

# Bayesian Methods for Multiple Change-Point Detection with Reduced Communication

Eyal Nitzan, *Member, IEEE*, Topi Halme, *Student Member, IEEE*, and Visa Koivunen, *Fellow, IEEE*

**Abstract**—In many modern applications, large-scale sensor networks are used to perform statistical inference tasks. In this paper, we propose Bayesian methods for multiple change-point detection using a sensor network in which a fusion center (FC) can receive a data stream from each sensor. Due to communication limitations, the FC monitors only a subset of the sensors at each time slot. Since the number of change points can be high, we adopt the false discovery rate (FDR) criterion for controlling the rate of false alarms, while minimizing the average detection delay (ADD). We propose two Bayesian detection procedures that handle the communication limitations by monitoring the subset of the sensors with the highest posterior probabilities of change points having occurred. This monitoring policy aims to minimize the delay between the occurrence of each change point and its declaration using the corresponding posterior probabilities. One of the proposed procedures is more conservative than the second one in terms of having lower FDR at the expense of higher ADD. It is analytically shown that both procedures control the FDR under a specified tolerated level and are also scalable in the sense that they attain an ADD that does not increase asymptotically with the number of sensors. In addition, it is demonstrated that the proposed detection procedures are useful for trading off between reduced ADD and reduced average number of observations drawn until discovery. Numerical simulations are conducted for validating the analytical results and for demonstrating the properties of the proposed procedures.

**Index Terms**—Sensor networks, Bayesian multiple change-point detection, communication limitations, average detection delay, false discovery rate

## I. INTRODUCTION

Large-scale sensor networks are prominent new tools in various applications, e.g. Internet of Things (IoT), cyber-physical systems such as power grids, environmental monitoring, and wireless communication. These sensor networks can be used to perform statistical inference tasks [1]–[4]. An important statistical inference problem is sequential change-point detection [5]–[13] in which one is interested in detecting a rapid change in the underlying probability model, anomaly or adversarial activity as quickly as possible subject to a false positive constraint. Sensor networks, where each sensor observes a different data stream and communicates with a fusion center (FC) or cloud, can be deployed to detect multiple change points in a monitored environment.

This work was partly supported by the Academy of Finland projects: (1) Statistical Signal Processing Theory and Computational Methods for Large Scale Data Analysis (2) Wi-FiUS project: Secure Inference in the Internet of Things.

E. Nitzan, T. Halme, and Visa Koivunen are with the Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland, e-mail: eyal.nitzan@aalto.fi, topi.halme@aalto.fi, and visa.koivunen@aalto.fi.

Multiple change-point detection is closely related to multiple hypothesis testing. A widely-used performance criterion in multiple hypothesis testing is the false discovery rate (FDR), where the FDR is the expected proportion of the number of false discoveries among all discoveries [14]–[16]. FDR control for multiple change-point detection has been considered in [3] and [4], [17] in the deterministic and Bayesian frameworks, respectively. These works assumed that *all* data streams are observed in parallel, which may not be feasible in large-scale sensor networks used in IoT. In the context of change-point detection, a Type I error (false positive) occurs if the detection procedure declares a change before the true change actually happens. In general, one would be interested in detecting the change point with minimum possible delay, while controlling the Type I error rate [6], [12]. In the Bayesian framework, the posterior probability of a change point having occurred, or some variation of it, is a commonly used test statistic [11], [12].

Several works have considered discrete time single change-point detection in which only a part of the observations is available. In [18], Bayesian change-point detection was considered by monitoring only a minimal number of sensors at each time slot, where the change detection problem was modeled as a Markov decision process. A Bayesian method to minimize the average detection delay (ADD) subject to constraints on both the probability of false alarm and the observation cost was proposed in [19], where an on-off observation control policy was selected along with the stopping time at which the change is declared. Deterministic versions of this work were developed in [20]–[22] under different settings. Deterministic change-point detection for high-dimensional data with missing elements was considered in [23]. In [24] and [25], quickest change detection problems with sampling right constraints were considered in the deterministic and the Bayesian frameworks, respectively. Quickest deterministic change-point detection with observation scheduling was considered in [26], where the decision maker chooses one of two different sequences of observations at each time slot. In [27], deterministic change-point detection in sensor network with communication rate constraints was studied and adaptive censoring strategies were developed for the sensors. Quickest deterministic change-point detection over multiple data streams was considered in [28], where the observer can only observe one data stream at each time slot.

In this paper, we consider the problem of rapidly detecting change points in multiple data streams [3], [4]. In particular, an FC receives statistically independent data streams from multiple sensors in a large-scale sensor network. Due to com-

munication limitations, at a given time slot the FC monitors only a subset of the active data streams for which change points have not been declared yet. The subset size has a fixed proportion with respect to (w.r.t.) the number of active data streams. We assume that each data stream has an associated random change point.

The contributions of this paper are:

- 1) A Bayesian sequential procedure, named the sequential maximum a-posteriori probability (S-MAP) procedure, is proposed. This procedure detects the change points in all of the data streams, while controlling the FDR. The proposed procedure is based on sequentially updating the sensors' posterior probabilities of change points having occurred. Then, at each time slot we choose to monitor a subset of the sensors with the highest posterior probabilities within the allowed proportion. This approach aims to minimize the time between change point occurrence and its declaration by monitoring the sensors for which change-point occurrence is most probable given the data. The S-MAP procedure uses the same Type I error constraints as in [4] and extends this work to communication constrained scenarios. The FDR control of the S-MAP procedure is established using analytical tools.
- 2) We develop an improved S-MAP (IS-MAP) procedure that is less conservative than the S-MAP procedure in the sense that it has a lower ADD but higher FDR than the S-MAP procedure. The decrease of the ADD is obtained by reducing the detection threshold values of the IS-MAP procedure compared to the S-MAP procedure. It is proved analytically that the FDR of the IS-MAP procedure is still controlled under the desired level despite its lower detection threshold values.
- 3) The asymptotic ADD behavior of the S-MAP and the IS-MAP procedures is established analytically for geometric prior distribution of the change points. It is shown that for any proportion value, both detection procedures are scalable in the sense that their asymptotic ADD does not increase with the number of data streams. In addition, the asymptotic ADD improvement that is obtained by using the IS-MAP procedure in comparison to the S-MAP procedure is characterized quantitatively.
- 4) We conduct simulations in order to evaluate the performance and to verify the established theoretical properties of the S-MAP and the IS-MAP procedures.
- 5) The S-MAP and the IS-MAP procedures are used for investigating the tradeoff between reducing the ADD and reducing the average number of observations (ANO) drawn until change points are declared. The proposed analysis can be useful for developing distributed statistical inference procedures using large-scale sensor networks in limited communication capability scenarios.

Preliminary results of this paper appear in our conference paper [29] and a deterministic version of the proposed detection methods appears in [30]. The remaining of this paper is organized as follows. In Section II, we formulate the Bayesian multiple change-point detection problem. The

S-MAP and the IS-MAP procedures are derived in Sections III and IV, respectively, and their FDR control property is proved. Asymptotic ADD analysis of the S-MAP and the IS-MAP procedures is conducted in Section V. Our simulations and conclusions appear in Sections VI and VII, respectively.

## II. BAYESIAN PROBLEM FORMULATION

We consider  $K$  statistically independent discrete time data streams denoted by  $\{X_n^{(k)}\}_{n=1}^\infty$ ,  $k \in [K] \triangleq \{1, \dots, K\}$ . For the  $k$ th data stream there is a random change point,  $t^{(k)} \geq 1$ ,  $\forall k \in [K]$ , where the prior distribution of each change point is assumed to be known. Commonly, geometric distribution is assumed as a prior for discrete time change-point detection [11], [31]. The change points are assumed to be independent and identically distributed (i.i.d.) among the data streams. For the  $k$ th data stream, given its change point,  $t^{(k)}$ , we assume that  $\{X_n^{(k)}\}_{n=1}^{t^{(k)}-1}$  are i.i.d. with known probability density  $f_0$  and  $\{X_n^{(k)}\}_{n=t^{(k)}}^\infty$  are i.i.d. with known probability density  $f_1$ . Due to communication limitations, at a given time slot we choose a subset of data streams to observe among the active data streams. Let  $K_n \in \mathbb{N}$  denote the number of active sensors at time slot  $n$ . We set a fixed proportion value  $q \in [0, 1]$  and observe  $\lceil qK_n \rceil \in \mathbb{N}$  of the active data streams, where  $\lceil \cdot \rceil$  is the ceiling operator. The actual data vectors that are sequentially observed by the FC are denoted by  $\{Y_n^{(s_n)}\}_{n=1}^\infty$ , where  $s_n \subset [K]$  is the subset of sensor indices that are monitored at time slot  $n$ . The filtration at time slot  $n$  is the  $\sigma$ -algebra generated by the random vectors  $Y_1^{s_1}, \dots, Y_n^{s_n}$ , which is denoted by  $\mathcal{F}_n \triangleq \sigma(Y_1^{s_1}, \dots, Y_n^{s_n})$ . In addition, we define the filtration of all the data as  $\mathcal{F}_\infty \triangleq \sigma(\{Y_m^{s_m}\}_{m=1}^\infty)$ . For  $k \in [K]$ , the event  $\{t^{(k)} \leq n\}$  stands for the case that change in the  $k$ th data stream has taken place before or at time slot  $n$ . We define the posterior probability of the event  $\{t^{(k)} \leq n\}$  using the observations up to time slot  $n$  as

$$\pi_n^{(k)} \triangleq P(t^{(k)} \leq n | \mathcal{F}_n), \quad n = 1, 2, \dots, \quad (1)$$

and  $\pi_0^{(k)} \triangleq 0$ . We also define the likelihood ratio (LR),

$$L(X) \triangleq \frac{f_1(X)}{f_0(X)}, \quad (2)$$

and denote the Kullback-Leibler divergence of  $f_1$  and  $f_0$  as  $D(f_1 || f_0)$ .

Under the assumption of i.i.d. change points, by using Bayes' rule we can recursively compute  $\pi_n^{(k)}$  as follows:

$$\pi_n^{(k)} = \begin{cases} \frac{L(X_n^{(k)}) (\pi_{n-1}^{(k)} + \rho_n (1 - \pi_{n-1}^{(k)}))}{L(X_n^{(k)}) (\pi_{n-1}^{(k)} + \rho_n (1 - \pi_{n-1}^{(k)})) + (1 - \rho_n) (1 - \pi_{n-1}^{(k)})}, & k \in s_n \\ \pi_{n-1}^{(k)} + \rho_n (1 - \pi_{n-1}^{(k)}), & k \notin s_n \end{cases} \quad (3)$$

$n \geq 1$ ,  $\forall k \in [K]$ , where  $\rho_n \triangleq P(t^{(k)} = n | t^{(k)} \geq n)$  depends on the prior distribution of the change point. In [12, Eq. (4.2)], the statistic  $\Lambda_n^{(k)} \triangleq \frac{\pi_n^{(k)}}{1 - \pi_n^{(k)}}$  is considered instead of  $\pi_n^{(k)}$  and corresponding recursive update formula is presented for single change-point detection under general prior distribution. In case  $k \in s_n$ , then at time slot  $n$  an observation is received from sensor  $k$  and  $\pi_n^{(k)}$  is computed using the observations received

before time slot  $n$ , the prior distribution of  $t^{(k)}$ , and the new observation  $X_n^{(k)}$ . The posterior update for the case  $k \notin s_n$  corresponds to the case in which at time slot  $n$  we do not receive an observation from sensor  $k$ . In this case,  $\pi_n^{(k)}$  is computed using only the observations received before time slot  $n$  and the prior distribution of  $t^{(k)}$ . It is shown in [25] that under mild conditions, the  $k$ th sensor posterior probability is a sufficient statistic for evaluating the  $k$ th stopping rule ADD and Type I error probability.

In the considered problem, we have to define multiple stopping rules  $T^{(k)}$ ,  $k \in [K]$ , where the event  $\{T^{(k)} \leq n\}$  is measurable w.r.t.  $\mathcal{F}_n$ . We define

$$\text{FDR} \triangleq \mathbb{E} \left[ \frac{V}{\max\{R, 1\}} \right], \quad (4)$$

where  $\mathbb{E}[\cdot]$  stands for the expectation. The term  $V$  is the number of false discoveries, i.e. the size of the subset of  $[K]$  s.t.  $T^{(k)} < t^{(k)}$ . The term  $R$  denotes the number of change points declared, i.e. the size of the subset of  $[K]$  s.t.  $T^{(k)} < \infty$ . We would like to control the FDR s.t. it will be no higher than a predefined tolerated level  $\alpha \in (0, 1)$ . The ADD for the  $k$ th data stream is defined as

$$\text{ADD}_k \triangleq \mathbb{E}[\max\{0, T^{(k)} - t^{(k)}\}]. \quad (5)$$

Since we consider multiple statistically independent data streams, we define the overall ADD as

$$\text{ADD} \triangleq \frac{1}{K} \sum_{k=1}^K \text{ADD}_k. \quad (6)$$

Assume that at time slot  $n$ , we have  $K_n$  active data streams. Then, we observe  $\lceil qK_n \rceil$  of them. We define

$$\text{ANO} \triangleq \mathbb{E} \left[ \frac{1}{K} \sum_{n=1}^{\sup_{k \in [K]} \{T^{(k)}\}} \lceil qK_n \rceil \right]. \quad (7)$$

The ANO definition extends the definition from [19], which is defined for single change point detection, i.e.  $K = 1$ . A difference between the definitions is that the ANO from [19] does not consider the observations drawn after the change point occurs, while the ANO definition in (7) takes into account all the observations drawn until change points are declared. This is in order to properly evaluate the communication burden caused by transmissions of data streams from the sensors to the FC. In the following section, we propose the S-MAP procedure, which is a Bayesian multiple change-point detection procedure that controls the FDR under the limitation on the proportion of sensors communicating their data streams to the FC.

### III. S-MAP DETECTION PROCEDURE

In this section, we propose a Bayesian detection procedure that is tasked to eventually discover all the random change points that occur in the monitored environment. At a given time slot, we consider each sensor individually and evaluate its posterior probability from (1) using the recursive formula from (3). At time slot  $n$ , we have  $K_n$  active data streams of which we observe only a subset of size  $\lceil qK_n \rceil \in \mathbb{N}$ . The developed

S-MAP procedure extends the method in [4] by proposing a rule for choosing the subset of  $\lceil qK_n \rceil$  data streams to observe. In the S-MAP procedure, we use the posterior probability from (1) as a test statistic, rather than the test statistic from [4], [13], which is based on a Bayesian version of the LR. The test statistic from [4], [13] is used under a very strong global false alarm probability constraint [13] that may be too conservative in terms of FDR control. Under the communication limitations, among the  $K_n$  active data streams, we choose to observe the  $\lceil qK_n \rceil$  data streams with the highest posterior probabilities of a change point having occurred. The motivation for the S-MAP approach is that we are interested in minimizing the time between the occurrence of a change point and its declaration using the sequentially updated posterior probabilities. The S-MAP procedure that monitors all of the active data streams, i.e. with  $q = 1$ , is denoted as the parallel procedure. In the following, we describe the proposed S-MAP procedure.

We construct a descending set of  $K$  thresholds  $Q_r$ ,  $r \in [K]$ , s.t. the detection on the  $k$ th data stream that samples until  $\pi_n^{(k)} \geq Q_r$  has a Type I error probability that is smaller than or equal to  $\frac{r\alpha}{K}$ , where  $\alpha \in (0, 1)$  is the predefined FDR tolerance level. Formally,

$$P(\exists n < t^{(k)} \text{ s.t. } \pi_n^{(k)} \geq Q_r) \leq \frac{r}{K} \alpha. \quad (8)$$

According to [12] and [31, p. 225], the choice

$$Q_r = 1 - \frac{r\alpha}{K} \quad (9)$$

ensures that (8) is satisfied. The proposed detection procedure is divided into sampling stages. Each sampling stage may take several time slots. In the beginning of a sampling stage, we gather all the active data streams and obtain observations from a subset of them, according to the S-MAP approach. This process is repeated at each time slot sequentially, until at least one active data stream posterior probability exceeds its corresponding threshold. Then, we declare changes for some of the active data streams, which are then eliminated from the active data streams set.

Let  $I_j$  denote the set of indices of active data streams with cardinality  $|I_j|$  at the beginning of the  $j$ th sampling stage and let  $n_j$  denote the time slot at the end of the  $j$ th sampling stage. Note that  $I_1 = [K]$  and  $n_0 = 0$ . The  $j$ th stage of sampling is described as follows:

- 1) Sample the  $\lceil q|I_j| \rceil$  data streams with the currently highest posterior probabilities.
- 2) Update the posterior probabilities of the sensors with active data streams using (3).
- 3) Sort the updated posterior probabilities in ascending order as  $\pi_n^{(i(n,l))}$ , where  $i(n,l)$  denotes the index of the  $l$ th ordered posterior probability at time slot  $n$ .
- 4) Repeat this process until time slot  $n_j$  in which at least one of the posterior probabilities is higher than its corresponding threshold, i.e.  $n_j = \min\{n > n_{j-1} : \exists l \in [|I_j|], \pi_n^{(i(n,l))} \geq Q_{K-l+1}\}$ .
- 5) Declare change points for the data streams  $i(n_j, l_j), i(n_j, l_j + 1), \dots, i(n_j, |I_j|)$ , where  $l_j = \min\{l \in [|I_j|] : \pi_{n_j}^{(i(n_j,l))} \geq Q_{K-l+1}\}$  and



remove these data streams from the set of active data streams.

- 6) Update  $I_{j+1}$  to be the set of indices of the remaining active data streams. Stop the procedure if  $|I_{j+1}| = 0$ .

In the following theorem, we show that we control the FDR of the S-MAP procedure to remain under the upper bound constraint  $\alpha \in (0, 1)$ .

**Theorem 1.** *For upper bound constraint  $\alpha \in (0, 1)$ , the S-MAP procedure satisfies*

$$FDR \leq \alpha. \quad (10)$$

*Proof:* Recall that we choose the thresholds  $Q_r$ ,  $r \in [K]$  from (9), s.t. (8) is satisfied. Thus, by following the lines of the FDR control proofs in [3], [4], we obtain that the FDR is controlled by the proposed S-MAP procedure under the upper bound constraint  $\alpha$ . ■

In the following section, we propose an alternative detection procedure that is less conservative than the S-MAP procedure in terms of FDR control. Therefore, the proposed alternative procedure has improved performance in terms of ADD and ANO compared to the S-MAP procedure.

#### IV. IMPROVING THE S-MAP PROCEDURE

In order to guarantee FDR control, the S-MAP procedure uses the false alarm constraints from (8), which are the same false alarm constraints as in [4] to guarantee FDR control. However, we show in this section that the false alarm constraints from [4] may be too conservative and the corresponding posterior probability threshold values may be too high. We propose the IS-MAP detection procedure, which is similar to the S-MAP procedure except that its threshold values are lower than the thresholds of the S-MAP procedure. Since the IS-MAP procedure uses lower threshold values, then for a fixed proportion,  $q$ , the ADD and ANO will decrease compared to the S-MAP procedure, i.e. the ADD and ANO performance will improve. Moreover, using the lower thresholds, we prove that we can still control the FDR under the desired level,  $\alpha$ . In the IS-MAP procedure, we construct a set of  $K$  thresholds  $Q_k$ ,  $k \in [K]$ , s.t. the detection on the  $k$ th data stream that samples until  $\pi_n^{(k)} \geq Q_k$  has an individual Type I error probability that is smaller than or equal to  $\alpha$ , where  $\alpha \in (0, 1)$  is the predefined FDR tolerated level. Formally,

$$P(\exists n < t^{(k)} \text{ s.t. } \pi_n^{(k)} \geq Q_k) \leq \alpha. \quad (11)$$

According to [12] and [31, p. 225], the choice

$$Q_k = Q = 1 - \alpha, \quad \forall k \in [K], \quad (12)$$

ensures that (11) is satisfied. Since the thresholds of the IS-MAP procedure are all equal to  $Q$ , its  $j$ th sampling stage can be written in a more compact form than the corresponding sampling stage of the S-MAP procedure. Let  $I_j$  denote the set of indices of active data streams with cardinality  $|I_j|$  at the beginning of the  $j$ th sampling stage and let  $n_j$  denote the time slot at the end of the  $j$ th sampling stage. The  $j$ th stage of sampling is described as follows:

- 1) Sample the  $\lceil q|I_j| \rceil$  data streams with highest posterior probabilities.
- 2) Update the posterior probabilities of the sensors with active data streams using (3).
- 3) Repeat this process until time slot  $n_j$  in which at least one of the posterior probabilities is higher than the threshold  $Q$ , i.e.  $n_j = \min\{n > n_{j-1} : \exists k \in I_j, \pi_n^{(k)} \geq Q\}$ .
- 4) Declare change points for all the data streams with indices in  $I_j$  whose posterior probabilities are higher than or equal to  $Q$  and remove these data streams from the set of active data streams.
- 5) Update  $I_{j+1}$  to be the set of indices of the remaining active data streams. Stop the procedure if  $|I_{j+1}| = 0$ .

In the following theorem, we show that the FDR of the IS-MAP procedure satisfies the desired upper bound constraint.

**Theorem 2.** *For upper bound constraint  $\alpha \in (0, 1)$ , the IS-MAP procedure satisfies*

$$FDR \leq \alpha. \quad (13)$$

*Proof:* The proof is given in Appendix A. ■

As mentioned previously, the proposed IS-MAP procedure is similar to the S-MAP procedure from Section III, except that the procedures use different thresholds in order to guarantee the FDR control. Since the thresholds of the IS-MAP procedure in (12) are smaller than the thresholds of the S-MAP procedure, then for a fixed proportion,  $q$ , the IS-MAP procedure will have a lower ADD and ANO than the S-MAP procedure, while the FDR of the IS-MAP procedure will be higher than the S-MAP FDR. It should be noted that in case of model uncertainty, FDR control is not guaranteed for the S-MAP and the IS-MAP procedures. Then, depending on the application, if ADD and ANO are more significant than FDR, the IS-MAP procedure should be implemented rather than the S-MAP procedure, while if FDR is more significant than ADD and ANO, then the S-MAP procedure may be preferred. In the following section, we analyze the asymptotic ADD behavior of the S-MAP and the IS-MAP procedures under the assumption of geometric prior distribution for the change points.

#### V. ADD ANALYSIS OF THE S-MAP AND THE IS-MAP PROCEDURES

In this section, we derive asymptotic lower and upper bounds on the ADD of the S-MAP and the IS-MAP procedures for  $\alpha \rightarrow 0$  and a fixed number of data streams  $K$ . Then, we characterize the behavior of these bounds as  $K \rightarrow \infty$ . For simplicity of the analysis, we assume that the prior distribution of each change point obeys a geometric distribution with common parameter  $\rho \in (0, 1)$ , i.e.

$$P(t^{(k)} = m) = \rho(1 - \rho)^{m-1}, \quad \forall m = 1, 2, \dots, \forall k \in [K]. \quad (14)$$

The geometric prior distribution is commonly assumed in change-point detection problems. This is a memoryless distribution that is both mathematically convenient and provides a reasonable model in practical applications [11], [31]. Under the assumption of i.i.d. change points with geometric priors,

it is shown in [19], [25] that the posterior probability of the  $k$ th sensor evolves in a sequential manner via the recursion

$$\pi_n^{(k)} = \begin{cases} \frac{L(X_n^{(k)}) (\pi_{n-1}^{(k)} + \rho(1 - \pi_{n-1}^{(k)}))}{L(X_n^{(k)}) (\pi_{n-1}^{(k)} + \rho(1 - \pi_{n-1}^{(k)})) + (1 - \rho)(1 - \pi_{n-1}^{(k)})}, & k \in s_n \\ \pi_{n-1}^{(k)} + \rho(1 - \pi_{n-1}^{(k)}), & k \notin s_n \end{cases} \quad (15)$$

$n \geq 1, \forall k \in [K]$ . It can be seen that the recursive formula in (15) is obtained by substituting  $\rho_n = \rho$  in (3).

Under communication limitations, the FC observes a subsequence of the complete observation sequence from each sensor. According to the maximum a-posteriori probability (MAP) approach, the indices of the monitored observations are random and determined online based on the proportion,  $q$ , and the posterior probability values of the active sensors at each time slot. Therefore, it is difficult to characterize the subsequence of observations obtained from each sensor. In order to obtain asymptotic bounds on the ADD of the S-MAP and the IS-MAP procedures, we begin by considering a single change-point detection with the posterior update from (15). Thus, we consider the observation sequence  $\{X_n\}_{n=1}^\infty$  with change point  $t$  and stopping rule of the form

$$T = \inf\{n \in \mathbb{N} : \pi_n \geq 1 - \eta\}, \eta \in (0, 1). \quad (16)$$

We assume that only a *subsequence* of the complete observation sequence is obtained. It is shown in [25] that for any subsequence of observations, the ADD of the stopping rule in the form of (16) as  $\eta \rightarrow 0$  satisfies

$$\text{ADD} \geq \frac{|\log \eta|}{D(f_1 \| f_0) + |\log(1 - \rho)|} (1 + o_\eta(1)) \quad (17)$$

and

$$\text{ADD} \leq \frac{|\log \eta|}{|\log(1 - \rho)|} (1 + o_\eta(1)), \quad (18)$$

where  $o_\eta(1) \rightarrow 0$  as  $\eta \rightarrow 0$ . The asymptotic ADD lower bound from (17) is attained when the complete observation sequence is available. The asymptotic ADD upper bound from (18) is attained when we do not take observations and the stopping rule is based only on the prior.

In the following theorem, using (17)-(18) we derive asymptotic lower and upper bounds on the ADDs of the S-MAP and the IS-MAP procedures as  $\alpha \rightarrow 0$ . These ADD bounds do not require any assumptions on the subsequence of observations obtained from each sensor.

**Theorem 3.** For  $\alpha \rightarrow 0$  and any proportion of observed sensors,  $q$ , we obtain

$$\text{ADD}_{S\text{-MAP}} \geq \frac{|\log \alpha|}{D(f_1 \| f_0) + |\log(1 - \rho)|} (1 + o_\alpha(1)), \quad (19)$$

$$\text{ADD}_{S\text{-MAP}} \leq \frac{\log K - \frac{1}{K} \log K! + |\log \alpha|}{|\log(1 - \rho)|} (1 + o_\alpha(1)), \quad (20)$$

$$\text{ADD}_{IS\text{-MAP}} \geq \frac{|\log \alpha|}{D(f_1 \| f_0) + |\log(1 - \rho)|} (1 + o_\alpha(1)), \quad (21)$$

and

$$\text{ADD}_{IS\text{-MAP}} \leq \frac{|\log \alpha|}{|\log(1 - \rho)|} (1 + o_\alpha(1)). \quad (22)$$

*Proof:* The proof is given in Appendix B. ■

For the ADD of the stopping rule,  $T$ , from (16) we can derive a tighter upper bound than (18) under some assumptions on the subsequence of observations obtained for the detection. Let us denote by  $\{X_{V_n}\}_{n=1}^\infty$  the subsequence of the complete observation sequence, where  $V_0 \triangleq 0$  and  $V_1, V_2, \dots$  are the integer time slots in which observations are obtained for the detection of the single change point,  $t$ , using the stopping rule,  $T$ . Equivalently, we sample the complete observation sequence with intervals

$$\zeta_n \triangleq V_n - V_{n-1} \geq 1, n \in \mathbb{N}. \quad (23)$$

In addition, we define

$$\zeta^{(N)} \triangleq \frac{1}{N} \sum_{n=1}^N \zeta_n = \frac{V_N}{N}, \quad (24)$$

which is the average length of intervals in which we sample  $N$  observations from the observation sequence, the stopping rule,

$$\Gamma \triangleq \inf\{n \in \mathbb{N} : \pi_{V_n} \geq 1 - \eta\}, \quad (25)$$

and the random change point,

$$\gamma \triangleq \inf\{n \in \mathbb{N} : V_n \geq t\}. \quad (26)$$

The stopping rule and change point from (25) and (26), respectively, represent the case in which we only count time slots where observations are obtained. The time slots,  $\{V_n\}_{n=1}^\infty$ , and intervals,  $\{\zeta_n\}_{n=1}^\infty$ , may be unknown. For the derivation of a tighter asymptotic upper bound on the ADD of the stopping rule,  $T$ , we only assume that the intervals are bounded, i.e. there exists  $1 \leq \mathcal{B} < \infty$  s.t.

$$\zeta_n \leq \mathcal{B}, \forall n \in \mathbb{N}, \quad (27)$$

there exists  $\zeta \in [1, \mathcal{B}]$  s.t.

$$\lim_{N \rightarrow \infty} \zeta^{(N)} = \zeta, \quad (28)$$

and

$$\mathbb{E}[\zeta^{(\Gamma)} \max\{0, \Gamma - \gamma\}] = \zeta \mathbb{E}[\max\{0, \Gamma - \gamma\}] (1 + o_\eta(1)). \quad (29)$$

From (23)-(24),  $\zeta^{(\Gamma)} = \frac{V_\Gamma}{\Gamma}$ ,  $\zeta^{(\gamma)} = \frac{V_\gamma}{\gamma}$ ,  $\zeta_\gamma \triangleq V_\gamma - V_{\gamma-1}$ , and the specific value of  $\zeta$  may be unknown. The assumption in (29) essentially requires that  $\Gamma \rightarrow \infty$  as  $\eta \rightarrow 0$ . In the following proposition, we derive an asymptotic ADD upper bound for the stopping rule,  $T$ , which is tighter than (18).

**Proposition 4.** Assume that (27)-(29) are satisfied. Then, as  $\eta \rightarrow 0$  the ADD of the stopping rule  $T$  from (16) satisfies

$$\text{ADD} \leq \frac{|\log \eta|}{\frac{1}{\zeta} D(f_1 \| f_0) + |\log(1 - \rho)|} (1 + o_\eta(1)). \quad (30)$$

*Proof:* The proof is given in Appendix C. ■

It should be noted that a special case of (30) with  $\zeta_n = \zeta < \infty, n \in \mathbb{N}$ , was proved in [25].

Assume that each stopping rule in the S-MAP procedure satisfies the ADD upper bound in (30) with  $\zeta = g_k^{\text{S-MAP}}$  and that each stopping rule in the IS-MAP procedure satisfies the ADD upper bound in (30) with  $\zeta = g_k^{\text{IS-MAP}}, \forall k \in [K]$ . In addition,

assume that  $g^* \triangleq \sup\{\{g_k^{\text{S-MAP}}\}_{k=1}^K, \{g_k^{\text{IS-MAP}}\}_{k=1}^K\} < \infty$ . Then, in a similar manner to the derivation of the upper bounds in (20) and (22), we obtain tighter asymptotic ADD upper bounds for the S-MAP and the IS-MAP procedures, given by

$$\text{ADD}_{\text{S-MAP}} \leq \frac{\log K - \frac{1}{K} \log K! + |\log \alpha|}{\frac{1}{g^*} D(f_1||f_0) + |\log(1-\rho)|} (1 + o_\alpha(1)) \quad (31)$$

and

$$\text{ADD}_{\text{IS-MAP}} \leq \frac{|\log \alpha|}{\frac{1}{g^*} D(f_1||f_0) + |\log(1-\rho)|} (1 + o_\alpha(1)), \quad (32)$$

respectively.

In (19), (20), and (31) and in (21), (22), and (32), we obtained asymptotic ADD bounds for the S-MAP and the IS-MAP procedures, respectively. For any fixed proportion,  $q$ , of observed data streams and for sufficiently small  $\alpha \neq 0$  these bounds hold. We characterize the behavior of these bounds as  $K$  increases towards  $\infty$  in order to investigate the scalability of the S-MAP and the IS-MAP procedures, as the number of data streams increases. Let

$$\text{ADD}_{\text{LB}}^{(\alpha)} \triangleq \frac{|\log \alpha|}{D(f_1||f_0) + |\log(1-\rho)|} \quad (33)$$

denote the asymptotic ADD lower bound for both the S-MAP and the IS-MAP procedures. It can be seen that this lower bound is a finite constant w.r.t.  $K$ .

We denote the asymptotic ADD upper bounds for the S-MAP procedure as

$$\text{ADD}_{\text{S-MAP,UB}}^{(\alpha,K)} \triangleq \frac{\log K - \frac{1}{K} \log K! + |\log \alpha|}{|\log(1-\rho)|}. \quad (34)$$

and

$$\text{ADD}_{\text{S-MAP,UB}}^{(\alpha,g^*,K)} \triangleq \frac{\log K - \frac{1}{K} \log K! + |\log \alpha|}{\frac{1}{g^*} D(f_1||f_0) + |\log(1-\rho)|}. \quad (35)$$

Consider the sequence  $\{\log K - \frac{1}{K} \log K!\}_{K=1}^\infty$ . Using [32, Eq. (5)] and Stirling's approximation (see e.g. [32], [33]) and applying some algebraic manipulations, it can be verified that this sequence is monotonically increasing and converges to 1. Thus, we obtain that  $\text{ADD}_{\text{S-MAP,UB}}^{(\alpha,K)}$  and  $\text{ADD}_{\text{S-MAP,UB}}^{(\alpha,g^*,K)}$  are monotonically increasing with  $K$  and converge to a finite constant, i.e.

$$\lim_{K \rightarrow \infty} \text{ADD}_{\text{S-MAP,UB}}^{(\alpha,K)} = \frac{1 + |\log \alpha|}{|\log(1-\rho)|} \quad (36)$$

and

$$\lim_{K \rightarrow \infty} \text{ADD}_{\text{S-MAP,UB}}^{(\alpha,g^*,K)} = \frac{1 + |\log \alpha|}{\frac{1}{g^*} D(f_1||f_0) + |\log(1-\rho)|}. \quad (37)$$

In a similar manner to (34)-(35), we denote

$$\text{ADD}_{\text{IS-MAP,UB}}^{(\alpha)} \triangleq \frac{|\log \alpha|}{|\log(1-\rho)|} \quad (38)$$

and

$$\text{ADD}_{\text{IS-MAP,UB}}^{(\alpha,g^*)} \triangleq \frac{|\log \alpha|}{\frac{1}{g^*} D(f_1||f_0) + |\log(1-\rho)|}. \quad (39)$$

The upper bounds in (38)-(39) are finite constants w.r.t.  $K$ .

The sequence  $\{\log K - \frac{1}{K} \log K!\}_{K=1}^\infty$  is nonnegative and thus,

$$\text{ADD}_{\text{IS-MAP,UB}}^{(\alpha)} \leq \text{ADD}_{\text{S-MAP,UB}}^{(\alpha,K)} \quad (40)$$

and

$$\text{ADD}_{\text{IS-MAP,UB}}^{(\alpha,g^*)} \leq \text{ADD}_{\text{S-MAP,UB}}^{(\alpha,g^*,K)}. \quad (41)$$

In addition, by comparing (38)-(39) to (34)-(35) as  $K \rightarrow \infty$ , we obtain

$$\begin{aligned} \lim_{K \rightarrow \infty} \frac{\text{ADD}_{\text{IS-MAP,UB}}^{(\alpha,K)}}{\text{ADD}_{\text{S-MAP,UB}}^{(\alpha,K)}} &= \lim_{K \rightarrow \infty} \frac{\text{ADD}_{\text{IS-MAP,UB}}^{(\alpha,g^*)}}{\text{ADD}_{\text{S-MAP,UB}}^{(\alpha,g^*,K)}} \\ &= \frac{|\log \alpha|}{1 + |\log \alpha|} < 1, \end{aligned} \quad (42)$$

where the second equality is obtained by substituting (36)-(39). The results in (40)-(42) demonstrate the ADD improvement obtained by using the IS-MAP procedure instead of the S-MAP procedure.

The presented asymptotic ADD results hold for any proportion value,  $q$ . However, it is expected that the S-MAP ADD and the IS-MAP ADD will increase as the proportion of monitored sensors decreases. An intuitive explanation for this phenomenon is as follows: For fixed  $\pi_{n-1}^{(k)}$ , the posterior probability in (15) is monotonically nondecreasing with the LR,  $L(\cdot)$ . After a change occurs, we receive samples from  $f_1$ . By taking the expectation of the difference  $L(X) - 1$  w.r.t.  $f_1$  and using  $E_{f_1}[L(X)] = E_{f_0}[L^2(X)]$  and  $E_{f_0}[L(X)] = 1$ , we obtain

$$E_{f_1}[L(X) - 1] = E_{f_0}[(L(X) - 1)^2] \geq 0. \quad (43)$$

The case  $L(X) = 1$  corresponds to the case in which we choose not to monitor the corresponding sensor. Thus, as the number of observations increases, the threshold will usually be exceeded in an earlier time slot and consequently, the ADD will usually be lower. An advantage of observing only a small subset of sensors is that the ANO for the detection task may decrease, which reduces the communication burden. Consequently, we identify a tradeoff between the ADD and the ANO. We will investigate this tradeoff in Section VI.

## VI. NUMERICAL SIMULATIONS

In this section, we evaluate the performance of the proposed S-MAP and IS-MAP procedures in terms of FDR, ADD, and ANO. In addition, the analytical results from Sections III-V are verified in the simulations. The simulation results are based on 1000 Monte Carlo runs. We generate the true change points independently for each sensor from a geometric distribution with parameter  $\rho = 0.01$  and assume that we know this parameter when applying the procedure. It should be noted that in case  $\rho$  is unknown then by assuming a sufficiently low value for  $\rho$ , the FDR of the S-MAP and the IS-MAP procedures may still be controlled under the desired upper bound. The reason is that the posterior probabilities from (15) decrease as  $\rho$  decreases. If the assumed value of  $\rho$  is lower than the true value of  $\rho$ , the change-points will usually be declared in later time slots than in the case in which the true value of  $\rho$  is used. Thus, the FDR will not increase. In all

cases, we set the FDR upper bound as  $\alpha = 0.1$ .

For comparison purposes, we implement and evaluate the performance of two additional procedures. The first procedure is a simplified version of the S-MAP procedure, which is referred to as the simple procedure. This procedure simplifies S-MAP from Section III by replacing the method of choosing the subset of sensors to monitor. In the simple procedure, at each time slot we randomly choose a subset of active sensors with consecutive indices to monitor within the allowed proportion. Following the FDR control proofs in [3], [4], it can be shown that the simple procedure controls the FDR under the predefined upper bound. This procedure is implemented in order to verify that the MAP approach for choosing the subset of sensors to monitor, as used in the S-MAP procedure, improves the ADD performance compared to randomly choosing this subset, as used in the simple procedure. The second method implemented for comparisons is the fully parallel procedure of [4], named D-FDR, that observes all the data streams. The FDR control of the D-FDR procedure is shown in [4]. In this procedure, the following test statistic is used

$$G_n^{(k)} = \sum_{m=1}^{\infty} P(t^{(k)} = m) \prod_{i=m}^n L(X_i^{(k)}), \quad n = 1, 2, \dots \quad (44)$$

This test statistic is the average LR (ALR) between the hypotheses that the change occurs at  $t^{(k)} = m < \infty$  and that the change never occurs,  $t^{(k)} = \infty$ . This ALR test statistic is recursively updated according to the following formula:

$$G_n^{(k)} = G_{n-1}^{(k)} L(X_n^{(k)}) + P(t^{(k)} \geq n+1)(1 - L(X_n^{(k)})), \quad (45)$$

where  $G_0^{(k)} \triangleq 1$ . For  $q = 1$ , the D-FDR procedure is similar to the S-MAP procedure except that it uses the ALR test statistic, rather than the posterior probability test statistic, with the thresholds

$$Q_r = \frac{K}{r\alpha}, \quad r \in [K], \quad (46)$$

in order to guarantee the same false positive constraints as in (8). Assume that for the  $k$ th data stream, the corresponding threshold is  $Q_{r_k} = \frac{K}{r_k\alpha}$ ,  $r_k \in [K]$ . It is shown in [13] that in this case, using the ALR test statistic with the threshold  $Q_{r_k}$  is equivalent to using the posterior probability test statistic with the threshold

$$Q_{r_k}^* = 1 - p(t^{(k)} \geq n+1) \frac{r_k\alpha}{K}. \quad (47)$$

Thus, from (9), (12), and (47), the posterior probability thresholds of the D-FDR procedure are higher than the posterior probability thresholds of the S-MAP and the IS-MAP procedures. Consequently for  $q = 1$ , the ADD and ANO of the S-MAP and the IS-MAP procedures will be lower than the ADD and ANO of the D-FDR procedure.

In Subsection VI-A, we consider multiple change-point detection with known Gaussian distributions and in Subsection VI-B, we consider a general model under some uncertainty and use  $p$ -values [14], [15], [34]–[37] from each sensor as observations for the multiple change-points detection. It should be noted that in the simulations, we assume that we have a sufficient number of observations for declaring the changes so there are no Type II errors corresponding to infinite ADD.

#### A. Gaussian distribution scenario

We consider Gaussian distributions with a change in the mean and set  $f_0 = \mathcal{N}(0, 1)$  and  $f_1 = \mathcal{N}(1, 1)$  as depicted in Fig. 1. First, for  $K = 10, 100, 200, 500, 1000$ , we examine the FDR control of the proposed S-MAP and IS-MAP procedures with  $\{q = 0.05m\}_{m=1}^{20}$ , where  $q \in [0, 1]$  is the proportion of monitored sensors. The proportion  $q = 1$  corresponds to the parallel versions of the S-MAP and the IS-MAP procedures that observe all the active data streams at each time slot. Due to space limitations, we do not present tables of all the estimated FDR results. The resulting minimum and maximum estimated FDR values of the S-MAP procedure are 0.028 and 0.037, respectively, while the resulting minimum and maximum estimated FDR values of the IS-MAP procedure are 0.058 and 0.068, respectively. Consequently, both procedures control the FDR under the upper bound  $\alpha = 0.1$ . These results confirm the analytical results in Theorems 1 and 2. The S-MAP FDR values are lower than the IS-MAP FDR values, since the S-MAP procedure is more conservative and uses higher thresholds than the IS-MAP procedure. For both the S-MAP and the IS-MAP procedures there is still a gap between the FDR values and the upper bound  $\alpha$ . This result follows from the choices of thresholds in (9) and (12) for the S-MAP and the IS-MAP procedures, respectively, that neglect the overshoot in the stopping rule [12].

In Fig. 2, we evaluate the ADD of the procedures: D-FDR, S-MAP with  $q = 0.5, 1$ , simple procedure with  $q = 0.5$ , and IS-MAP with  $q = 0.5, 1$  versus  $K = 10, 100, 200, 500, 1000$ . It can be seen that all the considered procedures have an approximately constant ADD as  $K$  increases, which verifies the analytical results in Section V. The parallel version of the IS-MAP procedure, i.e. for  $q = 1$ , has the lowest ADD. Moreover, it can be seen that the IS-MAP procedure with  $q = 0.5$  outperforms the parallel version of the S-MAP procedure and the D-FDR procedure. These results demonstrate the advantage of using the IS-MAP procedure instead of the S-MAP or the D-FDR procedures in terms of ADD. The simple procedure with  $q = 0.5$  has the highest ADD among the considered procedures implying that the proposed MAP approach is desirable for choosing the sensors to monitor at each time slot within the allowed proportion. In Fig. 3, we evaluate the ANO versus  $K$  of the procedures: D-FDR, S-MAP with  $q = 0.5, 1$ , and IS-MAP with  $q = 0.5, 1$ . It can be seen that IS-MAP with  $q = 0.5$  has the lowest and the D-FDR has the highest ANO. In addition, it can be seen that for all the procedures, the ANO is approximately a constant w.r.t.  $K$ .

In the upper and middle plots of Fig. 4, we plot the ADDs and ANOs, respectively, of the S-MAP and the IS-MAP procedures for  $K = 1000$  versus the proportion values  $\{q = 0.05m\}_{m=1}^{20}$ . It can be seen that for any of the considered proportions, the IS-MAP procedure has lower ADD and ANO than the S-MAP procedure. In addition, for both procedures the ADD decreases as the proportion increases, while the ANO increases approximately linearly as the proportion increases. Thus, we notice a tradeoff between ADD and ANO as we change the proportion value,  $q$ . It can be seen that for both procedures there is no significant increase in ADD when the



proportion decreases from  $q = 1$  to  $q = 0.3$ , whereas the ANO increases significantly as we increase  $q$  towards 1. This result implies that in this example it may be a waste of resources to monitor all the active data streams in parallel. In the lower plot of Fig. 4, we plot a curve connecting the ADD-ANO points of the S-MAP and the IS-MAP procedures from the upper and middle plots of Fig. 4. It can be seen that in this example there is a clear tradeoff between the ADD and ANO, i.e. as the proportion,  $q$ , increases the ADD becomes lower, while the ANO becomes higher.

In order to evaluate the performance of the procedures using both the ADD and the ANO as criteria, we define a weighted risk,

$$(1 - c)\text{ADD} + c\text{ANO}, \quad (48)$$

where  $c \in [0, 1]$  sets the weighting between the ADD and the ANO. For  $c = 0$  we are only interested in the ADD, while for  $c = 1$  we are only interested in the ANO. In the upper plot of Fig. 5, we compare the weighted risks of the S-MAP and the IS-MAP procedures with different proportions  $\{q = 0.05m\}_{m=1}^{20}$  versus the proportion size for  $c = 0.2$ . It can be seen that the weighted risk of the IS-MAP procedure is lower than the weighted risk of the S-MAP procedure. For both the S-MAP and the IS-MAP procedures, the best tradeoff among the considered proportions is achieved with the proportion  $q = 0.3$ . Thus, when both the ADD and the ANO are taken into account it may not be necessary to monitor all the active data streams in parallel, i.e. to choose  $q = 1$ .

In the lower plot of Fig. 5, for both the S-MAP and the IS-MAP procedures, we present the best proportion among the proportions  $\{q = 0.05m\}_{m=1}^{20}$  in terms of the weighted risk in (48), i.e. the proportion with lowest risk, versus the weighting coefficient  $c$ . It can be seen that for both procedures, as  $c$  increases the best proportion does not increase. Moreover, in most of the considered cases the best proportion decreases as  $c$  increases. Thus, as we put a higher weight on the ANO compared to the ADD we should usually choose a lower proportion of data streams to observe. In addition, as we change  $c$  from 0 to 0.1 there is a rapid decrease in the optimal proportion from  $q = 1$  to  $q = 0.45$  and  $q = 0.4$  in the S-MAP and the IS-MAP procedures, respectively. This result implies that even a small positive weight on the ANO leads to a much smaller proportion value than  $q = 1$  for which the lowest weighted risk is obtained among the considered proportions.

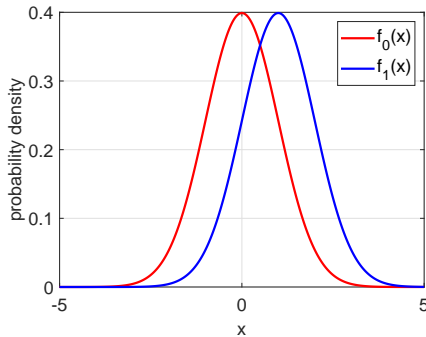


Fig. 1. Gaussian distributions: Probability densities  $f_0 = \mathcal{N}(0, 1)$  and  $f_1 = \mathcal{N}(1, 1)$ .

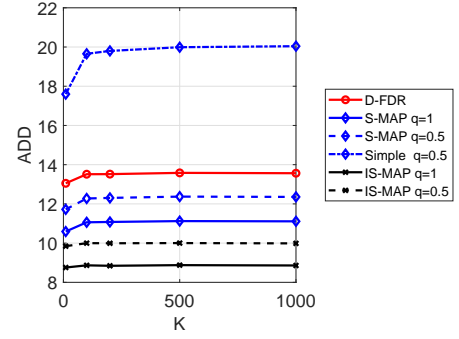


Fig. 2. Gaussian distributions: The ADD of the procedures D-FDR, S-MAP with  $q = 0.5, 1$ , simple, and IS-MAP with  $q = 0.5, 1$ , versus  $K = 10, 100, 200, 500, 1000$ . It can be seen that all the considered procedures have an approximately constant ADD as  $K$  increases. The parallel version of the IS-MAP procedure has the lowest ADD. Moreover, it can be seen that the IS-MAP procedure with  $q = 0.5$  outperforms the parallel version of the S-MAP procedure and the D-FDR procedure. These results demonstrate the improved ADD performance of the IS-MAP procedure compared to the S-MAP and the D-FDR procedures. The simple procedure with  $q = 0.5$  has the highest ADD implying that the MAP approach is more useful than random choice when choosing the subset of sensors to monitor.

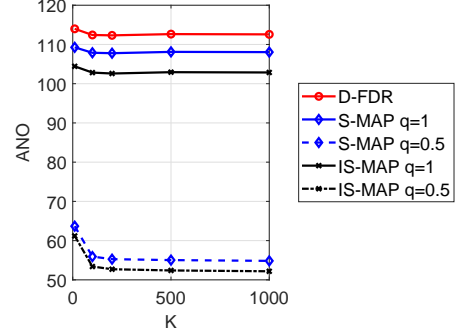


Fig. 3. Gaussian distributions: The ANO of the procedures D-FDR, S-MAP with  $q = 0.5, 1$ , and IS-MAP with  $q = 0.5, 1$ , versus  $K$ . It can be seen that IS-MAP with  $q = 0.5$  has the lowest ANO, while the D-FDR has the highest ANO. For all the considered procedures, the ANO is approximately a constant w.r.t.  $K$ .

### B. General model with uncertainty and known $p$ -values

Due to bandwidth limitations, in many distributed detection applications the sensors communicate to the FC condensed information about their observations in the form of a local decision and/or sufficient statistic. In this case, significantly less data needs to be communicated. Moreover, the local distributions at each sensor may be different and local decision statistics from each sensor may be easier to fuse than fusing the raw data from all the sensors. A common local decision statistic is the  $p$ -value [15], [35], [37], which is the probability of obtaining test results at least as extreme as the results observed during the test assuming that the null hypothesis is correct. The  $p$ -value is general and is not necessarily obtained from the Gaussian distribution. It is a tool for deciding whether to reject the null hypothesis. When the  $p$ -value approaches zero, it is more likely that the alternative hypothesis is true [31, p. 63], [34].

In this example, we assume that the  $p$ -values are accurately calculated by each sensor based on its local observations. The  $p$ -values from each sensor are communicated to the FC for the



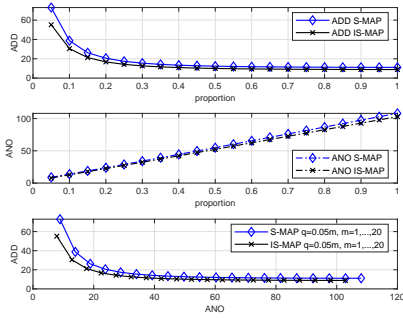


Fig. 4. Gaussian distributions: (Top and middle) The ADDs and ANOs of the S-MAP and the IS-MAP procedures for  $K = 1000$  versus the proportion values  $\{q = 0.05m\}_{m=1}^{20}$ . For any of the considered proportions, the IS-MAP procedure has lower ADD and ANO compared to the S-MAP procedure. For both procedures the ADD decreases as the proportion increases, while the ANO increases approximately linearly as the proportion increases. It can be seen that there is no significant increase in ADD when the proportion decreases from  $q = 1$  to  $q = 0.3$ , whereas the decrease in ANO is more substantial. Thus, with a small proportion we may attain a sufficiently small ADD and significantly decrease the communication burden. (Bottom) A curve connecting the ADD-ANO points of the S-MAP and the IS-MAP procedures from the upper and middle plots of this figure. In this example, there is a clear tradeoff between the ADD and the ANO, i.e. as the the ADD is lower the ANO is higher.

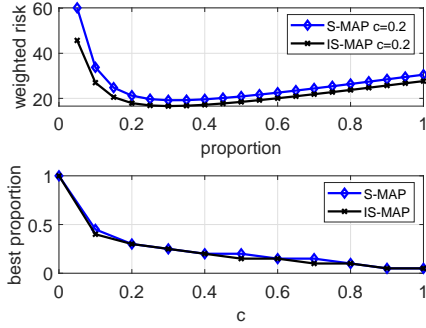


Fig. 5. Gaussian distributions: (Top) The weighted risks of the S-MAP and the IS-MAP procedures with proportions  $\{q = 0.05m\}_{m=1}^{20}$  versus the proportion values for  $c = 0.2$ . The weighted risk of the IS-MAP procedure is lower than the weighted risk of the S-MAP procedure. For both the S-MAP and the IS-MAP procedures, the best tradeoff among the considered proportions is achieved with the proportion  $q = 0.3$ . Thus, in this example, when both the ADD and the ANO are taken into account it is not necessary to monitor all the active data streams in parallel and a lower weighted risk can be attained with a much lower proportion. (Bottom) The best proportion among the proportions  $\{q = 0.05m\}_{m=1}^{20}$  of the S-MAP and the IS-MAP procedures in terms of the weighted risk in (48) is presented versus the weighting coefficient  $c$ . For both procedures, as  $c$  increases the best proportion value decreases, or stays the same in a few cases. Thus, as we put a higher weight on the ANO compared to the ADD we should usually choose a lower proportion of data streams to observe for achieving a lower weighted risk. In addition, there is a rapid decrease in the optimal proportion from  $q = 1$  to  $q = 0.45$  (S-MAP) and  $q = 0.4$  (IS-MAP), as we change  $c$  from 0 to 0.1 and thus, it can be seen that even a small positive weight on the ANO leads to a small proportion value,  $q < 0.5$ , for which the lowest weighted risk is obtained among the considered proportions.

multiple change-points detection. Under the null hypothesis the  $p$ -value is uniformly distributed on  $[0, 1]$  and thus, we set  $f_0 = U(0, 1)$ . Usually, under the alternative hypothesis the  $p$ -value follows a distribution that has high density for small  $p$ -values and the density decreases as the  $p$ -values increase towards 1 [36], [38]. A commonly assumed distribution for the  $p$ -value under the alternative hypothesis is the beta distribution [36]–[38]. Therefore, we set  $f_1^{(b_k)} = \beta(1, b_k)$ , i.e.  $f_1^{(b_k)}(X) \propto$

$(1 - X)^{b_k - 1}$ ,  $X \in [0, 1]$ , and  $f_1^{(b_k)}(X) = 0$ ,  $X \notin [0, 1]$ , where  $b_k$  is a parameter of the  $k$ th data stream probability density under the alternative hypothesis,  $\forall k \in [K]$ . For each sensor, We consider uncertainty in the value of the parameter  $b_k$ , where it is only known that  $b_k \in [b_{\min}, b_{\max}]$ ,  $\forall k \in [K]$ , and  $b_{\min}, b_{\max}$  are known. The true and unknown value of  $b_k$  for each sensor is set by randomly choosing a number in the interval  $[b_{\min}, b_{\max}]$ .

Due to the uncertainty in  $f_1^{(b_k)}$ ,  $k \in [K]$ , we implement all the procedures in this example with a generalized LR (GLR),

$$L_G(X) = \frac{\max_{b \in [b_{\min}, b_{\max}]} f_1^{(b)}(X)}{f_0(X)}$$
, instead of the actual LR, where we set  $b_{\min} = 10$  and  $b_{\max} = 20$ . For each data stream, given the observation we compute the corresponding GLR and use its value instead of the unknown LR. The probability densities,  $f_0$  and  $f_1$  with  $b = b_{\min} = 10$  and  $b = b_{\max} = 20$ , are depicted in Fig. 6. It should be noted that since the true  $f_1^{(b_k)}$ ,  $k \in [K]$ , is smaller than or equal to  $\max_{b \in [b_{\min}, b_{\max}]} f_1^{(b)}$ , the true LR is smaller than the implemented GLR and thus, the resulting FDR may be higher than the predefined upper bound.

We perform similar simulations as in Subsection VI-A. For  $K = 10, 100, 200, 500, 1000$ , we examine the FDR values of the proposed S-MAP and IS-MAP procedures with different proportions  $\{q = 0.05m\}_{m=1}^{20}$ . The resulting minimum and maximum estimated FDR values of the S-MAP procedure are 0.034 and 0.059, respectively. The resulting minimum and maximum estimated FDR values of the IS-MAP procedure are 0.064 and 0.102, respectively. Consequently, due to the model uncertainty and the maximization of  $f_1^{(b)}$  w.r.t.  $b \in [b_{\min}, b_{\max}]$ , some of the resulting FDR values of the IS-MAP procedure are slightly higher than  $\alpha = 0.1$ . This result demonstrates that since the S-MAP procedure is more conservative than the IS-MAP procedure in terms of FDR control then, the S-MAP procedure can be viewed as more robust than the IS-MAP procedure under the assumed model uncertainty.

**Remark 1.** In order to attempt to still maintain the FDR control of the IS-MAP procedure under the desired upper bound, we also implement it with  $\rho_{\text{sim}} = 0.005$ , which is lower than the true value,  $\rho = 0.01$ , under which the random change points are generated. As previously explained, in this case the FDR of the IS-MAP procedure will be lower at the expense of higher ADD. The resulting minimum and maximum estimated FDR values of the IS-MAP procedure are 0.035 and 0.056, respectively. Thus, all the IS-MAP estimated FDR values are below the predefined upper bound and FDR control is maintained. In addition, it can be seen that alternating the value of  $\rho_{\text{sim}}$  compared to the true  $\rho$  is a tool for controlling the tradeoff between FDR and ADD in case of model uncertainty.

In Fig. 7, we evaluate the ADD of the procedures: D-FDR, S-MAP with  $q = 0.5, 1$ , simple procedure with  $q = 0.5$ , and IS-MAP with  $q = 0.5, 1$  versus  $K = 10, 100, 200, 500, 1000$ . It can be seen that under the model uncertainty, all the considered procedures still have an approximately constant ADD as  $K$  increases, which is in accordance with the analytical results in Section V. The parallel version of the IS-MAP procedure has the lowest ADD. In addition, the

IS-MAP procedure with  $q = 0.5$  outperforms the parallel version of the S-MAP procedure and the D-FDR procedure, demonstrating the advantage of using the IS-MAP procedure rather than the S-MAP or the D-FDR procedures in terms of ADD. The simple procedure with  $q = 0.5$  has the highest ADD among the considered procedures. Thus, even under the model uncertainty, there is an advantage in terms of ADD in using the proposed MAP approach for choosing the monitored sensors rather than randomly choosing the subset of sensors to monitor. In Fig. 8, we evaluate the ANO versus  $K$  of the procedures: D-FDR, S-MAP with  $q = 0.5, 1$ , and IS-MAP with  $q = 0.5, 1$ . It can be seen that IS-MAP with  $q = 0.5$  has the lowest ANO, whereas the D-FDR has the highest one. In all the considered procedures, the ANO is approximately a constant w.r.t.  $K$ .

In the upper and middle plots of Fig. 9, we plot the ADDs and ANOs, respectively, of the S-MAP and the IS-MAP procedures for  $K = 1000$  versus the proportion values  $\{q = 0.05m\}_{m=1}^{20}$ . It can be seen that for any of the considered proportions, the IS-MAP procedure has lower ADD and ANO than the S-MAP procedure. In addition, for both procedures the ADD decreases as the proportion increases, while the ANO increases as the proportion increases. Similar to the previous example, it can be seen that there is no significant increase in ADD when the proportion decreases from  $q = 1$  to  $q = 0.3$ . The ANO increases significantly as  $q$  increases towards 1. In the lower plot of Fig. 9, we plot a curve connecting the ADD-ANO points of the S-MAP and the IS-MAP procedures from the upper and middle plots of Fig. 9. It can be seen that under the model uncertainty we still have a clear tradeoff between the ADD and ANO and the ADD decreases as the ANO increases.

In the upper plot of Fig. 10, we compare the weighted risks from (48) of the S-MAP and the IS-MAP procedures with proportions  $\{q = 0.05m\}_{m=1}^{20}$  versus the proportion size for  $c = 0.2$ . It can be seen that the weighted risk of the IS-MAP procedure is lower than the weighted risk of the S-MAP procedure. For both the S-MAP and the IS-MAP procedures, the best tradeoff among the considered proportions is achieved with the proportion  $q = 0.2$ . Thus, under the model uncertainty, it is still not desirable to monitor all the active data streams in parallel, when both ADD and ANO are taken into account. In the lower plot of Fig. 10, for both the S-MAP and the IS-MAP procedures, we present the best proportion among the proportions  $\{q = 0.05m\}_{m=1}^{20}$  in terms of the weighted risk in (48) versus the weighting coefficient  $c$ . Similarly to the previous example, for both procedures, as we increase  $c$  the best proportion value decreases or does not increase. We also noticed a rapid decrease in the optimal proportion from  $q = 1$  to  $q = 0.25$ , as we change  $c$  from 0 to 0.1.

## VII. CONCLUSION

In this paper, we developed methods for Bayesian multiple change-point detection in sensor network with limitations on the proportion of sensors that can be monitored in parallel. We proposed the S-MAP detection procedure in which observations are received only from a subset of sensors with highest posterior probabilities of change points having oc-

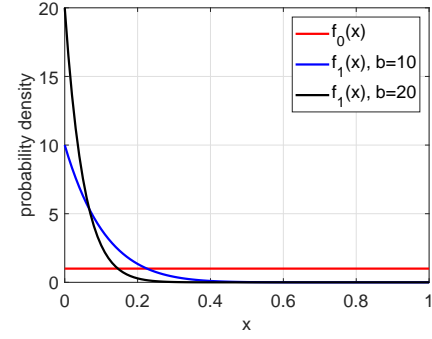


Fig. 6. General model with uncertainty and known  $p$ -values: Probability density  $f_0 = U(0, 1)$  that corresponds to the distribution of a  $p$ -value under the null hypothesis,  $f_1 = \beta(1, b)$ ,  $b = b_{\min} = 10$ ,  $b = b_{\max} = 20$ . The beta distribution is a common assumption for a  $p$ -value under the alternative hypothesis. We assume that  $b \in [b_{\min}, b_{\max}]$  is unknown and depict the corresponding probability densities with the lowest possible value of  $b$ ,  $b_{\min} = 10$ , and the highest possible value of  $b$ ,  $b_{\max} = 20$ .

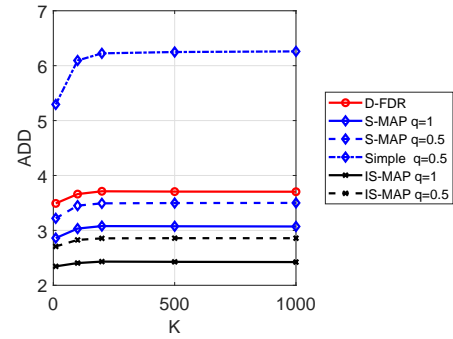


Fig. 7. General model with uncertainty and known  $p$ -values: The ADD of the procedures D-FDR, S-MAP with  $q = 0.5, 1$ , simple, and IS-MAP with  $q = 0.5, 1$ , versus  $K = 10, 100, 200, 500, 1000$ . It can be seen that even under the model uncertainty all the considered procedures have an approximately constant ADD as  $K$  increases. The parallel version of the IS-MAP procedure has the lowest ADD. Similarly to the previous example, the IS-MAP procedure with  $q = 0.5$  outperforms the parallel version of the S-MAP procedure and the D-FDR procedure. The simple procedure with  $q = 0.5$  has the highest ADD. Thus, even under model uncertainty, the MAP approach outperforms a random choice approach for choosing the subset of sensors to monitor.

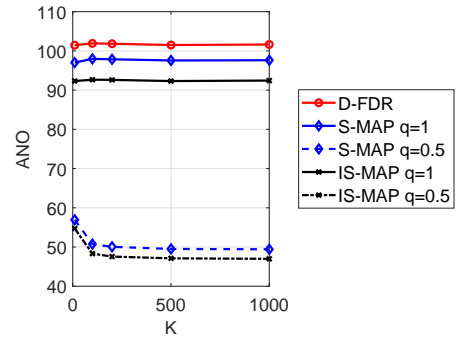


Fig. 8. General model with uncertainty and known  $p$ -values: The ANO of the procedures D-FDR, S-MAP with  $q = 0.5, 1$ , and IS-MAP with  $q = 0.5, 1$ , versus  $K$ . Under the model uncertainty, all the considered procedures still have an approximately constant ANO w.r.t.  $K$ , similarly to the previous example in which there is no model uncertainty. The IS-MAP with  $q = 0.5$  has the lowest ANO, while the D-FDR has the highest ANO.

curred, within the allowed proportion. In addition, we proposed an improved procedure named the IS-MAP procedure that requires lower thresholds than the S-MAP procedure and

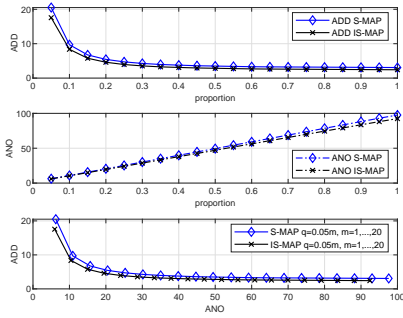


Fig. 9. General model with uncertainty and known  $p$ -values: (Top and middle) The ADDs and ANOs of the S-MAP and the IS-MAP procedures for  $K = 1000$  versus the proportion values  $\{q = 0.05m\}_{m=1}^{20}$ . The IS-MAP procedure has lower ADD and ANO compared to the S-MAP procedure. For both procedures the ADD decreases as the proportion increases, while the ANO increases approximately linearly as the proportion increases. There is no significant increase in ADD when the proportion decreases from  $q = 1$  to  $q = 0.3$ , while the decrease in ANO is more significant. (Bottom) A curve connecting the ADD-ANO points of the S-MAP and the IS-MAP procedures from the upper and middle plots of this figure. Under the model uncertainty there is still a clear tradeoff between the ADD and ANO, i.e. as the ADD decreases the ANO increases.

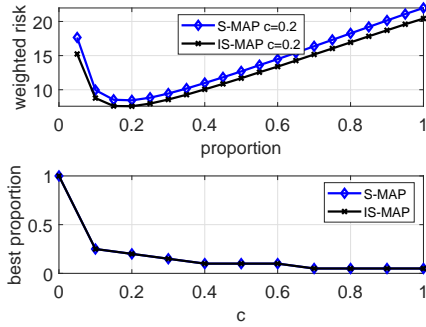


Fig. 10. General model with uncertainty and known  $p$ -values: (Top) The weighted risks of the S-MAP and the IS-MAP procedures with proportions  $\{q = 0.05m\}_{m=1}^{20}$  versus the proportion size for  $c = 0.2$ . The weighted risk of the IS-MAP procedure is lower than the weighted risk of the S-MAP procedure. For both the S-MAP and the IS-MAP procedures, the best tradeoff among the considered proportions is achieved with the proportion  $q = 0.2$ . Thus, in a similar manner to the previous example, when both the ADD and the ANO are taken into account it is not desirable to monitor all the active data streams in parallel and a lower weighted risk can be attained with a much lower proportion. (Bottom) The best proportion among the proportions  $\{q = 0.05m\}_{m=1}^{20}$  of the S-MAP and the IS-MAP procedures in terms of the weighted risk in (48) is presented versus the weighting coefficient  $c$ . In this case, the S-MAP and the IS-MAP procedures have the same best proportions. Similar to previous example, as  $c$  increases the best proportion value, among the considered proportions, decreases or does not increase. A rapid decrease is noticed in the optimal proportion from  $q = 1$  to  $q = 0.25$ , as we change  $c$  from 0 to 0.1.

attains lower ADD and ANO. It has been shown that both the proposed procedures control the FDR at a predefined level and achieve an ADD that asymptotically remains a constant as the number of sensors in the network increases. The S-MAP procedure is more conservative than the IS-MAP procedure in terms of FDR control, and thus, in the FDR control sense, the S-MAP procedure is more robust to model uncertainty than the IS-MAP procedure. In the simulations, we have first considered i.i.d. Gaussian observations with a change in the mean and then we have considered a general model with some model uncertainty in which  $p$ -values from each sensor are used as observations to perform the change-

points detection task. Our simulations in both cases show that the proposed S-MAP and IS-MAP procedures achieve a practically constant ADD as the number of sensors increases. The S-MAP procedure outperforms a corresponding simple procedure in terms of ADD demonstrating the benefit of the MAP approach compared to randomly choosing the subset of sensors to monitor. We have also used the S-MAP and IS-MAP procedures to study the tradeoff between ADD and ANO in multiple change-point detection. Under a joint weighted risk on the ADD and ANO with a positive weight on both figures of merit, we found that in all the considered cases observing all the data streams, i.e. setting  $q = 1$ , does not provide the best tradeoff between the ADD and ANO. In fact, the best tradeoff can be obtained with proportion  $q \ll 1$ , which implies that setting a small proportion, e.g.  $q = 0.3$ , can significantly reduce the communication burden, i.e. the ANO, while maintaining a low ADD. A Topic for future research is the derivation of novel procedures with FDR control capabilities for non-parametric [39] multiple change-point detection under communication limitations.

## APPENDIX A PROOF OF THEOREM 2

In this appendix, the FDR control of the IS-MAP procedure is proved. The number of change points declared,  $R$ , is known given the filtration of all the data,  $\mathcal{F}_\infty$ . Thus, using the law of total expectation, we can rewrite the FDR from (4) as

$$\text{FDR} = \mathbb{E} \left[ \frac{\mathbb{E}[V|\mathcal{F}_\infty]}{\max\{R, 1\}} \right]. \quad (49)$$

Recall that  $V$  is the number of false discoveries, i.e. the size of the subset of  $[K]$  s.t.  $T^{(k)} < t^{(k)}$ . Thus,  $V$  can be written as

$$V = \sum_{k=1}^K \mathbf{1}_{\{t^{(k)} > T^{(k)}\}}, \quad (50)$$

where  $\mathbf{1}_A$  is the indicator function of the event  $A$ . By substituting (50) in (49) and using the linearity of the expectation operator, we obtain

$$\begin{aligned} \text{FDR} &= \mathbb{E} \left[ \frac{\sum_{k=1}^K \mathbb{E}[\mathbf{1}_{\{t^{(k)} > T^{(k)}\}}|\mathcal{F}_\infty]}{\max\{R, 1\}} \right] \\ &= \mathbb{E} \left[ \frac{\sum_{k=1}^K \mathbb{E}[\mathbf{1}_{\{t^{(k)} > T^{(k)}\}}|\mathcal{F}_{T^{(k)}}]}{\max\{R, 1\}} \right], \end{aligned} \quad (51)$$

where the second equality is obtained since the stopping times,  $\{T^{(k)}\}_{k \in [K]}$ , are known given  $\mathcal{F}_\infty$  and for  $T^{(k)} < \infty$  we stop observing the  $k$ th data stream after  $T^{(k)}$ , i.e. after change point declaration for the  $k$ th data stream. Rewriting the expected indicator functions in (51) as conditional probabilities, we obtain

$$\begin{aligned} \text{FDR} &= \mathbb{E} \left[ \frac{\sum_{k=1}^K (1 - P(t^{(k)} \leq T^{(k)}|\mathcal{F}_{T^{(k)}}))}{\max\{R, 1\}} \right] \\ &= \mathbb{E} \left[ \frac{\sum_{k=1}^K (1 - \pi_{T^{(k)}}^{(k)})}{\max\{R, 1\}} \right], \end{aligned} \quad (52)$$

where the second equality is obtained by substituting (1) into the first equality. In case  $T^{(k)} = \infty$ , then  $\pi_{T^{(k)}}^{(k)} = 1$  and thus,

$$1 - \pi_{T^{(k)}}^{(k)} = 0, \forall k \in [K], \text{ s.t. } T^{(k)} = \infty. \quad (53)$$

On the other hand, in case  $T^{(k)} < \infty$  then, at time slot  $T^{(k)}$  the event  $\{\pi_{T^{(k)}}^{(k)} \geq Q\}$  occurs. Consequently,

$$1 - \pi_{T^{(k)}}^{(k)} \leq 1 - Q = \alpha, \forall k \in [K], \text{ s.t. } T^{(k)} < \infty, \quad (54)$$

where the last equality is obtained by substituting  $Q = 1 - \alpha$  from (12) into the term  $1 - Q$ . The term  $R$  is the cardinality of the subset of  $[K]$  s.t.  $T^{(k)} < \infty$ . Thus, by substituting (53)-(54) in (52), one obtains

$$\text{FDR} \leq \mathbb{E} \left[ \frac{\max\{R, 1\}\alpha}{\max\{R, 1\}} \right] = \alpha. \quad (55)$$

#### APPENDIX B PROOF OF THEOREM 3

In this appendix, we derive asymptotic lower and upper bounds on the ADD of the S-MAP and the IS-MAP procedure. For any data stream, the lowest possible threshold of the S-MAP procedure from (9) is  $Q_K = 1 - \alpha$ , i.e. change point cannot be declared before the posterior probability is higher than or equal to  $1 - \alpha$ . Thus, from (17)

$$\text{ADD}_{\text{S-MAP},k} \geq \frac{|\log \alpha|}{D(f_1||f_0) + |\log(1 - \rho)|} (1 + o_\alpha(1)), \quad (56)$$

$\forall k \in [K]$ . It can be seen that the asymptotic lower bound in (56) is independent of  $k$ . Thus, by substituting (56) in (6), we obtain (19).

According to the S-MAP procedure we can find a threshold for the  $k$ th data stream,  $Q_{r_k} = 1 - \frac{r_k \alpha}{K}$ ,  $r_k \in [K]$ , which is different from the thresholds of the other data streams. For this threshold, the change of the  $k$ th data stream is declared at the first time slot in which this threshold is exceeded or even before this threshold is exceeded. Thus, from (18),

$$\text{ADD}_{\text{S-MAP},k} \leq \frac{\log \frac{K}{r_k \alpha}}{|\log(1 - \rho)|} (1 + o_\alpha(1)), \forall k \in [K]. \quad (57)$$

By substituting (57) in (6), we obtain

$$\text{ADD}_{\text{S-MAP}} \leq \left( \frac{1}{K} \sum_{k=1}^K \frac{\log \frac{K}{r_k \alpha}}{|\log(1 - \rho)|} \right) (1 + o_\alpha(1)). \quad (58)$$

Since the thresholds are different, we obtain

$$\sum_{k=1}^K \log r_k = \sum_{k=1}^K \log k = \log K!. \quad (59)$$

By substituting (59) into (58) and reordering, we obtain (20).

In the IS-MAP procedure, for any data stream the threshold is  $Q = 1 - \alpha$  from (12). Thus, using (17) and (18), we obtain

$$\text{ADD}_{\text{IS-MAP},k} \geq \frac{|\log \alpha|}{D(f_1||f_0) + |\log(1 - \rho)|} (1 + o_\alpha(1)) \quad (60)$$

and

$$\text{ADD}_{\text{IS-MAP},k} \leq \frac{|\log \alpha|}{|\log(1 - \rho)|} (1 + o_\alpha(1)), \quad (61)$$

respectively,  $\forall k \in [K]$ . The asymptotic lower and upper bounds in (60) and (61), respectively, are independent of  $k$  and thus, by substituting (60) and (61) in (6), we obtain (21) and (22), respectively.

#### APPENDIX C PROOF OF PROPOSITION 4

In this appendix, we derive the asymptotic ADD upper bound from (30) under the assumption that (27)-(29) are satisfied. Using the definition of  $\gamma$  from (26), it can be seen that the prior distribution of  $\gamma \in \mathbb{N}$  is

$$\begin{aligned} P(\gamma = m) &= P(V_{m-1} < t \leq V_m) \\ &= P(t \leq V_m) - P(t \leq V_{m-1}). \end{aligned} \quad (62)$$

Under the geometric prior assumption on  $t$  we obtain

$$P(t \leq m) = 1 - (1 - \rho)^m, m \in \mathbb{N}. \quad (63)$$

By substituting (63) in (62), one obtains

$$P(\gamma = m) = (1 - \rho)^{V_{m-1}} - (1 - \rho)^{V_m}. \quad (64)$$

Using (64), we obtain

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{-\log P(\gamma \geq m+1)}{m} &= \lim_{m \rightarrow \infty} \frac{-\log((1 - \rho)^{V_m})}{m} \\ &= \left( \lim_{m \rightarrow \infty} \frac{V_m}{m} \right) |\log(1 - \rho)| \\ &= \zeta |\log(1 - \rho)|, \end{aligned} \quad (65)$$

where the third equality is obtained by substituting (24) and (28) into the second equality. Using the definition of  $\gamma$  from (26), we obtain that on  $\{\gamma = n\}$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=n}^{n+N-1} \log L(X_{V_i}) = D(f_1||f_0). \quad (66)$$

From the definitions of the stopping rule,  $\Gamma$ , and the change point,  $\gamma$ , in (25) and (26), respectively, and from (65) and (66), it can be seen that the detection of  $\gamma$  using  $\Gamma$  based on the sequence  $\{X_{V_n}\}_{n=1}^\infty$  is a Bayesian change-point detection procedure that satisfies the conditions of Theorem 3 in [12]. Thus, using this Theorem, we obtain the following asymptotic upper bound on the ADD of  $\Gamma$ :

$$\mathbb{E}[\max\{0, \Gamma - \gamma\}] \leq \frac{|\log \eta|}{D(f_1||f_0) + \zeta |\log(1 - \rho)|} (1 + o_\eta(1)). \quad (67)$$

Next, we consider the stopping rule

$$T^* = \inf\{V_n, n \in \mathbb{N} : \pi_{V_n} \geq 1 - \eta\} = V_\Gamma. \quad (68)$$

In a similar manner to  $T$ , the stopping rule  $T^*$  uses the posterior update from (15), but can only take values from the subsequence  $\{V_n\}_{n=1}^\infty$  rather than  $\mathbb{N}$ . Therefore,  $T \leq T^*$  and consequently

$$T - t \leq T^* - t = V_\Gamma - V_\gamma + V_\gamma - t, \quad (69)$$

where the equality follows from (68). From (23) and (26) we obtain that

$$V_\gamma - t \leq \zeta_\gamma - 1. \quad (70)$$



In addition, using (24) we can write

$$V_\Gamma = \Gamma \zeta^{(\Gamma)} \text{ and } V_\gamma = \gamma \zeta^{(\gamma)}. \quad (71)$$

By substituting (70)-(71) into the right hand side of (69), one obtains

$$\begin{aligned} T - t &\leq \zeta^{(\Gamma)}(\Gamma - \gamma) + \gamma(\zeta^{(\Gamma)} - \zeta^{(\gamma)}) + \zeta_\gamma - 1 \\ &\leq \zeta^{(\Gamma)}(\Gamma - \gamma) + |\gamma(\zeta^{(\Gamma)} - \zeta^{(\gamma)}) + \zeta_\gamma - 1|. \end{aligned} \quad (72)$$

Using (23) and (27), we obtain

$$1 \leq \zeta^{(N)} \leq \mathcal{B}, \quad \forall N \in \mathbb{N}. \quad (73)$$

Substituting (27) and (73) in (72), one obtains

$$T - t \leq \zeta^{(\Gamma)}(\Gamma - \gamma) + \gamma(\mathcal{B} - 1) + \mathcal{B} - 1. \quad (74)$$

From (64), it can be verified that

$$\mathbb{E}[\gamma] \leq \frac{1}{\rho}. \quad (75)$$

By using (29), (67), (74) and (75), we obtain that the ADD of  $T$  satisfies

$$\text{ADD} \leq \frac{\zeta |\log \eta|}{D(f_1 || f_0) + \zeta |\log(1 - \rho)|} (1 + o_\eta(1)) \quad (76)$$

and consequently (30) is obtained.

## REFERENCES

- [1] O. Hadjiladis, H. Zhang, and H. V. Poor, "One shot schemes for decentralized quickest change detection," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3346–3359, July 2009.
- [2] S. Chaudhari, V. Koivunen, and H. V. Poor, "Autocorrelation-based decentralized sequential detection of OFDM signals in cognitive radios," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2690–2700, July 2009.
- [3] J. Chen, W. Zhang, and H. V. Poor, "Non-Bayesian multiple change-point detection controlling false discovery rate," in *Proc. of the IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 31–35.
- [4] —, "On parallel sequential change detection controlling false discovery rate," in *Proc. of the 50th Asilomar Conference on Signals, Systems and Computers*, Nov. 2016, pp. 107–111.
- [5] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.
- [6] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability & Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.
- [7] G. Lorden, "Procedures for reacting to a change in distribution," *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.
- [8] M. Pollak, "Optimal detection of a change in distribution," *The Annals of Statistics*, vol. 13, no. 1, pp. 206–227, 1985.
- [9] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *The Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, 1986.
- [10] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2917–2929, 1998.
- [11] H. V. Poor and O. Hadjiladis, *Quickest Detection*. Cambridge University Press, 2008.
- [12] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *Theory of Probability & Its Applications*, vol. 49, no. 3, pp. 458–497, 2005.
- [13] A. G. Tartakovsky, "Asymptotic optimality in Bayesian changepoint detection problems under global false alarm probability constraint," *Theory of Probability & Its Applications*, vol. 53, no. 3, pp. 443–466, 2009.
- [14] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.
- [15] B. Efron, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge University Press, 2012, vol. 1.
- [16] A. Farcomeni, "A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion," *Statistical Methods in Medical Research*, vol. 17, no. 4, pp. 347–388, 2008.
- [17] J. Chen, W. Zhang, and H. V. Poor, "An FDR-oriented approach to multiple sequential fault detection and isolation," in *Proc. of the 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct. 2017, pp. 629–633.
- [18] K. Premkumar and A. Kumar, "Optimal sleep-wake scheduling for quickest intrusion detection using wireless sensor networks," in *Proc. of the 27th IEEE Conference on Computer Communications (INFOCOM)*, Apr. 2008, pp. 1400–1408.
- [19] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection with on-off observation control," *Sequential Analysis*, vol. 31, no. 1, pp. 40–77, 2012.
- [20] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection in minimax settings," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6917–6931, Oct. 2013.
- [21] —, "Data-efficient minimax quickest change detection with composite post-change distribution," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 5172–5184, Sep. 2015.
- [22] T. Banerjee and V. V. Veeravalli, "Data-efficient minimax quickest change detection in a decentralized system," *Sequential Analysis*, vol. 34, no. 2, pp. 148–170, 2015.
- [23] Y. Xie, J. Huang, and R. Willett, "Change-point detection for high-dimensional time series with missing data," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 1, pp. 12–27, Feb 2013.
- [24] J. Geng and L. Lai, "Non-Bayesian quickest change detection with stochastic sample right constraints," *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 5090–5102, Oct. 2013.
- [25] J. Geng, E. Bayraktar, and L. Lai, "Bayesian quickest change-point detection with sampling right constraints," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6474–6490, Oct. 2014.
- [26] X. Ren, K. H. Johansson, and L. Shi, "Quickest change detection with observation scheduling," *IEEE Trans. Autom. Control*, vol. 62, no. 6, pp. 2635–2647, June 2017.
- [27] X. Ren, K. H. Johansson, D. Shi, and L. Shi, "Quickest change detection in adaptive censoring sensor networks," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 1, pp. 239–250, Mar. 2018.
- [28] J. Geng and L. Lai, "Quickest change-point detection over multiple data streams via sequential observations," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Apr. 2018, pp. 4404–4408.
- [29] T. Halme, E. Nitzan, H. V. Poor, and V. Koivunen, "Bayesian multiple change-point detection with limited communication," *Accepted to the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2020.
- [30] E. Nitzan, T. Halme, H. V. Poor, and V. Koivunen, "Deterministic multiple change-point detection with limited communication," *Accepted to the Conference on Information Sciences and Systems (CISS)*, Mar. 2020.
- [31] P. Moulin and V. V. Veeravalli, *Statistical Inference for Engineers and Data Scientists*. Cambridge University Press, 2018.
- [32] J. Sándor and L. Debnath, "On certain inequalities involving the constant  $e$  and their applications," *Journal of Mathematical Analysis and Applications*, vol. 249, no. 2, pp. 569–582, 2000.
- [33] H. Robbins, "A remark on Stirling's formula," *The American Mathematical Monthly*, vol. 62, no. 1, pp. 26–29, 1955.
- [34] T. L. Bailey and M. Gribskov, "Combining evidence using p-values: application to sequence homology searches," *Bioinformatics (Oxford, England)*, vol. 14, no. 1, pp. 48–54, 1998.
- [35] B. Efron, "Large-scale simultaneous hypothesis testing: the choice of a null hypothesis," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 96–104, 2004.
- [36] S. Pounds and S. W. Morris, "Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values," *Bioinformatics*, vol. 19, no. 10, pp. 1236–1242, 2003.
- [37] T. Halme, M. Glz, and V. Koivunen, "Bayesian multiple hypothesis testing for distributed detection in sensor networks," in *Proc. of the IEEE Data Science Workshop (DSW)*, June 2019, pp. 105–109.
- [38] N. A. Heard and P. Rubin-Delanchy, "Choosing between methods of combining p-values," *Biometrika*, vol. 105, no. 1, pp. 239–246, 2018.
- [39] T. S. Lau, W. P. Tay, and V. V. Veeravalli, "A binning approach to quickest change detection with unknown post-change distribution," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 609–621, Feb. 2019.