arXiv:1207.2386v1 [math.ST] 10 Jul 2012

Sequential Multi-Sensor Change-Point Detection

Yao Xie,* David O. Siegmund†‡

July 11, 2012

Abstract

We develop a mixture procedure for monitoring parallel streams of data for a change-point that affects only a subset of them, without assuming a spatial structure relating the data streams to one another. Observations are assumed initially to be independent standard normal random variables. After a changepoint the observations in a subset of the streams of data have non-zero mean values. The subset and the post-change means are unknown. The procedure we study uses stream specific generalized likelihood ratio statistics, which are combined to form an overall detection statistic in a mixture model that hypothesizes an assumed fraction p_0 of affected data streams. An analytic expression is obtained for the average run length (ARL) when there is no change and is shown by simulations to be very accurate. Similarly, an approximation for the expected detection delay after a change-point is also obtained. Numerical examples are given to compare the suggested procedure to other procedures for unstructured problems and in one case where the problem is assumed to have a well defined geometric structure. Finally we discuss sensitivity of the procedure to the asssumed value of p_0 and suggest a generalization.

1 Introduction

Single sequence problems of change-point detection have a long history in industrial quality control, where an observed process is assumed initially to be in control and at a change-point becomes out of control. It is desired to detect the change-point with as little delay as possible subject to the

*Yao Xie (Email: yao.c.xie@gmail.com) was with the Department of Electrical Engineering at Stanford University, and it currently with the Department of Electrical and Computer Engineering at Duke 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.

constraint that false detections occurring before the true change-point are very rare. Outstanding early contributions are due to Page [Pag54], [Pag55], Shiryaev [Shi63], and Lorden [Lor71].

We assume there are parallel streams of data subject to change-points. More precisely suppose that for each $n = 1, \ldots, N$, we make observations $y_{n,t}$, $t = 1, 2, \ldots$. The observations are mutually independent within and across data streams. At a certain time κ, there are changes in the distributions of observations made at a subset $\mathcal{N} \subset \{1, \ldots, N\}$ of cardinality $|\mathcal{N}| \leq N$. Also denote by $\mathcal{N}^c$ the set of unaffected data streams. The change-point κ, the subset $\mathcal{N}$ and its size, and the size of the changes are all unknown. As in the case of a single sequence $N = 1$, 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, a surrogate for the frequency of false alarms is the average-run-length (ARL), defined to be the expected time before incorrectly announcing a change of distribution when none has occurred.

It may be convenient to imagine that the data streams represent observations made by a collection of N sensors, that the change-point is the onset of a localized signal that can be detected by sensors in the neighborhood of the signal. However, the problem is *unstructured* in the sense that there is no model that relates the changes seen at the different sensors. In a *structured* problem there exists a profile determining the relative magnitudes of the changes observed by different sensors, say according to their distance from the location of a signal, cf. [SY08]. For discussions of unstructured problems and applications see [TV08], [Mei10], [CGV+10], [PRT03], and [LLR09].

The detection problem of particular interest in this paper involves the case that N is large and $|\mathcal{N}|$ is relatively small. To achieve efficient detection, the detection procedure should use insofar as possible only useful information from affected sensors and suppress noise from the unaffected sensors.

In analogy with the well known CUSUM statistic (e.g., Page [Pag54], [Pag55], Lorden [Lor71]), Mei [Mei10] recently proposed a multisensor procedure based on sums of the CUSUM statistic from individual sensors. He then compares the sum with a suitable threshold to determine a stopping rule. While the distributions of the data, both before and after the change-point, are completely general, they are also assumed to be completely known. The method is shown to minimize asymptotically the expected detection delay for a given false alarm rate, when the threshold value (and hence the constraint imposed by the ARL) becomes infinitely large. The procedure fails to be asymptotically optimal when the specified distributions are incorrect. Tartakovsky and Veeravalli proposed a different 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,

both procedures assume the change-point is observed by all sensors. When only a subset of sensors observe the change-point, these procedures include noise from the unaffected sensors in the detection statistic, which may lead to long detection delays.

In this paper, we develop a *mixture* procedure that achieves good detection performance in the case of an unknown subset of affected sensors and incompletely specified post-change distributions. The key feature of the proposed procedure is that it incorporates an assumption about the fraction of affected sensors when computing the detection statistic. We assume that the individual observations are independent and normally distributed with unit variance, and that the changes occur in their mean values. At the tth vector of observations $(y_{n,t}, n = 1, \ldots, N)$, the mixture procedure first computes a generalized likelihood ratio (GLR) statistic for each individual sensor under the assumption that the change-point occurs at $\kappa \leq t$. Then the local GLR statistics are combined via a mixture model that depends on p_0, the hypothesized fraction of affected sensors. The local statistics are then summed and compared with a detection threshold. To characterize the performance of the mixture procedure, we derive analytic approximations for its ARL and expected detection delay, which are evaluated by comparing the approximations to suitable simulations. Since evaluation of the ARL by simulation is quite time consuming, the analytic approximation to the ARL proves very useful in determining a suitable detection threshold. The proposed procedure is then compared to competing procedures via simulation and is shown to be very competitive in unstructured problems. It is also shown to be reasonably robust to the choice of p_0, and a hierarchical mixture procedure is suggested for cases when a single value of p_0 seems inadequate.

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 remainder of the paper is organized as follows. In Section 2 we establish our notation and formulate the problem more precisely. In Section 3 we review several detections procedures and introduce the proposed 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 performs well compared to other procedures in the unstructured problem. We also compare the mixture procedure to that suggested by [SY08] in a structured problem. Finally Section 7 concludes the paper with some discussion.

2 Assumptions and Formulation

Given N sensors, for each $n = 1, 2, \cdots, N$, the observations from the nth sensor are given by $y_{n,t}$, $t = 1, 2, \cdots$. Assume that different observations are mutually independent and normally distributed with unit variances. Under the hypothesis of no change, they have zero means. 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 $\mathcal{N}$ of $\{1, 2, \ldots, N\}$ with cardinality $|\mathcal{N}|$ such that for $n \in \mathcal{N}$, the observations $y_{n,t}$ at sensors *affected* by the change-point, have means equal to $\mu_n > 0$ for all $t > \kappa$, while observations from the unaffected sensors keep the same standard normal 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 $\mathcal{N}$ and the values of μ_n, although this dependence is suppressed in the notation. The fraction of affected sensors is given by $p = |\mathcal{N}|/N$.

Our goal is to define a stopping rule T such that for all sufficiently large prescribed constants $c > 0$, $\mathbb{E}^\infty\{T\} \geq c$, while asymptotically $\mathbb{E}^\kappa\{T - \kappa | T > \kappa\}$ is a minimum. Ideally, the minimization would hold uniformly in the various unknown parameters: κ, $\mathcal{N}$ 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.

3 Detection Procedures

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

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

We assume that each sensor is affected by the change with probability p_0 (independently from one sensor to the next). The global log likelihood of all N sensors is

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

The expression (2) suggests several change-point detection rules.

One possibility is to set μ_n equal to a nominal change, say $\delta > 0$, which would be important to

detect, and define the stopping rule

$$T_1 = \inf\left\{ t : \max_{0\le k\le t} \sum_{n=1}^{N} \log\{1 - p_0 + p_0 \exp[\ell_n^+(t,k,\delta)] \ge b \right\}, \tag{3}$$

where x^+ denotes the positive part of x. Here thresholding by the positive part plays the role of dimension reduction by limiting the current considerations only to sequences that appear to be affected by the change-point.

Another possibility is to replace μ_n by its maximum likelihood estimator, as follows. The maximum likelihood estimate of the post-change mean as a function of the current number of observations t and putative change-point location k is given by

$$\hat{\mu}_n = \left(\sum_{i=k+1}^{t} y_{n,i} \right)^+ /(t-k). \tag{4}$$

Substitution into (1) gives the log generalized likelihood ratio (GLR) statistic. Putting

$$\begin{aligned} S_{n,t} &= \sum_{i=1}^{t} y_{n,i}, \\ U_{n,k,t} &= (t-k)^{-1/2}(S_{n,t} - S_{n,k}), \end{aligned} \tag{5}$$

we can write the log GLR as

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

We define the stopping rule

$$T_2 = \inf\left\{ t : \max_{0\le k< t} \sum_{n=1}^{N} \log\left(1 - p_0 + p_0 \exp[(U_{n,k,t}^+)^2/2]\right) \ge b \right\}. \tag{7}$$

Remark. In what follows we use a *window limited* version of (7), where the maximum is restricted to $m_0 \le t-k < m_1$ for suitable $m_0 < m_1$. The role of m_1 is two-fold. On the one hand it reduces the memory requirements to implement the stopping rule, and on the other it effectively establishes a minimum level of change that we want to detect. For asymptotic theory given below, we assume that $b \to \infty$, with m_1/b also diverging. More specific guidelines in selecting m_1 are discussed below. In the numerical examples that follow we take $m_0 = 1$. In practice a slightly larger value can be used to provide protection against outliers in the data, although it may delay detection in cases involving very large changes.

The detection rule (7) is motivated by the suggestion of [ZYS11] for a similar fixed sample change-point detection problem.

For the special case $p_0 = 1$, (7) becomes the (global) GLR procedure, which for $N = 1$ was studied by [SV95]. It is expected to be efficient if the change-point affects a large fraction of the sensors. At the other extreme, if only one or a very small number of sensors is affected by the change-point, a reasonable procedure would be

$$T_{\max} = \inf \left\{ t : \max_{0 \le k < t} \max_{1 \le n \le N} (U^+_{n,k,t})^2/2 \ge b \right\}. \tag{8}$$

The stopping rule $T_{\max}(p_0)$ can also be window limited.

Still other possibilities are suggested by the observation that a function of y of the form $\log[1 - p_0 + p_0 \exp(y)]$ large only if y is large, and then this function is approximately equal to $[y + \log(p_0)]^+$. This suggests the stopping rules

$$T_3 = \inf \left\{ t : \max_{0 \le k < t} \sum_{n=1}^{N} [\ell_n(t, k, \delta) + \log(p_0)]^+ \ge b \right\} \tag{9}$$

and

$$T_4 = \inf \left\{ t : \max_{0 \le k < t} \sum_{n=1}^{N} [(U^+_{n,k,t})^2/2 + \log(p_0)]^+ \ge b \right\}, \tag{10}$$

or a suitably window limited version.

Mei ([Mei10]) suggests the stopping rule

$$T_{\text{Mei}} = \inf \left\{ t : \sum_{n=1}^{N} \max_{0 \le k < t} \ell_n(t, k, \delta) \ge b \right\}, \tag{11}$$

which simply adds the classical CUSUM statistics for the different sensors. Note that this procedure fails to take advantage of the assumption that all distributions affected by the change-point change simultaneously. As we shall see below this failure has a negative impact on the efficiency of the procedure, although it might prove beneficial in differently formulated problems. For example, there may be a time delay before the signal is perceived at different sensors, or there may be different signals occurring at different times in the proximity of different sensors. In these problems, Mei's procedure, which allows changes to occur at different times, could be useful.

The procedure suggested by Tartakovsky and Veeravalli [TV08] is defined by the stopping rule

$$T_{\mathrm{TV}} \triangleq \inf\left\{ t : \max_{0\le k<t} \sum_{n=1}^{N} \ell_n(t,k,\delta) \ge b \right\}. \tag{12}$$

This stopping rule resembles $T_3(p_0)$ with $p_0 = 1$, but with one important difference. After a change-point the statistics of the unaffected sensors have negative drifts that tend to cancel the positive drifts from the affected sensors. This can lead to a large expected detection delay. Use of the positive part, $[\ell_n(t,k,\delta)]^+$, in the definitions of our stopping rules is designed to avoid this problem.

Different thresholds b are required for each of these detection procedures to meet the ARL requirement.

4 Properties of the Detection Procedures

In this section we develop theoretical properties of the detection procedures T_1 to T_4, with emphasis on T_2. We use two standard performance metrics: (i) the expected value of the stopping time T when there is no change, usually called the average run length or ARL and (ii) the expected detection delay in the extreme case where a change occurs immediately at $\kappa = 0$. This provides an upper bound on the expected detection delay when a change occurs later in the sequence of observations. The approximation to the ARL will be shown below to be very accurate, which is fortunate since its simulation can be quite time consuming.

4.1 Average run length when there is no change

The ARL is the expected value of the stopping time T when there is no change-point. It will be convenient to use the following notation. Let $g(x,p_0) = \log(1 - p_0 + p_0 \exp[(x^+)^2/2])$, and put

$$\psi(\theta) = \log \mathbb{E}\{\exp[\theta g(U,p_0)]\}, \tag{13}$$

where U has a standard normal distribution. Also let

$$\gamma(\theta) = \frac{1}{2}\theta^2 \mathbb{E}\{[\dot{g}(U,p_0)]^2 \exp[\theta g(U,p_0) - \psi(\theta)]\}. \tag{14}$$

Here the dot denotes differentiation (with respect to the first argument for a function of more than one variable). Let

$$f(N,\theta,p_0)=\frac{\theta[2\pi\ddot{\psi}(\theta)]^{1/2}}{\gamma(\theta)N^{1/2}}\exp\{N[\theta\dot{\psi}(\theta)-\psi(\theta)]\}. \tag{15}$$

Denote the standard normal density function by $\phi(x)$ and its distribution function by $\Phi(x)$. Also let $\nu(x)=2x^{-2}\exp\left[-2\sum_1^\infty n^{-1}\Phi\left(-|x|n^{1/2}/2\right)\right]$ (cf. [Sie85], p. 82). For numerical purposes a simple, accurate approximation is given by (cf. [SY07])

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

To state our approximation to the ARL of T_2, we assume that $N\to\infty$ and $b\to\infty$ with b/N fixed. Consider the window limited mixture stopping rule (7) with $m_1=o(b^r)$ for some positive integer r and define θ by $\dot{\psi}(\theta)=b/N$. Then

$$\mathbb{E}^\infty\{T_2\}\sim f(N,\theta,p_0)/\int_{[2N\gamma(\theta)/m_1]^{1/2}}^{[2N\gamma(\theta)/m_0]^{1/2}}y\nu^2(y)dy. \tag{16}$$

Remark. The integrand in the approximation is integrable at both 0 and ∞ by virtue of the relations $\nu(y)\to1$ as $y\to0$, and $\nu(y)\sim2/y^2$ as $y\to\infty$.

The following calculations provide support for the approximation in (16). For detailed proofs in similar problems see [SV95] or [SY08]. From arguments similar to those used by [ZYS11], we can show that

$$\begin{aligned}&\mathbb{P}^\infty\{T_2\le m\}\\&=\mathbb{P}^\infty\left\{\max_{t\le m,m_0\le t-k\le m_1}\sum_{n=1}^N g(U_{n,k,t};p_0)\ge b\right\}\\&\sim N^2e^{-N[\theta\dot{\psi}(\theta)-\psi(\theta)]}[2\pi N\ddot{\psi}(\theta)]^{-1/2}|\theta|^{-1}\gamma^2(\theta)\int_{m_0/m}^{m_1/m}\nu^2\left([2N\gamma(\theta)/(mt)]^{1/2}\right)(1-t)dt/t^2.\end{aligned} \tag{17}$$

Here it is assumed that m is large, but small enough that the right hand side of (17) converges to 0 when $b\to\infty$. Assume the maximum window size m_1 is small compared to m. Changing variables in the integrand and using the notation (15), we can re-write this approximation as

$$\mathbb{P}^\infty\{T_2\le m\}\sim m\int_{[2N\gamma(\theta)/m_1]^{1/2}}^{[2N\gamma(\theta)/m_0]^{1/2}}y\nu^2(y)dy/f(N,\theta,p_0). \tag{18}$$

From the arguments in [SV95] or [SY08] (see also [Ald88]), we see that T_2 is asymptotically ex-

ponentially distributed. Hence if λ denotes the factor multiplying m on the right hand side of (18), then for still larger m, in the range where $m\lambda$ is bounded away from 0 and ∞, $\mathbb{P}^\infty\{T_2 \leq m\} - [1 - \exp(-\lambda m)] \to 0$. Consequently, uniform integrability implies that $\mathbb{E}^\infty\{T_2\} \sim \lambda^{-1}$, which is equivalent to (16).

A simultation illustrating the fact that T_2 is approximately exponentially distributed is given in Section 4.3.

Remarks. (i) The result we have used from [ZYS11] was motivated by a problem involving changes that could be positive, or negative, or both and in that paper it was assumed that the function $g(x, p_0)$ is twice continuously differentiable in x. The function $g(x, p_0)$ defined above for one-sided changes, which are more natural in our problem, and have been used in the formulations of earlier authors, does not have this much smoothness. We can, however, approximate the indicator of $x > 0$ by $\Phi(rx)$, as $r \to \infty$, and thus use in place of x^+ in the definition of g the smooth function $\int_{-\infty}^{x} \Phi(rv)dv = x\Phi(rx) + r^{-1}\phi(rx)$, which converges uniformly to x^+ as $r \to \infty$. Using the cited result, then letting $r \to \infty$ and interchanging limits produces (17). An alternative approach would be simply to define $g(x, p_0)$ to be appropriate for a one sided change while having the required smoothness in x. A simple example is $g(x, p_0) = \log[1 - p_0 + p_0 \exp(x^2/2)\Phi(x)]$.

(ii) The fact that the stopping times T studied in this paper are asymptotically exponentially distributed can be very useful in simulating the ARL. It is not necessary to simulate the process until T, which can be computationally time consuming, but only until we are able to estimate $\mathbb{P}^\infty\{T \leq m\}$ with a reasonably small percentage error. For the numerical examples given later, we have occasionally used this shortcut with the value of m that makes this probability 0.1 or 0.05.

4.2 Expected Detection Delay

After a change-point occurs, we are interested in the expected number of additional observations required for detection. For the detection rules considered in this paper, the maximum expected detection delay over $\kappa \geq 0$ is attained at $\kappa = 0$. Hence we consider this case.

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 we let U denote a standard normal random variable. Recall that $\mathcal{N}$ denotes the set of sensors at which there is a change, $M = |\mathcal{N}|$ is the cardinality of this set, and $p = |\mathcal{N}|/N$ is the true fraction of sensors that are affected by the change. For each $n \in \mathcal{N}$ the mean value changes from 0 to $\mu_n > 0$, and for $n \in \mathcal{N}^c$ the distribution remains the same as before the change-point.

Let

$$\Delta = \left(\sum_{n \in \mathcal{N}} \mu_n^2 \right)^{1/2}. \tag{19}$$

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 see to a first order approximation that the expected detection delay is $2b/\Delta^2$, provided that the maximum window size, m_1, is large compared to this quantity. In the following derivation we assume $m_1 \gg 2b/\Delta^2$.

In addition, let

$$\tilde{S}_t \triangleq \sum_{i=1}^{t} z_i \tag{20}$$

be a random walk where the increments z_i are independent and identically distributed with mean $\Delta^2/2$ and variance Δ^2. Let $\tau = \min \left\{ t : \tilde{S}_t > 0 \right\}$. Our approximation to the expected detection delay given below depends on two related quantities. The first is

$$\rho(\Delta) = \frac{1}{2} \mathbb{E}\{\tilde{S}_\tau^2\} / \mathbb{E}\{\tilde{S}_\tau\}, \tag{21}$$

for which exact computational expressions and useful approximations are available in [Sie85]. In particular

$$\rho(\Delta) = \mathbb{E}\{z_1^2\}/(2\mathbb{E}\{z_1\}) - \sum_{i=1}^{\infty} i^{-1} \mathbb{E}\{\tilde{S}_i^-\} = \Delta^2/4 + 1 - \sum_{i=1}^{\infty} i^{-1} \mathbb{E}\{\tilde{S}_i^-\}, \tag{22}$$

where $(x)^- = -\min\{x, 0\}$. The second quantity is $\mathbb{E}\{\min_{t \geq 0} \tilde{S}_t\}$, which according to (Problem 8.14 in [Sie85]) is given by

$$\mathbb{E}\left\{ \min_{t \geq 0} \tilde{S}_t \right\} = \rho(\Delta) - 1 - \Delta^2/4. \tag{23}$$

The following approximation refines the first order result for the expected detection delay. As $b \to \infty$, with other parameters held fixed,

$$\mathbb{E}^0\{T_2\} = 2\Delta^{-2} \left[b + \rho(\Delta) - |\mathcal{N}| \log p_0 - |\mathcal{N}|/2 + \mathbb{E}\left\{ \min_{t \geq 0} \tilde{S}_t \right\} - (N - |\mathcal{N}|)\mathbb{E}\{g(U, p_0)\} + o(1) \right]. \tag{24}$$

The following calculation provides the ingredients for a proof of (24). For details in similar problems involving a single sequence, see [PS75] and [SV95]. For convenience we write $T = T_2$. Let $k_0 = b^{1/2}$. For $k < T - k_0$, we can write the detection statistic at the stopping time T as follows,

up to a term that tends to zero exponentially fast in probability:

$$
\begin{aligned}
Z_{k,t} =& \sum_{n=1}^{N} g(U_{n,k,T}, p_0) \\
=& \sum_{n\in\mathcal{N}} g(U_{n,k,T}, p_0) + \sum_{n\in\mathcal{N}^c} g(U_{n,k,T}, p_0) \\
=& \sum_{n\in\mathcal{N}} \log\left(p_0 \exp\left\{(U^+_{n,k,T})^2/2\right\}\left[1 + \frac{1-p_0}{p_0}\exp\left\{-(U^+_{n,k,T})^2/2\right\}\right]\right) + \sum_{n\in\mathcal{N}^c} g(U_{n,k,T}, p_0) \\
=& \sum_{n\in\mathcal{N}} \left[\log p_0 + (U^+_{n,k,T})^2/2\right] + \sum_{n\in\mathcal{N}^c} g(U_{n,k,T}, p_0) + \sum_{n\in\mathcal{N}} \log\left(1 + \frac{1-p_0}{p_0}\exp\left\{-(U^+_{n,k,T})^2/2\right\}\right) \\
=& |\mathcal{N}|\log p_0 + \sum_{n\in\mathcal{N}} (U^+_{n,k,T})^2/2 + \sum_{n\in\mathcal{N}^c} g(U_{n,k,T}, p_0) + o(1) \\
=& |\mathcal{N}|\log p_0 + \sum_{n\in\mathcal{N}} [(S_{n,T} - S_{n,k})^+]^2/2(T-k) + \sum_{n\in\mathcal{N}^c} g(U_{n,k,T}, p_0) + o(1).
\end{aligned}
\tag{25}
$$

The residual term $\sum_{n\in\mathcal{N}_a} \log\left(1 + (1-p_0)\exp\left\{-(U^+_{n,k,T})^2/2\right\}/p_0\right)$ tends to zero exponentially fast when $b \to \infty$, because when $b \to \infty$, $T \to b/\Delta$, and $n \in \mathcal{N}$, $(U^+_{n,k,T})^2$ grows on the order of $\mu_n^2(T-k) > \mu_n^2 k_0 = \mu_n^2\sqrt{b}$.

We then use the following simple identity to decompose the second term in (25) for the affected sensors into two parts:

$$
\begin{aligned}
(S^+_{n,t})^2/2t &= S^2_{n,t}/2t - (S^-_{n,t})^2/2t \\
&= \mu_n(S_{n,t} - \mu_n t/2) + (S_{n,t} - \mu_n t)^2/2t - (S^-_{n,t})^2/2t.
\end{aligned}
\tag{26}
$$

From the preceding discussion, we see that $\max_{0\le k<T-k_0} Z_{k,T}$ is on the order of b, while $\max_{T-k_0\le k<T} Z_{k,T}$ is on the order of $k_0 = b^{1/2}$. Hence with overwhelming probability the max over all k is attained for

$k < T - k_0$, so from (26) and (25) we have

$$
\begin{aligned}
&\max_{0\le k<T} Z_{k,t} \\
=&\max_{0\le k<T-k_0} \sum_{n=1}^{N} g(U_{n,k,T}, p_0) + o(1) \\
=&|\mathcal{N}| \log p_0 + \max_{0\le k<T-k_0} \left[\sum_{n\in\mathcal{N}} \mu_n[(S_{n,T} - S_{n,k}) - (T-k)\mu_n/2] + \sum_{n\in\mathcal{N}} [(S_{n,T} - S_{n,k}) - (T-k)\mu_n]^2/[2(T-k)] \right. \\
&\left. -[(S_{n,T} - S_{n,k})^-]^2/2(T-k) + \sum_{n\in\mathcal{N}^c} g(U_{n,k,T}, p_0) \right] + o(1) \\
=&|\mathcal{N}| \log p_0 + \sum_{n\in\mathcal{N}} \mu_n \left(S_{n,T} - T\mu_n/2\right) + \\
&\max_{0\le k<T-k_0} \left[-\sum_{n\in\mathcal{N}} \mu_n \left(S_{n,k} - k\mu_n/2\right) + \sum_{n\in\mathcal{N}} [(S_{n,T} - S_{n,k}) - (T-k)\mu_n]^2/[2(T-k)] \right. \\
&\left. - \sum_{n\in\mathcal{N}} [(S_{n,T} - S_{n,k})^-]^2/[2(T-k)] + \sum_{n\in\mathcal{N}^c} g\left(U_{n,k,T}, p_0\right) \right] + o(1).
\end{aligned}
\tag{27}
$$

The following lemma forms the basis for the rest of the derivation. The proof is omitted here; for details see ([Xie11]) or ([SV95]) for the special case $N = 1$.

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

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

where $o_p(1)$ converges to 0 in probability.

By taking expectations in (27), letting $b \to \infty$ and using Lemma 1, we have

$$\begin{aligned}
&\mathbb{E}^0\left\{\max_{0\le k<T}\sum_{n=1}^{N} g(U_{n,k,T},p_0)\right\}\\
=&\mathbb{E}^0\left\{|\mathcal{N}|\log p_0+\sum_{n\in\mathcal{N}}\mu_n(S_{n,T}-T\mu_n/2)+\sum_{n\in\mathcal{N}}\frac{(S_{n,T}-T\mu_n)^2}{2T}+\sum_{n\in\mathcal{N}^c} g(U_{n,0,T},p_0)+\right. \qquad (28)\\
&\left.\max_{0\le k<k_0}\left[-\sum_{n\in\mathcal{N}}\mu_n(S_{n,k}-k\mu_n/2)\right]\right\}+o(1).
\end{aligned}$$

We will compute each term on the right hand side of (28) separately. We will need the lemma due to Anscombe and Doeblin (see Theorem 2.40 in [Sie85]), which states that the standardized randomly stopped sum of random variables are asymptotically normally distributed under quite general conditions.

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

$$\mathbb{E}^0\left\{\sum_{n\in\mathcal{N}}\mu_n(S_{n,T}-T\mu_n/2)\right\}=\mathbb{E}^0\{T\}\Delta^2/2. \qquad (29)$$

(ii) By the Anscombe-Doeblin Lemma, $(S_{n,T}-T\mu_n)/T^{1/2}$ is asymptotically normally distributed with zero mean and unit variance. Hence $\sum_{n\in\mathcal{N}}(S_{n,T}-T\mu_n)^2/T$ is asymptotically a sum of independent χ^2_1 random variables, so

$$\mathbb{E}^0\left\{\sum_{n\in\mathcal{N}}(S_{n,T}-T\mu_n)^2/2T\right\}=|\mathcal{N}|/2+o(1). \qquad (30)$$

(iii) Similarly,

$$\mathbb{E}^0\left\{\sum_{n\in\mathcal{N}^c} g(U_{n,0,T},p_0)\right\}\to(N-|\mathcal{N}|)\mathbb{E}^0\{g(U,p_0)\}. \qquad (31)$$

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

Having evaluated the right hand of (28), we now consider the left-hand side. The first order asymptotic behavior of the process $\sum_{n=1}^{N} g(U_{n,k,T},p_0)$ is the same as that of $\sum_{n\in\mathcal{N}}\mu_n(S_{n,T}-T\mu_n/2)$,

which has drift $\Delta^2/2$ and variance Δ^2. By writing

$$\mathbb{E}^0\left\{\max_{0\le k<T}\sum_{n=1}^{N} g(U_{n,k,T},p_0)\right\} = b + \mathbb{E}^0\left\{\max_{0\le k<T}\sum_{n=1}^{N} g(U_{n,k,T},p_0) - b\right\}, \tag{32}$$

and using nonlinear renewal theory to evaluate the expected overshoot of the process of (20) over the boundary ([Sie85], Chapter IX), we obtain

$$\mathbb{E}^0\left\{\max_{0\le k<T}\sum_{n=1}^{N} g(U_{n,k,T},p_0) - b\right\} \to \rho(\Delta). \tag{33}$$

Remark. An examination of the preceding argument indicates that the accuracy in practice of the approximation may be quite variable. In particular, variability in the μ_n will be reflected in variability in the accuracy of the expansion of $g(U_{n,k,t},p_0)$. In particular if some of the μ_n are close to 0, there will not be the clear separation required by the derivation of the indices n into terms where the approximation applies and terms where it does not, with the result that the approximation may break down.

4.3 Accuracy of the Approximations

We start with examining the accuracy of our approximations for the ARL and the expected detection delay in (16) and (24). For a Monte Carlo experiment we use $N = 100$ sensors, $m_1 = 200$ and $\mu_n = 1$ for all affected data streams. The comparisons between the theoretical and Monte Carlo ARLs for different values of p_0 are given in Table 1. Numerical results obtained from 500 Monte Carlo trials are given in Table 1, where they show that the approximation in (16) is quite accurate.

Table 1: Average run length (ARL) of $T_2(p_0)$, $m_1 = 200$.

p_0	b	Theory	Monte Carlo
0.3	31.2	5001	5504
0.3	32.3	10002	10221
0.1	19.5	5000	4968
0.1	20.4	10001	10093
0.03	12.7	5001	4830
0.03	13.5	10001	9948

Figure 1 illustrates the fact that (for example) $T_2(0.1)$ is approximately exponentially distributed.

Results for the expected detection delay obtained from 500 Monte Carlo trialsare given in Table 2. The approximation for the expected detection delay does not appear to be as accurate as the

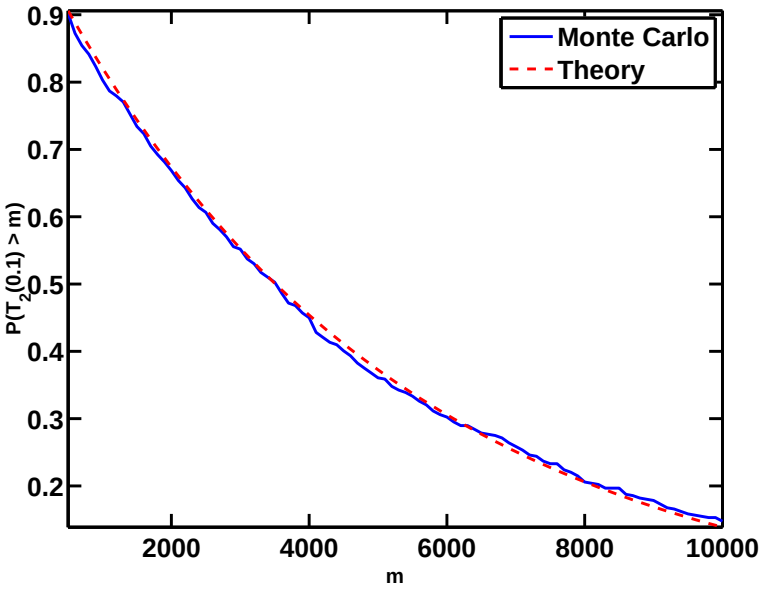

Figure 1: The tail probability $\mathbb{P}\{T_2(0.1) > m\}$. Theoretic values are calculated from (16).

approximation for the ARL. Since the expected detection delay requires considerably less computational effort to simulate and needs to be known only roughly when we choose design parameters for a particular problem, we are less concerned about its accuracy.

Table 2: Expected detection delay of the mixture procedure with ARL ≈ 5000, $\mu = 1$, and $m_1 = 200$.

p	p_0	b	Theory	Monte Carlo
0.3	0.3	31.2	3.5	3.2
0.1	0.3	31.2	6.2	6.5
0.3	0.1	19.5	5.2	3.6
0.1	0.1	19.5	7.2	6.7
0.03	0.1	19.5	13.9	14.3
0.03	0.03	12.7	13.9	14.2

We have performed considerably more extensive simulations that yield results consistent with the small experiments reported in Tables 1 and 2. Since the parameter p_0 defining T_2 must be chosen subjectively, it is interesting to observe that Table 2 suggests this procedure is reasonably robust with respect to the choice of p_0, and choosing p_0 somewhat too large is less costly than choosing p_0 too small. More extensive calculations bear out this observation. We return to the problem of choosing p_0 in Section 5.1.

5 Numerical Comparisons

In this section, we compare the expected detection delays for several procedures when their ARLs are all approximately 5000. The thresholds are given in Table 3, where we assume $N = 100$, and $m_1 = 200$ for those procedures for which a limited window size is appropriate. The procedure (7) is denoted by $T_2(p_0)$. For Mei's procedure we put $\delta_n = 1$. The procedures (9) are denoted by $T_3(p_0, \delta)$. Recall that $T_2(1)$ uses the generalized likelihood ratio statistic and $T_3(1, \delta)$ is similar to the procedure proposed by Tartakovsky and Veeravalli (2008), but we have inserted the positive part to avoid the problems discussed above with this procedure when a relatively small number of sensors are affected by the changepoint. The expected detection delays are obtained from 500 Monte Carlo trials and are listed in Table 4. For some entries values from our asymptotic approximation are given in parentheses.

Note that the max procedure (8) has the smallest detection delay when $p = 0.01$, but it has the largest delay for p greater than 0.1. The procedures defined by T_2 and by T_3 are comparable. Mei's procedure performs well when p is large, but poorly when p is small.

Table 3: Thresholds for ARL ≈ 5000, $m_1 = 200$.

Procedure	b	Monte Carlo ARL
Max	12.8	5041
$T_2(1)$	53.5	4978
$T_2(0.1)$	19.5	5000
Mei	88.5	4997
$T_3(.1, 1)$	12.4	4948
$T_3(1, 1)$	41.6	4993

Table 4: Expected Detection Delays with $N = 100$ obtained from 500 Monte Carlo trials. Thresholds for ARL 5000 are listed in Table 3. Theoretical approximations for the expected detection delay are in parentheses.

p	Method	DD, $\mu = 1$	DD, $\mu = 0.7$	DD, $\mu = 1.3$
0.01	max	25.5	49.6	16.3
	$T_2(1)$	52.3(56.9)	105.5(114.6)	32.9(34.1)
	$T_2(.1)$	31.6(32.5)	59.4(64.9)	20.3 (19.7)
	Mei	53.2	103.8	38.1
	$T_3(.1, 1)$	29.1	63.3	19.1
	$T_3(1, 1)$	82.0	213.7	53.3
0.03	max	18.1	33.3	11.6
	$T_2(1)$	18.7(19.3)	35.8(38.4)	12.6(11.7)
	$T_2(.1)$	14.2(13.9)	26.7(27.5)	9.3(8.5)
	Mei	23.0	41.6	16.4
	$T_3(.1)$	13.4	26.9	9.2
	$T_3(1)$	27.2	66.0	16.3
0.05	max	15.5	28.4	9.7
	$T_2(1)$	12.2(11.6)	21.8(23.0)	7.9(7.1)
	$T_2(.1)$	10.4(10.1)	18.9(19.9)	6.9(6.2)
	Mei	15.7	26.9	11.4
	$T_3(.1, 1)$	9.8	18.6	7.0
	$T_3(1, 1)$	15.5	38.8	9.0
0.1	max	12.6	23.0	8.4
	$T_2(1)$	6.7(5.9)	11.8(11.3)	4.7(3.7)
	$T_2(.1)$	6.7(7.2)	11.6(14.1)	4.6(4.5)
	Mei	9.6	15.4	7.4
	$T_3(.1, 1)$	7.1	11.9	5.3
	$T_3(1, 1)$	6.8	15.7	4.6
0.3	max	9.6	16.7	6.6
	$T_2(1)$	3.0(2.0)	4.4(3.5)	2.4(1.4)
	$T_2(.1)$	3.5(5.2)	5.6(10.1)	2.7(3.3)
	Mei	4.9	7.0	4.0
	$T_3(.1, 1)$	4.6	6.7	3.9
	$T_3(1, 1)$	3.0	4.3	2.5
0.5	max	8.6	14.4	5.8
	$T_2(1)$	2.3	3.0	2.0
	$T_2(.1)$	2.8	4.0	2.1
	Mei	3.8	5.0	3.0
	$T_3(.1, 1)$	4.0	5.4	3.3
	$T_3(1, 1)$	2.3	3.0	2.0
1	max	7.2	12.1	5.1
	$T_2(1)$	2.0	2.0	2.0
	$T_2(.1)$	2.0	2.6	2.0
	Mei	3.0	3.4	2.3
	$T_3(.1, 1)$	3.4	4.3	3.0
	$T_3(1, 1)$	2.0	2.1	2.0

5.1 Parallel Mixture Procedure

The procedures considered above depend on a parameter p_0, which presumably should be chosen to be close to the unknown true fraction p. While Table 4 suggests that the value of p_0 is fairly robust, to achieve robustness over a wider range of the unknown parameter p, we consider a parallel procedure that combines several procedures, each using a different p_0 to monitor a different range of p values. The thresholds of these individual procedures will be chosen so that they have the same ARL. For example, we can use two different values of p_0, say a small $p_0 = p_1$ and a large $p_0 = p_2$, and then choose thresholds b_1 and b_2 to obtain the same ARL for these two procedures. The parallel procedure claims a detection if at least one of the component procedures reaches its threshold, specifically

$$T_{\text{parallel}} \triangleq \min\{T_2(p_1), T_2(p_2)\}. \tag{34}$$

To compare the performance of the parallel procedure with that of a single T_2, we consider a case with $N = 400$ and $m_1 = 200$. For the single mixture procedure we use the intermediate value $p_0 = 0.10$ and threshold value $b = 44.7$, so $\mathbb{P}^{\infty}\{T_2 \leq 1000\} \approx 0.10$ and hence the ARL ≈ 10000. For the parallel procedure we consider the values $p_1 = 0.02$ and $p_2 = 0.33$. For the threshold values $b_1 = 21.2$ and $b_2 = 87.7$, respectively, we have $\mathbb{P}^{\infty}\{T_2(p_i) \leq 1000\} \approx 0.05$. By the Bonferroni inequality $\mathbb{P}^{\infty}\{\min[T_2(p_1), T_2(p_2)] \leq 1000\} \leq 0.1$, so conservatively $\mathbb{E}^{\infty}\{T_{\text{parallel}}\} \geq 10000$. Table 5 shows that the expected detection delays of the parallel procedure are usually smaller than those of the single procedure, particularly for very small or very large p. Presumably these differences are magnified in "larger" problems.

Table 5: Comparison of Parallel and Simple Procedures

Procedure	p	μ	Expected Detection Delay
$T_2(.1)$	0.1	0.7	6.5
	0.005	1.0	27.1
	0.005	0.7	54.5
	0.25	0.3	12.0
Parallel Procedure	0.1	0.7	6.4
	0.005	1.0	22.9
	0.005	0.7	45.8
	0.25	0.3	10.5

Simulations indicate that because of dependence between the two statistics used to define the parallel procedure, the ARL is actually somewhat larger than the Bonferroni approximation suggested. Since the parallel procedure becomes increasingly attractive in larger problems, which

provide more room for improvement over a single choice of p_0, but which are also increasingly difficult to simulate, it would be interesting to develop a more accurate theoretical approximation for the ARL.

An attractive alternative to the parallel procedure would be to use a weighted linear combination for different values of p_0 of the statistics used to define T_2 or T_3. Our approximation for the ARL can be easily adapted, but some modest numerical exploration suggests that the expected detection delay is not improved as much as for the parallel procedure.

6 Profile-based Procedure for Structured Problems

Up to now we have assumed there is no spatial structure relating the change-point amplitudes at difference sensors. In this section we will consider briefly a *structured problem*, where there is a parameterized profile of the amplitude of the signal seen at each sensor that is based on the distance of the sensor to the source of the signal. Assuming we have some knowledge about such a profile, we can incorporate this knowledge into the definition of an appropriate detection statistic. Our developments follow closely the analysis in ([SY08]).

Assume the location of the nth sensor given by its coordinates x_n, $n = 1, \cdots, N$ at points in Euclidean space, which for simplicity we take to be on an equi-spaced grid. We assume that the source is located in a region $\mathcal{D}$, which is a subset of the ambient Euclidean space. In our example below we consider two dimensional space, but three dimensions would also be quite reasonable. Assume the change-point amplitude at the nth sensor is determined by the expression

$$\mu_n = \sum_{m=1}^{M} r_m \alpha_{z_m}(x_n), \tag{35}$$

where M is the number of sources, $z_m \in \mathcal{D}$ is the (unknown) spatial location of the mth source, $\alpha_z(x)$ is the profile function, and the scalar r_m is an unknown parameter that measures the strength of the mth signal. The profile function describes how the signal strength of the mth point source has decayed at the nth sensor. We assume some knowledge about this profile function is available. For example, $\alpha_z(x)$ is often taken to be a decreasing function of the Euclidean distance between z and x. The profile may also depend on finitely many parameters, such as the rate of decay of the function. See [Rab94] or [SSSW03] for examples in a fixed sample context.

If the parameters r_m are multiplied by a positive constant and the profile $\alpha_{z_m}(x_n)$ divided by the same constant, the values of μ_n do not change. To avoid this lack of identifiability, it is convenient

to assume that for all z the profiles have been standardized to have unit Euclidean norm, i.e., $\sum_x \alpha_z^2(x) = 1$ for all z.

6.1 Profile-based procedure

Under the assumption that there is at most one source, say at z, for observations up to time t with a change-point assumed to equal k, the log likelihood function for observations from all sensors (1) is

$$\ell(t, k, r, z) = \sum_{n=1}^{N} [r\alpha_z(x_n)(S_{n,t} - S_{n,k}) - r^2(t-k)\alpha_z^2(x_n)/2]. \tag{36}$$

When maximized with respect to r this becomes

$$\frac{1}{2}\left[\{\sum_n \alpha_z(x_n)U_{n,k,t}\}^+\right]^2. \tag{37}$$

Maximizing the function (37) with respect to the putative change-point k and the source location z, we obtain the log GLR statistic and a profile-based stopping rule of the form

$$T_{\text{profile}} = \inf\left\{t : \max_{t-m_1 \le k \le t-m_0} \max_{z\in\mathcal{D}} \left[\{\sum_n \alpha_z(x_n)U_{n,k,t}\}^+\right]^2 \ge b\right\}. \tag{38}$$

If the model is correct, (38) is a matched-filter type of statistic.

6.2 Theoretical ARL of profile-based procedure

Using the result presented in [SY08], we can derive an approximation for the ARL of the profile-based procedure. We consider in detail a special case where $d = 2$ and the profile is given by a Gaussian function

$$\alpha_z(x) = \frac{1}{\sqrt{2\pi\beta}} e^{-\frac{1}{4\beta}||x-z||^2}, \qquad x \in \mathbb{R}^2, \beta > 0. \tag{39}$$

The parameter $\beta > 0$ controls of rate of profile decay and is assumed known. With minor modifications one could also maximize with respect to a range of values of β.

Although the sensors have been assumed to be located on the integer lattice of two-dimensional Euclidean space, it will be convenient as a very rough approximation to assume that summation over sensor locations x can be approximated by integration over the entire Euclidean space. With this approximation, $\sum_x \alpha_z^2(x)$, which we have assumed equals 1 for all z, becomes $\int_{\mathbb{R}^2} \alpha_z^2(x)dx$, which by (39) is readily seen to be identically 1. The approximation is reasonable if β is large, so

the effective distance between points of the grid is small, and the space $\mathcal{D}$, assumed to contain the signal, is well within the set of sensor locations (so edge effects can be ignored and the integration extended over all of $\mathbb{R}^2$).

It will be convenient to use the notation

$$\langle f, g \rangle = \int_{\mathbb{R}^2} f(x)g(x)dx. \tag{40}$$

Let $\dot{\alpha}_z$ denote the gradient of α_z with respect to z. Then according to [SY08],

$$\mathbb{P}^\infty\{T_{\text{profile}} \leq m\} \sim m \exp(-b/2)(b/4\pi)^{3/2}2^{1/2} \int_{(b/m_1)^{1/2}}^{(b/m_0)^{1/2}} u\nu^2(u)du \int_{\mathcal{D}} [\det|\langle \dot{\alpha}_z, \dot{\alpha}_z^\top \rangle|^{1/2}dz, \tag{41}$$

To evaluate the last integral in (41), we see from (39) that $\dot{\alpha}_z$ satisfies

$$\dot{\alpha}_z(x) = \alpha_z(x)(x - z)/(2\beta). \tag{42}$$

Hence by (40) $\langle \dot{\alpha}_z, \dot{\alpha}_z^\top \rangle$ is a 2×2 matrix of integrals, which can be easily evaluated, and its determinant equals $1/(16\beta^4)$. Hence the last integral in (41) equals $|\mathcal{D}|/(4\beta^2)$ where $|\mathcal{D}|$ denotes the area of $\mathcal{D}$. Arguing as above from the asymptotic exponentiality of T_{profile}, we find that an asymptotic approximation for the average run length is given by

$$\mathbb{E}^\infty\{T_{\text{profile}}\} \sim \exp(b/2) \left(\frac{b}{4\pi}\right)^{-3/2} \cdot \left[\int_{(b/m_1)^{1/2}}^{(b/m_0)^{1/2}} u\nu^2(u)du \cdot |\mathcal{D}|/(2^{1/2}\beta^2)\right]^{-1}. \tag{43}$$

6.3 Numerical examples

In this section we briefly compare the unstructured detection procedure based on T_2 with the profile procedure in the special case that the assumed profile is correct.

Assume that the profile is given by the Gaussian function (39) with parameter $\beta = 1$ and both procedures are window-truncated with $m_0 = 1$, $m_1 = 100$. The number of sensors is $N = 625$ distributed over a 25×25 square grid with center at the origin. In this situation, approximately $p = 0.016$ sensors are affected. In the specification of T_2 we take $p_0 = 0.05$.

The thresholds are chosen so that the average run lengths when there is no changepoint are approximately 5000. Using (41), we obtain $P_\infty\{T_{\text{profile}} \leq 250\} = 0.050$ for $b = 29.5$. From 500 Monte Carlo trials we obtained the threshold 26.3, so the theoretical approximation appears to be slightly conservative.

To deal with a failure to know the true rate of decay of the signal with distance, we could maximize over β, say for $\beta \in [0.5, 5]$. A suitable version of (41) indicates the threshold would be 33.8. This slight increase to the threshold suggests that failure to know the appropriate rate of decay of the signal with distance leads to a relatively moderate loss of detection efficiency.

For comparisons of the expected detection delay, we used for the profile procedure the threshold 26.3, given by simulation, while for $T_2(0.05)$ we used the analytic approximation, which our studies have shown to be very accurate. Table 6 compares the expected detection delay of the profile based procedure with that of the mixture procedure. As one would expect from the precise modeling assumptions, the profile procedure is substantially more powerful.

In many cases there will be little scientific basis for the assumed profile, so it would also be interesting to compare the structured and the unstructured problems when the assumed profile differs moderately or substantially from the true profile, perhaps in the number of sources of the signals, their shape, the rate of decay, or the locations of the sensors.

Table 6: Comparison of Expected Detection Delays (EDD)

	b	EDD $r = 1$	EDD, $r = 1.5$
Profile-based procedure	26.3	25.65±10.3	12.33±4.4
Unstructured procedure	39.7	78.26±30.9	35.79±11.1

7 Discussion

For an unstructured multi-sensor change-point problem we have suggested and compared a number of sequential detection procedures. We assume that the pre- and post- change samples are normal distributed with known variance and that both the post-change mean and the set of affected sensors are unknown. For performance analysis, we have derived approximations for the average run length (ARL) and the expected detection delay, and have shown that these approximations have reasonable accuracy. Our principal procedure depends on an assumed fraction of sensors affected by the change-point. We show numerically that the procedures are fairly robust with respect to discrepancies between the actual and the hypothesized fractions, and we suggest a parallel procedure based on two or more hypothesized fractions to increase this robustness.

In a structured problem, we have shown that knowledge of the correct structure can be implemented to achieve large improvements in the expected detection delay. An open question is the extent to which failure to hypothesize the appropriate structure compromises these improvements.

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.

[Lor71] G. Lorden. Procedures for reacting to a change in distribution. *Annuals of Mathematical Statistics*, 42:1897–1908, Dec. 1971.

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

[Pag54] E. S. Page. Continuous inspection schemes. *Biometrica*, 41(1/2):100–115, June 1954.

[Pag55] E. S. Page. A test for a change in a parameter occuring at an unknown point. *Biometrika*, 42:523–527, 1955.

[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.

[Shi63] W. A. Shiryaev. On optimal methods in quickest detection problems. *Theory Prob. Appl.*, 8:22 – 46, Jan. 1963.

[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.

[Xie11] Yao Xie. *Statistical Signal Detection with Multi-Sensor and Sparsity*. Ph. D. Thesis, Stanford University, 2011.

[ZYS11] N. Zhang, B. Yakir, and D. O. Siegmund. Detecting simultaneous variant intervals in aligned sequences. *Ann. Appl. Statis.*, 5:645–668, 2011.