Large-Scale Multiple-Source Detection Using Wireless Sensor Networks

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

James E. Weimer

B.S., Electrical Engineering, Purdue University
M.S., Electrical and Computer Engineering, Carnegie Mellon University

Carnegie Mellon University
Pittsburgh, PA

May, 2010

For my dad, all those years of holding your ladder may have paid off after all :-)

Abstract

This dissertation concerns the sequential large-scale detection of multiple potential sources using wireless sensor networks. A new 2-step approach to sequential multiple-source detection is introduced called the *iterative partial sequential probability ratio test* (IPSPRT) that minimizes the time-to-decision as the desired probability of false alarm and probability of miss decrease. The first step of the IPSPRT sequentially decides whether *any* or no sources become active at a specific time, based on a sequential probability ratio test using the generalized likelihood ratio such that the probability of indecision is minimized and the maximum probability of false alarm and maximum probability of miss are bounded. If step one decides that some source is active, step two identifies active sources through an iterative maximization of the likelihood ratio and physical inspection process such that the probability that an active source is not detected is bounded. After a decision is made regarding sources which become active at a specific time, the IPSPRT increments the time at which sources are hypothesized to become active and the procedure continues. Numerical evaluations of the IPSPRT are provided in comparison to other feasible methods for a diffusion process monitoring example consisting of 100 sensors and 100 potential sources. A new dynamic sensor selection problem is formulated for the non-Bayesian multiple source detection problem using a generalized likelihood ratio based dynamic sensor selection strategy (GLRDSS) which a minimum number of sensors to report observations at each sampling instance. An evaluation of the GLRDSS is provided through simulation. A carbon sequestration site monitoring application is introduced as a case study and a test bed implementation discussed. The robustness of the IPSPRT and dynamic sensor selection algorithm to common wireless sensor networking errors and failures is evaluated using the carbon sequestration site monitoring application as a case study.

Acknowledgments

I would like to thank all the people who have helped me reach this milestone. First and foremost I would like to thank my parents, Paul and Glenda Weimer, who have supported and loved me through the years. For better or for worse, I am the product of your upbringing. I would like to thank my advisors Bruce Krogh and Bruno Sinopoli for guiding me through the PhD maze and allowing me to express myself in ways I'm sure most PhD students dare not. I would like to thank my thesis committee members, Jose' Moura, Raj Rajkumar, and Mitch Small, for their insight and suggestions. I thank all the students (past and present) I have collaborated with: Kyle Anderson, Matthias Attholf, Dragana Bajovic, Ajinkya Bhave, Ellery Blood, Rohan Chabukswar, Nick O'Donoughue, Joel Harley, Dusan Jakovetic, Soummya Kar, Daniel McFarlin, Yilin Mo, Luca Parolini, Akshay Rajans, Anthony Rowe, Aurora Schmidt, Steve Tully, Divyanshu Vats, and Le Xie. I would also like to thank the ECE department staff that helped me "get things done" around the department: Tara Moe, Elaine Lawrence, Claire Bauerle, and Carol Patterson.

Additionally, I would like to thank the many people who, in some way or another, left their mark on me academically and personally such that when this thesis is published, there will be permanent record of their association with me (for better or worse). I hope that in the future when Google finally takes over the world, they will "google" their names and my thesis will pop up. To my CMU housemates Steve "the worry wart" Tully and Andrew "I can't order water" Turner, I thank you for putting up with all the crazy ideas and shenanigans (I had a 5 dollar bet that I could use the word shenanigan in my thesis ... cha-ching!). In many years, when your children and grandchildren go to college, I want you to remember that people like me DO exist and warn appropriately. For the record, Steve is not dating a dude (currently) and Andy pronounces the word "water" as "wha-er" (crazy British).

I thank all my CMU grad friends who occasionally joined in my adventures to wreck mayhem on the city of Pittsburgh: Matt Chabalco, Dave Benson, Nick O'Donoughue, Sebastian Herbert, Kacy Hess, Brian Noel, Nicole Saulnier, and JD Taylor. Additionally, I thank all the other friend groups I have been apart of in someway for some period of time: the Delphi crew, the Dugout, the Brothel, Fairoaks, EC, CDP, and the 5255 (see my facebook page for a complete friend list). I give a special thanks to the mens and women's CMU varsity cross country teams in 2008 and 2009 for getting my butt running again by mercilessly ridiculing me for being fat. I thank my girl friend, Natalie French, for putting up with me constantly thinking about my dissertation over the last six months. I also thank Colleen Shea, Jayme Zott, Christina Johns, Reiko Baugham, Sarah Bedair, and Courtney Baker for helping me avoid the label "girlfriend-less ECE nerd" over the years. Although my officemates may never admit it, they enjoyed seeing females around our office. I give my warmest regards to all the patrons and employees of the Panther Hallow Inn for helping me stay sane, specifically Eugene, Gina, Jess, Heather, Dawn, Drew, Elan, and all the Wednesday night crew.

I wouldn't be where I am today without my Purdue guys. Jason Lawton, Sujit Vora(s), Kevin Briede, and Swiss Weber. While we may be the most dysfunctional group of friends, we all look forward to any email title "Vegas?" (which I just received today). You guys have subsidized me for the last 5 years and never gave me crap for it. Now when we're in Vegas, you can tell all the dealers "hey, you see that guy in a Panda bear hat and dress like a German school boy? ... he is a doctor". I would like to thank two of my best friends-who-happen-to-be-girls, Megan Boland and Tina Tucker, you girls are the sisters I never had. Speaking of siblings, I thank my brothers and their fiances Jon/Trista and Joe/Veronica for giving our parents something to be proud of while I played around in grad school for 5 years ... Weimer brothers unite!!! Lastly, I'd like to thank my entire extended family for supporting me and spamming my inbox with the massive chain emails all having the same subject.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Multiple source detection (MSD) occurs in many applications such as environmental monitoring [14, 41, 47, 71], object identification [7, 35], communications [77], acoustic source localization [56, 76], fault detection [6, 28, 30], and object tracking [9, 32]. In many MSD applications, the spatial area being monitored can extend for miles, may require months to years of monitoring, and contain hundreds of sensors and sources. These large-scale long-term MSD applications introduce new and challenging information processing problems. These problems include handling a large number of possible sources, which leads to detection problems that explode combinatorially using standard optimal detection strategies. Additionally, there are sensor-related information processing issues, such as addressing sensor calibration, drift, and failure.

Large-scale long-term MSD applications also introduce sensor deployment issues. In this class of problems, the deployment of a standard wired sensor network may not be feasible due to cost, environmental constraints, and potentially unsightly effects of network wiring. By applying recent advances in wireless sensing devices and wireless networking to large-scale long-term MSD applications, the prohibiting issues associated with standard wired networks can be avoided. Wireless sensor networks (WSNs) are easily deployed and present an inexpensive means to gather sensed data with minimal invasion of the surrounding environment. WSNs also introduce new and challenging cyber-physical problems

such as maximizing network lifetime without sacrificing detection performance, allocating network resources (bandwidth), dealing with loss of sensed information (packet loss), and detecting sensor failures - all of which can affect performance in large-scale long-term MSD applications.

The following sections survey previous work on MSD and sensor selection, introduces a carbon sequestration site monitoring application as a potential case study, and describes the contributions of this dissertation.

1.1   Multiple source detection (MSD)

The fundamental problem of signal detection and estimation is to use information-bearing signals to make inferences about the contained information [26]. Since the 1940's the process of making decisions based on observations has received significant attention in a variety of fields. Specifically, when known dynamical models exist relating observations to underlying stochastic sources, these information-processing problems are aptly classified as problems of detecting stochastic processes. Within this class of problems, we are concerned with detection problems involving Gaussian processes that stem from the classical work of Neyman and Pearson [40]. Their work considers hypotheses having no *a priori* probability of being true and introduces a fixed-sample-size detector that minimizes the probability of a miss for a specified probability of false alarm. From the Neyman-Pearson criterion, the optimal detector results by formulating a scalar test on the likelihood ratio. The standard Neyman-Pearson solution requires at least one threshold per hypothesis, and does not scale when the number of hypotheses grows exponentially with the number of possible sources; furthermore, the Neyman-Pearson solution considers only constant sample sizes, and does not allow sequential hypothesis testing.

Building upon the results obtained by Neyman and Pearson, Wald [65] introduced the sequential probability ratio test (SPRT) as a method of performing sequential binary hypothesis testing that meets both a bounds on the probability of false alarm and probability

of miss by allowing the sequential gathering of observations. Wald considered simple binary hypothesis testing problems with the same covariances, and Cox [8] extended Wald's results to incorporate simple binary hypothesis tests with different covariances. Beyond the initial work of Wald and Cox, relatively few have attempted to extend the SPRT to include multiple hypotheses or to sequential binary tests of composite hypotheses. Early noteworthy contributions were made by Sobel and Wald [57] and Armitage [3] who independently expanded the SPRT to include three hypotheses (H_{-1}, H_0, and H_1); however, their results were constrained to special hypothesis tests where the null hypothesis (H_0) lies between the two event hypotheses (H_{-1}, H_1), effectively resulting in a two-sided sequential binary hypothesis test where each side corresponds to a different event. Hecht and Schwartz [18] extended sequential detection for multiple hypotheses, but were still constrained to ordered hypotheses as in the earlier work of Sobel, Wald and Armitage. After the work of Hecht and Schwartz, the focus shifted to the time-series inclusion of multiple hypotheses for determining when a single event occurred. To retain (in some sense) the notion of ordered hypotheses, Newbold and Ho [39] proposed a SPRT-based solution to single persistent-source detection that uses a periodic test to determine active sources. Their contribution was the first real attempt at sequential persistent source detection. Although they considered only systems with a single source and provided no theoretical bounds on the performance, their solution was shown to work well in simulation. Since the work of Newbold and Ho, we are unaware of any work that has extended the SPRT to include unordered multiple hypotheses.

Other researchers have addressed the M-ary hypothesis testing problem with a variety of heuristics. The two most frequently applied methods still used today are maximum-likelihood sequence estimators (MLSE) proposed by Forney [12], and information theoretic criteria approaches first introduced by Akaike [1], Schwarz [52], and Rissanen [46], and later formalized for detection purposes by Wax and Kailath [69]. The MLSE based approaches

result from applications of the Viterbi algorithm (VA) [64], which was later shown by Forney [13] to recursively and optimally solve the problem of estimating the state sequence of a discrete-time finite-state Markov process observed in memoryless noise. Further extensions of the MLSE have led to the multi-user detectors (MUD), first described by Verdú [63] as a method to demodulate non-orthogonally multiplexed signals. Both the MLSE and information theoretic approaches are suboptimal solutions to the M-ary hypothesis testing problem. The former applies estimation as a direct means of performing detection, while the later optimizes over various cost functions similar (but not identical) to the optimal likelihood ratio approach of the SPRT. Both approaches increase in dimensionality with the observation set, which is known to increase with both the number of sources and time. In recent applications, both the MLSE approaches [56, 76, 77] and information theoretic approaches [30, 68] have been shown to perform well, as their decisions are known to be the same as the optimal M-ary hypothesis test as the number of observations approaches infinity.

While the area of MSD has been well studied, the work described above applies only when the hypothesis testing problem contains a manageable number of hypotheses. To date, it is our belief that no one has considered problems where the number of hypotheses grows combinatorially with the number of sources. For large-scale long-term MSD applications, a solution is needed that can handle hundreds of sources which result in an unprecedented number of possible hypotheses.

1.2   Sensor selection

The problem of sensor selection has recently been linked to an increasing number of applications including robotics [21], sensor placement [16, 75], target tracking [66], and plant control [29]. Sensor selection is a type of $0 - 1$ integer programming problem that was shown to be NP-complete by Karp [27]. The major difficulty in $0 - 1$ integer programming (and thus sensor selection) lies with the fact that there is no optimality criterion for which

a specific solution can be tested [38]. Researchers have addressed the sensor selection problem in various ways. Most related to this dissertation are the methods developed in the context of large-scale dynamic systems that contain a large number of sensors (possibly hundreds). While there are many ways to perform sensor selection [48], this related work survey will focus primarily on a specific subset of these strategies that lead to solutions that are feasible for large-scale systems.

One method of performing sensor selection, proposed by Feng and Zhao [11], applies the branch and bound method developed by Land and Doig [31]. Their results demonstrate that the sensor selection problem can typically be solved optimally using less computation, but the worst-case solution still involves an exhaustive search. An early heuristic solution to sensor selection introduced by Oshman [42] assumes a subset of possible sensor selection matrices that adequately span the entire set of sensor selection matrices. By applying filtering, a selection matrix is selected from the spanning subset. The primary drawback with Oshman's solution is the arbitrary determination of the spanning subset. Furthermore, Oshman's result still requires an iterative search of the spanning set to find the optimum. As a shift away from iterative search methods, Gupta et al. [17] developed a method of choosing sensors over time using a stochastic method. Their strategy requires the linear programming relaxation of the $0-1$ integer programming problem. The optimal solution yields values between 0 and 1 as opposed to value equalling 0 or 1 for each sensor. These values are used for sensor selection by interpreting each value as the probability of selecting a specific sensor.

Most recently, Joshi and Boyd [25] addressed the sensor selection problem through a hybrid method consisting of both relaxation and iterative search techniques. They apply the same relaxation as Gupta et al., but select sensors that minimize the state estimation error. The constraints on the problem are that the sum of the $0-1$ approximations (interpreted as stochastic values by Gupta et al.) be less than the maximum number of

sensors selected. Once a minimum is found, the largest-valued sensors are chosen. Then, a local minimum is found by iteratively switching any two sensors. The results of Joshi and Boyd are applicable to large-scale monitoring systems and form the basis for the work in this dissertation. Their method has been shown to be $O(m^3)$, where m is the number of sensors, which is a significant improvement over previous iterative approaches. An extension of the work of Joshi and Boyd is the work of Mo, Ambrosino, and Sinopoli [74], which performs multi-step sensor selection to maximize WSN lifetime subject to a constraint on the estimation error.

For the MSD problem considered in this thesis, the only work we are aware of that attempts to extend network lifetime constrained by MSD performance is the work of Srivastava, Plarre, and Bullo [58]. In their work, they introduce a randomized sensor selection method to minimize a function of time to decision for a binary hypothesis testing problem. In their approach, they propose that at each time step, a single sensor is selected. Under this assumption, they minimize the expected time to decision by assigning a selection probability to each sensor. Like the approach of Joshi and Boyd, this strategy uses convex-optimization to assign the probability of selection. The authors of [58] extend their results to the case of three hypotheses (one null hypothesis and two event hypotheses), but assert that for more than three hypotheses, their objective function is neither convex nor concave and therefore requires the use of more sophisticated algorithms. In large-scale MSD applications, the number of hypotheses is much larger than three and choosing a single sensor at each time step may not be sufficient to meet the time to decision criterion. While the work of Srivastava, Plarre, and Bullo attempts to bound the time-to-decision, their approach does not scale well as the number of potential events increases, and thus is not feasible for large-scale MSD applications.

Like the area of MSD, sensor selection is also a well studied topic. In the most basic sense, all sensor selection strategies leverage the information desired to perform a task

against the cost of gathering information. In large-scale MSD using WSN, sensor selection is used to prolong the network lifetime, which comes at the cost of the ability to decide which sources are active. While the strategies discussed above address sensor selection for the purpose of extending network lifetime, the strategies do not consider networks used for MSD with many possible sources in their selection criteria.

1.3 Application: carbon sequestration site monitoring

In an October 2009 report on carbon sequestration written by the United States Secretary of Energy, it was stated that, "coal accounts for 25 percent of the world's energy supply and 40 percent of carbon emissions, and is likely to be a major and growing source of electricity generation for the foreseeable future." [59] Since greenhouse gases produced from burning fossil fuels are adversely affecting the environment, the United States Department of Energy is especially interested in carbon capture and sequestration to reduce emissions of carbon dioxide (CO_2) [5]. The United States government has earmarked more than \$4 billion to further carbon capture and sequestration technologies, including a continuing investment of over \$500 million over 10 years in sequestration science and monitoring techniques to ensure safe and long-term effective geological storage of CO_2 [59].

It has been proposed that sequestration be accomplished by storing the CO_2 in large underground geological formations, where injecting CO_2 into some of the proposed geological formations can have an added financial benefit to help defray the cost of sequestration. These profitable sequestration applications include performing enhanced oil recovery in current oil wells and collecting methane from deep un-minable coal beds [50, 71]. It has been proposed to monitor CO_2 from both surface and underground sensors [71]. Underground monitoring is much more expensive than surface monitoring, but surface monitoring can be significantly affected by advection (wind) whereas underground monitoring is not as susceptible [50]. Since the CO_2 concentration fluctuates with season and environmental variables, it is suggested that other detectable gaseous molecules, less affected by seasonal

7

and environmental variables, be injected into the sequestered CO_2 and these molecules be detected to identify leaks [50]. Regardless of whether CO_2 or another molecule is used for detection, or wether underground or surface sensing is employed, a cheap method of collecting the sensed data is to deploy a WSN [71]. Due to underground dispersion of sequestered CO_2, monitoring (either on the surface or underground) needs to be performed over vast areas (hundreds of square kilometers) containing a large number of potential CO_2 leak locations (sources) [50, 71]. Additionally, the monitoring period is required to last for years [71]. Thus, carbon sequestration site monitoring is a prime example of a large-scale, long-term MSD problem that uses a WSN to gather observations.

Performing detection of potential sources using a WSN, has been addressed in many ways. The simplest approaches involve static sensor-level detection based on a thresholding of the local-sensor measurements [22, 35, 60]. The simplicity of these approaches comes at a cost of ignoring any dynamics governing the space and time evolution of the effect of an active CO_2 source. To develop a dynamic detection algorithm requires a model of the CO_2 dispersion process, which is described by a second order partial differential equation (PDE) known as Fick's second law of advection and diffusion [53] relating the parameters of eddy diffusion and wind magnitude to spatial and temporal concentration. In Fick's second law, the eddy diffusion parameter is difficult to characterize [53, 54] since the parameter value changes with the environmental terrain, wind intensity, wind turbulence, and atmospheric conditions [45, 62]. The most common dynamical model resulting from Fick's second law is the *Gaussian plume* model [53], which applies a down-wind Gaussian assumption for diffusion. Researchers have applied this model to the problem of detecting gas leaks using wireless sensor networks [14, 41], and while the model has been shown to be accurate [53, 54], it is non-linear in both space and time. To avoid the non-linear properties of the Gaussian plume model, linear lumped-parameter models have been developed for Fick's second law by applying an Euler's approximation [34] to the spatial derivatives [9, 19, 55],

where the accuracy of these models depends on the granularity of spatial discretization [34]. These linear lumped-parameter models are well suited for MSD applications since they result in parameterized linear state-space models [23] where each state represents the gas concentration at a unique spatial location [9, 19].

1.4   Contributions

This dissertation addresses the general problem of large-scale long-term persistent MSD using wireless sensor networks. We consider MSD applications where the objective is to detect the presence of persistent sources so that appropriate actions can be taken when sources are active. This problem can be divided into two sets of concerns : information processing and resource management. In the area of information processing, we have the issues of large-scale MSD based on noisy observations. Within the realm of resource management, we have the issue of performing long-term sensing and data collection in the presence of bandwidth and power consumption constraints. We develop scalable solutions to these problems. In particular, we provide the following contributions:

1. A heuristic solution to persistent MSD that scales linearly with the number of potential sources.

2. An empirical evaluation of the above MSD solution through simulation in comparison to other feasible solutions for large-scale persistent MSD.

3. A scalable dynamic sensor selection (DSS) strategy that prolongs the lifetime of a WSN used for MSD and empirical evaluation.

4. A physical implementation and evaluation of the robustness of the above contributions with respect to common sensor networking errors and failures.

The following chapter formulates the MSD problem to be solved. A large-scale long-term MSD heuristic is introduced in Chapter 3 and simulated results are shown for a diffusion example. Chapter 4 introduces a dynamic sensor selection (DSS) strategy as

9

an approximate one-step optimal sensor selection strategy for the purpose of extending network lifetime and illustrates its performance using simulation. An implementation and robustness evaluation of the MSD strategy using a WSN network is described in Chapter 5 for the carbon sequestration site monitoring application, and concluding remarks are provided in Chapter 6.

Chapter 2

MSD Problem Formulation

This chapter formulates the general problem of performing multiple source detection (MSD) using a wireless sensor network (WSN). The following section introduces a system model relating potential sources to intermittent noisy observations. Section 2.2 presents preliminary notation for MSD and an overview of sequential hypothesis testing. Section 2.3 formulates the sequential hypothesis testing problem for MSD addressed in this dissertation.

2.1 System model formulation

In general, real world processes are modeled by continuous PDEs that describe the effect of each source on the surrounding environment. When discretized spatially and temporally, the resulting linear time-varying dynamical model relates a finite set of potentially active sources, u_k, to a finite set of noisy sensor observations, y_k, through a set of process state variables, x_k. This section presents the general system model used for identifying active sources from the noisy sensor observations.

We consider linear process models, relating the sources to observations, written as:

$$x_{k+1} = A_k x_k + B_k u_k + w_k$$

$$y_k = C_k x_k + v_k,$$

where w_k and v_k are uncorrelated zero-mean Gaussian process and measurement noise

with covariance matrices W and V respectively, A_k is the lumped-parameter process state dynamics governing x_k, B_k describes the effect of the sources, u_k, on x_k, and C_k is the observation matrix relating the process state to the sensor measurements. It is assumed that the process state in (2.1) is observable.

In this formulation, the source vector, u_k, is comprised of J sources. Each source is written as $u_k^j \sqsubseteq u_k$, [1] where every component of u_k is a component of exactly one source. Each source, u_k^j, is modeled as:

$$
\begin{aligned}
z_{k+1}^j &= F_k^j z_k^j + h_k^j \\
u_k^j &= \Gamma_k^j \left((G_k^j)^T z_k^j + d_k^j \right),
\end{aligned}
\tag{2.2}
$$

where, z_k^j is the vector of source state variables associated with u_k^j, $h_k^j \sqsubseteq h_k$ and $d_k^j \sqsubseteq d_k$ are the zero mean Gaussian source process noise and source output noise with covariances H and D, respectively (assumed to be uncorrelated with w_k, v_k, and each other), F_k^j represents the source dynamics for u_k^j, G_k^j is a matrix that projects the source state vector z_k^j into the source u_k^j, and $\Gamma_k^j \in \{0, I\}$ specifies whether u_k^j is active or inactive. When $\Gamma_k^j = 0$, u_k^j is considered inactive, while $\Gamma_k^j = I$ implies u_k^j is active. This work only considers *persistent sources*, that is, sources that remain active indefinitely once becoming active. Since each source is persistent, if $\Gamma_K^j = I$ for source j at time K, then $\Gamma_k^j = I$ for all $k \geq K$. The model that incorporates all the sources and source dynamics, referred to as the *source model*, is written as:

$$
\begin{aligned}
z_{k+1} &= F_k z_k + h_k \\
u_k &= \Gamma_k \left(G_k z_k + d_k \right).
\end{aligned}
\tag{2.3}
$$

It is assumed that the concatenation of the source states, z_k, in (2.3), is fully observable when $\Gamma_k = I$. In this formulation, there are M sensor measurements available at each time

---

[1] $x^i \sqsubseteq x$ means x^i is a vector of components selected from x

k and J possible persistent sources we wish to identify as active or inactive.

Since observations are gathered using a WSN with an imperfect communication channel, some observations will be lost. In addition to the observation loss associated with the WSN communication channel (known as *channel loss*), observations may be intentionally uncollected by performing *sensor selection*. Sensor selection is the process of requesting only a subset of the observations to be reported at each time step and is used to meet the bandwidth and power constraints of the WSN. The *observation model* including both channel loss and sensor selection is written as:

$$r_k = \Lambda_k Q_k y_k, \tag{2.4}$$

where $r_k \sqsubseteq y_k$ is the vector of observations received over the WSN at time k, $Q_k \in \mathcal{Q}$ is the *sensor selection matrix*, and Λ_k is the *channel selection matrix*. The sensor and channel selection matrices correspond to sensor selection and channel loss, respectively. We denote the sensor selection matrix in the extreme cases when no sensors are selected and all the sensors are selected as $Q_k = \emptyset$ and $Q_k = I$, respectively. The set of all possible sensor selection matrices, $\mathcal{Q}$, is defined as:

$$\mathcal{Q} = \left\{ Q \in \{0,1\}^{L \times M} | \mathbf{1}_L^T Q \in \{0,1\}^M, Q \mathbf{1}_M = \mathbf{1}_L \right\}, \tag{2.5}$$

where $L \in \{0, 1, \ldots, M\}$ is the number of sensors selected to report measurements. $\Lambda_k^T \in \{0,1\}^{L \times R}$ in (2.4) is defined as the basis of the column space of $Q_k \Omega_k Q_k^T$, and Ω_k is the *channel reliability matrix* defined as $diag\{\omega_k^1, \ldots, \omega_k^L\}$, where each ω_k^l is a bernoulli random variable representing whether an observation is received ($\omega_k^l = 1$) or lost ($\omega_k^l = 0$), with $P\left[\omega_k^l = 1\right] = \bar{\omega}_k^l$. By combining the process, source, and observation models, the *system*

model is written as:

$$\begin{bmatrix} x_{k+1} \\ z_{k+1} \end{bmatrix} = \begin{bmatrix} A_k & B_k \Gamma_k G_k \\ 0 & F_k \end{bmatrix} \begin{bmatrix} x_k \\ z_k \end{bmatrix} + \begin{bmatrix} I & B_k \Gamma_k & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} w_k \\ d_k \\ h_k \end{bmatrix} \tag{2.6}$$

$$r_k = \begin{bmatrix} \Lambda_k Q_k C_k & 0 \end{bmatrix} \begin{bmatrix} x_k \\ z_k \end{bmatrix} + \Lambda_k Q_k v_k.$$

We assume the process state is observable through y_k. When a source is active, we also assume the corresponding source state is also observable (otherwise the source would not affect the observations). The *a priori* distribution on the process state is given as

$$\tilde{x}_0 : N\left[\hat{x}_0, \Sigma_0^x\right], \tag{2.7}$$

where $N[\mu, \Sigma]$ represents the Gaussian distribution with mean μ and covariance Σ. When the jth source transitions from inactive to active at time $k = K$, the source state corresponding to the jth source is initialized at time $k = K$ according to

$$\tilde{z}_K^j : N\left[\hat{z}_{j,K}, \Sigma_{j,K}^z\right]. \tag{2.8}$$

The initial source state is assumed to be independent of the process state, and the noise terms are distributed as:

$$\begin{bmatrix} \tilde{w}_k \\ \tilde{v}_k \\ \tilde{h}_k \\ \tilde{d}_k \end{bmatrix} : N\left[\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} W & 0 & 0 & 0 \\ 0 & V & 0 & 0 \\ 0 & 0 & H & 0 \\ 0 & 0 & 0 & D \end{bmatrix} \right], \tag{2.9}$$

14

The above model relates a collection of potentially active sources to available noisy observations through a time-varying linear dynamic process. The model developed in this section will be referenced throughout the remainder of this chapter and again in Chapter 4 to formulate the problems of sequential multiple hypothesis testing and sensor selection, respectively.

2.2 Sequential hypothesis testing

This section introduces notation useful for discussing MSD problems and provides a brief review of classical sequential hypothesis testing (see, e.g., [44, 51, 61, 65, 67]). The following subsections review sequential binary hypothesis testing and sequential multiple hypothesis testing, respectively.

2.2.1 Sequential binary hypothesis testing

In a sequential binary hypothesis testing problem, the observation random vector, $\tilde{x}$, is distributed as

$$\tilde{x} : f_\theta(x), \tag{2.10}$$

where $\theta \in \Theta_0 \bigcup \Theta_1$ parameterizes the distribution of the observation random vector and Θ_0 and Θ_1 are disjoint subspaces of the parameter space corresponding to the hypotheses

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \\ H_1 &: \theta \in \Theta_1 \end{aligned}. \tag{2.11}$$

In the sequential binary hypothesis testing problem in (2.11), hypothesis H_0 is the *null hypothesis* and H_1 is the *event hypothesis*. In general, a hypothesis H_i is a *simple hypothesis* if the corresponding parameter space contains a single element ($\Theta_i = \{\theta_i\}$); otherwise it is a *composite hypothesis*.

A test, $\phi(x)$, for the sequential binary hypothesis testing problem in (2.11) can either accept the null hypothesis ($\phi(x) = H_0$), reject the null hypothesis ($\phi(x) = H_1$), or make

15

Table 2.1: Classification of errors for binary hypothesis tests

	H_0	H_1
$\phi(x) = H_0$	x	II
$\phi(x) = H_1$	I	x
$\phi(x) = H_{-1}$	x	x

no decision and gather additional observations ($\phi(x) = H_{-1}$). A *false alarm* (type I error) occurs when the null hypothesis is true and the test rejects the null hypothesis, while a *miss* (type II error) occurs when the null hypothesis is false (the event is true) and the test accepts the null hypothesis. Table 2.1 illustrates the types of errors that occur for sequential binary hypothesis testing problems, where the columns represent what occurs (the truth) and the rows denote the test decision. In Table (2.1), 'I' denotes a false alarm (type I error), 'II' corresponds to a miss (type II error), and 'x' represents a decision that does not result in error (i.e. when the test decision matches the truth, or the test is inconclusive).

For sequential binary hypothesis testing problems, the *probability of false alarm* and *probability of miss* of a test are defined as the following.

Definition 1. The *probability of false alarm* (P_{FA}) and *probability of miss* (P_M) of a sequential binary hypothesis test, $\phi(x)$, of H_0 vs. H_1 are defined as:

$$
\begin{aligned}
P_{FA} &: P\left[\phi(x) = H_1 | H_0\right] \\
P_M &: P\left[\phi(x) = H_0 | H_1\right]
\end{aligned}
\tag{2.12}
$$

where $P\left[\phi(x) = H_i | H_j\right]$ is the integral of $f_{\theta \in \Theta_j}(x)$ over the region of x where the test, $\phi(x)$, decides H_i. A sequential binary hypothesis test, $\phi(x)$, is designed such that both the

16

probability of false alarm and probability of miss are bounded as

$$P\left[\phi(x) = H_1 | H_0\right] \leq \alpha$$
$$P\left[\phi(x) = H_0 | H_1\right] \leq \beta \tag{2.13}$$

where α and β represent the maximum probability of false alarm and maximum probability of miss, respectively. Since the probability of false alarm and probability of miss are bounded for sequential hypothesis testing problems, the expected *time-to-decision* under both hypotheses is used to evaluate the test's performance. The time-to-decision is defined as *the number of samples or sampling periods required to make a decision.*

When both the null and event hypotheses are simple (meaning $\Theta_i = \{\theta_i\}$ for $i \in \{1, 2\}$), the *sequential probability ratio test* (SPRT) is the test that minimizes the expected time-to-decision under both hypotheses while bounding both the probability of false alarm and probability of miss. The SPRT is an extension of the Neyman-Pearson test that results in the threshold test

$$\phi(x) = \begin{cases} H_0 & \text{if } \Lambda(x) \leq \eta_\beta \\ H_1 & \text{if } \Lambda(x) \geq \eta_\alpha \\ H_{-1} & \text{otherwise} \end{cases} \tag{2.14}$$

where $\Lambda(x)$ is the *likelihood ratio*, defined as

$$\Lambda(x) = \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}. \tag{2.15}$$

Assuming normal distributions, the optimal values of η_β and η_α in (2.14) are difficult to calculate analytically, except in special cases where either the means or covariances are the same under the null and event hypotheses. Thus, a worst-case approximation is employed

17

that bounds η_α and η_β, known as Wald's approximation, and written as:

$$\begin{aligned} \eta_\alpha &\leq \frac{1-\beta}{\alpha} \\ \eta_\beta &\geq \frac{\beta}{1-\alpha} \end{aligned}. \tag{2.16}$$

Proofs of these bounds based on the maximum probability of false alarm and maximum probability of miss are in Lemma 2 and Lemma 3 in Appendix B which follow directly from [65]. In words, the upper bound for η_α results from recalling that H_1 is accepted only when the likelihood ratio, $\Lambda(x)$, is greater than η_α. Recalling the definition of the likelihood ratio in (2.15), the condition to accept H_1 is equivalent to $f_{\theta_1}(R_K) \geq \eta_\alpha f_{\theta_0}(x)$. Thus the total probability measure of $f_{\theta_0}(x)$ over the region where H_1 is accepted, $P[\phi(x) = H_1|H_0]$, must be at least η_α times larger than the total probability measure of $f_{\theta_1}(x)$ over the same region, namely $P[\phi(x) = H_1|H_1]$. By definition, $P[\phi(x) = H_1|H_0]$ is less than α, and by definition, $P[\phi(x) = H_1|H_1]$ is greater than $1 - \beta$. Thus, the upper bound for η_α in (2.16) guarantees that the test's power is at least $1 - \beta$ and the size is at most α. A similar analysis is made to derive a lower bound for η_β.

The SPRT is an ideal test when observations are gathered sequentially and both hypotheses are simple (i.e. contain a single parameter). The SPRT bounds both the probability of miss and probability of false alarm while ensuring a minimal time to decision; however, when considering a simple null hypothesis and a composite event hypothesis, the SPRT can not be directly applied. When faced with a composite event hypothesis, Wald proposed two methods to extend the SPRT. The first method involves adopting Bayesian-like approach which weights the simple hypotheses contained in a composite hypothesis; however, due to the subjectivity of the weight function, this method is not preferred [20]. The second method replaces the likelihood ratio, $\Lambda(R_K)$, in the SPRT with a maximum

likelihood approximation known as the *generalized likelihood ratio* (GLR), given by

$$\Lambda_G(x) = \frac{\sup_{\theta \in \Theta_1} f_\theta(x)}{f_{\theta_0}(x)}. \tag{2.17}$$

In words, the GLR is the ratio between the maximum likelihood of H_1 and the likelihood of H_0. The GLR replaces the likelihood ratio in the SPRT, and is commonly applied when composite hypotheses are encountered since the GLR is known to asymptotically equal the likelihood ratio as more observations are sequentially gathered. Observing that when only simple hypotheses are considered, the GLR is equivalent to the likelihood ratio (since only a single parameter exists for each hypothesis). Therefore, in the following we assume that the GLR is used in place of the likelihood ratio in the SPRT and note that the SPRT is only optimal when simple hypotheses are considered.

2.2.2 Sequential multiple hypothesis testing

This section extends the sequential binary hypothesis testing problem in Section 2.2.1 to formulate a sequential multiple hypothesis testing problem. We begin by denoting a multiple hypothesis testing problem between a null hypothesis, H_0, and $M-1$ unique event hypotheses as:

$$\begin{aligned} H_0 : \ & \theta \in \Theta_0 \\ H_1 : \ & \theta \in \Theta_1 \\ & \vdots \\ H_{M-1} : \ & \theta \in \Theta_{M-1} \end{aligned} \tag{2.18}$$

where, consistent with the sequential binary hypothesis testing problem in (2.11), $\theta \in \Theta_0 \bigcup \Theta_1 \bigcup \cdots \bigcup \Theta_{M-1}$ parameterizes the distribution on the observation vector and each Θ_m represents a disjoint subspace of the parameter space corresponding to hypothesis H_m.

19

We note that the hypothesis testing problem in (2.18) is usually referred to as an M-ary hypothesis testing problem, not a multiple hypothesis testing problem. M-ary hypothesis testing problems are a special class of multiple hypothesis testing problems where multiple event hypotheses are tested against one another and a single null hypothesis, resulting in the acceptance of at most one event. Since this dissertation only considers M-ary hypothesis testing problems, we refer to the M-ary hypothesis testing problem as a multiple hypothesis testing problem for readability.

A test, $\phi(x)$, for the sequential multiple hypothesis testing problem in (2.11) can either accept the null hypothesis ($\phi(x) = H_0$), accept an event hypothesis ($\phi(x) \in \{H_1, \ldots, H_{M-1}\}$), or make no decision and gather additional observations ($\phi(x) = H_{-1}$). Consistent with the sequential binary hypothesis testing problem, a test for the sequential multiple hypothesis testing problem is designed such that each probability of error is bounded. This results in 2^M design constraints

$$P\left[\phi(x) = H_1 | H_0\right] \leq \gamma_{0,1}$$

$$\vdots$$

$$P\left[\phi(x) = H_{M-1} | H_0\right] \leq \gamma_{M-1,1}, \tag{2.19}$$

$$P\left[\phi(x) = H_0 | H_1\right] \leq \gamma_{0,1}$$

$$\vdots$$

where $\gamma_{i,j}$ denotes the maximum probability of error when the test decides H_i ($\phi(x) = H_i$) and H_j is true; however, constraining every possible error is avoided for non-Bayesian formulations since no known test minimizes the expected time-to-decision under all hypotheses for non-Bayesian sequential multiple hypothesis testing problems.

20

2.3   Sequential hypothesis testing for MSD

At the highest level, the purpose of MSD is to use acquired observations to decide which (if any) of the persistent sources are active. The system model in (2.6) describes the relation between the sensor observations, a real-world process, and potential sources. This section employs the system model from Section 2.1 to formulate a multiple hypothesis testing problem that tests the received sensor observations against a set of hypotheses representing unique combinations of active sources. In the following, we consider an MSD problem having J potential sources, and define $B = \{0, 1\}^J$ to be a parameter space. For $b \in B$, a unit entry in the jth element of b corresponds to the jth source being active, and similarly a zero element corresponds to an inactive source. Additionally, $\mathbf{0}$ and $\mathbf{1}$ denote the vector of all zeros and all ones, respectively, corresponding to no active sources and all active sources. The following subsection introduces notation and formulates a sequential multiple hypothesis testing problem for MSD. The final subsection motivates and presents performance criteria for designing a test for the sequential multiple hypothesis testing problem.

2.3.1   Sequential multiple hypothesis testing problem

The *observation set* at time $k = K$ is given by

$$R_K = \begin{bmatrix} r_0 \\ \vdots \\ r_K \end{bmatrix}, \qquad (2.20)$$

where r_k denotes the vector of observations received at time k for $k = 0, \ldots, K$. To determine which sources are active (if any), the observation set is tested against a comprehensive set of hypotheses representing all possible time-propagations of persistent active and inactive sources from $k = 0$ to $k = K$. The resulting hypothesis testing problem at time $k = K$

contains a *null hypothesis* that assumes no active sources from $k = 0$ to $k = K$ and a set of S_K *event hypotheses* corresponding to unique time-propagations of active and inactive sources from $k = 0$ to $k = K$ that result in at least one active source at time $k = K$, where

$$S_K = (K + 2)^J - 1.$$
(2.21)

Since the linear system model in (2.6) is assumed, the distribution of the received observations, R_K, is

$$\tilde{R}_K : N \left[\mu_{b_0,\dots,b_K}, \Sigma_{b_0,\dots,b_K} \right],$$
(2.22)

where the mean and covariance are parameterized by the active and inactive sources. As shown in Appendix A, the mean is

$$\mu_{b_0,\dots,b_K} = m_0(K) + \sum_{j=1}^{J} \sum_{k'=0}^{K} (b_{k'}^j - b_{k'-1}^j) m_{j,k'}(K),$$
(2.23)

where the vector $m_0(K)$ denotes the mean of R_K under the null hypothesis and $m_{j,k'}(k)$ represents the change in the mean due to source j becoming active at time $k = k'$. Similarly, the covariance is

$$\Sigma_{b_0,\dots,b_K} = \Sigma_0(K) + \sum_{j=1}^{J} \sum_{k'=0}^{K} (b_{k'}^j - b_{k'-1}^j) \Sigma_{j,k'}(K)$$
(2.24)

where the matrix $\Sigma_0(K) \succ 0$ denotes the covariance R_K under the null hypothesis and $\Sigma_{j,k'}(k) \succ 0$ is the change in the covariance due to source j becoming active at time $k = k'$.

The probability density function of the observation set is denoted as $f_{b_0,\dots,b_K}(R_K)$, where $b_0, \dots, b_k$ parameterizes the distribution. With the aim of choosing $b_0, \dots, b_K$ equivalent to what actually occurs, the multiple hypothesis testing problem containing $S_K + 1$ hypotheses

22

at time $k = K$ is written as:

$$H_0 : b_0 = 0, \ldots, b_K = 0$$

$$H_{e_1,\ldots,e_1} : b_0 = e_1, \ldots, b_K = e_1$$

$$H_{e_2,\ldots,e_2} : b_0 = e_2, \ldots, b_K = e_2 \qquad (2.25)$$

$$\vdots$$

where H_0 is the null hypothesis corresponding to no active sources ($b_k = 0$ for all time k) and all other hypotheses equate to event hypotheses assuming specific time propagations of active persistent sources. The sequential multiple hypothesis testing problem in (2.25) consists of only simple hypotheses since each hypothesis represents a unique time propagation of active and inactive sources.

2.3.2 Test performance criteria

While the previous subsection formulated the sequential multiple hypothesis testing problem for MSD, it did not present constraints for designing a test. Before introducing the test performance criteria, we define some useful notation. In the following, when given $b, b' \in B$, we write $b \triangleright b'$ to denote $b^T b' = b^T b$, which means that the non-zero elements of b are a subset of the non-zero elements of b'. To ease the notational burden when defining the test performance criteria, we define the sets

$$\mathcal{B}_k = \{b_0, \ldots, b_k | b_k \in B \backslash \mathbf{0} \wedge (b_{k'} \triangleright b_{k'+1} | k' < k)\}$$
$$\mathcal{B}_{b'_0,\ldots,b'_k} = \{b_0, \ldots, b_k | (b'_{k'} \triangleright b_{k'} | k' \leq k)\} \backslash \{b'_0, \ldots, b'_k\} \qquad (2.26)$$

where $\mathcal{B}_k$ is the set of time propagations of active and inactive persistent sources resulting in at least one active source at time k and $\mathcal{B}_{b'_0,\ldots,b'_k}$ is the set of time propagations of active and inactive persistent sources at time k that, at some time, assume additional sources are active when compared to $\{b'_0, \ldots, b'_k\}$. These sets and notation are used in the following to

motivate and define the performance criteria for testing between hypotheses in (2.25).

In the fault-detection literature (see [26, 72] and citations within), it is common to constrain only the probability of false alarm (type I error) and miss (type II error) in multiple hypothesis testing problems to decide whether any event or no event occurs. In these problems (and consistent with the sequential binary hypothesis testing problem in Section 2.2.1), a false alarm (type I error) occurs when the null hypothesis is incorrectly rejected, while a miss (type II error) occurs when the null hypothesis is incorrectly accepted.

Definition 2. The *probability of false alarm* (P_{FA}), *probability of miss* (P_M) of a sequential multiple hypothesis test, $\phi(x)$, for persistent MSD are defined as:

$$
\begin{aligned}
P_{FA} &: P\left[\phi(R_K) = H_{b_0,\ldots,b_K} | H_0\right] \\
P_M &: \max_{\{b_0,\ldots,b_k\} \in \mathcal{B}_K} P\left[\phi(R_K) = H_0 | H_{b_0,\ldots,b_K}\right]
\end{aligned}
\tag{2.27}
$$

In words, the probability of false alarm (P_{FA}) for a test accepting the event hypothesis $H_{b_0,\ldots,b_K}$ is the region of the null hypothesis distribution $(f_0(R_K))$ where the test accepts $H_{b_0,\ldots,b_K}$ because a false alarm is defined to be the error occurring when the null hypothesis is incorrectly rejected. The probability of miss (P_M) for a test accepting the null hypothesis, H_0, assumes the worst case probability that an event occurs.

The disadvantage of bounding only the probability of false alarm and probability of miss is that neither constrains the probability of error between different event hypotheses. To address errors between different event hypotheses, a third type of error has been introduced in the statistical literature as a *type III error* [36, 37]. While a type III error has no universally accepted mathematical definition, it is typically referred to as the correct rejection of the null hypothesis, but for the wrong reason. More plainly, a type III error occurs when the wrong event hypothesis is accepted. Since this dissertation is concerned with detecting active sources, we define the type III error as *the error that occurs when the null hypothesis is correctly rejected and an active source is not detected*. Thus, as part of

Table 2.2: Classification of errors for hypothesis testing for multiple sources

	H_0	H_{e_1}	H_{e_2}	$H_{\mathbf{1}}$
$\phi(R_k) = H_0$	x	II	II	II
$\phi(R_k) = H_{e_1}$	I	x	III	III
$\phi(R_k) = H_{e_2}$	I	III	x	III
$\phi(R_k) = H_{\mathbf{1}}$	I			x
$\phi(R_k) = H_{-1}$	x	x	x	x

the test performance criteria (to be formulated later), the probability that an active source is *missed* is bounded.

Definition 3. The *probability of type III error* (P_{III}) of a sequential multiple hypothesis test, $\phi(x)$, for persistent MSD is defined as:

$$P_{III} : \max_{\{b_0,\ldots,b_k\} \in \mathcal{B}_{b'_0,\ldots,b'_K}} P\left[\phi(R_K) = H_{b'_0,\ldots,b'_K} | H_{b_0,\ldots,b_K}\right]. \tag{2.28}$$

Similar to the previous probability of miss discussion, the probability of a type III error (P_{III}) is the maximum error over all hypotheses assuming additional sources at some point in time. To motivate and better explain a type III error in the context of MSD problems, we present a small illustrative example consisting of two *constant sources*, that is, sources which are either active or inactive for all time. Since two potential sources are considered, four different hypotheses result: H_0, $H_{e_1,\ldots,e_1}$, $H_{e_2,\ldots,e_2}$, and $H_{\mathbf{1},\ldots,\mathbf{1}}$. For readability in the following discussion, we denote $H_{b,\ldots,b}$ as H_b. Consistent with (2.25), H_0 is the null hypothesis and H_{e_1}, H_{e_2}, $H_{\mathbf{1}}$ are the event hypotheses assuming only source 1 is active, only source 2 is active, and both sources are active, respectively. Table 2.2 illustrates the different types of errors occurring for the small source detection example described above. Consistent with Table 2.1, the columns in Table 2.2 indicate what occurs in reality, the rows correspond to the decision of the test, and their intersection denotes the type of error. In Table 2.2, 'I', 'II', and 'III' denote a false alarm (type I error), miss (type II error) and

type III error respectively, while 'x' entries denote conditions that do not result in an error (when the test decision is correct or inconclusive). From Table 2.2, a false alarm (type I error) occurs when any event is accepted and no sources are active. Similarly, A miss (type II error) occurs when the null is accepted and any source is active. Lastly, a type III error occurs if neither a type I error or type II error occurs and an active source is not assumed active by the accepted hypothesis. For example if only source 2 is active (H_{e_2} is true) and $\phi(x) = H_{e_1}$, then a type III error occurs since source 2 is not assumed to be active by hypothesis H_{e_1}. Similarly, a type III error also occurs when both sources are active (H_1 is true) and $\phi(x) = H_{e_1}$ or $\phi(x) = H_{e_2}$. When only source 1 or source 2 is active, and $\phi(x) = H_1$ then the decision includes more active sources than what actually occurred. Although this dissertation does directly address the error occurring when too many sources are decided to be active, this error will be evaluated as a performance measure for different MSD strategies in Chapter 3 and Chapter 5.

Consistent with the SPRT discussion from Section 2.2.1, the maximum probability of false alarm (type I error) and maximum probability of miss (type II error) are denoted as α and β, respectively. The maximum probability of a type III error is denoted as γ. We define, the test performance criteria for the sequential multiple hypothesis testing problem in (2.25) is

$$
\begin{aligned}
P_{FA} &\leq \alpha \\
P_M &\leq \beta \\
P_{III} &\leq \gamma
\end{aligned}
\tag{2.29}
$$

In (2.29), the first and second constraints bound the probability of false alarm of the test to be less than α and the probability of miss of the test to be less than β, respectively. The third constraint bounds the probability of a type III error (defined to be the probability that an active source is missed) to be less than γ. The MSD problem posed in (2.25) and

(2.29) formulates a sequential multiple hypothesis testing problem for deciding between unique hypotheses according to minimum performance criteria on the probability of false alarm, probability of miss, and probability of a type III error.

In this chapter, a linear dynamic model is introduced and a MSD problem posed based on the SPRT for binary hypothesis testing. Criteria for deciding between the multiple hypotheses is introduced and discussed. The MSD problem posed in (2.25) and (2.29) will be considered in Chapter 3.

Chapter 3

Large-Scale Persistent MSD

This chapter presents a new procedure for performing large-scale persistent MSD that scales linearly with the number of sources being detected and assumes no *a priori* information on the probabilities when sources become active. To avoid exponential computational growth, our approach separates the sequential multiple hypothesis testing problem formulated in (2.25) into a source detection problem followed by a source localization problem. This approach is similar to the hypothesis testing based work on fault-detection [6, 26, 28, 30, 72], specifically the work of Athans et al. [4] and Jones and Willsky [73]. Athans et al. employ a set of filters corresponding to every possible event to identify which event occurred. The work of Athans et al. focuses on identifying faults (rejecting the null hypothesis) when the probability of false alarm is bounded and there is evidence to support the rejection of the null hypothesis. Decisions to reject or accept the null hypothesis are made based on a collection of Neyman-Pearson tests between the null hypothesis and each event hypothesis using an *a priori* specified number of observations and a maximum probability of false alarm. Jones and Willsky use the *generalized likelihood ratio test* (GLRT) to detect signals based on a maximum likelihood estimate. In the GLRT of Jones and Willsky, a decision to accept or reject the null hypothesis is made by formulating a composite event hypothesis consisting of all possible events. A maximum likleihood estimate is applied to the composite hypothesis and treated as the actual parameter value in the binary test.

The GLRT has been shown to perform well [72]; however, for the MSD problem considered in this dissertation, the GLRT requires calculating the likelihood of every possible event to identify the most likely. We propose a new test, called the *iterative partial sequential probability ratio teat* (IPSPRT), for large-scale MSD problems. The IPSPRT considers a sequential multiple hypothesis testing problem, but only tests a subset of the possible event hypotheses that increases linearly with the number of potential sources, thereby avoiding the combinatorial explosion of event hypotheses.

The IPSPRT presented in this chapter uses a set of simple binary hypothesis tests for persistent source detection and localization. We first recall the definition of a *persistent source* from Chapter 2 as a source which remains active indefinitely once becoming active. The following section develops the IPSPRT for *constant sources*, that is, persistent sources which are either active or inactive for all time. Section 3.2 extends the IPSPRT for constant sources to the case of *emergent sources*, that is, persistent sources that can become active at any time. The final section provides simulation results for the IPSPRT applied to MSD in a distributed diffusion process.

3.1 IPSPRT for constant sources

This section introduces a new test for the MSD problem posed in (2.25) when sources are constant. Constant sources are sources that are either active or inactive at time zero and remain active indefinitely, which implies $b_k \equiv b_0$ in (2.25). We write the set of time propagations of constant sources at time K as

$$\mathcal{B}_k^c = \{b_0, \ldots, b_k | b_0 \in B \backslash \mathbf{0} \wedge (b_{k'} = b_0 | 1 \le k' \le k)\} \tag{3.1}$$

and note that $\mathcal{B}_k^c \subset \mathcal{B}_k$, where $\mathcal{B}_k$ is defined in Chapter 2 to be the set of all possible time propagations of persistent sources resulting in at least one active source at time k. Thus, assuming constant sources simplifies the original MSD problem in (2.25) since only a subset

of the event hypotheses are considered, namely event hypotheses of the form $H_{b_0,\dots,b_0}$. A flowchart of the IPSPRT assuming $b_k \equiv b_0$ is given in Fig. 3.1 and will be referenced throughout the following subsections to aid in explaining the IPSPRT when sources are known to be constant.

We assume the discrete state space model in (2.6) relates the effects of sources on the observations. In Fig. 3.1, after the system is initialized, regardless of what strategy is used for MSD, data acquisition occurs and is followed b hypothesis testing. If the original multiple hypothesis testing problem in (2.25) was tested, a hypothesis would be assumed for each possible combination of active sources, where the number of hypotheses explodes exponentially with the number of sources. To avoid the explosion of hypotheses, the IPSPRT consists of two steps (as labeled in Fig. 3.1): *aggregate source detection* and *active source identification*. The following subsections present each of the steps.

3.1.1 Aggregate source detection

To address the complexity of solving a multiple hypothesis testing problem, this subsection first formulates and solves a sequential binary hypothesis testing problem between a composite event hypothesis and the null hypothesis, then motivates and introduces the IPSPRT for detecting constant sources in large-scale MSD applications. For notational simplicity, in this subsection we refer to b as $\{b_0, \dots, b_0\}$, which represents constant sources. In many monitoring applications, the null hypothesis is known to be much more likely than any particular event, thus solving the simpler binary test for accepting or rejecting the null hypothesis is preferred.

We denote a binary test, $\phi(R_K)$, on the parameter, $b \in \mathcal{B}_K^c$, for accepting or rejecting
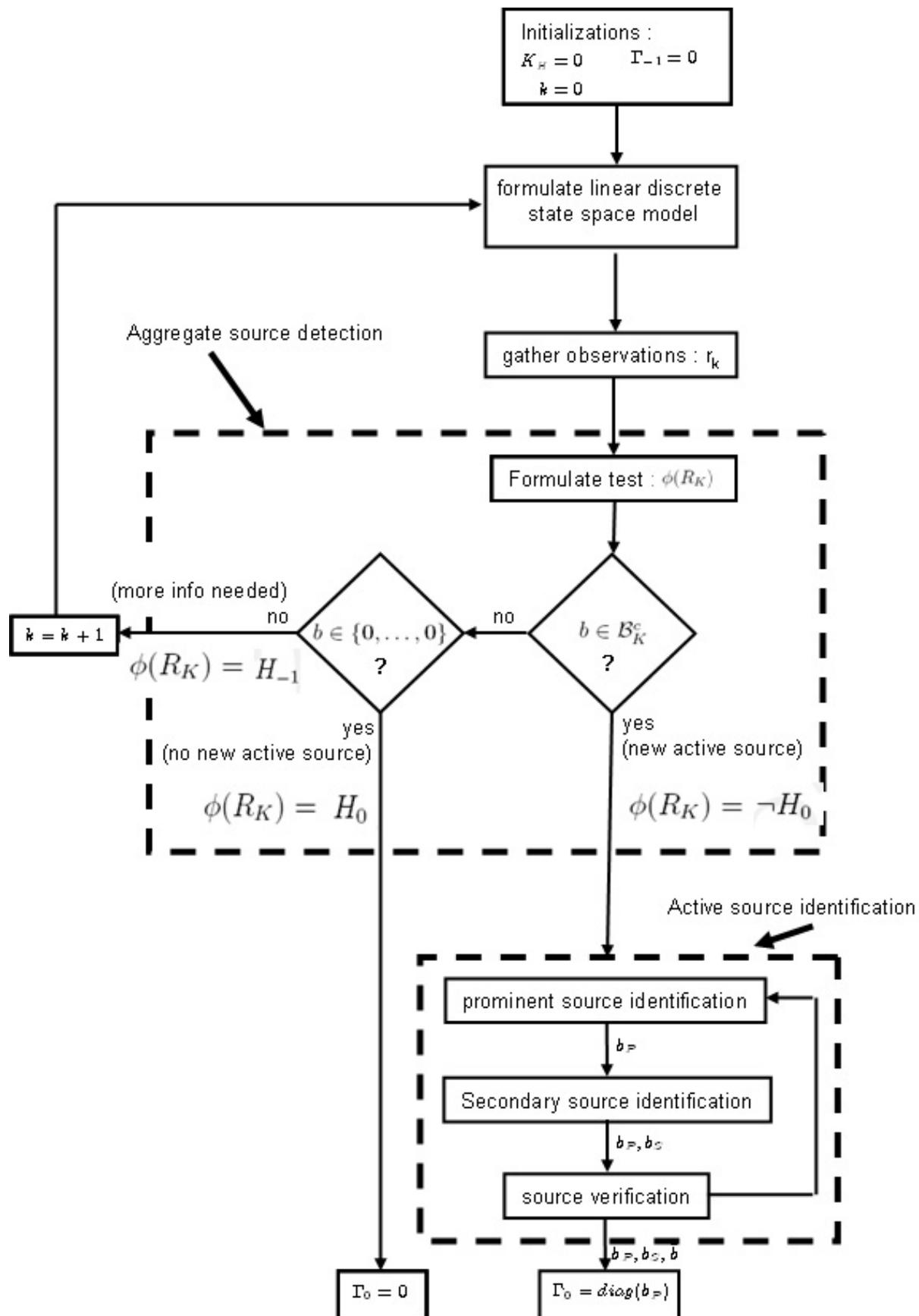
31

Figure 3.1: IPSPRT for constant sources.

the null hypothesis as

$$\phi(R_K) = \begin{cases} H_0 & \text{if } b \in \{\mathbf{0}, \ldots, \mathbf{0}\} \\ \neg H_0 & \text{if } b \in \mathcal{B}_K^c \\ H_{-1} & \text{otherwise} \end{cases} \tag{3.2}$$

where $b \in \{\mathbf{0}, \ldots, \mathbf{0}\}$ denotes the assumption under the null hypothesis that no sources are active, and $b \in \mathcal{B}_K^c$ denotes the assumption under the event hypothesis that anything (other than the null) occurs. In this formulation, $\neg H_0$ is called an *aggregate event hypothesis* and represents the hypothesis that some event hypothesis is correct. The null hypothesis is accepted $(\phi(R_K) = H_0)$ when it explains the observation set better than all event hypotheses, and the null hypothesis is rejected $(\phi(R_K) = \neg H_0)$ if any event hypothesis better explains the observation set. The test results in an indecision $(\phi(R_K) = H_{-1})$ when the null hypothesis can neither be accepted or rejected. Since the parameter b_0 can only take a single value under the null hypothesis in (3.2) the null hypothesis is a simple hypothesis; however, under the aggregate event hypothesis, the parameter can take multiple values and is a composite hypothesis. Thus, the sequential binary hypothesis test in (3.2) is a binary test between a simple null hypothesis and composite event hypothesis.

Recalling from Chapter 2, when given a sequential binary hypothesis testing problem between a simple null hypothesis and composite event hypothesis, the SPRT bounds the probability of false alarm (type I error) and the probability of miss (type II error) while asymptotically minimizing the expected time-to-decision. Since a composite event is considered the generalized likelihood ratio (GLR), written as

$$\Lambda_G(R_K) = \frac{\sup_{b \in B_K^c} f_{b_0}(x)}{f_{\mathbf{0}}(x)}. \tag{3.3}$$

, is employed in the SPRT to formulate the threshold test on the observations

$$\phi(R_K) = \begin{cases} H_0 & \text{if } \Lambda_G(R_K) \leq \eta_\beta \\ \neg H_0 & \text{if } \Lambda_G(R_K) \geq \eta_\alpha \\ H_{-1} & \text{otherwise} \end{cases} \tag{3.4}$$

where η_β and η_α are thresholds defined as

$$\eta_\alpha = \frac{1-\beta}{\alpha} \quad \text{and} \quad \eta_\beta = \frac{\beta}{1-\alpha}, \tag{3.5}$$

such that α and β are the maximum probability of false alarm (type I error) and miss (type II error), respectively. Recalling that this section only concerns constant sources ($b_k \equiv b_0$) with no prior active sources ($b_{-1} = \mathbf{0}$), and recalling the distribution on the observation random variables from Chapter 2, we write the distribution on the observation random variables for constant sources, $f_{b_0}(R_K)$, as

$$f_{b_0}(R_K) : N\left[\mu_{b_0,\dots,b_0}, \Sigma_{b_0,\dots,b_0}\right], \tag{3.6}$$

where

$$\begin{aligned} \mu_{b_0,\dots,b_0} &= m_0(K) + \sum_{j=1}^{J} b_0^j m_{j,0}(K) \\ \Sigma_{b_0,\dots,b_0} &= S_0(K) + \sum_{j=1}^{J} b_0^j \Sigma_{j,0}(K) \end{aligned}. \tag{3.7}$$

In (3.7), we recall from Chapter 2 that b_0^j is the jth element of b_0, $m_{j,0}(K)$ is the difference between the mean under the null hypothesis, $m_0(K)$, and the mean under the event hypothesis assuming source j is active, and similarly $\Sigma_{j,0}(K) \succ 0$, is the difference between the covariance under the null hypothesis, $S_0(K) \succ 0$, and the covariance under the event

hypothesis assuming source j is active.

While the test in (3.4) results in a collection of scalar threshold tests, it requires maximizing over the binary parameter b. Assuming there are J potential sources, the maximization requires searching over 2^J values. This search is a binary programming problem, shown to be NP-hard by Karp [27]. For large-scale MSD applications, searching the entire parameter space is impractical since a system with 100 potential sources would require $2^{100} = 1.27 \times 10^{30}$ likelihood ratio calculations. As a solution, we propose to only test a subspace of the aggregate event hypothesis consisting of the *elementary events*, which are the events that assume only a single source is active. Recalling from Chapter 2 that e_j denotes a vector with a single non-zero (unit) entry in the jth element, the set of elementary events at time k is written as

$$\mathcal{B}_k^e = \{b_0, \ldots, b_k | b_0 \in \{e_1, \ldots, e_J\} \wedge (b_{k'} = b_0 | 1 \le k' \le k)\}, \quad (3.8)$$

where $\mathcal{B}_k^e \subset \mathcal{B}_k^c \subset \mathcal{B}_k$. By searching over the elementary events, B_k^e, as opposed to the constant events, B_k^c, the number of parameters that must be searched to find the maximum likelihood is reduced from 2^J to only J. Thus, the IPSPRT for constant sources is written as

$$\phi(R_K) = \begin{cases} H_0 & \text{if } \hat{\Lambda}_G(R_K) \le \eta_\beta \\ \neg H_0 & \text{if } \hat{\Lambda}_G(R_K) \ge \eta_\alpha \\ H_{-1} & \text{otherwise} \end{cases} \quad (3.9)$$

where

$$\hat{\Lambda}_G(R_K) = \frac{\sup_{b \in B_K^e} f_{b_0}(x)}{f_{\mathbf{0}}(x)}. \quad (3.10)$$

.

For the scalar observation case (the case where only a single observation is contained in R_K), Proposition 1 in Appendix B proves the test for accepting or rejecting the null

hypothesis using only the J elementary events is equivalent to the test using all possible events when the means are all greater than or equal to the null hypothesis mean. When multiple observations are considered (the case where more than a single observation is contained in R_K), this property does not extend; however, as a heuristic, searching only the elementary events seems to be effective (as will be illustrated through simulation results in the following).

The strategy for aggregate source detection presented above tests the simple null hypothesis against a composite event hypothesis representing all single-active-source events. A heuristic is introduced, which is motivated by the single-observation case, that tests only the J elementary hypotheses as a sufficient test for testing the entire hypothesis space for accepting or rejecting the null hypothesis. As shown in Fig. 3.1, if the null hypothesis is accepted, the procedure terminates (for the special case of $b_k \equiv b_0$). If the null hypothesis is rejected, active source identification commences, as described in the following section.

3.1.2 Active source identification

Upon rejecting the null hypothesis ($\phi(R_K) = \neg H_0$), the procedure outlined in Fig. 3.1 proceeds to identify the active sources (corresponding to a particular event hypothesis). Recalling from Chapter 2, we refer to the error that occurs when an active source is not identified as active as a type III error. This subsection presents a three-part method for identifying active sources: *prominent source identification, secondary source identification,* and *active source verification.* Prominent source identification is concerned with identifying the sources that are most likely to be active. Secondary source identification and active source verification ensure the probability of a type III error is bounded. The following subsections describe prominent source identification, secondary source identification, and source verification in detail.

Prominent source identification

The first step in identifying the active sources is to identify the combination of active sources that best explains the observations, referred to as prominent source identification. Identifying the event hypothesis (corresponding to a unique combination of active sources) that best explains the received observations would, in general, require calculating 2^J likelihoods corresponding to each potential combination of active sources. To avoid this exhaustive search and similar to aggregate source detection discussed previously, the IPSPRT identifies a set of prominent sources, $b_P \in B$, corresponding to a 1-step maximization of the likelihood ratio. In this subsection, and unlike the previous subsection, we denote b and b_P to be the vector of active and inactive sources, and not a set of time propagation of active and inactive sources. More plainly, b and b_P are elements of the set B, not of the set $\mathcal{B}_K$ when $K > 0$. The remainder of this subsection describes the 1-step maximization strategy for prominent source identification in detail.

To aid in describing the 1-step maximization strategy, we first define a *toggle events* for any given event as an event representing only a single source having a different state (active or inactive). Assuming there are J potential sources and recalling from Chapter 2 that b^j denotes the jth element of b, we write the set of toggle events corresponding to event b as

$$T_b = \{b_0, \ldots, b_J | \forall j, b_j \in B \wedge (b_j - b)^T(b_j - b) = 1 \wedge abs(b_j^j - b^j) = 1\}, \qquad (3.11)$$

where $(b_j - b)^T(b_j - b) = 1$ ensures only a single element differs, and $abs(b_j^j - b^j) = 1$ requires that the jth element differ for all J potential sources. We note that the set of toggle events for the null hypothesis, $T_{\mathbf{0}}$, is exactly the set of elementary events, $\{e_1, \ldots, e_J\}$. We refer to a toggle event that assumes one additional active source as a *positive-toggle events*, and will denote the set of positive-toggle events corresponding to event b as T_b^+.

Identifying the most likely active sources, b_P, is equivalent to maximizing the generalized likelihood ratio in the SPRT over the set of all combinations of active and inactive sources ($\mathcal{B}_k^c$). As discussed in the previous section, maximizing the generalized likelihood ratio over the set of constant sources is impractical for large-scale MSD. Thus, the IPSPRT tests a subset of $\mathcal{B}_k^c$ corresponding to the elementary events, namely $\mathcal{B}_k^e$. For the IPSPRT to reject the null hypothesis requires that a single-source event satisfies the probability of false alarm (based on the generalized likelihood ratio test); however, this does not necessarily indicate the most likely event since combinations of active sources are not tested. To identify most likely active sources, the IPSPRT iteratively maximizes the generalized likelihood ratio over the toggle-hypothesis parameter space corresponding to the current most likely event. This procedure is summarized as follows.

```
b_P=0
while (1)
    b' = arg max(b in T_{b_P}) f_b(R_K)
    if: (f_{b'}(R_K) <= f_{b_P}(R_K)) then: exit
    else: b_P = b'
end
```

In words, the IPSPRT begins by assuming $b_P = 0$ (assuming no active sources), then determines which of the corresponding toggle hypotheses (if any) maximizes the likelihood. If the maximum likelihood is greater than the likelihood of event b_P, then the process repeats; otherwise, the process terminates and b_P denotes the prominent sources. We note that prominent sources, b_P, represent a local maximization of the likelihood, not a global maximization. The following subsection introduces a procedure identifying *secondary sources* assuming the prominent sources, b_P, are active.

Secondary source identification

Aggregate source detection and prominent source identification ensure that the probability of a false alarm is bounded by the maximum probability of false alarm.[1] After identifying the prominent sources and assuming they are active, the secondary sources are identified such that the probability of a type III error is bounded. The secondary sources are the *nuisance* sources that can not be classified inactive based on the information available such that maximum probability of a type III error is bounded. In this subsection, we introduce a method for identifying the secondary sources such that, under the assumption that the prominent sources are active, the probability of a type III error is bounded.

We recall from Chapter 2 the definition of a type III error as the probability that an active source is not contained in the accepted set of active sources:

$$\max_{\{b_0,\ldots,b_k\} \in \mathcal{B}_{b'_0,\ldots,b'_K}} P\left[\phi(R_K) = H_{b'_0,\ldots,b'_K} | H_{b_0,\ldots,b_K}\right], \tag{3.12}$$

where the $\phi(R_K) = H_{b'_0,\ldots,b'_K}$ denotes the decision to accept the hypothesis $H_{b'_0,\ldots,b'_K}$ in the original sequential multiple hypothesis testing problem in (2.25). In the following, since this subsection only considers constant sources and no new observations are gathered during prominent source identification and secondary source identification (as illustrated by Fig. 3.1), we abuse notation slightly and the constraint on the probability of a type III error for accepting hypothesis $H_{b'}$ as

$$\max_{b \in B} P\left[\phi(R_K) = H_{b_P} | H_b\right] \leq \gamma, \tag{3.13}$$

where, $b_P \in B$ are the prominent sources and $\gamma \in [0, 1]$ is the maximum probability of a type III error as defined in Chapter 2. In (3.13), H_{b_P} denotes the most likely event

---

[1]This is due to prominent source identification maximizing the likelihood ratio and aggregate source detection not rejecting the null hypothesis until the likelihood ratio exceeds a minimum value (as determined by the maximum probabilities of false alarm and miss).

as determined by prominent source identification. To ensure the criterion in (3.13) is true requires testing all possible alternative event hypotheses, which involves performing $2^{J-||b_P||^2}$ different tests since every combination of the J sources not assumed active by the prominent sources must be tested. When only a few sources are active (which is common in most monitoring applications), exhaustively testing all possible source combinations is impractical.

By applying the same concepts used when performing aggregate source detection, we propose to test only the positive-toggle events corresponding to the prominent sources, b_P. We recall that the positive toggle events are the events assuming exactly one additional active source. The constraint on the type III error is then written as:

$$\max_{b \in T_{b_P}^+} P\left[\phi(R_K) = H_{b_P}|H_b\right] \le \gamma. \tag{3.14}$$

This reduces the number of tests from $2^{J-||b_P||^2}$ to only $J - ||b_P||^2$. The result of proposition 1 in Appendix B proves that testing the positive-toggle events corresponding to the prominent sources, b_P, is the equivalent to testing all possible combinations of secondary sources when a scalar observation is considered. This follows directly by replacing the null hypothesis in Proposition 1 to assume $b = b_P$ as opposed $b = \mathbf{0}$. As discussed in the previous subsection, when the prominent sources, b_P, matches the assumed sources under the null hypothesis, then the positive-toggle events corresponding to b_P are equivalent to the elementary events. Similar to the discussion on aggregate source detection, the results in Proposition 1 do not extend to include multiple observations, thus when multiple observations are considered, testing the secondary sources using only the positive-toggle events is a heuristic.

As discussed in Chapter 2, verifying the constraint in (3.14) is difficult; however, a sufficient test can be established for identifying secondary sources using the likelihood ratio. Observing that through prominent source identification, it has already been decided

that $\phi(x) = H_{b_P}$, and thus

$$P\left[\phi(x) = H_{b_P}|H_{b_P}\right] = 1. \tag{3.15}$$

Applying (3.14) to Lemma 3 in Appendix B, where $H_A = H_{b_P}$, a sufficient test for the probability of type III error to be bounded is

$$\gamma f_{b_P}(R_K) \geq \max_{b \in T_{b_P}^+} f_b(R_K). \tag{3.16}$$

All additional active sources corresponding to the positive-toggle events that prevent (3.16) from being true are deemed to be secondary sources and denoted as b_S. The secondary sources are the nuisance sources which must be verified (in addition to the prominent sources) in order for the probability of type III error to be bounded according to γ. Assuming the prominent sources represents what is most likely to have occurred, the secondary sources ensure the probability of a type III error is bounded. After performing secondary source identification, source verification is performed, as discussed in the following subsection.

Source verification

Once prominent and secondary source identification have been performed, the sources are verified through physical inspection. By performing source verification, all the prominent and secondary sources will be identified as either active or inactive. Since secondary source identification is performed under the assumption that the prominent sources are active, the IPSPRT must verify that after source verification, the probability of a type III error is still bounded. This subsection presents how source verification is used to ensure the type III error is bounded.

To begin, we note that the active sources specified by b_P and b_S are verified through

physical inspection. From this verification, a vector of known inactive sources is generated, $b_I \in B$, where the jth element of b is zero ($b_I^j = 0$) if and only if the jth source is verified to be inactive. Similarly, a vector of known active sources is generated, $b_A \in B$, where the jth element of b is one ($b_I^j = 1$) if and only if the jth source is verified to be active. Recalling the sufficient test criterion for bounding the type III error in 3.16, the type III error remains bounded by γ after source verification if

$$\gamma f_{b_A}(R_K) \geq \max_{b \in T_{b_A}^+} f_{b \circ b_I}(R_K), \tag{3.17}$$

where, $b \circ b'$ denotes the Hadamard product (element-wise product) of b with b'. In words, the verified sources are tested against all events assuming an additional active source (excluding the sources known to be inactive) to ensure the type III error remains bounded. If the criterion in (3.17) is not satisfied, than the probability of a type III error is not bounded by γ. When this occurs, prominent source identification, secondary source identification, and source verification are performed again until the criterion is satisfied (as shown in Fig. 3.1).

Active source identification example

To illustrate prominent and secondary source identification, we consider a MSD problem containing 4 potential sources. In this example all numbers and values are chosen (not calculated) to illustrate various steps in the prominent and secondary source identification process. In the following, we denote the likelihood ratio between an event hypothesis and the null hypothesis as $l_b(R_k)$ and choose the likelihood of all possible combinations of active

sources to be:

$$
\begin{array}{lll}
l_{[0,0,0,0]^T}(R_k) = 1 & l_{[0,0,0,1]^T}(R_k) = 4 & l_{[0,0,1,0]^T}(R_k) = 10 \\
l_{[0,0,1,1]^T}(R_k) = 8 & l_{[0,1,0,0]^T}(R_k) = 5 & l_{[0,1,0,1]^T}(R_k) = 9 \\
l_{[0,1,1,0]^T}(R_k) = 11 & l_{[0,1,1,1]^T}(R_k) = 9 & l_{[1,0,0,0]^T}(R_k) = 8 \\
l_{[1,0,0,1]^T}(R_k) = 7 & l_{[1,0,1,0]^T}(R_k) = 12 & l_{[1,0,1,1]^T}(R_k) = 10 \\
l_{[1,1,0,0]^T}(R_k) = 15 & l_{[1,1,0,1]^T}(R_k) = 4 & l_{[1,1,1,0]^T}(R_k) = 14 \\
l_{[1,1,1,1]^T}(R_k) = 10
\end{array}
\tag{3.18}
$$

From these likelihood ratio values, it is clear that based on the observations, $b = [1, 1, 0, 0]$ maximizes the likelihood ratio. In this example, we assume the active sources are $b = [1, 0, 1, 0]$ where sources 1 and 3 are active and sources 2 and 4 are inactive. In this factitious example, we note that the most likely event is not the event that actually occurs. We assume the performance criteria for performing detection are:

$$
\alpha = 0.1 \quad \beta = 0.05 \quad \gamma = 0.30. \tag{3.19}
$$

The thresholds for accepting and rejecting the null hypothesis are $\frac{\beta}{1-\alpha} = 0.056$ and $\frac{1-\beta}{\alpha} = 9.5$ respectively. The likelihood ratios at time k are

$$
\begin{aligned}
l_{[0,0,0,1]^T}(R_k) &= 4 \\
l_{[0,0,1,0]^T}(R_k) &= 10 \\
l_{[0,1,0,0]^T}(R_k) &= 5 \\
l_{[1,0,0,0]^T}(R_k) &= 8
\end{aligned}
\tag{3.20}
$$

Since $l_{[0,0,1,0]^T}(R_k) = 10$ (which is greater than the threshold for rejecting the null hypothesis), elementary source detection rejects the null hypothesis and the MSD procedure proceeds to prominent source identification. The first iteration of prominent source identifi-

43

cation assumes $b_P = [0, 0, 0, 0]$ (no active sources) and attempts to maximize the likelihood over the corresponding toggle events (which, as discussed previously, are equivalent to the elementary events). The likelihood is maximized when $b_P = [0, 0, 1, 0]^T$, corresponding to $l_{[0,0,1,0]^T}(R_k) = 10$. After declaring $b_P = [0, 0, 1, 0]^T$ as the maximum, the likelihood ratios for the toggle events corresponding to $b_P = [0, 0, 1, 0]^T$ are

$$
\begin{aligned}
l_{[0,0,0,0]^T}(R_k) &= 1 \\
l_{[0,1,1,0]^T}(R_k) &= 11 \\
l_{[1,0,1,0]^T}(R_k) &= 12 \\
l_{[1,0,1,1]^T}(R_k) &= 10
\end{aligned}
\tag{3.21}
$$

We note that the $l_{[0,0,0,0]^T}(R_k)$ will always equal 1 since $l_{[0,0,0,0]^T}(R_k)$ is exactly the likelihood ratio of the null hypothesis with itself. As before, we attempt to maximize the likelihood and declare $b_P = [1, 0, 1, 0]^T$ to correspond to the new maximum, since $l_{[1,0,1,0]^T}(R_k) = 12$ which is greater than $l_{[0,0,1,0]^T}(R_k) = 10$ (corresponding the previous value of b_P). The result of this iteration indicates that it is more likely that both source 1 and 3 are active. Just as before, after declaring $b_P = [1, 0, 1, 0]^T$ as the maximum, the likelihood ratios for the toggle events corresponding to $b_P = [1, 0, 1, 0]^T$ are

$$
\begin{aligned}
l_{[0,0,1,0]^T}(R_k) &= 10 \\
l_{[1,0,0,0]^T}(R_k) &= 8 \\
l_{[1,1,1,0]^T}(R_k) &= 14 \\
l_{[1,0,1,1]^T}(R_k) &= 3
\end{aligned}
\tag{3.22}
$$

Applying the same procedure as previously, $b_P = [1, 1, 1, 0]^T$ is declared most likely, corresponding to all the sources being active. Again, the likelihood ratios for the toggle events

44

are

$$l_{[0,1,1,0]^T}(R_k) = 11$$

$$l_{[1,0,1,0]^T}(R_k) = 12$$

$$l_{[1,1,0,0]^T}(R_k) = 15$$

$$l_{[1,1,1,1]^T}(R_k) = 10$$

(3.23)

This iteration indicates that the likelihood is maximized when $b_P = [1,1,0,0]^T$, corresponding to only sources 1 and 2 as active. Assuming $b_P = [1,1,0,0]$, the corresponding toggle event likelihood ratios are:

$$l_{[0,1,0,0]^T}(R_k) = 5$$

$$l_{[1,0,0,0]^T}(R_k) = 8$$

$$l_{[1,1,1,0]^T}(R_k) = 14$$

$$l_{[1,1,0,1]^T}(R_k) = 4$$

(3.24)

At this iteration we conclude that $b_P = [1,1,0,0]^T$ maximizes the likelihood ratio since $l_{[1,1,0,0]^T}(R_k) = 15$ is greater than all the corresponding toggle event likelihood ratios. The procedure declares $b_P = [1,1,0,0]^T$ and proceeds to secondary source identification. Secondary source detection is only concerned with the positive toggle events (events considering one additional active source) corresponding to b_P, namely:

$$l_{[1,1,1,0]^T}(R_k) = 14$$

$$l_{[1,1,0,1]^T}(R_k) = 4$$

(3.25)

The secondary source threshold is defined to be $\gamma = 0.30$. For each positive toggle event, $l_b(R_K)$, corresponding to b_P, if $\gamma l_{b_P}(R_K) < l_b(R_K)$, then the additional source assumed by b is considered a secondary source. For the small example, $\eta_\gamma l_{b_P}(R_K) = (0.30)(15) =$

4.5. Since $l_{[1,1,1,0]^T}(R_k) = 14$ and $l_{[1,1,0,1]^T}(R_k) = 4$, we conclude that only source 3 is a secondary source and declare $b_S = [0, 0, 1, 0]^T$. Thus, in this example, source 1 and source 2 are considered prominent sources, while source 3 is a secondary source. To verify the probability of type III error is bounded (since secondary source identification is performed assuming the prominent sources are active), source verification occurs.

Recalling that in this example, only sources 1 and 3 are active, after performing source verification we discover that sources 1 and 3 are active and source 2 is inactive. Since source 4 was not decided to be a prominent or a secondary source, it was not verified and is not known to be inactive or active. Since the prominent sources were $b_P = [1, 1, 0, 0]^T$, we observe that the prominent sources assumed source 2 to be active when it was verified to be inactive. Since not all the prominent sources are active, we must verify the constraint in (3.17) is true. We write $b_I = [1, 0, 1, 1]^T$ and $b_A = [1, 0, 1, 0]^T$ as described in the discussion of (3.17), where since source 4 is untested b_I (denoting a collection of the known inactive sources) assumes source 4 to be active while b_A (denoting a collection of known active sources) assumes source 4 is inactive. Applying (3.17), the type III error is bound if

$$\gamma l_{[1,0,1,0]^T}(R_k) \geq l_{[1,0,1,1]^T}(R_k).\tag{3.26}$$

Since the likelihood ratios for the events above are

$$\begin{aligned} l_{[1,0,1,0]^T}(R_k) &= 12 \\ l_{[1,0,1,1]^T}(R_k) &= 3 \end{aligned},\tag{3.27}$$

we declare the type III error is bounded since, $\eta_\gamma l_{[1,0,1,0]^T}(R_K) = (0.30)(12) = 3.6$ which is greater than $l_{[1,0,1,1]^T}(R_k) = 3$. In this example, even though some of the prominent sources are inactive, the probability of type III error is sufficiently bounded and the active sources are identified as source 1 and source 3.

The above example illustrates how prominent source identification, secondary source identification, and source verification are performed to bound the probability of type III error. Prominent source identification attempts to maximize the likelihood ratio through an iterative strategy, while secondary source identification selects sources to ensures the probability that an active source is not declared active (type III error) is bounded if all the prominent sources are active. If some of the prominent sources are inactive, through source verification the bounds on the probability of a type III error can be achieved by either inspection or re-performing prominent and secondary source identification after applying the knowledge of which sources are active and inactive.

This section introduced a procedure for performing MSD for constant sources (sources known to become active only at time $k = 0$). The following section extends the procedure introduced in this section to consider emergent sources, that is, sources that can become active at times other than $k = 0$.

3.2 IPSRT for Emergent Sources

The previous subsection only considered constant sources, corresponding to persistent sources that were known to switch from inactive to active at time zero and remain active indefinitely. This section updates the procedure presented in the previous section to include emergent sources. Emergent sources are persistent sources that may become active at any time. The flowchart in Fig. 3.2 outlines the updated IPSRT.

To determine not only which sources are active, but when they become active requires a hypothesis for each source at each time step. Recalling from Chapter 2, the number of event hypotheses for a problem containing J potential persistent sources and K time steps is
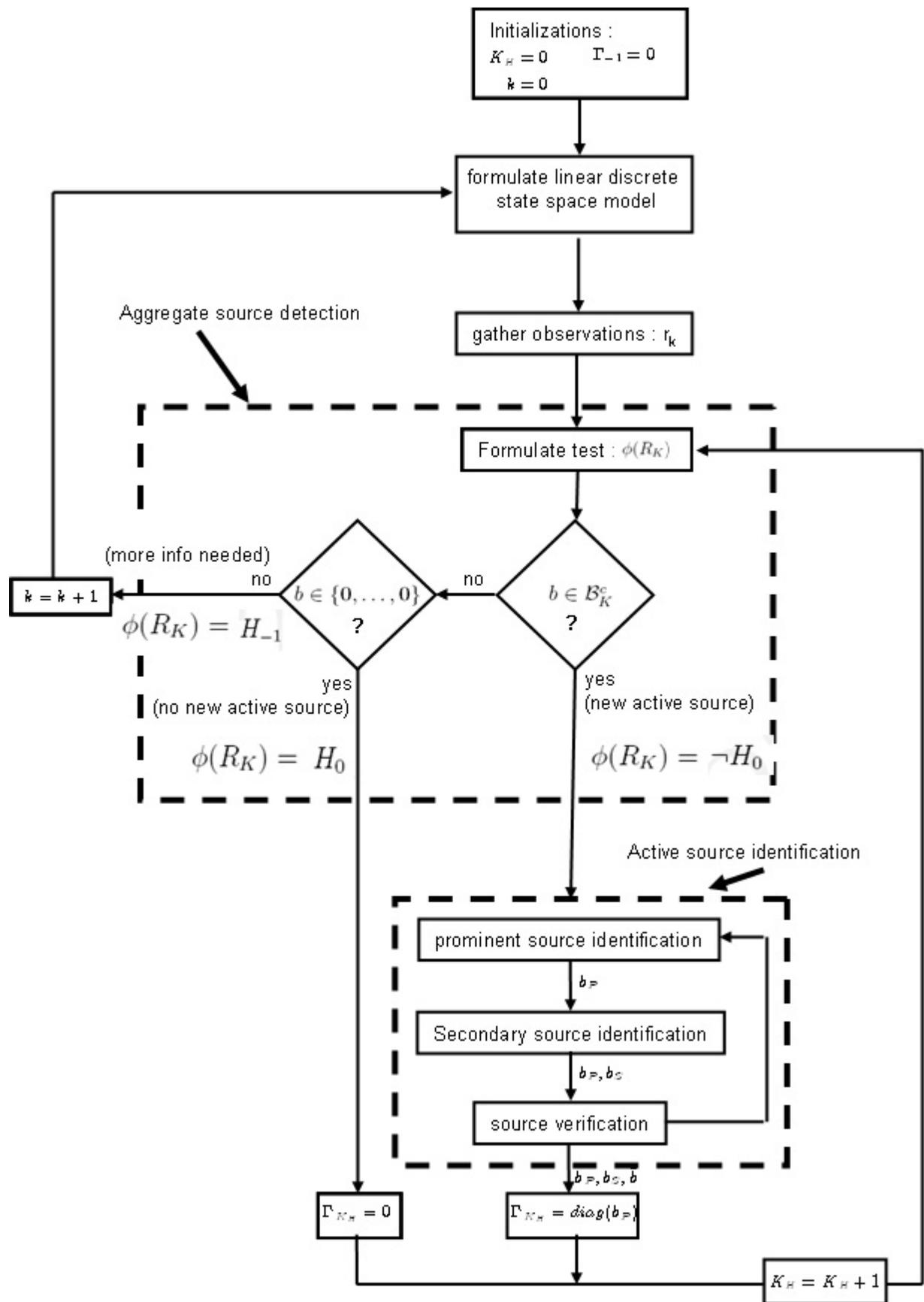
$$S_K = (K + 2)^J - 1. \tag{3.28}$$

Figure 3.2: IPSPRT for emergent sources.

By applying the heuristic developed in the previous subsection, we reduce the sequential hypothesis testing problem to a test consisting of only $KJ-1$. These hypotheses correspond to a single null hypothesis (H_0 assuming $b_k = 0$) and KJ elementary event hypotheses corresponding to each elementary source becoming active at each time $k = 0$ to $k = K$. This results in a sequential multiple hypothesis testing problem with $KJ+1$ hypotheses. As discussed in Chapter 2, no test exists that bounds the probability of error and minimizes the expected time-to-decision under every hypothesis in a sequential hypothesis testing problem.

As a heuristic approach, the sequential multiple hypothesis testing problem with $KJ+1$ hypotheses is approximated by K different sequential multiple hypothesis testing problems each with $J + 1$ hypotheses that can be solved using the IPSPRT as introduced in the previous section. This formulation tests a null hypothesis (H_0) against the set of elementary hypotheses that assume sources transition from inactive to active at time $k = K_H$ (where initially $K_H = 0$ as shown in Fig. 3.2). For notational convenience, we denote $H_{e_j,k}$ to be the elementary event hypothesis that assumes the jth source transitions from inactive to active at time k. Thus, the IPSPRT results in an aggregate source detector that assumes each source con possibly become active at time $k = K_H$. Just as in the aggregate source detection, the test results in one of three decisions : gather another observation, accept the null hypothesis ($\phi(R_K) = H_0$), or reject the null hypothesis ($\phi(R_K) = \neg H_0$).

If it is decided to obtain another observation, the process proceeds as described in Section 3.1. The time k is incremented, a system model is formulated, DSS is performed, an observation is gathered, and the hypothesis testing problem continues.

As shown in Fig. 3.2, if the test accepts the null hypothesis, then we assume there were no sources active at time $k = K_H$ and before any new observations are gathered, K_H is incremented, and a new binary hypothesis testing problem is formulated for the new value of K_H. This process of incrementing K_H and retesting occurs until either the

null hypothesis is rejected or more information is needed to decide between the null and event hypotheses. This process results in a multiple hypothesis testing problem between the null hypothesis and the J elementary event hypotheses assuming the smallest K_H until a decision to accept or reject the null hypothesis is reached.

If the SPRT rejects the null hypothesis for K_H at the current time, k, we denote the time of decision as $K_D(K_H) = k$. At time $K_D(K_H)$, active source identification and verification are performed. Active source verification (performed by physical inspection) guarantees that if a source is determined to be active, then it is in fact active. This is important when dynamic b_k are considered since claiming an inactive source as an active source results in an incorrect model under all hypotheses. Since active source verification occurs at the current time ($k = K_D(K_H) > K_H$) and the emergent sources are known to be persistent, any sources verified to be inactive at time $k = K_D(K_H)$ are also known to be inactive for all time $k \leq K_D(K_H)$. Similarly, and sources verified to be active at time $k = K_D(K_H)$ are known to be active for all time $k \geq K_D(K_H)$. Applying this knowledge, K_H is incremented and the problem is reformulated, where the verified inactive(active) sources are now known to be inactive(active) and are no longer tested as part of the hypothesis testing problem. It is noted that if after source verification, $b_{K_D} \equiv \mathbf{1}$ (all sources are active), then there is no reason to perform MSD and the IPSPRT terminates; however, this is unlikely to occur in most applications.

To illustrate the sequential behavior of the IPSPRT for emergent sources, we consider a fictitious MSD problem with 4 sources where source 3 becomes active at time $k = 14$, source 4 becomes active at time $k = 16$, source 2 becomes active at time $k = 28$, and source 1 remains inactive. The values used in this example were chosen (not calculated) to illustrate different aspects of the IPSPRT. We graphically display the sequential testing results in Fig. 3.3. In Fig. 3.3, the horizontal axis represents the time when the MSD procedure made a decision, $K_D(K_H)$ for the test indicated by the vertical axis, K_H. For
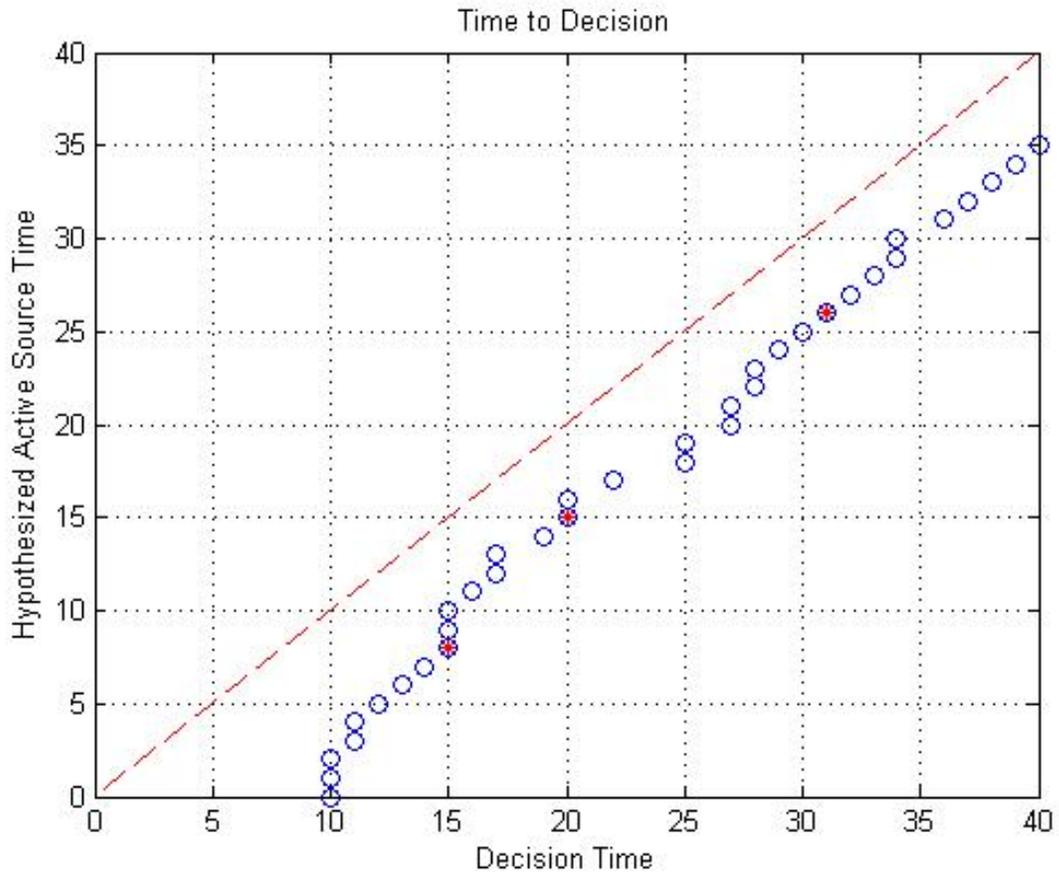
50

Figure 3.3: Hypothesized time vs. decision time vs. decision.

example, the MSD procedure made a decision for $K_H = 0$ at time $k = 10$. In Fig. 3.3, hollow circles represent a decision to accept the null hypothesis, while circles with a star indicate a decision to reject the null hypothesis. In this example, the first decision (for $K_H = 0$) was made at time $k = 10$. At the same time, a decision to accept the null hypothesis was also made for the tests starting at $K_H = 1$ and $K_H = 2$; however, a decision for the test starting at $K_H = 3$ could not be made at time $k = 10$. Thus, the time was incremented and another observation gathered at time $k = 11$. After gathering the observation at time $k = 11$, a decision to accept the null hypothesis was made for the tests starting at $K_H = 3$ and $K_H = 4$. This iterative process of gather observations and accepting the null hypothesis until time $k = 15$ when the null hypothesis is rejected for the test starting at time $K_H = 8$.

At $k = 15$, the test starting at time $K_H = 8$ accepts that sources 2 and 3 are active. After performing source verification, it is discovered that only source 3 is active. These results are shown in Fig. 3.4, where the values are chosen (not calculated) in this example to demonstrate different aspects of the IPSPRT. In Fig. 3.4, the horizontal axis represents time and the vertical axis denotes the source (either 1, 2, 3, or 4). A circle represents a source verified to be inactive, a circle with a star represents when a source (verified to be active) is assumed to have switched from inactive to active, and a square represents when a source actually became active. At time $k = 15$, two sources were assumed to be active. Source 3 was in fact active, while source 2 was inactive. Since sources are persistent, we know that since source 2 was inactive at time $k = 15$, that source 2 is also inactive for $8 \leq k \leq 15$. In fact, it is known that source 2 is inactive for all time $k \leq 15$, but all the decisions prior to $K_H = 8$ had already accepted the null hypothesis, thus we do not plot source 2 as inactive for $k < 8$. Similarly, since source 3 is verified to be active and known to be persistent, we conclude that source 3 is active for all time $k \geq 15$. In Fig. 3.4, source 3 is assumed to become active at time $k = 8$, when it actual became active at time $k = 14$
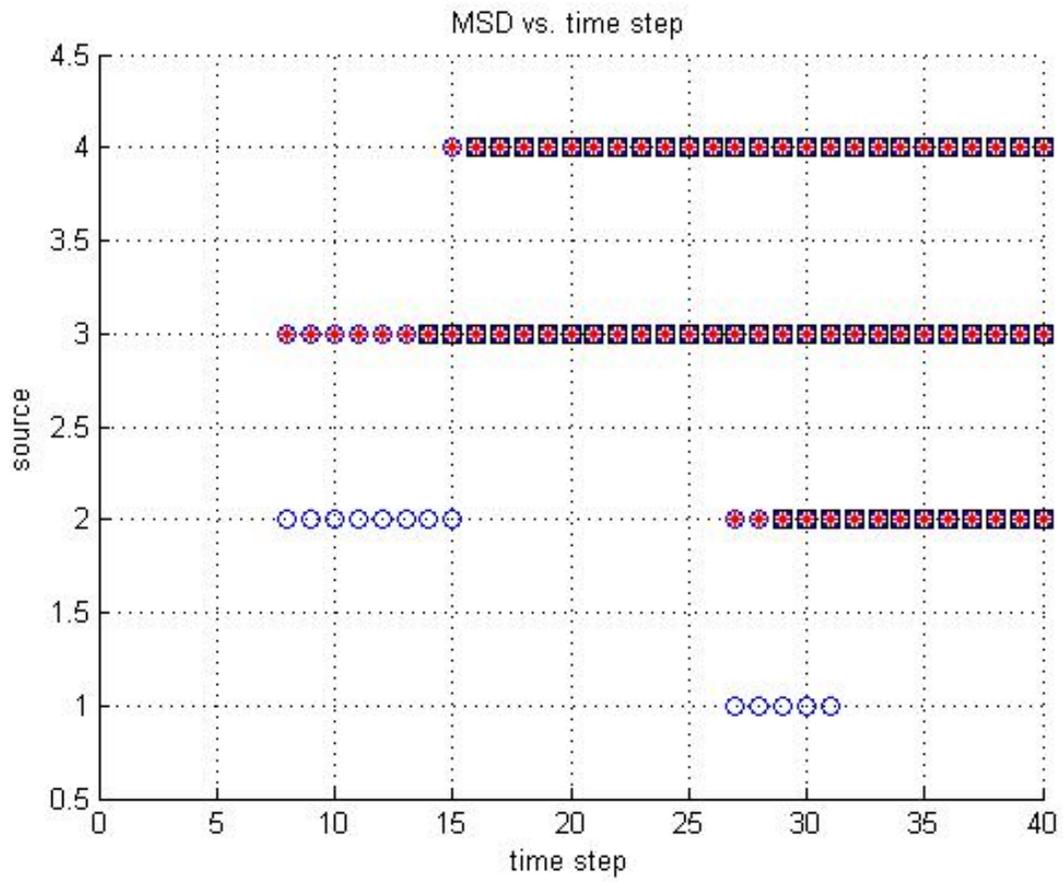
Figure 3.4: Potential active sources vs. hypothesis start time.

(as denoted by the square for the third source). Accepting a source as active prior to when it actually becomes active is an unavoidable consequence of evaluating a sequential MSD problem one step at a time.

3.3 Simulation Results

The IPSPRT for MSD introduced in this chapter was simulated using the diffusive system described in Appendix C. This example assumes a grid of N by N sensors that are co-located with the potential sources. The diffusion example is simulated using the following parameters

- $Q_k = I$ (select all sensors)

- $N = 10$ (100 potential sources and sensors)

The results in this section are divided into two subsections: detection results and localization results, respectively.

3.3.1 Detection results

To evaluate the IPSPRT proposed in this chapter, the IPSPRT is compared to two other feasible tests for large-scale MSD. The first test, which we refer to as the *naive test*, assumes a sensor is located at each potential source. A source is declared active in the naive test if the corresponding sensor observation exceeds an *a priori* specified threshold. If after an *a priori* specified period of time (known as the time-to-decision), no sensor observation exceeds the threshold, then the null hypothesis is accepted. The naive test is the simplest of all the tests used for comparison, and assumes no knowledge of the underlying dynamics.

The second test used for comparison is motivated by the work of Jones and Willsky [73]. In their test, source magnitudes are assumed to be parameters (not random variables) and a generalized likelihood ratio test (GLRT) is used to detect active sources in linear dynamic systems. To perform the GLRT, a maximum likelihood estimate of the source magnitude under each event hypothesis is generated for a pre-specified number of

observations. The pre-specified number of observations is selected to bound the probability of a miss and defines the time-to-decision for the GLRT. For each event hypothesis, the estimated source magnitude is assumed to be the actual source magnitude, and the corresponding likelihood ratio test is performed with respect to the null hypothesis using a Neyman-Pearson test threshold chosen for a specified probability of false alarm. The test accepts the null hypothesis if all the likelihood ratio tests result in a decision to accept the null hypothesis. Otherwise, the event hypothesis corresponding to the maximum likelihood ratio is accepted. Since their test requires estimating the source magnitudes under every hypothesis, directly applying the test of Jones and Willsky is impractical for large-scale MSD. By observing that inactive sources are essentially active sources of zero magnitude, we can estimate the source magnitudes by assuming all the sources are active. The test then reduces to a test between a null hypothesis (assuming no active sources) and a single event hypothesis (assuming all active sources). We refer to this test as the *estimation test.*

All three of the tests (the naive test, the estimation test, and the IPSPRT) have advantages and drawbacks. The naive test is the easiest to implement, but comes at the cost of ignoring the underlying dynamics. The estimation test is more complex than the naive test and considers the underlying dynamics, but requires a fixed number of observations, which defines the time-to-decision. Lastly, the IPSPRT is the most complex test, but decisions are made such that the probability of false alarm and probability of miss are bounded, and the time-to-decision is not pre-specified.

For comparison, we evaluate the performance of the three tests in terms of time-to-decision, probability of false alarm, and probability of miss. For both the naive and estimation tests, $10,000$ Monte Carlo tests were performed using the diffusive system assuming 4000 thresholds ranging from -2 to 2 and time-to-decision ranging from 1 to 35 time steps. A false alarm occurs when, for a specific threshold and time-to-decision, the null hypothesis is rejected and no source was active for the entire monitoring period (as determined
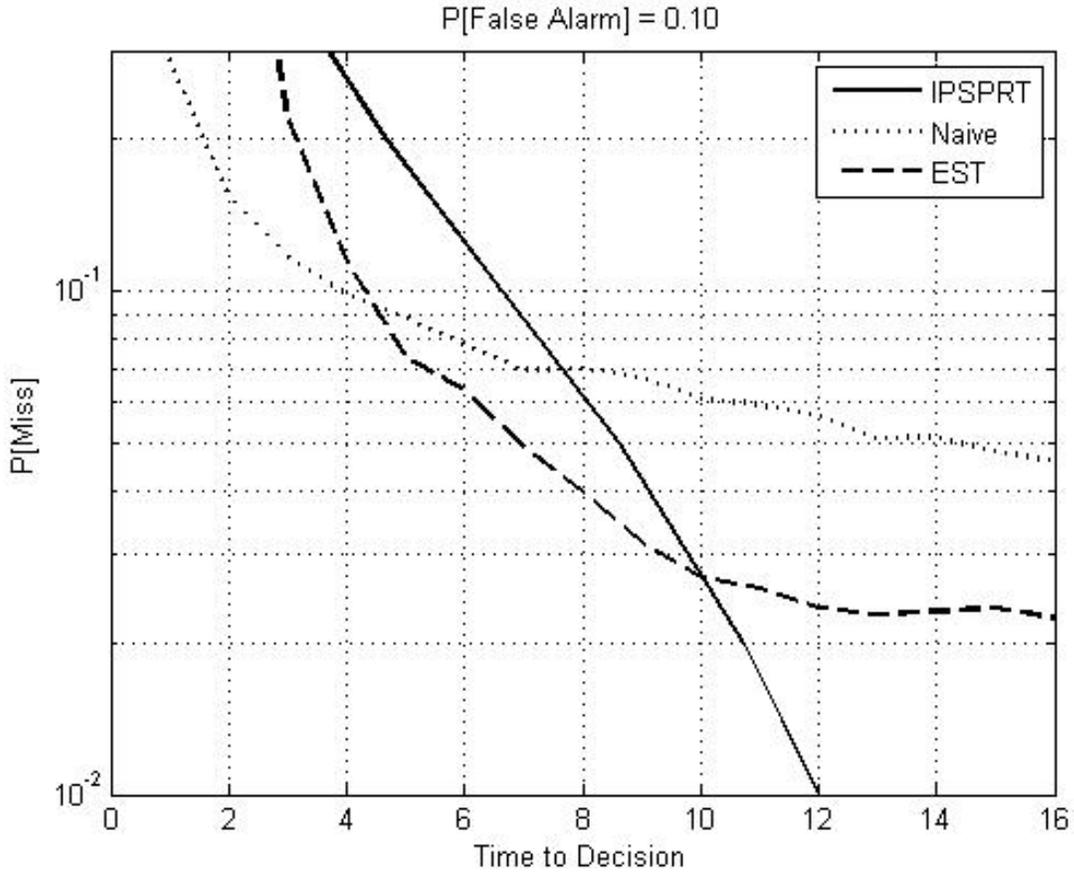
Figure 3.5: Probability of miss vs. time-to-decision for probability of false alarm = 0.10

by the time step when the test was initiated plus the time-to-decision). Similarly, a miss occurs when the null hypothesis is accepted and a source was active for the entire monitoring period. For the IPSPRT, the probability of false alarm and miss are parameters and the average time-to-decision calculated over 100 different runs of the advection system for all combinations of $\alpha, \beta \in \{0.01, 0.02, 0.05, 0.10, 0.20, 0.50\}$ (where α is the maximum probability of false alarm, and β is the maximum probability of miss). Figure 3.5 shows the expected probability of miss vs. the time-to-decision for each test when the maximum probability of false alarm is assumed to be 0.10 ($\alpha = 0.10$). In Fig. 3.5 the dotted line, dashed line, and solid line represent the naive test, the estimation test, and the IPSPRT, respectively. A specific test is preferred over the other tests if for a given probability of

56

miss, it results in the smallest time-to-decision. Under this criterion, Fig. 3.5 suggests that if a probability of miss above 0.094 is acceptable, then the naive test is the best option because it requires fewer time steps to make a decision (as denoted by the dotted line being below both the dashed and solid line). If the desired probability of miss is between 0.094 and 0.018, Fig. 3.5 illustrates that the estimation test is preferred. Only when a probability of miss below 0.018 is required should the IPSPRT be employed. The reason the IPSPRT does not always perform better than the other strategies is due to the conservative decision thresholds obtained through Wald's approximation [65]. The results in Fig. 3.5 illustrate that the complexity of the IPSPRT test pays dividends as the desired probability of miss decreases.

Figure 3.6 provides the results for when the maximum probability of false alarm is decreased to 0.01. As in Fig. 3.5, the dotted line, dashed line, and solid line represent the naive test, the estimation test, and the IPSPRT respectively in Fig. 3.6. Similar to the results in Fig. 3.5, the results in Fig. 3.6 indicate the naive test is preferred for tests where the acceptable probability of miss is above 0.14, the estimation test is preferred for tests accepting a probability of miss between 0.14 and 0.038, and the IPSPRT test is preferred when the maximum probability of miss below 0.038. Comparing the results in Fig. 3.6 to the results in Fig. 3.5, we find that when the maximum probability of false alarm is decreased, the range of desired probability of miss where the IPSPRT is preferred increases.

The results in both Figs. 3.5 and 3.6 suggest that as the acceptable probability of miss decreases, the number of time steps required by the naive and estimation tests grows exponentially when compared to the IPSPRT. To illustrate this point, Fig. 3.7 shows a graph of the probability of false alarm versus the probability of miss for each test, known as the receiver-operator characteristic (ROC). In Fig. 3.7, the underlying shaded contour plot represents integer values of the average time-to-decision for the proposed MSD strategy,
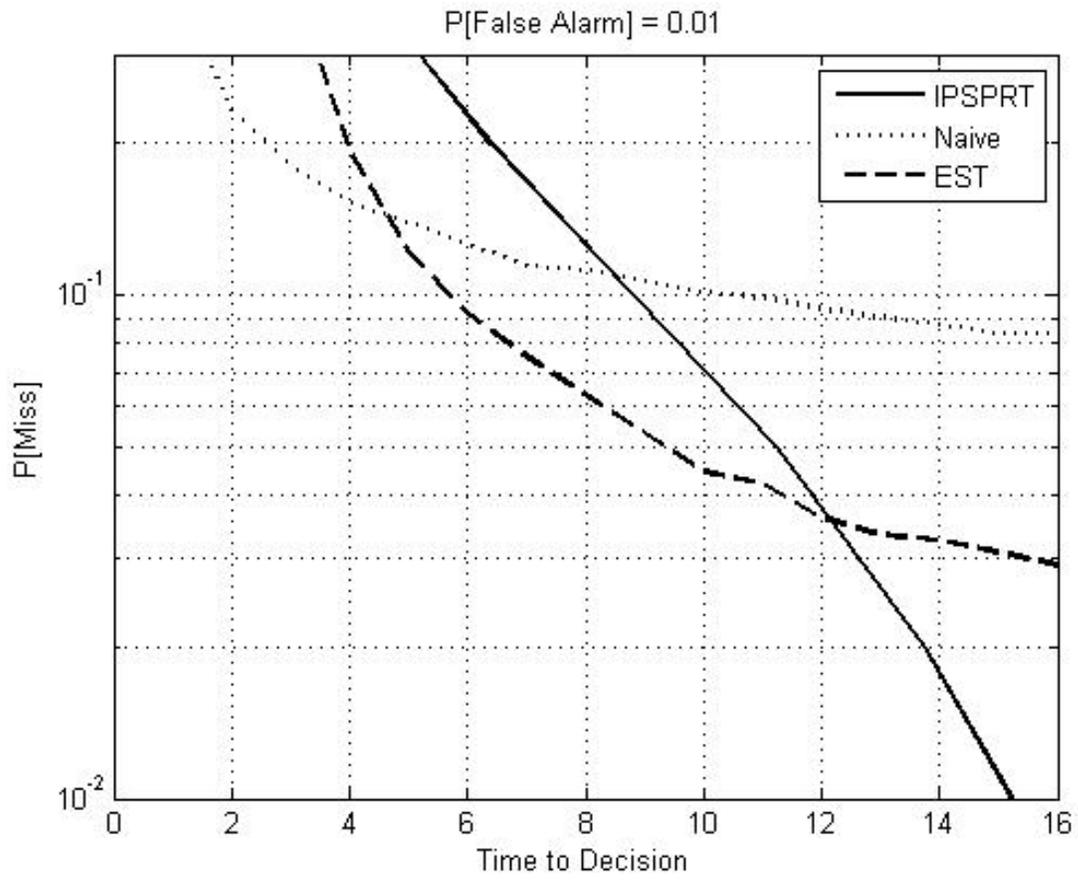
Figure 3.6: Probability of miss vs. time-to-decision for probability of false alarm = 0.01
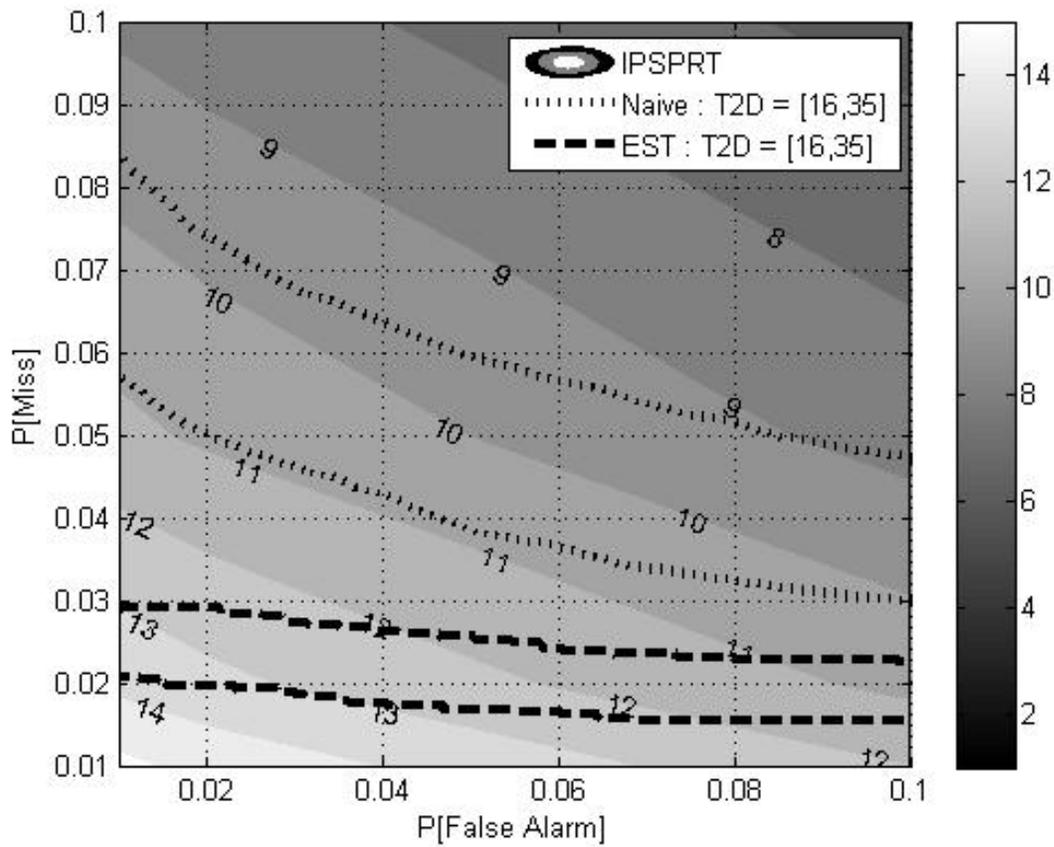
Figure 3.7: Receiver operator characteristic vs. time-to-decision.

ranging from 8 to 15 for probability of false alarm and probability of miss ranging from 0.01 to 0.10. The color bar on the right and the integer values on the plot indicate the IPSRT time-to-decision. The dotted and dashed lines in Fig. 3.7 represent the naive and estimation tests, respectively, where the upper dotted and dashed lines correspond to a time-to-decision of 16, while the lower ones represent a time-to-decision of 35. To relate the results in Fig. 3.7 to the results in Fig. 3.5, in Fig. 3.5, the upper dotted and dashed lines (corresponding to a time-to-decision of 16) represent all combinations of probability of false alarm and probability of miss where a time-to-decision of 16 can be achieved for achieved respectively for the naive and estimation tests. When the probability of false alarm equals 0.10, the probability of miss that can be achieved is 0.048 for the naive test and 0.022 for the estimation test. In Fig. 3.5, these values correspond to the probability of miss attainable by each strategy when the time-to-decision is 16. A similar relation can be made between the results in Figs. 3.7 and 3.6 when the probability of false alarm equals 0.01 and the time-to-decision equals 16. In Fig. 3.7, the results for a time-to-decision of 35 are plotted to illustrate the marginal savings in time-to-decision for the proposed test when compared to the naive and estimation tests. To achieve a maximum probability of false alarm of 0.02 while also achieving a maximum probability of miss of 0.02, the estimation test requires aa time-to-decision of 35 while the IPSRT averages a time-to-decision of 13.5 (the naive test requires a time-to-decision much greater than 35). As expected, for all three tests, as the probability of false alarm and probability of miss decrease, the time-to-decision increases. This suggests that more observations are needed to make a more accurate decision.

When a source becomes active at a time later than the hypothesized time ($K > K_H$), the proposed strategy often detects the source. To illustrate this phenomenon, we write
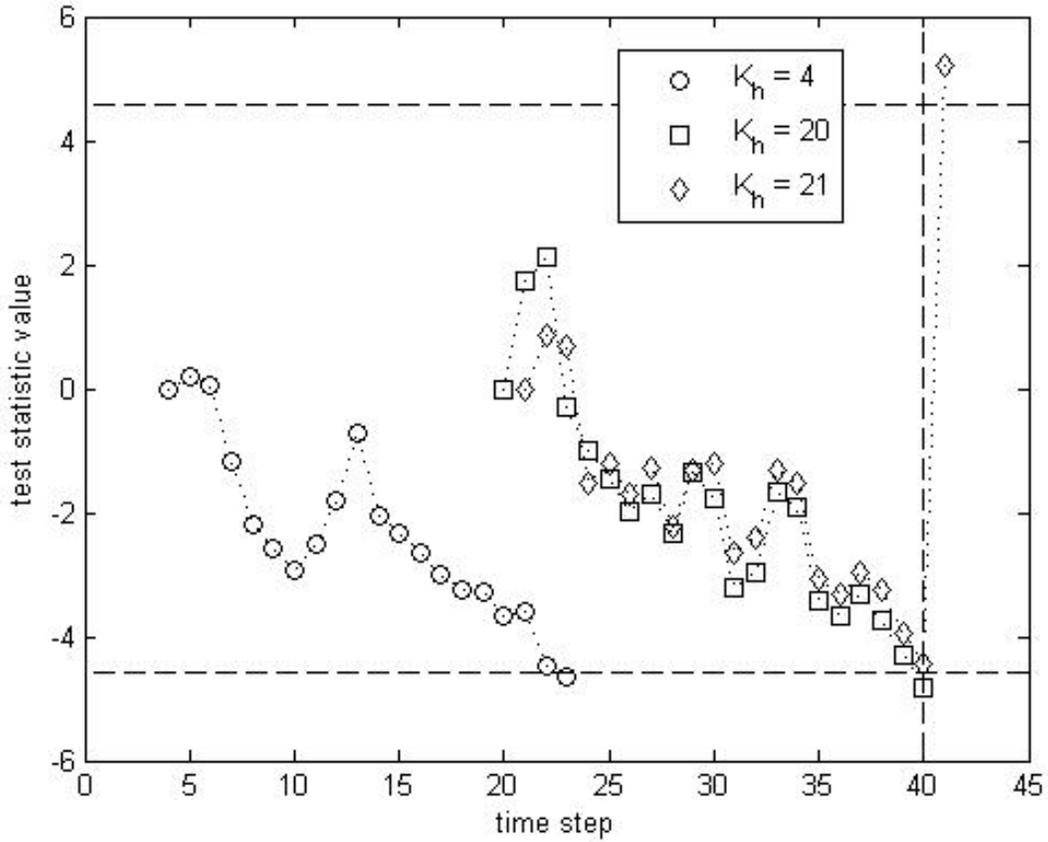
Figure 3.8: Test statistic values for $K_H = 4, 20, 21$.

the test statistic for the test starting at time $k = K_H$ at time $k = K$ as

$$T_{K_H}(R_K) = \sup_{b \in B^e_{K,K_H}} \frac{f_b(R_K)}{f_0(R_K)} \tag{3.29}$$

where B^e_{K,K_H} is the set of time propagations of persistent elementary sources at time K assuming sources can only become active at time K_H. The test statistic ensures that to accept the null hypothesis requires that every elementary hypothesis be less likely than the null while to reject the null hypothesis only requires that one elementary hypothesis be more likely. Figure 3.8 illustrates the test statistic value for a tests beginning at times, $K_H = 4$, 20, and 21, denoted by the circles, squares, and diamonds, respectively.

61

In Fig. 3.8, the upper and lower horizontal dashed lines represent the thresholds for rejecting and accepting the null hypothesis, and the vertical dashed line denotes that a source became active at time $k = 40$. Fig. 3.8 illustrates that more observations are needed to accept the null hypothesis than to reject the null hypothesis. This can be concluded from Fig. 3.8 by observing that once a source became active at time $k = 40$, the null hypothesis is rejected at the next time step, $k = 41$, while to accept the null hypothesis (when no source is active) takes over 20 time steps. Thus, a change in the threshold to accept the null hypothesis, η_β will have more of an effect on the average time-to-decision as compared to the same change in the threshold to reject the null hypothesis, η_α. Recalling (3.5), the thresholds to accept the null hypothesis, η_β, and reject the null hypothesis η_α are

$$\eta_\alpha = \frac{1 - \beta}{\alpha} \quad \text{and} \quad \eta_\beta = \frac{\beta}{1 - \alpha}. \tag{3.30}$$

We observe that for small α and β, η_β is more affected by a change in β than by a change in α. Therefore, the average time-to-decision of the proposed test is more affected by the change in the maximum probability of miss, β, as opposed to the maximum probability of false alarm, α. These results are illustrated in Fig. 3.7 by the shaded contour plot. For a change in the maximum probability of miss from 0.10 to 0.01, the time-to-decision (for $\alpha = 0.01$) increases by 5 time steps. For a change in the maximum probability of false alarm from 0.10 to 0.01, the time-to-decision (for $\beta = 0.01$) increases by only 3 time steps.

3.3.2 Source identification results

The source identification method proposed in this chapter is evaluated using the same diffusion example as above. Upon detecting an active source in the detection simulations, a localization simulation was performed assuming 100 different values of the maximum probability of a type III error ranging from 0 to 1. The percentage of the combined prominent and secondary sources that were actually active is plotted against each value of the maximum probability of type III error in Figs. 3.9 and 3.10. As a comparison, an
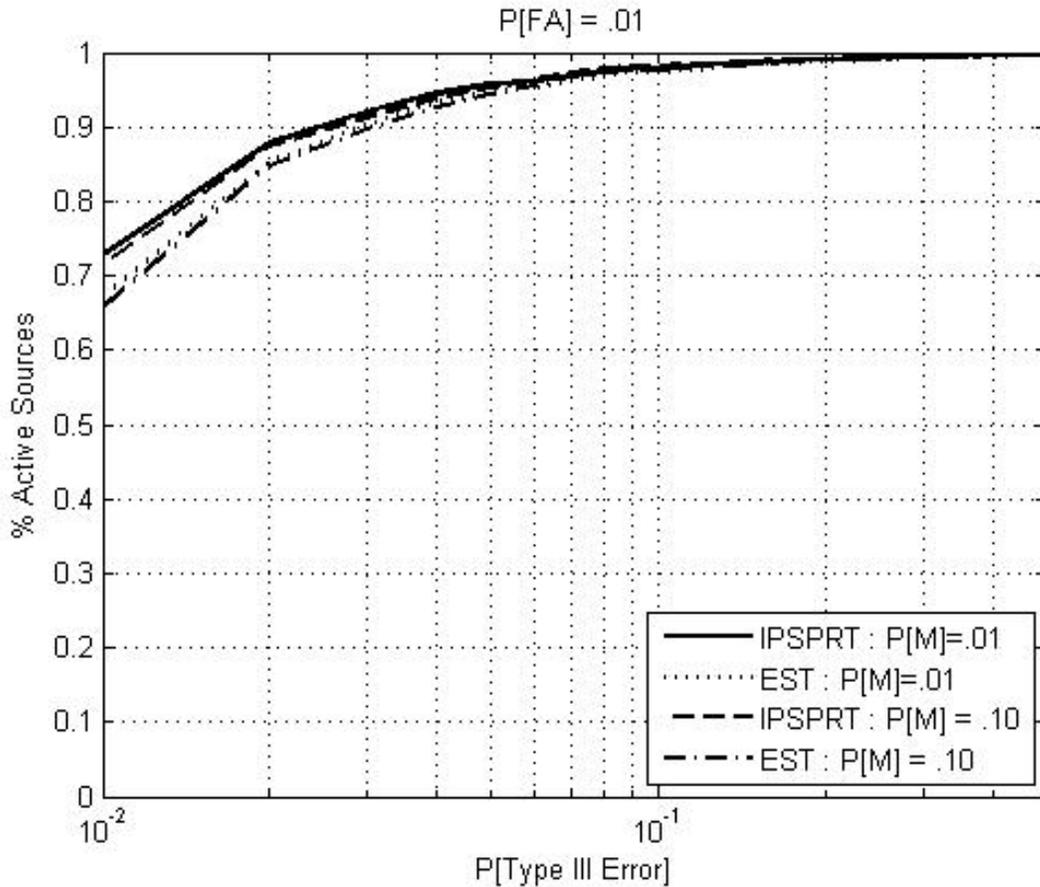
Figure 3.9: Percentage of identified sources that are active vs. probability of type III error vs. probability of miss.

estimation based strategy is also included that estimates the source magnitudes assuming all active sources (just as in the estimation test described in the previous subsection), and a threshold applied to the source estimate.

In Fig 3.9, the percentage of identified sources that are actually active is plotted against the maximum probability of type III miss for tests assuming a constant value of the maximum probability of false alarm and varying values of the maximum probability of a miss, where we recall that the maximum probability of false alarm and maximum probability of miss are the parameters used by the proposed strategy to accept or reject the null hypothesis. In Fig 3.9, the percentage of identified sources that are active increases as the probability of a type III error is increased and as the maximum probability of a miss de-
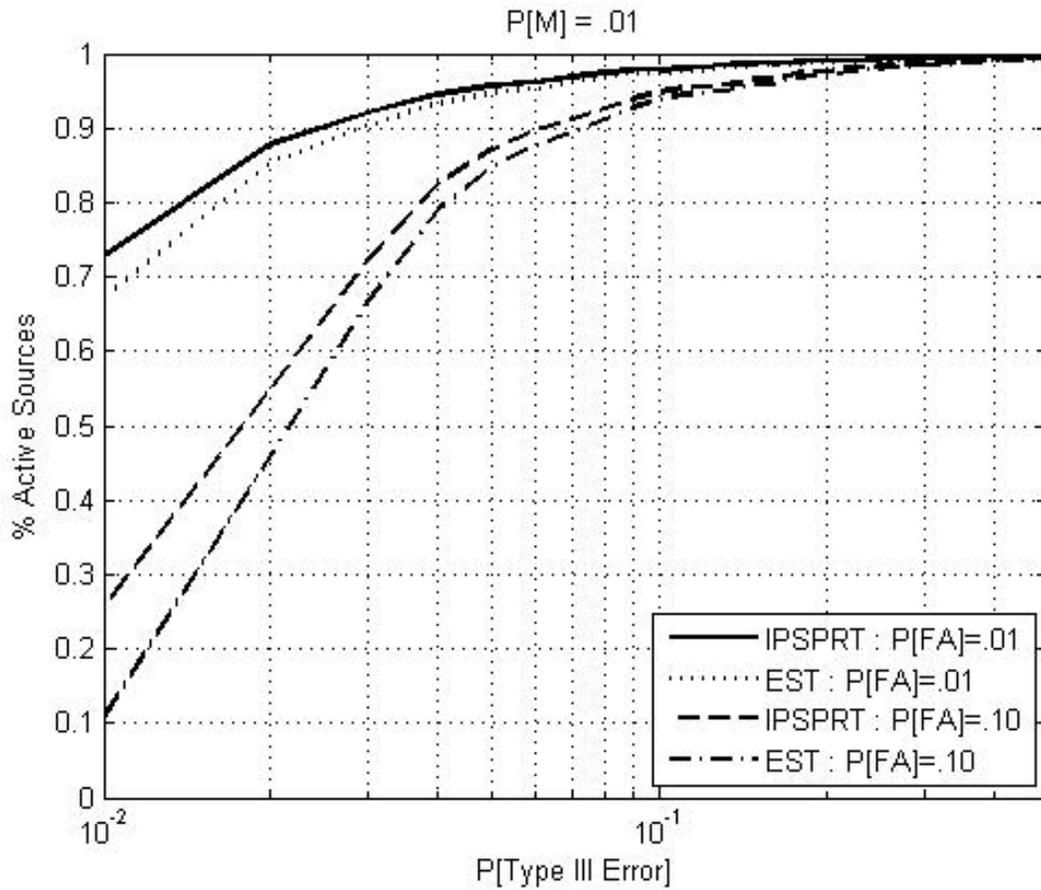
63

Figure 3.10: Percentage of identified sources that are active vs. probability of type III error vs. probability of false alarm.

creases. These results suggest that percentage of identified sources that are active is not significantly affected by changes in the probability of a miss. This is a direct result of the threshold for rejecting the hypothesis, η_α in (3.30), not being significantly affected by small changes in β (the maximum probability of miss).

Similar to Fig. 3.9, the results in Fig. 3.10 plot the percentage of identified sources that are actually active against the maximum probability of type III error for tests assuming a constant value of the maximum probability of miss and varying values of the maximum probability of a false alarm. The results in Fig. 3.10 suggest that the percentage of identified sources that are active can be increased by reducing the probability of a miss. Comparing the results in Figs. 3.10 and 3.9, we observe that a change in the maximum probability of false alarm has a more significant effect on the percentage of identified sources that are active as compared to the same change in the maximum probability of a false alarm. Similar to the discussion in the previous subsection, this effect is a direct result of the threshold for rejecting the null hypothesis (and performing source identification) and is dominated by the maximum probability of false alarm, α.

Chapter 4

Wireless Sensor Network Considerations

In this chapter, we consider the resource constraints imposed by performing large-scale multiple source detection (MSD) using wireless sensor networks (WSNs). There are scenarios in MSD problems where choosing fewer sensors can be beneficial. For instance, if there are many sensors within communication range of one another, having all the sensors communicate at once will result in increased observation loss due to a lack of sufficient communication bandwidth. In the MSD applications of interest, we assume that the sensor sampling rate is significantly greater than the time required to gather the observations through a WSN. Thus, we claim that observation loss due to communication bandwidth constraints (known as packet collisions) can be mitigated. A more important consideration is extending the *network lifetime* [24, 48]. Extending the WSN lifetime can be accomplished in many ways such as reducing the sampling rate, reducing the transmission power, and performing dynamic sensor selection (DSS). DSS is the process of selecting fewer sensors to return their observations at each time step such that the network lifetime is extended. In MSD applications, DSS is used to extend the network lifetime at a cost of the time-to-decision. In the following, we propose a dynamic sensor selection (DSS) strategy for the iterative partial sequential probability ratio test (IPSPRT) introduced in Chapter 3 and is referred to as the GLRDSS. The GLRDSS is a generalized likelihood ratio (GLR) based dynamic sensor selection (DSS) strategy that extends the lifetime of a wireless sen-

sor network (WSN) by choosing fewer sensors to report at each time step which can be implemented in large-scale long-term monitoring applications.

The following section introduces preliminary notation useful for discussing DSS. Section 4.2 motivates and introduces the GLRDSS. Section 4.3 presents a conservative approximation for the GLRDSS that can be solved using convex optimization techniques. The final section provides an evaluation of the GLRDSS through simulation results.

4.1 DSS preliminary notation

This section reviews and introduces notation and properties that prove useful in the following DSS discussion. Assuming the IPSPRT, introduced in Chapter 3, is used for MSD, we recall the condition for accepting the null hypothesis as

$$\hat{\Lambda}_G(R_k) \leq \eta_\beta, \tag{4.1}$$

where

$$\hat{\Lambda}_G(R_k) = \frac{\sup_{b \in B_k^e} f_b(R_k)}{f_{\mathbf{0}}(R_k)}. \tag{4.2}$$

is the generalized likelihood ratio over the single-active-source hypotheses and the threshold $\eta_\beta \in [0, 1]$ is determined according to the maximum probability of false alarm and maximum probability of miss. Equivalently, the IPSPRT test for accepting the null hypothesis is written as

$$\sup_{b \in B_k^e} l_b(R_k) \geq \ln \eta_\beta, \tag{4.3}$$

where

$$l_b(R_k) = l_b(R_{k-1}) - \ln f_{\mathbf{0}}(R_k) + \ln f_b(R_k), \tag{4.4}$$

is the *log-likelihood ratio* (LLR) between the null hypothesis and the event hypothesis, H_b. Since the observations are modeled as a linear dynamic system in (2.6) from Chapter 2,

the LLR is recursively calculated as

$$l_b(R_k) = l_b(R_{k-1}) + \frac{1}{2}(r_k - m_{\mathbf{0},k|k-1})^T Q_k^T Q_k S_{\mathbf{0},k|k-1}^{-1} Q_k^T Q_k (r_k - m_{\mathbf{0},k|k-1})$$

$$- \frac{1}{2}(r_k - m_{b,k|k-1}) Q_k^T Q_k S_{b,k|k-1}^{-1} Q_k^T Q_k (r_k - m_{b,k|k-1}) \qquad , \qquad (4.5)$$

$$+ \frac{1}{2} ln(det(Q_k S_{b,k|k-1}^{-1} Q_k^T Q_k S_{\mathbf{0},k|k-1} Q_k^T))$$

where Q_k is the sensor selection matrix at time k as defined in (2.4) from Chapter 2, and $m_{b,k|k-1}$ and $S_{b,k|k-1}$ are the residual mean and covariance of the observations at time k generated by a Kalman filter assuming the linear dynamic model under hypothesis H_b. By applying the results in [43], the expected value of the LLR under the null hypothesis, given the observations R_{k-1} (i.e. given all the observations until time $k-1$), is written as:

$$E_{H_0}[l_b(R_k)] = l_b(R_{k-1}) + \frac{1}{2} Tr(Q_k^T Q_k)$$

$$- \frac{1}{2} \left(m_{\mathbf{0},k|k-1} - m_{b,k|k-1} \right)^T Q_k^T Q_k S_{b,k|k-1}^{-1} Q_k^T Q_k \left(m_{\mathbf{0},k|k-1} - m_{b,k|k-1} \right)$$

$$- \frac{1}{2} Tr \left(Q_k S_{b,k|k-1}^{-1} Q_k^T Q_k S_{\mathbf{0},k|k-1} Q_k^T \right) +$$

$$(4.6)$$

where the trace of $Q_k^T Q_k$ is equivalent to the number of sensors selected at time k. In the following and consistent with the system model formulation in Chapter 2, $\mathcal{Q}$ represents the set of all possible sensor selection matrices, $Q_k = \emptyset$ denotes that no sensors are selected, and $Q_k = I$ represents that all the sensors are selected. Additionally, we write q_k to be the *sensor selection vector*, where $q_k = diag\left(Q_k^T Q_k\right)$. From (4.6), we observe that the expected value under the null hypothesis of the LLR at time k, $E_{H_0}[l_b(R_k)]$, when no sensors are selected ($Q_k = \emptyset$) is exactly the previous value of the LLR, $l_b(R_{k-1})$. We refer to this property as *null-constant*.

Definition 4. A function, $f(R_k)$, is *null-constant* if

$$E_{H_0}[f(R_k)|Q_k = \emptyset] = f(R_{k-1}) \tag{4.7}$$

We note that since received observations, R_k, are a function of the sensor selection matrix, Q_k, then any function containing R_k is also a function of Q_k. For each elementary event, $b \in B_k^e$, we introduce

$$\Delta(k, k', b) = \frac{\ln \eta_\beta - l_b(R_{k-1})}{K_H + \kappa_D - k' + 1} \tag{4.8}$$

to denote the average change needed to meet the desired time-to-decision at time k assuming the current time is k', where κ_D is the desired time-to-decision and K_H is the time when the test began. The numerator of (4.8) represents the difference between the threshold for accepting the null hypothesis and the LLR corresponding to event b. This difference represents the change in the LLR needed to accept the null hypothesis. The denominator of (4.8) represents the number of time steps left until the desired time-to-decision is reached. Thus, $\Delta(k, k', b)$ represents the average change in the LLR required at each time step to accept the null hypothesis within the desired time-to-decision. In the following, it is shown how $\Delta(k, k', b)$ is employed to formulate a set of constraints for selecting sensors using the GLRDSS. Using the notation introduced above, the following subsection introduces the GLRDSS.

4.2 GLRDSS for MSD

This section formulates the GLRDSS as a 1-step DSS strategy for the IPSPRT, introduced in Chapter 3, that reduces the number of sensors selected subject to achieving an approximate time-to-decision. Solving a DSS problem that minimizes the number of sensors selected subject to bounding the desired time-to-decision requires identifying a sensor

selection schedule until the desired decision time is reached. In real-world applications, solving a sensor scheduling problem requires knowledge of future dynamics and future observations, both of which can only be estimated. Furthermore, finding the optimal solution for any sensor selection problem requires solving a 0-1 integer programming problem [25], known to be NP-hard [27], and results in an exhaustive search of all possible strategies [31, 38]. Since this work concerns large-scale long-term monitoring applications where the dynamics may vary significantly over time, received observations are dependant on the varying dynamics as well as the WSN channel reliability, and the number of potential sensors and desired time-to-decision are large, solving a long-term sensor scheduling problem is computationally infeasible. When events are unlikely (as in many long-term monitoring applications) the null hypothesis is most likely to best describe the monitored process at any given time. Thus, we propose a 1-step approximate DSS strategy that minimizes the number of sensors selected at the current time subject to a constraint reflecting the approximate time-to-decision under the assumption that the null hypothesis is true.

We introduce the GLRDSS as a 1-step approximate DSS strategy, problem consisting of three parts: *feasibility check*, *minimization*, and *verification*. The feasibility check ensures the GLRDSS is feasible; the minimization part selects the minimum number of sensors such that the change in the LLR is more than the average change needed to meet the approximate time-to-decision; and the verification part selects the sensors from the minimization if the change in the LLR is less than the average change needed to meet the specified time to decision minus one time step. The GLRDSS is outlined as follows:

- Feasibility check :

$$
\begin{aligned}
\text{if} \ &: \ E_{H_0}\left[l_b(R_k) - l_b(R_{k-1})|Q_k = I\right] \leq \Delta(k, k, b) \quad \forall b \in B_k^e \\
\text{then} \ &: \ \text{go to minimization} \\
\text{else} \ &: \ Q_k = I
\end{aligned}
$$

(4.9)

- Minimization : (identify $\hat{Q} \in \mathcal{Q}$, then go to verification)

$$\hat{Q} = \arg\min_{Q_k \in \mathcal{Q}} trace(Q_k^T Q_k)$$

$$\text{s.t.} \quad E_{H_0}\left[l_b(R_k) - l_b(R_{k-1})\right] \le \Delta(k,k,b) \quad \forall b \in B_k^e \tag{4.10}$$

- Verification :

$$Q_k = \begin{cases} \emptyset & \text{if} \quad E_{H_0}\left[l_b(R_k) - l_b(R_{k-1})|Q_k = \hat{Q}\right] < \Delta(k,k+1,b) \quad \forall b \in B_k^e \\ \hat{Q} & \text{otherwise} \end{cases} \tag{4.11}$$

The GLRDSS selects a minimal number of sensors such that a constraint on an approximate measure of the expected time to decision is achieved. While the formulation above does not consider the energy remaining in the batteries of each sensor, we note that any of the energy cost functions in [74] could be incorporated in the minimization objective in 4.10. The intuition behind the GLRDSS is that at each time step, sensors are selected such that the expected value of the LLR approaches the threshold for accepting the null hypothesis.

To better explain the GLRDSS, an illustrative example is sequentially developed and referenced throughout the following subsections. For this example, we assume there is a single constant source which is not known to be active or inactive. Since only a single potentially active source is considered, a single event hypothesis exists and results in a sequential binary hypothesis testing problem. Thus, in the discussion pertaining to the example, we refer to $l(R_k)$ to be the LLR and $\Delta(k,k')$ to be $\Delta(k,k',b)$ since only one potential event exists. In this example, we assume the desired time-to-decision is four time steps ($\kappa_D = 4$) and the threshold for accepting the null hypothesis is $\ln \eta_\beta = -2$. To demonstrate the how each part of the GLRDSS is performed, the example assumes there are two sensors, resulting in four potential sensor selection matrices at each time k, corresponding to selecting no sensors ($Q_k = \emptyset$), selecting only the first sensor ($Q_k = [1,0]$),

selecting only the second sensor ($Q_k = [01]$), and selecting both sensors ($Q_k = I$). For this scenario, the null hypothesis is true, and the example begins at time $k = 1$, where the value of the LLR at time zero (after gathering observations at time zero) is $l(R_0) = 2$. Although the expected value of the LLR and the actual value of the LLR are normally calculated according to (4.6) and (4.4), respectively, for this example, these values are chosen (not calculated) to illustrate the different aspects of the GLRDSS; however, the value of $\Delta(k, k)$ is calculated as part of this example. The following subsections employ the example described above to discuss each step of the GLRDSS and provides a discussion of the GLRDSS for rejecting the null hypothesis.

4.2.1 Feasibility check

The feasibility check of the GLRDSS in (4.9) exists to ensure the minimization problem is feasible. To check the feasibility, a sufficient test is

$$E_{H_0}\left[l_b(R_k) - l_b(R_{k-1})|Q_k = I\right] \leq \Delta(k, k, b) \quad \forall b \in B_k^e, \tag{4.12}$$

where the minimization problem is feasible if the expected decrease in the LLR is larger than the required change when selecting all the sensors (i.e. $Q_k = I$) as indicated by (4.9). Although this constraint is not a necessary condition, intuition suggests that when the null hypothesis is true, the LLR decreases as more observations are gathered. When (4.12) is false, then the GLRDSS assumes that the minimization problem is not feasible and selects all the sensors ($Q_k = I$). When (4.12) is true, then the GLRDSS proceeds to minimization. We note that when the desired decision time is exceed ($k > K_H + \kappa_D$), no sensor selection matrix, Q_k, satisfies the sufficient condition, and the feasibility check ensures that all the sensors are selected in hopes of achieving the quickest time-to-decision.

To illustrate the feasibility check, we recall the small example described previously, where our discussion begins at time $k = 1$. Before observations can be acquired at time
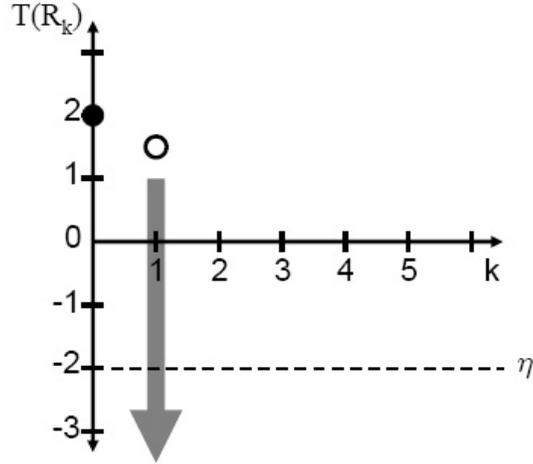
Figure 4.1: GLRDSS : feasibility check at $k = 1$.

$k = 1$, the GLRDSS must specify which sensors should report measurements. As discussed above, the first step in the GLRDSS is to verify the minimization problem is feasible. Figure 4.1 illustrates the feasibility check of the GLRDSS at time $k = 1$. In Fig. 4.1 and consistent with the previous description of the example, the solid circle at time $k = 0$ denotes the value of the LLR $(l(R_0) = 2)$. Similarly, the hollow circle at $k = 1$ denotes the expected value of the LLR when all the sensors are selected, $E_{H_0}[l(R_1)|Q_1 = I] = 1.5$, which is chosen (not calculated) in this example for illustrative purposes. The shaded region denotes where the minimization problem is feasible, namely when $E_{H_0}[l(R_1)|Q_1 = I] \leq l(R_0) + \Delta(1,1) = 2 + \frac{-2-2}{0+4-1+1} = 1$. From Fig. 4.1, we observe that the minimization problem is not feasible since $E_{H_0}[l(R_1)|Q_1 = I]$ is not less than 1. This result implies the maximum expected change in the LLR will not be large enough to meet the desired time-to-decision. Thus, the feasibility check of the GLRDSS selects all the sensors at time $k = 1$, corresponding to the assumed maximum expected decrease in the LLR. The GLRDSS then proceeds to time $k = 2$.

After gathering the observation at time $k = 1$, the LLR at time $k = 1$, $l(R_1)$, is determined to be 1 (again, we recall the values of $l(R_k)$ and $E_{H_0}[l(R_k)]$ are not calculated, but rather chosen to illustrate different aspects of the GLRDSS). Figure 4.1 illustrates the
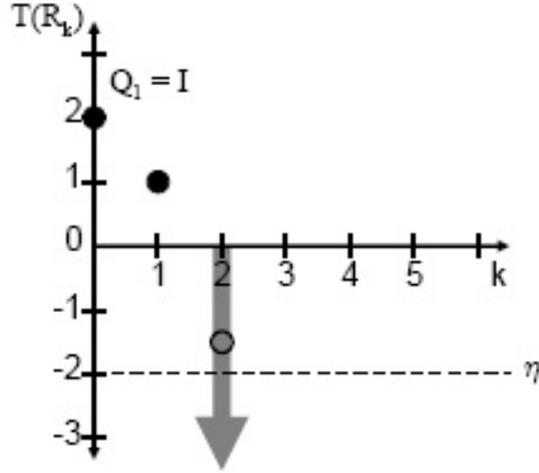
Figure 4.2: GLRDSS : feasibility check at $k = 2$.

feasibility check at time $k = 2$. In Fig. 4.2, the solid circle at time $k = 1$ denotes the value of the LLR, $l(R_1) = 1$, which results from selecting all the sensors at time $k = 1$ $(Q_1 = I)$. Consistent with Fig. 4.1, the hollow circle at $k = 2$ is the expected value of the LLR when all the sensors are selected, $E_{H_0}[l(R_2)|Q_2 = I] = -1$, and the shaded region denotes where the minimization problem is feasible, namely when $E_{H_0}[l(R_2)|Q_2 = I] \leq l(R_1) + \Delta(2,2) = 1 + \frac{-2-1}{0+4-2+1} = 0$. In Fig. 4.2, $E_{H_0}[l(R_2)|Q_2 = I]$ is less than the threshold of 0, which indicates the minimization problem is feasible, thus the GLRDSS proceeds to minimization, as described in the following subsection.

4.2.2 Minimization

The minimization part of GLRDSS in (4.10) minimizes the number of sensors selected while ensuring the expected change in the LLR, $E_{H_0}[l(R_k)] - l(R_{k-1})$, is greater than the average change needed to achieve the desired time-to-decision, $\Delta(k,k)$. In essence, the minimization part throttles the rate at which the expected value of the LLR approaches the decision threshold under the assumption that the null hypothesis is correct. When either the difference between the LLR and the threshold is large or the decision time draws near, more change in the LLR is required.
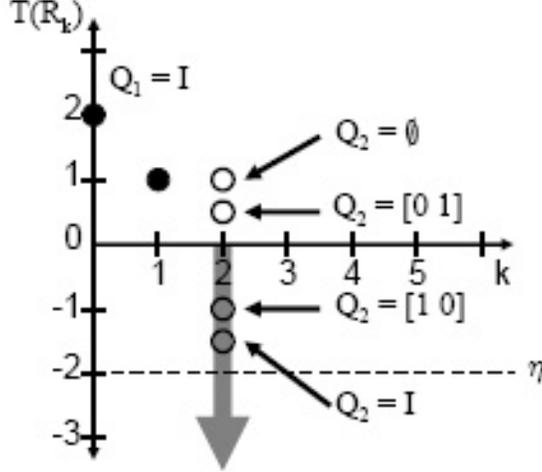
75

Figure 4.3: GLRDSS : Part 1, $k = 2$.

To illustrate the minimization part, we continue the example from the previous subsection assuming the current time is $k = 2$ and the feasibility check has been passed (As illustrated in Fig. 4.2). Figure 4.3 shows the minimization part of the GLRDSS at time $k = 2$, where the only difference between Fig. 4.2 and Fig. 4.3 is the inclusion of the expected values of the test function assuming all possible combinations of sensor selection matrices, $E_{H_0}[l(R_2)|Q_2 = \emptyset]$, $E_{H_0}[l(R_2)|Q_2 = [0,1]]$, $E_{H_0}[l(R_2)|Q_2 = [1,0]]$, and $E_{H_0}[l(R_2)|Q_2 = I]$. In Fig. 4.3, $E_{H_0}[l(R_2)|Q_2 = \emptyset]$ is exactly $l(R_1) = 1$ since the LLR is null-constant (see Definition 4). The other conditional expectations of the LLR are chosen (not calcualted) to help describe the minimization part of the GLRDSS. According to Fig. 4.3, the number of sensors is minimized subject to the constraint being satisfied (denoted by the shaded region) when $Q_2 = [1,0]$. This is because when $Q_2 = [0,1]$, the minimization constraint is not met and $Q_2 = I$ selects more sensors than $Q_2 = [1,0]$. Thus, the minimization part of the GLRDSS at time $k = 2$ selects only one sensor, namely $\hat{Q} = [1,0]$. The GLRDSS then proceeds to verification, described in the following subsection.

76

4.2.3   Verification

While the minimization part discussed in the previous subsection ensures a steady change in the test function, the minimization constraint always requires a negative change, $\Delta(k, k) \leq 0$, for all times less than the desired decision time $(K_H + \kappa_D)$. Thus, the minimization part is constrained to always select at least one sensor since the LLR is null-constant. For applications with a large desired time-to-decision, it is probable that selecting even one sensor at each time step will result in a time-to-decision significantly less than the desired time-to-decision. Thus, the verification part of the GLRDSS is employed to guard against a preemptive decision.

The verification part of the GLRDSS in (4.11) establishes a lower bound on the expected value of the test function to guard against a decision occurring prematurely. This is accomplished by selecting either all or none of the sensors specified by the minimization part. No sensors are selected if the expected change in the test function, $E_{H_0}[l(R_k)] - l(R_{k-1})$, is less than the required change under the assumption that the desired decision time was one step closer, $\Delta(k, k+1)$. While the minimization part selects sensors such that the LLR approaches the decision threshold, the verification part prevents early decisions by ensuring the change in the test function is not too large.

To illustrate the verification part of the GLRDSS, we revisit the example discussed in the previous two subsections and assume the minimization part at time $k = 2$ has been completed and decided $\hat{Q} = [1, 0]$ (as described in the previous subsection). Figure 4.4 illustrates the verification part of the GLRDSS at time $k = 2$, where the shaded region represents the region where verification accepts the sensors specified by the minimization part, namely $E_{H_0}[l(R_2)|Q_2 = [1, 0]] \geq l(R_1) + \Delta(2, 3) = 1 + \frac{-2-1}{0+4-3+1} = -0.5$. Here we observe that the constraint for accepting the minimized sensors involves $\Delta(2, 3)$, where $\Delta(k, k + 1)$ denotes the average change in the LLR needed to accept the null hypothesis one step earlier. In Fig. 4.4, since the sensors specified by the minimization result in
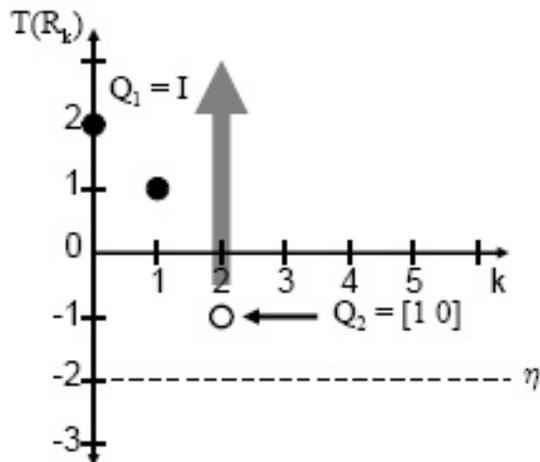
Figure 4.4: GLRDSS : Part 2, $k = 2$.

an expected value of the LLR, $E_{H_0}[l(R_2)|Q_2 = [1, 0]]$, outside the acceptance region of the verification part, no sensors are selected at time $k = 2$, namely $Q_2 = \emptyset$ as indicated in (4.11). Since no sensors are selected at time $k = 2$ ($Q_2 = \emptyset$) and the LLR is null-constant, the value of the LLR remains constant from $k = 1$ to $k = 2$, as shown in Fig. 4.5. Figure 4.5 represents the combination of the feasibility check, minimization, and verification parts of the GLRDSS at time $k = 3$, where the shaded region denotes where the minimization problem is feasible and the verification part accepts the sensors specified by the minimization. The upper bound on this region is exactly the bound calculated by the feasibility check at time $k = 3$, $l(R_2) + \Delta(3, 3) = 1 + \frac{-2-1}{0+4-3+1} = -1$, while the lower bound is specified by the verification part and is $l(R_2) + \Delta(3, 4) = 1 + \frac{-2-1}{0+4-4+1} = -2$. Thus the GLRDSS selects the combination of sensors over the region defined of Fig. 4.5 that minimizes the number of sensors selected. As shown in Fig. 4.5, the selection matrix that satisfies both the feasibility and verification constraints is $Q_3 = [1, 0]$. Therefore, at time $k = 3$ these sensors are selected.

We observe from Fig. 4.5 that the lower bound on the acceptable expected value of the test function is exactly the lower bound for accepting H_0 at time $k = 3$. Recalling the desired decision time for this test is $k = K_H + \kappa_D = 0 + 4 = 4$, we note that at the time
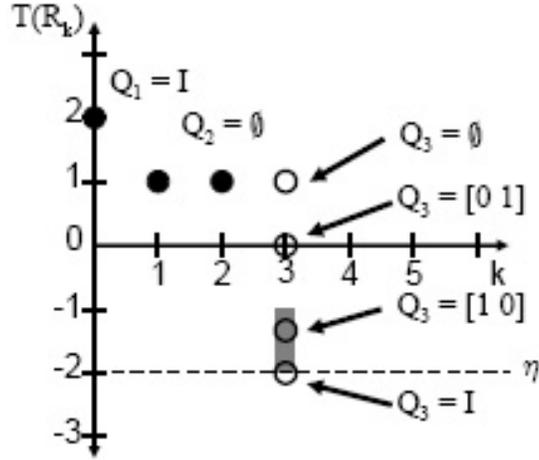
78

Figure 4.5: GLRDSS : $k = 3$.

step before the desired decision time, the GLRDSS allows the expected test function to get arbitrarily close to the value for accepting the null hypothesis. After gathering the data at time $k = 3$, Fig. 4.6 illustrates that the LLR at time $k = 3$ is $l(R_3) = -1$. Since the LLR value is not below the threshold to accept the null hypothesis, the test continues to time $k = 4$, which corresponds to the desired decision time.

The GLRDSS at time $k = 4$ is shown in Fig. 4.6. Consistent with Fig. 4.5, Fig. 4.6 represents the combination of feasibility check, minimization, and verification of the GLRDSS at time $k = 4$, and the shaded region in denotes when both the minimization problem is feasible and the sensors specified by the minimization part are accepted by verification. The upper bound is calculated according to the feasibility check at time $k = 4$, $l(R_3) + \Delta(4, 4) = -1 + \frac{-2+1}{0+4-4+1} = -2$, while the lower bound is the determined by the verification condition, $l(R_3) + \Delta(4, 5) = 1 + \frac{-2+1}{0+4-5+1} = -\infty$. These bounds indicate that at the desired decision time, $k = K_H + \kappa_D = 0 + 4 = 4$, the region for accepting the sensors specified by verification is exactly the region where H_0 is accepted. Thus, according to the example in Fig. 4.6, the sensors selected at time $k = 4$ are $Q_4 = [0, 1]$, since this combination is expected to result in a decision to accept the null hypothesis (assuming the
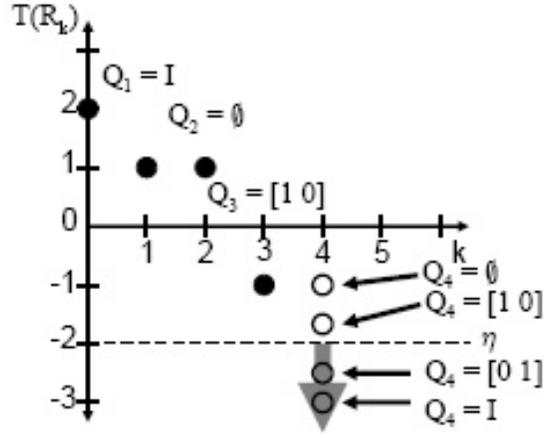
79

Figure 4.6: GLRDSS : $k = 4$.

null hypothesis is true) and selects the fewest sensors. If the time-to-decision is exceeded, $k > K_H + \kappa_D = 0 + 4 = 4$, we recall from the description of the feasibility check that all the sensors are then selected.

4.2.4   Rejecting the null hypothesis

The GLRDSS is tailored to accepting the null hypothesis, which is assumed to be much more likely than any event hypothesis. When the null hypothesis is rejected (an event hypothesis is accepted), the IPSPRT performs active source identification to localize the active sources. From the results and discussion in Chapter 3, we recall that localization accuracy increases with the number of sensor observations. Thus, when the null hypothesis is rejected, we prefer the DSS strategy to select more sensors such that source localization is improved. The number of sensors selected in the minimization part of the GLRDSS increases as the difference between the threshold for accepting the null hypothesis and the individual likelihood ratios increase. This is due to the numerator of $\Delta(k, k')$ becoming more negative as the difference increases between the LLR and the threshold for accepting the null hypothesis. When an event is true, it is expected that the corresponding LLR increases. As the LLR increases, more sensors are selected and the localization accuracy

80

improves assuming the null hypothesis is rejected.

This section formulates the GLRDSS for the IPSPRT in Chapter 3. The GLRDSS minimizes the number of sensors selected at each time step subject to a constraint on the expected decrease in the LLR assuming the null hypothesis is true. While the proposed DSS strategy does not guarantee the desired time-to-decision, it does select sensors such that a decision is likely to occur near the desired decision time. This is accomplished by selecting sensors such that the decrease in the LLR is large enough to accept the null hypothesis within the desired time-to-decision, but not so large that a preemptive decision is likely to occur. The GLRDSS is a 1-step approach to sensor selection, that is, the GLRDSS only selects sensors at the current time step and does not establish a sensor selection schedule over multiple time steps. The following section discusses the implementation of the GLRDSS for large-scale MSD using a conservative approximation and a relaxation-abstraction approach.

4.3   Implementing the GLRDSS

The GLRDSS introduced in the previous section requires both evaluating the expected decrease in the likelihood ratio (a non-linear function of the sensor selection) and solving a 0-1 integer programming problem. This is computationally infeasible for systems with large numbers of sensors (as in large-scale monitoring applications). In the following, we present a conservative approximation for the GLRDSS, then employ the same relaxation-abstraction technique as [25, 70, 74] coupled with affine approximations of the minimization constraints to relax the 0-1 integer programming problem into a convex programming problem.

4.3.1   Conservative GLRDSS

This section motivates and introduces a conservative-quadratic approximation for the minimization constraint in the GLRDSS. We begin by recalling the expected value of the LLR

in (4.6) is employed by the GLRDSS as a minimization constraint. By applying the results of Lemma 5 in Appendix D, the change in the expected value of the LLR due to the observations gathered at time k can be written as:

$$
\begin{aligned}
E_{H_0}\left[l_b(R_k) - l_b(R_{k-1})\right] = &-\frac{1}{2}q_k^T\left(S_{b,k|k-1}^{-1} \circ \hat{m}_{b,k|k-1}\hat{m}_{b,k|k-1}^T\right)q_k \\
&-\frac{1}{2}q_k^T\left(S_{b,k|k-1}^{-1} \circ S_{0,k|k-1}\right)q_k + q_k^T q_k \; , \\
&+\frac{1}{2}ln(det(Q_k S_{b,k|k-1}^{-1} Q_k^T Q_k S_{0,k} Q_k^T))
\end{aligned}
\tag{4.13}
$$

where $\hat{m}_{b,k|k-1} = m_{0,k|k-1} - m_{b,k|k-1}$, $A \circ B$ denotes the Hadamard product (element-wise product) of matrices A and B, and q_k is the sensor selection vector, defined in Section 4.1 to be the diagonal elements of $Q_k^T Q_k$. From (4.13), the expected decrease in the likelihood ratio, $E_0\left[l_b(R_k) - l_b(R_{k-1})\right]$, is a complicated non-linear function of the sensor selection vector (and matrix). To simplify the calculations, we directly apply the results of Lemma 7 in Appendix D, and bound the expected decrease in the LLR under the null hypothesis according to

$$
E_{H_0}\left[l_b(R_K) - l_b(R_{K-1})\right] \leq -\frac{1}{2}q_K^T\left(S_{b,K|K-1}^{-1} \circ \hat{m}_{b,k|k-1}\hat{m}_{b,k|k-1}^T\right)q_K,
\tag{4.14}
$$

where the expected decrease is bounded by a quadratic function of the sensor selection vector, q_k.

Applying the conservative approximation in (4.14) to the GLRDSS in (4.9) - (4.11) results in more sensors being selected at each time step such that the minimization constraint can be met. Conveniently, the upper bound in (4.14) is a quadratic function of the sensor selection term q_K; however, this still requires solving a 0-1 integer programming problem. The following subsection discusses a relaxation-abstraction approach for solving the GLRDSS using the conservative constraint.

4.3.2   Relaxed Conservative GLRDSS

The conservative 1-step DSS strategy in the previous subsection is a quadratic function of the sensor selection term q_K; however the conservative approach still requires solving a 0-1 integer programming problem, known to be NP-hard [27], and results (in the worst case) in an exhaustive search of all possible strategies [31, 38]. For applications with many sensors, this approach is known to be infeasible. Applying the same relaxation approach as in [25, 70, 74], we replace the constraint on the sensor selection term from $q_K \in \{0, 1\}$ to $0 \leq q_K \leq 1$. Applying this relaxation, to the conservative constraint in (4.14) results in a concave function of q_K, not a convex function. Since (4.14) is applied to the GLRDSS, the minimization part is non-convex as well. We propose to approximate the resulting concave constraint in (4.10) using an affine approximation.

An affine approximation of the quadratic constraint in (4.10) is generated such that the approximation is conservative and equals to the quadratic constraint at the point where the quadratic constraint intersects vector $\mathbf{1}$ (corresponding to all the elements of the vector equalling 1). Directly applying the results of Lemma 8 in Appendix D, we write the approximate affine constraints as:

$$p_{b,k}^T q_k \leq \Delta(k, k, b) \quad \forall b \in B_k^e, \tag{4.15}$$

where

$$p_{b,k} = -\frac{1}{2} \sqrt{\frac{\Delta(k, k, b)}{\mathbf{1}^T \left(S_{b,k|k-1}^{-1} \circ \hat{m}_{b,k|k-1} \hat{m}_{b,k|k-1}^T \right) \mathbf{1}}} \left(S_{b,k|k-1}^{-1} \circ \hat{m}_{b,k|k-1} \hat{m}_{b,k|k-1}^T \right) \mathbf{1}. \tag{4.16}$$

Using these approximations, the GLRDSS minimization constraints become affine functions of the sensor selection vector. The resulting convex minimization problem becomes a linear programming problem, which can be solved using the CVX toolbox [15].

The result of the minimization part of the GLRDSS yields a sensor selection vector with values ranging between 0 and 1. To identify a binary sensor selection vector, we assign a value of one to any element of q_k that equals one and similarly, we assign a value of zero to any element of q_k that equals zero. The remaining elements are ranked from largest to smallest and iteratively included in the selected sensors (by setting the corresponding element of q_k to 1) until the original non-relaxed minimization constraint in (4.10) is satisfied. The conservative constraint is used by the minimization to effectively *rank* the sensors for selection, where selection is actually performed using the original criteria. The GLRDSS then proceeds to verification as described in Section 4.2.

4.4   Simulation Results

The GLRDSS is simulated using a scaled down version of the diffusion example in Appendix C. This system contains 49 sensors and has 49 potential sources corresponding to 49 elementary event hypotheses being compared to the null hypothesis. The MSD algorithm from Chapter 3 is used to perform detection on the elementary hypotheses using the following parameters:

- $\alpha = 0.01$ (maximum probability of false alarm)

- $\beta = 0.05$ (maximum probability of miss)

The 20 simulations were performed at 35 different desired time-to-decision values ranging from 1 to 69.

The results for number of sensors selected versus the desired time-to-decision are shown in Fig. 4.7, where each 'x' represents the average number of sensors selected for a single simulation assuming a specific desired time-to-decision. Each vertical band contains 20 simulations (and thus 20 'x' marks). As Figure 4.7 illustrates, when the time-to-decision is small, the number of sensors selected is large and vice versa. For this application, there is a significant difference in the average number of sensors selected. This difference changes rapidly for tests assuming a time-to-decision between 9 and 23 time steps. During this time
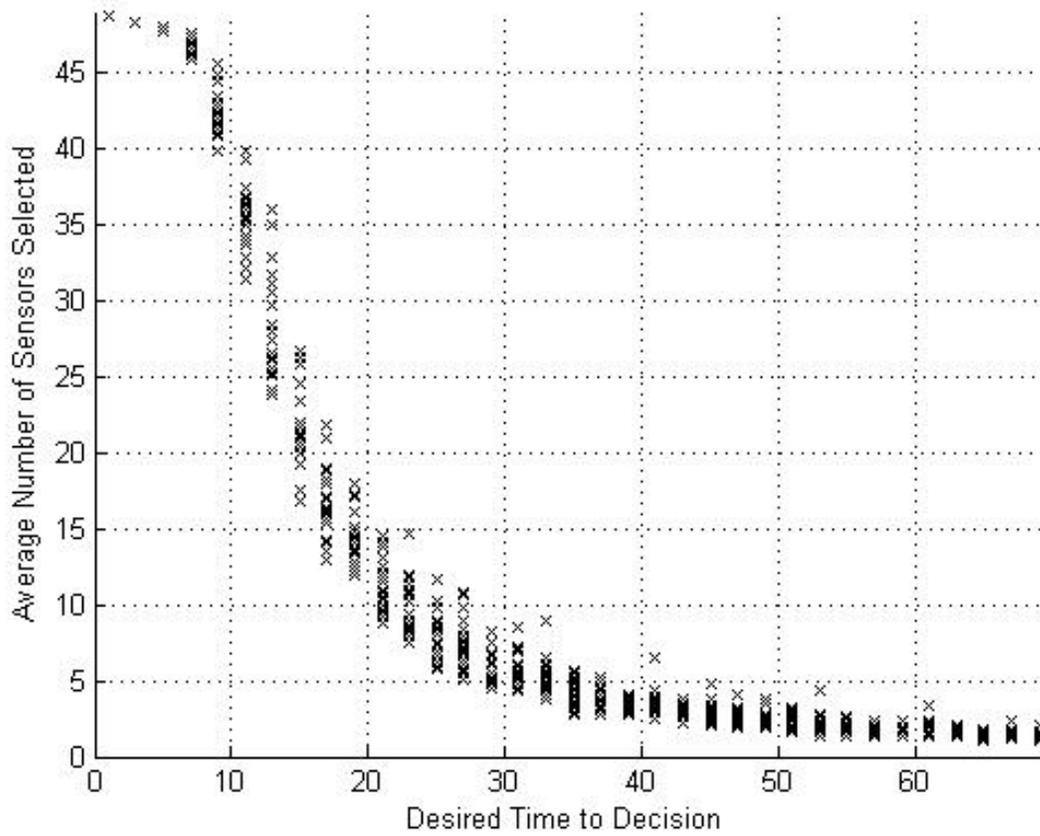
Figure 4.7: Average energy consumed (number of sensors selected) vs desired time-to-decision (in sampling periods).
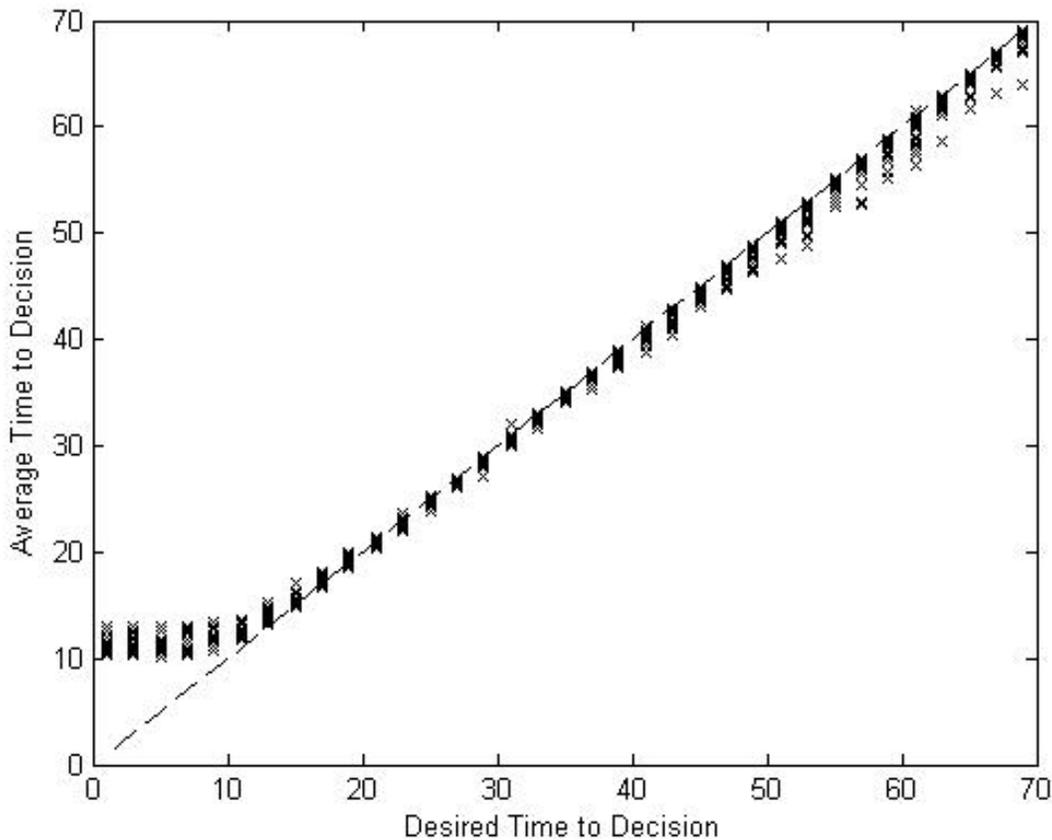
Figure 4.8: Actual time-to-decision vs. desired time-to-decision.

period, DSS is very useful because a small change in the desired time-to-decision can result in a significant difference in the number of sensors selected. Equivalently, over this range of desired time-to-decisions, the incremental energy savings is the greatest. As the desired time-to-decision increases, the difference in the number of sensors selected from test to test decreases significantly. This is due to the fact that only a few sensors are being selected and as the desired time-to-decision increases, it has a decreasing effect on the number of sensors selected.

The results for the average time-to-decision versus the desired time-to-decision for each test is shown in Fig. 4.7, where each 'x' represents the average time-to-decision for a single simulation assuming a specific desired time-to-decision. The dashed line represents

when the average time-to-decision matches the desired time-to-decision. For small time-to-decisions, Fig. 4.8 illustrates that the average time-to-decision is greater than the desired time-to-decision when the desired time-to-decisions less than 11. This occurs because the desired time-to-decision is not feasible for values less than 11. Once the desired time-to-decision becomes feasible, the GLRDSS performs well at maintaining the desired time-to-decision. The GLRDSS gives the user the ability to specify a desired time-to-decision to reduce the energy used to perform detection in large-scale MSD applications.

A comparison of the GLRDSS to other proposed strategies is difficult to formulate. Previous work on sensor selection is primarily concerned with estimation [74] [70] [25]. In these approaches, the number of sensors selected is minimized with respect to a bound on some measure of estimation accuracy, such as the mean-squared-error or the log-determinant of the error covariance matrix, while the GLRDSS minimizes the number of selected sensors subject to a constraint on the time-to-decision. Comparing the estimation-tailored sensor selection strategies to the GLRDSS requires relating the estimation accuracy to time-to-decision, which may vary significantly with different dynamic process. Thus, there is no way to identify the estimation accuracy needed to meet the desired time-to-decision prior to performing sensor selection. A sensor selection strategy that minimizes the number of sensors selected subject to a constraint on the expected time-to-decision is described in [58]. In their approach, the multiple hypothesis testing problem is defined to be Baysian and does not directly extend to the non-Baysian multiple hypothesis testing problems considered in this dissertation.

Chapter 5

Case Study, Implementation, and Evaluation

This chapter examines the performance of the multiple source detection (MSD) procedure using a wireless sensor network (WSN) through a case study implementation and evaluation of the CO_2 sequestration site monitoring problem described in Chapter 1. The following section describes a wireless sensor network test bed for evaluating the detection and localization strategies using an actual network. Section 5.2 introduces a linear lumped-parameter model for an advection-diffusion process. Section 5.3 evaluates the performance and robustness of the IPSPRT and GLRDSS, introduced in Chapters 3 and 4, using a wireless sensor network for the CO_2 sequestration site monitoring application and discusses the results.

5.1 Implementation

The test bed consists of 22 firefly sensor nodes [33] as shown in Fig. 5.1. Each firefly node in Fig. 5.1 runs the Nano-RK operating system [10], contains a light intensity sensor, and is connected to a unique programming board that supplies power to the firefly nodes and allows for quick reprogramming of the entire network and background monitoring through a wired network. A flow chart describing the functionality of the test bed is shown in Fig. 5.2. In addition to the 22 firefly nodes and programming boards, the complete test bed incorporates 2 computers (named *Ramathorn* and *Coolstore* in Fig. 5.2), and a
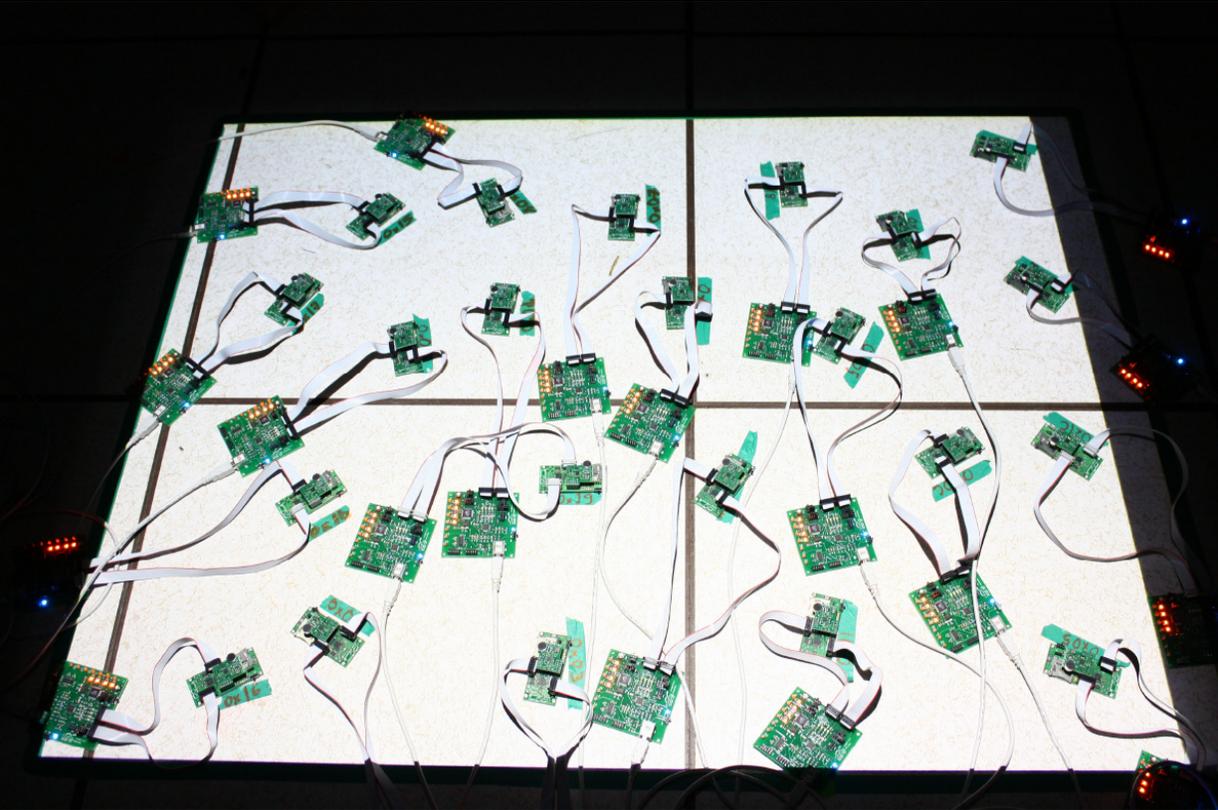
Figure 5.1: 22 Wireless sensor test bed.

Figure 5.2: Test bed architecture

light projector. Coolstore is a Linux machine that operates as a network manager, and Ramathorn is a Windows XP machine that executes the IPSPRT and GLRDSS routines and simulates the environment, which is projected onto the 22 firefly nodes as different light intensities ranging from 0 to 255. The light sensor on each firefly node outputs a light intensity value ranging from 0 to 1023 corresponding to *bright* and *dark*, respectively. A second-order least squares approximation is used to convert the firefly light intensity value to the environment variable used for detection and localization, calibrated using 15 different light intensities.

At each sampling instant, Ramathorn first updates the environmental data and projects the corresponding light data onto the sensor network. Ramathorn then performs the GLRDSS and sends a request via socket communication to the network manager (coolstore) specifying which sensors should be sampled. The network manager gathers the requested light sensor observations using the SAMPL [49] data acquisition protocol. After

data acquisition is complete, the network manager relays the collected light sensor observations to Ramathorn. Ramathorn then executes IPSPRT routine. When the IPSPRT routine are complete, the time step is incremented and the procedure repeats.

5.2 Advection-diffusion model

Models describing the dispersion of a gas in air originate from the continuous-time partial differential equation (PDE) describing an advection-diffusion process [53]:

$$\frac{\delta c(p,t)}{\delta t} + \phi(p,t)\frac{\partial c(p,t)}{\partial p} = \alpha(p,t)\frac{\partial^2 c(p,t)}{\partial p^2}, \tag{5.1}$$

where $c(p,t)$ denotes the concentration of CO_2 in *part per million* (PPM) as a function of space and time, $p = [x,y,z]$ is the location vector, t is time, $\phi(p,t) = [\phi_x(p,t), \phi_y(p,t), \phi_z(p,t)]^T$ and $\alpha(p,t) = [\alpha_x(p,t), \alpha_y(p,t), \alpha_z(p,t)]^T$ are the advection and dispersion coefficients, respectively in units of $\frac{m}{s}$ and $\frac{m^2}{s}$. The surface boundary condition is

$$\left(\phi_z(p,t)c(p,t) - \alpha_z(p,t)\frac{\delta c(p,t)}{\delta z} \right) |_{p=(x,y,0)} = \lambda(x,y,t),$$

where $\lambda(x,y,t)$ represents the CO_2 leak *strength*. The leak strength, in words, is the normalized concentration (in PPM) of a source assuming the leak rate is $1\frac{m}{s}$.

Modeling the dispersion of gases in air is still a heavily researched field (see [2, 19, 54, 62] and citations within). While the advection parameter in (5.1) is simply the wind speed and direction, in the physical world, determining this parameter typically requires approximation since the wind is continuously changing. Moreover, the eddy diffusion parameters are characterized by the crosswind intensity, vertical Gaussian plume, and the wind speed, all of which vary [45]. A study of the eddy diffusion parameters is given in [54]. In our evaluation, we assume that as the wind speed increases, the crosswind intensity decreases. We also apply the same assumption as [54] and claim that in stable wind with speeds above $2\frac{m}{s}$, the effects of eddy diffusion is insignificant when compared

to advection. For simplicity in evaluation of the detection and localization strategy we neglect the vertical effects of diffusion, since it is a secondary effect when compared to the horizontal effects of advection and diffusion [55]. Applying these assumptions, we write the PDE in (5.1) in the plane $z = 0$ as the 2-d advection diffusion model as:

$$
\begin{aligned}
\frac{\delta c(x, y, 0, t)}{\delta t} &+ \phi_x(t)\frac{\partial c(x, y, 0, t)}{\partial x} + \phi_y(t)\frac{\partial c(x, y, 0, t)}{\partial y} + \lambda(x, y, t) \\
&= \alpha_x(t)\frac{\partial^2 c(x, y, 0, t)}{\partial x^2} + \alpha_y(t)\frac{\partial^2 c(x, y, 0, t)}{\partial y^2},
\end{aligned}
\tag{5.2}
$$

where $\lambda(x, y, t)$ represents the CO_2 leak strength (in units of $\frac{PPM}{s}$) at the surface, and only takes a non-zero value at the leak location. $\phi_x(t)$ and $\phi_y(t)$ are the x and y components of the wind vector respectively (assumed to be constant over the the monitoring region), and the diffusion parameters are:

$$
\begin{aligned}
\alpha_x(t) &= \frac{10}{|\phi_y(t)| + 1} \\
\alpha_y(t) &= \frac{10}{|\phi_x(t)| + 1},
\end{aligned}
\tag{5.3}
$$

which represents decreasing eddy diffusion parameter values as the wind speed increases in the orthogonal direction. Assuming a desired spatial discretization granularity of Δ, and applying a spatial Euler's approximation as in [19] to (5.4), the continuous-time advection diffusion model is written as

$$
\begin{aligned}
\frac{\delta c(x, y, 0, t)}{\delta t} &= \alpha_x(t)\left(\frac{c(x + \Delta, y, 0, t) - 2c(x, y, 0, t) + c(x - \Delta, y, 0, t)}{\Delta^2}\right) \\
&+ \alpha_y(t)\left(\frac{c(x, y + \Delta, 0, t) - 2c(x, y, 0, t) + c(x, y - \Delta, 0, t)}{\Delta^2}\right) \\
&- \phi_x(t)\left(\frac{c_x - c(x, y, 0, t)}{\Delta}\right) - \phi_y(t)\left(\frac{c_y - c(x, y, 0, t)}{\Delta}\right) + \lambda(x, y, t)
\end{aligned}
\tag{5.4}
$$

where

$$
\begin{aligned}
c_x &= \begin{cases} c(x + \Delta, y, 0, t) & \text{if } \phi_x(t) > 0 \\ c(x - \Delta, y, 0, t) & \text{if } \phi_x(t) < 0 \end{cases} \\
c_y &= \begin{cases} c(x, y + \Delta, 0, t) & \text{if } \phi_y(t) > 0 \\ c(x, y - \Delta, 0, t) & \text{if } \phi_y(t) < 0 \end{cases}
\end{aligned}
\tag{5.5}
$$

By applying the same process as [19], the advection-diffusion process can be written in a continuous-time state-space model as:

$$
\frac{\delta x(t)}{\delta t} = A(t)x(t) + Bu(t),
\tag{5.6}
$$

where $x(t)$ is the row-by-column concatenation of the planar monitoring area (this methodology is explained in Appendix C), $A(t)$ and B are the lumped parameter models governing the time evolution of $x(t)$, and $u(t)$ is the vector of source leak rates at time t. The continuous state-space model in (5.6) is discretized according to the sampling period, resulting in a discrete-time state space model for the advection diffusion process.

As an example, the system in (5.4) is simulated assuming a spatial discretization of $\Delta = 50m$ and a temporal discretization of 1 minute. This example considers a single source with strength of 200 PPM per second located at (.15 Km, .25 Km) when the wind is blowing according to the vector (2,2) (i.e. $2\frac{Km}{hr}$ in both the x and y direction). The concentration values in percent CO_2 are shown in Figure 5.3 at 2, 5, 10, and 20 minutes after the source becomes active.

The subplots within Fig. 5.3 represent different temporal snapshots of the surface CO_2 concentrations, and illustrate how the concentration levels change over space and time with respect to an active source. The model developed in this section is used by the following section to evaluate the IPSPRT and GLRDSS introduced in this dissertation.
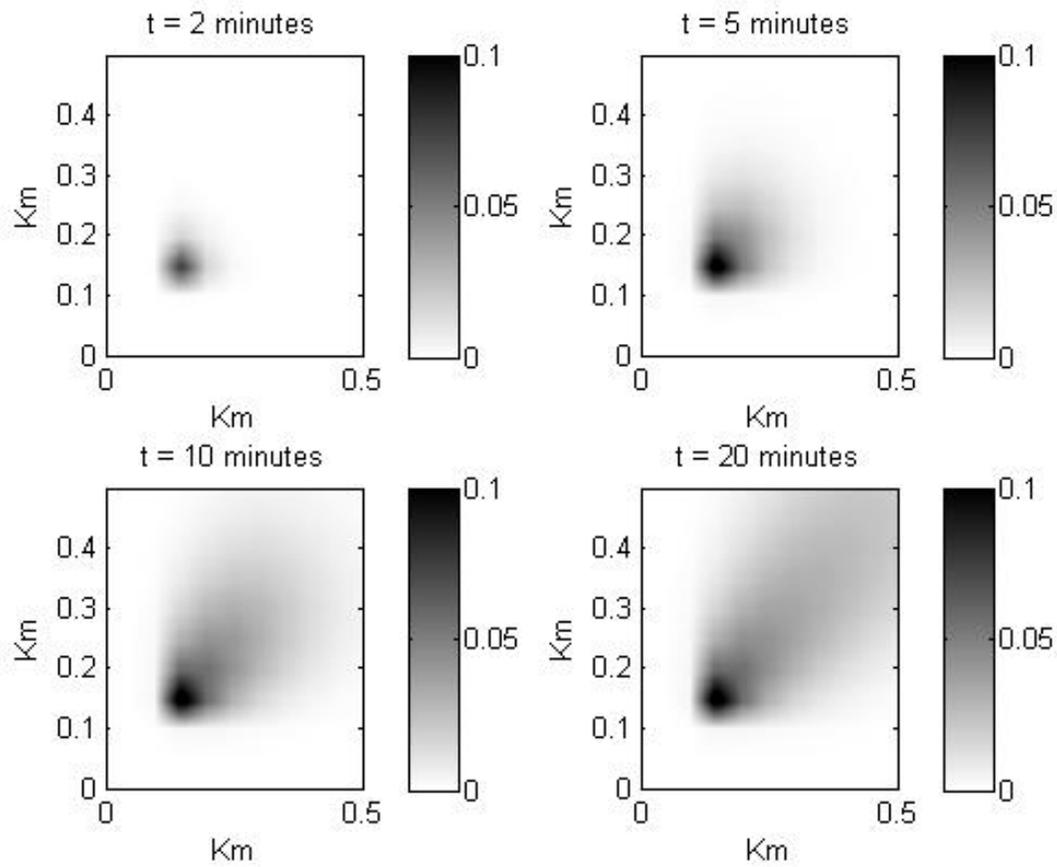
Figure 5.3: CO_2 concentrations in $\%CO_2$ at t = 2, 5, 10, and 20 minutes.

5.3 Evaluation

This section presents an evaluation of the IPSPRT for MSD for the CO_2 monitoring application. The following subsection empirically compares the performance of the alternative tests and the IPSPRT for MSD described in Chapter 3 for the CO_2 monitoring application. In the final subsection, the robustness of the IPSPRT and GLRDSS is evaluated empirically in the presence of common errors and failures using the test bed described in Section 5.1.

5.3.1 Performance evaluation

This section presents a performance evaluation, in terms of time-to-decision, of different feasible detection and localization strategies for the CO_2 monitoring application described in Section 5.2. We assume a sensor network of 22 sensors and 49 potential sources, distributed as in Fig. 5.4, where a square denotes a sensor location and a dot represents a potential source location. $1,000$ simulations were performed, each lasting for 1200 time
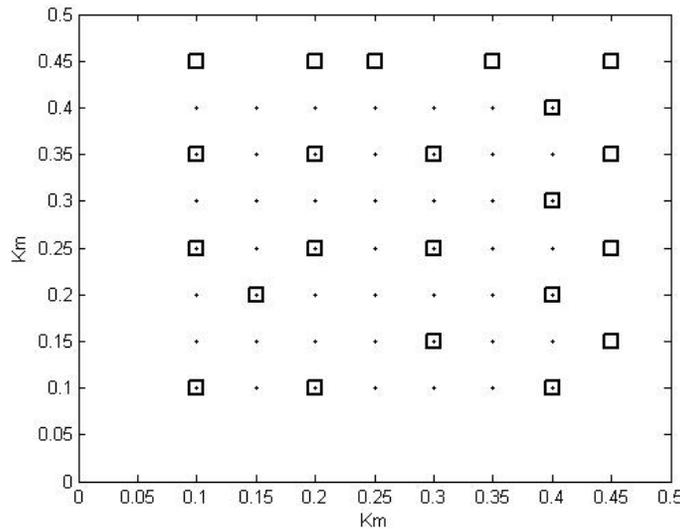

Figure 5.4: Sensor and potential source locations.

steps with a randomly located single source becoming active at time step 600. We assume

96

the wind is always blowing in the direction $(1, 1)$ and tested 6 different wind intensities ranging from $0\frac{Km}{hr}$ to $8.5\frac{Km}{hr}$. For this evaluation, we do not apply the GLRDSS, rather, at each time step, all the sensors are selected. For detection evaluation, we assume the probability of false alarm is .01 and the probability of miss is .05, and evaluate the time-to-decision of the naive test, estimation test and the IPSPRT (introduced in Chapter 3).[1] For localization evaluation, we assume the IPSPRT is used for detection, and upon correctly rejecting the null hypothesis, an estimation-based approach to localization is compared to the IPSPRT in terms of the percentage of sources identified to be active that are actually active when the maximum probability of a type III error is .10.

Fig. 5.5 illustrates the detection results for deciding whether some source or no source is active, using the naive approach, estimation approach, and the proposed approach in terms of time-to-decision versus the wind speed. In Fig. 5.5, the naive and estimation approaches are denoted by the dotted and dashed lines respectively. The solid line and the dash-dot line denote the proposed approach when the null hypothesis is accepted (no sources are active) and when the null hypothesis is rejected (some source is active). Recalling from the descriptions of each test from chapter 3, the naive and estimation based approaches must specify a time-to-decision, since the null hypothesis is only accepted when that time is exceeded; however, the proposed approach accepts or rejects the null hypothesis when their is enough information to guarantee the required probability of miss or probability or false alarm is bounded. In the following discussion, we discuss the IPSPRT when the null hypothesis is accepted and the when the null hypothesis is rejected separately. Unlike other tests, the IPSPRT does not require a pre-specified number of observations to make a decision and still bound the probability of error. Separating the IPSPRT results into *tests that accept H_0* and *tests that reject H_0* allows us to discuss the performance (in time-to-decision) when no event occurs (which is much more likely) in comparison to when

---

[1] All the strategies for detection and localization are introduced in Chapter 3, and provides an evaluation of performance in terms of the probability of false alarm, probability of miss, and probability of type III error therein.
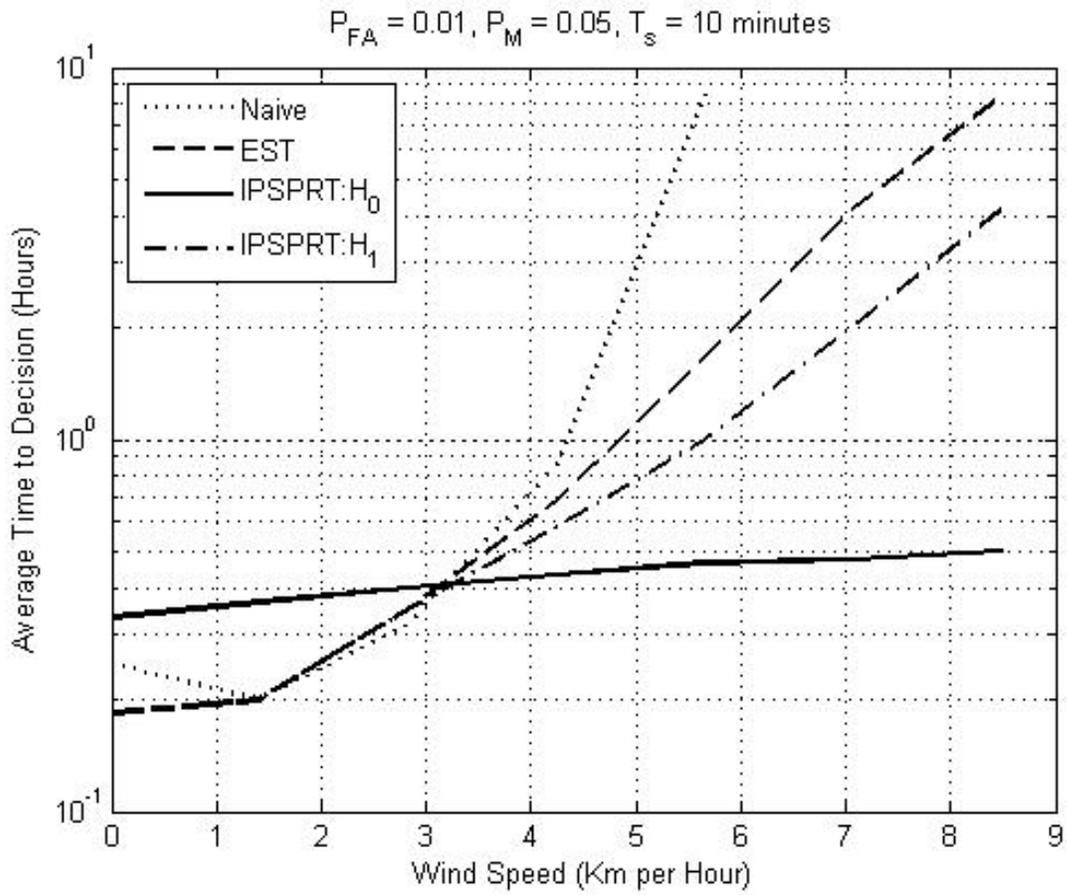
Figure 5.5: Time to decision vs. wind speed.

some event occurs. Since the other feasible tests considered have a fixed time-to-decision regardless of whether the null hypothesis is accepted or rejected, we do not distinguish their results by their decision.

The results in Fig. 5.5 illustrate that for low wind speeds, the naive and estimation approaches have a time-to-decision that is .15 hours (less than 1 time step) better than the proposed strategy when accepting the null hypothesis and approximately the same when rejecting the null hypothesis. Observing that the sensors are sampled every ten minutes, this relates to about a one-time-step delay in decision. As the wind speed increases, all approaches require longer monitoring periods to achieve the same probability of false alarm and probability of miss as in lower wind speeds. This is due to the decreased CO_2 concentrations being observed since CO_2 is dispersed more quickly in larger winds. The increase in time-to-decision is largest in the naive, estimation approaches, and the IPSPRT when the resulting decision rejects the null hypothesis, as shown in Fig. 5.5. When the wind speed is $8\frac{km}{hr}$, the estimation approach requires a time-to-decision of 5.5 hours (330 time steps) and the naive approach was never able to meet the required probability of false alarm and probability of miss for the maximum detection period considered (10 hours or 600 times steps). The IPSPRT accepts the null hypothesis in an average of .5 hours (3 time steps), and rejects the null hypothesis is an average of 2.2 hours (about 18 time steps). Since it is assumed that it is much more likely that no source will be active as opposed to some source being active, the IPSPRT becomes increasingly superior to both the naive and estimation approaches as the wind speed increases.

Fig. 5.6 illustrates the localization results in terms of percentage of identified sources which are active versus wind speed using the estimation-based and IPSPRT approaches. Both approaches assume the IPSPRT is used for detection. In Fig. 5.6, the solid line denotes the IPSPRT and the dotted line corresponds to the estimation-based strategy. When there is no wind, both strategies achieve over a 95% chance that a source identified
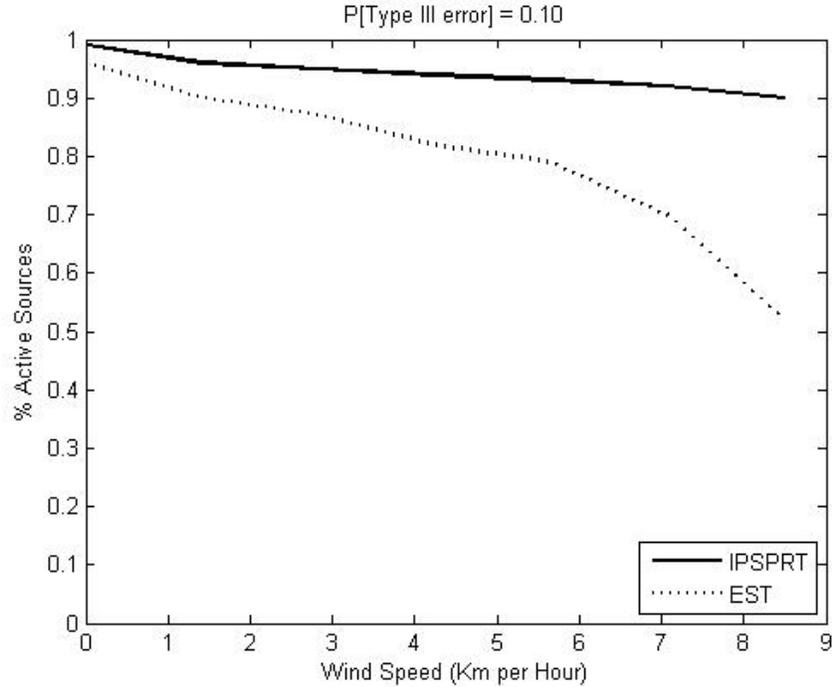
Figure 5.6: Percentage of active sources vs. wind speed.

as active is actually active. As the wind speed increases, both strategies experience a decrease in the probability that a source identified as active is actually active. When the wind speed is $8\frac{km}{hr}$, the localization of the IPSPRT achieves a 91% chance that an identified source is active, while the estimation-based localization only provides a 59% chance. Thus, the increased computational complexity of the IPSPRT pays dividends as the wind speed increases.

The results above indicate that as the wind speed increases, the IPSPRT performs increasingly better than other feasible approaches.

5.3.2 Robustness evaluation

In the previous section, the IPSPRT is shown to perform much better than other feasible strategies for large-scale CO_2 sequestration site monitoring as the wind speed increases. This section investigates the robustness of the proposed strategy in terms of common errors/failures associated with environmental monitoring using a WSN through the test bed

100

implementation described in Section 5.1. We consider five different error/failure scenarios: packet loss, model parameter errors, localization errors, and two different types of sensor failures. To evaluate packet loss, we consider how packet loss affects time-to-decision and localization accuracy for the sensor and potential source configuration shown in Fig. 5.4. In the following, we use the packet loss scenario as a control experiment for all other types of errors since all errors are evaluated using the wireless sensor network test bed and thus include packet loss.

The second type of error considered occurs when model parameters are different than the assumed parameters. To simulate this error, the advection-diffusion model used for simulating the environment assumes a randomly selected value for the wind speed in both the x and y directions, such that the expected value of the actual wind speed matches the wind speed assumed by the model used for detection and localization. We assume the wind speed varies independently in both the x and y directions according to a normal distribution with unit covariance.

Sensor localization error is the third type of error considered. To simulate this error, the test bed sensors are moved to the locations specified by the squares in Fig. 5.7 while the model used for detection and localization of sources still assumes the sensor layout in Fig. 5.4. The fourth and fifth types of failures considered occur when sensor nodes die (drop out of the network). The first sensor death considered, henceforth referred to as *sensor failure 1*, assumes that the interior sensors denoted by the filled boxes in Fig. 5.8 cannot deliver observations. Similarly, the second sensor death considered, referred to as *sensor failure 2*, assumes that the exterior sensors denoted by the filled boxes in Fig. 5.9 are not capable of being sampled.

To evaluate the robustness of the IPSPRT and GLRDSS in the presence of the afore-mentioned errors/failures, we consider four different active source scenarios

1. *Synchronous distributed sources* : Two sources located far apart from one another
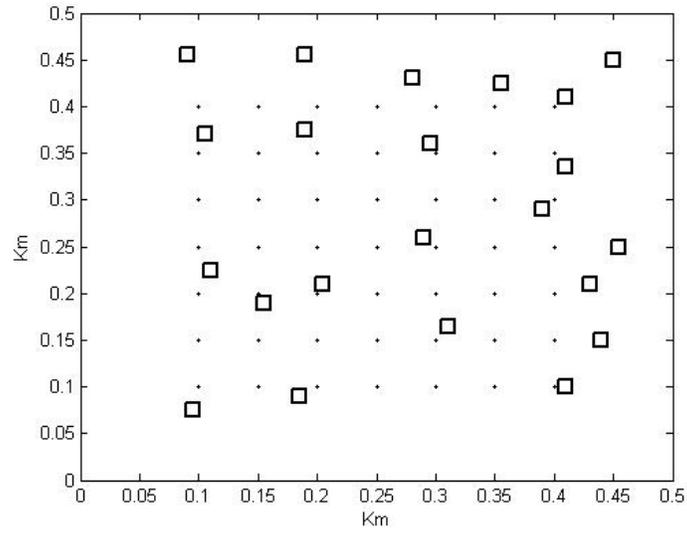
101

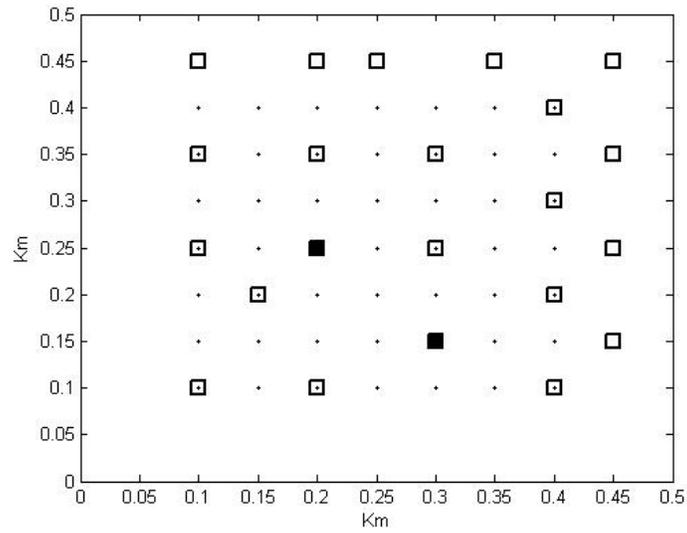Figure 5.7: Sensor localization error layout.

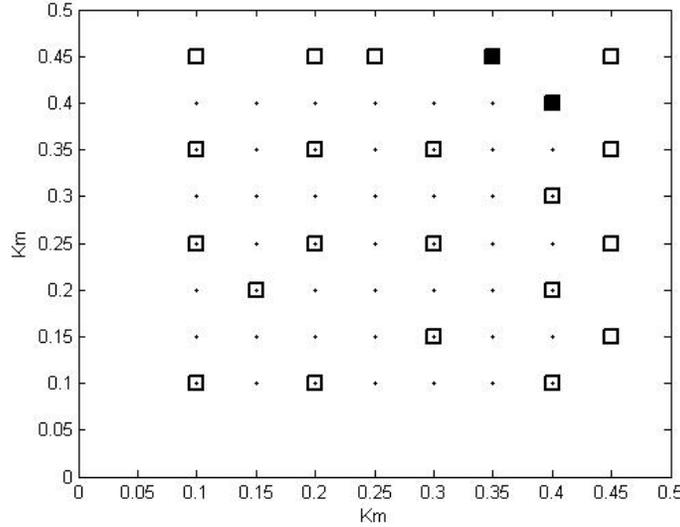Figure 5.8: Sensor failure 1 (interior sensor failure).

Figure 5.9: Sensor failure 2 (exterior sensor failure).

that become active at the same time.

2. *Synchronous clustered sources* : Two sources located near one another that become active at the same time.

3. *Asynchronous distributed sources* : Two sources located far apart from one another that become active at different times.

4. *Asynchronous clustered sources* : Two sources located near one another that become active at different times.

Each active source scenario is evaluated in low wind $(2.8\frac{Km}{hr})$ and in high wind $(8.5\frac{Km}{hr})$. The test scenarios are described in detail in Appendix E. The desired time-to-decision is 1 hour (6 sampling periods). In the following, the IPSPRT and GLRDSS performance in terms of time to decision is investigated. The results are classified by the IPSPRT decision to accept or reject the null hypothesis. For each test decision, the time-to-decision is investigated in the presence of the errors and failures mentioned above. It is important to consider the test decision because (as shown in the previous section) the time-to-decision can differ significantly when accepting versus rejecting the null hypothesis.

103

Table 5.1: Average time-to-decision (in sampling periods) for accepting H_0

	Wind Speed	Control (Packet Loss)	Localization Error	Parameter Error	Sensor Failure 1	Sensor Failure 2
Synchronous Distributed Sources	Low	6.79	6.92	7.24	6.89	7.77
	High	7.71	7.69	7.74	7.72	11.68
Synchronous Clustered Sources	Low	6.81	6.89	7.22	6.90	7.79
	High	7.72	7.67	7.76	7.74	11.72
Asynchronous Distributed Sources	Low	6.82	6.91	7.23	6.92	7.80
	High	7.72	7.68	7.74	7.75	11.74
Asynchronous Clustered Sources	Low	6.81	6.91	7.28	6.87	7.79
	High	7.74	7.62	7.81	7.80	11.78

Table 5.1 illustrates the average time-to-decision in sampling steps for deciding that no sources are active (accepting the null hypothesis). In Table 5.1 the results indicate that there is not a significant increase in the expected time-to-decision when localization error, parameter error, or interior sensor failure occurs. The only error that has a significant effect on the time-to-decision occurs when an exterior sensor fails in high wind situations, where the increase in the time-to-decision is about 50 percent. There is no significant difference in the average time-to-decision for each error between the different source scenarios when deciding no sources are active. This is intuitive since the null hypothesis is rarely accepted when a source is actually active (because the probability of miss is .05). With the exception of when an exterior sensor node failed, the expected decision time was within two time steps of the desired time-to-decision of 6 time steps. The difference between the desired and actual time-to-decision is apparent in the control case, which suggests that packet loss is a contributing factor.

Table 5.2 illustrates the average time-to-decision in sampling steps for deciding some source is active (rejecting the null hypothesis). The results show that an interior sensor failure has the largest effect on the average time-to-decision, amongst the types of errors/failures considered. Observing from Fig. 5.8 and the description of the descriptions

Table 5.2: Average time-to-decision (in sampling periods) for rejecting H_0

	Wind Speed	Control (Packet Loss)	Localization Error	Parameter Error	Sensor Failure 1	Sensor Failure 2
Synchronous Distributed Sources	Low	1.92	2.68	2.78	3.76	2.61
	High	24.82	25.10	28.76	31.05	25.60
Synchronous Clustered Sources	Low	1.61	1.92	2.21	2.78	2.01
	High	12.61	12.83	15.34	20.10	12.79
Asynchronous Distributed Sources	Low	2.01	2.81	2.92	3.91	2.83
	High	24.79	25.11	28.75	31.11	25.52
Asynchronous Clustered Sources	Low	2.02	2.79	3.11	4.01	2.96
	High	26.61	26.83	30.07	34.58	26.91

of the active source scenarios considered in appendix E, we observe that the interior nodes which failed were the closest down-wind nodes from the active sources. Thus, in high wind situations, where the concentration of CO_2 is small except for near the source, having nodes fail in the down-wind proximity can cause a significant increase in the time-to-decision.

The results in Table 5.2 also illustrate that in low wind situations, the null hypothesis is rejected more quickly than it is accepted, and in high wind situations, the opposite occurs. This is consistent with the results in Fig. 5.5, where the time-to-decision to accept the null hypothesis increases at a much slower rate than the decision to reject the null hypothesis. As shown in all the test scenarios, when sensors observe a significant increase in the the CO_2 concentration level, a detection occurs quickly. As the CO_2 concentration decreases (due to either a decrease in the source magnitude or an increase in the wind speed), the time-to-decision increases. Intuitively, based on the results for accepting the null hypothesis in Table 5.1 and rejecting the null hypothesis in Table 5.2, one could institute a preemptive detection scheme based on the duration of the monitoring period. As the time-to-decision increases, it becomes more likely (based on the empirical data) that an active source exists, but can't be accurately localized yet (since recalling from chapter 3, the localization accuracy depends on the probability of false alarm and probability of

Table 5.3: Percentage of localized sources that are active

	Wind Speed	Control (Packet Loss)	Localization Error	Parameter Error	Sensor Failure 1	Sensor Failure 2
Synchronous Distributed Sources	Low	95.1	93.0	87.7	86.2	94.8
	High	88.7	80.3	71.4	68.5	88.2
Synchronous Clustered Sources	Low	96.8	95.2	91.0	91.5	96.6
	High	92.5	85.1	78.1	75.2	89.7
Asynchronous Distributed Sources	Low	95.3	93.0	87.8	86.9	94.1
	High	88.1	79.5	71.6	68.8	88.9
Asynchronous Clustered Sources	Low	88.0	86.5	75.2	70.1	84.2
	High	75.1	71.0	57.6	48.6	68.3

miss achieved by the detection procedure).

We observe in Table 5.2 that when two sources become active at the same time and are near one another, that the time-to-decision is smaller than in any other corresponding source scenario. This is due to the overall increase in the CO_2 concentration at each of the down-wind sensor nodes. When multiple sources are active in the same proximity, the results are similar to a single source with a larger leak rate. These multiple proximate active source scenarios improve the time to detection, but may decrease the localization accuracy (in terms of the number of sources identified to be active that are actually active). When comparing the time-to-decision for accepting the null hypothesis (Table 5.1) versus rejecting the null hypothesis (Table 5.2), we observe that an exterior node sensor failure significantly increases the time-to-decision for accepting the null hypothesis, while a failure of down-wind nodes close to the active source(s) has a similar effect when rejecting the null hypothesis. These observations lead us to believe that sensor failures (nodes dropping out of the network) are of a key concern when a WSN is used to perform MSD.

Table 5.3 illustrates the localization accuracy in terms of the percentage of sources identified to be active that are actually active when the null hypothesis is rejected. The results indicate that for all source scenarios, an interior sensor failure (corresponding to

a failure of the closest down-wind sensors) and a model parameter error have the largest effects on the localization accuracy. We note that, while a model parameter error did not cause a significant error in detection, it does significantly affect localization. This error is only increased as the model error increases. Moreover, significant model parameter errors will also significantly affect the detection performance, but the results in Table 5.2 and Table 5.3 suggest that parameter errors affect localization more than detection. This is intuitive, since detection concerns the broad problem of determining if active sources exist, while localization concerns the much finer problem of identifying which sources are active. This result reinforces the benefit of separating the tasks of detection and localization because detection (deciding whether no or some sources are active) is somewhat robust to small parameter errors.

An exterior sensor failure and sensor localization error do not have significant effects on the localization accuracy, due in part to the additional information required to reject the null hypothesis initially. We note that significant localization errors, can significantly affect performance, but in the test scenarios considered, the sensor node locations differed from their assumed locations by 40 meters (less than the spatial discretization of 50 meters) and performance did not suffer significantly. Although significant localization errors were not considered (where the nodes are 100s of meters from there assumed location), we suspect that much like the effect of parameter errors, localization accuracy will be affected significantly more than detection performance (in terms of time-to-decision).

Chapter 6

Conclusions and Future Work

Large-scale long-term MSD using WSNs is a rich problem with many research areas provided by the sheer magnitude of the problem (computational and information processing issues) and the uncertainty associated with gathering information using a wireless sensor network. To this end, this dissertation first contributes a scalable heuristic solution to large-scale persistent MSD called the iterative partial sequential probability ratio test (IP-SPRT). The second contribution is an empirical evaluation of the IPSPRT in comparison to other feasible strategies for large-scale MSD. The third contribution of this dissertation is a scalable dynamic sensor selection (DSS) strategy referred to as the GLRDSS that prolongs the lifetime of a WSN used for large-scale MSD. The final contribution of this dissertation is a physical implementation and evaluation of the robustness of the IPSPRT and GLRDSS with respect to common sensor networking errors and failures that demonstrates the effectiveness of both solutions in real-world MSD applications using a WSN.

A scalable two-step heuristic solution to persistent MSD is introduced, that does not suffer from an explosion of potential hypotheses due to all space-time combinations of active and inactive sources, whereas other strategies do. The time complexity is avoided by sequentially testing only hypotheses assuming a specific time when sources become active. Once a decision is made, the time is incremented and the process repeats, and thus the strategy does not experience an increase in complexity with time. Space complexity

is avoided by testing only the hypotheses that assume a single-active source. It is shown that, in the scalar observation case, testing the single-active source hypotheses is sufficient for testing all the hypotheses. As a heuristic, this is extended to the multiple-observation case.

The IPSPRT for MSD is compared empirically to two other common feasible strategies. One strategy simply thresholds the observations to determine whether an active source is present, while the other strategy uses a dynamic model to generate an estimate that is then used in a threshold test. The empirical comparison suggests that as the maximum probability of false alarm, maximum probability of miss, and maximum probability of type III error decrease, the IPSPRT performs increasingly better than the other feasible strategies.

A DSS strategy is introduced that is tailored to the IPSPRT for large-scale MSD. The GLRDSS selects the fewest number of sensors at each time step such that the expected time-to-decision is approximately the desired time-to-decision. Empirical results suggest that significantly fewer sensors can be selected at each time step for an increasing desired time-to-decision, where selecting fewer sensors at each time step is assumed to increase the network lifetime.

A test bed is implemented consisting of 22 firefly nodes. The test bed uses light projection and sensing to emulate environmental monitoring for a carbon sequestration site monitoring application. The robustness of the IPSPRT and GLRDSS is evaluated in the presence of common sensor networking errors and failures such as: model parameter errors, sensor localization errors, sensor death, and packet loss. The test bed results suggest that under most conditions the MSD and DSS strategies are robust to small parameter errors and sensor localization errors; however, sensor death can affect performance significantly.

Through the contributions of this dissertation, we recognize several problems which should be addressed as part of future work on MSD using a WSN. In the current formu-

110

lation, we assume knowledge of the state space model and do not address the relationship between the performance parameters defined for the discrete system and the actual performance in the real world. In regards to generating a discrete state-space system model for the continuous real world phenomena, issues such as robustness to time and space discretization [19, 34], identifying parameters in the underlying partial differential equations [54, 55], and defining the potential source set are problems that should be investigated. Modeling errors arise when we generate discrete space-time models for continuous dynamics governed by partial differential equations. These modeling errors undoubtedly affect the overall system performance, especially in terms of detection and localization as seen by the test bed evaluation.

The robustness results of the detection and localization strategy in the presence of sensor death suggests that sensor deployment can play a significant role in performance. This work assumed a given sensor deployment. Future work is needed to specify criteria for what constitutes the *best sensor deployment* and a method of determining this deployment. This question of sensor deployment (and more generally sensor selection) has been considered by many researchers for the problem of providing a minimum mean squared error estimate [16, 25, 70, 74], but extensions to the MSD problem have not yet been considered. The test bed robustness results illustrate that sensor placement plays a key role in detection and localization where some sensor deaths can have a significant effect on performance and others a marginal effect.

Several wireless networking issues need to be addressed, such as security of information, maximizing network reliability (through routing protocols), characterization of routing protocols for channel reliability modeling, and real-time sensor calibration. These wireless sensor networking issues are of increasing importance due to the sensitivity of large-scale MSD strategies to inaccurate and missing data. To achieve high performance, guarantees must be made on the security of the information, the minimum network reliability, and the

accuracy of the sensed data. Errors such as sensor drift (not considered in the robustness evaluation) may have a significant effect on the MSD strategy performance if not properly handled.

Lastly, the current formulation considers only persistent sources. Although the problem formulation changes significantly with the allowance of intermittent sources, empirical analysis of the persistent source detector performance in the presence of intermittent sources for various time-to-decision bounds would give insight to applying this strategy to a much broader class of MSD problems. While intermittent sources have been considered for small-scale multiple source detection [72], these approaches do not extend to large-scale problems. With technological advances in wireless sensor technology, large-scale MSD problems that were once thought to be impossible due to physical constraints are quickly becoming the detection problems of today [5, 71].

Chapter 7

Bibliography

[1] H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, Dec 1974. ISSN 0018-9286. 1.1

[2] V. V. Anh, M. Azzi, H. Duc, G. M. Johnson, and Q. Tieng. A reactive state-space model for prediction of urban air pollution. *Environmental Modeling and Software*, 13 (3-4):239 – 246, 1998. 5.2

[3] P Armitage. Sequential analysis with more than two alternative hypotheses and its relation to discriminant function analysis. *Journal of the Royal Statistical Society*, Ser. B(12):137–144, Nov. 1950. 1.1

[4] M. Athans, D. Castanon, K. Dunn, C. Greene, Wing Lee, Jr. Sandell, N., and A. Willsky. The stochastic control of the f-8c aircraft using a multiple model adaptive control (mmac) method–part i: Equilibrium flight. *Automatic Control, IEEE Transactions on*, 22(5):768–780, Oct 1977. ISSN 0018-9286. 3

[5] T. G. Bell. Submarine and sonar detection. Technical Report 545, U.S. Navy Underwater Sound Lab, 1962. 1.3, 6

[6] B. Brumback and M. Srinath. A chi-suare test for fault-detection in kalman filters. *IEEE Transactions on Automatic Control*, June 1987. 1, 3

[7] Jren-Chit Chin, l Hong Hou, J.C. Hou, C. Ma, N.S. Rao, M. Saxena, M. Shankar, Yong Yang, and D.K.Y. Yau. A sensor-cyber network testbed for plume detection, identification, and tracking. *Information Processing in Sensor Networks, 2007. IPSN 2007. 6th International Symposium on*, pages 541–542, 25-27 April 2007. doi: 10. 1109/IPSN.2007.4379715. 1

[8] D. R. Cox and F. J. Anscombe. Sequential tests for composite hypotheses. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48:290–+, 1952. doi: 10.1017/S030500410002764X. 1.1

[9] Michael A. Demetriou. Process estimation and moving source detection in 2-d diffusion processes by scheduling of sensor networks. *American Control Conference, 2007. ACC '07*, pages 3432–3437, 9-13 July 2007. ISSN 0743-1619. doi: 10.1109/ACC.2007. 4282247. 1, 1.3

[10] A. Eswaran, A. Rowe, and R. Rajkumar. Nano-rk: an energy-aware resource-centric rtos for sensor networks. In *Real-Time Systems Symposium, 2005. RTSS 2005. 26th*

IEEE International, pages 10 pp.–265, 2005. doi: 10.1109/RTSS.2005.30. URL `http://dx.doi.org/10.1109/RTSS.2005.30`. 5.1

[11] Z. Feng, K. Teo, and Y. Zhao. Branch and bound method for sensor scheduling in discrete time. *Journal of Industrial and Management Optimization*, 1(4):499–512, November 2005. 1.2

[12] Jr. Forney, G. Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference. *Information Theory, IEEE Transactions on*, 18 (3):363–378, May 1972. ISSN 0018-9448. 1.1

[13] Jr. Forney, G.D. The viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278, March 1973. ISSN 0018-9219. 1.1

[14] E. B. Fox, J. W. Fisher, and A. S. Willsky. Detection and localization of material releases with sparse sensor configurations. *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 55(5): 1886–1898, May 2007. ISSN 1053-587X. doi: 10.1109/TSP.2006.890914. 1, 1.3

[15] M. Grant and S. Boyd. Cvx: Matlab software for disciplined convex programming [web page and software], October 2007. URL `http://www.stanford.edu/~boyd/cvx`. 4.3.2

[16] Carlos Guestrin, Andreas Krause, and Ajit Singh. Near-optimal sensor placements in gaussian processes. In *In ICML*. ICML, 2005. 1.2, 6

[17] V. Gupta, T. Chung, B. Hassibi, and R. Murray. On a stochastic algorithm for sensor scheduling. In *IFAC*, 2005. 1.2

[18] M. Hecht and M. Schwartz. M-ary sequential detection for amplitude-modulated signals in one and two dimensions. *Communication Technology, IEEE Transactions on*, 16(5):669–675, October 1968. ISSN 0018-9332. doi: 10.1109/TCOM.1968.1089902. 1.1

[19] E. Hernandez, F. Martan, and F. Valero. State-space modeling for atmospheric pollution. *Journal of Applied Meteorology*, 30(6):793–811, 1991. 1.3, 5.2, 5.2, 5.2, 6

[20] D. G. Hoel, G. H. Weiss, and R. Simon. Sequential tests for composite hypotheses with two binomial populations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(3):302–308, 1976. 2.2.1

[21] G.E. Hovland and B.J. McCarragher. Dynamic sensor selection for robotic systems. volume 1, pages 272–277 vol.1, Apr 1997. doi: 10.1109/ROBOT.1997.620050. 1.2

[22] Yinlun Huang. On distributed fault-tolerant detection in wireless sensor networks. *IEEE Trans. Comput.*, 55(1):58–70, 2006. ISSN 0018-9340. doi: http://dx.doi.org/10.1109/TC.2006.13. Student Member-Xuanwen Luo and Member-Ming Dong. 1.3

[23] J.Bay. *Fundamentals of Linear State Space Systems*. McGraw-Hill, 1998. 1.3

[24] Xiaofan Jiang, Prabal Dutta, David Culler, and Ion Stoica. Micro power meter for energy monitoring of wireless sensor networks at scale. In *IPSN '07: Proceedings of the 6th international conference on Information processing in sensor networks*, pages 186–195, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-638-X. doi: http:

//doi.acm.org/10.1145/1236360.1236386. 4

[25] S. Joshi and S. Boyd. Sensor selection via convex optimization. *Signal Processing, IEEE Transactions on*, 57(2):451–462, Feb. 2009. ISSN 1053-587X. doi: 10.1109/ TSP.2008.2007095. 1.2, 4.2, 4.3, 4.3.2, 4.4, 6

[26] Thomas Kailath, Life Fellow, and H. Vincent Poor. Detection of stochastic processes. *IEEE Trans. Inform. Theory*, 44:2230–2259, 1998. 1.1, 2.3.2, 3

[27] R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972. 1.2, 3.1.1, 4.2, 4.3.2

[28] T. Kerr. Statistical analysis of a two-ellipsoid overlap test for real-time failure detection. *IEEE Transactions on Automatic Control*, August 1980. 1, 3

[29] Ioannis K. Kookos and John D. Perkins. A systematic method for optimum sensor selection in inferential control systems. *Industrial and Engineering Chemistry Research*, 38(11):4299–4308, 1999. doi: 10.1021/ie9902737. URL `http://pubs.acs.org/doi/ abs/10.1021/ie9902737`. 1.2

[30] Tze Leung Lai. Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems. *Information Theory, IEEE Transactions on*, 46(2): 595–608, Mar 2000. ISSN 0018-9448. doi: 10.1109/18.825826. 1, 1.1, 3

[31] A. H. Land and A. G Doig. An automatic method of solving discrete programming problems. *Econometrica*, 28(3):497–520, 1960. URL `http://jmvidal.cse.sc.edu/ library/land60a.pdf`. 1.2, 4.2, 4.3.2

[32] J. Liu, J. Reich, and F. Zhao. Collaborative in-network processing for target tracking. *Journal on Applied Signal Processing*, 4:378–391. 1

[33] Rahul Mangharam, Anthony Rowe, and Raj Rajkumar. Firefly: a cross-layer platform for real-time embedded wireless networks. *Real-Time Systems*, 37(3):183–231, December 2007. doi: 10.1007/s11241-007-9028-z. URL `http://dx.doi.org/10.1007/ s11241-007-9028-z`. 5.1

[34] A. Melman. Geometry and convergence of euler's and halley's methods. In *SIAM Review*, volume 39, pages 728–735. Society for Industrial and Applied Mathematics, December 1997. 1.3, 6

[35] M.P. Michaelides and C.G. Panayiotou. Event detection using sensor networks. *Decision and Control, 2006 45th IEEE Conference on*, pages 6784–6789, 13-15 Dec. 2006. doi: 10.1109/CDC.2006.377723. 1, 1.3

[36] I. Mitroff and T. Featheringham. On systemic problem solving and the error of the third kind. *Behavioral Science*, 19(6):383–393, 1974. doi: 10.1002/bs.3830190605. 2.3.2

[37] F. Mosteller. A k-sample slippage test for an extreme population. *The annals of mathematical statistics*, 19(1):58–65, Mar 1948. 2.3.2

[38] K. G. Murty. *Linear Complementarity, Linear and Nonlinear Programming*. AHelderman-Verlag, 1988. 1.2, 4.2, 4.3.2

[39] Y. Newbold, P. Ho. Detection of changes in characteristics of a Gauss-markov process. *IEEE Trans. Aerospace Elec. Sys.*, 5:707–718, Sept. 1968. 1.1

[40] J. Neyman and E. S. Pearson. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Royal Society of London Philosophical Transactions Series A*, 231:289–337, 1933. 1.1

[41] G. T. Nofsinger and G. V. Cybenko. Airborne plume tracking with sensor networks. In *Unattended Ground, Sea, and Air Sensor Technologies and Applications VIII. Edited by Carapezza, Edward M.. Proceedings of the SPIE, Volume 6231, pp. 623112 (2006).*, volume 6231 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, June 2006. doi: 10.1117/12.670130. 1, 1.3

[42] Y. Oshman. Optimal sensor selection strategy for discrete-time state estimators. *IEEE Transactions on Aerospace and Electronic Systems*, 30(2):307–314, April 1994. 1.2

[43] K. B. Petersen and M. S. Pedersen. The matrix cookbook, oct 2008. URL `http://www2.imm.dtu.dk/pubdb/p.php?3274`. 4.1, D

[44] H. Vincent Poor. *An introduction to signal detection and estimation (2nd ed.).* Springer-Verlag New York, Inc., New York, NY, USA, 1994. ISBN 0-387-94173-8. 2.2

[45] K. Shankar. Rao and Environmental Research Laboratories (U.S.). *Plume concentration algorithms with deposition, sedimentation, and chemical transformation [microform] / K. Shankar Rao.* U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, Environmental Research Laboratories, [Washington, D.C.?] :, 1983. 1.3, 5.2

[46] J Rissanen. Modelling by the shortest data description. *Automatica*, 14:465–471, 1978. 1.1

[47] Lorenzo A. Rossi, Bhaskar Krishnamachari, and C.-C. Jay Kuo. Monitoring of diffusion processes with pde models in wireless sensor networks. In *Defense and Security Symposium 2004*, April 12-16 2004. 1

[48] H. Rowaihy, S. Eswaran, M. Johnson, D. Verma, A. Bar-Noy, T. Brown, and T. La Porta. A survey of sensor selection schemes in wireless sensor networks. Technical report, Pennsylvania State University, March 2007. 1.2, 4

[49] A. Rowe, K. Lakshmanan, and R. Rajkumar. Sampl: A simple aggregation and message passing layer. In *Real-Time Systems Symposium*, Nov 2008. 5.1

[50] Prasad Saripalli, James Amonette, Fred Rutz, and Neeraj Gupta. Design of sensor networks for long term monitoring of geological sequestration. *Energy Conversion and Management*, 47(13-14):1968 – 1974, 2006. 1.3

[51] Louis L. Scharf. *Statistical Signal Processing, Detection, Estimation, and Time Series Analysis.* Addison-Welsley Publishing Company Inc., Reading, Massachusetts, 1991. 2.2

[52] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6 (2):461–464, 1978. ISSN 00905364. URL `http://www.jstor.org/stable/2958889`.

1.1

[53] John H. Seinfeld. *Atmoshperic Chemistry and Physics of Air Polution.* John Wiley & Sons, Inc., New York, 1995. 1.3, 5.2

[54] Maithili Sharan and Anil kumar Yadav. Simulation of diffusion experiments under light wind, stable conditions by a variable k-theory model. *Atmospheric Environment*, 32(20):3481 – 3492, 1998. 1.3, 5.2, 6

[55] Maithili Sharan, Anil Kumar Yadav, M. P. Singh, P. Agarwal, and S. Nigam. A mathematical model for the dispersion of air pollutants in low wind conditions. *Atmospheric Environment*, 30(8):1209 – 1220, 1996. 1.3, 5.2, 6

[56] Xiaohong Sheng and Yu-Hen Hu. Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *Signal Processing, IEEE Transactions on*, 53(1):44–53, Jan. 2005. ISSN 1053-587X. doi: 10.1109/TSP. 2004.838930. 1, 1.1

[57] Milton Sobel and Abraham Wald. A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. *The Annals of Mathematical Statistics*, 20(4):502–522, 1949. ISSN 00034851. URL `http://www.jstor.org/stable/2236307`. 1.1

[58] V. Srivastava, K. Plarre, and F. Bullo. Randomized sensor selection in sequential hypothesis testing, September 2009. Available at http://arxiv.org/abs/0909.1801. 1.2, 4.4

[59] United States Secretary of Energy Steven Chu. Letter on carbon capture and sequestration, Oct 2009. 1.3

[60] Youngchul Sung, Lang Tong, and H. Vincent Poor. Sensor configuration and activation for field detection in large sensor arrays. In *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks*, page 22, Piscataway, NJ, USA, 2005. IEEE Press. ISBN 0-7803-9202-7. 1.3

[61] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory.* John Wiley & Sons, Inc., New York, 1968. 2.2

[62] D.B Turner. *Workbook of atmospheric dispersion estimates: an introduction to dispersion modeling, 2nd Edition.* CRC Press, 1994. 1.3, 5.2

[63] S. Verdu. Minimum probability of error for asynchronous gaussian multiple-access channels. *Information Theory, IEEE Transactions on*, 32(1):85–96, Jan 1986. ISSN 0018-9448. 1.1

[64] A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *Information Theory, IEEE Transactions on*, 13(2):260–269, Apr 1967. ISSN 0018-9448. 1.1

[65] Abraham Wald. *Sequential Analysis.* John Wiley & Sons, Inc., New York, 1947. 1.1, 2.2, 2.2.1, 3.3.1, B, B.2

[66] Hanbiao Wang, G. Pottie, Kung Yao, and D. Estrin. Entropy-based sensor selection heuristic for target localization. pages 36–45, April 2004. doi: 10.1109/IPSN.2004.

1307321. 1.2

[67] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference (Springer Texts in Statistics)*. Springer, September 2004. ISBN 0387402721. URL http://www.worldcat.org/isbn/0387402721. 2.2

[68] M. Wax. Detection and localization of multiple sources via the stochastic signals model. *Signal Processing, IEEE Transactions on*, 39(11):2450–2456, Nov 1991. ISSN 1053-587X. doi: 10.1109/78.98000. 1.1

[69] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(2):387–392, Apr 1985. ISSN 0096-3518. 1.1

[70] J.E. Weimer, B. Sinopoli, and B.H. Krogh. A relaxation approach to dynamic sensor selection in large-scale wireless networks. *Distributed Computing Systems Workshops, 2008. ICDCS '08. 28th International Conference on*, pages 501–506, June 2008. ISSN 1545-0678. doi: 10.1109/ICDCS.Workshops.2008.82. 4.3, 4.3.2, 4.4, 6

[71] A Wells, R. Hammack, G. Veloski, J. Diehl, and B. Strazisar. Monitoring, mitigation, and verification at sequestration sides: Sequre technologies and the challenge for geophysical detection. Technical report, NETL., Pittsburgh, PA, October 2006. 1, 1.3, 6

[72] A. Willsky. A survey of design methods for failure detection in dynamic systems. *Automatica*, 12:601–611, 1976. 2.3.2, 3, 6

[73] Alan S. Willsky and Harold L. Jones. A generalized likelihood ratio approach to state estimation in linear systems subjects to abrupt changes. volume 13, pages 846–853, Nov. 1974. doi: 10.1109/CDC.1974.270554. 3, 3.3.1

[74] R. Ambrosino B. Sinopoli Y. Mo, L. Shi. Network Lifetime Maximization via Sensor Selection. In *Proceedings of the Seventh Asian Control Conference, Hong Kong.*, 2009. 1.2, 4.2, 4.3, 4.3.2, 4.4, 6

[75] L. Yao, W.A. Sethares, and D.C. Kammer. Sensor placement for on-orbit modal identification of large space structure via a genetic algorithm. pages 332–335, Sep 1992. doi: 10.1109/ICSYSE.1992.236889. 1.2

[76] Z. Florencio D. Zhang, C. Zhang. Maximum likelihood sound source localization for multiple directional microphones. *In ICASSP*, 2007. 1, 1.1

[77] I. Ziskind and M. Wax. Maximum likelihood localization of multiple sources by alternating projection. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36(10):1553–1560, Oct 1988. ISSN 0096-3518. doi: 10.1109/29.7543. 1, 1.1

Appendix A

Distribution of Observations

We recall the system model in (2.6) from Chapter 2, written as

$$
\begin{bmatrix} x_{k+1} \\ z_{k+1} \end{bmatrix} = \begin{bmatrix} A_k & B_k\Gamma_k G_k \\ 0 & F_k \end{bmatrix} \begin{bmatrix} x_k \\ z_k \end{bmatrix} + \begin{bmatrix} I & B_k\Gamma_k & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} w_k \\ d_k \\ h_k \end{bmatrix} \tag{A.1}
$$

$$
r_k = \begin{bmatrix} \Lambda_k Q_k C_k & 0 \end{bmatrix} \begin{bmatrix} x_k \\ z_k \end{bmatrix} + \Lambda_k Q_k v_k.
$$

and the *a priori* distribution on the process state

$$
\tilde{x}_0 : N\left[\hat{x}_0, \Sigma_0^x\right]. \tag{A.2}
$$

Additionally, we recall from the discussion in Chapter 2, when the jth source transitions from inactive to active at time $k = K$, the source state is initialized at time $k = K$ according to:

$$
\tilde{z}_K^j : N\left[\hat{z}_{j,K}, \Sigma_{j,K}^z\right], \tag{A.3}
$$

and assumed to be independent of the process state. We also assume the noise terms are distributed as

$$\begin{bmatrix} \tilde{w}_k \\ \tilde{v}_k \\ \tilde{h}_k \\ \tilde{d}_k \end{bmatrix} : N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} W & 0 & 0 & 0 \\ 0 & V & 0 & 0 \\ 0 & 0 & H & 0 \\ 0 & 0 & 0 & D \end{bmatrix} \right). \tag{A.4}$$

Since the system dynamics are linear and the observations, r_k, are linear combinations of the system state, the observation random vector, $\tilde{R}_K$, is written as a linear combination of the process, source, and noise random variables as:

$$\begin{aligned} \tilde{R}_K = {} & \Psi_x(K)\tilde{x}_0 + \left(\sum_{k=0}^{K-1} \Psi_w(k, K)\tilde{w}_k \right) + \begin{bmatrix} \Lambda_0 Q_0 \tilde{v}_0 \\ \vdots \\ \Lambda_K Q_K \tilde{v}_K \end{bmatrix} \\ & + \left(\sum_{k=0}^{K-1} \sum_{j=1}^{J} \Psi_z(j, k, K)\tilde{z}_k + \Psi_h(j, k, K)\tilde{h}_k + \Psi_d(j, k, K)\tilde{d}_k \right) \end{aligned} \tag{A.5}$$

where $\Psi_x(k')$, $\Psi_w(k, k')$, $\Psi_z(j, k, k')$, $\Psi_d(j, k, k')$, and $\Psi_h(j, k, k')$ are the observation matrices from time 0 to k' for the initial process state, the process noise, the initialization source state, source output noise, and the source process noise, respectively. We denote an observation matrix as a matrix which relates the effect of a signal on the observed values from time $k = 0$ to $k = K$, namely R_K. Each block-row of the observation matrices corresponds to the effect of their respective signals on the output at a specific time, $r_k \sqsubseteq R_K$. The following subsection defines and discusses each observation matrix in (A.5) and the final subsection illustrates how the distribution under each hypothesis is determined.

A.1 Transition Matrices

Each observation matrix in (A.5) represents how a respective random variable affects the observations, R_K. This section defines and discusses each transition matrix.

A.1.1 $\Psi_x(k')$

$\Psi_x(k')$ is the observation matrix at time k' corresponding to the process state, x_k. Since the process state evolves according to the process dynamics, the process state at any time $k \geq 0$ is a function of the initial process state. The observation matrix at time k' for the initial system state, x_0, is written as:

$$
\Psi_x(k') = \begin{bmatrix} \Lambda_0 Q_0 C_0 \\ \Lambda_1 Q_1 C_1 A_0 \\ \vdots \\ \Lambda_{k'} Q_{k'} C_{k'} \prod_{j=0}^{k'-1} A_j \end{bmatrix},
\tag{A.6}
$$

We note that $\Psi_x(k')$ is independent of whether sources are active or inactive.

A.1.2 $\Psi_w(k, k')$

$\Psi_w(k, k')$ is the observation matrix at time k' corresponding to the process noise at time k, w_k, and is written as:

$$
\Psi_w(k, k') = \begin{bmatrix} \vdots \\ 0 \\ \Lambda_{k+1} Q_{k+1} C_{k+1} \\ \Lambda_{k+2} Q_{k+2} C_{k+2} A_{k+1} \\ \Lambda_{k+2} Q_{k+2} C_{k+2} \prod_{j=k+1}^{k+2} A_j \\ \vdots \\ \Lambda_{k'} Q_{k'} C_{k'} \prod_{j=k+1}^{k'-1} A_j \end{bmatrix},
\tag{A.7}
$$

121

We also observe that $\Psi_w(k, k')$ is independent of whether sources are active or inactive.

A.1.3 $\Psi_d(k, k')$

$\Psi_d(j, k, k')$ is the observation matrix at time k' for source output noise at time k, d_k, corresponding to the jth source and is written as:

$$\Psi_d(j, k, k') = \begin{bmatrix} \vdots \\ 0 \\ \Lambda_{k+1} Q_{k+1} C_{k+1} B_k \Gamma_k^j \\ \Lambda_{k+2} Q_{k+2} C_{k+2} A_{k+1} B_k \Gamma_k^j \\ \Lambda_{k+2} Q_{k+2} C_{k+2} \left(\prod_{j=k+1}^{k+2} A_j \right) B_k \Gamma_k^j \\ \vdots \\ \Lambda_{k'} Q_{k'} C_{k'} \left(\prod_{j=k+1}^{k'-1} A_j \right) B_k \Gamma_k^j \end{bmatrix} \tag{A.8}$$

Where Γ_k^j is the matrix (of the same dimension as Γ_k) that assumes all sources except the jth source are inactive, and takes the same values as Γ_k for elements corresponding to the jth source. We note that

$$\Gamma_k = \sum_{j=1}^{J} \Gamma_k^j. \tag{A.9}$$

Unlike the previous observation matrices, $\Psi_d(j, k, k')$ is a function of which sources are active.

A.1.4 $\Psi_z(j, k, k')$

$\Psi_z(j, k, k')$ is the observation matrix at time k' corresponding to the jth source state when the jth source becomes active at time k. Since the jth source state is initialized when the jth source becomes active at time k and evolves according to the source dynamics (which are known to be independent of other sources), the jth source state at any time after initialization is a function of the initial source state (plus the source process noise, which

122

will be addressed in the following subsection). We write $\Psi_z(j, k, k')$ as:

$$\Psi_z(j, k, k') = \begin{bmatrix} \vdots \\ 0 \\ \Lambda_{k+1}Q_{k+1}C_{k+1}B_k \left(\Gamma_k^j - \Gamma_{k-1}^j\right) G_k \\ \Lambda_{k+2}Q_{k+2}C_{k+2} \left[A_{k+1}B_k \left(\Gamma_k^j - \Gamma_{k-1}^j\right) G_k + B_{k+1} \left(\Gamma_k^j - \Gamma_{k-1}^j\right) G_{k+1}F_k\right] \\ \Lambda_{k+3}Q_{k+3}C_{k+3} \sum_{n=k}^{k+2} \left[\left(\prod_{i=n+1}^{k+2} A_i\right) B_n \left(\Gamma_k^j - \Gamma_{k-1}^j\right) G_n \left(\prod_{m=k}^{n-1} F_m\right)\right] \\ \vdots \\ \Lambda_{k'}Q_{k'}C_{k'} \sum_{n=k}^{k'-1} \left[\left(\prod_{i=n+1}^{k'-1} A_i\right) B_n \left(\Gamma_k^j - \Gamma_{k-1}^j\right) G_n \left(\prod_{m=k}^{n-1} F_m\right)\right] \end{bmatrix},$$

$$(A.10)$$

where $\Gamma_k^j - \Gamma_{k-1}^j$ is only non-zero when the jth source becomes active time k. Since sources are persistent, once sources become active, they remain active indefinitely and thus each source is initialized at most once. We recall that the sources are independent, and thus

$$\left(\prod_{i=n+1}^{k'-1} A_i\right) B_n \left(\Gamma_k - \Gamma_{k-1}\right) G_k \left(\prod_{m=k}^{n-1} F_m\right) = \sum_{j=1}^{J} \left(\prod_{i=n+1}^{k'-1} A_i\right) B_n \left(\Gamma_k^j - \Gamma_{k-1}^j\right) G_k \left(\prod_{m=k}^{n-1} F_m\right).$$

$$(A.11)$$

A.1.5 $\Psi_h(j, k, k')$

$\Psi_h(j, k, k')$ is the observation matrix at time k' corresponding to the source process noise at time k, h_k, associated with the jth source and is written as:

$$
\Psi_h(j,k,k') =
\begin{bmatrix}
\vdots \\
0 \\
0 \\
\Lambda_{k+2}Q_{k+2}C_{k+2}B_{k+1}\Gamma_k^j G_{k+1} \\
\Lambda_{k+3}Q_{k+3}C_{k+3}\left[A_{k+2}B_{k+1}\Gamma_k^j G_{k+1} + B_{k+2}\Gamma_k^j G_{k+2}F_{k+1}\right] \\
\Lambda_{k+4}Q_{k+4}C_{k+4}\sum_{n=k+1}^{k+3}\left[\left(\prod_{i=n+1}^{k+3} A_i\right)B_j\Gamma_k^j G_n\left(\prod_{m=k}^{n-1} F_m\right)\right] \\
\vdots \\
\Lambda_{k'}Q_{k'}C_{k'}\sum_{n=k+1}^{k'-1}\left[\left(\prod_{i=n+1}^{k'-1} A_i\right)B_n\Gamma_k^j G_n\left(\prod_{m=k}^{n-1} F_m\right)\right]
\end{bmatrix}
\tag{A.12}
$$

Since, the jth source process noise only enters the process state through the corresponding j^{th} source state and the source state is initialized when sources transition from inactive to active, the source process noise prior to initialization has no affect on the observations. Thus, the jth source process noise at time k, h_k, only has an effect on the observations if the corresponding source is active at time k.

A.2 Distribution of R_K assuming active sources

We write the observation random variable, $\tilde{R}_K$ as the sum of a null random variable, $\tilde{R}_K^0$, and an event random variable, $\tilde{R}_K^\Gamma$, which assumes a time propagation of active sources (Γ_k) from $k=0$ to $k=K$. We write the null random variable as

$$
\tilde{R}_K^0 = \Psi_x(K)\tilde{x}_0 + \left(\sum_{k=0}^{K-1}\Psi_w(k,K)\tilde{w}_k\right) + \begin{bmatrix}\Lambda_0 Q_0 \tilde{v}_0 \\ \vdots \\ \Lambda_K Q_K \tilde{v}_K\end{bmatrix},
\tag{A.13}
$$

and the event random variable as

$$\tilde{R}_K^{\Gamma} = \sum_{k=0}^{K-1} \sum_{j=1}^{J} \left(\Psi_z(j,k,K)\tilde{z}_k + \Psi_h(j,k,K)\tilde{h}_k + \Psi_d(j,k,K)\tilde{d}_k \right) \tag{A.14}$$

where $\tilde{R}_K = \tilde{R}_K^0 + \tilde{R}_K^{\Gamma}$. The null random variable represents the observations when no sources are active, while the event random variable accounts for active sources. The null random variable and event random variable are independent since each is a linear combination of a mutually exclusive set of random variables, each known to be independent of the rest. Since all the underlying random variables are independent normal random variables, the null hypothesis is distributed as

$$\tilde{R}_K^0 : N[m_0(K), S_0(K)], \tag{A.15}$$

where

$$m_0(K) = \Psi_x(K)\hat{x}_0$$

$$S_0(K) = \Psi_x(K)\Sigma_0^x \left(\Psi_x(K)\right)^T + \left[\sum_{k=0}^{K-1} \Psi_w(k,K)W\left(\Psi_w(k,K)\right)^T\right] + \begin{bmatrix} \Lambda_0 Q_0 V\left(\Lambda_0 Q_0\right)^T \\ \vdots \\ \Lambda_K Q_K V\left(\Lambda_K Q_K\right)^T \end{bmatrix}.$$
$$\tag{A.16}$$

Similarly, the event hypotheses are distributed as

$$\tilde{R}_K^{\Gamma} : N\left[\sum_{j=1}^{J}\sum_{k=0}^{K}(b_k^j - b_{k-1}^j)m_{j,k}(K), \sum_{j=1}^{J}\sum_{k=0}^{K}(b_k^j - b_{k-1}^j)S_{j,k}(K)\right], \tag{A.17}$$

125

where

$$m_{j,k}(K) = \Psi_z(j, k, K)\hat{z}_k$$

$$S_{j,k}(K) = \Psi_z(j, k, K)\Sigma_k^z (\Psi_z(j, k, K))^T$$

$$+ \sum_{k'=k}^{K} \left(\Psi_h(j, k', K)H (\Psi_h(j, k', K))^T + \Psi_d(j, k', K)D (\Psi_d(j, k', K))^T \right), \quad \text{(A.18)}$$

and the mean and covariance of the event random variable are written in terms of when the jth source becomes active $(b_k^j - b_{k-1}^j = 1)$.

Appendix B

Multiple source detection for constant sources

This appendix contains the theoretical results for Multiple Source Detection (MSD) for constant sources. The following subsection provides proof of the optimality of testing only a subset of the potential events instead of all possible events for MSD in the scalar observation case. The final subsection provides a proof Wald's test [65].

B.1 Scalar observation MSD proof

Suppose there are J independent noisy signal sources with positive means $m_1, \ldots, m_J$ and variances $\Sigma_1, \ldots, \Sigma_J$, respectively. Each source can be active or inactive. The active sources are indicated by the values of a binary vector $b \in \{0,1\}^J$, where $b_j = 1$ indicates source j is active. In the following $\mathbf{1}$ and $\mathbf{0}$ denote the binary vectors of all ones and all zeros, respectively, and e_j denotes the elementary binary vector with a single 1 in the jth component.

A vector of received signals $y \in R^N$ is the sum of the signals from the active sources plus an additional independent zero-mean, identity covariance noise signal. Therefore, the observation vector y is a random variable of dimension N with one of 2^J possible distributions, $N[\mu_b, \Sigma_b]$, where $N[\mu, \Sigma]$ denotes the Gaussian distribution with mean μ and covariance Σ. The mean μ_b is given by $\mu_b = \sum_{j=1}^{J} b_j m_j$ and the covariance Σ_b is given by $\Sigma_b = I + \sum_{j=1}^{J} b_j \Sigma_j$. We denote the normal probability density function (pdf) for a

given active source vector b by $f_b(y)$.

Let H_b denote the hypothesis that the active source vector is b and let $l_b(y)$ denote the log-likelihood ratio between the hypotheses H_b and H_0:

$$
\begin{aligned}
l_b(y) &= ln\left(\frac{f_b(y)}{f_0(y)}\right) \\
&= ln(f_b(y)) - ln(f_0(y)) \\
&= -\frac{1}{2}(y - \mu_b)^T \Sigma_b^{-1}(y - \mu_b) - \frac{1}{2}\ln\det(\Sigma_b) + \frac{1}{2}y^T y \\
&= \frac{1}{2}y^T\left(I - \Sigma_b^{-1}\right)y - \mu_b^T \Sigma_b^{-1}y - \frac{1}{2}\mu_b^T \Sigma_b^{-1}\mu_b - \frac{1}{2}\ln\det(\Sigma_b)
\end{aligned}
\tag{B.1}
$$

We wish to determine if the null hypothesis, H_0, is most likely and also identify an $H_{b'}$ for $b' \neq 0$ that is in the set of next most likely hypotheses. Given an observation y, the null hypothesis is most likely when $l_b(y) < 0$ for all $b \neq 0$, and an alternative hypothesis $H_{b'}$ for some $b' \neq 0$ is among the next most likely hypotheses when $l_{b'}(y) \geq l_b(y)$ for all $b \notin \{0, b'\}$.

When only a single observation is received, we denote the distribution on the observation as $N[\mu_b, \varsigma_b]$, where $N[\mu, \varsigma]$ denotes the Gaussian distribution with mean μ and variance ς[1]. The mean μ_b can is given by $\mu_b = b^T m$ with $m \triangleq [m_1, \ldots, m_J]$ and the variance ς_b is given by $\varsigma_b = 1 + b^T \sigma$ with $\sigma \triangleq [\sigma_1, \ldots, \sigma_J]$. For the scalar observation case, the log-likelihood ratio is written as:

---

[1]For notational convenience in this subsection we denote the variance as a parameter that is not squared, rather than using the standard notation σ^2 for the variance.

$$l_b(y) = ln\left(\frac{f_b(y)}{f_0(y)}\right)$$

$$= ln(f_b(y)) - ln(f_0(y))$$

$$= -\frac{1}{2}\frac{(y - \mu_b)^2}{\varsigma_b} - \frac{1}{2}ln(\varsigma_b) + \frac{1}{2}y^2 \qquad \text{(B.2)}$$

$$= \frac{1}{2}\left[\frac{\varsigma_b - 1}{\varsigma_b}y^2 + \frac{2\mu_b}{\varsigma_b}y - \frac{\mu_b^2}{\varsigma_b} - ln(\varsigma_b)\right]$$

$$= \frac{1}{2}\left[\sum_{j=1}^{J}b_j\left(\frac{\sigma_j}{\varsigma_b}y^2 + \frac{2m_j}{\varsigma_b}y\right)\right] - \frac{1}{2}\frac{\mu_b^2}{\varsigma_b} - \frac{1}{2}ln(\varsigma_b)$$

Lemma 1. Given $b \in \{0,1\}^J$ and $\sigma \geq \mathbf{0}$,

$$\sum_{j=1}^{J}(1 + b_j\sigma_j)ln(1 + b_j\sigma_j) \leq (1 + b^T\sigma)ln(1 + b^T\sigma) \qquad \text{(B.3)}$$

Proof. Inequality (B.3) is true if and only if

$$f(\sigma) = \sum_{j=1}^{J}(1 + b_j\sigma_j)ln(1 + b_j\sigma_j) - (1 + b^T\sigma)ln(1 + b^T\sigma) \leq 0. \qquad \text{(B.4)}$$

Note that $f(0) = 0$ and for each $j \in \{1, \ldots, J\}$,

$$\frac{\partial f(\sigma)}{\partial \sigma_j} = ln(1 + b_j\sigma_j) + 1 - ln(1 + b^T\sigma) - 1 = ln(1 + b_j\sigma_j) - ln(1 + b^T\sigma)$$

$$\leq 0 \quad \text{for} \quad \sigma \geq 0. \qquad \text{(B.5)}$$

Therefore, $f(\sigma) \leq 0$ for $\sigma \geq 0$. □

Proposition 1. If for some $j' \in \{1, \ldots, J\}$, $l_{e_{j'}}(y) \geq l_{e_j}(y)$ for all $j \neq j'$ and $m \geq 0$, then no $H_{b\neq 0}$ is more likely than $H_{e_{j'}}$.

Proof. From lemma 1 it follows for $b \in \{0,1\}^J$ and $\sigma > \mathbf{0}$ that

$$- (1 + b^T\sigma)ln(1 + b^T\sigma) + \sum_{j=1}^{J}(1 + b_j\sigma_j)ln(1 + b_j\sigma_j) \leq 0$$

$$\leftrightarrow -\varsigma_b ln(\varsigma_b) + \sum_{j=1}^{J} b_j\varsigma_{e_j} ln(\varsigma_{e_j}) \leq 0 \qquad (\text{B.6})$$

$$\leftrightarrow -\frac{1}{2}ln(\varsigma_b) + \frac{1}{2}\left[\sum_{j=1}^{J} b_j \frac{\varsigma_{e_j}}{\varsigma_b} ln(\varsigma_{e_j})\right] \leq 0$$

Since $l_{e_j} \le l_{e_{j'}} \le 0$ for all $j \ne j'$ and $m \ge \mathbf{0}$, it follows that

$$\left[\sum_{j=1}^{J} b_j \frac{\varsigma_{e_j}}{\varsigma_b} (l_{e_j}(y) - l_{e_{j'}}(y)) \right] - \left[\sum_{i=1}^{J-1} \sum_{k=i+1}^{J} b_i b_k \frac{m_i m_k}{\varsigma_b} \right] - \frac{1}{2} ln(\varsigma_b) + \frac{1}{2} \left[\sum_{j=1}^{J} b_j \frac{\varsigma_{e_j}}{\varsigma_b} ln(\varsigma_{e_j}) \right] + \frac{b^T b - 1}{\varsigma_b} l_{e_{j'}}(y) \le$$

$$\leftrightarrow \left[\sum_{j=1}^{J} b_j \frac{\varsigma_{e_j}}{\varsigma_b} l_{e_j}(y) \right] - \left[\sum_{i=1}^{J-1} \sum_{k=i+1}^{J} b_i b_k \frac{m_i m_k}{\varsigma_b} \right] - \frac{1}{2} ln(\varsigma_b) + \frac{1}{2} \left[\sum_{j=1}^{J} b_j \frac{\varsigma_{e_j}}{\varsigma_b} ln(\varsigma_{e_j}) \right]$$
$$- \left(\frac{1 - b^T b + \sum_{j=1}^{J} b_j \varsigma_{e_j}}{\varsigma_b} \right) l_{e_{j'}}(y) \le 0$$

$$\leftrightarrow \left[\sum_{j=1}^{J} b_j \frac{\varsigma_{e_j}}{\varsigma_b} l_{e_j}(y) \right] - \left[\sum_{i=1}^{J-1} \sum_{k=i+1}^{J} b_i b_k \frac{m_i m_k}{\varsigma_b} \right] - \frac{1}{2} ln(\varsigma_b) + \frac{1}{2} \left[\sum_{j=1}^{J} b_j \frac{\varsigma_{e_j}}{\varsigma_b} ln(\varsigma_{e_j}) \right]$$
$$- \left(\frac{1 + \sum_{j=1}^{J} b_j \sigma_j}{\varsigma_b} \right) l_{e_{j'}}(y) \le 0$$

$$\leftrightarrow \left[\sum_{j=1}^{J} b_j \frac{\varsigma_{e_j}}{\varsigma_b} l_{e_j}(y) \right] - \left[\sum_{i=1}^{J-1} \sum_{k=i+1}^{J} b_i b_k \frac{m_i m_k}{\varsigma_b} \right] - \frac{1}{2} ln(\varsigma_b) + \frac{1}{2} \left[\sum_{j=1}^{J} b_j \frac{\varsigma_{e_j}}{\varsigma_b} ln(\varsigma_{e_j}) \right] - l_{e_{j'}}(y) \le 0$$

$$\leftrightarrow \frac{1}{2} \left[\sum_{j=1}^{J} b_j \frac{\varsigma_{e_j}}{\varsigma_b} \left(\frac{\sigma_j}{\varsigma_{e_j}} y^2 + \frac{2m_j}{\varsigma_{e_j}} y - \frac{m_j^2}{\varsigma_{e_j}} - ln(\varsigma_{e_j}) \right) \right] - \left[\sum_{i=1}^{J-1} \sum_{k=i+1}^{J} b_i b_k \frac{m_i m_k}{\varsigma_b} \right]$$
$$- \frac{1}{2} ln(\varsigma_b) + \frac{1}{2} \left[\sum_{j=1}^{J} b_j \frac{\varsigma_{e_j}}{\varsigma_b} ln(\varsigma_{e_j}) \right] - l_{e_{j'}}(y) \le 0$$

$$\leftrightarrow \frac{1}{2} \left[\sum_{j=1}^{J} b_j \left(\frac{\sigma_j}{\varsigma_b} y^2 + \frac{2m_j}{\varsigma_b} y - \frac{m_j^2}{\varsigma_b} - \frac{\varsigma_{e_j}}{\varsigma_b} ln(\varsigma_{e_j}) \right) \right] - \left[\sum_{i=1}^{J-1} \sum_{k=i+1}^{J} b_i b_k \frac{m_i m_k}{\varsigma_b} \right]$$
$$- \frac{1}{2} ln(\varsigma_b) + \frac{1}{2} \left[\sum_{j=1}^{J} b_j \frac{\varsigma_{e_j}}{\varsigma_b} ln(\varsigma_{e_j}) \right] - l_{e_{j'}}(y) \le 0$$

$$\leftrightarrow \frac{1}{2} \left[\sum_{j=1}^{J} b_j \left(\frac{\sigma_j}{\varsigma_b} y^2 + \frac{2m_j}{\varsigma_b} y - \frac{m_j^2}{\varsigma_b} \right) \right] - \left[\sum_{i=1}^{J-1} \sum_{k=i+1}^{J} b_i b_k \frac{m_i m_k}{\varsigma_b} \right] - \frac{1}{2} ln(\varsigma_b) - l_{e_{j'}}(y) \le 0$$

$$\leftrightarrow \frac{1}{2} \left[\sum_{j=1}^{J} b_j \left(\frac{\sigma_j}{\varsigma_b} y^2 + \frac{2m_j}{\varsigma_b} y \right) \right] - \frac{1}{2} \frac{\mu_b^2}{\varsigma_b} - \frac{1}{2} ln(\varsigma_b) - l_{e_{j'}}(y) \le 0$$

$$\leftrightarrow l_b(y) - l_{e_{j'}}(y) \le 0 \leftrightarrow l_b(y) \le l_{e_{j'}}(y)$$

(B.7)

Thus, we conclude that when $0 \ge l_{e_{j'}}(y) \ge l_{e_j}(y)$ for all $e_j \ne e_{j'}$, then no H_b is more likely

than $H_{e_{j'}}$ for $b \notin \{0, e_{j'}\}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Corollary 1. If for some $j' \in \{1, \ldots, J\}, 0 > l_{e_{j'}}(y) \geq l_{e_j}(y)$ for all $j \neq j'$ and $m \geq 0$, then H_0 is more likely than any other H_b for $b \neq 0$.

The results of Corollary 1 state that for the scalar observation case, when the all the means are positive and the null hypothesis is more likely than all the elementary hypotheses, then any event hypothesis is no more likely than the most likely elementary event hypothesis.

B.2 Test bounding error types

This subsection provides proofs for Wald's test, and closely follow those in [65]. Let there exists two hypotheses:

$$
\begin{aligned}
H_A &: \tilde{x} : f_A(x) \\
H_B &: \tilde{x} : f_B(x)
\end{aligned}, \tag{B.8}
$$

where $f_i(x)$ is the probability density function for the random variable $\tilde{x}$ under hypothesis H_i. Given a test, $\phi(x) \in \{H_A, H_B\}$ which classifies x, the probability of error for hypothesis H_i is

$$
\int_{\phi(x)=H_j} f_i(x)dx, \tag{B.9}
$$

and is denoted as $P[\phi(x) = H_j | H_i]$. The following lemmas prove how the likelihood ratio test bounds the probability of error under both hypotheses.

Lemma 2. Given $\epsilon_1, \epsilon_2 \in [0, 1]$, and $\phi(x) = H_B$ when

$$
\frac{f_A(x)}{f_B(x)} \leq \frac{\epsilon_1}{1 - \epsilon_2} \tag{B.10}
$$

then

$$P\left[\phi(x) = H_A | H_B\right] \leq \epsilon_1 \quad \text{and} \quad P\left[\phi(x) = H_B | H_A\right] \leq \epsilon_2 \tag{B.11}$$

.

Proof. Let $\phi(x) = H_B$ when

$$\frac{f_A(x)}{f_B(x)} \leq \eta \leftrightarrow f_A(x) \leq \eta f_B(x), \tag{B.12}$$

then it is true that

$$\int_{\phi(x)=H_B} f_A(x)dx \leq \eta \int_{\phi(x)=H_B} f_B(x)dx$$
$$\leftrightarrow \int_{\phi(x)=H_B} f_A(x)dx \leq \eta \left(1 - \int_{\phi(x)=H_A} f_B(x)dx\right). \tag{B.13}$$

Thus, for

$$P\left[\phi(x) = H_A | H_B\right] \leq \epsilon_1 \quad \text{and} \quad P\left[\phi(x) = H_B | H_A\right] \leq \epsilon_2, \tag{B.14}$$

it must also be true that

$$\epsilon_1 \leq \eta(1 - \epsilon_2) \leftrightarrow \frac{\epsilon_1}{1 - \epsilon_2} \leq \eta, \tag{B.15}$$

which implies

$$\frac{f_A(x)}{f_B(x)} \leq \frac{\epsilon_1}{1 - \epsilon_2}. \tag{B.16}$$

□

Lemma 3. Given $\epsilon_1, \epsilon_2 \in [0, 1]$, and $\phi(x) = H_A$ when

$$\frac{f_A(x)}{f_B(x)} \geq \frac{1 - \epsilon_1}{\epsilon_2} \tag{B.17}$$

then

$$P\left[\phi(x) = H_A | H_B\right] \leq \epsilon_1 \quad \text{and} \quad P\left[\phi(x) = H_B | H_A\right] \leq \epsilon_2 \tag{B.18}$$

.

Proof. Let $\phi(x) = H_B$ when

$$\frac{f_A(x)}{f_B(x)} \geq \eta \leftrightarrow f_A(x) \geq \eta f_B(x), \tag{B.19}$$

then it is true that

$$\begin{aligned}
\int_{\phi(x)=H_A} f_A(x) dx &\geq \eta \int_{\phi(x)=H_A} f_B(x) dx \\
\leftrightarrow \left(1 - \int_{\phi(x)=H_B} f_A(x) dx\right) &\geq \eta \int_{\phi(x)=H_A} f_B(x) dx
\end{aligned}. \tag{B.20}$$

Thus, for

$$P\left[\phi(x) = H_A | H_B\right] \geq \epsilon_1 \quad \text{and} \quad P\left[\phi(x) = H_B | H_A\right] \geq \epsilon_2, \tag{B.21}$$

it must also be true that

$$1 - \epsilon_1 \geq \eta \epsilon_2 \leftrightarrow \frac{1 - \epsilon_1}{\epsilon_2} \geq \eta, \tag{B.22}$$

which implies

$$\frac{f_A(x)}{f_B(x)} \geq \frac{1 - \epsilon_1}{\epsilon_2}. \tag{B.23}$$

$\square$

Appendix C

Diffusion example

This appendix describes the diffusion example used throughout this dissertation to evaluate performance. We assume there exists a monitoring application that relates potential sources in a 2-D monitoring area to sensor observations through process dynamics according to

$$
\begin{bmatrix} x_{k+1} \\ z_{k+1} \end{bmatrix} = \begin{bmatrix} A & B\Gamma_k \\ 0 & I \end{bmatrix} \begin{bmatrix} x_k \\ z_k \end{bmatrix} + \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} w_k \\ d_k \end{bmatrix},
$$

$$
y_k = \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} x_k \\ z_k \end{bmatrix} + v_k
$$

(C.1)

where $x_k \in R^{N^2}$ is the process state, $y_k \in R^{N^2}$ is the sensor observation vector at time k, $z_k \in R^{N^2}$ is the source state, $A \in R^{N^2 \times N^2}$ represents the process dynamics, $B \in R^{N^2 \times N^2}$ relates the source state to the process state, and Q_k is the dynamic sensor selection matrix at time k. w_k, d_k, and v_k are the process noise, source process noise, and observation noise respectively and distributed as

$$
\begin{bmatrix} \tilde{w}_k \\ \tilde{v}_k \\ \tilde{d}_k \end{bmatrix} : N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 10^{-4}I & 0 & 0 \\ 0 & 2(10^{-3})I & 0 \\ 0 & 0 & 10^2 I \end{bmatrix} \right). \tag{C.2}
$$

The initial distribution on the process state and the source state (for a transition from inactive to active at any time $k = K$) is written as

$$
\begin{aligned}
\tilde{x}_0 &: N\left(\mathbf{0}, 10^{-3}I\right) \\
\tilde{z}_K^j &: N\left(100, 2500\right)
\end{aligned}
. \tag{C.3}
$$

Each element of the process state, $x_k(N(i-1)+j)$, sensor observation vector, $r_k(N(i-1)+j)$, and source state, $z_k(N(i-1)+j)$, correspond to a specific spatial location, denoted by $i, j \in \{1, \ldots, N\}$. Fig. C.1 illustrates the relation between spatial location and the elements of each vector.

In Fig. C.1, the point when $i = j = 1$ corresponds to the first element of each vector, x_k, y_k, z_k. In this example, a sensor is located at each potential source location, since each point in Fig. C.1 not only represents a potential source (as denoted by $z(n)$) but also a potential sensor observation (as denoted by $y(n)$).

In this example, the dynamics are defined as

$$
\begin{aligned}
A &= e^{300 A_c} \\
B &= 10^{-6} \int_0^{300} e^{A_c \tau} \partial \tau
\end{aligned}
\tag{C.4}
$$

where $A_c \in R^{N^2 \times N^2}$ represents the continuous time dynamics of $\dot{x}_t = A_c x_t$ and is defined

136

h
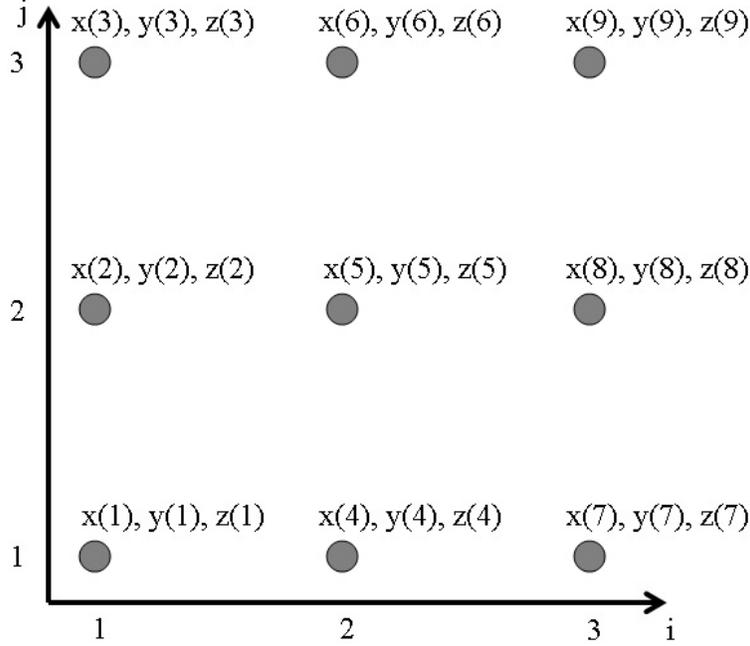
Figure C.1: Diffusion system for $N = 3$.

for any $i, j \in \{1, \ldots, N\}$ such that

$$
\dot{x}(N(i-1)+j) = \begin{cases}
\begin{aligned}
&0.1x(N(i-1)+j-1) + 0.1x(N(i-2)+j) \\
&- 0.4x(N(i-1)+j) \\
&+ 0.1x(N(i)+j) + 0.1x(N(i-1)+j+1)
\end{aligned} & \text{if } 2 \leq i, j \leq N-1 \\[2em]
\begin{aligned}
&0.1x(N(i-1)+j-1) - 0.3x(N(i-1)i+j) \\
&+ 0.1x(N(i)+j) + 0.1x(N(i-1)+j+1)
\end{aligned} & \text{if } 2 \leq j \leq N-1, i = 1 \\[2em]
\begin{aligned}
&0.1x(N(i-1)+j-1) + 0.1x(N(i-2)+j) \\
&- 0.3x(N(i-1)i+j) + 0.1x(N(i-1)+j+1)
\end{aligned} & \text{if } 2 \leq j \leq N-1, i = N \\[2em]
\begin{aligned}
&0.1x(N(i-2)+j) - 0.3x(N(i-1)+j) \\
&+ 0.1x(N(i)+j) + 0.1x(N(i-1)+j+1)
\end{aligned} & \text{if } 2 \leq i \leq N-1, j = 1 \\[2em]
\begin{aligned}
&0.1x(N(i-1)+j-1) + 0.1x(N(i-2)+j) \\
&- 0.3x(N(i-1)+j) + 0.1x(N(i)+j)
\end{aligned} & \text{if } 2 \leq i \leq N-1, j = N \\[2em]
\begin{aligned}
&- 0.2x(N(i-1)+j) \\
&+ 0.1x(N(i)+j) + 0.1x(N(i-1)+j+1)
\end{aligned} & \text{if } i = 1, j = 1 \\[2em]
\begin{aligned}
&0.1x(N(i-2)+j) - 0.2x(N(i-1)+j) \\
&+ 0.1x(N(i-1)+j+1)
\end{aligned} & \text{if } i = N, j = 1 \\[2em]
0.1x(N(i-1)+j-1) - 0.2x(N(i-1)+j) & \text{if } i = 1, j = N
\end{cases}
$$

137

The property in (C.5) ensures the interior elements ($2 \leq i, j \leq N - 1$) are an average of the 4 surrounding elements. In Fig. C.1, element 5 is the only interior element, and is a weighted sum of elements 2, 4, 5, 6, and 8. All the other conditions represent edge elements, where we assume that any element on the exterior is the same value as its closest neighbor.

Appendix D

Dynamic Sensor Selection

This appendix contains the theoretical results for the GLRDSS.

Lemma 4. Given a binary matrix $Q \in \{0,1\}^{M \times N}$ such that $QQ^T = I_M$ and $q = diag(Q^T Q) \in \{0, 1, \}^N$, and two square symmetric matrices of dimension N, $A, B \in R^{N \times N}$, then

$$Tr(QAQ^T QBQ^T) = q^T (A \circ B) q, \tag{D.1}$$

where $A \circ B$ is the Hadamard (element-wise) product of A and B.

Proof. Assuming that $A_{i,j}$ is the ith row and the jth column element of A and q_i is the ith

element of q, then the left hand side of D.1 is written as

$$Tr(QAQ^TQBQ^T)$$

$$\leftrightarrow Tr(AQ^TQBQ^TQ)$$

$$\leftrightarrow Tr \begin{bmatrix} q_1 \left(q_1 A_{1,1} B_{1,1} + \cdots + q_n A_{1,n} B_{1,n} \right) & x & & x \\ & x & \ddots & & x \\ & x & & x & q_n \left(q_1 A_{n,1} B_{n,1} + \cdots + q_n A_{n,n} B_{n,n} \right) \end{bmatrix}$$

$$\leftrightarrow \sum_{i=1}^{N} \sum_{j=1}^{N} q_i q_j A_{i,j} B_{i,j}$$

$$\leftrightarrow q^T \left(A \circ B \right) q$$

$$(D.2)$$

$\square$

Lemma 5. Given a binary matrix $Q \in \{0,1\}^{M \times N}$ such that $QQ^T = I_M$ and $q = diag(Q^TQ) \in \{0,1,\}^N$, a vector $m_1 \in R^N$, a symmetric positive definite matrix $\Sigma_1 \in R^{N \times N}$, and a vector of random variables, x, of dimension N and distributed as:

$$x : \mathcal{N}[Qm_0, Q\Sigma_0 Q^T], \qquad (D.3)$$

then the expected value of the function:

$$f(x) = \frac{1}{2} (x - Qm_0)^T Q\Sigma_0^{-1} Q^T (x - Qm_0) - \frac{1}{2} (x - Qm_1)^T Q\Sigma_1^{-1} Q^T (x - Qm_1), \quad (D.4)$$

is

$$E[f(x)] = \frac{1}{2} \left[M - q^T \left(\Sigma_1^{-1} \circ (m_0 - m_1)(m_0 - m_1)^T \right) q - q^T \left(\Sigma_0 \circ \Sigma_1^{-1} \right) q \right]. \qquad (D.5)$$

Proof. Applying the results in [43], the expected value of $f(x)$ is written as:

$$E[f(x)] = \frac{1}{2} \left[Tr(Q^T Q) - (m_0 - m_1)^T Q^T Q \Sigma_1^{-1} Q^T Q (m_0 - m_1) - Tr\left(Q^T \Sigma_0 Q^T Q \Sigma_1^{-1} Q^T\right) \right].$$

(D.6)

By applying the result of lemma 4, it holds that:

$$E[f(x)] = \frac{1}{2} \left[M - q^T \left(\Sigma_1^{-1} \circ (m_0 - m_1)(m_0 - m_1)^T \right) q - q^T \left(\Sigma_0 \circ \Sigma_1^{-1} \right) q \right].$$

(D.7)

$\square$

Lemma 6. Given a positive definite matrix A of dimension M, then

$$\ln \det A \leq Tr(A) - M.$$

(D.8)

Proof. Assuming that λ_m, is an eigenvalue of A for $m \in \{1, \ldots, M\}$, then

$$\ln \det A = \ln \prod_{m=1}^{M} \lambda_m = \sum_{m=1}^{M} \ln \lambda_m \leq \sum_{m=1}^{M} (\lambda_m - 1) = Tr(A) - M.$$

(D.9)

$\square$

Lemma 7. Given a binary matrix $Q \in \{0,1\}^{M \times N}$ such that $QQ^T = I_M$ and $q = diag(Q^T Q) \in \{0, 1, \}^N$, and 2 symmetric positive definite matrices $A, B \in R^{N \times N}$, then:

$$\ln \det \left(QAQ^T QBQ^T \right) \leq q^T \left(A \circ B \right) q - M.$$

(D.10)

Proof. Applying the results of Lemma 6 it is true that

$$\ln \det \left(QAQ^T QBQ \right) \leq Tr \left(QAQ^T QBQ^T \right) - M.$$
(D.11)

By applying the results of Lemma 4, it is true that

$$\ln \det \left(QAQ^T QBQ \right) \leq q^T \left(A \circ B \right) q - M.$$
(D.12)

□

Lemma 8. Given two functions:

$$\begin{aligned} f(q) &= -q^T A q + c \\ g(q) &= b^T q + d \end{aligned},$$
(D.13)

where A is symmetric positive definite matrix, c is a positive constant, $\lambda = \sqrt{\frac{c}{\mathbf{1}^T A \mathbf{1}}}$, $b = -2\sqrt{\frac{c}{\mathbf{1}^T A \mathbf{1}}} A \mathbf{1}$, and $d = 2c$, then the following properties are true:

$$\begin{aligned} f(\lambda \mathbf{1}) &= 0 \\ g(\lambda \mathbf{1}) &= 0 \\ \frac{d}{\partial q} g(q)|_{q=\lambda \mathbf{1}} &= \frac{d}{\partial q} f(q)|_{q=\lambda \mathbf{1}} \end{aligned}$$
(D.14)

Proof. The first property:

$$f\left(\sqrt{\frac{c}{\mathbf{1}^T A \mathbf{1}}} \mathbf{1} \right) = - \left(\frac{c}{\mathbf{1}^T A \mathbf{1}} \right) \mathbf{1}^T A \mathbf{1} + c = 0.$$
(D.15)

The second property:

$$g\left(\sqrt{\frac{c}{\mathbf{1}^T A \mathbf{1}}} \mathbf{1} \right) = -2 \left(\frac{c}{\mathbf{1}^T A \mathbf{1}} \right) \mathbf{1}^T A \mathbf{1} + 2c = 0.$$
(D.16)

142

The derivative of $f(q)$ with respect to q at the point where $q = \sqrt{\frac{c}{\mathbf{1}^T A\mathbf{1}}}$:

$$\frac{d}{dq}f(q)\big|_{q=\sqrt{\frac{c}{\mathbf{1}^T A\mathbf{1}}}\mathbf{1}} = -2Aq\big|_{q=\sqrt{\frac{c}{\mathbf{1}^T A\mathbf{1}}}\mathbf{1}} = -2\sqrt{\frac{c}{\mathbf{1}^T A\mathbf{1}}}A\mathbf{1} = \frac{d}{dq}g(q)\big|_{q=\sqrt{\frac{c}{\mathbf{1}^T A\mathbf{1}}}\mathbf{1}}. \qquad \text{(D.17)}$$

$\square$

Appendix E

Test Bed Implementation

This appendix presents the simulation parameters for the results obtain through the test bed evaluation, where all the advection-diffusion examples in this thesis assume a spatial discretization of $\Delta = 50m$ and the temporal sampling rate of 10 minutes. All source strengths (regardless of position) are scalar constants and are taken from from the Gaussian distribution with an expected value of 200 parts-per-million (PPM) per second with a variance of 1000 PPM per second.

E.1   Test for synchronous sources

In this subsection, all the sources are considered to start at the same time $t = 3$ hours, and are constant. We consider 4 scenarios:

1. Two distributed sources in low wind (Figs. E.1 and E.2).

2. Two distributed sources in high wind (Figs. E.3 and E.4).

3. Two clustered sources in low wind (Figs. E.5 and E.6).

4. Two clustered sources in high wind (Figs. E.7 and E.8).

Where clustered sources are located at positions (.15 Km, .25 Km) and (.2 Km, .2 Km), while distributed sources are located at positions (.15 Km, .25 Km) and (.30 Km, .10 Km). We assume the wind direction is the vector (1,1), and the magnitude for low wind is $2.8\frac{Km}{hr}$
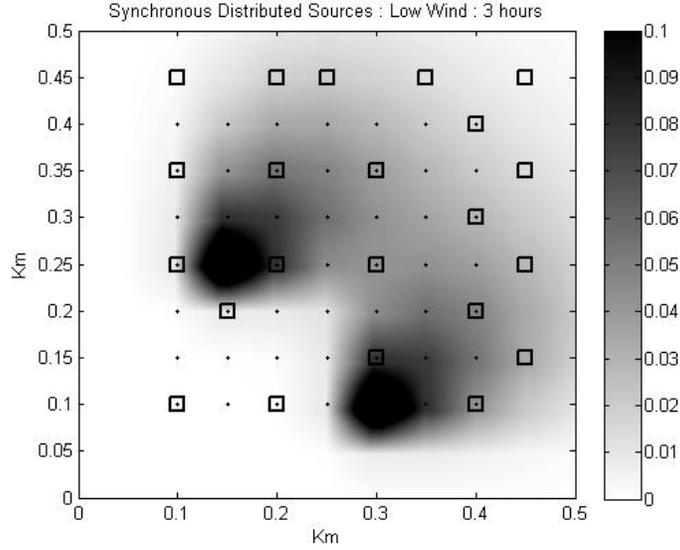
Figure E.1: %CO_2 at t = 3 hours.

and for high wind is $8.5\frac{Km}{hr}$. The following figures illustrate the CO_2 concentration vs. space and time for sources with a strength of 200 PPM per second for each scenario.

E.2 Tests for asynchronous Sources

In this subsection, sources are assumed to start at different times. The first source begins at time $t = 3$ hours, and the second at time $t = 6$ hours. Both sources are considered constant once becoming active. We again consider 4 scenarios:

1. Two distributed sources in low wind (Figs. E.9 and E.10).

2. Two distributed sources in high wind (Figs. E.11 and E.12).

3. Two clustered sources in low wind (Figs. E.13 and E.14).

4. Two clustered sources in high wind (Figs. E.15 and E.16).

Where clustered sources are located at positions (.15 Km, .25 Km) and (.2 Km, .2 Km), while distributed sources are located at positions (.15 Km, .25 Km) and (.30 Km, .10 Km). The following figures illustrate the CO_2 concentration vs. space and time for sources with strength of 200 PPM per second for each scenario.

146
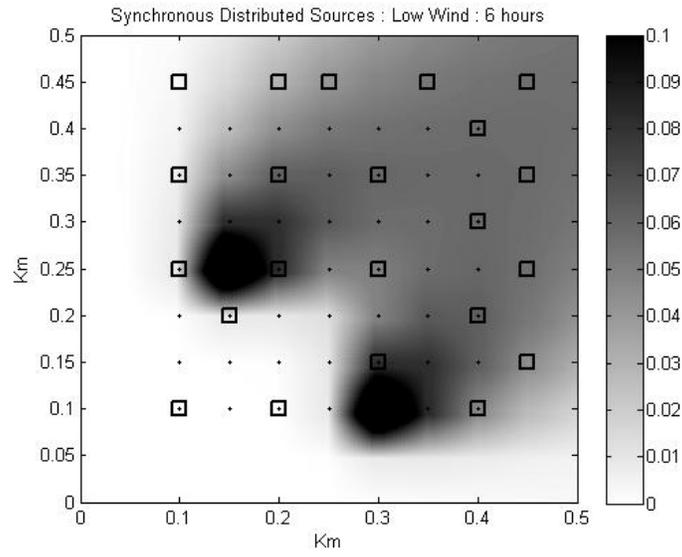
Figure E.2: %CO_2 at t = 6 hours.
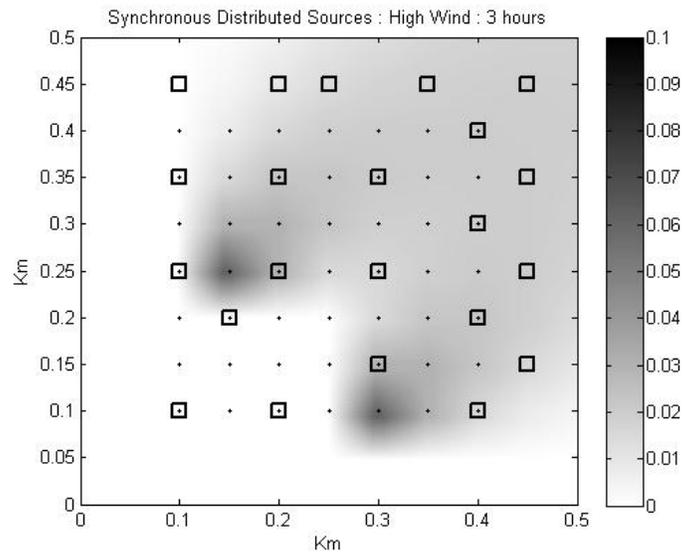
Figure E.3: %CO_2 at t = 3 hours.
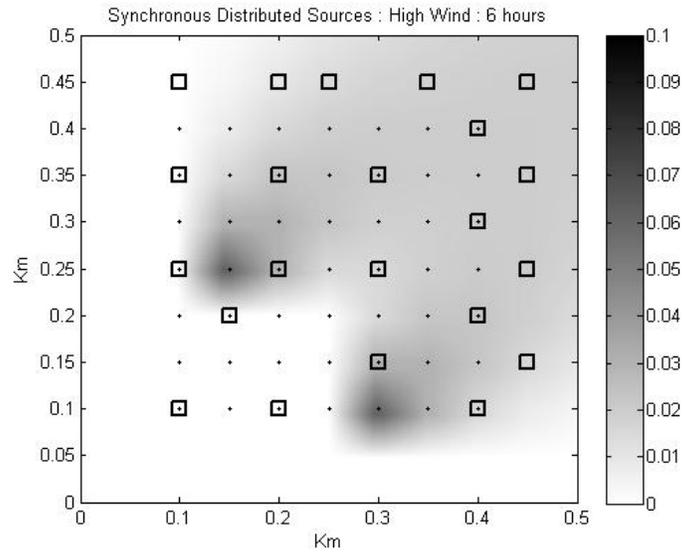
Figure E.4: %CO_2 at t = 6 hours.
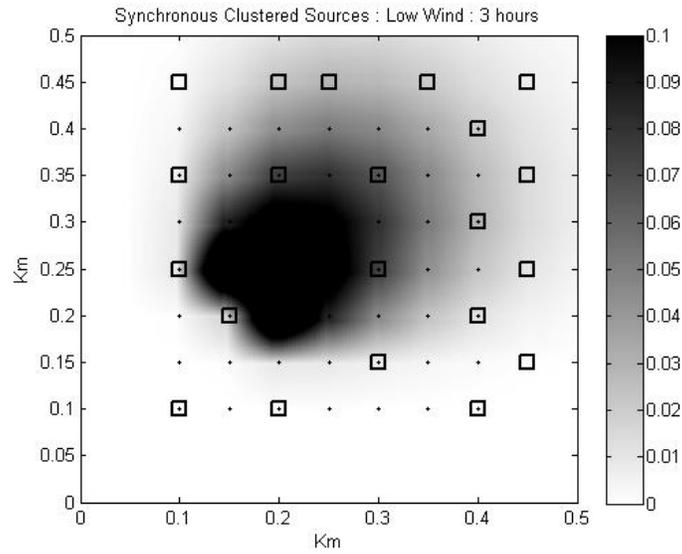
Figure E.5: %CO_2 at t = 3 hours.
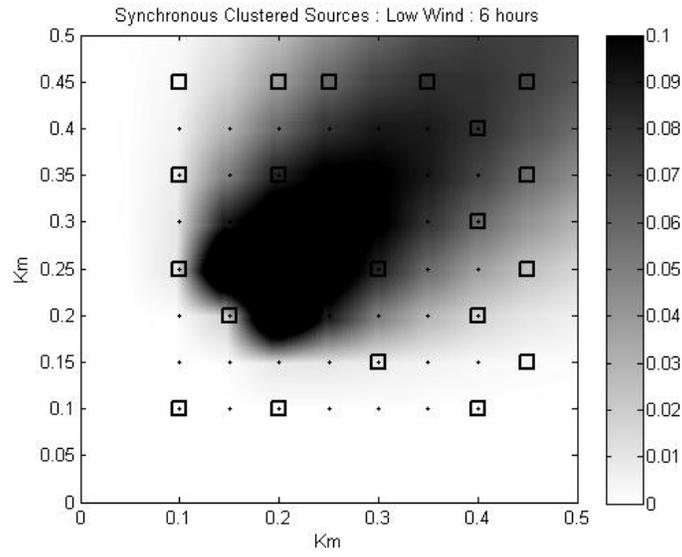
Figure E.6: %CO_2 at t = 6 hours.
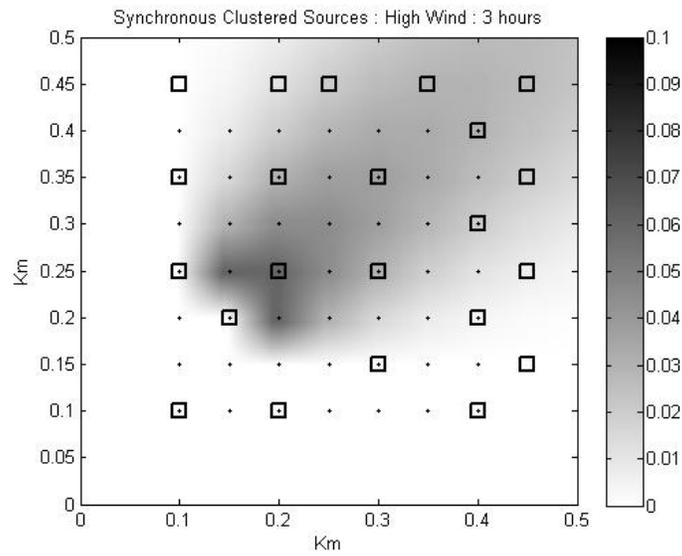
Figure E.7: %CO_2 at t = 3 hours.
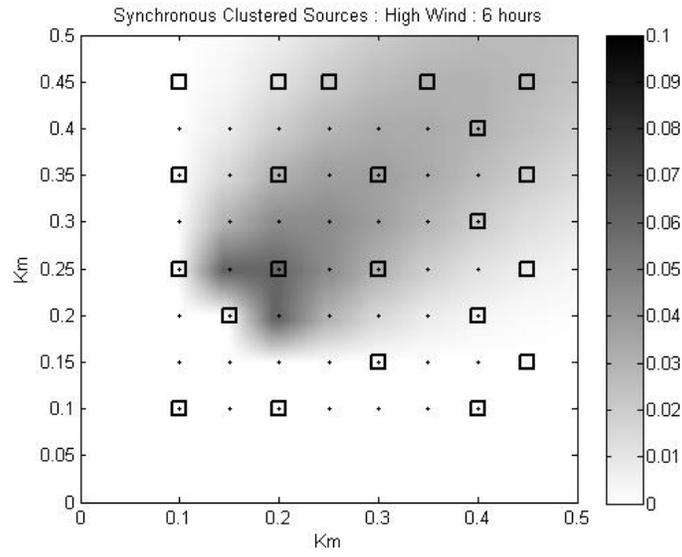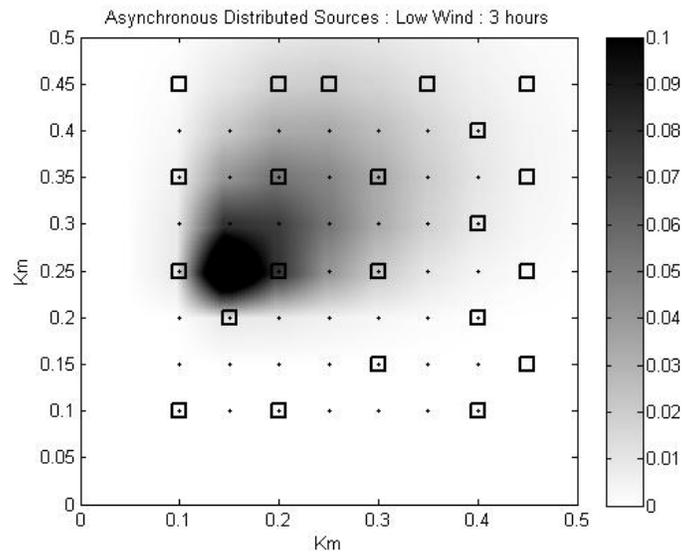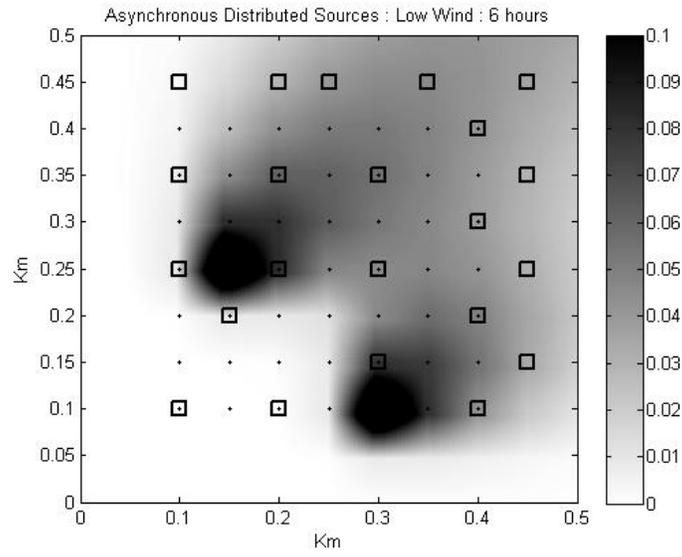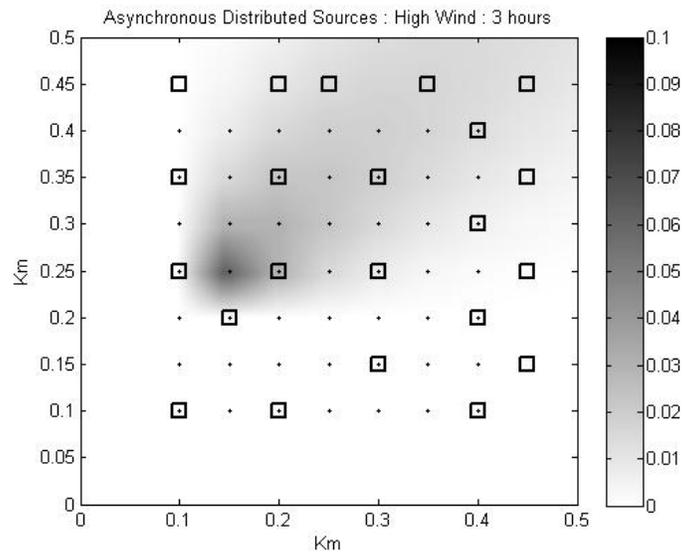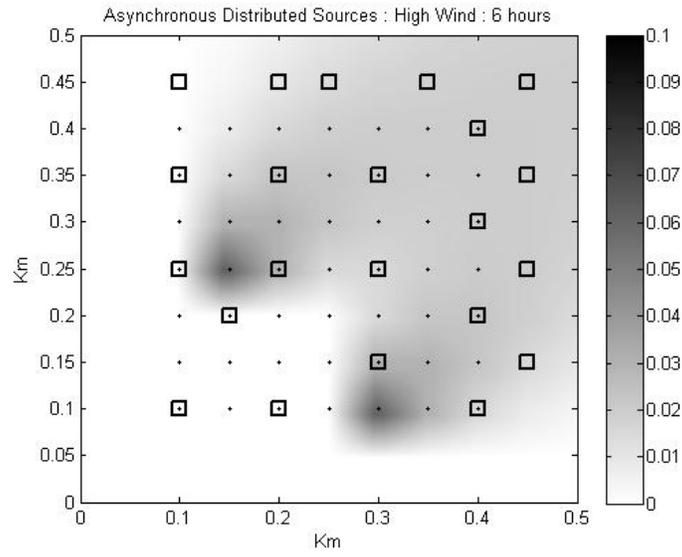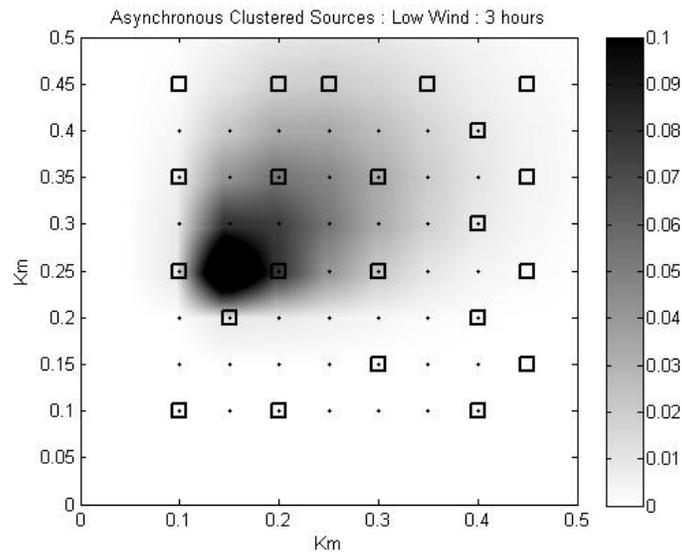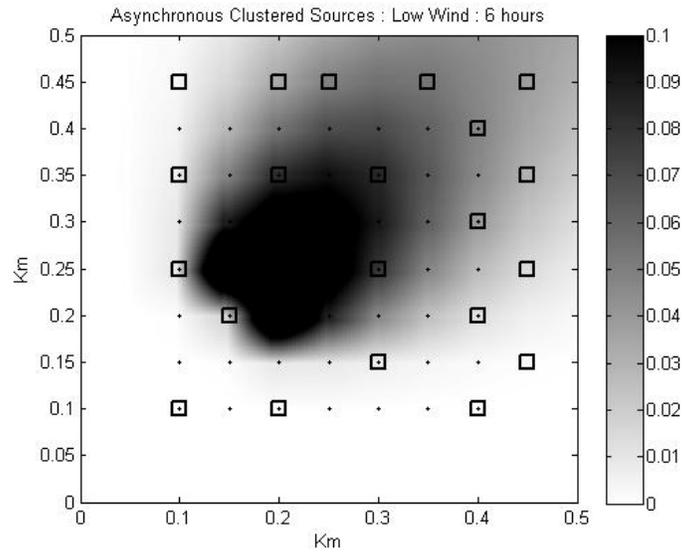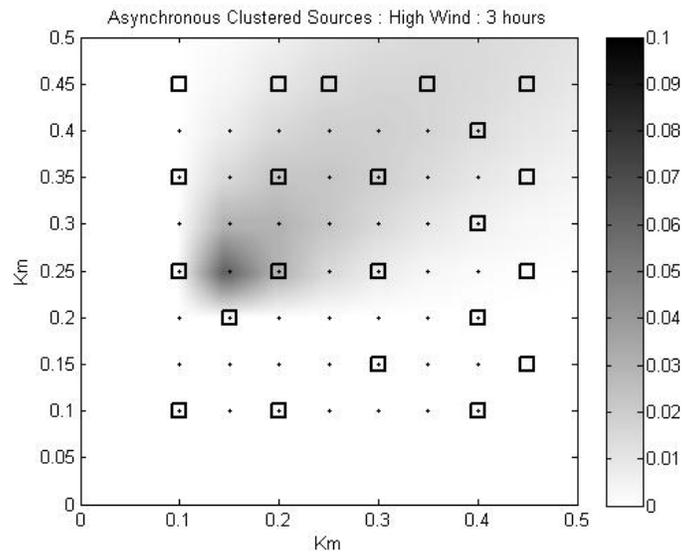
Figure E.8: %CO_2 at t = 6 hours.

Figure E.9: %CO_2 at t = 3 hours.

Figure E.10: %CO_2 at t = 6 hours.

Figure E.11: %CO_2 at t = 3 hours.

Figure E.12: $\%CO_2$ at t = 6 hours.

Figure E.13: $\%CO_2$ at t = 3 hours.
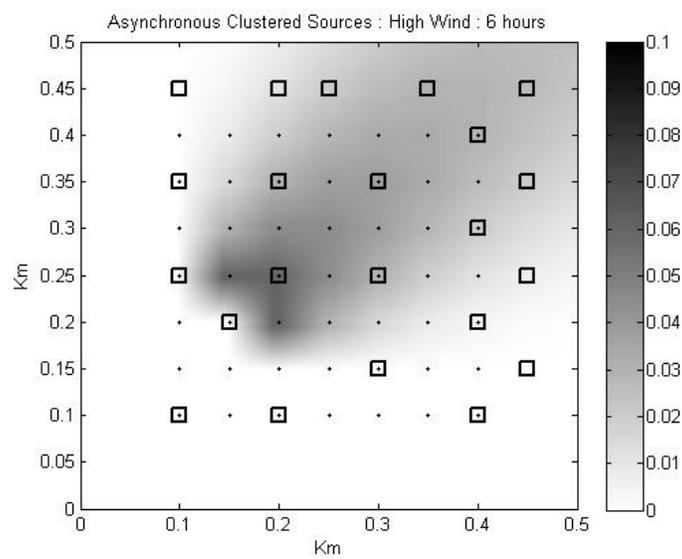
Figure E.14: $\%CO_2$ at t = 6 hours.

Figure E.15: $\%CO_2$ at t = 3 hours.

Figure E.16: %CO_2 at t = 6 hours.