18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

Change Detection with an Unknown Sensor Subset: More Information is Not Always Better

Marco Guerriero, Dayu Huang, Jayakrishnan Unnikrishnan, Michael Lexa, Satish Iyengar and Fred Wheeler
Sensor and Signal Analytics, Software Sciences & Analytics,
GE Global Research Center, Niskayuna, NY, USA

***Abstract*—We study the problem of detecting changes in the environment based on observations taken by multiple sensors under the setting in which the change affects an unknown subset of the sensors. We explore different approaches to this problem and relate different stopping rules for change detection with multiple sensors in a single framework. We introduce four new different stopping rules: T_{MAP}, $T_{SOFT\text{–}MAP}$, T_{ML} and T_{Order}. While the first three rules rely on the information contained in the posterior probability of the sensor being affected by the change and on the Maximum Likelihood (ML) estimator of the sensor being affected, respectively, the last one is based on the order statistic of the local likelihood ratios at the sensors. We show that: i) T_{ML}, which is based on the "scan statistic", is equivalent to the Bayesian stopping rule $T_3(p_0 = 1)$ that was recently introduced by Xie and Siegmund; ii) T_{Order} is the counterpart of T_{ML} when the cardinality of the affected sensors is known. Surprisingly, the additional information about the cardinality does not always lead to better detection performance. A derivation of an upper bound for the false alarm rate for T_{Order} is given and a comparative numerical analysis of the different stopping rules is provided in order to relate their performances.**

***Index Terms*—Change Detection, Multiple Sensors, Scan Statistic, Sparse Estimation.**

I. Introduction

Change detection dates back over eighty years with the work of Shewart [1] in the context of quality control and significant independent contributions were subsequently made by Page and Shirayev [2, 3] and by Lorden and Pollak [4, 5] in the intervening years. Nowadays there is a huge demand for change detection in a wide range of important applications [6, 7] such as quality control engineering, diagnostic and prognostic, condition-based maintenance, industrial internet [8], tracking, intrusion detection, surveillance and monitoring. The problem of change detection for multi-sensor situation was first introduced by Tartakovsky and Veeravalli in [9] and further explored in [10, 11]. The change-points can be simultaneous in time across all sensors as in [12] or they can occur at different times as in [11]. Mei [11] introduced a procedure that is based on the sum of the local CUSUM statistics from local sensors, while Tartakovsky et al in [12] proposed the CUSUM of the sum of the likelihood ratios at the local sensors. Both procedures have optimal asymptotic properties; however, when the change is not observed by all sensors the performance of both procedures degrades. In fact, when only a subset of sensors senses the change, the "noise" produced by the unaffected sensors can cause longer detection delays.

In this paper, we study multi-sensor change detection problems where ***an unknown subset of sensors*** are affected by a change, but where the change is assumed to occur at the same time among those affected. As previously mentioned, multi-sensor (also known as multi-channel) change detection has been extensively investigated in the case where all sensors see the change [11, 12]. Less explored in the literature is the situation where only a subset of sensors observes the change.

A. Related Work

The problem of estimating the correct portion of the sensors experiencing the change, adds a new challenge to the problem of change detection with multiple sensors. This estimation problem falls under the broader problem of estimating the support set of a vector which is currently a very active area research for those working in hypothesis testing (see [13] and references therein) and sparse estimation [14–17]. In the context of multi-sensor change detection many different aspects have been investigated. Mei [18] introduced a detection scheme based on the sum of the top-*r* local CUSUM statistics with *r* being the number of affected sensors. This procedure is effective when the sensors change their state at different times, and in fact, has been shown to be asymptotically optimal. However, as reported in [19], the assumption of concurrence of the change points across the sensors, has a negative impact on the efficiency of this procedure. Xie and Siegmund [19] recently introduced a Bayesian mixture test statistic which incorporates prior knowledge on the fraction p_0 of the affected sensors. In their paper it is shown that the procedure is robust to the choice of p_0. In [20] the authors compared the performance of an Adaptive CUSUM [21] with a Parallel CUSUM [22]. Currently, Banerjee and Veeravalli [23] are studying a data-efficient solution for this problem.

B. Summary of results and contributions

In our framework, we introduce an auxiliary boolean variable that indicates whether a sensor is affected by the change. For the Bayesian approach, this variable is considered to be a hidden (or unobserved) *random* quantity that has an associated prior distribution. For maximum likelihood approach, this variable is considered to be a hidden *nonrandom* quantity. Depending on which viewpoint is adopted, different stopping rules can be derived. We introduce four new stopping rules: T_{MAP}, $T_{SOFT\text{–}MAP}$, T_{ML}, and T_{Order}. While the first two rules rely on a Bayesian model for the affected sensors, the last two assume that the affected sensors are members of a nonrandom, but unknown, subset. Interestingly, this framework reveals several connections among seemingly unrelated stopping rules. T_{ML} is equivalent to the Bayesian stopping rule $T_3(p_0 = 1)$, recently introduced by Xie and Siegmung in [19]. We also demonstrate that T_{ML} and T_{Order} are based on the *scan statistic* [13, 24] for unknown and known cardinality of the affected sensors, respectively. One would expect that the additional information represented by knowing the cardinality[1] of the affected sensors, would improve the change detection performance of any detector. However, after a careful numerical analysis, we report a counter-intuitive result. In the case of change in mean with Gaussian observations, the prior knowledge on the sparsity, does not necessarily imply higher change detection performance. Surprisingly, T_{Order} does not uniformly outperform T_{ML}: In case of moderate sparsity, the T_{ML} is better. Finally, we illustrate the analysis developed in this paper using several numerical examples.

II. Problem Formulation

We consider a scenario with N sensors where each sensor $n \in \{1,\ldots,N\}$ makes measurements y_i^n, $i \in \{1,2,\ldots\}$. We assume that an unknown subset of sensors $\mathcal{S}$ experience a change in their measurements at some common but unknown time k and that the measurements are mutually independent within and across the sensors. We are interested in a multi-sensor change detection test that detects a change-point, if one occurs, as quickly as possible while maintaining a low false alarm rate. Formally, we are interested in finding effective stopping rules for the test,

$$\begin{aligned} H_\infty :& y_i^n \sim f_0, i \in \{1,2\ldots\}, \forall n \in \{1,\ldots,N\} \\ \text{and } H_k :& y_i^n \sim f_0, i \in \{1,\ldots,k-1\}, n \in \mathcal{S}^c \\ & y_i^n \sim f_1, i \in \{k,\ldots\}, n \in \mathcal{S}. \end{aligned} \tag{1}$$

The observations y_i^n under both hypotheses are initially independent and identically distributed (i.i.d.). Under H_k, however, the distribution of the observations within $\mathcal{S}$ switches from the probability density function f_0 to f_1 at some unknown time k. We use the notation $p_k(\cdot)$ to denote the likelihood function under H_k and $E_k[\cdot]$ to denote the expectation operator when the change happens at time k. Likewise, $p_\infty(\cdot)$ and $E_\infty[\cdot]$ denote the likelihood function and the expectation operator under H_∞. We denote the log-likelihood ratio between H_k and H_∞ by LL_k. We also refer to the cardinality of $\mathcal{S}$ as $|\mathcal{S}|$, or equivalently, use a sparsity index $\gamma = (1 - |\mathcal{S}|/N)$. When γ is close to one, few sensors experience a change; when a γ close to zero, most of them experience a change.

To compare the performance of the different stopping rules, we use two classical figures of merit: the worst-case detection delay (WDD) and the false alarm rate (FAR) defined as [4]

$$\text{WDD}(T) = \sup_{k\geq 1} \operatorname{ess\,sup} E_k \left[(T-k+1)^+ \,|\mathcal{F}_{k-1} \right], \tag{2}$$

$$\text{FAR}(T) = \frac{1}{E_\infty[T]}, \tag{3}$$

[1]Henceforth, we will associate the concept of cardinality of the subset of sensors with the notion of *sparsity*; the smaller the cardinality, the sparser the problem is.

where x^+ denotes the positive part of x, that is $\max(0,x)$ and $\mathcal{F}_k$ denotes the sigma algebra [25] generated by $(y_1^n, y_2^n, \ldots, y_k^n)$. Our goal is to design stopping rules [26] for detecting the change-point k as quickly as possible while maintaining the low false alarm rates.

III. Review of existing stopping rules

This section briefly reviews several relevant stopping rules. We believe it is important to analyze their underlying *model assumptions* that were used to obtain them. In fact, a closer look at these assumptions might provide useful insights that help to derive new stopping rules and relate them to the existing ones. In [19], Xie introduced the mixture stopping rule T_1 defined as

$$T_1 = \inf\left\{ t : \max_{k<t} \sum_{n=1}^{N} \log\left(1 - p_0 + p_0 \exp(l_n^+(t,k)\right) \geq b \right\}, \tag{4}$$

where the log-likelihood ratio $l_n(t,k)$ of the observations accumulated by time $t > k$ at the sensor n is given by

$$l_n(t,k) = \sum_{i=k}^{t} \log \frac{f_1(y_i^n)}{f_0(y_i^n)}. \tag{5}$$

Here p_0 is the fraction of affected sensors, that is $\frac{|\mathcal{S}|}{N}$. Based on the observation that $\log(1-p_0+p_0\exp(x))$ is large only if x is large and that it is approximately equal to $[x+\log(p_0)]^+$, Xie also proposed the stopping rule,

$$T_3 = \inf\left\{ t : \max_{k<t} \sum_{n=1}^{N} [l_n(t,k) + \log(p_0)]^+ \geq b \right\}. \tag{6}$$

As noted in [19], the stopping rule $T_3(p_0)$ with $p_0 = 1$ is similar to the procedure proposed by Tartakovsky and Veeravalli in [12],

$$T_{Sum} = \inf\left\{ t : \max_{k<t} \sum_{n=1}^{N} l_n(t,k) \geq b \right\}, \tag{7}$$

but with a very important difference represented by the positive part $l_n^+(t,k)$, which has the intrinsic feature of filtering out the "noise" from the unaffected sensors. Another stopping rule which is effective when only one or a very small number of sensors is affected by the change point is [9]:

$$T_{Max} = \inf\left\{ t : \max_{k<t} \max_{n=1}^{N} l_n(t,k,\mu) \geq b \right\}. \tag{8}$$

We conclude this section with a rule introduced by [11],

$$T_{Mei} = \inf\left\{ t : \sum_{n=1}^{N} \max_{k<t} l_n(t,k) \geq b \right\}. \tag{9}$$

However, as discussed in [19], this last procedure is not very effective in our current setting where the affected sensors change state simultaneously. For this reason, we will not further consider the T_{Mei} rule in our comparative analysis.

IV. New testing procedures

To mathematically express which sensors are affected by the change we introduce the *hidden* variables $z_n, n = 1, \dots, N$,

$$z_n = \begin{cases} 1 & \text{if sensor } n \text{ is affected by the change,} \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

These variables are not directly observed, but information about their values can be estimated from the sensors' post-change observations. In what follows, we examine novel Bayesian and maximum likelihood formulations, and in addition, propose a stopping rule that uses partial knowledge of $\mathcal{S}$. Specifically, it leverages knowledge of the cardinality of $\mathcal{S}$, i.e., the *number* of affected sensors.

A. Bayesian approach: z_n modeled as a random variable

For the Bayesian formulation, we assign prior probability distributions to $\mathbf{z} = [z_1, \dots, z_N]$ under each hypothesis:

$$p_k(\mathbf{z}) = \prod_{n=1}^{N} p_0^{z_n}(1-p_0)^{1-z_n} \tag{11}$$

$$p_\infty(\mathbf{z}) = \begin{cases} 1 & \text{for } \mathbf{z} = \mathbf{0} \\ 0 & \text{else.} \end{cases} \tag{12}$$

This model is equivalent to the one used by Xie [19] where each sensor n is affected by the change with probability p_0 (independently from one sensor to the next).

Let $\mathbf{y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]$ denote the augmented vector of the observations where $\mathbf{y}_n = [y_1^n, \dots, y_k^n]$. The joint likelihood of $(\mathbf{y}, \mathbf{z})$ under H_k is

$$\begin{aligned} p_k(\mathbf{y}, \mathbf{z}) &= p_k(\mathbf{z}) p_k(\mathbf{y}|\mathbf{z}) \\ &= \prod_{n=1}^{N} p_0^{z_n}(1-p_0)^{1-z_n} \\ &\left[z_n \prod_{i=1}^{k-1} f_0(y_i^n) \prod_{i=k}^{t} f_1(y_i^n) + (1-z_n) \prod_{i=1}^{t} f_0(y_i^n) \right], \end{aligned} \tag{13}$$

and under H_∞ is

$$\begin{aligned} p_\infty(\mathbf{y}, \mathbf{z}) &= p_\infty(\mathbf{y}) p_\infty(\mathbf{z}) \\ &= \prod_{n=1}^{N} \prod_{i=1}^{t} f_0(y_i^n). \end{aligned} \tag{14}$$

The log-likelihood ratio is then

$$\begin{aligned} LL_k(\mathbf{y}, \mathbf{z}) &= \log \frac{p_k(\mathbf{y}, \mathbf{z})}{p_\infty(\mathbf{y}, \mathbf{z})} \\ &= \sum_{n=1}^{N} z_n \log(p_0) + (1-z_n)\log(1-p_0) \\ &+ \log \left[z_n \prod_{i=k}^{t} \frac{f_1(y_i^n)}{f_0(y_i^n)} + (1-z_n) \right]. \end{aligned} \tag{15}$$

Since $\mathbf{z}$ is not directly observed, we can replace it by its maximum a posterior (MAP) estimate, $\hat{z}^{MAP} = \max_{\mathbf{z}} p_k(\mathbf{z}|\mathbf{y})$. It is straightforward to show that for each n

$$\hat{z}_n^{MAP}(k) = \begin{cases} 1 & \text{when } \frac{p_0}{1-p_0} \prod_{i=k}^{t} \frac{f_1(y_i^n)}{f_0(y_i^n)} \geq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

An alternative expression for $\hat{z}_n^{MAP}$ can be found by considering the posterior probability

$$\begin{aligned} w_n(k) &= p_k(z_n = 1|\mathbf{y}_n) \\ &= \frac{p_k(\mathbf{y}_n|z_n = 1) p_k(z_n = 1)}{p_k(\mathbf{y}_n)} \\ &= \frac{1}{1 + \frac{1-p_0}{p_0} \prod_{i=k}^{t} \frac{f_0(y_i^n)}{f_1(y_i^n)}}. \end{aligned} \tag{17}$$

Comparing (17) to (16), one sees that (16) is equivalent to

$$\hat{z}_n^{MAP}(k) = \begin{cases} 1 & \text{if } p_k(z_n = 1|\mathbf{y}_n) \geq 0.5 \\ 0 & \text{if } p_k(z_n = 1|\mathbf{y}_n) < 0.5. \end{cases} \tag{18}$$

Substituting the MAP estimator into the log-likelihood ratio (15) yields the MAP stopping rule

$$\begin{aligned} T_{MAP} = \inf \Bigg\{ t : \max_{k<t} \sum_{n=1}^{N} & \hat{z}_n^{MAP}(k) \log(p_0) \\ &+ (1 - \hat{z}_n^{MAP}(k)) \log(1-p_0) \\ &+ \log \left[\hat{z}_n^{MAP}(k) \prod_{i=k}^{t} \frac{f_1(y_i^n)}{f_0(y_i^n)} + (1 - \hat{z}_n^{MAP}(k)) \right] \geq b \Bigg\}. \end{aligned} \tag{19}$$

Another possible stopping rule results by substituting a "soft" estimate of z_n into the log-likelihood ratio. In particular, we propose to insert the posterior probability $w_n(k) = p_k(z_n = 1|\mathbf{y}_n)$ for z_n. Thus instead of using the measurements to definitively decide whether a sensor is affected by a change, we use the data to weight each local likelihood ratio. The resulting soft MAP stopping rule is

$$\begin{aligned} T_{SOFT\text{–}MAP} = \inf \Bigg\{ t : \max_{k<t} \sum_{n=1}^{N} & w_n(k) \log(p_0) \\ &+ (1 - w_n(k)) \log(1-p_0) \\ &+ \log \left[w_n(k) \prod_{i=k}^{t} \frac{f_1(y_i^n)}{f_0(y_i^n)} + (1 - w_n(k)) \right] \geq b \Bigg\}. \end{aligned} \tag{20}$$

Remark. By marginalizing the joint distributions as per equations (13) and (14) over z we can define the log-likelihood ratio $LL_k(\mathbf{y})$ as

$$\begin{aligned} LL_k(\mathbf{y}) &= \log \frac{p_k(\mathbf{y})}{p_\infty(\mathbf{y})} \\ &= \sum_{n=1}^{N} \log \left(1 - p_0 + p_0 \prod_{i=k}^{t} \frac{f_1(y_i^n)}{f_0(y_i^n)} \right) \end{aligned} \tag{21}$$

This log-likelihood ratio is used in the mixture stopping rule T_1 in (4). This clearly indicates the Bayesian nature of Xie's stopping rule.

B. ML approach: z_n modeled as nonrandom variable

Under the assumption that z_n is a deterministic unknown variable, the global likelihood functions under the post-change hypothesis H_k and the pre-change hypothesis H_∞, can be written as

$$p_k(\mathbf{y};\mathbf{z}) = \prod_{n=1}^{N}\left[z_n \prod_{i=1}^{k-1} f_0(y_i^n) \prod_{i=k}^{t} f_1(y_i^n) + (1-z_n)\prod_{i=1}^{t} f_0(y_i^n)\right] \quad (22)$$

and

$$p_\infty(\mathbf{y}) = \prod_{n=1}^{N}\prod_{i=1}^{t} f_0(y_i^n), \quad (23)$$

respectively. The semicolon notation in $p_k(\mathbf{y};\mathbf{z})$ indicates that $\mathbf{z}$ is a nonrandom quantity and distinguishes from the Bayesian case $p_k(\mathbf{y},\mathbf{z})$. The global log-likelihood ratio $LL_k(\mathbf{y};\mathbf{z})$ is

$$LL_k(\mathbf{y};\mathbf{z}) = \log \frac{p_k(\mathbf{y};\mathbf{z})}{p_\infty(\mathbf{y})} = \sum_{n=1}^{N} \log\left[z_n \prod_{i=k}^{t} \frac{f_1(y_i^n)}{f_0(y_i^n)} + (1-z_n)\right]. \quad (24)$$

We can show that the ML estimator $\hat{z}_n^{ML}(k)$ which maximizes (22), is given by

$$\hat{z}_n^{ML}(k) = \begin{cases} 1 & \text{if } \prod_{i=k}^{t} \frac{f_1(y_i^n)}{f_0(y_i^n)} \geq 1 \\ 0 & \text{otherwise.} \end{cases} \quad (25)$$

This suggests the following stopping rule

$$T_{ML} = \inf\left\{t : \max_{k<t} \sum_{n=1}^{N} \log\left[\hat{z}_n^{ML}(k) \prod_{i=k}^{t} \frac{f_1(y_i^n)}{f_0(y_i^n)} + (1-\hat{z}_n^{ML}(k))\right] \geq b\right\} \quad (26)$$

which using the notation $l_n(t,k)$ can be rewritten as

$$T_{ML} = \inf\left\{t : \max_{k<t} \sum_{n=1}^{N} \log\left[\hat{z}_n^{ML}(k) \exp(l_n(t,k)) + (1-\hat{z}_n^{ML}(k))\right] \geq b\right\}. \quad (27)$$

The maximum likelihood stopping rule thus computes the estimates $\hat{z}_n^{ML}(k)$ by thresholding the local CUSUM statistics to determine whether or not they should be included in the overall stopping rule. Conceptually, this process should help eliminate the inclusion of local statistics that do not exhibit a change and that could delay detection.

Remark. The summation defining the ML stopping rule in (26) can be rewritten as

$$\sum_{n=1}^{N} \log\left[\hat{z}_n^{ML}(k)\exp(l_n(t,k)) + (1-\hat{z}_n^{ML}(k))\right] = \sum_{n=1}^{N} \hat{z}_n^{ML}(k) l_n(t,k) = \sum_{n=1}^{N} l_n^+(t,k). \quad (28)$$

By substituting (28) into (27) and comparing the resulting stopping rule to (6), one sees that T_{ML} is equivalent to $T_3(p_0)$ when $p_0 = 1$, or in other words, when one assumes that all the sensors exhibit a change.

Remark. It is also interesting to note that T_{ML} is equivalent to T_{MAP} when $p_0 = 0.5$, or in other words, when one chooses a "non-informative prior" [27] for z_n. This is seen by comparing (16) and (25) and observing that $\hat{z}_n^{MAP} = \hat{z}_n^{ML}$ when $p_0 = 0.5$. The equivalence of the stopping rules is seen simply by substituting $\hat{z}_n^{ML}(k)$ into the MAP stopping rule with $p_0 = 0.5$.

$$\begin{aligned} T_{MAP} &= \inf\Big\{t : \max_{k<t} \sum_{n=1}^{N} \hat{z}_n^{ML}(k)\log(0.5) + (1-\hat{z}_n^{ML}(k))\log(0.5) + \log\big[\hat{z}_n^{ML}(k) \times \prod_{i=k}^{t} \frac{f_1(y_i^n)}{f_0(y_i^n)} + (1-\hat{z}_n^{ML}(k))\ \big] \geq b\Big\} \\ &= \inf\Big\{t : \max_{k<t} \sum_{n=1}^{N} \log(0.5) + \log\big[\hat{z}_n^{ML}(k) \times \prod_{i=k}^{t} \frac{f_1(y_i^n)}{f_0(y_i^n)} + (1-\hat{z}_n^{ML}(k))\ \big] \geq b\Big\} \\ &= \inf\Big\{t : \max_{k<t} \sum_{n=1}^{N} \log\big[\hat{z}_n^{ML}(k) \times \prod_{i=k}^{t} \frac{f_1(y_i^n)}{f_0(y_i^n)} + (1-\hat{z}_n^{ML}(k))\ \big] \geq b'\Big\} \end{aligned} \quad (29)$$

with $b' = b - N\log(0.5)$.

Remark. A connection can be also be made to the so-called *maximum test* or *scan statistic* [13, 24]. The stopping rule based on the scan statistic is defined as

$$T_{Scan} = \inf\left\{t : \max_{k<t}\max_{\mathcal{S}\in\Omega} \sum_{n\in\mathcal{S}} l_n(t,k) \geq b\right\} \quad (30)$$

where Ω is the power set [28] of the set $\{1,..,N\}$. By using

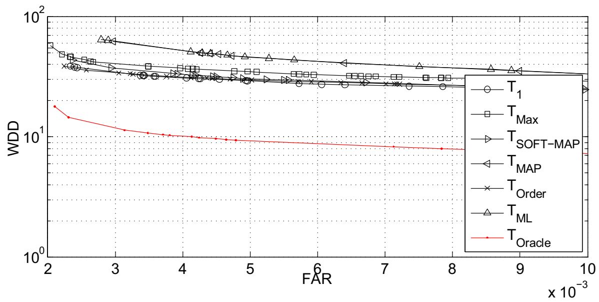
$$\max_{\mathcal{S}\in\Omega} \sum_{n\in\mathcal{S}} l_n(t,k) = \sum_{n=1}^{N} l_n^+(t,k) \quad (31)$$

we see that T_{ML} is equivalent to T_{Scan}.

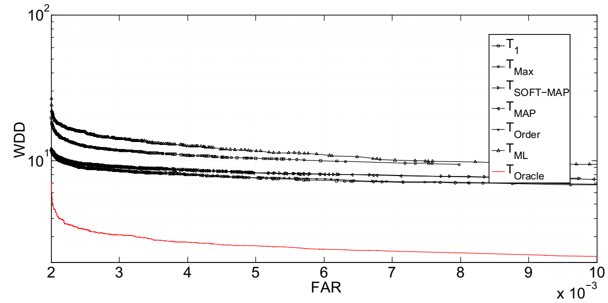
C. Stopping rule when the cardinality $\mathcal{S}$ is known

In this section, we introduce a new stopping rule which we call T_{Order}. This new rule is based on the scan statistic when the cardinality of the affected sensors $|\mathcal{S}| = M$ is known. It can be easily shown that

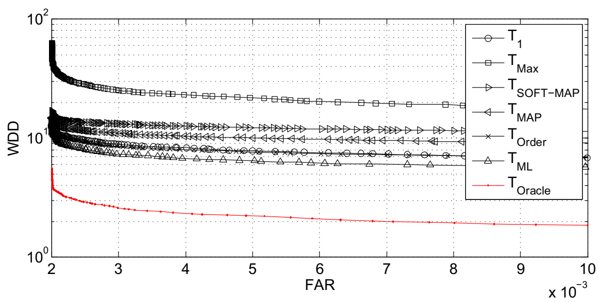
$$\max_{\mathcal{S}\in\Omega:|\mathcal{S}|=M} \sum_{n\in\mathcal{S}} l_n(t,k) = \sum_{n=1}^{M} l_{[n]}(t,k). \quad (32)$$



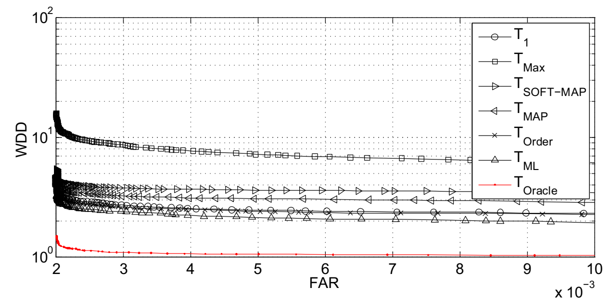
(a) $\mu = 0.5$, $\gamma = 0.96$, $N = 100$, $M = 4$ and $p_0 = 0.04$.



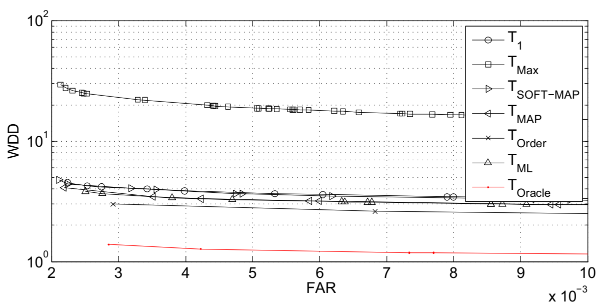
(b) $\mu = 1.0$, $\gamma = 0.96$, $N = 100$, $M = 4$ and $p_0 = 0.04$.



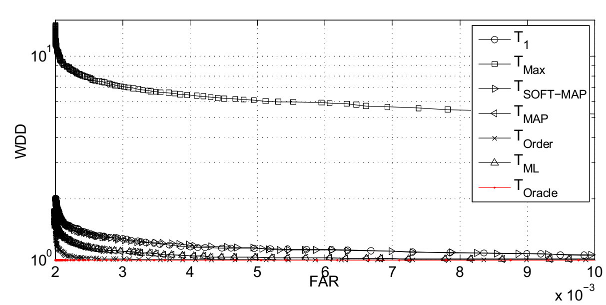
(c) $\mu = 0.5$, $\gamma = 0.80$, $N = 100$, $M = 20$ and $p_0 = 0.20$.



(d) $\mu = 1.0$, $\gamma = 0.80$, $N = 100$, $M = 20$ and $p_0 = 0.20$.



(e) $\mu = 0.5$, $\gamma = 0.5$, $N = 100$, $M = 50$ and $p_0 = 0.5$.



(f) $\mu = 1.0$, $\gamma = 0.5$, $N = 100$, $M = 50$ and $p_0 = 0.5$.

Fig. 1. Numerical comparison of different stopping rules for different values of μ and sparsity index γ. All results were created with 300 Monte Carlo runs.

where $l_{[n]}(t,k)$ denotes the n^{th} order statistic of $l_n(t,k)$, or n^{th} largest value of the sample $(l_1(t,k),...,l_N(t,k))$. Using this expression, we define T_{Order} as

$$T_{Order} = \inf\left\{t : \max_{k<t} \sum_{n=1}^{M} l_{[n]}(t,k) \geq b\right\}. \tag{33}$$

The word "order" in the rule's name reflects its connection with order statistics.

The T_{Order} rule resembles the detection scheme proposed by Mei in [18]. However these two approaches strongly differ in the sense that Mei was focusing on the setting in which the changes occur at different times at different sensors, whereas we assume that changes happen simultaneously across all affected sensors. Thus Mei's rule adds the top $|\mathcal{S}|$ local CUSUM statistics, while the T_{Order} is the CUSUM of the top $|\mathcal{S}|$ local log-likelihood ratios $l_n(t,k)$. Rewriting T_{Max} as

$$T_{Max} = \inf\left\{t : \max_{k<t} l_{[1]}(t,k,\mu) \geq b\right\}. \tag{34}$$

we see that in contrast to T_{Order}, T_{Max} neglects useful information from other affected sensors. On the other hand, the stopping rule T_{Sum} given in (7) incorporates noisy information also from the sensors which are not affected by the change. The stopping rule T_{Order} makes the best out of those opposite rules by incorporating the additional knowledge on the cardinality of the affected sensors. In Appendix A, an upper bound for the FAR for T_{Order} is provided.

Before presenting simulation results, we introduce an oracle stopping rule T_{Oracle} that will be used as a benchmark for the performance evaluation of the other stopping rules. It assumes full knowledge of the set of affected sensors.

$$T_{Oracle} = \inf\left\{t : \max_{k<t} \sum_{n\in\mathcal{S}} l_n(t,k,\mu) \geq b\right\} \qquad (35)$$

V. Numerical Comparison

In this section we compared the performance of the different stopping rules we introduced, in terms of detection delay and false alarm curve where WDD(T) is plotted against FAR(T) for a given stopping rule T. We consider the problem formalized in (1) with a change in mean for Gaussian observations with unit variance. Mathematically we have:

$$f_0(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}, f_1(y) = \frac{1}{\sqrt{2\pi}} e^{-(y-\mu)^2/2}. \qquad (36)$$

Figure 1 reports the numerical comparison between the different stopping rules introduced in this paper. We should consider that in the current simulation set-up, the unknown subset of affected sensors $\mathcal{S}$ does not change within each Montecarlo run. Therefore the numerical result cannot be interpreted as an indicator of average performance for different values of the sparsity index γ. Not surprisingly, different conclusions can be drawn from our numerical analysis, depending upon how sparse the problem is. When few sensors see the change, regardless of how big the size of the change μ is, the Bayesian stopping rules which are T_1, T_{MAP} and $T_{SOFT-MAP}$ perform better than the T_{ML} (please refer to Figures 1(a) and 1(b)). As the number of affected sensors increases (see Figures 1(c) and 1(d)), the T_{ML} seems to be superior to all the other stopping rules. Figures 1(e) and 1(f) illustrate that in low sparsity regime (half of the sensors experience the change), the T_{Order} is the best. We can also observe that when $p_0 = 0.5$, T_{ML} and T_{MAP} are equivalent as it has been shown in section IV-B. A close numerical inspection of the performance of T_{ML} and its counterpart T_{Order}, is given in Figures 2 and 3.

In Figure 2, we report the WDD for the T_{ML} and T_{Order} as function of sparsity index γ, for $N = 100$, $\mu = 0.5$ and FAR $= 0.1$. This rather small value of μ makes the detection problem hard to solve; we want to compare these two rules systems just in this challenging scenario. As the T_{Order} utilizes more information (prior knowledge on the cardinality S) than T_{ML}, one would expect T_{Order} being superior to T_{ML} in terms of WDD uniformly for all the sparsity levels. Surprisingly, this is not the case. For low-medium values of the sparsity index, the T_{Order} is better, until the performance intersection point, where the T_{ML} starts to outperform the T_{Order} in the medium-high sparsity region. In Figure 3 we display the same simulation scenario but for a different value of FAR $= 0.005$. In this case two intersection points divide the performance plane in three regions: low-medium, medium-high and very high sparsity regimes. At this moment we don't have an understanding of this very interesting phenomenon.

We can just comment that T_{Order} and T_{ML} are not known to be optimal tests. In the space of suboptimal algorithms, the results are not as predictable as in the space of optimal tests: It is not always possible to find intuitive explanations of the results.

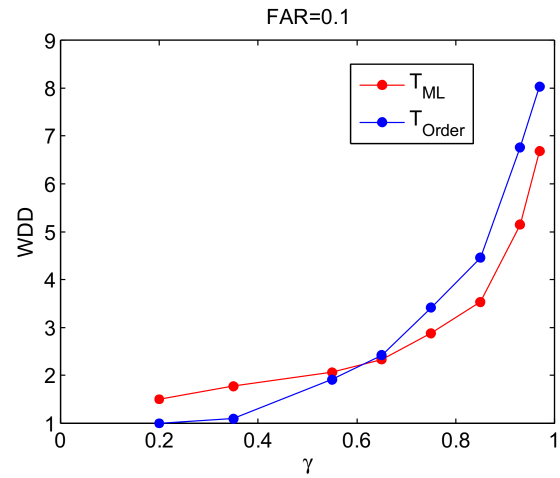


Fig. 2. WDD as function of γ. Here we used 300 Monte Carlo runs, $N = 100$, $\mu = 0.5$ and FAR $= 0.1$.

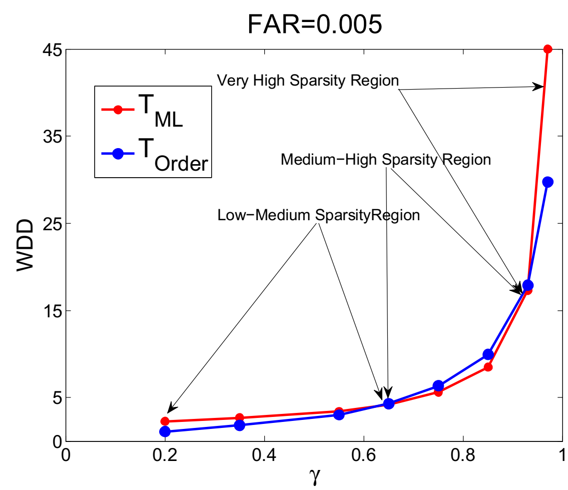


Fig. 3. WDD as function of γ. Here we used 300 Monte Carlo runs, $N = 100$, $\mu = 0.5$ and FAR $= 0.005$.

VI. Conclusions

Change detection with multiple sensors is an active research area. In this paper we addressed the change detection problem with an unknown subset of the affected sensors. We introduced different stopping rules and related them in a single framework. One particular and interesting finding is that the additional information about the cardinality of the affected sensors does not always lead to better detection performance.

Further investigations might include the analytical characterization of the performance of the proposed stopping rules. This analysis might also help understand (at least from a mathematical point of view) the counter-intuitive phenomenon we observed about the performance of T_{Order} and T_{ML}. It would be of interest, as a natural next challenge, to consider spatially arranged sensors where the subset $\mathcal{S}$ of affected sensors is likely to be co-located. To model the spatial dependency across the affected sensors, one could assign a different prior to the vector $\mathbf{z}$ as opposed to the prior in eq. (11). Incorporating such spatial dependency into the test statistic will be the key for the design of new stopping rules.

APPENDIX A
T_{Order}: AN UPPER BOUND FOR THE FAR

Proposition A.1.

$$FAR(T_{Order}) \leq \binom{N}{M} e^{-b}$$

Proof of Proposition A.1. The proof follows the technique used in Theorem 1 in [9]. For any non empty set $\mathcal{S} \in \Omega$,

$$R^{\mathcal{S}}(t) = \sum_{k=1}^{t} \exp\{\sum_{n \in \mathcal{S}} l_n(t,k)\}$$

The average of $R^{\mathcal{S}}(t)$ over all set of cardinality M is denoted by:

$$R(t) = \binom{N}{M}^{-1} \sum_{\mathcal{S} \in \Omega : |\mathcal{S}| = M} R^{\mathcal{S}}(t)$$

Define the auxiliary stopping time

$$\widetilde{T}(b) = \inf\{t : R(t) \geq e^b\} \quad (37)$$

It is easy to show that $R(t)$ provides an upper bound on the test statistic used in T_{Order} after a normalization:

$$R(t) \geq \binom{N}{M}^{-1} \max_{k<t} \exp\{\sum_{n=1}^{M} l_{[n]}(t,k)\}$$

Consequently,

$$T_{Order}(b) \geq \widetilde{T}(b - \log \binom{N}{M}) \quad (38)$$

Furthermore, by following the same steps as in [9], it can be verified that under H_∞, $R^S(t) - t$ is a martingale with mean zero. It then follows from the optional stopping theorem [7] that

$$\mathrm{E}_\infty[R(\widetilde{T})] = \mathrm{E}_\infty[\widetilde{T}]$$

Consequently

$$\mathrm{E}_\infty(\widetilde{T}) = \mathrm{E}_\infty R(\widetilde{T}) \geq e^b,$$

where the inequality follows from the definition of $\widetilde{T}$ in (37).

Combining this with (38) leads to the desired bound

$$\mathrm{E}_\infty[T_{Order}(b)] \geq \mathrm{E}_\infty[\widetilde{T}(b - \log \binom{N}{M})] \geq \binom{N}{M}^{-1} e^b.$$

□

REFERENCES

[1] W. A. Shewhart, "Economic control of quality of manufactured product," *American Society for Quality Control*, 1931.

[2] E. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.

[3] A. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability & Its Applications*, vol. 8, pp. 22–46, 1963.

[4] G. Lorden, "Procedures for reacting to a change in distribution," *The Annals of Mathematical Statistics*, vol. 42, pp. 1897–1908, 1971.

[5] M. Pollak, "Optimal detection of a change in distribution," *The Annals of Statistics*, vol. 13, pp. 206–227, 1985.

[6] H. V. Poor, , and O. Hadjiliadis, *Quickest Detection*. Cambridge University Press, 2009.

[7] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[8] P. Evans and M. Annunziata, "Industrial internet: Pushing the boundaries of minds and machines," GE General Electric Company, Tech. Rep., 2012. [Online]. Available: http://www.ge.com/docs/chapters/IndustrialInternet.pdf/

[9] A. G. Tartakovsky and V. Veeravalli, "An efficient sequential procedure for detecting changes in multichannel and distributed systems," in *Proceedings of IEEE International Conference on Information Fusion*, Jul. 2002, pp. 41–48.

[10] A. G. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Change-Point Detection*. Statistics, CRC Press, 2014.

[11] Y. Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, p. 419433, 2010.

[12] A. G. Tartakovsky and V. Veeravalli, "Asymptotically optimal quickest change detection in distributed sensor systems," *Sequential Analysis*, vol. 27, p. 441475, 2008.

[13] L. Addario-Berry, N. Broutin, Devroye, and L. G. Lugosi, "On combinatorial testing problems," *The Annals of Statistics*, vol. 38, p. 30633092, 2010.

[14] A. Jung, Z. Ben-Haim, F. Hlawatsch, and Y. Eldar, "Unbiased estimation of a sparse vector in white gaussian noise," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7856–7876, Dec. 2011.

[15] R. M. Castro, "Adaptive sensing performance lower bounds for sparse signal detection and support estimation," *Bernoulli*, vol. 20, no. 4, pp. 2217–2246, 2014.

[16] J. Haupt, R. M. Castro, and R. Nowak, "Distilled sensing: Adaptive sensing for sparse detection and estimation," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6222–6235, Sep 2011.

[17] E. Tanczos and R. Castro, "Adaptive sensing for estimation of structured sparse signals," *to appear in IEEE Transactions on on Information Theory*, 2015. [Online]. Available: http://arxiv.org/pdf/1311.7118v1.pdf

[18] Y. Mei, "Quickest detection in censoring sensor networks," in *International Symposium on Information Theory (ISIT)*, Aug. 2011, pp. 2148–2152.

[19] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," *The Annals of Statistics*, vol. 41, no. 2, pp. 670–692, 2013.

[20] S. Kumar, B. Sai Kiran, A. Pachai, and S. Bhashyam, "Algorithms for change detection with unknown number of affected sensors," in *Communications (NCC), 2013 National Conference on*, Feb 2013.

[21] C. Li, H. Dai, and H. Li, "Adaptive quickest change detection with unknown parameter," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2009.

[22] I. Nikiforov, "Suboptimal quadratic change detection scheme," *IEEE Transactions on Information Theory*, vol. 46, pp. 2095–2107, Sep 2000.

[23] T. Banerjee and V. Veeravalli, "Data-efficient quickest outlying sequence detection in sensor networks," *(submitted) to IEEE Transactions on Signal Processing*, Nov 2014. [Online]. Available: http://arxiv.org/pdf/1411.0183v2.pdf

[24] J. Glaz, J. Naus, and S. Wallenstein, *Scan Statistics*. Springer, 2001.

[25] P. Billingsley, *Convergence of Probability Measures*. 2nd ed., Wiley, 1999.

[26] B. Ghosh and P. Sen, *Handbook of Sequential Analysis*. Hardcover, 1991.

[27] C. Robert, *The Bayesian Choice*. New York: Springer-Verlag, 2001.

[28] K. J. Devlin, *Fundamentals of contemporary set theory*. Springer-Verlag, 1979.