, October 2018.

Yuan Wang and Yajun Mei. Large-scale multi-stream quickest change detection via shrinkage post-change estimation. *IEEE Transactions on Information Theory*, 61(12):6926–6938, 2015.

Alan S Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12(6):601–611, 1976.

Xiaochen Xian, Andi Wang, and Kaibo Liu. A Nonparametric Adaptive Sampling Strategy for Online Monitoring of Big Data Streams. *Technometrics*, 60(1):14–25, January 2018.

Yao Xie and David Siegmund. Sequential multi-sensor change-point detection. In *2013 Information Theory and Applications Workshop (ITA)*, pages 1–20. IEEE, 2013.

H. Yan, K. Paynabar, and J. Shi. Image-based process monitoring using low-rank tensor decomposition. *IEEE Transactions on Automation Science and Engineering*, 12(1):216–227, 2015.

Hao Yan, Kamran Paynabar, and Jianjun Shi. Anomaly detection in images with smooth background via smooth-sparse decomposition. *Technometrics*, 59(1):102–114, 2017.

Hao Yan, Kamran Paynabar, and Jianjun Shi. Real-Time Monitoring of High-Dimensional Functional Data Streams via Spatio-Temporal Smooth Sparse Decomposition. *Technometrics*, 60(2):181–197, April 2018.

Hao Yan, Marco Grasso, Kamran Paynabar, and Bianca Maria Colosimo. Real-time detection of clustered events in video-imaging data with applications to additive manufacturing. *arXiv preprint arXiv:2004.10977*, 2020.

Xiaowei Yue, Hao Yan, Jin Gyu Park, Zhiyong Liang, and Jianjun Shi. A wavelet-based penalized mixed-effects decomposition for multichannel profile detection of in-line raman spectroscopy. *IEEE Transactions on Automation Science and Engineering*, 15(3):1258–1271, 2017.

Wanrong Zhang and Yajun Mei. Bandit change-point detection for real-time monitoring high-dimensional data under sampling control. *arXiv preprint arXiv:2009.11891*, 2020.