Sequential Multi-Sensor Change-Point Detection

Yao Xie,* David O. Siegmund†‡

September 22, 2011

Abstract

We develop a mixture procedure for sequential multi-sensor change-point detection problems without assuming a spatial structure. We assume that the change-point amplitudes observed by the sensors are non-homogeneous, and that the subset of affected sensors as well as their post-change means are unknown. The mixture procedure incorporates some prior information - the fraction of affected sensors p when forming the detection statistics. In particular, the mixture procedure obtains a local generalized likelihood ratio (GLR) statistic from each sensor, then uses soft-thresholding on these local GLR to suppress noise from the unaffected sensors, sums the outputs, and then compares the result with a detection threshold. We derive a closed-form average run length (ARL) as well as the expected detection delay for the mixture procedure. These closed-form approximations are numerically verified to have high accuracy. We also show that the mixture procedure does not require an accurate estimate of p. Finally, we demonstrate using numerical examples that for a given ARL, compared with other procedures for unstructured problems, the mixture procedure achieves a smaller expected detection delay; compared with procedures for structured problems, when there are multiple change-points, the mixture procedure achieves a similar expected detection delay with lower complexity but using less prior information.

1 Introduction

In multi-sensor change-point detection problems, multiple sensors monitor the abrupt emergence of a signal, which is also known as the change-point detection problem. It is convenient to imagine

*Yao Xie (Email: yaoxie@stanford.edu) is with the Department of Electrical Engineering at Stanford University.

†David O. Siegmund (Email: siegmund@stanford.edu) is with the Department of Statistics at Stanford University.

‡This work is partially supported by Stanford General Yao-Wu Wang Graduate Fellowship.

that the sensors are spatially distributed and can "see" signals that arise in their proximity, but not those arising far away. Suppose we have N sensors monitoring a random field. At a certain time κ, the distributions of the random field change, and consequently there are changes in the distributions of observations made at M of the sensors in a subset $\mathcal{N}_a$, where $1 \leq M \leq N$. The change-point κ, the subset and the size of the subset of sensors affected by the change-point, and the size of the changes seen at these sensors are all unknown. The goal is to detect the change-point as soon as possible after it occurs, while keeping the frequency of false alarms as low as possible. In the change-point detection literature, the frequency of false alarms is usually replaced by a related performance metric, the average-run-length (ARL), which is the expected time before incorrectly announcing a change of distribution when none has occurred.

In this detection problem it may or may not be reasonable to assume a known profile which determines the relative change-point amplitudes observed by different sensors according to their distance from the change-point location. This profile is modeled as the spatial structure of the random field. Based on if we can make an assumption about the spatial structure, there are two types of problems. The unstructured problems, which have no spatial models relating the signal to the observations at the various sensors. For example, in network intrusion detection [LLR09], the attack affects distribution of a group of computers with different IP addresses. However, these IP addresses of the affected computers may have no relationship in terms of their geographic vicinity. The non-structured problem have been discussed by [TV08], [Mei10], [CGV+10], [PRT03], and [LLR09], with variations depending on the envisioned applications.

The structured problems has a spatial structure that relates the signal to observations at various sensors. The spatial structure is determined by the known sensor locations throughout the region of interest, as well as the location of the emerging signal centered inside the region and decaying according to a known function of the distance from this center to each sensor. Examples of structured problems with fixed number of samples include epidemiological monitoring for disease outbreaks [Rab94], fMRI scanning [SSSW03], and on-line monitoring of the emergence of a signal with known profile [SY08] in a sequential setting.

A multi-sensor detection problem of particular interest in this paper involves a case that N is large and the number of affected sensors M is relatively small. Moreover, the number and the subset of affected sensors are unknown. This problem arises in many applications such as sensor network surveillance, where the number of sensors is large and not all of them observe a change-point. The affected sensors and their observation amplitudes are unknown. These unknown parameters poses a challenge when we look for an efficient detection procedure that has low expected detection delay and

a large ARL. To achieve such a goal, the detection procedure should use only the useful informations from the affected sensors, ignores noise from the unaffected sensors, as well as considers the unknown and non-homogeneous change-point amplitudes at different sensors. Hence, the detection procedure has to identify which sensors are affected. However, searching for unknown affected sensors across a large number of sensors is exponentially complex.

For this problem, Mei proposed a procedure [Mei10] that sums the CUSUM statistic from individual sensors and compares the sum with a suitable threshold, assuming the post-change distributions are completely prescribed. This method asymptotically minimizes the expected detection delay for a desired false alarm rate, when the threshold value becomes infinitely large. However, In practice, knowledge about the post-change distributions is usually not available. When the true distributions deviate from the assumed ones, the procedure in [Mei10] is no longer asymptotically optimal. Tartakovsky and Veeravalli proposed a different procedure (referred to as the TV procedure) [TV08] that sums the local likelihood ratio statistic before forming a CUSUM statistics. They also assume the post-change distributions are completely prescribed. Moreover, Mei's procedure and the TV procedure both assume the change-point is observed by all sensors. When only a subset of sensors observe the change-point, the Mei's and the TV procedures include noise from the unaffected sensors into the detection statistic and may incur long detection delays.

In this paper, we develop a mixture procedure that achieves good detection performance in the case of unknown subset of affected sensors and incompletely specified post-change distributions. The key feature of the mixture procedure is that it incorporates prior information about the fraction of affected sensors when computing the detection statistic. We make the frequently used assumptions that the individual data observations are independent and normally distributed with unit variance, and that the changes occur in their mean values. The mixture procedure first computes a generalized likelihood ratio (GLR) statistic for individual sensor under the assumption that the change-point occurs at $k \leq t$. Then the local GLR statistics are soft-thresholded using a thresholding function that depends on p_0, as a guess for the fraction of affected sensors p (which equals M/N), to suppress noise from the non-affected sensors. The soft-thresholded local GLRs are then summed and compared with a proper detection threshold. To characterize the performance of the mixture procedure, we derive closed-form expressions for its ARL and detection delay. The closed-form ARL provides a convenient way to determine a threshold b for a desired false alarm rate. We demonstrate numerically that the mixture procedure is not sensitive to the choice of p. We also compare the mixture procedure with other multi-sensor detection procedures numerically. The results show that for unstructured problems the mixture procedure has smaller expected detection delay. For the

structured problems, the mixture procedure can achieve performance similar to the profile-based procedure. Although the mixture procedure assumes less no prior information about the signal spatial structure whereas the profile-based method makes such assumptions. This indicates that the one parameter p used in the mixture procedure efficiently captures the useful side information for multi-sensor change-point detection. Although we assume throughout that the observations are normally distributed, the model can be generalized to an exponential family of distributions satisfying some additional regularity conditions.

The reminder of the paper is organized as follows. In Section 2 we establish our notations and present the system model and problem formulation. In Section 3 we review several detections procedures and introduce the new mixture procedure. In Section 4 we derive the theoretical ARL and expected detection delay of the mixture procedure, and demonstrate with numerical examples that these approximations are reasonably accurate. In Section 5, we demonstrate that the mixture procedure has good performance compared with other procedures in the unstructured setting and that it has comparable performance to the profile-based method when there are multiple change-points in the spatial structured setting. Finally Section 7 concludes the paper with some discussions.

2 Model and Formulation

Given N sensors, the observations from the nth sensor are given by y_t^n, $t = 1, 2 \cdots$. Assume that different data streams are mutually independent. Under the null hypothesis, the observations y_t^n are identically distributed with zero means and unit variances. The probability and expectation of this case are denoted by $\mathbb{P}^\infty$ and $\mathbb{E}^\infty$, respectively. Alternatively, there exists a change-point $\kappa \geq 0$ and a subset $\mathcal{N}_a$ of $\{1, 2, \ldots, N\}$ with cardinality M such that for $n \in \mathcal{N}_a$, the observations of affected sensors y_t^n, $n \in \mathcal{N}_a$, have means equal to $\mu_n \neq 0$ for all $t > \kappa$ and unit variances, while observations from the unaffected sensors y_t^n, $n \in \mathcal{N}_a^c$ remain the same distribution. The probability and expectation of this case are denoted by $\mathbb{P}^\kappa$ and $\mathbb{E}^\kappa$, respectively. Note that this probability depends on $\mathcal{N}_a$ and the values of μ_n, although this dependence is suppressed in the notation. Hence the fraction of affected sensors is given by $p = M/N$.

Our goal is to define a stopping rule T such that for a prescribed large constant $c > 0$, asymptotically as $c \to \infty$, $\mathbb{E}^\infty(T) \geq c$, while asymptotically $\mathbb{E}^\kappa(T - \kappa | T > \kappa)$ is a minimum. Ideally, the minimization would hold uniformly in the various unknown parameters: κ, $\mathcal{N}_a$ and the μ_n. Since this is clearly impossible, in Section 5 we will compare different procedures through numerical examples computed under various hypothetical conditions.

Since the observations are independent, for an assumed value of the change-point $\kappa = k$ and sensor $n \in \mathcal{N}_a$, the log-likelihood of observations accumulated by time $t > k$ is given by

$$\ell_n(t, k, \mu_n) = \sum_{i=k+1}^{t} \left(\mu_n y_i^n - \mu_n^2/2\right). \tag{1}$$

When μ_n are unknown, we use their maximum likelihood estimates as their substitutes, and form the generalized likelihood ratio (GLR) statistics. To find the maximum likelihood estimates, we first take derivative of the log-likelihood ratio $\ell_n(t, k, \mu_n)$ with respect to μ_n, set it to zero, and solve for the solution. The maximum likelihood estimate of the post-change mean μ_n as a function of the current number of observations t and the putative change-point location k is given by

$$\hat{\mu}_n = \left(\sum_{i=k+1}^{t} y_i^n\right)/(t-k). \tag{2}$$

Substitution gives the log GLR statistic

$$\ell_n(t, k, \hat{\mu}_n) = \left(\sum_{i=k+1}^{t} y_i^n\right)^2/[2(t-k)]. \tag{3}$$

It will be convenient to put:

$$\begin{aligned} S_t^n &= \sum_{i=1}^{t} y_i^n, \\ U_{k,t}^n &= (t-k)^{-1/2}(S_t^n - S_k^n). \end{aligned} \tag{4}$$

Then we can write the log GLR as

$$\ell_n(t, k, \hat{\mu}_n) = (U_{k,t}^n)^2/2. \tag{5}$$

If we require the post-change means to be positive, $\mu_n > 0$, the maximum-likelihood estimate of the post-change mean is given by $\hat{\mu}_n = (t-k)^{-1/2}(U_{k,t}^n)^+$, where $(x)^+ = x$ if $x \geq 0$, and is equal to zero when $x < 0$. The corresponding log GLR statistic is given by

$$\ell_n(t, k, \hat{\mu}_n) = [(U_{k,t}^n)^+]^2/2. \tag{6}$$

In the following, we will focus on the case where μ_n can be either positive or negative. The detection

procedure for this case is called the *two-sided* procedure. The detection procedure derived assuming μ_n is positive is called the *one-sided* procedure. The one-sided procedure can be derived from the two-sided procedure, by using expression (6) rather than (5) for the log-GLR statistic in the detection procedures.

3 Detection Procedures for Non-Structured Problems

Assume the fraction of affected sensors to be p_0, which may differ from the true fraction value p. We consider the following mixture procedure that uses a soft thresholding function with parameter p_0,

$$g(x; p_0) = \log[1 - p_0 + p_0 \exp(x^2/2)]. \tag{7}$$

The mixture procedure applies this function of (7) on the local log GLR (5). The function $g(x; p_0)$ impose a soft threshold on the input argument. The threshold level is controlled by the value p_0. The smaller the p_0, and higher the level of thresholding the function applies on the local GLR statistic. When $p_0 = 1$, $g(x; p_0) = x^2/2$, in which case the function does no soft thresholding on the local GLR since we believe all sensors observe the change-point.

Then the procedure sums the results to form the detection statistic, which is given by

$$\sum_{n=1}^{N} \log\left(1 - p_0 + p_0 \exp[(U_{k,t}^n)^2/2]\right). \tag{8}$$

The associated stopping rule for the mixture procedure T_{mix} is defined as

$$T_{\text{mix}} = \inf\left\{t : \max_{0 \le k < t} \sum_{n=1}^{N} \log\left(1 - p_0 + p_0 \exp[(U_{k,t}^n)^2/2]\right) \ge b\right\}. \tag{9}$$

This soft-thresholding function has been used in the previous work [ZYS10] and the motivation for choosing such a function can be found therein. In particular, applying the threshold function (7) attenuates noise while preserves the change-point information. To see this, when the nth sensor observes the change-point which occurs at k_0, then for $n \in \mathcal{N}_a$, the observations y_n^t, $t > k_0$, have positive means μ_n. From (4), the process $U_{k,t}^n$ also has positive drift for $t > k \ge k_0$. Thus as t increases, $\exp[(U_{k,t}^n)^2/2]$ in (8) grows fast and dominates the other term, which leads to $p_0 \exp[(U_{k,t}^n)^2/2] \gg 1 - p_0$. Then applying $g(x; p_0)$ to the local GLR of the affected sensor becomes $g(U_{k,t}^n; p_0) \approx -\log p_0 + (U_{k,t}^n)^2/2$, for $n \in \mathcal{N}_a$. On the other hand, for the unaffected sensors $n \in \mathcal{N}_a^c$, $\exp[(U_{k,t}^n)^2/2]$ is a bounded random variable since it is due to noise. Hence $g(U_{k,t}^n; p_0)$ is also a

bounded random variable for $n \in \mathcal{N}_a^c$ and the effect the noise term is attenuated through scaling the term by p_0 and taking a log.

For the special case $p_0 = 1$, (9) becomes the GLR procedure studied in [SV95]:

$$T_{\text{GLR}} = \inf \left\{ t : \max_{0 \le k < t} \sum_{n=1}^{N} (U_{k,t}^n)^2/2 \ge b \right\}, \tag{10}$$

which is efficient when the change-point affects a large fraction of the sensors. At the other extreme, if the set $\mathcal{N}_a$ is very small, a reasonable procedure would be

$$T_{\max} = \inf \left\{ t : \max_{0 \le k < t} \max_{1 \le n \le N} (U_{k,t}^n)^2/2 \ge b \right\}. \tag{11}$$

The stopping rule for Mei's procedure [Mei10] for change-point detection is given by

$$T_{\text{Mei}} = \inf \left\{ t : \sum_{n=1}^{N} \max_{0 \le k < t} [\tilde{\mu}_n (S_t^n - S_k^n) - \tilde{\mu}_n^2 (t-k)/2] \ge b \right\}, \tag{12}$$

where $\{\tilde{\mu}_n\}$ is a set of assumed post-change means seen by the sensors. The $\tilde{\mu}_n$ essentially establishes a minimum size of change of interest. Mei's procedure is a one-sided procedure, since it assumes a specific value for each μ_n. When the assumed μ_n is positive, but the true μ_n is negative, Mei's procedure will have a model mismatch.

We also consider a modification of the TV procedure studied in [TV08], with the stopping rule given by

$$T_{\text{TV}} \triangleq \inf \left\{ t : \max_{0 \le k < t} \sum_{n=1}^{N} [\tilde{\mu} (S_t^n - S_k^n) - \tilde{\mu}_n^2 (t-k)/2] \ge b \right\}. \tag{13}$$

Note that for this procedure, in the presence of a change-point, the statistics of the unaffected sensors have negative drifts that tend to decrease the positive drifts from the affected sensors, which causes a longer expected detection delay. We suggest the following modification to the TV procedure to eliminate such negative drifts

$$T_{\text{TV,modified}} \triangleq \inf \left\{ t : \max_{0 \le k < t} \sum_{n=1}^{N} [\tilde{\mu} (S_t^n - S_k^n) - \tilde{\mu}_n^2 (t-k)/2]^+ \ge b \right\}. \tag{14}$$

Comparing (12) and (14), we note that the detection statistic of Mei's procedure is greater than that of the modified TV procedure, since interchanging the order of maximizing with sum increases the value. The Modified TV procedure is also a one-sided procedure.

Both the Mei's and the TV procedures can be modified to be two-sided procedures. This can be done by assuming that the post-change mean can be μ_n or $-\mu_n$ equally likely.

The threshold b of each of the above procedures (9) - (14) are chosen to meet the ARL requirement.

Usually very small changes are not of interest, so we can modify the definitions of (9), (10), (11) and (14) by maximizing over $t - T_1 < k < t$, where T_1 is a suitable maximum window size. In the modified version we only examine possible change-points in the most recent T_1 observations. This also has the effect of simplifying the computations required to implement the algorithm. It is possible to require a minimum window size as well, say T_0, if we are concerned about possible outliers in the data. Or the individual "observations" may actually be means or medians of grouped data, to make the assumption of normality more reasonable. There are convenient recursive computational algorithms for each of the terms in the sum of (12), so window limiting plays no role for the Mei's procedure.

4 Performance of the Mixture Procedure

In this section we study the theoretical performance of the mixture procedure using two performance metrics: the average run length (ARL) when there is no change, and the expected detection delay in the extreme case where a change occurs immediately $k = 1$, which provides an upper bound on the expected detection delay when a change occurs later in the sequence $k > 1$.

4.1 Average Run Length When There Is No Change

The average-run-length (ARL) is the average length of intervals between two false-alarms when there is no change-point. It is a common performance metric for false-alarm rate of a sequential detection procedure. To characterize the ARL of the mixture procedure, we need the following quantities:

$$\psi(\theta) = \log \mathbb{E}[\exp\{\theta g(U; p_0)\}], \tag{15}$$

where U has a standard normal distribution;

$$\gamma(\theta) = \frac{1}{2}\theta^2 \mathbb{E}\left\{[\dot{g}(U; p_0)]^2 \exp[\theta g(U; p_0) - \psi(\theta)]\right\}, \tag{16}$$

where the dot denotes differentiation; and

$$f(N,\theta,p_0)=\frac{\gamma(\theta)N^{1/2}}{\theta\{2\pi\ddot{\psi}(\theta)\}^{1/2}}\exp\{-N[\theta\dot{\psi}(\theta)-\psi(\theta)]\}.$$

Denote the standard normal density function as $\phi(x)$ and its distribution function as $\Phi(x)$, with $\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, and $\Phi(x)=\int_{-\infty}^{x}\phi(z)dz$.

Theorem 1. *Assume $N\to\infty$ and $b\to\infty$. For θ satisfies $\dot{\psi}(\theta)=b/N$, we have*

$$\mathbb{E}^{\infty}(T_{\text{mix}})\sim\left[f(N,\theta,p)\int_{[2N\gamma(\theta)/T_1]^{1/2}}^{[2N\gamma(\theta)]^{1/2}}y\nu^2(y)dy\right]^{-1}, \tag{17}$$

where $\nu(x)=2x^{-2}\exp\left[-2\sum_1^{\infty}n^{-1}\Phi\left(-|x|\sqrt{n}/2\right)\right]$ (p. 82 [Sie85]) and it can be numerically approximated in terms of $\phi(x)$ and $\Phi(x)$ by [SY07]:

$$\nu(x)\approx\frac{(2/x)[\Phi(x/2)-0.5]}{(x/2)\Phi(x/2)+\phi(x/2)}.$$

Here the notation $x\sim y$ means $x/y\to c$ asymptotically for some constant c.

Remark: The integrand in the approximation is integrable at both 0 and ∞ by virtue of the relations $\nu(y)\to 1$ as $y\to 0$, and $\nu(y)\sim 2/y^2$ as $y\to\infty$. Theorem 1 is for the two-sided procedure. The ARL of the corresponding one-sided procedure is half of the expression in (17). This is because the tail probability of a two-sided procedure is half of that of the one-sided procedure.

The following heuristic calculations provide support for the approximation in Theorem 1. In [ZYS10] it was shown that

$$\begin{aligned}&\mathbb{P}^{\infty}\left\{\max_{t\le m,T_0\le t-k\le T_1}\sum_{n=1}^{N}\log\left(1-p_0+p_0\exp[(U_{k,t}^n)^2/2]\right)\ge b\right\}\\&\sim N^2e^{-N[\theta\dot{\psi}(\theta)-\psi(\theta)]}[2\pi N\ddot{\psi}(\theta)]^{-1/2}|\theta|^{-1}\gamma^2(\theta)\int_{\frac{T_0}{m}}^{\frac{T_1}{m}}\nu^2\left([2N\gamma(\theta)/(mt)]^{1/2}\right)(1-t)dt/t^2,\end{aligned} \tag{18}$$

where N and b diverge to ∞ at the same rate, and m is also large, but small enough that the right hand side of (18) converges to 0. We are primarily interested in the case where $T_0=1$, and the maximum window size T_1 is small compared to m. Changing variables in the integrand, we can re-write this approximation as

$$\mathbb{P}^{\infty}(T_{\text{mix}}\le m)\sim mf(N,\theta,p_0)\int_{[2N\gamma(\theta)/T_1]^{1/2}}^{[2N\gamma(\theta)]^{1/2}}y\nu^2(y)dy. \tag{19}$$

It can be justified using the arguments in [Ald88] and [SY08] that T_{mix} is asymptotically exponentially distributed. Hence if λ denotes the factor multiplying m on the right hand side of (19), then for larger m, in the range where $m\lambda$ is bounded away from 0 and ∞, $\mathbb{P}^{\infty}(T_{\text{mix}} \leq m) \sim [1 - \exp(-\lambda m)] \to 0$. Hence, provided there is uniform integrability, $\mathbb{E}^{\infty}(T_{\text{mix}}) \sim \lambda^{-1}$, which leads to (17) by plugging in the expression for the corresponding λ in (19).

4.2 Expected Detection Delay

After a change-point occurs, we are interested in the expected number of additional observations required for detection. The maximum expectation detection delay over $k \geq 0$ happens when the change-point occurs at the first observation. Hence we consider this scenario.

We continue to use the notation of the preceding section. In particular $g(x; p_0) = \log[1 - p_0 + p_0 \exp(x^2/2)]$, and U denotes a standard normal random variable. Recall that $\mathcal{N}_a$ denotes the set of sensors at which there is a change, M is the cardinality of this set, and $p = M/N$ is the true fraction of sensors that are affected by the change. For each $n \in \mathcal{N}_a$ the mean value changes from 0 to $\mu_n \neq 0$, and for $n \in \mathcal{N}_a$ the distribution remains the same before and after the change-point occurs. Let

$$\Delta = \left(\sum_{n \in \mathcal{N}_a} \mu_n^2 \right)^{1/2}, \tag{20}$$

with Δ^2 interpreted as the total energy of the change-point observed by the sensors.

Note that the Kullback-Leibler divergence of a vector of observations after the change-point from a vector of observations before the change-point is $\Delta^2/2$, which determines the asymptotic rate of growth of the detection statistic after the change-point. Using Wald's identity [Sie85], we know to the first order approximation the detection delay is $2b/\Delta^2$, provided that the maximum window size, T_1, is large compared to this quantity. In the following derivation we assume $T_1 \gg 2b/\Delta^2$.

In addition, let

$$\tilde{S}_t \triangleq \sum_{i=1}^{t} z_i \tag{21}$$

be a random walk and the increments z_i are independent and identically distributed with mean $\Delta^2/2$ and variance Δ^2. For $c > 0$, let $\tau = \min\{t : \tilde{S}_t > 0\}$. Our approximation to the expected detection delay given below depends on the following two quantities. Define

$$\rho(\Delta) = \frac{\mathbb{E}\left(\tilde{S}_\tau^2\right)}{2\mathbb{E}\left(\tilde{S}_\tau\right)}, \tag{22}$$

for which exact computational expressions and useful approximations are available in [Sie85]. One such expression for $\rho(\Delta)$ is given by

$$\rho(\Delta) = \frac{\mathbb{E}(z_1^2)}{2\mathbb{E}(z_1)} - \sum_{i=1}^{\infty} i^{-1}\mathbb{E}(Z_i^-) = \frac{\Delta^2}{4} + 1 - \sum_{i=1}^{\infty} i^{-1}\mathbb{E}(Z_i^-), \tag{23}$$

with $Z_i \sim \mathcal{N}(i\Delta^2/2, i\Delta^2)$. Another quantity is $\mathbb{E}\left\{\min_{t\geq 0} \tilde{S}_t\right\}$, which is given by (Problem 8.14 in [Sie85]).

$$\mathbb{E}\left\{\min_{t\geq 0} \tilde{S}_t\right\} = \rho(\Delta) - 1 - \Delta^2/4. \tag{24}$$

The following theorem refines this first order result from Wald's identity [Sie85] for the expected detection delay:

Theorem 2. *As $b \to \infty$, with other parameters held fixed,*

$$\mathbb{E}^0\{T_{\text{mix}}\} = 2\Delta^{-2}\left[b + \rho(\Delta) - M\log p_0 - M/2 + \mathbb{E}\left\{\min_{t\geq 0} \tilde{S}_t\right\} - (N-K)\mathbb{E}\{g(U)\} + o(1)\right], \tag{25}$$

where U is a normal random variance with zero mean and unit variance.

Remark: The above theorem is for the two-sided mixture procedure. For the corresponding one-sided mixture, the expected detection delay is approximately the same since the positive drifts of the statistics from the affected sensors dominate the expected detection delay.

The following heuristic argument provides a support to the argument in Theorem 2. For details of similar problems involving a single sequence, see [PS75] and [SV95]. In the following, for convenience we write $T = T_{\text{mix}}$.

For affected data streams, if k is small compared to t, $p_0 \exp[(U_{k,t}^n)^2/2] \gg 1 - p_0$. So we can write the detection statistic at the stopping time T up to terms that tend to 0 exponentially fast in probability as

$$\begin{aligned}
\sum_{n=1}^{N} g(U_{k,T}^n; p_0) &= \sum_{n\in\mathcal{N}_a} g(U_{k,T}^n; p_0) + \sum_{n\in\mathcal{N}_a^c} g(U_{k,T}^n; p_0) \\
&\approx \sum_{n\in\mathcal{N}_a} \left[\log p_0 + (U_{k,T}^n)^2/2\right] + \sum_{n\in\mathcal{N}_a^c} g(U_{k,T}^n; p_0) \\
&= M\log p_0 + \sum_{n\in\mathcal{N}_a} (U_{k,T}^n)^2/2 + \sum_{n\in\mathcal{N}_a^c} g(U_{k,T}^n; p_0).
\end{aligned} \tag{26}$$

We will use the simple identity

$$S_t^2/2t = \mu(S_t - \mu t/2) + (S_t - \mu t)^2/2t. \tag{27}$$

Using identity (27) and (26), we have

$$\begin{aligned}
&\max_{0\le k<T} \sum_{n=1}^{N} g(U_{k,T}^n; p_0) \\
&= M \log p_0 + \max_{0\le k<T} \left[\sum_{n\in\mathcal{N}_a} \mu_n[(S_T^n - S_k^n) - (T-k)\mu_n/2] \right. \\
&\quad \left. + \sum_{n\in\mathcal{N}_a} [(S_T^n - S_k^n) - (T-k)\mu_n]^2/[2(T-k)] + \sum_{n\in\mathcal{N}_a^c} g(U_{k,T}^n; p_0) \right] \qquad (28) \\
&= M \log p_0 + \sum_{n\in\mathcal{N}_a} \mu_n \left(S_T^n - T\mu_n/2\right) + \\
&\quad \max_{0\le k<T} \left[-\sum_{n\in\mathcal{N}_a} \mu_n \left(S_k^n - k\mu_n/2\right) + \sum_{n\in\mathcal{N}_a} \frac{[(S_T^n - S_k^n) - (T-k)\mu_n]^2}{2(T-k)} + \sum_{n\in\mathcal{N}_a^c} g\left(U_{k,T}^n; p_0\right) \right].
\end{aligned}$$

The following lemma forms the basis of the rest of the derivation (see Appendix A for details).

Lemma 3. *For $k_0 = b^{1/2}$, asymptotically as $b \to \infty$*

$$\begin{aligned}
&\max_{0\le k<T} \left[-\sum_{n\in\mathcal{N}_a} \mu_n \left(S_k^n - k\mu_n/2\right) + \sum_{n\in\mathcal{N}_a} \frac{[(S_T^n - S_k^n) - (T-k)\mu_n]^2}{2(T-k)} + \sum_{n\in\mathcal{N}_a^c} g(U_{k,T}^n; p_0) \right] = \\
&\sum_{n\in\mathcal{N}_a} [S_T^n - T\mu_n]^2/2T + \sum_{n\in\mathcal{N}_a^c} g(U_{0,T}^n; p_0) + \max_{0\le k<k_0} \left[-\sum_{n\in\mathcal{N}_a} \mu_n (S_k^n - k\mu_n/2) \right] + o(1).
\end{aligned}$$

By taking expectations on (28), letting $b \to \infty$ and use Lemma 3, we have

$$\begin{aligned}
&\mathbb{E}^0 \left\{ \max_{0\le k<T} \sum_{n=1}^{N} g(U_{k,T}^n; p_0) \right\} \\
=&\mathbb{E}^0 \left\{ M \log p_0 + \sum_{n\in\mathcal{N}_a} \mu_n (S_T^n - T\mu_n/2) + \sum_{n\in\mathcal{N}_a} [S_T^n - T\mu_n]^2/2T + \sum_{n\in\mathcal{N}_a^c} g(U_{0,T}^n; p_0) + \right. \qquad (29) \\
&\left. \max_{0\le k<k_0} \left[-\sum_{n\in\mathcal{N}_a} \mu_n (S_k^n - k\mu_n/2) \right] \right\} + o(1).
\end{aligned}$$

We will compute each term on the right hand side of (29) separately. We will need the lemma due

to Anscombe and Doeblin (see Theorem 2.40 in [Sie85]), which states that the randomly stopped average of random variables are asymptotically normal distributed under quite general conditions.

(i) By Wald's identity [Sie85]:

$$\mathbb{E}^0\left\{\sum_{n\in\mathcal{N}_a}\mu_n(S_T^n - T\mu_n/2)\right\} = \mathbb{E}^0\{T\}\Delta^2/2. \tag{30}$$

(ii) By the Anscombe-Doeblin Lemma, $(S_T^n - T\mu_n)/T^{1/2}$ is asymptotically normally distributed. Hence $\sum_{n\in\mathcal{N}_a}[S_T^n - T\mu_n]^2/T$ is asymptotically a sum of independent χ_1^2 random variables, so

$$\mathbb{E}^0\left\{\sum_{n\in\mathcal{N}_a}[S_T^n - T\mu_n]^2/2T\right\} = M/2 + o(1). \tag{31}$$

(iii) For the same reason as (ii):

$$\mathbb{E}^0\left\{\sum_{n\in\mathcal{N}_a^c} g(U_{0,T}^n)\right\} \to (N-M)\mathbb{E}_0[g(U)]. \tag{32}$$

(iv) The term $-\sum_{n\in\mathcal{N}_a}\mu_n(S_k^n - \mu_n k/2)$ $(k \geq 0)$ is a random walk with negative drift $-\Delta^2/2$ and variance Δ^2. Hence $\mathbb{E}^0\left\{\max_{0\leq k<k_0} -\sum_{n\in\mathcal{N}_a}\mu_n(S_k^n - k\mu_n/2)\right\}$ converges to the expected minimum of this random walk. And by symmetry, this random walk has the same distribution as the process $\min_{t\geq 0}\tilde{S}_t$ defined above.

Having finished evaluating the right hand of (29), we now consider the left-hand side. The first order asymptotic behavior of the process $\sum_{n=1}^N g(U_{k,T}^n; p_0)$ is the same as that of $\sum_{n\in N_c}\mu_n(S_T^n - T\mu_n/2)$, which has drift $\Delta^2/2$ and variance Δ^2. Equivalent, the process $\sum_{n=1}^N g(U_{k,T}^n; p_0)$ has the same distribution as the process $\tilde{S}_T$ in (21). By writing

$$\mathbb{E}^0\left\{\max_{0\leq k<T}\sum_{n=1}^N g(U_{k,T}^n; p_0)\right\} = b + \mathbb{E}^0\left\{\max_{0\leq k<T}\sum_{n=1}^N g(U_{k,T}^n; p_0) - b\right\}, \tag{33}$$

and using nonlinear renewal theory to evaluate the expected overshoot of the process of (21) over the boundary ([Sie85], Chapter IX), we obtain

$$\mathbb{E}^0\left\{\max_{0\leq k<T}\sum_{n=1}^N g(U_{k,T}^n; p_0) - b\right\} \to \rho(\Delta). \tag{34}$$

4.3 Accuracy of the Approximations

We start with examining the accuracy of our approximations of the ARL and the expected detection delay in Theorems 1 and 2 for the two-sided mixture procedure. For a Monte Carlo experiment we use $N = 100$ sensors, $T_1 = 100$ and $\mu_n = 1$ for affected data streams. All data streams are contaminated by independent normally distributed noise with zero mean and unit variance. The comparisons between the theoretical and Monte Carlo ARLs for different values of p_0 are given in Table 1. To expedite Monte Carlo evaluation of the ARL, we use the fact that T is asymptotically exponentially distributed. Hence, as indicated in (19), $\mathbb{P}^{\infty}(T \leq m) \sim \lambda_b m$, provided m is not too large, with the rate λ_b depending on threshold b. Hence, $\mathbb{E}(T) \sim \lambda_b$. Hence by repeatedly simulating a relatively small number of observations (500 for the cases studied below), we can estimate $\mathbb{E}^{\infty}(T)$ when it is about 5000 or 10000. Our numerical results (obtained from 500 Monte Carlo trials and shown in Table 1) demonstrate that the ARL approximation in Theorem 1 is quite accurate.

Table 1: Average run length (ARL) of the mixture procedure, $T_1 = 100$.

p_0	b	Theory	Monte Carlo
1	84.500	5001.1	4837.3
1	86.240	10000	9416.0
0.1	27.670	5000.1	4684.3
0.1	28.718	10003	9591.4
0.03	16.433	5000.3	4992.8
0.03	17.307	10005	9660.9

We start assuming the parameter in the mixture procedure equals true parameter value $p_0 = p$. Comparison of the theoretical versus Monte Carlo expected detection delays (Fig. 1) demonstrates that the approximation in Theorem 2 is reasonably accurate in this case. In another case, we assume no knowledge of the true parameter value and fix $p_0 = 0.1$, and hence there is a mismatch between p and the true parameter value $p_0 \neq p$). The comparison (Fig. 2) shows that the approximation to the expected detection delay of Theorem 2 is also reasonably accurate in this case. The values of the comparison are also listed in Table 2.

Table 2: Expected detection delay of the mixture procedure with ARL = 5000±5 and $T_1 = 100$.

p	p_0	b	Theory	Monte Carlo
0.3	0.1	27.514	5.5085	4.0940
0.3	0.3	47.056	4.0205	3.8720
0.1	0.1	27.514	7.8530	7.5420
0.01	0.1	27.514	37.6325	37.6430
0.01	0.01	11.283	27.4596	27.7540

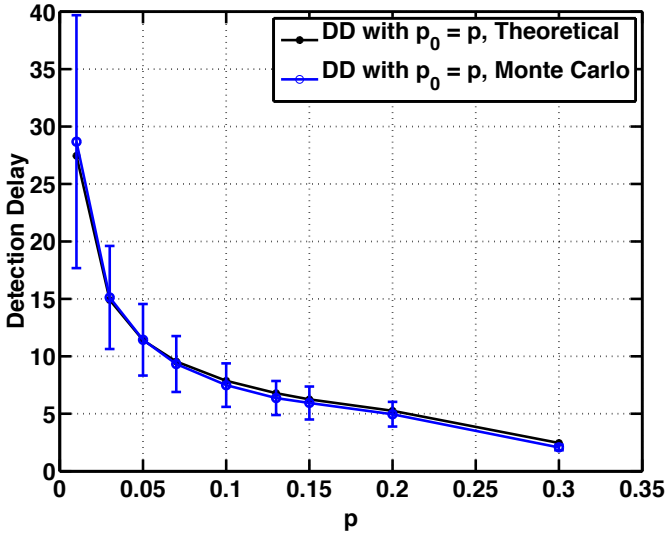

Figure 1: Theoretical versus expected detection delay (DD) of the mixture procedure, with $p_0 = p$, $T_1 = 100$, and different b such that ARL equals 5000.

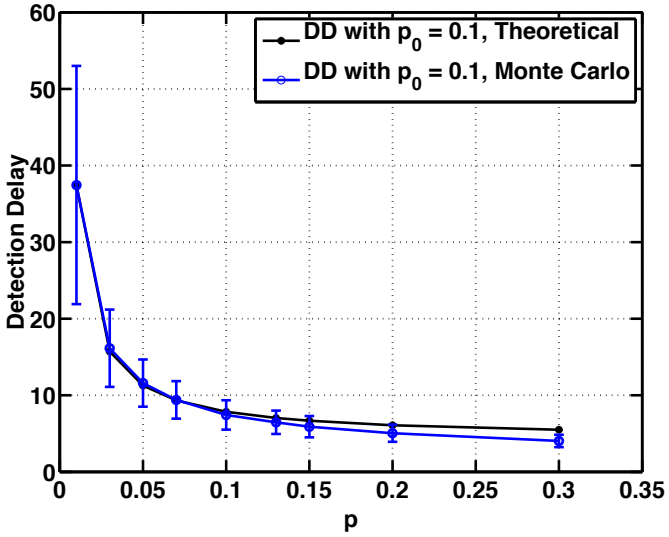

Figure 2: Theoretical versus expected detection delay of the mixture procedure, with $p_0 = 0.1$, $T_1 = 100$, and $b = 27.4$ such that ARL equals 5000.

5 Numerical Performance Analysis and Comparison

In this section, we will demonstrate the performance of the mixture procedure and compare it with other detection procedures.

5.1 Robustness to choice of p_0

First we demonstrate that the expected detection delay of the mixture procedure does not deviate much if the assumed p_0 is different from p. Again let $N = 100$ and all $\mu_n = 1$. We compare two mixture procedures. The first mixture procedure is ignorant of the true fraction. It sets $p_0 = 0.1$ regardless of p and chooses $b = 27.4$ so that ARL equals 5000. The second mixture procedure knows the true fraction of affected sensors, sets $p_0 = p$ and chooses threshold values for each p such that ARL equals 5000 for all p values. Fig. 3 shows the theoretical expected detection delays (from Theorem 2) of these two procedures when p varies from 0.01 to 0.3. Note that although the first procedure is ignorant of p and does not perform as well as the second procedure that knows p, the first procedure does very well in a broad band of values where $0.03 < p < 0.2$. We will return to this point in Section 5.4.

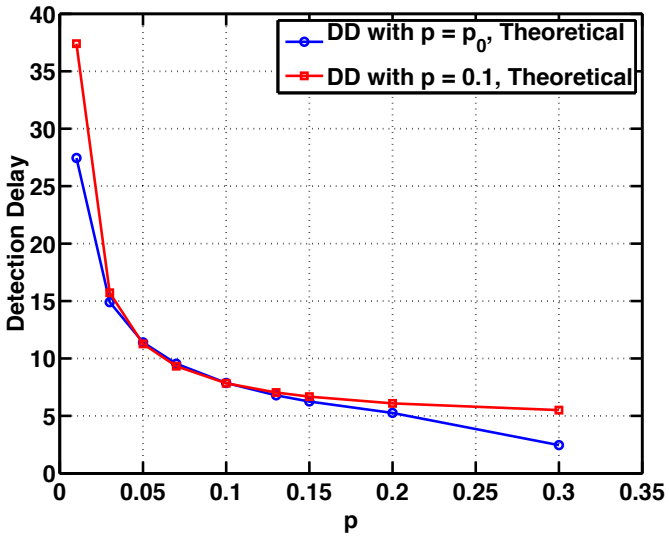

Figure 3: Comparison of the expected detection delay from Theorem 2, when there is no mismatch $p_0 = p$, and when $p = 0.1$ and there is a mismatch. The window size $T_1 = 100$, and the threshold values are chosen such that ARL equals 5000.

5.2 Effect of Window Size T_1

In approximating the detection delay, we assume the window size T_1 is large enough such that when the samples within the window contain a change-point, the mixture procedure of (9) can reach the

threshold. In this example, we study how large T_1 should be for this assumption to hold. Assume $N = 100$, $p_0 = p = 0.03$, and two scenarios with $\mu_n = 1$ and $\mu_n = 0.6$, respectively. We vary T_1 from 10 to 100 and find b using (17) such that ARL equals 5000. For this range of T_1 and required ARL, the threshold values are less than 17. By the first order analysis, T_1 should be greater than $2b/\Delta^2$ for the approximation in (2) to hold. This suggests T_1 to be greater than 12 for $\mu_n = 1$, and greater than 32 for $\mu_n = 0.6$, and is verified in Figure 4. The figure shows the Monte Carlo expected detection delay as T_1 increases. In Figure 4, the expected detection delay converges to the theoretical approximation when $T_1 > 20$ for $\mu_n = 1$ and $T_2 > 40$ for $\mu_n = 0.6$. Hence choosing T_1 greater than $4b/\Delta^2$ can reasonably guarantee the approximation in (2) to hold.

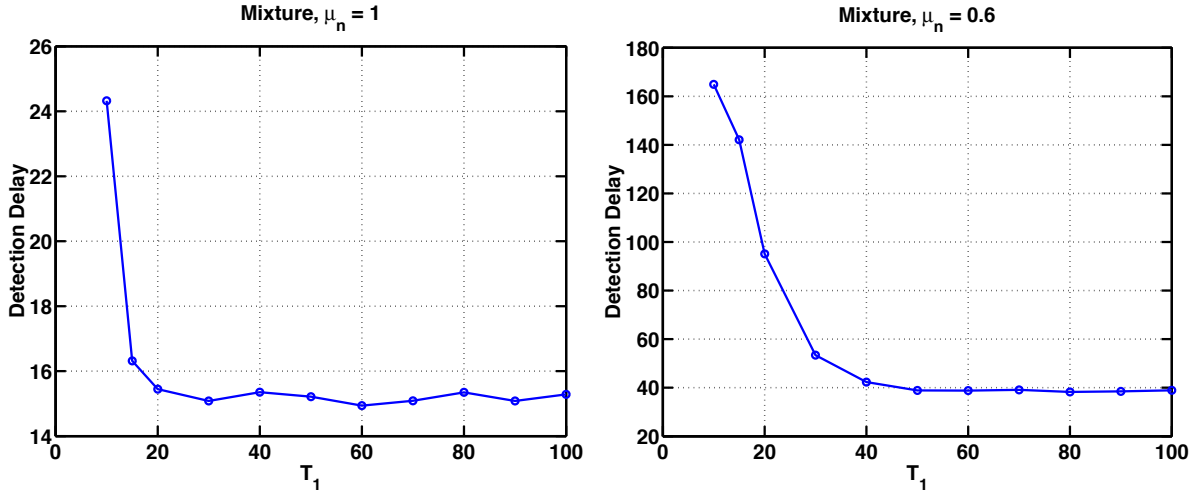

Figure 4: Expected detection delay versus T_1 for the mixture procedure with ARL being 5000.

5.3 Comparing Procedures

We compare the expected detection delays for the mixture procedure and other procedures, when their ARLs are all approximately 5000. We consider the one-sided procedures, with the change-amplitudes assumed to be positive (since the Mei's and TV's procedures are one-sided). The thresholds for the one-sided procedures are listed in Table 3. Assume the number of sensors $N = 100$ and $T_1 = 200$ for all procedures (except for the Mei's procedure, which has a recursive implementation and hence needs no windowing). The mixture procedure has no knowledge about the true fraction of affected sensors and it sets $p_0 = 0.1$. The Mei's and the modified TV procedures have no knowledge about the true change-point amplitudes and they use $\tilde{\mu}_n = 1$ for all scenarios. The expected detection delays are obtained from 500 Monte Carlo trials and are listed in Table 4. In the first two settings, we use $\mu_n = 1$ for all n or use $\mu_n = 0.7$ for all n. In the third case, the random μ_n are independent and identically distributed uniformly over [0.5, 1.5]. Once the μ_n's are chosen, they are fixed for 500 Monte Carlo trials. For different p, the total energy is normalized to one: $\sum_{n=1}^{N} \mu_n^2 = 1$, so that the change in the detection delay is not due to change of the total change-point energy. The expected detection delays when $\mu_n = 0.7$ are plotted in Fig. 5.

Note that the max procedure (11) has the smallest detection delay when $p = 0.01$, but it has the greatest delay for p greater than 0.1. The modified TV procedure has a slight advantage in detection delay when p approaches 0.3 and larger values, but much longer delay as p approaches to 0.01. This is expected since when most sensors are affected the modified TV procedure collects most energy of the change-point. When only one sensor is affected, the max procedure collects energy most efficiently in that it excludes noise from the unaffected sensors. The Mei's procedure performs well when p is approaches 0.3 and larger values, but it has longer delay than the mixture procedure then the mixture procedure. The mixture procedure has the smallest detection delay when p is greater than 0.05, and it is only slightly slower than the max procedure when $p = 0.01$. In this respect, the mixture procedure performs best over a wide range of p values.

Table 3: Thresholds for one-sided procedures with ARL 5000± 10, $T_1 = 200$.

Procedure	b	Monte Carlo ARL
Max	12.68	5000
GLR	53.085	4990
Mixture ($p_0 = 0.1$)	19.25 (19.483 - theo)	5000
Mei	88.525	4997.4
Modified TV	41.6	4993.4

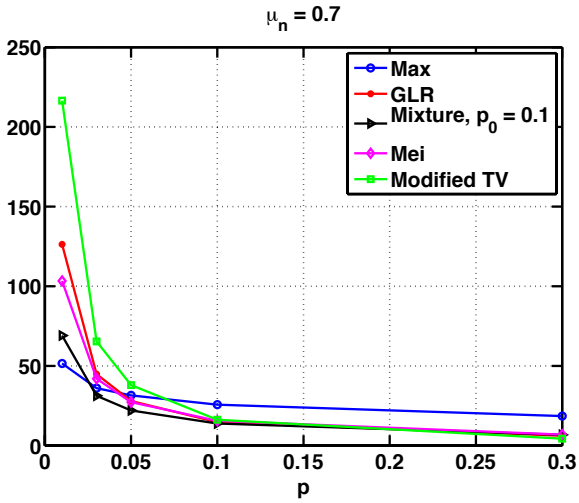

Figure 5: Comparison of expected detection delays for various μ and using $T_1 = 200$. The mixture procedure uses $p_0 = 0.1$, the Mei's and TV procedures use $\tilde{\mu}_n = 1$.

Table 4: Expected Detection Delays for the one-sided detection procedures with $N = 100$ and $T_1 = 200$, obtained from 500 Monte Carlo trials. The mixture procedure has $p_0 = 0.1$. Mei's and the Modified TV procedures use $\tilde{\mu}_n = 1$. The thresholds of these procedures with ARL 5000 are listed in Table 3.

p	method	DD, $\mu_n = 1$	DD, $\mu = 0.7$	DD, Random μ_n
0.01	max	25.5	49.648	7.822
	GLR	52.312	105.522	15.306
	mixture	31.954	59.39	9.328
	Mei	53.15	103.796	22.85
	Modified TV	81.968	213.6506	27.688
0.03	max	18.142	33.322	9.802
	GLR	18.666	35.842	15.18
	mixture	14.224	26.672	10.6
	Mei	22.992	41.56	20.37
	Modified TV	27.236	65.968	23.724
0.05	max	15.468	28.376	14.6
	GLR	12.206	21.838	14.382
	mixture	10.35	18.892	11.374
	Mei	15.748	26.9	18.544
	Modified TV	15.52	38.79	20.994
0.1	max	12.576	23	21.502
	GLR	6.724	11.782	13.844
	mixture	6.668	11.592	13.37
	Mei	9.646	15.37	17.79
	Modified TV	6.754	15.712	21.022
0.3	max	9.606	16.672	38.872
	GLR	3.008	4.402	11.866
	mixture	3.49	5.604	14.094
	Mei	4.9	6.956	16.578
	Modified TV	3.028	4.304	22.26
0.5	max	8.574	14.432	55.278
	GLR	2.262	2.992	10.37
	mixture	2.758	3.976	14.39
	Mei	3.842	4.98	14.684
	Modified TV	2.282	3.048	18.748
1	max	7.178	12.096	81.982
	GLR	2	2.046	8.034
	mixture	2.038	2.642	13.162
	Mei	2.992	3.426	11.906
	Modified TV	2	2.076	10.772

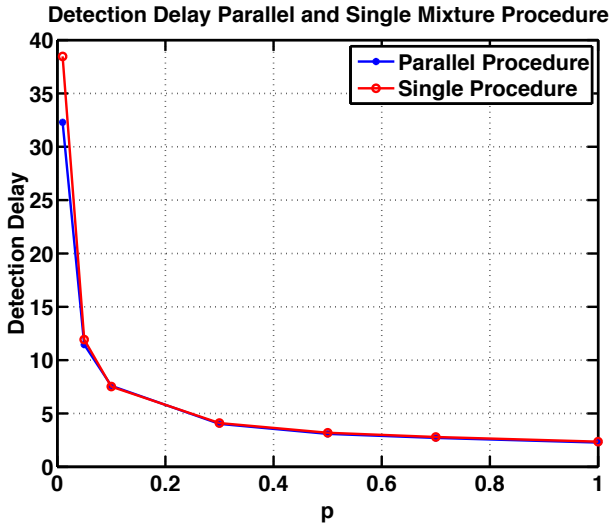

Figure 6: Expected detection delay of the parallel mixture procedure with $p_1 = 0.2$ and $p_2 = 0.03$ versus that of a single mixture procedure with $p_0 = 0.1$.

5.4 Parallel Mixture Procedure

Although we have demonstrated in Section 5.1 that the mixture procedure is not sensitive to the precision of p_0, to achieve robustness over a wider range of p, we consider a parallel mixture procedure that combines several mixture procedures, each using a different p_0 to monitor a different range of p values. The thresholds of these individual mixture procedures are chosen such that they have the same ARL. For example, we can use two mixture procedures with a small $p_0 = p_1$ and a large $p_0 = p_2$, respectively, and then choose thresholds b_1 and b_2 to obtain the same individual ARL. The parallel procedure claim a detection once any one of the mixture procedure reaches their threshold:

$$T_{\text{parallel}} \triangleq \min\{T_{\text{mixture}}(p_1), T_{\text{mixture}}(p_2)\}. \tag{35}$$

The ARL of the parallel mixture procedure is smaller than each individual ARL. As we will demonstrate in the following, the parallel procedure has a smaller expected detection delay than the individual mixture procedure over a wide range.

To compare the performance of the parallel procedure with that of the a single mixture procedure, we consider a case with $N = 100$ and $T_1 = 100$. The single mixture procedure uses $p_0 = 0.1$ and threshold value $b = 27.692$ such that ARL equals 5072.7. The parallel procedure uses two mixture procedures with $p_1 = 0.2$ and $p_2 = 0.03$ and threshold values $b_1 = 39.51$ and $b_2 = 17.027$ such that their individual ARLs are approximately 8000. The resulted ARL for parallel mixture procedure is 5075.6. Fig. 6 shows that the expected detection delays of the parallel procedure is

smaller than that of the single procedure, and is strict smaller at lower values of p.

6 Profile-based Procedure for Structured Problems

In discussions so far we have assumed there is no spatial structure relating the change-point amplitudes at difference sensors. In the following we will consider an alternative scenario, the structured problem. In structured problems we assume there is the parameterized profile of the sensor amplitudes. We have some knowledge about such a profile, and incorporate such profile into the definition of the likelihood ratio statistic.

Consider a multi-sensor system, with the location of the nth sensor given by its coordinates in the Euclidean space (x_n, y_n), $n = 1, \cdots, N$. Suppose we are only interested in the source located in a region $\mathcal{D} \subset \mathbb{R}^2$. Assume the change-point amplitude at the nth sensor is determined by the following expression

$$\mu_n = \sum_{m=1}^{P} r_m \alpha_{\boldsymbol{z}_m}(x_n, y_n), \tag{36}$$

where P is the number of sources, the d-dimensional vector $\boldsymbol{z}_m \in \mathcal{D} \subset \mathbb{R}^d$ is the (unknown) spatial location of the mth source over some region $\mathcal{D}$, and the scalar r_m is the (unknown) amplitude of each source. In (36), the profile function

$$\alpha_{\boldsymbol{z}_m}(x_n, y_n) : \mathbb{R}^2 \to \mathbb{R}^+, \tag{37}$$

describes how the mth change-point source decays at the nth sensor. We assume some knowledge about this profile function is available. For example, $\alpha_{\boldsymbol{z}}(x, y)$ can be a decreasing function of the Euclidean distance between $\boldsymbol{z}$ and the location of the sensor (x, y). In principle, $\alpha_{\boldsymbol{z}}(x, y)$ may also depend on finitely many parameters, such as the rate of decay of the function and an amplitude constant of the profile. See [Rab94] or [SSSW03] for examples in a fixed sample context.

If the change-point amplitudes r_m are multiplied by a positive constant and the profile $\alpha_{\boldsymbol{z}_m}(x_n, y_n)$ divided by the same constant, the values of profile μ_n do not change. To remove this ambiguity, it is mathematically convenient to assume that the change-point profile has been standardized to have unit norm. Define a vector $\boldsymbol{\alpha}_{\boldsymbol{z}} = [\alpha_{\boldsymbol{z}}(x_1, y_1), \cdots, \alpha_{\boldsymbol{z}}(x_N, y_N)]^\top$. The normalization of the profile function means that we require the profile function to satisfy

$$\boldsymbol{\alpha}_{\boldsymbol{z}}^\top \boldsymbol{\alpha}_{\boldsymbol{z}} = 1, \tag{38}$$

for any $\boldsymbol{z} \in \mathcal{D}$.

6.1 Profile-based procedure

In this section, we will derive the log GLR statistic by incorporating the profile function. First, assume that the profile function (36) is only due to one source

$$\mu_n = r\alpha_{\boldsymbol{z}_0}(x_n, y_n). \tag{39}$$

Using this profile, the log likelihood function for observations from all sensors (1) is given by the following:

$$\begin{aligned} l(t, k, r, \boldsymbol{z}) &= \sum_{n=1}^{N} \sum_{s=k+1}^{t} (r\alpha_{\boldsymbol{z}}(x_n, y_n) y_s^n - r^2 \alpha_{\boldsymbol{z}}^2(x_n, y_n)/2) \\ &= r(t-k)^{1/2} \boldsymbol{\alpha}_{\boldsymbol{z}}^\top \boldsymbol{U}_{k,t} - (t-k) r^2/2, \end{aligned} \tag{40}$$

with the vectors $\boldsymbol{y}_s = [y_s^1, \ldots, y_s^N]^\top$ and $\boldsymbol{U}_{k,t} = [U_{k,t}^1, \ldots, U_{k,t}^N]^\top$. In (40), we have used the property (38) that profile function has unit norm for any source location $\boldsymbol{z} \in \mathcal{D}$.

Then we solve for the maximum-likelihood estimate of r from (40). We take the derivative of the log likelihood function (40) with respect to r, set it to zero,

$$(t-k)^{1/2} \boldsymbol{\alpha}_{\boldsymbol{z}}^\top \boldsymbol{U}_{k,t} - (t-k) r = 0. \tag{41}$$

and solve for

$$\hat{r} = \frac{\boldsymbol{\alpha}_{\boldsymbol{z}}^\top \boldsymbol{U}_{k,t}}{(t-k)^{1/2}}. \tag{42}$$

Substitution of maximum likelihood estimate $\hat{r}$ into (40) leads to

$$l(t, k, \hat{r}, \boldsymbol{z}) = \frac{1}{2} (\boldsymbol{\alpha}_{\boldsymbol{z}}^\top \boldsymbol{U}_{k,t})^2. \tag{43}$$

We further maximize the function (43) with respect to the change-point time k and source location z and obtain the log GLR statistic. Using this log GLR statistic, the profile-based procedure is given by:

$$T_{\text{profile}} = \inf \left\{ t : \max_{t - T_1 \le k < t} \max_{\boldsymbol{z} \in \mathcal{D}} \frac{1}{2} (\boldsymbol{\alpha}_{\boldsymbol{z}}^\top \boldsymbol{U}_{k,t})^2 \ge b \right\}. \tag{44}$$

Note that (44) is a matched-filter type of statistic: the statistic $\boldsymbol{U}_{k,t}$ is matched to a profile $\boldsymbol{\alpha}_{\boldsymbol{z}}$.

When μ_n is given by the general form, a sum of multiple profile functions, the corresponding log

GLR statistic is more complex and is derived in Appendix B. However, we can show that when the sources are non-overlapping, the detection procedure based on the log GLR statistic for the general form of profile function (36) is equivalent to the procedure (44) (when there is only one source). In the following, we assume the sources are non-overlapping.

6.2 Theoretical ARL of profile-based procedure

In this section, we will derive the theoretical ARL of the profile-based procedure based on the result presented in [SY08]. We consider a special case of the general profile-based procedure in the two-dimensional space, $d = 2$. Assume the profile function is given by a two-dimensional Gaussian function:

$$\alpha_{\boldsymbol{z}}(x, y) = \frac{1}{\sqrt{2\pi\beta}} e^{-\frac{1}{4\beta}(x-z_x)^2 - \frac{1}{4}(y-z_y)^2}, \qquad (x, y) \in \mathbb{R}^2, \beta > 0, \tag{45}$$

for a sensor locates at (x, y) and a source locates at $\boldsymbol{z} = (z_x, z_y)$. The parameter $\beta > 0$ controls of rate of profile decay. Define the inner product of two function in L_2 as

$$\langle f(x, y), g(x, y) \rangle = \int \int_{\mathbb{R}^2} f(x, y) g(x, y) dx dy. \tag{46}$$

For short-hand we also write $\alpha_{\boldsymbol{z}}$ for $\alpha_{\boldsymbol{z}}(x, y)$. It can be verified that the norm of (45) is one:

$$\langle \alpha_{\boldsymbol{z}}, \alpha_{\boldsymbol{z}} \rangle = 1, \tag{47}$$

for any $\boldsymbol{z}$ within a region $\boldsymbol{z} \in \mathcal{D} \subset \mathbb{R}^2$. So this $\boldsymbol{\alpha}_{\boldsymbol{z}}$ approximately satisfies (38). This approximation is good if β is sufficiently large, the distance between points of the grid sufficiently small, and the boundary of the contour for all sensor locations is well approximated by integration over the entire Euclidean space [SY08].

In [SY08], theoretical ARL for the profile-based procedure with a general profile function is given. However, the profile-based procedure considered in [SY08] has a different form from (44). Moreover, the procedure in [SY08] has a different form for the threshold. The procedure in [SY08] is given by

$$T = \inf \left\{ t : \max_{t - T_1 < k < t} \max_{\boldsymbol{z} \in \mathcal{D}} \max_{r > 0} \sum_{s=k}^{t} \left(r \boldsymbol{\alpha}_{\boldsymbol{z}}^\top \boldsymbol{y}_s - \frac{r^2}{2} \right) \geq \log A + \log c \right\}, \tag{48}$$

The result in [SY08] (equation (14)) states that if the sensors locations are specified by a d-dimensional vector, the observations $\{y_s^n\}$ are i.i.d. and normally distributed with zero mean and variance σ^2, the profile function is given by $\alpha_{\boldsymbol{z}}$, the threshold is specified by $\log A + \log c$ for some

constant A and c, and magnitude $r \geq 0$ is nonnegative, then the tail-probability for the stopping time of the profile-based procedure (48) is given by

$$\mathbb{P}^{\infty}(T \leq t) = \frac{t}{Ac} \cdot \left(\frac{\log A}{2\pi\sigma^2}\right)^{(d+1)/2} \int_{\sqrt{2\log A/T_1}}^{\sqrt{2\log A}} \int_{\mathcal{D}} \frac{\nu(x)^2 x^d \sqrt{|\langle \dot{\alpha}_{\boldsymbol{z}}, \dot{\alpha}_{\boldsymbol{z}}^{\top} \rangle|}}{[I(x)]^{(d-1)/2}} dx d\boldsymbol{z}. \tag{49}$$

Here $\dot{\alpha}_{\boldsymbol{z}}$ is a two-dimensional vector.

If we set $\log A = b/2$, $c = 1$, $d = 2$, and replace the maximization $\max_{r>0}$ with $\max_r$, the procedure (48) of [SY08] is equivalent to the profile-based procedure (44) that we consider here. Hence we can use the result (49) in [SY08] for this special case. We have to also take into account that when the probability of hitting a threshold before t for a two-sided problems is twice of the probability for a one-sided problem. Hence, from (49), the tail probability of T_{profile} is given by

$$\mathbb{P}^{\infty}(T_{\text{profile}} \leq t) = 2 \cdot \frac{t}{e^{b/2}} \left(\frac{b}{4\pi}\right)^{3/2} \cdot \sqrt{2} \int_{\sqrt{b/T_1}}^{\sqrt{b}} x\nu^2(x) dx \cdot \int_D \sqrt{|\langle \dot{\alpha}_{\boldsymbol{z}}, \dot{\alpha}_{\boldsymbol{z}}^{\top} \rangle|} d\boldsymbol{z}, \tag{50}$$

where we have used the fact that for normal random variables with zero mean and unit variance, the K-L divergence is $I(x) = x^2/2$. The next step is to evaluate the last term in (50) that involves a integration, which corresponds to the volume of $\alpha_{\boldsymbol{z}}$ over D. By (45), we have

$$\dot{\alpha}_{\boldsymbol{z}} = \left[\frac{d\alpha_{\boldsymbol{z}}}{dz_x}, \frac{d\alpha_{\boldsymbol{z}}}{dz_y}\right]^{\top} = \frac{1}{\sqrt{2\pi\beta}} e^{-\frac{1}{4\beta}[(x-z_x)^2+(y-z_y)]^2} [(x-z_x)/(2\beta), (y-z_y)/(2\beta)]^{\top}. \tag{51}$$

Hence

$$\begin{aligned}
&\langle \dot{\alpha}_{\boldsymbol{z}}, \dot{\alpha}_{\boldsymbol{z}}^{\top} \rangle \\
&= \begin{bmatrix} \frac{1}{4\beta^2} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-z_x)^2} (x-z_x)^2 dx & \frac{1}{4\beta^2} \int\int \frac{1}{2\pi} e^{-\frac{1}{2}(x-z_x)^2 - \frac{1}{2}(y-z_y)^2} (x-z_x)(y-z_y) dx dy \\ \frac{1}{4\beta^2} \int\int \frac{1}{2\pi} e^{-\frac{1}{2}(x-z_x)^2 - \frac{1}{2}(y-z_y)^2} (x-z_x)(y-z_y) dx dy & \frac{1}{4\beta^2} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-z_y)^2} (y-z_y)^2 dy \end{bmatrix} \\
&= \begin{bmatrix} 1/(4\beta^2) & 0 \\ 0 & 1/(4\beta^2) \end{bmatrix},
\end{aligned}$$

the determinant $|\langle \dot{\alpha}_{\boldsymbol{z}}, \dot{\alpha}_{\boldsymbol{z}}^{\top} \rangle| = 1/(16\beta^4)$, and

$$\int_D \sqrt{|\langle \dot{\alpha}_{\boldsymbol{z}}, \dot{\alpha}_{\boldsymbol{z}}^{\top} \rangle|} d\boldsymbol{z} = |D|/(4\beta^2), \tag{52}$$

with $|D|$ denoting the area of D. From (52) we have

$$\mathbb{P}^{\infty}(T_{\text{profile}} \leq t) = \frac{t}{e^{b/2}} \left(\frac{b}{4\pi}\right)^{3/2} \cdot \sqrt{2} \int_{\sqrt{b/T_1}}^{\sqrt{b}} x\nu^2(x)dx \cdot \frac{D}{4\beta^2} \cdot 2. \tag{53}$$

Based on (53), the approximation for the ARL of the profile-based procedure is given by:

$$\mathbb{E}^{\infty}\{T_{\text{profile}}\} = e^{b/2} \left(\frac{b}{4\pi}\right)^{-3/2} \cdot \left[\sqrt{2} \int_{\sqrt{b/T_1}}^{\sqrt{b}} x\nu^2(x)dx \cdot \frac{D}{2\beta^2}\right]^{-1}. \tag{54}$$

Note that when $d = 1$, the profile-based procedure corresponds to the log GLR procedure (10). We can show that the theoretical ARL of the profile-based procedure with $d = 1$ is equal to the theoretical ARL of the log GLR procedure given in [SV95].

6.3 Numerical examples

6.3.1 Approximate method to simulate ARL

We can use two methods to simulate the ARL or expected detection delay of a sequential procedure. The first method has been used in Section 3, which is to directly simulate the detection statistic process. In other word, we generate N sequences of independent normal random variables, form the process from $t = 1$, and let the process continue until it hits the threshold, and record the stopping time. We such a realization a Monte Carlo trial. We repeat the trial for many times and compute the sample mean of the stopping times as the average run length or the expected detection delay. However, this simulation method is very time consuming, when the scale of the problem becomes large and the procedure is computationally intensive to evaluate for each time instance, such as the profile-based procedure. For the latter case, we consider using the following alternate to approximately simulate the ARL.

In the second method, for each trial, we also generate N sequences of independent random variables, start the process from $t = 1$. But we only let the process continue till a given time, say $t = 250$, and record the value of the process at that time. We repeat this trial for many times, and compute the frequency that the simulated processes exceed the threshold at the end-point $t = 250$. This frequency can be used to approximate the ARL of the procedure due to the following reason. If we assume the stopping time defined by the procedure is asymptotically exponential when the threshold b is large, then $\mathbb{E}^{\infty}\{T\} = \lambda$, and we have $\mathbb{P}^{\infty}(T \leq t) = 1 - e^{-t/\lambda} \approx t/\lambda$. From this relationship, we can estimate $\mathbb{E}^{\infty}\{T\} \approx t/\mathbb{P}^{\infty}(T \leq t)$. For example, when ARL is 5000, we have

$\lambda = 5000$. For $t = 250$, the tail probability is given by $\mathbb{P}^{\infty}(T \leq t) = t/\lambda = 250/5000 = 0.05$. Hence if we choose the threshold b such that $\mathbb{P}^{\infty}(T \leq 250) \approx 0.05$, the resulted ARL is approximately 5000.

6.3.2 Small example

Consider a two-dimensional version of the profile-based procedure. We assume that the profile is given by the Gaussian function (45) with parameter $\beta = 1/4$. We consider window-truncated detection procedures and choose $T_1 = 100$ for both the mixture procedure and the profile-based procedure.

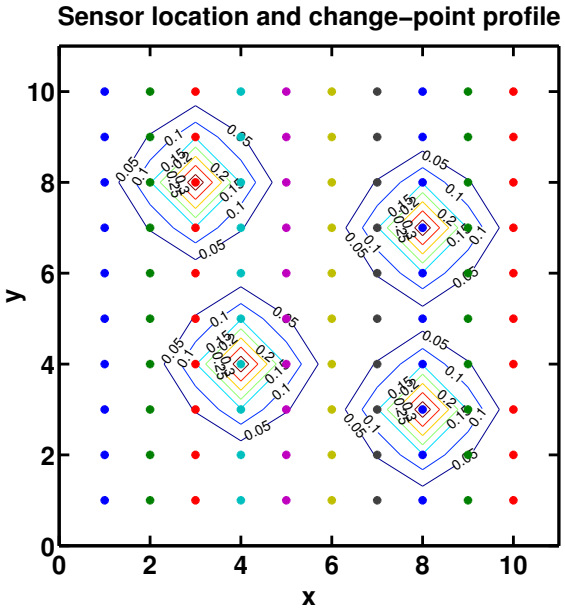

Figure 7: Four non-overlapping sources observed by a 10 by 10 sensor grid ($N = 100$). The sources have the Gaussian profile function given in (45) with $\beta = 1/4$.

Table 5: Expected detection delays for mixture versus profile-based procedures in the two dimensional example. The mixture procedure uses $p_0 = 0.1$ and $T_1 = 100$.

	Profile, $b = 26.7$		Mixture, $b = 27.682$	
	ARL	Delay	ARL	Delay
One-Source	5037.2	26.364	5039.6	43.234
Two-Source	5037.2	37.366	5039.6	49.498
Three-Source	5037.2	48.332	5039.6	52.666
Four-Source	5037.2	57.834	5039.6	57.980

In comparing the performance of the mixture procedure and that of the profile-based procedure, we consider four scenarios. There can be up to four non-overlapping sources with identical profile

function described by (45) with $\beta = 1/4$. Moreover, the profile-based procedure knows the true profile and uses the true profile in forming the detection statistic. For the one-source case, the source amplitude $r_1 = 1$ and the location is $z_1 = (4, 4)$. For the two-source case, the source amplitudes $r_1 = r_2 = 1/\sqrt{2}$ and locations are $z_1 = (4, 4)$ and $z_2 = (8, 7)$. For the three-source case, the source amplitudes are $r_p = 1/\sqrt{3}$, $p = 1, 2, 3$ and locations are $z_1 = (4, 4)$, $z_2 = (8, 7)$, and $z_3 = (3, 8)$. For the four-source case, the source amplitudes are $r_p = 1/2$, $p = 1, \cdots, 4$ and the source locations are $z_1 = (4, 4)$, $z_2 = (8, 7)$, $z_3 = (3, 8)$ and $z_4 = (8, 3)$. The amplitudes r_m of the sources in the multiple-signal cases are chosen such that the total energy of the sources is equal to one.

There are $N = 100$ sensors uniformly distributed on a 10 by 10 grid with locations on the integer points (x_n, y_n) with $x_n = 1, 2, \cdots, 10$ and $y_n = 1, 2, \cdots, 10$. The sensor locations as well as the signal amplitude contour when there are four sources are illustrated in Fig. 7.

When there is one source, approximately 10% sensors are affected by the change-point. Since the sources are non-overlapping, when there are two, three and four sources, the fractions of affected sensors are 20%, 30% and 40%, respectively. In all cases we set $p_0 = 0.1$ for the mixture procedure so that it assumes no knowledge of underlying source.

The thresholds of the profile-based and the mixture procedures are chosen such that their ARLs are about 5000. These thresholds are listed in Table 5. The threshold of the profile-based procedure is simulated from 500 Monte Carlo trials using the first simulation method in the previous section, Section 6.3.1. The threshold of the mixture procedure is obtained from Theorem 1.

We compare the expected detection delay of the profile-based method with that of the mixture method, which is listed in Table 5. The results in Table 5 demonstrate that although the mixture procedure detects a bit slower than the profile-based procedure in the first three cases, the mixture procedure detects equally quickly to the profile-based procedure in the four-source scenario. This is because with four sources, the energy of the change-point is distributed over four sources, but the profile-based procedure can only collect energy from one source. Although the mixture procedure also has a mismatch from the true underlying fraction of affected sensors (it assumes $p_0 = 0.1$ rather than the true $p = 0.4$), it is robust to the mismatch as we have demonstrated in Section 5.1. The profile-based procedure also collects more noise from searching for the unknown source location in the log GLR statistics. Hence in the last case, the mixture procedure can achieve a comparable performance but with much lower computational complexity than the profile-based procedure.

6.3.3 Large example

Again consider a two-dimensional version of the profile-based procedure. We assume that the profile is given by the Gaussian function (45) with parameter $\beta = 1$, so in this case the profile decays slower. We also assume the procedures are window-truncated with $T_1 = 100$ for both the mixture and the profile-based procedures.

In this example, there are more sensors than the small example. Assume $N = 625$ sensors distributed over a larger area. The locations of the sensors are (x_n, y_n) with $x_n = 1, 2, \cdots, 25$ and $y_n = 1, 2, \cdots, 25$. Again there can be up to four non-overlapping signal sources and the total energy of the sources are normalized to one for all scenarios. The sensor locations as well as the signal amplitude contour when there are four sources are illustrated in Fig. 8. When there is one source, approximately $p = 0.016$ sensors are affected. The mixture method assumes $p_0 = 0.01$ for all scenarios regardless of the number of actual sources.

The thresholds of these procedures are chosen such that their ARLs are approximately 5000. These thresholds are listed in Table 6. For this example, we obtain the threshold of the profile-based procedure from 500 Monte Carlo trials using the second simulation method in Section 6.3.1. The threshold of the mixture procedure is obtained from Theorem 1.

We also verified the accuracy of the second approximate simulation method using the theoretical result we derived in (53) of Section 6.2. The theoretical tail-probability in (53) shows that when $b = 30.64$, $P_\infty(T \le 250) = 0.0503$ (which corresponds to ARL of 4970.2). The tail-probability we obtained using 500 Monte Carlo trials shows that when $b = 30.01$, $P_\infty(T \le 250) = 0.0508$, which corresponds to ARL of 4921.3. This demonstrates that the second simulation method of Section 6.3.1 is quite accurate for the profile-based procedure in this case.

Table 7 listed the comparison of the expected detection delay of the profile-based procedure with that of the mixture procedure. The results demonstrate that for this larger example (and wider spread of sources), it is harder for the mixture procedure to obtain a comparable performance to that the profile-based procedure. On the other hand, the results also show that, in this example, the profile-based procedure takes advantage that it knows the exact profile of sources. When there is a mismatch between the assumed profile and the actual one, the performance of the profile-based procedure may degrade.

Table 6: $\mathbb{P}^\infty(T_{\text{profile}} \le 250)$ of Profile-Based Procedure

	b	Theoretical	Monte Carlo
ARL	30.64	0.0503	0.0410
ARL	30.01	0.0668	0.0508

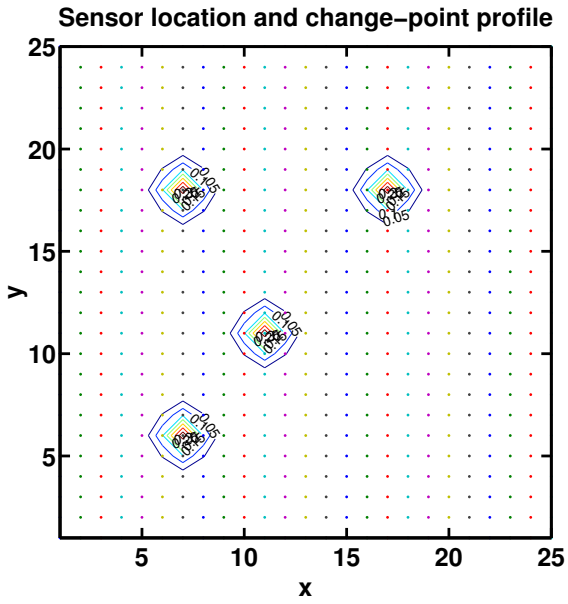

Figure 8: Four sources observed by a 25 by 25 sensor grid ($N = 625$). The sources have Gaussian profile function given in (45) with $\beta = 1$.

Table 7: Comparison of expected detection delay (EDD) with one source

	b	ARL	EDD $r = 1$	EDD, $r = 1.5$
Profile-based procedure	30.64	4970.2	29.2	13.5
Mixture procedure	25.356	4969.2	105.7	41.4

7 Conclusions

We have presented a mixture procedure for the unstructured multi-sensor sequential detection problem. We assume that the pre- and post- change samples are normal distributed, and that the post-change mean as well as the set of affected sensors are unknown. The mixture procedure first soft-thresholds the generalized likelihood ratio (GLR) statistic of each sensor, then sums the results up to form the detection statistic, and compares the statistic with a prescribed threshold. For performance analysis, we have derived the average run length (ARL) and the expected detection delay of the mixture procedure, and have shown that these approximations have reasonable accuracy. We have also shown that the mixture procedure does not require a precise knowledge on p. Moreover, we have compared numerically the mixture procedure with other procedures for the unstructured problem, and demonstrated that the mixture procedure has a lower expected detection delay than other procedures over a wide range of p values. For the structured problem, we have shown that the mixture procedure has a similar expected detection delay performance to the profile-based procedure when the change-point are multiple sources.

References

[Ald88] D. Aldous. *Probability approximations via the Poisson clumping heuristic*. Springer, 1 edition, Nov. 1988.

[CGV+10] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. C. M. Leung. Body area networks: a survey. *Mobile Networks and Applications*, DOI:10.1007/s11036-010-0260-8:1 – 23, Aug. 2010.

[LLR09] C. Levy-Leduc and F. Roueff. Detection and localization of change-points in high-dimensional network traffic data. *The Annuals of Applied Statistics*, 3(2):637–662, 2009.

[Mei10] Y. Mei. Efficient scalable schemes for monitoring a large number of data streams. *Biometrica*, 97(2):419 – 433, 2010.

[PRT03] A. Petrov, B. L. Rozovskii, and A. G. Tartakovsky. Efficient nonlinear filtering methods for detection of dim targets by passive systems. *Submitted to Multitarget-Multisensor Tracking: Applications and Advances*, IV, 2003.

[PS75] M. Pollak and D. Siegmund. Approximations to the expected sample size of certain sequential tests. *Ann. Statist.*, 3(6):1267 – 1282, 1975.

[Rab94] D. Rabinowitz. *Detecting clusters in disease incidence*, pages 255 – 275. Change-point Problems. IMS, Hayward, CA, 1994.

[Sie85] D. Siegmund. *Sequential Analysis: Tests and Confidence Intervals*. Springer Series in Statistics. Springer, 1985.

[SSSW03] K. Shafie, B. Sigal, D. Siegmund, and K. Worsley. Rotation space random fields with an application to fmri data. *Ann. Statist.*, 31:1732 – 1771, 2003.

[SV95] D. Siegmund and E. S. Venkatraman. Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Ann. Statist.*, 23(1):255 – 271, 1995.

[SY07] D. Siegmund and B. Yakir. *The statistics of gene mapping*. Springer, 2007.

[SY08] D. O. Siegmund and B. Yakir. Detecting the emergence of a signal in a noisy image. *Statistics and Its Inference*, 1:3–12, 2008.

[TV08] A. G. Tartakovsky and V. V. Veeravalli. Asymptotically optimal quickest change detection in distributed sensor. *Sequential Analysis*, 27(4):441–475, 2008.

[ZYS10] N. Zhang, B. Yakir, and D. O. Siegmund. Detecting simultaneous variant intervals in aligned sequences. *Submitted to Annals of Applied Statistics*, 2010.

A Proof of Lemma 3

Proof. We use the following identity:

$$\max_{0<k\le T} h_k = \max\left\{\max_{0<k\le k_0} h_k, \max_{k_0<k\le T} h_k\right\}, \tag{55}$$

which holds for any sequence h_k. In particular,

$$h_k = -\sum_{n\in\mathcal{N}_a} \mu_n(S_k^n - k\mu_n/2) + \sum_{n\in\mathcal{N}_a} [(S_T^n - S_k^n) - (T-k)\mu_n]^2/[2(T-k)] + \sum_{n\in\mathcal{N}_a^c} g(U_{k,T}^n; p_0). \tag{56}$$

First we evaluate the term $\max_{0<k\le k_0} h_k$ in (55). Since when $b\to\infty$, $T\sim b/\Delta$. For $0<k\le k_0$, $k/T < k_0/T = \sqrt{T}/T = \sqrt{\Delta/b} \to 0$. Also, $S_k^n/S_T^n \to k\Delta/(T\Delta) \to 0$ by the law of large number

and the previous argument. Hence we have

$$\begin{aligned}
&\sum_{n\in\mathcal{N}_a}[(S_T^n - S_k^n) - (T-k)]^2/[2(T-k)] \\
&= \sum_{n\in\mathcal{N}_a} \frac{[S_T^n\,(1 - S_k^n/S_T^n) - T\,(1-k/T)]^2}{2T\,(1-k/T)} \\
&\to \sum_{n\in\mathcal{N}_a} (S_T^n - T)^2/(2T).
\end{aligned} \tag{57}$$

Since $g(x;p_0) = \log[1 - p_0 + p_0\exp(x^2/2)]$ as a function of p_0 is monotonically increasing in p, we have

$$g(x;p_0) \le x^2/2. \tag{58}$$

Hence the last term in (56) is upper-bounded by:

$$g(U_{k,T}^n;p_0) = \log(1 - p_0 + p_0\exp[(U_{k,T}^n)^2/2]) \le (S_T^n - S_k^n)^2/[2(T-k)]. \tag{59}$$

Using a similar argument to (57), when $b\to\infty$, for $k\le k_0$, we have

$$(S_T^n - S_k^n)^2/[2(T-k)] \to (S_T^n)^2/(2T). \tag{60}$$

By the bounded convergence theorem,

$$\sum_{n\in\mathcal{N}_a^c} g(U_{k,T}^n;p_0) \to \sum_{n\in\mathcal{N}_a^c} \log\left\{1 - p_0 + p_0\exp[(S_T^n)^2/(2T)]\right\} = \sum_{n\in\mathcal{N}_a^c} g(U_{0,T}^n;p_0). \tag{61}$$

Hence when $b\to\infty$,

$$\max_{0<k\le k_0} h_k \to \max_{0<k\le k_0}\left\{-\sum_{n\in\mathcal{N}_a}\mu_n(S_k^n - k\mu_n/2)\right\} + \sum_{n\in\mathcal{N}_a}(S_T^n - T)^2/(2T) + \sum_{n\in\mathcal{N}_a^c} g(U_{0,T}^n;p_0). \tag{62}$$

Then we evaluate $\max_{k_0<k\le T} h_k$ in (55). Note that when $b\to\infty$,

$$-\sum_{n\in\mathcal{N}_a}\mu_n(S_k^n - k\mu_n/2) \to -\Delta k \le -\Delta k_0 = -\Delta\sqrt{T} \sim -\Delta b/\Delta \to -\infty, \tag{63}$$

and the other two terms are bounded with high probability. By Markov inequality:

$$\begin{aligned}&\mathbb{P}\left\{\sum_{n\in\mathcal{N}_a}[(S_T^n-S_k^n)-(T-k)\mu_n]^2/[2(T-k)]\geq c\right\}\\&\leq\mathbb{E}\left\{\sum_{n\in\mathcal{N}_a}[(S_T^n-S_k^n)-(T-k)\mu_n]^2/[2(T-k)]\right\}\Big/\, c=M/c,\end{aligned}\tag{64}$$

By monotonicity (58) and Markov inequality,

$$\mathbb{P}\left\{\sum_{n\in(\mathcal{N}_a)^c}g(U_{k,T}^n;p_0)\geq c\right\}\leq\mathbb{P}\left\{\sum_{n\in(\mathcal{N}_a)^c}(S_T^n-S_k^n)^2/[2(T-k)]\geq c\right\}\leq(N-M)/c.\tag{65}$$

Hence $\max_{k_0<k\leq T}h_k\to-\infty$ as $b\to\infty$. Plug this result and (62) into (55), we have

$$\max_{0<k\leq T}h_k=\max\left\{\max_{0<k\leq k_0}h_k,\max_{k_0<k\leq T}h_k\right\}\to\max_{0<k\leq k_0}h_k,\tag{66}$$

which concludes the proof for Lemma 3. □

B Multiple Overlapping Sources

When there are P sources, the profile is given by

$$\mu_n=\sum_{m=1}^{P}r_m\alpha_{\boldsymbol{z}_m}(x_n,y_n),\tag{67}$$

with r_m specifying the amplitude and $\boldsymbol{z}_m$ specifying the location of the source. The log-likelihood function is given by (40) with μ_n replaced by the new expression in (67):

$$\begin{aligned}l(t,k,r,\boldsymbol{z})&=\sum_{n=1}^{N}\sum_{i=k+1}^{t}\left[y_i^n\sum_{m=1}^{P}r_m\alpha_{\boldsymbol{z}_m}(x_n,y_n)\right]-\frac{1}{2}\left(\sum_{m=1}^{P}r_m\alpha_{\boldsymbol{z}_m}(x_n,y_n)\right)^2\\&=\left[\sum_{m=1}^{P}r_m(t-k)^{1/2}\boldsymbol{\alpha}_{\boldsymbol{z}_m}^\top\boldsymbol{U}_{k,t}\right]-\frac{t-k}{2}\left[\sum_{m=1}^{P}r_m^2\right]-\frac{t-k}{2}\left[\sum_{m\neq p}r_mr_p\boldsymbol{\alpha}_{\boldsymbol{z}_m}^\top\boldsymbol{\alpha}_{\boldsymbol{z}_p}\right].\end{aligned}\tag{68}$$

Setting the derivative of the log-likelihood function with respect to each r_m to zero gives us a

set of linear equations:

$$r_m + \frac{1}{2}\sum_{p\neq m} r_p \boldsymbol{\alpha}_{\boldsymbol{z}_m}^\top \boldsymbol{\alpha}_{\boldsymbol{z}_p} = (t-k)^{-1/2}\boldsymbol{\alpha}_{\boldsymbol{z}_m}^\top \boldsymbol{U}_{k,t}, \quad m = 1, \cdots, P. \tag{69}$$

We can write (69) as:

$$\boldsymbol{A}\boldsymbol{r} = \mathbf{b}_{k,t}, \tag{70}$$

with M equations and M unknowns. The matrix $\boldsymbol{A}$ has ones on the diagonal, and the entry on the mth row and the nth column is given by $\frac{1}{2}\boldsymbol{\alpha}_{\boldsymbol{z}_m}^\top \boldsymbol{\alpha}_{\boldsymbol{z}_n}$. The mth element of the vector $\mathbf{b}_{k,t}$ is given by $(t-k)^{-1/2}\boldsymbol{\alpha}_{\boldsymbol{z}_m}^\top \boldsymbol{U}_{k,t}$. If $\boldsymbol{A}$ is invertible (otherwise the sources cannot be uniquely identified), from this we can solve an estimate for the vector of source amplitudes $\mathbf{r} = \boldsymbol{A}^{-1}\mathbf{b}_{k,t}$ for a given set of source locations. By plugging the estimate $\boldsymbol{r}$ into the likelihood function (40), and maximizing the resulted statistic with respect to k and all possible source locations $(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_M)$ in the set $\mathcal{D}$, we obtain the GLR statistics for multiple overlapping sources.

If we assume the sources are non-overlapping, then $\boldsymbol{\alpha}_{\boldsymbol{z}_m}^\top \boldsymbol{\alpha}_{\boldsymbol{z}_p} = 0$, for $p \neq m$. As a result, (69) becomes (41) for each r_m, and

$$\hat{r}_m = \frac{\boldsymbol{\alpha}_{\boldsymbol{z}_m}^\top \boldsymbol{U}_{k,t}}{(t-k)^{1/2}}, \tag{71}$$

for a candidate source location $\boldsymbol{z}_m$. Using this amplitude estimate leads to the same GLR statistic as the one used in the one-source case in (40). Hence when the sources are non-overlapping, the profile-based detection procedure based on GLR statistic searches one source at a time. Then there are multiple maximum values of the GLR statistics near the location of the sources.