# Bayesian Quickest Detection of Propagating Spatial Events

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

***Abstract*—Rapid detection of spatial events that propagate across a sensor network is of wide interest in many modern applications. In particular, in communications, radar, IoT, environmental monitoring, and biosurveillance, we may observe propagating fields or particles. In this paper, we propose Bayesian sequential single and multiple change-point detection procedures for the rapid detection of such phenomena. Using a dynamic programming framework we derive the structure of the optimal single-event quickest detection procedure, which minimizes the average detection delay (ADD) subject to a false alarm probability upper bound. The multi-sensor system configuration is arbitrary and sensors may be mobile. In the rare event regime, the optimal procedure converges to a more practical threshold test on the posterior probability of the change point. A convenient recursive computation of this posterior probability is derived by using the propagation characteristics of the spatial event. The ADD of the posterior probability threshold test is analyzed in the asymptotic regime, and specific analysis is conducted in the setting of detecting random Gaussian signals affected by path loss. Then, we show how the proposed procedure is easy to extend for detecting multiple propagating spatial events in parallel in a multiple hypothesis testing setting. A method that provides strict false discovery rate (FDR) control is proposed. In the simulation section, it is demonstrated that exploiting the spatial properties of the event decreases the ADD compared to procedures that do not utilize this information, even under model mismatch.**

***Index Terms*—Sensor network, Bayesian spatial change-point detection, change propagation, average detection delay, false discovery rate, multiple hypothesis testing.**

## I. Introduction

SEQUENTIAL change-point detection, often referred to as quickest detection, is a fundamental statistical inference task [1], [2], [3], [4], [5], [6], [7], [8]. It is encountered in numerous applications, such as Internet of Things, environmental monitoring, biosurveillance, finance, radar, and wireless communications. Sensor networks are commonly used to rapidly detect a disruption or an event in the monitored physical environment [9], [10], [11], [12], [13], [14], [15]. Usually, in these sensor networks the sensors communicate with a fusion center (FC) or a cloud that performs statistical inference tasks based on the data or local statistics from the sensors. The network setting can be centralized [9], [10] where the FC has access to all the data from the sensors or decentralized [11], [12], [13], [14], [15], [16], [17] where the sensors perform local computations/inferences and may only send the results, e.g. some sufficient statistic, to the FC. Recently, quickest detection of multiple change points in parallel has gained wide interest [10], [18], [19], [20], [21]. Parallel multiple change points can be caused, for example, by multiple active radio transmitters, multiple sound sources, multiple radar targets, or multiple emitters of polluting particles.

In many cases, the event causing the change in the environment has some spatial properties. The event can be a moving target that appears in a surveillance system, propagating radio frequency or audio signals impinging distributed sensors, pollution emanating from a malfunctioning device, or the onset of an epidemic. The effect of the spatial event on a specific area of the network can be modeled in different ways depending on the underlying physical phenomenon of interest and parameters of the sensor system observing it, such as the sensor system configuration and sampling rate. For example, the event may instantaneously affect the sampling distributions of all the sensors in the vicinity of the event [8], [11]. Alternatively, the disruption may propagate across the sensors in the network or in some cluster of the network over a short time period [9], [14], [22]. Examples are propagation of polluting particles in environmental monitoring, seismic activity in earthquake monitoring, propagation of radio waves through space in communication or radar systems, and epidemic traveling wave in biosurveillance.

Several works have considered quickest detection while incorporating spatial information. In [9], Bayesian quickest detection was considered where the sensors are numbered and located with regular geometry and uniform displacements. The initial origin of the disruption was known to be at the first sensor and the disruption propagates through all the sensors as a Markov process in an order determined by the numbering. Extension of [9] to the case where the first sensor experiencing the initial change is unknown was proposed in [23]. Given the sensor observing the disruption first, a predetermined change propagation trajectory was assumed across the sensors. The work in [14], [15][1] considered a similar setting to [9], [23], with the difference that the change propagation pattern is assumed to be unknown. In addition, both

Manuscript received 12 May 2022; revised 22 November 2022; accepted 30 November 2022. Date of publication 19 December 2022; date of current version 3 January 2023. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Augusto Aubry. This work was supported by the Academy of Finland projects: (1) Statistical Signal Processing Theory and Computational Methods for Large Scale Data Analysis (2) WiFiUS project: Secure Inference in the Internet of Things. *(Corresponding author: Topi Halme.)*

The authors are with the Department of Signal Processing and Acoustics, Aalto University, 02150 Espoo, Finland (e-mail: topi.halme@aalto.fi; eyalnitzan1@gmail.com; visa.koivunen@aalto.fi).

Digital Object Identifier 10.1109/TSP.2022.3230334

[1]The basic idea of a change that propagates through an array with an unknown first sensor was considered in V. Raghavan, "Multi-sensor Change Propagation: Near-optimal Procedures and Applications," Proc. Intern. Workshop on Seq. Methodologies, Rouen, France, 2017.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

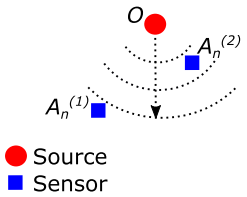



Fig. 1. Source wavefront propagation: Phenomenon emanating from the source $O$ towards the sensors $A_n^{(1)}$ and $A_n^{(2)}$ where $n$ is the current time slot.

centralized and decentralized settings were considered. In [22], Bayesian continuous-time single change-point detection with sensor networks was studied. The event was assumed to occur at a random time instant in a random location and gradually propagate through the sensor network with unknown velocity triggering interdependent change points. A numerical procedure was proposed to approximate the optimal but intractable Bayesian solution based on the approximated posterior probabilities. Non-Bayesian change-point detection with spatial information under different setups has been studied in [24], [25], [26], [27], [28], [29], [30], [31], [32], [33]. In [27], a setting where a moving anomaly affects one sensor at a time was considered. As the disruption moves over the network, the affected sensor changes with time according to either a known or an unknown probability model. Rapid detection of propagating phenomena has recently also been studied within the learning and adaptation framework [34]. In [34], a fully-flat network without a central unit is deployed for monitoring. Sensors must exchange information with their neighbors in order to accurately estimate the true state of nature in their vicinity. Due to the more complex network communication topology and interaction and information exchange among the neighboring sensor nodes, obtaining strong performance guarantees in terms of detection delay and false alarm rate, as is the goal under a quickest detection formulation, is difficult. Moreover, the delays in exchanging information among sensors may be significant compared to the propagation speed of the monitored phenomenon. Nonetheless, an information diffusion scheme that results in each sensor detecting changes faster than sensors working in isolation was derived.

In this paper, we propose a Bayesian method for the detection of a propagating spatial event. A discrete-time model is used in acquiring observations. It is assumed that the event propagates in a two-dimensional plane with area radius that increases either randomly or deterministically based on the laws of physics. This wavefront propagation model is illustrated in Fig. 1. It is relevant, for example, in wireless communications where radio waves that carry information are emitted from a transmitter, propagate through space with velocity equal to the speed of light, and detected by a receiver [35, Ch. 3]. The considered propagation model and change detection setup are also relevant in radar and seismic monitoring and localization applications where, for example, a point source or a target is generating or reflecting waveforms that propagate across a sensor array as a plane wave [36], [37]. Another application is in biosurveillance applications that attempt to detect outbreaks of epidemics [22], [38], [39], [40], [41], [42].

At each time slot, sensors that are located outside the disruption area obtain observations that follow a common null distribution (no signal present). Sensors that are located within the disruption area obtain observations that obey alternative distributions that may be different among the exposed sensors. We are interested in detecting the initial event as quickly as possible subject to statistical constraints on the rate of false alarms. We assume that the number of sensors and their locations are known at each time slot but may change in time, e.g. a mobile wireless sensor network [43] where the sensors might correspond to e.g. smart phones or drones. Since the main focus of this work is on detection of spatially localized events, it is assumed that all sensors are able to communicate individually with a common FC or cloud. The FC could be a base station (BS) serving different users and sensors in its coverage area. Modern wireless systems such as 5G have reasonably small coverage areas because of the higher frequencies they are using. Separate mobile access points that traverse the network and collect information from the sensors, as in the SENMA framework [44], [45], [46], are not required. Obviously, there exists a wide area of applications where such access points are useful, however, proving optimality properties may be difficult in settings where both the event and the access point are moving. To reduce the communication burden between the FC and the sensors, ideas from [47] could be considered as a future direction.

The related works described earlier do not explicitly and jointly take into account the sensor locations, the displacement between the sensor location and the location of the disruption source, and potential sensor mobility. In particular, to the best of our knowledge, the problem of quickest detection in discrete time in the practically relevant scenarios where the change is caused by a gradually expanding spatially localized event(s) has not been addressed earlier. The most related previous work [9], [14], [15], [23], [27], [28], [29] focuses on settings where the dynamics of the change-event are modeled as movement from *sensor(s)-to-sensor(s)*. The particular assumed sensor network topology (e.g. an array [9] or a graph [27]) results jointly from the spatial properties of the event, and the placement of the sensors. The event is then assumed to affect the sensors according to a model specified by this topology. In contrast, in this paper we consider potentially mobile sensor systems with completely arbitrary displacements and no regular sensing geometry. Therefore it is not in general possible to describe the propagation of the event with any fixed network topology or model. Whether a particular sensor is affected depends on its location relative to the source of the physical event, and the sensor locations may be arbitrary and change over time. Hence, the previous works cannot directly be applied to quickest detection of spatially phenomena emanating from a distinct and potentially unknown location, especially when sensor locations can vary in time.

The contributions of this paper are:

- We propose a dynamic programming framework for deriving a stopping time that exploits the spatial information and minimizes the average detection delay (ADD) under a constraint on the false alarm probability. An optimal detection procedure that minimizes the associated Bayes risk is derived.
- As the optimal procedure is hard to implement in practice, we propose a more practical detection procedure based on thresholding the posterior probability of the change

point having occurred. To theoretically justify the use of this simpler procedure, it is shown that the procedure is a limiting form of the optimal procedure in the rare event regime. A convenient recursive formula for computing the posterior probability is developed by using the radial change propagation pattern of the spatial event.

- The probability of false alarm (PFA) control of the thresholding procedure is established and its ADD is analyzed in the asymptotic regime. Conditions under which the threshold procedure achieves asymptotic optimality are provided. Furthermore, we analyze the procedure in the specific setting of quickest detection of attenuating random signals. An approximation for asymptotic ADD is provided, and asymptotic optimality is established.
- We extend the single-event detection procedure to the detection of multiple statistically independent spatial events in parallel by combining the developed posterior probability threshold procedure with a Multiple Hypothesis Testing (MHT) setup. The rate of false alarms is controlled using the False Discovery Rate (FDR) criterion, which is widely used for MHT [10], [48], [49]. This is highly relevant in many modern applications where high dimensional data must be processed in parallel and there may be multiple events taking place at the same time [10], [18], [21]. It is shown that the proposed parallel procedure strictly controls the FDR level.
- Simulations are conducted to verify the theoretical findings. It is clearly demonstrated that exploiting the spatial properties of the event decreases the ADD compared to procedures that do not utilize this information in both single and multiple cluster settings. This benefit is achieved even under model mismatch. It is demonstrated that the gain in performance is the largest when the event propagates sufficiently slowly compared to the sampling rate, or when the sensor displacements are large, and/or when the pre- and post-change probability models are different enough.

Preliminary results of this paper appear in conference papers [50] and [51]. This paper is organized as follows. In Section II, we formulate the quickest detection problem. A dynamic programming framework for the detection is formulated in Section III. Under the radial change propagation setup, a change-point detection procedure and its extension to multiple parallel change-point detection are presented in Sections IV and V, respectively. Our simulations and conclusions appear in Sections VI and VII, respectively.

*Notation:* Scalar random variables are denoted by normal font capital letters, with the exception of the change point $t$, which is also a random variable. Scalar constants, such as realizations of random variables, are denoted by normal font lowercase letters, with the exception of $L$, $M$, $N$, $K$ and $\mathfrak{R}$, which are constant integers, and $U_{n,m,r}$ which denotes an event. Boldface uppercase and lowercase letters are used for vector random variables and constants, respectively. For an integer $K$, we use $[K]$ to denote the set $\{0, 1, \ldots, K-1\}$ of cardinality $K$.

## II. Model and Problem Formulation

We begin by describing the model for a single spatial change-point detection problem. The model, relevant terms, and notations for multiple change-point detection in parallel will be described in Section V. Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space, where $\Omega$ is the sample space, $\mathcal{F}$ is the $\sigma$-algebra generated by $\Omega$, and $\mathbb{P}$ is a probability measure. The expectation operator with respect to $\mathbb{P}$ is denoted by $\mathbb{E}$.

At each time slot we have sensors in known locations but with arbitrary configuration within a domain of interest, $\mathcal{S} \subset \mathbb{R}^2$. The set of sensor locations at time slot $n$ is denoted by $\mathcal{A}_n$ and the corresponding number of sensors is $|\mathcal{A}_n|$. If $|\mathcal{A}_n| = 0$, then there are no observations received by the FC at time slot $n$. Unless otherwise stated, in this paper the sensor locations are considered known and deterministic.

We consider a centralized setting where every sensor communicates its observations or local decision statistics to the FC. At time slot $n$, the data transmitted by the sensors and received by the FC are random variables $X_n^{(a)}$, $a \in \mathcal{A}_n$. The realization of $X_n^{(a)}$ is denoted by $x_n^{(a)}$. In mobile scenarios, the location information $a$ is communicated to the FC in addition to the observation value. Alternatively, the FC can have a capability to reliably estimate the locations of the sensors. Uncertainty in the location estimate could be represented as a probability distribution, which could be averaged over in a Bayesian framework. However, for the purposes of this paper we assume for simplicity that reliable point-estimates of the sensor locations exist. We define the $|\mathcal{A}_n| \times 1$ data vector $\mathbf{X}_n$ that contains all the observations transmitted at time slot $n$, including the locations at which the observations were obtained by mobile sensors. It is assumed that at time slot $n$ the FC has access to the current and past observations, $I_n \triangleq (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, where $I_0$ is the empty set.

At a random time instant, $t$, a source becomes active and starts emitting a propagating signal/event from an unknown origin, $O$, causing a disruption in the domain of interest. It is assumed that the initial event time, $t$, has a geometric prior distribution with parameter, $\rho \in (0, 1)$, i.e.

$$\mathbb{P}(t = m) = \rho(1-\rho)^m,\ m \in \mathbb{N}_0, \tag{1}$$

where $\mathbb{N}_0 \triangleq \mathbb{N} \bigcup \{0\}$ and $\mathbb{N}$ is the set of positive integers. The geometric prior distribution is very common in change-point detection because it is a mathematically convenient memoryless distribution, which is also relevant in a variety of practical applications [7], [8], [9], [14].

We want to discover the initial event time, $t$, with minimal delay while controlling the PFA. In a multiple change-point setup, as will be described in Section V, the PFA is replaced by the FDR criterion in a MHT framework. Hereafter, we refer to $t$ as the change point even though it may not cause an instantaneous change in the received observations as in classic change-point detection, due to the fact that the sensors are in distinct locations and displaced from the signal source.

In order to reduce the complexity of the problem, we initially assume that $O \in \mathcal{O} \subset \mathcal{S}$ where $\mathcal{O} \triangleq \{o_0, \ldots, o_{M-1}\}$ is a finite set of possible source or emitter locations in $\mathcal{S}$ with cardinality $|\mathcal{O}| = M$. We assume that the initial event can occur in any of the possible locations, $o_m \in \mathcal{O}$, with equal probability

$$\mathbb{P}(O = o_m) = \frac{1}{M}.$$

The assumption of a uniform probability distribution is made for simplicity and is not necessary for the following derivations.

In practical settings it may not be realistic to expect that the source location can only appear within a known finite set of points. However, choosing the set $\mathcal{O}$ to be a sufficiently dense discretization of the 2D-plane can allow one to approximate the continuous field well and reduce the complexity, as is demonstrated in the simulation section. In this work the true location is initially assumed to lie within a known finite set in order to facilitate a dynamic programming solution. Additional signal processing may be applied to obtain point estimates of the location. That thoroughly studied source localization topic is outside the scope of this paper.

The chosen sampling rate and duration of the discrete time slot used in acquiring the observations can highly affect the sensitivity of the network to the spatial event, and the time resolution and delay of detecting the change. Generally, the time-domain sampling rate should be selected so that one can distinguish among differences in the disruption arrival times at different sensors. In the considered model, if the event propagates with a constant radial velocity, we model the sampling rate such that during each time slot the radius of the disruption area increases by a fixed unit, e.g. some fraction or multiple of the wavelength. In order to take a variety of random propagation effects into account, we allow some randomness in the propagation of the spatial event. For example, epidemic spread may have a high degree of stochasticity due to random movements and interactions among individuals [22]. Generally, propagation randomness can be due to randomness in the velocity [22, Eq. (6.8)], due to timing jitter [52], or due to reflections, non-homogeneous medium, scattering, and multipath [35].

Let $R_n$ denote the area radius of the propagating event at time slot $n$. For simplicity, we assume that the area radius can have only discrete integer values corresponding to a fixed distance unit. Let $\mathfrak{R} \in \mathbb{N}$ denote the smallest disruption area radius that covers the entire domain of interest, $\mathcal{S}$, regardless of the actual point of origin, $O \in \mathcal{O}$. Thus, we assume that $R_n \in [\mathfrak{R}+1]$. It is assumed that $R_n = 0$ when $n < t$, and $R_t = 1$, i.e. only when the initial change occurs, the event area radius expands by one unit. In addition, we assume that

$$\mathbb{P}(R_n = r+1 | R_{n-1} = r) = 1 - \mathbb{P}(R_n = r | R_{n-1} = r) = \rho_1, \tag{2}$$

$\forall r \in [\mathfrak{R}] \setminus \{0\}$, $n \in \mathbb{N}$. At each time slot after the initial change the radius of the disruption increases by one radius unit with probability $\rho_1 \in (0,1]$ and stays the same as in the previous time slot with probability $1-\rho_1$. As $\mathfrak{R}$ is the maximum radius of the affected region, if $R_n = \mathfrak{R}$ then $R_m = \mathfrak{R}$ for $m \geq n$. Using (1), we obtain

$$\begin{aligned}\mathbb{P}(R_n = 1 | R_{n-1} = 0) &= 1 - \mathbb{P}(R_n = 0 | R_{n-1} = 0) \\ &= \mathbb{P}(t = n | t \geq n) = \rho,\ n \in \mathbb{N}.\end{aligned} \tag{3}$$

In addition, we obtain $\mathbb{P}(R_0 = 0) = 1 - \rho$ and $\mathbb{P}(R_0 = 1) = \rho$.

At each sensor location it is assumed that the sensor observes the disruption only if the disruption is present in this location, i.e. the distance between the sensor location and the source location is smaller than the current area radius of the disruption. Assume that the disruption is emanating from a source at $O = o_m$. Then, if a sensor at location $a \in \mathcal{S}$ is not exposed to the disruption, it acquires a noise-only observation coming from a known null probability density function (pdf), $f_0$. Otherwise, if this sensor is exposed to the disruption, it receives an observation with known pdf, $f_1^{(a,o_m)}$, that may depend on $a \in \mathcal{S}$ and $o_m \in \mathcal{O}$. For example, the power of the received signal can affect the parameters of the alternative pdf, and due to path loss may depend on the displacement between the sensor and the source [35], [53]. In many applications, the $f_0$ density represents random noise only, the statistical properties of which can be either known from theory, or estimated from training data even locally for each sensor in the absence of signal. On the other hand, the exact $f_1$ distribution, influenced by the appearing signal, may not always be known in practice. The issue of dealing with uncertainty in the $f_0$ and $f_1$ distributions has been an active topic of research in the field of quickest detection, see for example [54], [55] and references therein. Therefore, in this work we consider the probability models to be known, and refer to the existing literature for solutions on handling any model uncertainty. Moreover, it will be observed in Section IV that for the PFA control it suffices to know only the $f_0$ distribution. In many detection problems, controlling the false positives is crucial so that the system is not overwhelmed with detections and subsequent tasks. Conditional on the true system state and the sensor locations, the observations at each time slot are assumed to be independent across the sensors, as well as independent of all previous observations. Since the individual sensors are distributed and in distinct locations, the sensor noise present in any physical measurement can be considered independent.

At each time slot, the FC decides whether the initial event has taken place or not based on the information, $I_n$, which is available at time slot $n$. To this end, it uses a stopping time, $T$, according to a predefined stopping rule. The delay in detection is quantified by the ADD,

$$\mathrm{ADD}(T) \triangleq \mathbb{E}[(T-t)^+], \tag{4}$$

where $x^+ \triangleq \max\{0, x\}$. The PFA is defined as

$$\mathrm{PFA}(T) \triangleq \mathbb{P}(T < t). \tag{5}$$

## III. Dynamic Programming for Optimal Stopping Time

In a similar manner to classic Bayesian change-point detection [8], our goal is to derive the stopping time

$$T_{\mathrm{opt}} = \arg \inf_{T \in \Delta_\alpha} \mathrm{ADD}(T), \tag{6}$$

where $\Delta_\alpha \triangleq \{T : \mathrm{PFA}(T) \leq \alpha\}$. Put into words, we want to find a stopping time with the smallest ADD among stopping times for which the PFA is not larger than $\alpha$, where $\alpha \in (0,1)$ is a predefined tolerated level of false alarms. In this section, we take a dynamic programming approach for solving (6).

### A. Finite Horizon

We begin by restricting the stopping time to a finite horizon $[0, N]$. Solving the constrained optimization problem in (6) can be approached by formulating a Lagrangian relaxation problem that minimizes the Bayes risk

$$B(T, c) \triangleq \mathbb{P}(T < t) + c \cdot \mathbb{E}[(T-t)^+] \tag{7}$$

over all admissible stopping times. The state of the system at time $n$ is denoted by $S_n \in \{(m,r) : m \in [M],\, r \in [\mathfrak{R}+1]\} \cup \Upsilon$, with $S_n = (m,r)$ meaning that at time $n$ the event originating from $O = o_m$ has radius $R_n = r$. The term $\Upsilon$ represents the terminal state that the system goes into after a change is declared. In case $R_n = 0$, the spatial event has not occurred yet. From the description in Section I it is clear that the system state $S_n$ evolves as a Markov process. Moreover, conditional on the system state and sensor locations, the observations are i.i.d. As such, the problem lends itself to a dynamic programming [56] solution.

Since $\{T < t\} \Leftrightarrow \{R_T = 0\}$, the Bayes risk in (7) can be expressed in additive form as [9], [14]

$$B(T,c) = \mathbb{P}\left(R_T = 0\right) + c \cdot \mathbb{E}\left[\sum_{n=0}^{T-1} \mathbb{P}\left(R_n \geq 1\right)\right]. \quad (8)$$

In a finite horizon, we denote the minimum expected cost-to-go from $n$ to $N$ by $\mathcal{J}_n^N(I_n)$, which is in general a function of all available information $I_n$ at time $n$. The cost-to-go function [56] obeys the backwards recursion

$$\mathcal{J}_n^N(I_n) = \min\left\{\mathbb{P}(R_n = 0|I_n), c \cdot \mathbb{P}(R_n \geq 1|I_n) + \mathbb{E}[\mathcal{J}_{n+1}^N(I_{n+1})|I_n]\right\}, \quad (9)$$

with

$$\mathcal{J}_N^N(I_N) = \mathbb{P}(R_N = 0|I_N). \quad (10)$$

In (9) the first term inside the minimum corresponds to the expected cost of stopping at $n$, and the second term denotes the expected cost of continuing the monitoring process. We denote the posterior probability of the event $\{S_n = (m,r)\}$ given $I_n$ by $p_{n,m,r}$. That is,

$$p_{n,m,r} \triangleq \mathbb{P}(O = o_m, R_n = r|I_n), \quad (11)$$

and

$$\boldsymbol{p}_n \triangleq [p_{n,0,0}, \ldots, p_{n,1,\mathfrak{R}}, p_{n,2,0}, \ldots, p_{n,2,\mathfrak{R}}, \ldots, p_{n,M-1,\mathfrak{R}}] \quad (12)$$

is a $M \cdot (\mathfrak{R}+1)$ dimensional vector that collects all of the probabilities of time $n$. In the next subsection, we present a recursive update formula for $\boldsymbol{p}_n$ that will be used in the dynamic programming solution.

### B. Posterior Probabilities Computation

At any time slot, $n$, the sample space of the considered setup, $\Omega$, can be partitioned as

$$\Omega = \bigcup_{m=1}^{M} \bigcup_{r=0}^{\mathfrak{R}} U_{n,m,r}, \quad (13)$$

where

$$U_{n,m,r} \triangleq \{O = o_m, R_n = r\} = \{S_n = (m,r)\} \quad (14)$$

are pairwise disjoint events and the events $\{O = o_m\}$ and $\{R_n = r\}$ are independent. In the following, we derive a convenient recursive formula for computing $p_{n,m,r} = \mathbb{P}(U_{n,m,r}|I_n)$. Repeated use of the Bayes rule allows us to write $p_{n,m,r}$ as

$$p_{n,m,r} = \frac{f(\mathbf{x}_n|U_{n,m,r})\mathbb{P}(U_{n,m,r}|I_{n-1})}{\sum_{l=1}^{M}\sum_{\tilde{r}=0}^{\mathfrak{R}} f(\mathbf{x}_n|U_{n,l,\tilde{r}})\mathbb{P}(U_{n,l,\tilde{r}}|I_{n-1})}. \quad (15)$$

Given the conditional independence of the observations we have the factorization

$$f(\mathbf{x}_n|U_{n,m,r}) = \prod_{a\in\mathcal{A}_n:\|a-o_m\|<r} f_1^{(a,o_m)}(x_n^{(a)}) \prod_{a\in\mathcal{A}_n:\|a-o_m\|\geq r} f_0(x_n^{(a)}). \quad (16)$$

In addition, according to the assumed propagation model, $R_{n-1}$ can only be equal to $R_n$ or less than $R_n$ by one. Therefore, by the law of total probability, Bayes rule, and (2)–(3), we can write

$$\mathbb{P}(U_{n,m,r}|I_{n-1}) = \mathbb{P}(U_{n,m,r}|U_{n-1,m,r-1})p_{n-1,m,r-1} + \mathbb{P}(U_{n,m,r}|U_{n-1,m,r})p_{n-1,m,r}. \quad (17)$$

The conditional probabilities that the radius increases by one radius unit during one time slot for different radius values are

$$\mathbb{P}(U_{n,m,r}|U_{n-1,m,r-1}) = \rho_1, \quad r \in [\mathfrak{R}+1]\backslash\{0,1\},$$

and $\mathbb{P}(U_{n,m,1}|U_{n-1,m,0}) = \rho$. The conditional probabilities that the radius stays the same during one time slot for different radius values are $\mathbb{P}(U_{n,m,\mathfrak{R}}|U_{n-1,m,\mathfrak{R}}) = 1$,

$$\mathbb{P}(U_{n,m,r}|U_{n-1,m,r}) = 1 - \rho_1,\ \forall r \in [\mathfrak{R}+1]\backslash\{0,\mathfrak{R}\},$$

and $\mathbb{P}(U_{n,m,0}|U_{n-1,m,0}) = 1 - \rho$. At $n = 0$, we obtain

$$\mathbb{P}(U_{0,m,r}) = 0,\ \forall r \in [\mathfrak{R}+1]\backslash\{0,1\},$$

$\mathbb{P}(U_{0,m,1}) = \frac{1}{M}\rho$, and $\mathbb{P}(U_{0,m,0}) = \frac{1}{M}(1-\rho)$. In particular, it is seen that $\mathbb{P}(U_{n,m,r}|U_{n-1,m,r-j})$, $j = 0,1$, is independent of $n$. It should be noted that the radius of the area of the spatial event can reach a radius $r$ no earlier than time slot $n = r-1$.

At time $n$, the probabilities $\boldsymbol{p}_n$ can be updated using only the probabilities at the previous time step $\boldsymbol{p}_{n-1}$, current observation vector $\mathbf{x}_n$, and prior information. Thus, even as data accumulates with time, the amount of computations required for computing $\boldsymbol{p}_n$ remains constant per time slot. In particular, at time slot $n$ the amount of computations required for computing $\boldsymbol{p}_{n,m,r}$ is $\mathcal{O}(M\mathfrak{R})$.

It is observed that $\boldsymbol{p}_n$ depends on $I_{n-1}$ only through $\boldsymbol{p}_{n-1}$ and by (10) we have $\mathcal{J}_N^N(I_N) = \mathcal{J}_N^N(\boldsymbol{p}_N)$. Then, a simple induction argument shows that $\boldsymbol{p}_n$ is a sufficient statistic for the program, i.e. the minimum expected cost-to-go from $n$ to $N$ can be expressed as a function of $\boldsymbol{p}_n$, and thus $\mathcal{J}_n^N(I_n) = \mathcal{J}_n^N(\boldsymbol{p}_n)$. We denote the posterior probability of the event having radius $r$ at time $n$ by

$$\pi_{n,r} \triangleq \mathbb{P}(R_n = r|I_n) = \sum_{m\in[M]} p_{n,m,r}. \quad (18)$$

The Bellman equations from (9) and (10) can then be expressed as

$$\mathcal{J}_n^N(\boldsymbol{p}_n) = \min(\pi_{n,0},\, c(1-\pi_{n,0}) + \mathcal{D}_n^N(\boldsymbol{p}_n)), \quad (19)$$

and

$$\mathcal{J}_N^N(\boldsymbol{p}_n) = \pi_{N,0}, \tag{20}$$

where

$$\mathcal{D}_n^N(\boldsymbol{p}_n) \triangleq \mathbb{E}[\mathcal{J}_{n+1}^N(\boldsymbol{p}_{n+1})|I_n], \tag{21}$$

can be expressed as a function of $\boldsymbol{p}_n$ similarly to [9], [14].

### C. Extension to Infinite Horizon

In this subsection, we remove the upper bound on $T$, and consider the case $N \to \infty$. We write $\mathcal{J}_n \triangleq \lim_{N\to\infty} \mathcal{J}_n^N$ for the cost-to-go function in the limit and similarly $\mathcal{D}_n \triangleq \lim_{N\to\infty} \mathcal{D}_n^N$. The limits are well defined, since $0 \leq \mathcal{J}_n^N(\boldsymbol{p}_n) \leq 1$ and $\mathcal{J}_n^N(\boldsymbol{p}_n) \geq \mathcal{J}_n^{N+1}(\boldsymbol{p}_n)$ for all $\boldsymbol{p}_n$, $n$ and $N$. Therefore, we obtain

$$\mathcal{J}_n(\boldsymbol{p}_n) = \min(\pi_{n,0}, c(1-\pi_{n,0}) + \mathcal{D}_n(\boldsymbol{p}_n)) \quad n \in \mathbb{N}_0, \tag{22}$$

where $\mathcal{D}_n$ and $\mathcal{J}_n$ are non-negative functions on the $M \times (\mathfrak{R}+1)$-dimensional simplex. It is then seen that the optimal stopping time $T_{\text{opt}}$ is of the form

$$T_{\text{opt}} = \inf\{n \in \mathbb{N}_0 : \pi_{n,0} < c(1-\pi_{n,0}) + \mathcal{D}_n(\boldsymbol{p}_n)\}, \tag{23}$$

where the change is declared the first time the posterior probability of the event not being present drops below $c(1-\pi_{n,0}) + \mathcal{D}_n(\boldsymbol{p}_n)$. In general, the structure of $\mathcal{D}_n$ is not explicitly known, hence no closed-form optimal solution exists. Furthermore a numerical approximation of the optimal stopping time is computationally challenging and may be hard to analyze.

An interesting special case is the regime where the initial disruption is a rare event, i.e. $\rho \to 0$. The following result establishes that in this scenario the optimal test $T_{\text{opt}}$ converges in probability to a simple threshold test on $\pi_{n,0}$ which provides an attractive solution for practical use.

*Theorem 1:* The optimal stopping rule in (23) converges in probability to a threshold test

$$T_Q \triangleq \inf\{n \in \mathbb{N}_0 : \pi_{n,0} \leq Q\}, \tag{24}$$

for a properly chosen $Q$ as $\rho \to 0$.

*Proof:* See Appendix A. ■

In the following section, we propose a procedure denoted as the radial propagation (RP) procedure for single change-point detection, which is based on the threshold test from (24).

## IV. Single Change-Point Detection Procedure for Radial Propagation

In the previous section, it was observed that in the limit $\rho \to 0$, the optimal Bayesian stopping rule converges to a simple threshold rule $T_Q$, defined in (24). In this subsection, we study the performance of this stopping rule for any $\rho$.

The following proposition provides an upper bound for the false alarm of probability of $T_Q$.

*Proposition 1:* The false alarm probability of $T_Q$ from (24) can be upper bounded with $\text{PFA}(T_Q) \leq Q$.

*Proof:* By combining (5) and (8), one obtains

$$\begin{aligned}\text{PFA}(T_Q) &= \mathbb{P}(R_{T_Q} = 0) = \mathbb{E}\left[\mathbb{P}(R_{T_Q} = 0|I_{T_Q})\right] \\ &= \mathbb{E}[\pi_{T_Q,0}|I_{T_Q}] \leq Q,\end{aligned}$$

where the second equality is obtained using the law of iterated expectations, the third equality is obtained from the definition of $\pi_{n,0}$ in (18) and the inequality from the definition $T_Q$. ■

*Remark:* It should be noted that the PFA upper bound of Proposition 1 is valid even if many of the model assumptions are violated. As $T$ is a stopping time, $\{T < t\} \in I_{t-1}$. As all observations in $I_{t-1}$ are generated from the pre-change model, it is clear from the definition of the probability of false alarm in (5) that the PFA depends only on the pre-change observations. Violations of the assumed post-change behavior, such as a misspecified $f_1$ or departures from the assumed propagation model do not impact the PFA. This is a useful property, since the post-change distributions (usually generated by signal + noise) are often more difficult to characterize than the pre-change (noise only), as training data may be available from the pre-change probability model only.

From here on, we refer to the stopping time $T_Q$ as the radial propagation (RP) procedure, where the stopping threshold is chosen to equal the false alarm upper bound $\alpha$,

$$T_{\text{RP}} \triangleq \inf\{n \in \mathbb{N}_0 : \pi_{n,0} \leq \alpha\}. \tag{25}$$

### A. Asymptotic Optimality

In full generality, the ADD of the RP procedure is tedious to analyze due to the unknown source origin point, the potential mobility of the sensors and their arbitrary locations at each time slot. In order to shed some light on the ADD of the RP procedure, we provide sufficient conditions under which the RP procedure is asymptotically optimal in the vanishing PFA regime $\alpha \to 0$.

In the asymptotic analysis we consider the case where the disruption propagates in a deterministic fashion with constant velocity, i.e. $\rho_1 = 1$, so that its area radius increases by one unit in each time slot up to the maximum radius, $\mathfrak{R}$. The observations $\mathbf{X}_n$ are conditionally independent with pre-change pdf $f_0$ and post-change pdf $f_1^{(a,o_m)}$, respectively, where $f_1^{(a,o_m)}$ may depend on the sensor location $a$ and source location $o_m$. To proceed, let us define for all $n$ the set $\eta(n,k,m) \triangleq \{a \in \mathcal{A}_n : \|a - o_m\| < n-k+1\}$ that contains the locations of sensors that observe the event at time $n$ assuming it took place at time $k$ at origin $m$. Since the event propagates with constant velocity, on $\{t = k, O = o_m\}$, the joint density of the observations received at time $n \geq k$ by the FC is

$$f_{1,k,m}(\mathbf{x}_n) \triangleq \prod_{a\in\eta(n,k,m)} f_1^{(a,o_m)}(x_n^{(a)}) \prod_{a\notin\eta(n,k,m)} f_0(x_n^{(a)}), \tag{26}$$

with the factorization given is a result of the conditional independence of the sensor data. For $\{t = \infty\}$, at time $n$ the joint pdf of the observations received by the FC is $f_0(\mathbf{x}_n) \triangleq \prod_{a\in\mathcal{A}_n} f_0(x_n^{(a)})$. The log-likelihood ratio of the hypotheses $\{t = k, O = o_m\}$ and $\{t = \infty\}$ at time $n$ is:

$$Z_n^{k,m} \triangleq \sum_{i=k}^{n} \log \frac{f_{1,k,m}(\mathbf{x}_i)}{f_0(\mathbf{x}_i)}, \; k \leq n. \tag{27}$$

In order to analyze the asymptotic detection delay, some conditions on the long-term behaviour of the log-likelihood ratio process $Z_n^{k,m}$ are required. It is assumed that there exists some $q_m$ such that for every $k$ and $m$, on $\{t = k, O = o_m\}$

$$\frac{1}{n} Z_{k+n}^{k,m} \longrightarrow q_m \quad \text{almost surely.} \tag{28}$$

Note that if the post-change distribution is independent of location, i.e. $f_1^{(a,o_m)} = f_1$, and the number of sensors remains constant over time $|\mathcal{A}_n| = L$, it follows from the strong law of large numbers, the i.i.d. assumption and finiteness of $\mathfrak{R}$ that $q_m = L \cdot D(f_1||f_0)$ for all $m$. The following Lemma provides an asymptotic lower bound for the ADD for any procedure $T$ that fulfills PFA$(T) \leq \alpha$.

*Lemma 1:* Suppose that $\rho_1 = 1$ and that (28) applies for all $m \in [M]$. Let $\Delta_\alpha \stackrel{\triangle}{=} \{T : \text{PFA}(T) \leq \alpha\}$. Then,

$$\inf_{T \in \Delta_\alpha} \text{ADD}(T) \geq \frac{1}{M} \sum_{m=0}^{M-1} \frac{|\log \alpha|}{q_m + |\log(1-\rho)|}(1 + o(1)), \tag{29}$$

where $o(1) \to 0$ as $\alpha \to 0$.

*Proof:* From the definition of ADD we have that

$$\text{ADD}(T) = \frac{1}{M} \sum_{m=0}^{M-1} \text{ADD}_m(T), \tag{30}$$

where $\text{ADD}_m(T) \stackrel{\triangle}{=} \mathbb{E}[(T-t)^+ | O = o_m]$ is the detection delay when the true source location is $o_m$. Conditional on the source location $o_m$ being known, the problem reduces to a standard Bayesian quickest detection formulation with a non-i.i.d. post-change distribution given by (26). A lower bound for the asymptotic detection delay of any procedure in the class $\Delta_\alpha$ in this setting was derived in [8]. Specifically, by [8, Thm. 1]

$$\inf_{T \in \Delta_\alpha} \text{ADD}_m(T) \geq \frac{|\log \alpha|}{q_m + |\log(1-\rho)|}(1 + o(1)). \tag{31}$$

The Lemma follows from combining (30) and (31). ■

It should be noted that the asymptotic lower bound from (29) is identical to the lower bound for the case of instantaneous change, where all the sensors are affected at the same time [11].

The almost sure converge of $Z_n^{k,m}$, as required in (28), is not sufficient for proving the asymptotic optimality of the threshold rule $T_{\text{RP}}$. Therefore, in the following theorem we impose some mild additional assumptions on the rate of convergence of $Z_n^{k,m}$ to $q_m$ and show that $T_{\text{RP}}$ is asymptotically optimal and attains the lower bound from (29). To this end, we define for $\epsilon > 0$ the random variable,

$$Q_\epsilon^{(k,m)} = \sup\left\{ n \in \mathbb{N} : \left| \frac{1}{n} Z_{k+n-1}^{k,m} - q_m \right| > \epsilon \right\},$$

which is the largest value of $n$ for which the absolute difference between $\frac{1}{n} Z_{k+n-1}^{k,m}$ and $q_m$ is larger than $\epsilon$. It is required that

$$\sum_{k=1}^{\infty} \mathbb{P}(t = k) \mathbb{E}[Q_\epsilon^{(k,m)} | t = k] < \infty, \forall \epsilon > 0, m \in [M]. \tag{32}$$

Similarly to [8, Eq. (3.22)], the condition in (32) is a joint condition on the convergence rates of $\frac{1}{n} Z_{k+n}^{k,m}$ for each $t = k$ and the prior distribution of the change point $t$. In particular, it is analogous to complete convergence [57] of $\frac{1}{n} Z_{t+n}^{t,m}$ to $q_m$ under the distribution of $t$.

*Theorem 2:* Suppose the conditions of Theorem 1 are satisfied and assume that (32) is satisfied. Then $T_{RP}$ is first-order asymptotically optimal in the limit $\alpha \to 0$, i.e.

$$\lim_{\alpha \to 0} \frac{\inf_{T \in \Delta_\alpha} \text{ADD}(T)}{\text{ADD}(T_{RP})} = 1, \tag{33}$$

where

$$\inf_{T \in \Delta_\alpha} \text{ADD}(T) = \frac{1}{M} \sum_{m=0}^{M-1} \frac{|\log \alpha|}{q_m + |\log(1-\rho)|}(1 + o(1)). \tag{34}$$

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

### B. Detection of Attenuating Signals

In this subsection, we show how the obtained asymptotic results can be used to accurately approximate the expected detection delay in practically relevant settings. We consider the case of detecting an attenuating random Gaussian signal in additive noise. This model is highly relevant in a variety of practical applications in e.g. wireless communications and radar [58]. Prior to the change, only zero-mean i.i.d. Gaussian noise with variance $\sigma^2$ is observed. At an unknown time $t$, a signal source becomes active somewhere in the field. If the signal does not have any known structure, it can be modelled as zero-mean Gaussian with variance $\gamma^2$, where $\gamma^2$ is the signal transmit power. In free space, radio wave power decreases as the inverse square of distance $d$ between the source and the receiver [35]. In most practical wireless settings, the path loss exponent, denoted here by $\theta$, is usually greater than 2 due to obstacles, reflectors and scatterers. Therefore, for a sensor at distance $d$ away from the source excluding antenna and frequency dependent factors, the observed signal is of the form $\mathcal{N}(0, \gamma^2/\tilde{d}^\theta)$, where $\tilde{d} = \max(d, 1)$. Denoting $f_1^{(d)}$ as the post-change distribution at distance $d$ from the source, we have $f_0 = \mathcal{N}(0, \sigma^2)$ and $f_1^{(d)} = \mathcal{N}(0, \sigma^2 + \gamma^2/\tilde{d}^\theta)$. Suppose for analysis purposes that the signal source location is known, that the domain of interest is a disk with large radius $R$ centered at the signal source, and that at each time step there are $L$ sensors located independently and uniformly at random within the disk. The following result establishes that $T_{\text{RP}}$ is asymptotically optimal in this setting, and provides a first order approximation of the asymptotic detection delay.

*Proposition 2:* Under the conditions described in Sction IV-B, $T_{RP}$ is first-order asymptotically optimal. Moreover, in the free-space conditions of path-loss exponent $\theta = 2$

$$\text{ADD}(T_{\text{RP}}) = \frac{|\log \alpha|}{L q_\phi + |\log(1-\rho)|}(1 + o(1)), \tag{35}$$

where $q_\phi = \frac{1}{2R^2}[\phi + \phi \log(\phi + 1) - (\phi + R^2) \log(1 + \frac{\phi}{R^2})]$ and $\phi = \gamma^2/\sigma^2$ is the SNR in linear scale.

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

Observe, that the expression for $Lq_\phi$ in (35) further simplifies in the limit $R^2, L \to \infty$. When $R^2, L \to \infty$ such that $L/R^2 \to$

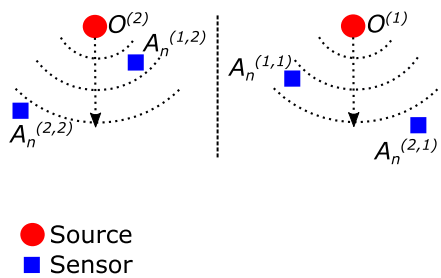



Fig. 2. Sources wavefront propagation: Phenomena emanating in two distinct clusters of sensors from the sources $O^{(1)}$ and $O^{(2)}$ towards the sensors at locations $A_n^{(1,1)}$, $A_n^{(2,1)}$ and $A_n^{(1,2)}$, $A_n^{(2,2)}$, respectively.

$\lambda$, we have

$$Lq_\phi \overset{R,L\to\infty}{\longrightarrow} \frac{\lambda}{2}\phi\log(\phi+1), \tag{36}$$

where $\lambda$ represents the average number of sensors per unit area.

## V. Extension to Multiple Change-Point Detection

In this subsection, we briefly describe how the simple structure of the RP stopping rule allows its use in settings when one is monitoring multiple separate fields and signal sources at once. This is a very relevant case in practice since in IoT, wireless networks or radar systems there may be multiple active signal sources, abrupt events or targets simultaneously and there is a need to strictly control the false positives in decision making while detecting changes rapidly. Suppose that there are $K \geq 2$ distinct clusters of sensors. For each cluster, $k \in [K]$, there may exist a random initial event (change point), at time $t^{(k)}$, that propagates and affects the sensors in the cluster according to the model in Section II. We allow the probability of no event in a cluster to be non-zero, where no event implies an infinite change point. The spatial events and sensor observations of the different clusters are assumed to be independent. This assumption, while restrictive in general, is reasonable in cases where the sensor clusters exist in spatially dispersed locations, and the events are spatially localized. The assumed setup is illustrated in Fig. 2. We would like to derive multiple stopping rules, $T^{(k)}$, $k \in [K]$, in order to discover all the change points, $t^{(k)}$, $k \in [K]$, respectively, while strictly controlling Type I errors. We employ a sequential multiple hypothesis testing framework for this purpose.

A practical assumption for any sequential detection procedure is that it must be stopped at some finite time instance. Thus, we allow the existence of a deadline $N_{\max}$ for the multiple change-point detection. If a change point in the $k$th cluster has not been declared before time slot $N_{\max}$, we declare that there is no spatial event in the $k$th cluster and set $T^{(k)} = \infty$. For detecting multiple change-points in parallel, the False Discovery Rate (FDR) is a relevant false alarm rate criterion [10], [21]. This criterion is defined as

$$\text{FDR} \triangleq \mathbb{E}\left[\frac{V}{\max(R,1)}\right]. \tag{37}$$

The term $V$ is the number of false discoveries (false alarms) under deadline, i.e. the size of the subset of $[K]$ s.t. $T^{(k)} < t^{(k)}$ and $T^{(k)} < N_{\max}$. The term $R$ denotes the number of discoveries under deadline, i.e. the size of the subset of $[K]$ s.t. $T^{(k)} < N_{\max}$. We would like to control the FDR s.t. it will be no higher than a predefined tolerated level $\alpha \in (0,1)$.

Taking into account the possibility of infinite change points, we denote by $K_f$ the random number of finite change points and define the overall ADD as

$$\text{ADD} \triangleq \mathbb{E}\left[\frac{1}{K_f}\sum_{k,t^{(k)}<\infty}(T^{(k)}-t^{(k)})^+\right], \tag{38}$$

where for $K_f = 0$ the argument of the expectation in (38) is zero. In case the *no change-point* probabilities are zero we can rewrite the ADD as

$$\text{ADD} \triangleq \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[(T^{(k)}-t^{(k)})^+]. \tag{39}$$

For the considered multiple statistically independent clusters, we will implement the following $K$ parallel stopping rules:

$$T_{\text{RP}}^{(k)} \triangleq \inf\left\{n \in \mathbb{N}_0 : \pi_{n,0}^{(k)} \leq \alpha\right\},\ k \in [K], \tag{40}$$

where $\pi_{n,0}^{(k)}$ is the posterior probability of a cluster change point having occurred in cluster $k$. A cluster change point is the time of an initial event in the cluster. The threshold choice in (40) guarantees FDR control under upper bound $\alpha$, in accordance with the parallel version of the IS-MAP procedure in [21].

In each cluster it is assumed that there is no change point with probability $p_\infty$ and with probability $1-p_\infty$ the prior distribution of the initial change point, $t^{(k)}$, is geometrically distributed with parameter $\rho$. Under the above assumptions, the change-point posterior probability update is similar to the one described in Section III-B and implemented for each cluster separately. However, some expressions for probabilities need to be re-derived. For simplicity of presentation, we omit the cluster index, $k \in [K]$, in the following expressions. For a specific cluster, by using (1) and the no change-point probability, $p_\infty$, we obtain

$$\begin{aligned}\mathbb{P}(R_n = 1|R_{n-1}=0) &= P(t=n|t\geq n)\\ &= \rho\frac{(1-p_\infty)(1-\rho)^{n-1}}{p_\infty+(1-p_\infty)(1-\rho)^{n-1}},\end{aligned} \tag{41}$$

$n \in \mathbb{N}$, where we recall that $P(R_n = 1|R_{n-1}=0) = 1 - P(R_n = 0|R_{n-1}=0)$. In addition to (41), the following expressions are rewritten to take into account the no change-point probability:

$$\mathbb{P}(R_0 = 0) = p_\infty + (1-p_\infty)(1-\rho),$$
$$\mathbb{P}(R_0 = 1) = (1-p_\infty)\rho,$$
$$\mathbb{P}(U_{n,m,1}|U_{n-1,m,0}) = \rho\frac{(1-p_\infty)(1-\rho)^{n-1}}{p_\infty+(1-p_\infty)(1-\rho)^{n-1}},$$
$$\mathbb{P}(U_{n,m,0}|U_{n-1,m,0}) = \frac{p_\infty+(1-p_\infty)(1-\rho)^{n}}{p_\infty+(1-p_\infty)(1-\rho)^{n-1}},$$

$$\mathbb{P}(U_{0,m,1}) = \frac{1}{M}(1 - p_\infty)\rho,$$

and

$$\mathbb{P}(U_{0,m,0}) = \frac{1}{M}(p_\infty + (1 - p_\infty)(1 - \rho)),$$

for $m \in [M]$.

## VI. Numerical Simulations

In this section, we evaluate the performance of the RP procedure in terms of PFA and FDR control and ADD performance under different radial propagation models and different multi-sensor configurations. In order to better understand the behavior of the RP procedure, we compare its performance to other procedures that either know the unobservable true source location, or deploy a more simplistic propagation model. Robustness to misspecification is tested by implementing a misspecified RP procedure that incorrectly assumes the event to propagate much faster than it does.

### A. Simple Gaussian Observation Model

We begin by considering a single cluster and a simple Gaussian observation model where $f_0 = \mathcal{N}(0, 1)$ and $f_1 = \mathcal{N}(0, 1 + \gamma^2)$, no matter the sensor location and the event origin point. In all experiments $L = 100$ sensors are randomly placed on the field at each time instance. The true source location is selected randomly from a uniform distribution over the field. Observe, that this is in contrast to the design-stage assumption that the true source locations lies in the finite set $\mathcal{O}$. The RP procedure is compared against two other procedures. The first one is an Oracle version of the RP procedure that knows the exact source location. The Oracle procedure is a special case of the RP procedure with $|\mathcal{O}| = 1$. The other procedure implemented for comparison purposes assumes that the event, once it appears, affects all sensors instantly [11]. We refer to this procedure as the Instant procedure. The Instant procedure is also a particular special case of the RP procedure, where one assumes $R_n = \Re$ for $n \geq t$ and $R_n = 0$ for $n < t$. It is to be expected that this procedure will provide inferior performance to the RP procedure, as it does not take the dynamic nature of the propagation into account. However, the comparison will provide insight into to the behavior of the RP procedure by highlighting the scenarios in which the performance gap between the properly specified and misspecified procedures is significant, and where the difference in performance is smaller.

We start by setting $\rho = 0.02$, $\rho_1 = 0.25$, $\gamma^2 = 1$ and considering a square spatial field $\mathcal{S} = [0, 10] \times [0, 10]$ where the sensors and sources are located. The set $\mathcal{O}$ used by the RP procedure is taken to be an equally spaced grid of $M$ points which covers the field of interest. In addition to the properly specified RP procedure, we implement a mismatched RP stopping rule (with $M = 50$), which correctly assumes that radius increases with probability $\rho_1$ but with increments of 5 times the true radius increment (1 unit). It corresponds to a setting where the real event propagates slower than assumed by the RP procedure.

TABLE I
Observed False Alarm Probabilities for Different Threshold Values $\alpha$

| | $\alpha = 0.1$ | $\alpha = 0.05$ | $\alpha = 0.01$ | $\alpha = 0.005$ |
|---|---|---|---|---|
| RP, $M = 10$ | 0.040 | 0.020 | 0.002 | 0.000 |
| RP, $M = 50$ | 0.034 | 0.012 | 0.002 | 0.002 |
| RP, $M = 100$ | 0.036 | 0.022 | 0.004 | 0.002 |
| RP, mismatched | 0.032 | 0.016 | 0.001 | 0.001 |
| Oracle | 0.024 | 0.016 | 0.004 | 0.004 |
| Instant | 0.008 | 0.002 | 0.001 | 0.001 |

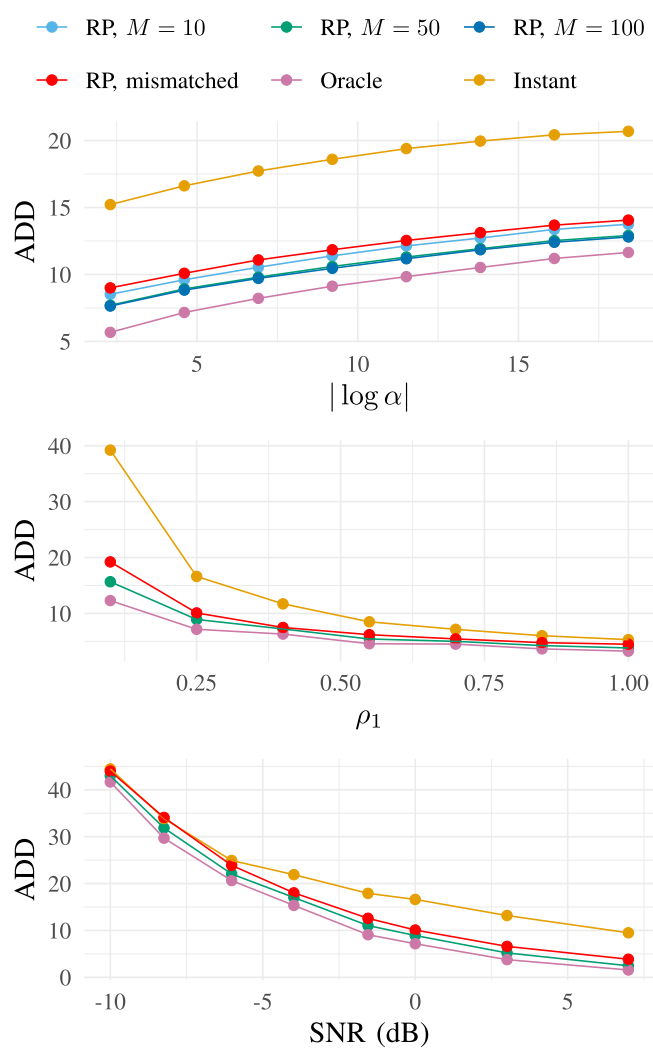



Fig. 3. Top: Average detection delay as a function of the PFA bound $\alpha$ for all procedures. Middle: ADD evaluated for different values of propagation parameter $\rho_1$, while keeping other parameters fixed. Bottom: ADD as a function of the SNR.

In Table I, observed false alarm probabilities of all procedures for different stopping thresholds $\alpha$ are displayed. It is confirmed that the theoretical PFA upper bound derived in Proposition 1 holds in all cases. In the top plot of Fig. 3, the PFA-ADD trade-off curves are plotted for the procedures, with the RP procedure implemented using source location grids of density $M = 10$, 50, and 100. For this small field, the performance of the RP procedure is comparable to the Oracle procedure. Furthermore, it is observed that increasing the density of the location grid in

the RP procedure improves performance. However, under this configuration for $M = 50$ and $M = 100$ the gap in performance is already indistinguishable. All versions of RP procedure, including the misspecified one, outperform the Instant procedure. In the middle plot of Fig. 3, the procedures are compared for varying values of the propagation parameter $\rho_1$, with $\alpha = 0.01$ fixed. For small values of $\rho_1$, the Instant procedure experiences performance loss in comparison to the others. This is because when $\rho_1$ is small the event will expand slowly with respect to the discrete-time sampling rate and thus remain spatially localized for a longer time, making it harder to detect for the Instant procedure. In general there is an inverse relationship between ADD and $\rho_1$ for all procedures, as a larger $\rho_1$ implies that the event will be visible to more sensors quicker. The RP and Oracle procedures provide near identical performance for all $\rho_1$. The mismatched RP procedure achieves lower ADD than the Instant procedure for all $\rho_1$ values. In the bottom plot of Fig 3, we fix $\rho_1 = 0.25$, $\alpha = 0.1$ and vary the signal power parameter $\gamma^2$. For unit noise variance, we have SNR (dB) $= 10\log_{10}(\gamma^2)$. It is observed that at low SNR regime the difference between the RP and Instant procedures is smaller, but for moderate and and high SNRs a clear gap in performance in favor of the RP procedure again emerges. Moreover, the difference in performance between the RP and Oracle procedures is small, and the size of the performance gap is relatively independent of SNR.

### B. Detection of Attenuating Radio Signals

In this subsection, we implement the attenuating signal model introduced in Section IV-B. To demonstrate the extension of the RP procedure to the detection of multiple events in parallel, we consider a setting with $K = 20$ distinct, independent sensor clusters each with 100 sensors as described in Section V. Prior to the change in a given cluster, all sensors observe noise only, so that $f_0 = \mathcal{N}(0, 1)$. When a signal source appears in the $k$th cluster at time $t^{(k)}$, it starts emitting an i.i.d. random signal modeled as $\mathcal{N}(0, \gamma^2)$. Due to path loss, the received signal strength attenuates according to a path-loss exponent $\theta$ of the distance $d$ from the source. The signal and noise are considered additive, hence for a sensor at distance $d$ from the signal source we have $f_1^{(d)} = \mathcal{N}(0, 1 + \gamma^2/\tilde{d}^\theta)$, where $\tilde{d} = \max(d_0, d/d_0)$ with $d_0$ being a reference distance where the received signal power equals $\gamma^2$. In the case of radio waves, the signal propagates at the speed of light $c$. The sensors take discrete time samples with some common sampling rate $f_s$. Therefore, the signal area radius expands in a deterministic manner (i.e. $\rho_1 = 1$) by $c/f_s$ meters in a single time step. We take each cluster area $\mathcal{S}_k, k \in [K]$ to be a square field with side length 5 km. The time at which the signal appears, $t^{(k)}$, is considered to have an exponential prior distribution with a mean (in seconds) of $\beta = 10$ in all sensor clusters. A routine computation utilizing the properties of the exponential and geometric distributions then shows that the sample index at which the emitted signal first appears obeys a geometric distribution with parameter $\rho = 1 - \exp(-1/(\beta f_s))$. The RP procedure is again compared against an Oracle procedure that knows the true signal source location in each cluster, and the exact propagation dynamics. Additionally, two versions of the Instant procedure are implemented. The first one (called Instant-Oracle) knows the true and unobservable source location in each cluster, but assumes that the event reaches all sensors in the cluster immediately. The other one (Instant) assumes similarly to the RP procedure that the source location in each cluster belongs to a finite set $\mathcal{O}$, and that the change is immediate everywhere in the cluster. Note that in the setting of Section VI-A knowledge of the true source location is not utilized in the Instant procedure since the appearance of the event was assumed to immediately change all sampling distributions from $f_0$ to $f_1$ no matter the source location. However, in this setting, the post-change sampling distribution $f_1^{(d)}$ depends on the distance of the sensor from the source. Therefore, knowing the true source location has value even if the propagation is assumed immediate. Consequently, we obtain an interesting comparison between the RP and the Instant-Oracle procedures, as the RP procedure is aware of the propagation dynamics, but the Instant-Oracle has knowledge of the true source location.

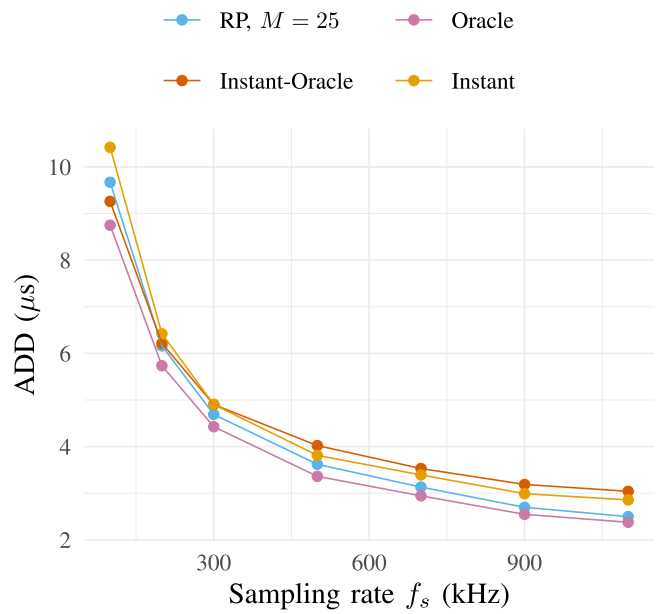



Fig. 4. Top: ADD evaluated for different values of propagation parameter $f_s$ in the attenuating signal and multiple cluster setting.

In all clusters, we set the signal power $\gamma^2 = 2$ at a reference distance of 500 m from the source, $\alpha = 0.01$ and the path loss exponent $\theta = 2$. The true source location of each cluster is sampled uniformly at random from $\mathcal{S}_k$. The sensor locations are also random and uniform, and assumed to remain stationary during the monitoring process. In Fig. 4, the procedures are compared for different values of the sampling rate $f_s$. The detection delay decreases for all procedures as the sampling rate increases, and the difference in ADD (in microseconds) between the RP and Oracle procedures shrinks as the sampling rate increases. It is observed, that for sufficiently high sampling rates the RP procedure achieves smaller detection delay than the Instant-Oracle procedure. When the sampling rate is high, accounting for the propagation dynamics is more valuable than theoretical knowledge of the source location, and vice versa when the propagation is rapid in comparison to the sampling rate. In Table II, the observed FDR values are displayed for

TABLE II
OBSERVED FALSE DISCOVERY RATES FOR DIFFERENT THRESHOLD VALUES $\alpha$

| | $\alpha = 0.1$ | $\alpha = 0.05$ | $\alpha = 0.01$ | $\alpha = 0.005$ |
|---|---|---|---|---|
| RP, $M = 10$ | 0.062 | 0.036 | 0.002 | 0.001 |
| Oracle | 0.048 | 0.020 | 0.006 | 0.005 |
| Instant-Oracle | 0.042 | 0.024 | 0.004 | 0.002 |
| Instant | 0.062 | 0.026 | 0.001 | 0.001 |

The RP procedure controls the FDR level below the specified threshold.

different choices of the stopping threshold $\alpha$ when $f_s = 1$ MHz. It is demonstrated that the RP procedure controls the FDR below the prespecified level $\alpha$.

## VII. CONCLUSION

In this paper, we proposed a method for Bayesian quickest detection of spatial events with radial propagation patterns using a mobile sensor network. First, we considered a single spatial event. A dynamic programming framework was used to derive the structure of the optimal stopping time in terms of ADD under upper bound constraint on the PFA. The optimal procedure has a complicated structure and implementing an approximation is computationally challenging and infeasible to analyze. Therefore, utilizing a limiting form of the optimal procedure we proposed the simpler RP procedure that employs a stopping threshold on the posterior probability of the change point of interest. It was shown both analytically and experimentally that the RP procedure controls the PFA under a prespecified upper bound, even if the post-change probability models are misspecified. In addition, we showed that under some conditions the proposed RP procedure coincides with an asymptotically optimal procedure in terms of ADD as the PFA upper bound $\alpha \to 0$. Then, we proposed an extension to parallel detection of multiple spatial events occurring in distinct clusters. The proposed method stems from a multiple hypothesis testing problem formulation and strictly controls FDR criterion while taking into account the spatial nature of the observed phenomena or fields. A posterior probability update expression for multiple change-point detection which takes into account a probability that no event appears was derived.

In the simulations it was observed that for phenomena that propagate slowly with regard to the sampling rate, the RP procedure vastly outperforms a procedure that assumes that the effect takes place instantly everywhere in the field. Similarly, in the high SNR regime the RP procedure provided significantly better performance than the Instant procedure. When the event propagates very quickly in relative to the sampling rate, or alternatively the SNR is very low, the performance gap was smaller, although still in favor of the RP procedure.

Topics for future research include the derivation of spatial procedures for multiple change-point detection and localization, where the locations of the signal sources are estimated using the observations. Additionally, extending the RP procedure to a non-Bayesian framework and studying its possible optimality properties is an interesting direction of future work.

## APPENDIX A
## PROOF OF THEOREM 1

Stemming from [9, Th. 2], our proof proceeds by showing that the optimal stopping time $T_{\text{opt}}$ can be written as

$$T_{\text{opt}} = \inf\left\{ n \in \mathbb{N}_0 : \pi_{n,0} < \frac{c + \Psi_n}{c + \rho} \right\}, \tag{42}$$

where $\Psi_n$ is a function such that $\frac{\Psi_n}{\rho} \to 0$ as $\rho \to 0$. The desired threshold test structure from (24) is then obtained in the limit $\rho \to 0$. Let us define

$$\Psi_n \triangleq \mathcal{D}_n(\boldsymbol{p}_n) - (1 - \rho)\pi_{n,0}. \tag{43}$$

Substituting this definition into (23) and rearranging gives (42). Then, the convergence of $\frac{\Psi_n}{\rho} \to 0$, as $\rho \to 0$, can be shown by introducing the transformation

$$q_{n,r} = \frac{\pi_{n,r}}{\rho \pi_{n,0}} \iff \pi_{n,r} = \frac{q_{n,r}}{\sum_{\tilde{r}=0}^{\mathfrak{R}} q_{n,\tilde{r}}}.$$

This expression allows for using the steps in [9, Th. 2] to complete the proof.

## APPENDIX B
## PROOF OF THEOREM 2

The proof is in two parts. First we define a set of $M$ stopping times $T^{(0)}, \ldots, T^{(M-1)}$, such that

$$T^{(m)} \triangleq \inf\{n : W_n^{(m)} \geq \nu\}, \tag{44}$$

for some threshold $\nu$, where $W_n^{(m)} \triangleq \mathbb{P}(t \leq n | I_n, O = o_m)$. In Lemma 2 below, we show that a stopping time defined as the minimum of these $M$ stopping times with thresholds $\nu = 1 - \alpha/M$ achieves the asymptotic ADD lower bound. Then, it is shown that $\text{ADD}(T_{\text{RP}}) \leq \text{ADD}(T^*)$, and the Theorem follows.

*Lemma 2:* Let $T^* = \inf\{T^{(0)}, \ldots, T^{(M-1)}\}$ and $\nu = 1 - \alpha/M$. Then

$$\text{ADD}(T^*) \leq \frac{1}{M} \sum_{m=0}^{M-1} \frac{|\log \alpha|}{q_m + |\log(1 - \rho)|}(1 + o(1)), \tag{45}$$

i.e. $T^*$ achieves the asymptotic ADD lower bound in (29).

*Proof:* Observe first from the definition of ADD that

$$\text{ADD}(T^*) = \frac{1}{M} \sum_{m=0}^{M-1} \text{ADD}_m(T^*), \tag{46}$$

where $\text{ADD}_m(T^*) \triangleq \mathbb{E}[(T^* - t)^+ | O = o_m]$. When $o_m$ is the true source location, the problem reduces to a Bayesian quickest detection task with a non-i.i.d. post-change distribution. It is shown in [8], that $T^{(m)}$ with a properly chosen stopping threshold isasymptotically optimal for minimizing $\text{ADD}_m$, so that

$$\text{ADD}_m\left(T^{(m)}\right) \leq \frac{|\log(1 - \nu)|}{q_m + |\log(1 - \rho)|}(1 + o(1)). \tag{47}$$

Since $|\log(1 - \nu)| = |\log \alpha| + \log M = |\log \alpha|(1 + o(1))$, and by definition $T^* \leq T^{(m)}$ for all $m$, combining (46) and (47) yields (45). ■

By Proposition 1, $\text{PFA}(T_{\text{RP}}) \leq \alpha$. Therefore, in order to prove Theorem 2, it is sufficient to show that $\text{ADD}(T_{\text{RP}}) \leq \text{ADD}(T^*)$. We have

$$1 - \pi_{n,0} = \sum_{m=0}^{M-1} W_n^{(m)} \mathbb{P}(O = o_m | I_n), \tag{48}$$

and

$$\begin{aligned}\mathbb{P}(O = o_m | I_n) &= \mathbb{P}(O = o_m, t \leq n | I_n) \\ &\quad + \mathbb{P}(O = o_m, t > n | I_n) \\ &= W_n^{(m)} \mathbb{P}(O = o_m | I_n) + \frac{\pi_{n,0}}{M}.\end{aligned} \tag{49}$$

The second equality in (49) follows from the fact that given $t > n$, the event $\{O = o_m\}$ is independent of $I_n$. Rearranging, one obtains $\mathbb{P}(O = o_m | I_n) = \pi_{n,0}/(M(1 - W_n^{(m)}))$. Hence,

$$1 - \pi_{n,0} = \frac{\pi_{n,0}}{M} \sum_{m=0}^{M-1} \frac{W_n^{(m)}}{1 - W_n^{(m)}}, \tag{50}$$

$$\pi_{n,0} = \frac{M}{M + \sum_{m=0}^{M-1} \frac{W_n^{(m)}}{1 - W_n^{(m)}}}. \tag{51}$$

Since the function $h(x) = M/(M + x)$ is decreasing in $x$ for $x > 0$, and $W_{T^*}^{(m)} \geq \nu = 1 - \alpha/M$ for some $m$ by definition of $T^*$, we obtain

$$\begin{aligned}\pi_{T^*,0} &\leq \frac{M}{M + \frac{1-\alpha/M}{\alpha/M}} = \frac{M\alpha}{M\alpha + M - \alpha} \\ &= \alpha + \frac{\alpha^2 - M\alpha^2}{M\alpha + M - \alpha},\end{aligned} \tag{52}$$

where the first equality is obtained by rearrangement, and the second equality by adding and subtracting $\alpha$ and rearranging. Since $M \geq 1$ and $\alpha \in [0, 1]$, the second term on the last line of (52) is non-positive, and hence $\pi_{T^*,0} \leq \alpha$. As $T_{\text{RP}} = \inf\{n : \pi_{n,0} \leq \alpha\}$ and $\pi_{T^*,0} \leq \alpha$, we obtain $T_{\text{RP}} \leq T^*$ and $\text{ADD}(T_{\text{RP}}) \leq \text{ADD}(T^*)$.

## Appendix C
## Proof of Proposition 2

We would like to use Theorem 2 to establish asymptotic optimality of $T_{\text{RP}}$ for quickest detection of the propagating signal, and approximate the detection delay in this setting.

To compute the constant $q$ defined in (28) and appearing in (34), observe that we have the partitioning

$$Z_{k+n}^k = \sum_{i=k}^{k+\mathfrak{R}-1} \log \frac{f_{1,k}(\mathbf{x}_i)}{f_0(\mathbf{x}_i)} + \sum_{i=k+\mathfrak{R}}^{k+n} \log \frac{f_{1,k}(\mathbf{x}_i)}{f_0(\mathbf{x}_i)}, \tag{53}$$

where $R$ is the number of time steps needed for the event to cover the entire region. On $\{t = k\}$ when $i \geq k + R$, the signal reaches all sensors, no matter where they are located within the domain. Therefore, for $i \geq k + R$

$$\mathbb{E}\left[\log \frac{f_{1,k}(\mathbf{x}_i)}{f_0(\mathbf{x}_i)}\right] = L\mathbb{E}_{f_1^{(d)}}\left[\log \frac{f_1^{(d)}(x)}{f_0(x)}\right], \tag{54}$$

where the latter expectation is over both the random location (in particular the random distance $d$ from the source) and the random observation generated from the post-change $f_1^{(d)}$ distribution. Since $R$ is a finite constant, by the strong law of large numbers and (53)–(54)

$$\frac{1}{n} Z_{k+n}^k \underset{n\to\infty}{\longrightarrow} L\mathbb{E}_{f_1^{(d)}}\left[\log \frac{f_1^{(d)}(x)}{f_0(x)}\right] \quad \text{a.s.} \tag{55}$$

By iterated expectation and a direct computation of the KL-divergence between two Gaussians we get

$$\mathbb{E}_{f_1^{(d)}}\left[\log \frac{f_1^{(d)}(x)}{f_0(x)}\right] = \frac{1}{2}\mathbb{E}_d\left[\frac{\phi}{\tilde{d}^\theta} - \log\left(1 + \frac{\phi}{\tilde{d}^\theta}\right)\right] \triangleq q_\phi, \tag{56}$$

where $\phi \triangleq \gamma^2/\sigma^2$ is the SNR in linear scale. As the sensor locations are uniform on the disk, $\mathbb{P}(d \leq s) = (s/R)^2$, for $0 \leq s \leq R$. Therefore,

$$q_\phi = \frac{1}{2R^2} \int_0^R \frac{\phi}{\tilde{s}^{\theta-1}} - s \log\left(1 + \frac{\phi}{\tilde{s}^\theta}\right) ds. \tag{57}$$

Evaluating the integral in (57) for a general path loss exponent $\theta$ is possible, but leads to convoluted results. In the commonly considered free-space conditions of $\theta = 2$, we have

$$q_\phi = \frac{1}{2\mathfrak{R}^2}\left[\phi + \phi \log(\phi + 1) - (\phi + R^2) \log\left(1 + \frac{\phi}{R^2}\right)\right]. \tag{58}$$

To apply Theorem 2, it remains to check that the joint convergence condition of both the observations and the prior distribution in (32) is satisfied. This is straightforward, since on $\{t = k\}$, $Z_{k+n}^{k+\mathfrak{R}}$ is a sum of i.i.d. random variables such that $n^{-1} Z_{k+n}^k$ converges almost surely to (36), and $\mathfrak{R}$ is a finite constant. Therefore, following [8, Sec. 4], (32) is established. The Proposition then follows from Theorem 2.

## References


[1] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

[2] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Probab. Appl.*, vol. 8, no. 1, pp. 22–46, 1963.

[3] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, 1971.

[4] M. Pollak, "Optimal detection of a change in distribution," *Ann. Stat.*, vol. 13, no. 1, pp. 206–227, 1985.

[5] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Stat.*, vol. 14, no. 4, pp. 1379–1387, 1986.

[6] 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, Nov. 1998.

[7] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[8] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *Theory Probab. Appl.*, vol. 49, no. 3, pp. 458–497, 2005.

[9] V. Raghavan and V. V. Veeravalli, "Quickest change detection of a Markov process across a sensor array," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1961–1981, Apr. 2010.

[10] J. Chen, W. Zhang, and H. V. Poor, "A false discovery rate oriented approach to parallel sequential change detection problems," *IEEE Trans. Signal Process.*, vol. 68, pp. 1823–1836, 2020.

[11] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1657–1665, May 2001.

[12] O. Hadjiliadis, 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, Jul. 2009.

[13] 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, Jul. 2009.

[14] M. N. Kurt and X. Wang, "Multisensor sequential change detection with unknown change propagation pattern," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 3, pp. 1498–1518, Jun. 2019.

[15] M. N. Kurt, V. Raghavan, and X. Wang, "Multi-sensor sequential change detection with unknown change propagation dynamics," 2017. [Online]. Available: https://arxiv.org/abs/1708.04722v1

[16] J. Li, D. Towsley, S. Zou, V. V. Veeravalli, and G. Ciocarlie, "A consensus-based approach for distributed quickest detection of significant events in networks," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, 2019, pp. 1–4.

[17] D. Li, S. Kar, and S. Cui, "Distributed quickest detection in sensor networks via two-layer large deviation analysis," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 930–942, Apr. 2018.

[18] J. Chen, W. Zhang, and H. V. Poor, "Non-Bayesian multiple change-point detection controlling false discovery rate," in *Proc. Int. Symp. Inf. Theory*, 2016, pp. 31–35.

[19] E. Nitzan, T. Halme, H. V. Poor, and V. Koivunen, "Deterministic multiple change-point detection with limited communication," in *Proc. IEEE 54th Annu. Conf. Inf. Sci. Syst.*, 2020, pp. 1–6.

[20] T. Halme, E. Nitzan, H. V. Poor, and V. Koivunen, "Bayesian multiple change-point detection with limited communication," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 5490–5494.

[21] E. Nitzan, T. Halme, and V. Koivunen, "Bayesian methods for multiple change-point detection with reduced communication," *IEEE Trans. Signal Process.*, vol. 68, pp. 4871–4886, 2020.

[22] M. Ludkovski, "Bayesian quickest detection in sensor arrays," *Seq. Anal.*, vol. 31, no. 4, pp. 481–504, 2012.

[23] D. Li, L. Lai, and S. Cui, "Quickest change detection and identification across a sensor array," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2013, pp. 145–148.

[24] D. Siegmund and B. Yakir, "Detecting the emergence of a signal in a noisy image," *Statist. Interface*, vol. 1, no. 1, pp. 3–12, 2008.

[25] D. Marangoni-Simonsen and Y. Xie, "Sequential changepoint approach for online community detection," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1035–1039, Aug. 2015.

[26] S. Li, Y. Xie, M. Farajtabar, A. Verma, and L. Song, "Detecting changes in dynamic events over networks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 2, pp. 346–359, Jun. 2017.

[27] G. Rovatsos, S. Zou, and V. V. Veeravalli, "Sequential algorithms for moving anomaly detection in networks," *Seq. Anal.*, vol. 39, no. 1, pp. 6–31, 2020.

[28] G. Rovatsos, G. V. Moustakides, and V. V. Veeravalli, "Quickest detection of moving anomalies in sensor networks," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 2, pp. 762–773, Jun. 2021.

[29] G. Rovatsos, V. V. Veeravalli, D. Towsley, and A. Swami, "Quickest detection of growing dynamic anomalies in networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 8926–8930.

[30] S. Zou, V. V. Veeravalli, J. Li, and D. Towsley, "Quickest detection of dynamic events in networks," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2280–2295, Apr. 2020.

[31] A. Wang, X. Xian, F. Tsung, and K. Liu, "A spatial-adaptive sampling procedure for online monitoring of big data streams," *J. Qual. Technol.*, vol. 50, no. 4, pp. 329–343, 2018.

[32] X. Xian, C. Zhang, S. Bonk, and K. Liu, "Online monitoring of big data streams: A rank-based sampling algorithm by data augmentation," *J. Qual. Technol.*, vol. 53, no. 2, pp. 135–153, 2021.

[33] M. Nabhan, Y. Mei, and J. Shi, "Correlation-based dynamic sampling for online high dimensional process monitoring," *J. Qual. Technol.*, vol. 53, no. 3, pp. 289–308, 2021.

[34] S. Marano and A. H. Sayed, "Decision learning and adaptation over multi-task networks," *IEEE Trans. Signal Process.*, vol. 69, pp. 2873–2887, 2021.

[35] E. Biglieri, A. J. Goldsmith, L. J. Greenstein, N. B. Mandayam, and H. V. Poor, *Principles of Cognitive Radio*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[36] Y. U. Lee, J. Choi, I. Song, and S. R. Lee, "Distributed source modeling and direction-of-arrival estimation techniques," *IEEE Trans. Signal Process.*, vol. 45, no. 4, pp. 960–969, Apr. 1997.

[37] O. Bar-Shalom and A. J. Weiss, "DOA estimation using one-bit quantized measurements," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 868–884, Jul. 2002.

[38] L. Lopez, F. A. B. Coutinho, M. N. Burattini, and E. Massad, "Modelling the spread of infections when the contact rate among individuals is short ranged: Propagation of epidemic waves," *Math. Comput. Model.*, vol. 29, no. 7, pp. 55–69, 1999.

[39] N. S. Trovão, M. A. Suchard, G. Baele, M. Gilbert, and P. Lemey, "Bayesian inference reveals host-specific contributions to the epidemic expansion of influenza a H5N1," *Mol. Biol. Evol.*, vol. 32, no. 12, pp. 3264–3275, 2015.

[40] N. J. Cunniffe, R. C. Cobb, R. K. Meentemeyer, D. M. Rizzo, and C. A. Gilligan, "Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in California," *Proc. Nat. Acad. Sci.*, vol. 113, no. 20, pp. 5640–5645, 2016.

[41] G. Soldi et al., "Quickest detection and forecast of pandemic outbreaks: Analysis of COVID-19 waves," *Comm. Mag.*, vol. 59, no. 9, pp. 16–22, Sep. 2021.

[42] P. Braca, D. Gaglione, S. Marano, L. M. Millefiori, P. Willett, and K. Pattipati, "Decision support for the quickest detection of critical covid-19 phases," *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, 2021.

[43] J. Cortes, S. Martinez, T. Karatas, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Trans. Robot. Automat.*, vol. 20, no. 2, pp. 243–255, Apr. 2004.

[44] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," in *Proc. IEEE Mil. Commun. Conf.*, 2003, vol. 1, pp. 688–693.

[45] G. Mergen, Q. Zhao, and L. Tong, "Sensor networks with mobile access: Energy and capacity considerations," *IEEE Trans. Commun.*, vol. 54, no. 11, pp. 2033–2044, Nov. 2006.

[46] S. Marano, V. Matta, P. Willett, and L. Tong, "Cross-layer design of sequential detectors in sensor networks," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4105–4117, Nov. 2006.

[47] P. Addesso, S. Marano, and V. Matta, "Sequential sampling in sensor networks for detection with censoring nodes," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5497–5505, Nov. 2007.

[48] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. R. Statist. Soc. Ser. B*, vol. 57, no. 1, pp. 289–300, 1995.

[49] B. Efron, *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*, vol. 1. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[50] T. Halme, E. Nitzan, and V. Koivunen, "Bayesian multiple change-point detection of propagating events," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 4515–4519.

[51] T. Halme, E. Nitzan, and V. Koivunen, "Quickest detection of propagating signals for flexible radio spectrum use," in *Proc. IEEE 55th Asilomar Conf. Signals, Syst., Comput.*, 2021, pp. 805–810.

[52] W. M. Lovelace and J. K. Townsend, "The effects of timing jitter and tracking on the performance of impulse radio," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 9, pp. 1646–1651, Dec. 2002.

[53] S. D. Chitte, S. Dasgupta, and Z. Ding, "Distance estimation from received signal strength under log-normal shadowing: Bias and variance," *IEEE Signal Process. Lett.*, vol. 16, no. 3, pp. 216–218, Mar. 2009.

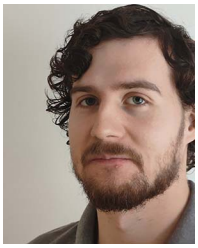
[54] J. Unnikrishnan, V. V. Veeravalli, and S. P. Meyn, "Minimax robust quickest change detection," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1604–1614, Mar. 2011.

[55] A. G. Tartakovsky and A. S. Polunchenko, "Quickest changepoint detection in distributed multisensor systems under unknown parameters," in *Proc. IEEE 11th Int. Conf. Inf. Fusion*, 2008, pp. 1–8.

[56] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Amsterdam, The Netherlands: Athena Scientific, 2000.

[57] P. L. Hsu and H. Robbins, "Complete convergence and the law of large numbers," *Proc. Nat. Acad. Sci. USA*, vol. 33, no. 2, pp. 25–31, 1947.

[58] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: A sequential change detection framework," in *Proc. IEEE Glob. Telecommun. Conf.*, 2008, pp. 1–5.



**Topi Halme** (Student Member, IEEE) received the B.Soc.Sc. and M.Sc. degrees in statistics from the University of Helsinki, Helsinki, Finland, in 2019 and 2021, respectively. He is currently working toward the Ph.D. degree in electrical engineering with Aalto University, Espoo, Finland. His research interests include change-point detection, multiple hypothesis testing, and sequential inference in sensor networks and spatial domains.

**Eyal Nitzan** (Member, IEEE) received the B.Sc. (*cum laude*), M.Sc. (*cum laude*), and Ph.D. degrees in electrical and computer engineering from the Ben-Gurion University of the Negev, Beer-Sheva, Israel, in 2012, 2014, and 2018, respectively. He was a Postdoctoral Researcher with the Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland. He is currently working with KIOXIA America, Inc. His research interests include estimation and detection theory, distributed inference, statistical signal processing, and circular statistics. He was the recipient of the Best Student Paper Award (1st prize) in IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing 2013, the Best Student Paper Award (3rd prize) in IEEE International Conference on Acoustics, Speech, and Signal Processing 2017, and the Best Student Paper Award (one of the three winners) in IEEE Statistical Signal Processing Workshop (SSP) 2018. He was also the recipient of the Negev scholarship awarded to excellent Ph.D. candidates in 2014, and the Israel's Ministry of Science, Technology, and Space Grant for active participation in international conference/workshop for doctoral students in 2017.

**Visa Koivunen** (Fellow, IEEE) received the D.Sc. degree (Hons.) in electrical engineering (EE) degree (Hons.) from the Department of Electrical Engineering, University of Oulu, Oulu, Finland. He was a Visiting Researcher with the University of Pennsylvania, Philadelphia, PA, USA, from 1991 to 1995, and an Adjunct Full Professor from 2003 to 2006. Since 1999, he has been a full Professor of signal processing with Aalto University (formerly HUT), Espoo, Finland. In 2010 he received the Academy Professor position and Aalto Distinguished Professor in 2020. During his sabbatical terms in 2006–2007 and 2013–2014, he was a Visiting Faculty with Princeton University, Princeton, NJ, USA. He has also been a Visiting Fellow with Nokia Research (2006–2012). Since 2010 he had many mini-sabbaticals with Princeton University. On his sabbatical term during 2022–2023, he is a Visiting Professor with EPFL, Lausanne, Switzerland.

His research interests include statistical signal processing, wireless comms, radar, multisensor systems and machine learning. He was the recipient of the IEEE SP Society Best Paper Award for the year 2007 (with J. Eriksson) and 2017 (with Zoubir, Muma and Chakhchouk) . He was with the Editorial Board for PROCEEDINGS OF THE IEEE, IEEE SIGNAL PROCESSING LETTERS, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and *IEEE Signal Processing Magazine*. He is an EURASIP Fellow. He was also the recipient of the 2015 EURASIP Technical Achievement Award for fundamental contributions to statistical signal processing and its applications in wireless communications, radar and related fields.