Bayesian Quickest Detection in Sensor Arrays

Michael Ludkovski

Department of Statistics and Applied Probability, University of California

Santa Barbara, CA, USA 93106

Abstract: We study Bayesian quickest detection problems with sensor arrays. An underlying signal is assumed to gradually propagate through a network of several sensors, triggering a cascade of inter-dependent change-points. The aim of the decision-maker is to centrally fuse all available information to find an optimal detection rule that minimizes Bayes risk. We develop a tractable continuous-time formulation of this problem focusing on the case of sensors collecting point process observations and monitoring the resulting changes in intensity and type of observed events. Our approach uses methods of nonlinear filtering and optimal stopping and lends itself to an efficient numerical scheme that combines particle filtering with Monte Carlo based approach to dynamic programming. The developed models and algorithms are illustrated with plenty of numerical examples.

Keywords: Bayesian Quickest Detection; Multiple Change-point Detection; Particle Filtering; Sensor Arrays.

Subject Classifications: 60G35; 62L15; 93E11; 65C35

1. INTRODUCTION

Quickest detection of signal changes is an important problem in a variety of disciplines ranging from bio-surveillance to network communications. In a common setting, several sensors are arranged in a specified sensor topology and monitor for a signal that may arise in a region of interest. With this physical layout, the signal gradually *propagates* through space, triggering distinct change-points as it reaches each sensor site. The task of the sensors is to fuse the sequentially collected information in order to detect the change-points as quickly as possible while controlling for probability of false alarms.

Address correspondence to M. Ludkovski, Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106-3110, USA; Tel: +1 (805) 893-5634; E-mail: ludkovski@pstat.ucsb.edu

A mechanistic description of a sequence of such disorders may be given by specifying the signal origin, start time, propagation dynamics and impact on sensor data. Alternatively, a purely statistical description can be made through specifying the relationship between the underlying change-points. In either case, the information fusion is fundamentally driven by the dependence structure of the change-points and the induced correlation among sensor observations. As a result, the associated detection problem is intrinsically multi-variate and much more complex than in the traditional single change-point models.

Quickest detection problems have been extensively analyzed in the past fifty years under a variety of probabilistic assumptions, including min-max, Bayesian and other settings, see e.g. the recent survey of Polunchenko and Tartakovsky (2012). In particular, optimal detection with multiple sensors was treated among others in Bayraktar and Poor (2008), Dayanik et al. (2008b,a), Tartakovsky et al. (2006). However, nearly all existing analysis is limited to the case of an identical change-point across all observation processes. This assumption rules out consideration of sensor topologies where the signal physically propagates through the sensor array. Optimality properties of detection rules such as CUSUM or Shiryaev-Roberts have been thoroughly studied in a variety of single change-point models, see for instance Feinberg and Shiryaev (2006), Moustakides (2004), Pollak and Tartakovsky (2009). In contrast, the case of multiple change-points has been considered only very recently. The influential work of Raghavan and Veeravalli (2010) investigated propagating change-points, but only under a restrictive Markovian mechanism with a known propagation order. Moreover, Raghavan and Veeravalli (2010) focused on the asymptotic Bayesian solution which admits a one-dimensional sufficient statistic and a simple threshold strategy. While theoretically convenient, these features cannot be maintained in realistic sensor networks. From a different direction, Hadjiliadis et al. (2009a) considered detection of the first disorder in a K-dimensional min-max setting, establishing asymptotic optimality of the multi-chart CUSUM rule as the probability of false alarms vanishes. In this paper, our aim is to construct the exact Bayesian optimal solution, rather than seeking other (asymptotic) notions of optimality. Thus, we focus on numerical approximation of the non-asymptotic optimal detection policy while maintaining as general a framework as possible.

Biosurveillance Application

Our main motivation for sensor array modeling arises from biosurveillance applications. The aim of biosurveillance is to monitor a range of prediagnostic and diagnostic data for the purpose of enhancing the ability of the public health infrastructure to detect, investigate, and respond to disease outbreaks. A typical example involves influenza surveillance in order to implement timely mitigation policies for the annual winter flu epidemics. Flu monitoring tracks the number of observed ILI (influenza-like-illness) cases in the

surveyed population, augmented with emergency room visits, laboratory tests, over-the-counter medication sales, etc. This is done on a local level by the city- or county-administered public health agencies. An epidemic traveling wave is commonly observed for new influenza strains (Riley 2007) (for example, the 2009 H1N1 pandemic originated in Mexico in April 2009 and gradually spread across the world over that summer); at the same time the inherently random movements and interactions between individuals introduce a high degree of stochasticity in epidemic spread. In order to localize outbreak detection, multi-site monitoring (e.g. in several neighboring counties) is strongly advocated. Thus, it is of great interest to efficiently fuse information from multiple spatially heterogeneous surveillance sites to pinpoint epidemic onset and pathogen properties. As noted by Raubertas (1989), a major open challenge is that "the use of spatial information is computationally demanding".

New Approach

Below we adopt the Bayesian formulation of quickest detection to develop a novel sequential algorithm for detection and identification of the unknown change-points. Our framework relies on the optimal stopping paradigm and allows us to efficiently obtain numerical solutions in generic models of sensor arrays. As our main setup we consider *centralized Bayesian detection* with continuous-time observations modeled as a point process. Use of continuous-time is not only convenient analytically but is also more faithful for asynchronous systems without a well-defined time scale. For instance, biosurveillance has been moving from discrete weekly or daily data collection to real-time syndromic monitoring, such as the Google FluTrends search engine data. Real-time information is bursty and consists of discrete events that punctuate periods of inactivity. We model such observations via a marked point process, linking to the theory of Poisson disorder problems. Because the eventual numerical implementation is in discrete time, our methods are in fact directly applicable (with obvious modifications) also in discrete-time models.

In line with the envisioned applications, we assume centralized information processing, with all sensors continuously relaying all observations to the fusion center. The latter then aggregates these multi-dimensional observations to infer current signal state and make the optimal detection decision. This situation occurs when information transmission is cheap and there is a clear hierarchy of decision making (e.g. federal biosurveillance protocols mandated by the CDC). With low-power sensors, such as in military applications or wireless networks, it might be too expensive for a sensor to continually communicate, making decentralized detection essential (Tartakovsky and Veeravalli 2005, 2008, Veeravalli et al. 1993, 1994, Unnikrishnan et al. 2011). Proposed solutions include one-shot communications (Hadjiliadis et al. 2009b), asynchronous communication protocols (Fellouris and Moustakides 2011), and distributed algorithms without central fu-

sion (Rajagopal et al. 2010).

Compared to existing approaches, our framework removes several important limitations by allowing for (i) generic signal propagation without any restrictive assumptions on network topology or ordering of the change-points; (ii) a true Bayesian formulation without existence of low-dimensional sufficient statistics or independence between change-point and observations; (iii) consideration of joint detection and identification objectives; (iv) flexible model specification through a fully simulation-based implementation that requires no intermediate analytical computations. The last point in particular means that the proposed algorithm code can be easily augmented in response to a modified model and is well suited to be published as an online package. Another benefit of our Bayesian approach is the ability to simultaneously optimize multiple objectives, such as balancing low detection delay vis-a-vis identification of the signal type. Identification of the signal is a common task in sensor arrays via triangulation of sensor measurements.

Summing up, our contribution is to develop a comprehensive Bayesian formulation of the detection problem in sensor arrays which faithfully accounts for signal propagation and sensor interaction. This framework is related to our previous work on Bayesian detection problems in the context of Poisson-type observations (Bayraktar and Ludkovski 2009, 2010, Ludkovski and Sezer 2012), as well as the author's computational tools (Ludkovski 2009, 2012). The tractability and flexibility of our approach is demonstrated with several examples in Section 6.

2. STOCHASTIC MODEL

Let $(\vec{\mathbf{Y}}_t) \equiv (Y_t^k)$ be the data streams at given sensor sites $k = 1, 2, \ldots, K$. For instance, in the context of biosurveillance, sites may refer to counties within a given state. The statistical properties of the information received at each site undergo a transition when the signal is present; the instants of disruption, henceforth called a change-point, are denoted by θ^k, $k = 1, \ldots, K$.

2.1. Propagation Model

A basic spatial spread of the signal can be mechanistically represented by a *wavefront* model illustrated in Figure 1. In that diagram, starting from its origin location $\mathcal{O} = (o^1, o^2)$ at initial epoch $\mathcal{T}_0 \geq 0$, the signal spreads outwards at constant radial velocity $\mathcal{V}$, eventually reaching the (known and fixed) sensor site locations $A_k = (a_k^1, a_k^2) \in \mathbb{R}^2$. Thus, the corresponding change-points are

$$\theta^k := \mathcal{T}_0 + \inf\{t \geq 0 : \|\mathcal{O} - A_k\| \leq \mathcal{V}t\} = \mathcal{T}_0 + \frac{\|\mathcal{O} - A_k\|}{\mathcal{V}}, \qquad k = 1, \ldots, K, \tag{2.1}$$

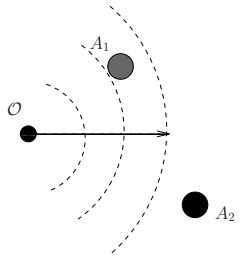


Figure 1: Signal front propagating through space. Signal starts at origin $\mathcal{O}$ and has constant radial velocity $\mathcal{V}$.

where $\|\cdot\|$ is the given metric for signal propagation, e.g. the Euclidean distance in $\mathbb{R}^2$ (see Figure 1) or a graph distance on a network. The system state is then described by the *disorder indicator process* $\vec{\mathbf{X}}_t \equiv (X_t^1, \ldots, X_t^K) \in \{0,1\}^K$ which encodes the present state of the signal at each site,

$$X_t^k := 1_{\{\theta^k \leq t\}}, \qquad k = 1, 2, \ldots, K. \tag{2.2}$$

To lighten notation, we will alternately think of $(\vec{\mathbf{X}}_t)$ as a process on the canonical space $E = \{e_k : k = 1, 2, \ldots, 2^K\}$, where $e_k \in \mathbb{R}^{2^K}$ is a vector with zero entries everywhere except a unit at the k-th coordinate, $e_{kj} = \delta_{kj}$, with δ the Dirac delta.

These signal parameters are summarized as $Z \equiv (\mathcal{O}, \mathcal{T}_0, \mathcal{V}) \in \mathcal{A} \subseteq \mathbb{R}^{|Z|}$, where $\mathcal{A}$ is the set of all possible signals, and are not observed. Rather, the observed information is solely based on the sensor data so far,

$$\mathcal{F}_t := \sigma(\vec{\mathbf{Y}}_s : s \leq t). \tag{2.3}$$

It is contrasted with the full filtration $\mathcal{G}_t := \mathcal{F}_t \vee \sigma(Z)$. Crucially, while $(\vec{\mathbf{X}}_t)$ is trivially $\mathcal{G}$-adapted, it is not measurable with respect to $(\mathcal{F}_t)$. Following the Bayesian paradigm, the parameters Z are treated as a random vector with a known prior distribution $Z \sim \tilde{\pi}$.

Remark 2.1. The disorder process $(\vec{\mathbf{X}}_t)$ is the key system state for quickest detection. As such, one could omit any description of the signal primitives Z and directly specify the distribution of $(\vec{\mathbf{X}}_t)$ through e.g. a copula on the marginal disorder dates θ^k. Indeed, the possible relative locations of $(\mathcal{O}, A_k)$ simply induce an implicit relationship between (X_t^k)'s. We find the use of copulas less convincing and more difficult to calibrate; moreover a copula approach typically precludes any Markovian description of $(\vec{\mathbf{X}}_t)$ which is available in the mechanistic description used below.

2.2. Observations Model

A variety of formulations are possible for the dynamics of observations $(\vec{\mathbf{Y}}_t)$. Motivated by sensor arrays that monitor asynchronous "lumpy" events, we model each coordinate (Y_t^k) via a doubly stochastic Poisson process driven by the signal. Precisely, we assume that (Y_t^k) is a point process with intensity $\Lambda^k(X_t^k; t, Z)$. In other words, we have that

$$dY_t^k = 1_{\{t \le \theta^k\}} dN^{k,0}(t) + 1_{\{t > \theta^k\}} dN^{k,1}(t), \qquad k = 1, \ldots, K, \tag{2.4}$$

where $(N^{k,j}(t))$ are conditionally independent counting processes with intensities $\Lambda^k(j; t, Z)$, $j = 0, 1$, where $\Lambda^k(j; t, z)$ are given functions. In the basic case where Λ's are independent of outbreak parameters Z and of t, $(N^{k,j}(t))$ become Poisson processes. Note that this formulation allows both the post-change and pre-change (which is less realistic perhaps) observation characteristics to depend on the changepoint.

As seen by the central decision-maker, the superposition property of counting processes allows to combine all the information sources into a single doubly-stochastic *marked* point process. Indeed, let $\mathcal{Y} = (\sigma_k, \nu_k)$ be a marked point process with arrival times $\sigma_k < \sigma_{k+1} \in \mathbb{R}_+$ and marks $\nu_k \in \{1, 2, \ldots, K\}$. The σ's aggregate all the events observed at the sensors, while the marks ν_k identify which sensor collected the k-th event. Then (Bremaud 1981), $\mathcal{Y}$ has intensity

$$\Lambda_t := \sum_{k=1}^{K} \Lambda^k(X_t^k; t, Z), \tag{2.5}$$

and conditional mark distribution

$$p(k; \vec{x}, t, Z) := \mathbb{P}\left\{\nu_\ell = k | \vec{\mathbf{X}}_{\sigma_\ell -} = \vec{x}, \sigma_\ell = t, Z\right\} = \frac{\Lambda^k(x^k; t, Z)}{\sum_{j=1}^{K} \Lambda^j(x^j; t, Z)}. \tag{2.6}$$

2.3. Performance Criterion

We take the point of view of a policy maker who aims to detect signal presence in the array, i.e. the *minimum* of the change-points

$$\Theta := \min(\theta^1, \ldots, \theta^K),$$

by raising the alarm at decision time $\tau \le \infty$. Since θ^k are not directly observed, we require that the decision is based on available information, namely $\tau \in \mathcal{S}$, where $\mathcal{S}$ denotes the set of all $\mathcal{F}$-adapted stopping times.

Given Bayesian priors, the key objective of the policy maker is to achieve quickest detection while maintaining a bound on false alarm frequency. Additionally, the policy maker accounts for the *quality* of her

detection by making an announcement $d \in \mathbb{R}^{|Z|}$ at the detection date and comparing it to the true signal-type Z. Formally, the controller wishes to minimize a weighted average of expected (linear) detection delay (ADD), probability of false alarms (PFA), and announcement error. For concreteness, we will base our criteria on the detection delay $(\tau-\Theta)^+$, where $a^+ := \max(a, 0)$, the probability of the false alarm $\{\tau < \Theta\}$, and the classification error $f(d, Z)$, where f is a given mis-identification penalty function and the announcement $d \in \mathcal{F}_\tau$ is based on the observations up to τ only.

Denote by $\tilde{\pi}$ the distribution of $Z = (\mathcal{O}, \mathcal{T}_0, \mathcal{V})$ (which equivalently induces some distribution $\vec{\pi}_0$ on $\vec{\mathbf{X}}_0$). We use $\mathbb{P}_{\vec{\pi}_0}$ and $\mathbb{E}_{\vec{\pi}_0}$ to denote the corresponding probability measure and expectation. The Bayesian quickest detection problem is to compute

$$\begin{aligned} V(\tilde{\pi}) &:= \inf_{\tau\in\mathcal{S}, d\in\mathcal{F}_\tau} \mathbb{E}_{\tilde{\pi}} \left\{(\tau-\Theta)^+ + c1_{\{\tau<\Theta\}} + f(d, Z)\right\} \\ &= \inf_{\tau\in\mathcal{S}, d\in\mathcal{F}_\tau} \mathbb{E}_{\vec{\pi}_0} \left\{\int_0^\tau 1_{\{\vec{\mathbf{X}}_s\neq\vec{0}\}} ds + c1_{\{\vec{\mathbf{X}}_\tau=\vec{0}\}} + f(d, Z)\right\}, \end{aligned} \tag{2.7}$$

where $\vec{0} = (0, 0, \ldots, 0) \in \mathbb{R}^K$ is the zero element in E. The first term on the right-hand-side in (2.7) is the average detection delay, the second term is the likelihood of a false alarm given that outbreak is declared at τ, and the third term is the accuracy of identifying the *type* of the outbreak. The parameter c is the tunable penalty for false alarms; small c will induce aggressive detection, while $c \to \infty$ is the case where false alarms are not tolerated; similarly the function f is the tunable penalty for mis-classification.

Remark 2.2. More generally, we may consider the problem of detecting any hitting time of $(\vec{\mathbf{X}}_t)$, i.e. $\Theta := \inf\{t : \vec{\mathbf{X}}_t \in G\}$, $G \subseteq E$. For example, the setting of (2.7) corresponds to $G \equiv \{\vec{0}\}^c$ while $G = \{\vec{x} : x^k = 1\}$ corresponds to detecting the disorder in the k-th channel only. Finally, we could also straightforwardly generalize to linear combinations of objectives such as in (2.7), asking the controller to choose a sequence of pairs (τ_ℓ, d_ℓ) to match some Θ_ℓ's.

If $(\vec{\mathbf{X}}_t)$ was observed (i.e. $\mathcal{F}$-adapted), then the optimal detection rule would simply be $(\tau^* = \Theta, d^* = Z)$; the crux of the problem is therefore to construct a good approximation to Θ using information flow $(\mathcal{F}_t)$ only. From a control perspective, this means that to minimize Bayes risk requires solving a partially observable optimal stopping problem. Indeed, the costs in (2.7) are not measurable with respect to the decision variables (τ, d). Accordingly, the solution approach (Peskir and Shiryaev 2006) is to first perform *filtering* of the latent state $(\vec{\mathbf{X}}_t)$ by computing the posterior distribution

$$\Pi_t(D) := \mathbb{P}_{\vec{\pi}_0}\left\{\vec{\mathbf{X}}_t \in D|\mathcal{F}_t\right\} \qquad D \subseteq \{0, 1\}^K. \tag{2.8}$$

For a Borel set $\mathcal{E}$, denote by $\mathcal{M}(\mathcal{E})$ the set of all probability measures supported by $\mathcal{E}$. Then we may identify Π_t as an element of $\mathcal{M}(\{0, 1\}^K) \equiv \Delta_{2^K} := \{\vec{\pi} \in [0, 1]^{2^K} : \sum_i \pi^i = 1\}$, i.e. a $(2^K - 1)$-dimensional

stochastic process on the simplex Δ_{2^K}. Alternatively, in terms of the model primitives we can consider the filter $(\tilde{\Pi}_t)$, which is a diffuse-measure-valued process on $\mathcal{M}(\mathcal{A})$, such that for any Borel set $\tilde{D} \in \mathcal{B}(\mathcal{A})$,

$$\tilde{\Pi}_t(\tilde{D}) := \mathbb{P}_{\tilde{\pi}}\{Z \in \tilde{D} | \mathcal{F}_t\}. \tag{2.9}$$

An important advantage of using $(\tilde{\Pi}_t)$ over (Π_t) is the Markov property. Indeed, since $(\vec{\mathbf{X}}_t)$ is in general not Markov, neither is its conditional distribution. On the other hand, since Z is constant, the full filter $(\tilde{\Pi}_t)$ is trivially Markov. In other words, to gain the Markov property, one must lift from the restricted filter (Π_t) to the full filter $(\tilde{\Pi}_t)$. In particular, it is immediate that one can evaluate the conditional probability of any $\mathcal{G}$-measurable function $g(z)$ as

$$\tilde{\Pi}_t(g) := \mathbb{E}_{\tilde{\pi}}\{g(Z)|\mathcal{F}_t\} = \int_{\mathcal{A}} g(z)\tilde{\Pi}_t(dz).$$

An important example which allows to project down from $\tilde{\Pi}_t$ to Π_t is $\tilde{\Pi}_t(1_{\{\theta^k \leq t\}}) = \Pi_t(\{X_t^k = 1\})$.

Taking conditional expectations with respect to $(\mathcal{F}_t)$ in (2.7), we end up with the Bayesian performance functional $V(\tilde{\pi}) = \inf_{\tau,d} J_B(\tau, \tilde{\pi})$ with

$$J_B(\tau; \tilde{\pi}) := \mathbb{E}\left\{ \int_0^\tau H^1(s, \tilde{\Pi}_s)\, ds + H^2(\tau, \tilde{\Pi}_\tau) \,\middle|\, \tilde{\Pi}_0 = \tilde{\pi} \right\}, \tag{2.10}$$

where $H^1(t, \tilde{\pi}) := \tilde{\pi}(1_{\{\Theta \leq t\}})$ and H^2 optimizes over the announcement d,

$$H^2(t, \tilde{\pi}) := c\tilde{\pi}(1_{\{\Theta > t\}}) + \inf_d \tilde{\pi}(f(d, Z)).$$

A Bayesian optimal detection rule τ^* is the optimizer in (2.10) and can be viewed as the functional mapping histories to decisions $\tau^* : \mathcal{F}_t \to \{stop, continue\}$. Because the state variable is Markov, the detection rule is simplified to a function of the current $\tilde{\Pi}_t$. This point of view highlights the key challenges in working with (2.10), namely the need to (i) characterize and solve the evolution equations of the filter process $(\tilde{\Pi}_t)$ and (ii) overcome the curse of dimensionality associated with optimizing over the state space of $\tilde{\Pi}_t$. In fact, without further assumptions, $\tilde{\Pi}_t$ is infinite-dimensional making V a functional on the non locally compact space $\mathcal{M}(\mathcal{A})$. Thus, a complete Bayesian solution requires consideration of non-Markov optimal stopping (if working with (Π_t)) or infinite-dimensional Markov optimal stopping problems. The resulting complexity has earned this approach the stigma of analytical and computational intractability.

3. Solution Method

In general, there are no sufficient statistics for the measure-valued process $(\tilde{\Pi}_t)$. The only special cases we are aware of consist of either (a) taking $(\vec{\mathbf{X}}_t)$ be Markov or (b) making the support $\mathcal{A}$ of Z to be finite. Otherwise, the key to characterizing $(\tilde{\Pi}_t)$ are stochastic filtering techniques (Bain and Crisan 2009). Namely,

$(\tilde{\Pi}_t)$ satisfies a variant of the Kushner-Stratonovich nonlinear filtering equation. These equations are typically analytically intractable, and we therefore seek numerical approximations. An efficient and flexible approach to computing $\tilde{\Pi}_t$ is to apply Sequential Monte Carlo methods (SMC), also known as particle filters, which approximate $\tilde{\Pi}_t$ with an empirical particle cloud (Doucet et al. 2001). The main mechanism of SMC consists of a mutation-selection procedure applied to an interacting particle system.

In terms of the control step, since the state variable is $\tilde{\Pi}_t$, analytic characterizations, through, e.g. quasi-variational inequalities, of the resulting value function $V(\tilde{\pi})$ are difficult to come by. Instead we recall the probabilistic characterization of V through its dynamic programming equations. Precisely, define for any stopping time $\sigma > 0$ the monotone operator $\mathcal{J}$ acting on a measurable test function $v : \mathcal{M}(\mathcal{A}) \to \mathbb{R}$ via

$$\mathcal{J}v(\tilde{\pi}) = \inf_{\tau \in \mathcal{S}} \mathbb{E}_{\tilde{\pi}} \left\{ \int_0^{\tau \wedge \sigma} H^1(s, \tilde{\Pi}_s)\, ds + 1_{\{\tau \leq \sigma\}} H^2(\tau, \tilde{\Pi}_\tau) + 1_{\{\tau > \sigma\}} v(\tilde{\Pi}_\sigma) \right\}. \tag{3.1}$$

Then guided by the Bellman optimality principle we have that

Lemma 3.1. *$V(\tilde{\pi})$ is the largest fixed point of $\mathcal{J}$ smaller than $H^2(0, \tilde{\pi})$ and one can approximate V as*

$$V = \lim_{n \to \infty} V_n, \qquad \textit{where } V_n := \mathcal{J} V_{n-1}, \qquad \textit{with} \quad V_0(\tilde{\pi}) = H^2(0, \tilde{\pi}).$$

Moreover, the optimal stopping rule is given by

$$\tau^* = \inf\{t : V(\tilde{\Pi}_t) \geq H^2(t, \tilde{\Pi}_t)\}, \tag{3.2}$$

and can be approximated through $\tau^n = \inf\{t : V_n(\tilde{\Pi}_t)) \geq H^2(t, \tilde{\Pi}_t)\}$.

The optimal stopping problem in (3.1) leads to a representation of $V(\tilde{\Pi}_t)$ as the Snell envelope corresponding to the reward functional $J_B(\tau; \cdot)$ in (2.10). In turn, this gives rise to a novel approach to resolve challenge (ii) through the use of Monte Carlo dynamic programming (MCDP) methods (Egloff 2005, Ludkovski 2009) to solve the dynamic programming equations within a stochastic simulation/regression framework. A simplified version of this approach was studied in Ludkovski (2012) for a special case of (2.7) with a single change-point θ modulating a multivariate jump-diffusion $(\vec{\mathbf{Y}}_t)$. The resulting Monte Carlo algorithm first uses particle filtering to obtain a high-dimensional approximation $(\hat{\Pi}_t^{(N)})$ to the true $(\tilde{\Pi}_t)$ with arbitrarily small errors as the number of particles $N \to \infty$, and then applies MCDP to solve the optimal stopping problem for $(\hat{\Pi}_t^{(N)})$. Overall, we end up with a fully simulation-based solution of the Bayesian formulation, which seamlessly merges SMC inference methods and MCDP for the optimal stopping step.

3.1. General Setup

The schematic of Figure 1 and the ensuing formulation was chosen for simplicity; a much more general setup can be considered within the Bayesian paradigm. Let $\vec{\mathbf{X}}_t = (X_1, \ldots, X_K)$ be a stochastic process

on $\{0,1\}^K$ corresponding to sensor change-point indicators, and $(\vec{\mathbf{Y}}_t)$ be the stochastic process for observations. In addition, let the random vector Z be the collection of all relevant model parameters. Let $\mathcal{H}_t := \sigma((\vec{\mathbf{X}}_s, \vec{\mathbf{Y}}_s), s \leq t) \vee \sigma(Z)$ be the full filtration. Then the Bayesian framework treats $(\vec{\mathbf{X}}_t, \vec{\mathbf{Y}}_t)$ as a coupled process (note that we no longer assume that $\vec{\mathbf{X}}_t$ is $\sigma(Z)$-measurable) which is to be estimated based on observed information $\mathcal{F}_t$ only.

To wit, the evolution of the pure jump process $\vec{\mathbf{X}}_t$ is summarized by its transition times T_ℓ and the discrete skeleton chain (χ_ℓ),

$$\vec{\mathbf{X}}_t = \sum_{\ell=1}^{\infty} \chi_\ell 1_{\{T_\ell \leq t < T_{\ell+1}\}}, \qquad \chi_\ell \in \{0,1\}^K.$$

We assume that T_ℓ and χ_ℓ are described respectively through an intensity function $\vec{\mu}_t$ and jump distribution $\nu(\cdot)$. A minimal requirement is that $(\vec{\mu}_t)$ is $\mathcal{H}$-adapted and the distribution ν of χ_ℓ is $\mathcal{H}_{\sigma_\ell -}$-measurable.

Depending on the actual dependence structure of $(\vec{\mathbf{X}}_t)$ three main cases can now be identified. Mechanistic disorder models correspond to $(\vec{\mathbf{X}}_t)$ being a *deterministic* function of Z (i.e. $\sigma(Z)$-measurable) and can be used to model a random phenomenon that propagates through the sensor array as in Figure 1. The case where $(\vec{\mathbf{X}}_t)$ is an *autonomous Markov* process (possibly in an enlarged space, to allow for time-dependence and self-exciting features) models the situation of an exogenous epidemic process observed through conditionally independent $(\vec{\mathbf{Y}}_t)$. Finally, the case where $\vec{\mu}_t$ and/or ν depend on $(\vec{\mathbf{Y}}_t)$ corresponds to feedback effects where the change-points are (partially) driven by the observations.

Moreover, we can also generalize to more complex observation processes $(\vec{\mathbf{Y}}_t)$. For example, diffusion or jump-diffusion observations can be treated in completely analogous manner and could be used to e.g. model high-frequency biosurveillance data, such as search engine queries or online sales of flu medications. The basic requirement is tractability of the conditional likelihoods $p(\vec{\mathbf{Y}}_t | \vec{\mathbf{X}}_s, 0 \leq s \leq t, Z)$, and conversely of $p(\vec{\mathbf{X}}_t | \vec{\mathbf{Y}}_s, 0 \leq s \leq t, Z)$. These probabilities are always available in closed-form whenever $(\vec{\mathbf{Y}}_t)$ is a marked point process, see (4.2) below. We further discuss some of the possible setups in Section 5.

3.2. Relationship to a Markovian Model

Let us compare the setup of Figure 1 to the simplified model where the disorder indicator process $(\vec{\mathbf{X}}_t)$ is directly taken to be Markov. This is the original approach of Raghavan and Veeravalli (2010) and can be viewed as a Hidden Markov model where the change-point times θ^k are given in terms of hitting times of a latent state $(\vec{\mathbf{X}}_t)$. In the case of two sites $K = 2$, this means that the signal $\vec{\mathbf{X}}_t = (X_t^1, X_t^2) \in E := \{00, 01, 10, 11\}$ is an autonomous Markov chain with known generator Q. The disorder times are $\theta^k = \inf\{t \geq 0 : X_t^k = 1\}$; the observations $(\vec{\mathbf{Y}}_t)$ remain as in (2.4) and all the parameters $Z \equiv (\mu_{ij}, \Lambda^k(j))$ are

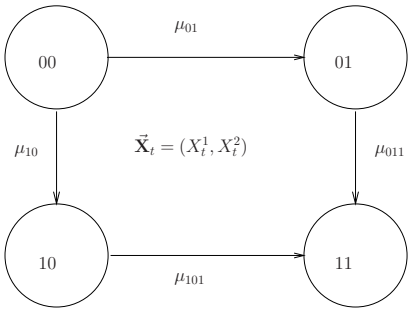


Figure 2: Evolution of the signal state $\vec{\mathbf{X}}_t$. The corresponding infinitesimal generator $Q = (q_{ij})$ is given by

$$Q = \begin{pmatrix} -\mu_{01} - \mu_{10} & \mu_{01} & \mu_{10} & 0 \\ 0 & -\mu_{101} & 0 & \mu_{101} \\ 0 & 0 & -\mu_{011} & \mu_{011} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

known. The resulting evolution is illustrated in Figure 2.

In this case, signal detection reduces to a *finite-dimensional* problem, namely filtering $(\vec{\mathbf{X}}_t)$ is equivalent to finding the posterior distribution (3-dim vector) $\vec{\Pi}_t \equiv (\Pi_t^{\ell_1\ell_2})$,

$$\Pi_t^{\ell_1\ell_2} := \mathbb{P}_{\vec{\pi}_0}\left\{\vec{\mathbf{X}}_t = \ell_1\ell_2 | \mathcal{F}_t\right\}, \qquad \ell_k \in \{0,1\},$$

which provides a full characterization of the system state at date t. The Bayesian formulation can now be analytically characterized (subject to verifying technical conditions) through e.g. quasi-variational inequalities satisfied by the value function. We present an example of such semi-analytical approach in Section 6.1 where $\vec{\Pi}_t$ admits explicit dynamics.

However, the assumption that $(\vec{\mathbf{X}}_t)$ is a finite-state Markov chain imposes severe restrictions (e.g. exponential inter-disorder times) in terms of calibrating the model. Also, unless symmetry is assumed, the number of parameters to specify Q and the modulation of Λ by $\vec{\mathbf{X}}_t$ grows exponentially in the number of sensors K. In Raghavan and Veeravalli (2010) a fixed order was imposed on the relative occurrence of θ^k, leading to a K-dimensional sufficient statistic. Such an assumption corresponds to knowing exactly the identity of the first sensor to observe the disorder and might be unrealistic in many applications.

4. NUMERICAL ALGORITHM

In this Section we propose a new method to approximate the value function of (2.7). Illustrative examples are then provided in Section 6.

4.1. Particle Filtering

To compute the filter (2.9) we utilize sequential Monte Carlo approach to approximate $\tilde{\Pi}_t \simeq \hat{\Pi}_t^{(N)}$, where the discrete measure $\hat{\Pi}_t^{(N)}$ consists of N particles,

$$\hat{\Pi}_t^{(N)} := \frac{1}{W(t)} \sum_{n=1}^{N} w^n(t) \delta_{z^n(t)}(\cdot). \tag{4.1}$$

Above $w^n(t) \in \mathbb{R}_+$ are the particle weights, $W(t)$ is a normalizing constant, and $z^n(t) = (o^n, T^n, v^n)$ are the particle version of all the unknown parameters Z. In other words, any posterior probability is approximated via

$$\mathbb{P}_{\tilde{\pi}}\{Z \in \tilde{D}|\mathcal{F}_t\} \simeq \frac{1}{W(t)} \sum_{n:z^n(t)\in\tilde{D}} w^n(t), \qquad \tilde{D} \subseteq \mathcal{A}.$$

The SMC algorithm is now specified through the recursive evolution of the particles $(w^n(t), z^n(t))_{n=1}^N$, allowing for a sequential (online) update of the particle filter as new information is collected. This evolution is given by the genetic mutation-selection steps. In general, the particles are supposed to mimic Z, so that $z^n(t) = z^n$ is static. Given z^n we can also compute the associated change-point $\theta^{n,k}$, $k = 1, \ldots, K$, and the dynamic disorder indicator $\vec{x}_t^{n,k} := 1_{\{\theta^{n,k}\leq t\}}$. The weights $w^n(t)$ then correspond to the likelihood of observations $(\vec{\mathbf{Y}}_s)_{s\leq t}$ given the particle history $(\vec{x}^n(s))_{s\leq t}$. Using the properties of doubly stochastic marked point processes, we have

$$w^n(t) = w^n(s) \cdot \exp\left(-\int_s^t \Lambda(\vec{x}_u^n)\, du; z^n\right) \cdot \prod_{k:s\leq\sigma_k\leq t} \Lambda(\vec{x}_{\sigma_k}^n) p(\nu_k; \vec{x}_{\sigma_k}^n, z^n). \tag{4.2}$$

As information is collected, most particles will diverge from observations and their weights will collapse $w^n(t) \to 0$. To avoid the resulting particle degeneracy, the SMC approach applies sequential resampling to multiply "good" particles, and cull poor ones, ensuring particle diversity. Thus, we introduce re-sampling instances R_k, $k = 1, \ldots$, at which we draw (with replacement) from the atomic measure $\hat{\Pi}_{R_k-}^{(N)}$ according to the weights $w^n(R_k-)$ and then reset the particle weights to $w^n(R_k) = 1$. A popular approach to select the R_k's is to use the Effective Sample Size (ESS) measure of particle diversity, $ESS(t) = \{\sum_{n=1}^N (w^n(t))^2\}^{-1}$ and resample as soon as ESS drops below a threshold $R_k = \inf\{t \geq R_{k-1} : ESS(t) \leq \overline{ess}\}$.

Due to the static nature of Z and accordingly z^n above, the above resampling in fact does not fully resolve particle degeneracy, since it will simply produce identical copies $z^{n'}$ of particles whose parent was z^n. Thus, as $t \to \infty$, all the particles will eventually become identical. This particle collapse phenomenon is well-known when using SMC for static parameter estimation and calls for making the particles truly dynamic.

Recall that the key quantity for detection are the sensor disorder indicators $(\vec{\mathbf{X}}_t)$ which track the change-points θ^k. Consequently, we adjust our SMC algorithm to treat the particle-specific $\vec{x}_t^n$ (and hence $z^n(t)$) as a dynamic process. The main idea is that while the particles $z^n(t)$ still mimic Z, we assume that until the first disorder $\Theta^n := \min_{k\in\{1,\ldots,K\}}(\theta^{n,k})$, $z^n(t)$ can be resampled freely to maintain particle diversity. Formally, at the resampling date R_k we regenerate the particle parameters according to:

$$z^n(R_k) = \begin{cases} z' \quad \text{where } z' \sim p(Z|\Theta > R_k) \quad \text{if } \Theta^n > R_k, \\ z^n(R_k-) \quad \text{otherwise}. \end{cases} \tag{4.3}$$

In other words, until the particle-specific first disorder time Θ^n the primitive parameters are kept "dormant" and repeatedly resampled (conditionally on $\{t < \Theta\}$). As a result, at any instant t, the particles that have no signal presence are guaranteed to be diverse. Practically, we implement the sampling on first line of (4.3) as rejection sampling by sampling unconditionally from the prior of Z until the condition $\Theta \geq t$ is satisfied.

A further technique to overcome particle degeneracy in post-disorder particles is artificial particle enrichment. Namely, we introduce artificial moves for the constant Z-components following the Liu-West kernel shrinkage scheme (Liu and West 2001). Thus, individual particles experience small moves in time, such that the overall marginal means and variance of the particle filter are kept constant. We refer to our recent work (Ludkovski 2012) for further details and analysis; we find that this method reduces particle degeneracy appreciably at a small computational cost. Algorithm 1 below summarizes particle filtering in sensor arrays. For simplicity it assumes that resampling takes place at arrival dates $R_k = \sigma_k$.

Algorithm 1 Particle Filtering in Sensor Arrays

Input: observations trajectory $(\vec{y}_t)$ consisting of arrival dates (σ_k) and marks (ν_k)

Sample $z^n \sim \tilde{\pi}$, $n = 1, \ldots, N$

Set $w^n(0) = 1$, $n = 1, \ldots, N$

for $k = 1, \ldots$ **do**

Carry out Liu-West step on z^n for $\{n : \vec{x}^n_{\sigma_k} = \vec{0}\}$

for each particle $n = 1, \ldots, N$ **do**

If $\Theta^n > \sigma_k$, re-sample $z^n \sim p(Z|\Theta > \sigma_k)$

Compute disorder indicator process $(\vec{x}^n_t)$ on the interval $t \in (\sigma_k, \sigma_{k+1})$

Calculate weights $w^n(\sigma_{k+1}) \leftarrow w^n(\sigma_k) \cdot \Lambda(\vec{x}^n_{\sigma_{k+1}}) p(\nu_{k+1}; \vec{x}^n_{\sigma_{k+1}}, z^n) \cdot \exp\left(-\int_{\sigma_k}^{\sigma_{k+1}} \Lambda(\vec{x}^n_s, z^n)\, ds\right)$

end for

if $ESS(\sigma_{k+1}) < \overline{ess}$ **then**

Re-sample $n' \propto w^n(\sigma_{k+1})$ for $n' = 1, \ldots, N$

Update $z^n \leftarrow z^{(n')}$

Reset weights $w^n(\sigma_{k+1}) \leftarrow 1$

end if

end for

4.2. Monte Carlo Dynamic Programming

Equipped with the filter of Z, Bayesian sequential detection reduces to solving the optimal stopping problem (2.10) with the Markovian state variable $\tilde{\Pi}_t$. Because $(\tilde{\Pi}_t)$ is high-dimensional, analytic approaches for obtaining V become computationally intractable. Instead we use a robust simulation-based method. Recall that for a discrete-time problem with finite horizon T defined by

$$V^{\Delta}(0,\tilde{\pi};T) := \inf_{\tau\in\mathcal{S}^{\Delta}(T)} \mathbb{E}_{\tilde{\pi}}\left\{\int_0^{\tau} H^1(s,\tilde{\Pi}_s)\,ds + H^2(\tau,\tilde{\Pi}_\tau)\right\}, \tag{4.4}$$

where $\mathcal{S}^{\Delta}(T) = \{\tau\in\mathcal{S} : \tau\in\{0,\Delta t, 2\Delta t,\ldots,(T/\Delta t)\Delta t\}\}$, Bellman's optimality principle implies that

$$V^{\Delta}(t,\tilde{\Pi}_t) = \mathbb{E}\left\{\sum_{s=t/\Delta t}^{\tau^*(t)/\Delta t-1} H^1(s\Delta t,\tilde{\Pi}_{s\Delta t})\Delta t + H^2(\tau^*(t),\tilde{\Pi}_{\tau^*(t)})\big|\,\mathcal{F}_t\right\}; \tag{4.5}$$

$$\text{where}\quad \tau^*(t) = t1_{\{S_t\}} + \tau^*(t+\Delta t)1_{\{S_t^c\}},$$

$$S_t := \left\{H^2(t,\tilde{\Pi}_t) < H^1(t,\tilde{\Pi}_t)\Delta t + \mathbb{E}\left\{V^{\Delta}(t+\Delta t,\tilde{\Pi}_{t+\Delta t})|\,\mathcal{F}_t\right\}\right\}, \tag{4.6}$$

and where $\tau^* = \tau^{*,\Delta}(t)$ is the optimal stopping time conditioned on not stopping before t, and S_t^c is the complement of the set S_t.

By the Markov property, the conditional expectation $\mathbb{E}\left\{V^{\Delta}(t+\Delta t,\tilde{\Pi}_{t+\Delta t})\big|\,\mathcal{F}_t\right\} =: \hat{E}(t,\tilde{\Pi}_t)$ is a function of the measure-valued $\tilde{\Pi}_t$ for some functional $\hat{E} : \mathbb{R}_+\times\mathcal{M}(\mathcal{A})\to\mathbb{R}$. The MCDP method first replaces $V(t+\Delta t,\tilde{\Pi}_{t+\Delta t})$ in the last term of (4.6) with an *empirical pathwise continuation value* $v_{t+\Delta t}$ (computed according to (4.5)). It then implements (4.6) by replacing the conditional expectation operator $\mathbb{E}[\cdot|\mathcal{F}_t]$ (characterized as the L^2-minimizer) with an L^2-projection onto the $span(B_i(\tilde{\Pi}_t) : i = 1,\ldots,r)$,

$$\mathbb{E}\left\{V^{\Delta}(t+\Delta t,\tilde{\Pi}_{t+\Delta t})|\mathcal{F}_t\right\} \simeq \sum_{i=1}^{r}\alpha^i(t)B_i(\tilde{\Pi}_t), \tag{4.7}$$

where $(B_i(\tilde{\pi}))_{i=1}^r$ are the basis functions and $\alpha^i(t)$ the corresponding regression coefficients. This is implemented through a *cross-sectional regression* of a Monte Carlo collection $(v^m_{t+\Delta t})_{m=1}^M$ to find (α^i). Comparing the regression prediction $\sum_i\alpha^i(t)B_i(\tilde{\Pi}_t)$ and the immediate payoff $H^2(t,\tilde{\Pi}_t)$ we then construct the approximate stopping region S_t for (4.6).

Finally, since we do not have access to $(\tilde{\Pi}_t)$, we instead work with the approximate filter $\hat{\Pi}^{(N)}$. Thus, we simulate M realizations (y_t^m) of $(\vec{\mathbf{Y}}_t)$, generating $(\hat{\Pi}_t^{(N),m})$ along each Monte Carlo path using the particle filter above. We then approximate $B_i(\tilde{\Pi}_t^m)\simeq B_i(\hat{\Pi}_t^{(N),m})$ and using backward recursion implement (4.6) by regressing the empirical $(v^m_{t+\Delta t})$ against the simulated $\{B_i(\tilde{\Pi}_t^{(N),m})\}_{m=1}^M$ to obtain the empirical regression coefficients from the simulation of size M, $\alpha^{(M),\cdot}(t)$, and the approximate value function $V^{\Delta}(0,\vec{\pi};T,M,N,r,\Delta t)$.

4.3. Stationary Solution

Returning to our original problem, general theory (Ludkovski 2009) implies that $V^{\Delta}(0, \vec{\pi}; T, M, N, r, \Delta t) \to V$ as $r \to \infty$, $N \to \infty$, $M \to \infty$, $\Delta t \to 0$, $T \to \infty$. To obtain a stationary stopping rule on infinite horizon, we (i) set $\hat{\pi_0} \sim \tilde{\Pi}_{T'} | \Theta > T'$ for T' large enough (this corresponds to a quasi-stationary distribution of the filter conditional on no disorder yet; this distribution can be sampled from by generating paths of $(\vec{\mathbf{Y}}_t)$ constrained to $\Theta > T'$ and filtering along them on $[0, T']$); (ii) solve the finite horizon problem $V^{\Delta}(\hat{\pi}_0; T)$ for T large using the MCDP algorithm; (iii) using the stopping region S_0 and a fresh set of Monte Carlo simulations, evaluate the performance of the resulting detection rule

$$V(\vec{0}) \simeq \mathbb{E}_{\vec{0}} \left\{ (\hat{\tau} - \Theta)^+ + f(d^*, Z) \right\} + c \mathbb{P}_{\vec{0}} \{ \hat{\tau} < \Theta \}, \qquad \hat{\tau} := \inf \left\{ t \geq 0 : \hat{\Pi}_t^{(N)} \in S_0 \right\}.$$

Algorithm 2 summarizes the full procedure in pseudo-code.

4.4. Choice of Algorithm Parameters

The Bayesian detection rule is a map between $\tilde{\Pi}_t$ and the stopping decision. This suggests that to obtain good tests, it is first and foremost important to identify the key features in $\tilde{\Pi}_t$. For example, the posterior probability of no disorders, $\pi(\vec{0})$ directly drives the immediate payoff H^2 and is certainly an important quantity. In the MCDP method, this translates into *parametrizing* candidate tests in terms of the summary statistics used. For instance, if we take $r = 1$, $B_1(t, \tilde{\pi}) = \tilde{\pi}(\{\vec{\mathbf{X}}_t = \vec{0}\})$, the resulting detection test consists of declaring alarms based solely on $\mathbb{P}\{\Theta \leq t | \mathcal{F}_t\}$. In general, this cannot be optimal, since it would imply ignoring the other information in $(\tilde{\Pi}_t)$ while the latter has no finite-dimensional Markovian representation. (We remark that in the model of Raghavan and Veeravalli (2010), a similar one-dimensional detection rule was shown to be asymptotically optimal.)

The choice of the basis functions $B_i(\tilde{\pi})$ and their number r is heuristic. First, we expect r to be large enough so that the resulting projection of the conditional expectation is well-approximated, which can be empirically verified by varying r and checking that the results remain stable. Second, while one could automate the choice of B_i by selecting some functional family and picking the first r terms, we observe that practically speaking, customization to the given problem at hand is desirable in the sense of allowing smaller r. For instance, it is recommended to include the terminal cost $H_2(t, \tilde{\Pi}_t)$ as one of the basis functions. Finally, we find that in practice the choice of the quantities N and M that control algorithm running time can be done independently; namely first pick number of particles N to achieve minimal filtering error; then pick number of simulations M large enough so that the Monte Carlo variance of the MCDP step (monitored

through the regression coefficients $\vec{\alpha}^{(M)}$) is acceptably small. Note that M is primarily determined by r which in turn typically needs to grow (slowly) in the number of sensors K.

5. RELATED MODELS

Our framework of applying stochastic filtering and optimal stopping techniques on quickest detection problems is highly flexible and can handle a variety of modified models. Below we briefly discuss several other cases that fit into the general framework described in Section 3.1.

5.1. Further Examples of Signal/Observation Setups

Local Identification: Rather than identifying the global signal presence, it is often of interest to carry out local detection for a particular changepoint θ^k. The other sensors are then used as secondary information sources and the performance measure is modified to, e.g.,

$$V_1(\tilde{\pi}) := \inf_{\tau \in \mathcal{S}} \mathbb{E}_{\tilde{\pi}} \left\{ \int_0^\tau 1_{\{X_s^1 = 1\}} \, ds + c 1_{\{X_\tau^1 = 0\}} \right\},$$

where now detection delay and false alarms are defined with respect to θ^1 only. Solution of this problem uses the identical method of Algorithms 1 and 2.

Signal Strength: As pointed out earlier, the observed arrival rates $\Lambda(\vec{\mathbf{X}}_t; t, Z)$ may be a function of the parameters Z. For instance, in physical systems with well-defined signal origin, the signal strength observed by each sensor would depend on the distance between signal origin $\mathcal{O}$ and sensor location A_k. In homogenous media, we might therefore model

$$\Lambda^k(X_t^k; t, Z) = \Lambda^k(0) + \frac{a X_t^k}{\|\mathcal{O} - A_k\|^2}.$$

Note that here the statistical profile after disorder, being a function of unobserved $\mathcal{O}$, is no longer exactly known. Of course, such dependence would help to improve detection; a similar adaptive disorder problem for a single change-point was studied in Ludkovski (2012).

A further possibility is time-variable signal strength. For instance, in biosurveillance (Y_t^k) corresponds to count of infected cases at site k; once an epidemic begins at θ^k, the ensuing infectivity rate is nonlinear

$$\Lambda_t^k = \lambda^k(0) + \Lambda((t - \theta^k)^+),$$

where Λ is some specified known function (e.g. through a corresponding ODE). Such dependence between Λ and Z is straightforwardly incorporated into the filtering Algorithm 1. Since each particle is already equipped with its copy z^n of Z, all that is required is appropriate adjustment of the weights $w^n(t)$ in (4.2).

Transient Signals: We could also consider problems with transient signals. For instance, in radar communications the signal corresponds to a target moving through space. Given a radar detection radius r, a target affects observations at sensor k at instant t only if $\|D_t - A^k\| \leq r$, where $D_t = \mathcal{O} + \mathcal{V}t$ is the (for simplicity 1-dim.) location of the target. Thus, each sensor is exposed to two disorder times $\theta^{k,1}$ and $\theta^{k,2}$ with the statistical profile

$$\Lambda_t^k = \begin{cases} \Lambda^k(1) & \text{if} \quad \theta^{k,1} \leq t < \theta^{k,2}; \\ \Lambda^k(0) & \text{otherwise.} \end{cases}$$

In this setup, $(\vec{\mathbf{X}}_t)$ can therefore transition both to "higher" and "lower" (in terms of number of disorders present) states; the rest of Algorithm 1 remains the same. Clearly, temporary disorder makes detection more difficult.

5.2. Dependence between Change-Points and Observations

The traditional Bayesian formulation treats the underlying change-points θ^k as conditionally independent (given true signal parameters Z) of the observed (Y_t^k). However, this is not essential in our framework which allows complex couplings between disorder state $\vec{\mathbf{X}}_t$ and observations $(\vec{\mathbf{Y}}_t)$.

To illustrate the possibilities, we consider an interacting extension of the Markov model of Section 3.2 using a system of bivariate Hawkes processes for each sensor $k = 1, \ldots, K$. We take (Y_t^k) to be a doubly stochastic Poisson process with known pre- and post-disorder intensities $\Lambda^k(0)$ and $\Lambda^k(1)$ and arrivals (σ_ℓ^k), $\ell = 1, \ldots$. The *transition rate* $\mu^k(t)$ of (X_t^k) now has a feedback effect from arrivals in $\vec{\mathbf{Y}}$,

$$\mu^k(t) = \mu_0^k + \sum_{\ell : \sigma_\ell \leq t} a^k e^{-\beta^k (t - \sigma_\ell^k)}, \tag{5.1}$$

for some known constants $a^k \in \mathbb{R}, \beta^k \in \mathbb{R}_+$, $k = 1, 2, \ldots, K$. Thus, the transition rate of (X_t^k) increases by a^k after each arrival σ_ℓ^k; this effect dissipates exponentially at rate β^k. It follows that if $a^k > 0$ then the change-point θ^k is likely to be "triggered" by a cluster of observed events, correlating $(\vec{\mathbf{Y}}_t)$ and $(\vec{\mathbf{X}}_t)$.

As a motivation, consider a biosurveillance setup where (Y_t^k) is the count of observed infections for an endemic pathogen, and θ^k is the (unknown) instant when the disease goes epidemic. Recently, such models have been proposed for understanding the spread of avian H5N1 flu in human populations (Bettencourt and Ribeiro 2008). Currently, avian flu is only (rarely) transmitted from animals to humans; however each time a human is infected, further virus adaptation may result, enabling direct human-to-human transmission and causing widespread epidemic. This creates a positive feedback between observed infections $(\vec{\mathbf{Y}}_t)$ and the epidemic change-point with $a^k > 0$ in (5.1).

On a filtering level, dealing with (5.1) requires straightforward adjustments in the particle filtering Algorithm 1. Evaluating the likelihood of observations $(\vec{\mathbf{Y}}_t)$ conditional on $\vec{\mathbf{X}}_t$ remains elementary. Simulating $(\vec{\mathbf{X}}_t)$ conditional on a trajectory of $\vec{\mathbf{Y}}_t$ can be done in the case of (5.1) through a variant of the Poisson thinning algorithm (Lewis and Shedler 1979). Namely, we equip each particle with a time-dependent intensity function $\mu^{n,k}(t)$ which is updated as new events σ_ℓ arrive and in turn is used to simulate the particle-specific disorder time $\theta^{n,k}$. Given $\theta^{n,k}$, the algorithm then proceeds as before to assign weights $w^n(t)$ and perform resampling on the particle cloud. We note that with (5.1), $(\vec{\mathbf{X}}_t, Z)$ is no longer Markovian, but the tuple $(\vec{\mathbf{X}}_t, \vec{\mathbf{Y}}_t, \vec{\mu}_t)$ is Markov, since the dynamics of the current intensity $\mu^k(t)$ depend only on $\vec{\mathbf{Y}}_t$.

6. NUMERICAL EXAMPLES

6.1. Markovian Model

As a first example we consider a 2-sensor Markovian model, $K = 2$, which provides a semi-analytical benchmark. The disorder indicator process $(\vec{\mathbf{X}}_t)$ takes values in $E = \{00, 01, 10, 11\}$ and has the infinitesimal generator

$$Q = \begin{pmatrix} -0.3 & 0.15 & 0.15 & 0 \\ 0 & -0.5 & 0 & 0.5 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \text{with} \quad \mathbb{P}\{\vec{\mathbf{X}}_0 = 00\} = 1. \tag{6.1}$$

For observations, we take a basic doubly stochastic Poisson model for $(\vec{\mathbf{Y}}_t)$ with the intensities

$$\vec{\lambda}^1 = (\Lambda^1(j)) = \begin{pmatrix} 3 \\ 5 \end{pmatrix} \quad \text{and} \quad \vec{\lambda}^2 = \begin{pmatrix} 5 \\ 10 \end{pmatrix}, \qquad j = 0, 1. \tag{6.2}$$

Thus, the fused statistical profile is summarized by

$$\vec{\Lambda} = (\Lambda(i))_{i \in E} = \begin{pmatrix} 8 \\ 13 \\ 10 \\ 15 \end{pmatrix}, \qquad \vec{p} = (p(\nu, i))_{\nu \in \{1,2\}, i \in E} = \begin{pmatrix} 3/8 & 5/8 \\ 3/13 & 10/13 \\ 5/10 & 5/10 \\ 5/15 & 10/15 \end{pmatrix}. \tag{6.3}$$

The corresponding filter $(\vec{\Pi}_t)$ of $(\vec{\mathbf{X}}_t)$ is 3-dimensional. The following lemma, which is a direct application of the results in Ludkovski and Sezer (2012), characterizes the evolution of $(\vec{\Pi}_t)$.

Lemma 6.1. *The posterior filter $(\vec{\Pi}_t)$ follows the piecewise-deterministic dynamics*

$$\begin{cases} \vec{\Pi}_t = \vec{u}(t-\sigma_k; \vec{\Pi}_{\sigma_k}) \qquad \text{for } \sigma_k \le t < \sigma_{k+1}; \\ \vec{\Pi}_{\sigma_{k+1}} = J_{\nu_{k+1}} \vec{\Pi}_{\sigma_{k+1}-}, \end{cases} \tag{6.4}$$

where the jump operator J is defined coordinate-wise by

$$(J_\nu \pi)^i := \frac{\pi^i \lambda^i p(\nu; i)}{\sum_{j\in E} \pi^j \lambda^j p(\nu; j)}, \qquad i \in E \tag{6.5}$$

with $\lambda^i = \sum_{k=1}^K \Lambda^k(i)$ the arrival intensity in regime $i \in E$ and the vector field $\vec{u}$ given by $u^i(t,\pi) = \frac{\rho^i(t,\pi)}{\sum_{j\in E}\rho^j(t,\pi)}$, with $\vec{\rho}(t,\pi) = \pi e^{t(Q-L)}$ the unnormalized likelihood processes, and $L = \operatorname{diag}(\lambda^1, \dots, \lambda^4)$ the diagonal matrix of all the state intensities λ^i's.

We consider a simple Bayesian risk minimization problem

$$J_B(\tau; \vec{\pi}_0) = \mathbb{E}_{\vec{\pi}_0}\left\{ \int_0^\tau [1 - \Pi_s^{00}] ds + c\Pi_\tau^{00} \right\}, \tag{6.6}$$

where $\Pi_s^{00} = \mathbb{P}\{s < \theta_1 \wedge \theta_2 | \mathcal{F}_s\}$ is the probability of no disorders observed yet and there is no mis-identification penalty $f \equiv 0$. Thus, the Bayesian risk minimization is equivalent to solving the three-dimensional optimal stopping problem $V(\vec{\pi}_0) := \inf_{\tau\in\mathcal{S}} J_B(\tau; \vec{\pi}_0)$. Using Lemma 6.1, we obtain that the value function $V(\vec{\pi})$, $\vec{\pi} \in \Delta_4$ satisfies the variational inequality (in viscosity sense and subject to further technical assumptions)

$$\min\left(\mathcal{L}V(\vec{\pi}) + c(1-\pi^{00}), \pi^{00} - V(\vec{\pi})\right) = 0, \tag{6.7}$$

where

$$\mathcal{L}V(\vec{\pi}) := \sum_{i\in E}\left(\sum_{j\in E} q_{ji}\pi^j - \lambda^i\pi^i + \pi^i \sum_{j\in E}\lambda^j\pi^j\right)\frac{\partial V}{\partial \pi^i} + \sum_{\nu=1}^K \left[(V(J_\nu\vec{\pi}) - V(\vec{\pi})) \sum_{i\in E} \pi^i\lambda^i p(\nu, i)\right].$$

Efficient numerical solution of V is possible using the methods of Dayanik et al. (2008a), Ludkovski and Sezer (2012) which rely on applying Lemma 3.1 and iteratively computing $\mathcal{J}V_n$ using a two-step algorithm consisting of deterministic optimization and interpolation.

Remark 6.1. Another semi-analytic special case corresponds to pure diffusion observations where (Y_t^k) are Brownian motions with drift driven by $(\vec{\mathbf{X}}_t)$. This is the observations setup in Raghavan and Veeravalli (2010). Classical techniques (Bain and Crisan 2009, Ch 3) imply that in this case the filter $(\vec{\Pi}_t)$ satisfies the Zakai equation of nonlinear filtering. The resulting value function $V(\vec{\pi})$ can be again obtained via the quasi-variational inequality (6.7) where now the generator $\mathcal{L}$ is a second-order elliptic differential operator. At least for $K = 2$ these equations can be solved using standard pde methods.

The left panel of Figure 3 shows a sample path of the inference problem corresponding to (6.1)-(6.3). Recall that Lemma 6.1 means that $(\vec{\Pi}_t)$ is piecewise-deterministic; in particular between events, the conditional probability of no disorder rises (Π_t^{00} increases) while at event times σ_k the conditional probabilities of disorders rise, with the precise amount contingent on the observed mark ν_k. We observe that while the intensity of arrivals grows by nearly 60% once both sensors see the signal, there is considerable detection delay and the signal is quite faint. In Figure 3 the filter reacts to the disorder at sensor 1 around $t \simeq 3.4$, and to the disorder at sensor 2 around $t \simeq 4.7$ which confirms the presence of a signal and leads to $\Pi_t^{00} \simeq 0$ for $t > 5$. We note that between the two disorders $\theta^1 \leq t \leq \theta^2$ the filter only weakly detects the true state $\vec{\mathbf{X}}_t = (1, 0)$ ($\vec{\Pi}_t^{10}$ never rises above 30%). Overall, this example illustrates the large degree of noise present in a typical model and the complexity of the filtering problem.

Using the algorithm of Ludkovski and Sezer (2012) we proceed to solve for the resulting Bayes risk (6.6) by a basic fixed-mesh discretization of the state space $\Delta_4 \ni \vec{\Pi}_t$ and computing the value functions $V_n(\vec{\pi})$ until $\|V_n - V_{n-1}\|_\infty \leq 10^{-4}$. The right panel of Figure 3 shows the resulting stopping region S for $c = 10$. As expected, the decision maker stops once Π_t^{00} is "low enough". However, as a testament to the interaction between the sensors, the optimal detection rule is not simply $\inf\{t : \Pi_t^{00} \leq b\}$ for some threshold b, but forms a nontrivial surface in the simplex Δ_4. In particular for $\Pi_t^{00} \in [0.08, 0.11]$, the stopping decision is determined by the other $\vec{\Pi}_t$-coordinates.

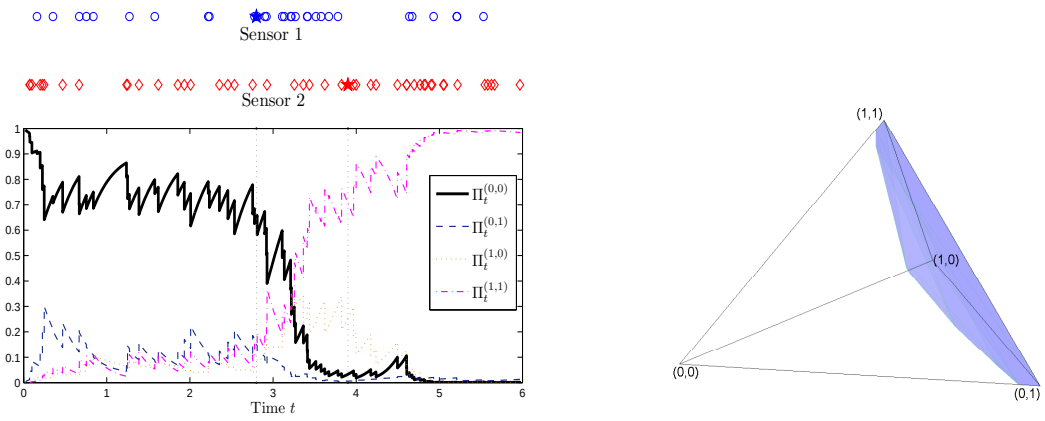


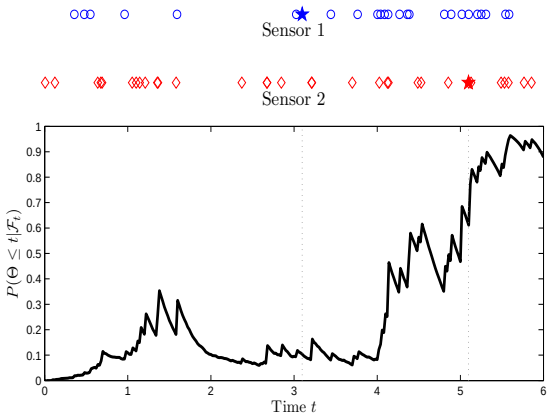
Figure 3: Left: Sample path of the observed arrivals (Y_t^k), $k = 1, 2$ (top) and corresponding 3-dimensional filter $(\vec{\Pi}_t)$ (bottom) obtained from Lemma 6.1. The true disorder dates θ^k are indicated with stars and dashed vertical lines. Right: stopping region $S = \{\vec{\pi} \in \Delta_4 : V(\vec{\pi}) = H^2(0, \vec{\pi})\}$.

6.2. Wavefront Example

As our second example we take a one-dimensional wavefront model for $Z = (\mathcal{O}, \mathcal{T}_0, \mathcal{V})$. Namely $\mathcal{O} \in [0, 2.5]$ with $A_1 = 1$ and $A_2 = 2$ and $\|\mathcal{O} - A_k\| \equiv |\mathcal{O} - A_k|$. We assume independent marginal distributions

$$\mathcal{T}_0 \sim Exp(0.3), \qquad \mathcal{O} \sim Unif(0, 2.5), \qquad \mathcal{V} = \begin{cases} 0.5 \text{ with prob. } 0.5; \\ 1 \text{ with prob. } 0.5. \end{cases} \tag{6.8}$$

The left panel of Figure 4 shows a sample path of the resulting particle filter $(\hat{\Pi}_t^{(N)})$. From our numerical experiments, $N = 2000$ produces a good approximation to the true disorder probabilities Π_t; each path of such $(\hat{\Pi}_t^{(N)})_{t\in[0,10]}$, takes about half a second to generate.



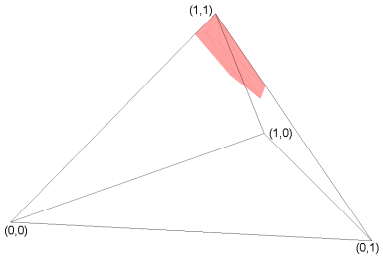


Figure 4: Left: Sample path of the observations (Y_t^k) (top) and corresponding posterior probability of disorder $\mathbb{P}\{\Theta \le t|\mathcal{F}_t\}$ computed using $(\hat{\Pi}_t^{(N)})$ with $N = 2000$. (bottom). The true disorder dates θ^k are indicated with the dashed vertical lines. Right: detection rule projected onto the simplex defined by the posterior probabilities $\Pi_t(1_{\{i\}}) = \mathbb{P}\{\vec{\mathbf{X}}_t = i|\mathcal{F}_t\}$, $i \in E$. The shaded volume indicates the respective stopping region S.

We proceed to study the stopping rule for the pure Bayesian risk minimization problem (again without classification objectives). We use $\Delta t = 0.05$, $T = 8$, $M = 32,000$ and $N = 2000$ particles with the four $r = 4$, basis functions

$$B_1(\tilde{\pi}) = \mathbb{P}\{\Theta \le t|\mathcal{F}_t\}, \quad B_2(\tilde{\pi}) = \mathbb{P}\{\theta^1 \le t|\mathcal{F}_t\},$$
$$B_3(\tilde{\pi}) = \mathbb{P}\{\theta^2 \le t|\mathcal{F}_t\}, \quad B_4(\tilde{\pi}) = \mathbb{P}\{\Theta \le t|\mathcal{F}_t\}^2.$$

Table 1 presents some summary results as we vary the cost of false alarms c. We recall that in this case the total Bayes risk can be decomposed into the probability of false alarm (PFA) and expected detection delay

c	$V(\vec{0})$	$\mathbb{E}_{\vec{0}}\{\tau^*\}$	PFA	EDD
5	1.79	5.05	0.064	1.48
10	2.09	5.51	0.036	1.73
20	2.36	5.86	0.015	2.06

Table 1: Solution of the Bayesian risk minimization problem (2.10) for a range of false alarm costs c. We decompose the value function as $V(\vec{0}) = \mathbb{E}_{\vec{0}}\{(\tau^* - \theta)^+\} + c\mathbb{P}_{\vec{0}}\{\theta \leq \tau^*\} =: \text{EDD} + c \cdot \text{PFA}$.

(EDD). As expected, higher c reduces PFA and increases EDD, as well as the average time until first alarm. For example for $c = 10$, the PFA is about 3.6% while the detection delay is 1.73 time units (corresponding to about 20 arrivals after disorder). The right panel of Figure 4 shows the resulting stopping region S for $c = 10$ projected onto the 3-dim. simplex $\vec{\Pi}_t \in \Delta_4$. Perhaps surprisingly, this region is much smaller (i.e. detection rule is more conservative) than in the Markovian model above and only involves the corner region around $\vec{\mathbf{X}}_t = (1, 1)$. However, we note that it is almost impossible to have $\mathbb{P}\{\vec{\mathbf{X}}_t = (1, 0)|\mathcal{F}_t\}$ or $\mathbb{P}\{\vec{\mathbf{X}}_t = (0, 1)|\mathcal{F}_t\}$ close to one, so in fact the corresponding corner regions are never visited by the paths of $(\tilde{\Pi}_t)$; thus they are also not explored during the MCDP step and the accuracy of the stopping rule is not guaranteed to be high there. In other words, while the computed stopping rule is clearly not accurate around $\mathbb{P}\{\vec{\mathbf{X}}_t = (1, 0)|\mathcal{F}_t\} \simeq 1$, this has negligible effect on its Bayes risk or its (approximate) optimality. We also remark that the exact stopping region is given in terms of the full $\tilde{\Pi}_t$ and the plot is therefore only a projection of this infinite-dimensional object. Since the chosen basis functions $B_i(\tilde{\pi})$ live on Δ_4, Figure 4 does provide a faithful visualization of this approximation.

6.3. Detection and Identification with Multiple Sensors

Our final example explores the impact of deploying variable numbers of sensors for quickest detection and also multi-objective Bayes risk. A signal arises at $\mathcal{T}_0$ from location $\mathcal{O}$ according to the prior distribution in (6.8). As the signal propagates, the background intensity at location A rises from $\Lambda^k(0; Z) = 5$ to a new $\mathcal{O}$-dependent intensity of $\Lambda^k(1; Z) = 5 + \frac{10}{1+|\mathcal{O}-A|^2}$. The aim of the decision maker is to (i) detect the local change-point θ^1 at location $A_1 = 1$ as soon as possible and (ii) identify the signal origin $\mathcal{O}$. For this purpose, she can deploy $K = 1, 2, 4$ sensors at further locations $A_2 = 2, A_3 = 0.5, A_4 = 1.5$. We use a quadratic identification penalty $f(d, Z) = |d - \mathcal{O}|^2$ for $d \in [0, 2]$, which leads to $d^*_\tau = \mathbb{E}\{\mathcal{O}|\mathcal{F}_\tau\}$, so that

	$c_2 = 0$			$c_2 = 10$		
K	$V(\vec{0})$	EDD	$\mathbb{V}ar\{\mathcal{O}\|\mathcal{F}_{\tau^*}\}$	$V(\vec{0})$	EDD	$\mathbb{V}ar\{\mathcal{O}\|\mathcal{F}_{\tau^*}\}$
1	1.18	0.90	0.310	3.51	1.36	0.192
2	0.79	0.76	0.064	1.42	0.85	0.054
4	0.65	0.39	0.038	1.02	0.44	0.032

Table 2: Solution of the Bayesian risk minimization problem with K sensors arranged in a line. The value function is given by (6.9) with $c_1 = 5$. Here EDD := $\mathbb{E}_{\vec{0}}\{(\tau^* - \theta^1)^+\}$.

the full objective is

$$V(\vec{0}) = \inf_{\tau \in \mathcal{S}} \Big(\mathbb{E}_{\vec{0}} \left\{ (\tau - \theta^1)^+ \right\} + c_1 \mathbb{P}_{\vec{0}} \left\{ \tau < \theta^1 \right\} + c_2 \mathbb{V}ar \left\{ \mathcal{O} | \mathcal{F}_\tau \right\} \Big). \tag{6.9}$$

We solve (6.9) using our method with $M = 24000$ and $N = 2000$ and up to $r = 11$ basis functions of the form $\mathbb{P}\{\theta^k \leq t|\mathcal{F}_t\}$, $\mathbb{P}\{\min_k \theta^k \leq t|\mathcal{F}_t\}$ and $\mathbb{V}ar\{\mathcal{O}|\mathcal{F}_t\}$, as well as squares and pairwise products of these posterior probabilities. Table 2 shows that the resulting minimal Bayes risk is highly sensitive to the number of sensors employed. As expected, with more sensors, the detection delay decreases and the posterior variance of $\mathcal{O}$ shrinks by an order of magnitude when comparing $K = 1$ and $K = 4$. We observe that the variable post-disorder intensity together with further sensors more than halves detection delay compared to the previous example.

7. CONCLUSION

Above we have developed a stochastic model for quickest detection in sensor arrays. The key to our formulation is a Bayesian point of view which translates change-point detection into a nonlinear filtering step followed by an optimal stopping step. By approximating the full posterior distribution $(\tilde{\Pi}_t)$ to an arbitrarily high precision, our method remains faithful to the true non-Markovian system dynamics. At the same time, parametrizing the stopping tests using the basis functions $B_i(\tilde{\pi})$ corresponds to employing natural stopping tests that can be easily understood by the policy maker.

Our approach lends itself to a robust numerical implementation that can be easily customized and extended. We focused on a common case where the disorders are triggered mechanistically due to gradual signal propagation; as a result it is natural to carry the inference on the latent "primitive" system parameters. For simplicity we used location and (radial) velocity as our main such parameters; in reality a large range of other specifications (such as signal strength, etc.) could be considered. Perhaps the most interesting

possibility is modeling the interaction between observations and change-points along the lines sketched in Section 5.2. A more detailed analysis of this case will be presented in a separate forthcoming paper.

Acknowledgment

This research was partially supported by the Hellman Family Foundation Research Grant. I am deeply grateful to the editor and the anonymous referee for many suggestions that improved the final manuscript.

REFERENCES

A. Bain and D. Crisan (2009). *Fundamentals of stochastic filtering*, volume 60 of *Stochastic Modelling and Applied Probability*. New York: Springer.

E. Bayraktar and M. Ludkovski (2009). Optimal tracking of a hidden Markov chain under point process observations. *Stochastic Processes and Applications*, 119: 1792–1822.

E. Bayraktar and M. Ludkovski (2010). Inventory management with partially observed nonstationary demand. *Annals of Operations Research*, 176: 7–39.

E. Bayraktar and H. Poor (2008). Quickest detection of a minimum of two Poisson disorder times. *SIAM Journal on Control and Optimization*, 46: 308–331.

L. Bettencourt and R. Ribeiro (2008). Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One*, 3: e2185.

P. Bremaud (1981). *Point Processes and Queues*. New York: Springer.

S. Dayanik, V. Poor, and S. Sezer (2008a). Sequential multi-hypothetis testing for compound Poisson processes. *Stochastics*, 80: 19–50.

S. Dayanik, V. Poor, and S. Sezer (2008b). Multisource Bayesian sequential change detection. *Annals of Applied Probability*, 15: 552–590.

A. Doucet, N. de Freitas, and N. Gordon, editors (2001). *Sequential Monte Carlo methods in practice*. Statistics for Engineering and Information Science. New York: Springer-Verlag.

D. Egloff (2005). Monte Carlo algorithms for optimal stopping and statistical learning. *Annals of Applied Probability*, 15: 1396–1432.

E. Feinberg and A. Shiryaev (2006). Quickest detection of drift change for Brownian motion in generalized Bayesian and minimax settings. *Statistics & Decisions*, 24: 445–470.

G. Fellouris and G. V. Moustakides (2011). Decentralized sequential hypothesis testing using asynchronous communication. *IEEE Transactions on Information Theory*, 57: 534–548.

O. Hadjiliadis, T. Schaefer, and H. Poor (2009a). Quickest detection in coupled systems. In *Proceedings of the 48th IEEE Conference on Decision and Control*, pp. 4723–4728. IEEE.

O. Hadjiliadis, H. Zhang, and H. V. Poor (2009b). One shot schemes for decentralized quickest change detection. *IEEE Transactions on Information Theory*, 55: 3346–3359.

P. Lewis and G. Shedler (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26: 403–413.

J. Liu and M. West (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo methods in practice*, A. Doucet, N. de Freitas, and N. Gordon, eds, Stat. Eng. Inf. Sci., pp. 197–223. New York: Springer.

M. Ludkovski (2009). A simulation approach to optimal stopping under partial observations. *Stochastic Processes and Applications*, 119: 4061–4087.

M. Ludkovski (2012). Monte Carlo methods for adaptive disorder problems. In R. Carmona, P. D. Moral, P. Hu, and N. Oudjane, editors, *Numerical Methods in Finance*, volume 12 of *Springer Proceedings in Mathematics*, pp. 83–112. New York: Springer.

M. Ludkovski and S. Sezer (2012). Finite horizon decision timing with partially observable Poisson processes. *Stochastic Models*, 28: 207–247.

G. Moustakides (2004). Optimality of the CUSUM procedure in continuous time. *The Annals of Statistics*, 32: 302–315.

G. Peskir and A. N. Shiryaev (2006). *Optimal Stopping and Free-boundary problems*. Zurich: Birkhauser-Verlag, Lectures in Mathematics.

M. Pollak and A. G. Tartakovsky (2009). Asymptotic exponentiality of the distribution of first exit times for a class of Markov processes with applications to quickest change detection. *Theory of Probability and its Applications*, 53: 430–442.

A. S. Polunchenko and A. G. Tartakovsky (2012). State-of-the-art in sequential change-point detection. *Methodology and Computing in Applied Probability*, (To Appear). Available at DOI: 10.1007/s11009-011-9256-5.

V. Raghavan and V. V. Veeravalli (2010). Quickest change detection of a Markov process across a sensor array. *IEEE Transactions on Information Theory*, 56: 1961–1981.

R. Rajagopal, X. Nguyen, S. Ergen, and P. Varaiya (2010). Simultaneous sequential detection of multiple interacting faults. Technical report, Arxiv preprint arXiv:1012.1258.

R. Raubertas (1989). An analysis of disease surveillance data that uses the geographic locations of the reporting units. *Statistics in Medicine*, 8: 267–271.

S. Riley (2007). Large-scale spatial-transmission models of infectious disease. *Science*, 316: 1298.

A. G. Tartakovsky and V. V. Veeravalli (2005). General asymptotic Bayesian theory of quickest change detection. *Theory of Probability and its Applications*, 49: 458–497.

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

A. G. Tartakovsky, B. L. Rozovskii, R. B. Blažek, and H. Kim (2006). Detection of intrusions in information systems by sequential change-point methods. *Statistical Methodology*, 3: 252–293.

J. Unnikrishnan, V. V. Veeravalli, and S. P. Meyn (2011). Minimax robust quickest change detection. *IEEE Transactions on Information Theory*, 57: 1604–1614.

V. V. Veeravalli, T. Basar, and H. V. Poor (1993). Decentralized sequential detection with a fusion center performing the sequential test. *IEEE Transactions on Information Theory*, 39: 433–442.

V. V. Veeravalli, T. Başar, and H. V. Poor (1994). Decentralized sequential detection with sensors performing sequential tests. *Mathematics of Control, Signals, and Systems (MCSS)*, 7: 292–305.

Algorithm 2 Disorder Detection in Sensor Arrays

Input: M (number of paths); N (number of particles per path); Δt (time step for Snell envelope); $B_i(\tilde{\Pi})$ (regression basis functions); r (number of basis functions)

for $m = 1, 2, \ldots, M$ **do**

Simulate a trajectory (y_t^m) of the observation process $(\vec{\mathbf{Y}}_t)$ on $[0, T]$

Sample N particles forming $\hat{\Pi}_0^{(N),m}$ from the prior $\tilde{\pi}$ of Z

Use the particle filter Algorithm 2 to compute $\hat{\Pi}_t^{(N),m}$ along the path (y_t^m) for $t = 0, \Delta t, \ldots, T$

Initialize $v_T^m = H_2(T, \hat{\Pi}_T^{(N),m})$, and $\tau_T^m = T$

end for

for $t = T - \Delta t, \ldots, \Delta t, 0$ **do**

Evaluate the basis functions $B_i(\hat{\Pi}_t^{(N),m})$, for $i = 1, \ldots, r$ and $m = 1, \ldots, M$

Regress

$$\vec{\alpha}^{(M)}(t) \triangleq \underset{(\alpha^1,\ldots,\alpha^r)\in\mathbb{R}^r}{\arg\min} \sum_{m=1}^{M} \left| v^m(t+\Delta t) - \sum_{i=1}^{r} \alpha^i B_i(\hat{\Pi}_t^{(N),m}) \right|^2$$

for $m = 1, \ldots, M$ **do**

Set $h^{1,m}(t) = H_1(t, \hat{\Pi}_t^{(N),m})$ and $h^{2,m}(t,) = H_2(t, \hat{\Pi}_t^{(N),m})$

Set $q^m(t) = h_t^{1,m}\Delta t + \sum_{i=1}^{r} \alpha^{(M),i}(t) B_i(\hat{\Pi}_t^{(N),m})$

Set $\hat{S}(t) = \{m : h^{2,m}(t) < q^m(t)\}$ { Empirical Stopping Region}

Set $v^m(t) = q^m(t) 1_{\{\hat{S}^c(t)\}} + h^{2,m}(t) 1_{\{\hat{S}(t)\}}$

Update $\tau_t^m = \tau_{t+\Delta t}^m 1_{\{\hat{S}^c(t)\}} + t 1_{\{\hat{S}(t)\}}$

end for

end for

Generate an independent fresh set of simulations $(y_t^{m'})$, $m' = 1, \ldots, M$

Initialize with $\hat{\Pi}_0^{(N),m'} = \tilde{\pi}$ and find $(\hat{\Pi}_t^{(N),m'})$ along each observation path using Algorithm 1

for $m' = 1, \ldots, M$ **do**

Compute empirical stopping time $\tau^{m'} := \inf\{t : y_t^{m'} \in \hat{S}(0)\}$

Compute empirical stopping cost $v^{m'} := \sum_{j=1}^{\tau^{m'}-1} h^{1,m'}(j\Delta t) + h^{2,m}(\tau^{m'})$

Compute empirical detection announcement $d^{m'}$

end for

return $V(\tilde{\pi}) \simeq \frac{1}{M} \sum_{m'=1}^{M} v^{m'}$.