

Multiscale, Statistical Anomaly Detection Analysis and Algorithms for Linearized Inverse Scattering Problems *

Eric L. Miller
The Communications and
Digital Signal Processing Center
Department of Electrical and Computer Engineering
235 Forsyth
Northeastern University
360 Huntington Ave.
Boston, MA 02115
Tel: (617) 373-8386
FAX: (617) 373-8627
email: elmiller@cdsp.neu.edu

Alan S. Willsky
Laboratory for Information and
Decision Systems
Department of Electrical Engineering
and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts, 02139

December 13, 1995

Abstract

In this paper we explore the utility of multiscale and statistical techniques for detecting and characterizing the structure of localized anomalies in a medium based upon observations of scattered energy obtained at the boundaries of the region of interest. Wavelet transform techniques are used to provide an efficient and physically meaningful method for modeling the non-anomalous structure of the medium under investigation. We employ decision-theoretic methods both to analyze a variety of difficulties associated with the anomaly detection problem and as the basis for an algorithm to perform anomaly detection and estimation. These methods allow for a quantitative evaluation of the manner in which the performance of the algorithms is impacted by the amplitudes, spatial sizes, and positions of anomalous areas in the overall region of interest. Given the insight provided by this work, we formulate and analyze an algorithm for determining the number, location, and magnitudes associated with a set of anomaly structures. This approach is based upon the use of a Generalized, M-ary Likelihood Ratio Test to successively subdivide the region as a means of localizing anomalous areas in both space and scale. Examples of our multiscale inversion algorithm are presented using the Born approximation of an electrical conductivity problem formulated so as to illustrate many of the features associated with similar detection problems arising in fields such as geophysical prospecting, ultrasonic imaging, and medical imaging.

*This work was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-92-J-0002, and the Advanced Research Project Agency under Air Force Grant F49620-93-1-0604 and the Office of Naval Research under Grant N00014-91-J-1004. The work of the first author was also supported in part by a US Air Force Laboratory Graduate Fellowship

1 Introduction

The goal of many applied problems is the recovery of information regarding the structure of a physical medium based upon measurements of scattered radiation collected at the boundaries [8, 15, 19, 43, 48]. For some of these tomographic-type inverse problems, one seeks a complete description (in the form of an image in two dimensions or a volumetric rendering in 3D) of the structure of the medium. In other cases, however, the full reconstruction is not needed; rather, the ultimate objective is to extract the structure of areas in the medium which are, in some sense, anomalous; that is, regions where the nature of the medium differs from some prior set of expectations. This *anomaly detection problem* arises, for example, in geophysical prospecting where in many instances the fundamental issue is the determination of oil bearing regions in the earth and medical imaging where tumor detection is of import.

As discussed in [27, 29, 31, 33, 34, 44, 45] for many of the application areas previously cited, methods for solving the anomaly detection problem typically proceed by initially generating the full, pixel-by-pixel reconstruction and subsequently post-processing the results to determine the nature of anomalous structures. The necessity of generating a solution to the so-called “full inverse problem” however makes these schemes rather unattractive. Indeed, for many interesting applications, obtaining a full reconstruction of the medium presents a collection of well-known and extensively studied challenges [2, 3, 40] which suggest that solving this problem as the first step toward localizing anomalies should be avoided. In this paper we demonstrate the utility of a multiscale framework for explicitly solving the spatial anomaly detection problem in the context of linearized inverse scattering (also known as diffraction tomography [15]) applications.

The basis for solving the anomaly detection problem is the use of wavelet transforms and the statistical theories of optimal estimation and detection to develop both efficient algorithms for anomaly detection and localization and analytical insight into the nature of the problem and the limits of performance that result from the fundamental physics relating the characteristics of the medium to the observations. In [39, 40], we introduced the use of wavelet transforms and multiresolutional statistical techniques for overcoming many of the challenges associated with the solution of full reconstruction, linearized inverse electrical conductivity problems. Many of the results in [39, 40] followed from the use of multiscale, statistical regularization methods for the incorporation of prior knowledge into the inversion routine. The use of such prior statistical models automatically implies an assumption of some type of statistical regularity on the field and therefore fails to capture adequately the presence of anomalies or localized inhomogeneities. Thus, roughly stated, the problem considered in this paper is the detection, localization, and estimation of such anomalies superimposed on a background of known statistical structure and observed indirectly through the scattering measurements.

The consideration of the anomaly detection problem raises a variety of questions beyond those arising in the full reconstruction inverse problem. How many anomalies are there? Where are they located? What are their sizes? What are their amplitudes? Given answers to the first three of these problems, the fourth is a variant of the full inverse problem in which we focus our attention on determining the magnitudes of only the previously identified anomalous regions. The determination of the number, sizes and locations of the anomalous regions is, however, a potentially daunting collection of tasks as a result of the vast number of combinations of anomaly structures which, in principle, must be explored in the generation of a solution.

Over the past decade, significant work has been performed in the area of anomaly detection from tomographic-type measurements. In [44], Rossi and Willsky were concerned primarily with the use of estimation-theoretic analysis and algorithmic methods for determining the location of

a single object of known size and structure given noisy and sparse computed tomography (CT) measurements. Recently, these results have been extended by Devaney and co-workers [16, 17, 46] in consideration of diffraction tomography (DT) and exact scattering applications. More closely related to the problem of interest in this paper is the work of Bresler, Fessler and Macovski. In [5], the authors examined a 3D reconstruction problem from CT measurements in which the first step of their algorithm required the localization of an unknown number of anomalies of unknown structure. The solution to this problem presented in [5] was to estimate the required parameters for a pre-determined, maximum number of anomalies knowing that further processing would eliminate falsely identified anomalous regions.

In this paper, we present a scale-recursive algorithm for anomaly detection and characterization given DT-type data. Here, the tools of optimal hypothesis testing are used to make a sequence of anomaly detection and localization decisions starting at coarse scales, thereby allowing for the detection of spatially large anomaly structures and providing coarse localization of finer scale anomalies, and then moving to finer ones. This algorithm is significant for two reasons. First, this approach provides a computationally efficient and accurate means of localizing areas of anomalous behavior. Second, the anomaly characterization algorithm may be viewed as a highly efficient first stage in a larger image processing application. Specifically, the output of the algorithm could be refined (for example via the methods described in [5, 44] generalized to the case of diffraction tomography) by higher level processing stages concerned with issues such as identification, classification, or imaging. Toward this end, in Section 6.3, we present one way in which knowledge of the anomaly structures can be used to supplement the information in the prior statistical model in order to improve the output of a least-squares, pixel-by-pixel reconstruction of the region of interest.

In addition to the development of the scale-recursive processing algorithm, by using these same statistical techniques, we provide analysis of the anomaly detection problem that not only yields overall performance limits, but also guides the detection procedure. For example, we are able to define and determine the statistical distinguishability of a small scale, large amplitude anomaly from a larger scale, but smaller magnitude structure or a pair of closely spaced anomalies from a single, broader anomalous region. The use of the results from this analysis can then tell us at what scale and in which regions to terminate our detection procedure, i.e. when finer scale localization is unwarranted given the available data.

In Section 2, we present an overview of the particular anomaly detection problem of interest in this work. The formal definition of the anomaly detection problem as one of optimal hypothesis testing and a review of results from statistical decision theory is provided in Section 3. In Section 4 we demonstrate the utility of our framework in characterizing the *detectability* of an anomaly. Section 5 is devoted to the question of the *distinguishability* of anomalies as a function of their relative positions and structures. In Section 6 we develop and analyze a scale-recursive algorithm for anomaly detection, localization, and estimation, and present the results of its performance under a variety of experiment conditions. Conclusions reached in this paper and directions for further work are presented in Section 7.

2 A Multiscale Framework for Inverse Scattering

2.1 The Scattering Problem

The context in which we develop our anomaly detection algorithm is a low-frequency, two-dimensional inverse electrical conductivity problem illustrated in Figure 1 and similar to problems arising in the field of geophysical prospecting [23, 24, 48] and medical imaging using electrical impedance

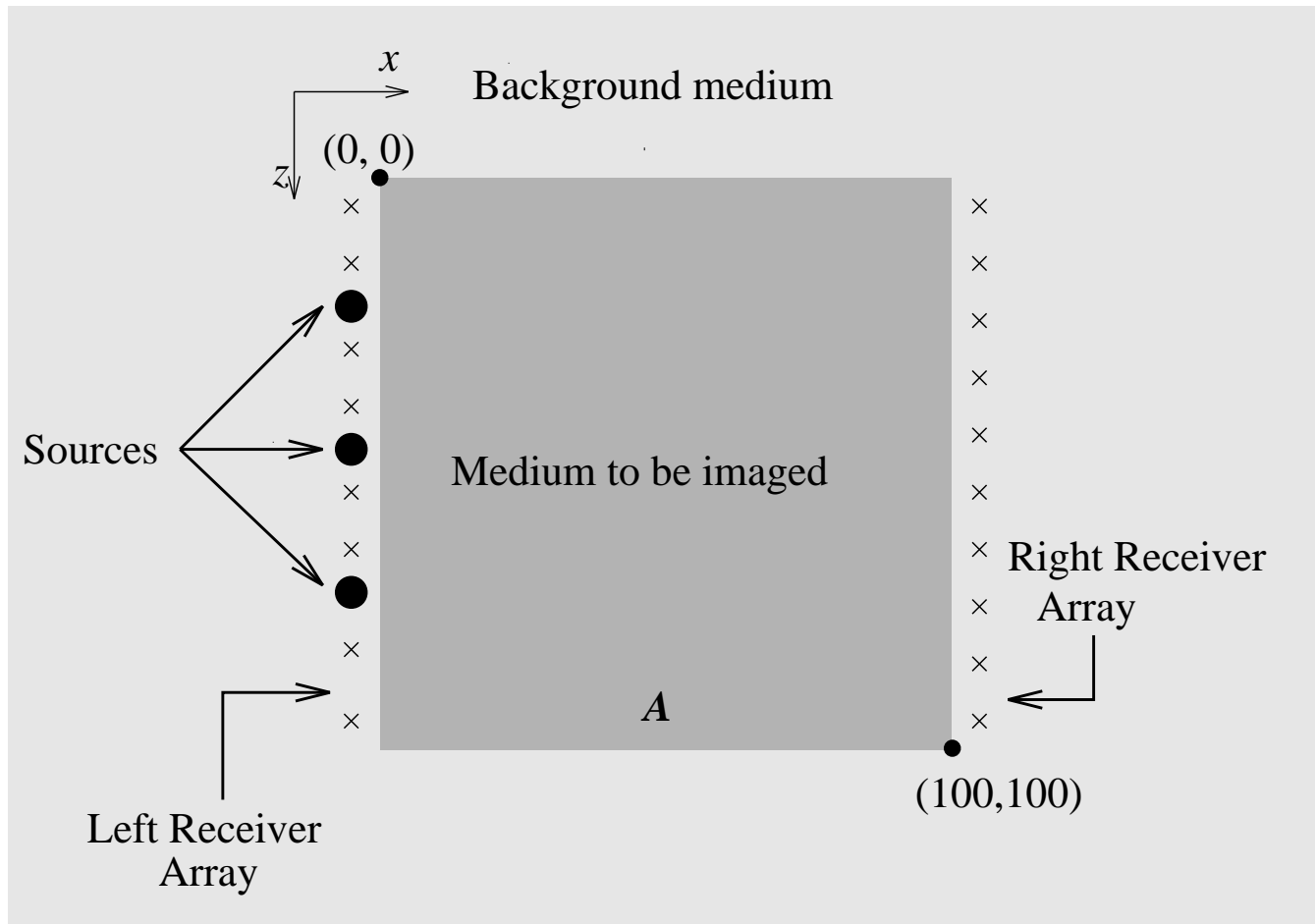


Figure 1: Configuration of inverse conductivity problem. The electromagnetic sources (indicated by the black circles) emit time-harmonic waves into a lossy medium which subsequently are scattered by conductivity inhomogeneities located in the darkly shaded rectangle, A . The secondary fields are observed at one or both receiver arrays located on either vertical edge of region under investigation. Based upon these observations, the objective of the inverse problem is the reconstruction of the conductivity perturbation.

tomography [18–20, 22, 29–31, 43]. Here, we have an array of electromagnetic line-sources oriented perpendicularly to the page emitting time-harmonic, waves into a lossy medium. The electrical properties of this environment are assumed to be decomposed into the sum of an infinite, known, and constant background and a conductivity perturbation, g , with support restricted to region A in Figure 1. The fields from the transmitters are scattered by g , and the secondary fields are observed at one or both of the receiver arrays positioned on the vertical edges of region A . Based upon these observations, the objective of the problem is to detect and localize areas in the region of interest where the structure of g is, in a sense to be defined below, anomalous.

We consider the collection of eighteen scattering experiments defined in Table 1 where each such experiment produces a vector of measurements comprised of the in-phase and quadrature

components of the observed scattered field obtained over one of the two receiver arrays due to energy put into the medium from one of the sources operating at a particular frequency. As is shown in [9], the use of the first Born approximation yields the following linear relationship between the vector of observations associated with the i^{th} scattering experiment, y_i , and a discrete representation of the two dimensional conductivity anomaly, g

$$y_i = T_i g + n_i \quad i = 1, 2, \dots, 18 \quad (1)$$

where the matrices T_i encompass the (linearized) physics and n_i is an additive, zero-mean, uncorrelated, random vector representing the noise in the data. That is, the i^{th} noise is modeled as $n_i \sim \mathcal{N}(0, r_i I)$ where I is an appropriately sized identity matrix.¹ The discrete representation of the conductivity g is constructed using the so-called “pulse” set of basis functions where the conductivity is assumed to be piecewise constant over an $N_{g,x} \times N_{g,z}$ grid of square pixels covering A [26]. For future reference, we define the “stacked” system of data

$$y = Tg + n \quad (2)$$

where $y^T = [y_1^T \ y_2^T \ \dots \ y_{18}^T]$ with T and n defined accordingly.

Experiment number	Source Position	Frequency of source (Hz)	Receiver Array
1 – 6	0:20:100	$f_{HI} = 10,000$	Left
7 – 12	0:20:100	$f_{MID} = 1,000$	Left
13 – 18	0:20:100	$f_{LO} = 100$	Right

Table 1: Data set definitions for observation processes of interest in the paper. The notation $x : y : z$ indicated that the sources are distributed in y increments along a line from x to z .

2.2 A Multiscale Representation of the Problem

The detection techniques developed in Sections 4 – 6 are based upon a linear model relating multiresolution representations of g and n_i to a multiresolution representation the data, y_i . A scale-space representation of the problem has been chosen for two reasons. First, the matrices T_i in (1) are of the class which are made sparse in the wavelet transform domain [1,4] thereby lowering the computational complexity of the detection algorithm in Section 6. Although not considered extensively in this work, such computational benefits are explored in [41]. Second, as we discuss below, a collection of useful and physically meaningful models for the non-anomalous behavior of the conductivity field are specified easily in the wavelet domain.

Following the work in [39,40], orthonormal, discrete wavelet transform (DWT) [14] operators (matrices) \mathcal{W}_i and \mathcal{W}_g are used to move from physical to scale space in the following manner

$$\eta_i = \mathcal{W}_i y_i = (\mathcal{W}_i T_i \mathcal{W}_g^T)(\mathcal{W}_g g) + \mathcal{W}_i n_i \equiv \Theta_i \gamma + \nu_i \quad (3)$$

where $\mathcal{W}_g^T \mathcal{W}_g = \mathcal{W}_i^T \mathcal{W}_i = I$ follows from the orthonormality of the wavelet transformation [14,35]. There are a variety of reasons why we may wish to use different transforms for the data than for g . First, from Figure 1, each data set is to be collected over a 1D array of receivers. Hence, \mathcal{W}_i will act on a one dimensional signal while \mathcal{W}_g is used to transform the 2D conductivity profile. Additionally, it may be the case that the lengths of each data record vary from one observation process to the next. Finally, analogously to the physical space case, we define the stacked systems

$$\eta = \Theta \gamma + \nu \quad (4)$$

where $\eta = [\eta_1^T \ \eta_2^T \ \dots \ \eta_{18}^T]^T$, Θ and ν are defined analogously and $\nu \sim \mathcal{N}(0, R)$ with $R = \text{diag}(r_1 I, r_2 I, \dots, r_{18} I)$.

¹The notation $x \sim \mathcal{N}(m, P)$ indicates that the random vector x has a Gaussian probability distribution with mean m and covariance matrix P .

2.3 Multiscale Prior Models

Recently there has been significant work in the use of fractal models for describing the spatial distribution of geophysical quantities. In [13], Crossley and Jensen explore the propagation of acoustic radiation in the Earth's crust using a velocity model composed of the sum of a deterministic profile and a fractal perturbation. In considering the distribution of hydraulic conductivity, Brewer and Wheatcraft [6] employ a wavelet-based model very similar to the one described below as a means of interpolating coarse scale observations of hydraulic conductivity to finer scales. Brown [7] relates both the electrical and hydraulic conductivities in the earth to a self-similar model for the height distribution in rock fractures and studies the resulting fluid and current flow patterns through such a formation. Finally, the propagation of electromagnetic radiation through media with fractal characteristics has been studied extensively by Jaggard and co-workers [32].

With this work as motivation, we use a stochastic, fractal-type model to describe the spatial distribution of the electrical conductivity in the absence of anomalies. While there are many self-similar models which may be used to describe the conductivity, results of Wornell [50], Tewfik [47], and Chou et al. [10–12] suggest that there exist a wide range of statistical models specified *directly* in the wavelet transform domain possessing the desired modeling characteristics and simple structures thereby making them quite attractive for use in signal and image processing applications.

Under the particular wavelet-based model of interest in this paper, the wavelet coefficients of the non-anomalous conductivity field, denoted by the vector $\tilde{\gamma}$, are taken to be uncorrelated, Gaussian random variables. That is, $\tilde{\gamma}$ is distributed according to

$$\tilde{\gamma} \sim \mathcal{N}(0, P_0) \quad (5)$$

where P_0 is a *diagonal* matrix whose nonzero entries are the variances of the corresponding wavelet coefficients. While a detailed description of the internal structure of P_0 is presented in [35, 50], the fractal-type behavior of the process is obtained by taking the variance of the wavelet coefficients to vary exponentially with scale. Coefficients in $\tilde{\gamma}$ governing the coarsest scale behavior of the conductivity have relatively large variances while fine scale components possess smaller variances.

3 Anomaly Detection as a Hypothesis Testing Problem

3.1 A Model for the Conductivity

The objective of the anomaly detection problem is to determine those areas in A where the behavior of g is anomalous in that in these regions g differs from some prior set of beliefs regarding the manner in which the conductivity is expected to behave. Thus, the conductivity g is decomposed as

$$g = \tilde{g} + \bar{g} \quad (6)$$

where \tilde{g} represent that portion of g consistent with our prior assumptions and \bar{g} encompasses the anomalous behavior of the conductivity; that is, the perturbation of the conductivity away from its non-anomalous structure. In the wavelet transform domain, (6) takes the form

$$\gamma = \mathcal{W}_g g = \mathcal{W}_g \tilde{g} + \mathcal{W}_g \bar{g} \equiv \tilde{\gamma} + \bar{\gamma}. \quad (7)$$

As will be seen in Sections 4–6, considerable insight into the anomaly detection problem is obtained through performance analysis carried out using anomaly structures of varying sizes (i.e. spatial scales) located in different regions of A . Also, the primary intent of the detection algorithm presented in Section 6 is to localize quickly and efficiently regions where anomalies are suspected to exist. As region A is pixelated into an $N_{g,x} \times N_{g,z}$ grid and because we perform anomaly localization through a process of spatial subdivision, we are lead naturally to consider a representation in which anomalous regions are defined to be superpositions of rectangular subsets of A .

Referring to Figure 2, the structure of the i^{th} anomaly in A is defined by its magnitude, a_i , its size, and its location in A . The area of an anomaly defines its *scale* in that small scale anomalies are

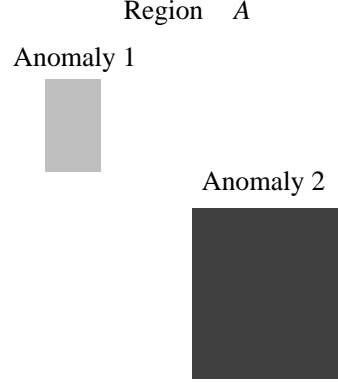


Figure 2: General structure of anomalous regions of interest in this paper. The magnitudes, a_1 and a_2 of the two anomalies shown here are proportional to the color of the corresponding rectangles.

correspondingly small in area and similarly for larger scale anomaly structures. Mathematically, the form for the anomalous behavior of the conductivity over the region A is

$$\bar{g} = \sum_{j=1}^{N_a} b_j a_j = Ba. \quad (8)$$

Here, N_a is the number of anomalous regions located in A , a_j is a scalar defining the magnitude of each anomaly, and b_j represents the discrete indicator function over the j^{th} rectangular region in \bar{g} . In (8), the column vector a represents the collection of anomaly amplitude coefficients while B is the matrix whose j^{th} column is b_j . In the wavelet transform domain, (8) is written as

$$\bar{\gamma} = \sum_{j=1}^{N_a} (\mathcal{W}_g b_j) a_j \equiv \mathcal{B}a \quad (9)$$

where $\mathcal{B} = [\mathcal{W}_g b_1 \ \mathcal{W}_g b_2 \ \dots \ \mathcal{W}_g b_{N_a}]$. Finally, use of (7) and (9) in (4) yields the following relationship among the anomaly structures, the non-anomalous background \tilde{g} or $\tilde{\gamma}$, and the data

$$\eta = \Theta \bar{\gamma} + \Theta \tilde{\gamma} + \nu = \Theta \mathcal{B}a + \Theta \tilde{\gamma} + \nu \quad (10)$$

where, because $\tilde{\gamma}$ and ν are taken to be uncorrelated,

$$P_\eta = E[\eta \eta^T] - E[\eta]E[\eta^T] = \Theta P_0 \Theta^T + R. \quad (11)$$

Note that the analysis methods and algorithmic techniques presented in this work are based entirely on an observation model of the form in (10). In particular, the results in this paper are not dependent upon the assumption of rectangular anomalies; rather structures with arbitrary shapes and orientations can be employed in principle through the appropriate specification of the matrix \mathcal{B} . Nonetheless, as will be seen in Sections 4–6 of this paper, rectangular structures prove to be highly useful for obtaining significant insight into the nature of the anomaly characterization problem and as the basis for an algorithm designed to extract this information from observed scattered fields.

To provide a normalized notion of the overall size of an anomaly, we define an SNR-type quantity called the *anomaly-to-background ratio* (ABR) which provides a measure of the energy in an anomaly relative to that of \tilde{g} . Mathematically, we have for an anomaly \bar{g} composed of a single rectangular region defined by the column vector b and with amplitude a

$$ABR^2 = \frac{\text{Power in } \bar{g}}{\text{Expected power in } \tilde{g}} = \frac{a^2 (b^T b)}{\text{tr}(\bar{P}_0)} \quad (12)$$

where $\text{tr}(M)$ is the trace of the matrix M and $\bar{P}_0 = \mathcal{W}_g^T P_0 \mathcal{W}_g$ is the covariance matrix of \tilde{g} .

As described in [40], under the Born approximation used to obtain (1), $g = \mathcal{W}_g^T \gamma$ represents a perturbation about a known, constant background conductivity, g_0 . From physical principles, the overall conductivity, $g_0 + g = g_0 + \tilde{g} + Ba$ must be greater than zero. Thus, in theory the elements of a may assume both positive as well as negative values so long as the positivity constraint is satisfied. To simplify matters, in this paper we assume that the a_i are strictly greater than zero corresponding to regions of locally higher conductivity than the background.

3.2 The M-ary Hypothesis Testing Problem

In Section 6, we consider a statistical decision-theoretic methodology for reconstructing $\bar{\gamma}$ which is based upon a sequence of M-ary Generalized Likelihood Ratio Tests (GLRT) as a means of localizing an unknown number of anomalous regions in A . The mathematical description of each such test begins with the formulation of the following M hypotheses, H_i for $i = 0, 1, 2, \dots, M-1$, corresponding to M different configurations of anomalous areas

$$H_i: \quad \eta = \Theta \mathcal{B}_i a_i + \Theta \tilde{\gamma} + \nu \quad i = 0, 1, 2, \dots, M-1. \quad (13)$$

Note that from (13) under H_i we have, $\eta \sim \mathcal{N}(\Theta \mathcal{B}_i a_i, P_\eta)$ where P_η is given by (11).

The hypothesis test is implemented as a rule which when given the data, indicates which of the H_i is true. Because it will be the case in Section 6 that the a_i are taken to be deterministic but unknown parameters, a standard likelihood ratio test solution to the hypothesis testing problem [49] cannot be employed in this context. Rather, we use a Generalized Likelihood Ratio Test (GLRT) [49] for performing the test. This procedure requires first that an estimate of each a_i be computed assuming that H_i is correct. As this problem is, in general, ill-posed, we choose here to use the following regularized, least squares estimate

$$\hat{a}_i = (\mathcal{B}_i^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_i + \alpha I)^{-1} \mathcal{B}_i^T \Theta^T P_\eta^{-1} \eta. \quad (14)$$

where the parameter α is used to control the degree of regularization.

Given \hat{a}_i , the hypothesis testing rule employed in this paper is

$$\text{Choose } H_i \text{ with } i = \begin{cases} 0 & \max_j L_j(\eta) < 0 \\ \arg \max_j L_j(\eta) & \text{otherwise} \end{cases} \quad (15)$$

where

$$L_j(\eta) = l_j(\eta) - l_0(\eta) \quad j = 1, 2, \dots, M-1 \quad (16)$$

and for $j = 0, 1, 2, \dots, M-1$

$$l_j(\eta) = \eta^T P_\eta^{-1} \Theta \mathcal{B}_j \hat{a}_j - \frac{1}{2} \hat{a}_j^T \mathcal{B}_j^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_j \hat{a}_j. \quad (17)$$

3.3 The Binary Hypothesis Testing Case

While the algorithm for extracting anomaly information is based upon the M-ary GLRT, much of the analysis of the anomaly detection problem is performed in the context of the *binary hypothesis testing* (BHT) framework in which two alternatives, $\bar{\gamma}_0 = \mathcal{B}_0 a_0$ and $\bar{\gamma}_1 = \mathcal{B}_1 a_1$, are compared.² Traditionally, the analysis of the BHT centers around the probability of detection, P_d and the false alarm probability, P_f . For the linear-Gaussian model considered in this work, it is shown in [49] that P_d and P_f are related to the various quantities defining the structure of the problem via

$$d = \text{erfc}^{\star 1}(P_f) - \text{erfc}^{\star 1}(P_d) \quad (18)$$

²Note that in the contexts where the binary testing scenario is to be explored, the values of a_0 and a_1 are assumed known so that a generalized test is not required.

where

$$d^2 = (\bar{\gamma}_1 - \bar{\gamma}_0)^T \Theta^T P_\eta^{-1} \Theta (\bar{\gamma}_1 - \bar{\gamma}_0) \quad (19)$$

$$\text{erfc}_*(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt. \quad (20)$$

Thus, based upon (18), we see that our ability to distinguish between two anomaly structures is intimately related to the Fisher discriminant, d , which has the interpretation of a “signal-to-noise” ratio [49]. Note that for a given P_f , larger d results in larger P_d and therefore better performance.

From (19) we observe that the performance of the binary hypothesis test is a function of both the geometric configurations, as captured in the matrices \mathcal{B}_i , and the magnitudes, a_i , of the two candidate anomaly structures. To better understand the role of these two factors, consider the case in which $\bar{\gamma}_i$ corresponds to a single rectangular region so that each \mathcal{B}_i is a column vector and each a_i is a scalar. Substituting (9) into (19) and expanding the quadratic yields

$$\delta_1^2 a_1^2 - 2\delta_{1,0} a_1 a_0 + \delta_0^2 a_0^2 - (\text{erfc}_*^1(P_f) - \text{erfc}_*^1(P_d)) = 0 \quad (21)$$

where

$$\delta_j^2 = \mathcal{B}_j^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_j \quad \text{for } j = 0, 1 \quad (22)$$

$$\delta_{1,0} = \mathcal{B}_1^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_0. \quad (23)$$

In [37], it is shown that when viewed as a function of a_0 and a_1 , (21) defines an ellipse the form of which is illustrated in Figure 3.³ This ellipse indicates that, given the geometry of the candidate anomalies, \mathcal{B}_0 and \mathcal{B}_1 , there are only certain combinations of a_0 and a_1 which will result in performance below that level dictated by a particular P_d and P_f . In fact, these points are precisely those that lie inside the plotted ellipse. Also, there exists a minimum level, $a_{1,0}^{min}$ (depending on the geometric structures of *both* anomalies) such that for $\bar{\gamma}_1 = \mathcal{B}_1 a_1$ with $a_1 > a_{1,0}^{min}$, the binary hypothesis test will achieve or exceed the P_d and P_f performance figures *independent* of a_0 .

4 Detectability Analysis

The first issue we address in conjunction with the anomaly detection problem is that of the detectability of an anomaly as a function of location, spatial size, and amplitude. After defining a particular collection of anomaly structures, a set of binary hypothesis testing problems are explored in which H_0 corresponds to there being no anomaly in the region while under H_1 , a particular member of our anomaly collection is assumed to be present. The objective of the detectability analysis is to determine the minimum magnitude each such structure must possess to guarantee a prespecified level of performance from the binary hypothesis test.

Detectability is of interest due to the physics governing the relationship between the observations, η , and the conductivity, γ and the constrained experimental conditions in which data are collected only along the vertical edges of A . From these facts, it is not expected that arbitrarily small (in scale and magnitude) anomalies will be detectable with arbitrary precision throughout A . Rather, we anticipate that small anomalies should be readily detected only close to the observation points while interior to region A small scale structures would require significantly larger magnitudes to be as detectable as their counterparts closer to the edges.

With this intuition in mind, we consider a family of anomaly structures generated by a set of dyadic tessellations of A . For example, with $N_{g,x} = N_{g,z} \equiv N_g = 16$, we take as \mathcal{J}_1 the set of N_g^2 indicator functions which are one over single pixels in A and zero elsewhere. Analogously, \mathcal{J}_2 is the collection of $N_g^2/4$ characteristic functions over disjoint 2×2 sized regions of A . Thus, in general

³For illustrative purposes only, in Figure 3 it is assumed that the major axis of the ellipse is oriented at an angle less than 90° from the a_0 axis. While this is not necessarily the case, the analysis which follows is independent of which axis is the major and which the minor.

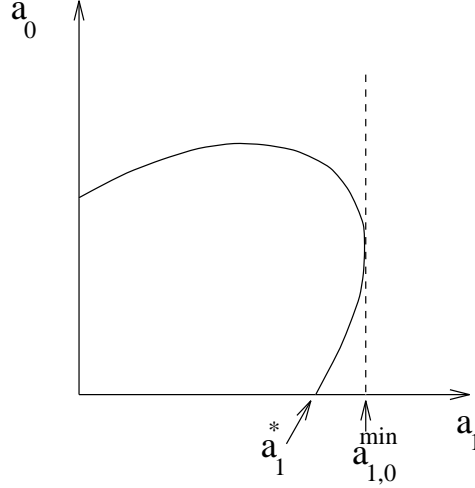


Figure 3: The structure of the ellipse defined by (21). The axes represent the magnitudes of anomaly structures in a binary hypothesis testing problem. As discussed in Section 2, a_0 and a_1 are taken to be nonnegative so that only the first quadrant is shown in this illustration. Here a_1^* is the minimum amplitude of $\tilde{\gamma}_1$ required to detect this structure when the alternate hypothesis is $\tilde{\gamma}_0 = 0$ for a BHT with prespecified P_d and P_f . The value $a_{1,0}^{min}$ is the minimum intensity of $\tilde{\gamma}_1$ required to ensure that for *any* $\tilde{\gamma}_0$ the performance of the resulting BHT meets or exceeds that defined by P_d and P_f .

\mathcal{J}_m (for m an integral power of 2) is the set of $(N_g/m)^2$ non-overlapping square regions of size $m \times m$ completely covering A . Finally, we define \mathcal{J} as the union of all \mathcal{J}_m .

To begin our analysis of detectability, for each anomaly structure in \mathcal{J} , we consider a collection of binary hypothesis testing problems where the two hypotheses in the j^{th} problem correspond to the situations in which no anomaly is present in A or a scaled version of the j^{th} element of \mathcal{J} is in A . Recalling (13), these alternatives take the form

$$H_0 : \eta = \Theta \tilde{\gamma} + \nu \quad (24a)$$

$$H_{1,j} : \eta = \Theta \mathcal{B}_j a_j + \Theta \tilde{\gamma} + \nu. \quad (24b)$$

The goal of our detectability analysis then is to determine for each anomaly structure in \mathcal{J} , the minimum value of a_j , denoted a_j^* , such that the above hypothesis test attains a certain level of performance as specified by P_d and P_f .

The primary quantity used to characterize the performance of the binary hypothesis test in (24a)–(24b) is the Fisher discriminant discussed in the previous section which here takes the form

$$d_j^2 = a_j^2 (\mathcal{B}_j^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_j) \equiv a_j^2 \delta_j^2 \quad (25)$$

where δ_j^2 is defined in (22) and represents the Fisher discriminant for the unit amplitude anomaly over the j^{th} member of \mathcal{J} . Now, for a given P_d and P_f , (18) and (25) are combined to give the following expression for a_j^* :

$$a_j^* = \frac{\text{erfc}^{-1}(P_f) - \text{erfc}^{-1}(P_d)}{\delta_j}. \quad (26)$$

In Figure 4, a_j^* are plotted for all anomalies in \mathcal{J} for the case in which data from the 18 experiments described in Table 1 at an SNR of 10 are available and where P_d is set to 0.95 and P_f is 0.05. In this work, the SNR associated with the anomaly-free observation process $\eta_i = \Theta_i \tilde{\gamma} + \nu_i$ with $\nu_i \sim (0, r_i^2 I)$ and $\gamma \sim (0, P_0)$ is defined as

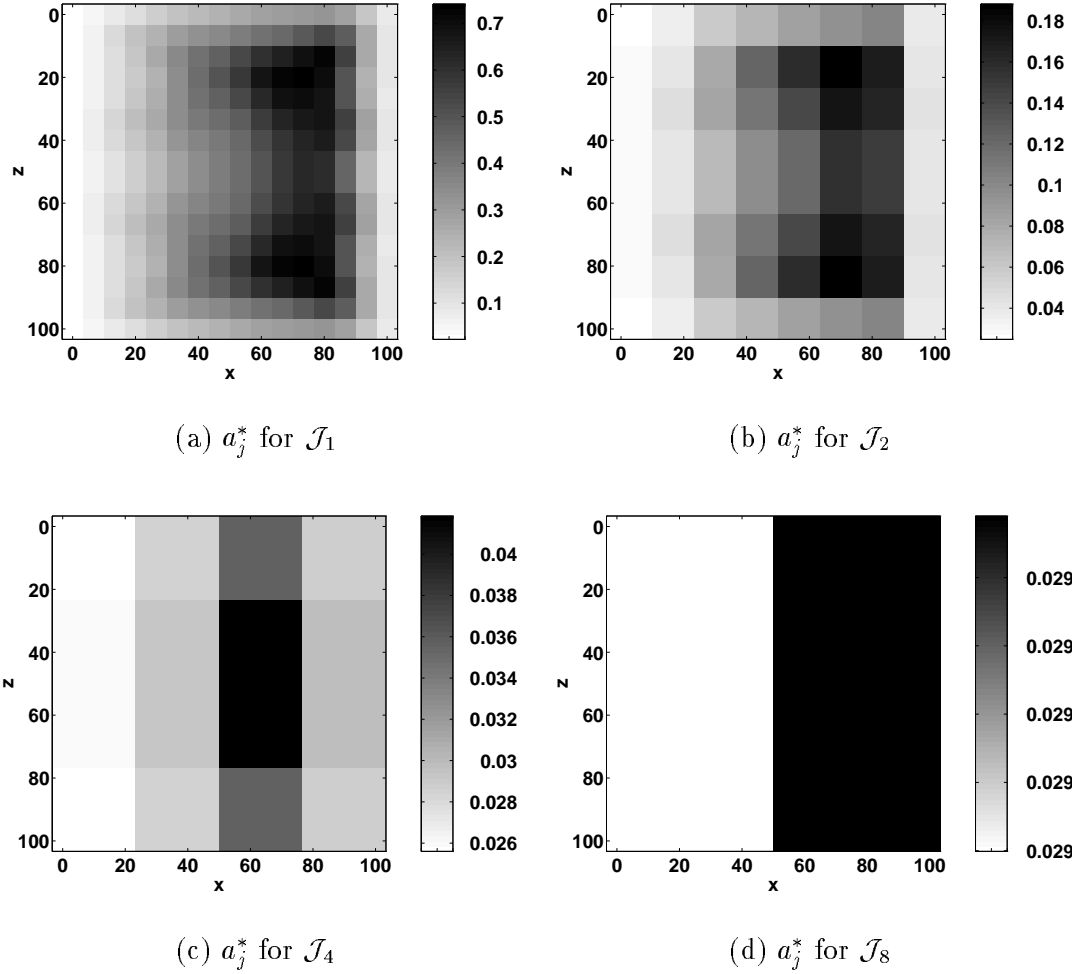


Figure 4: Value of a_j^* for all anomaly structures in \mathcal{J} where the data from the experiments described in Table 1 at an SNR of 10 are used as input to the likelihood ratio test. Here, we have $P_d = 0.95$ and $P_f = 0.05$. Note that the scales in these images are all different with a^* decreasing significantly as the size of the anomalies increases.

$$SNR_i^2 = \frac{\text{Power per pixel in } \Theta_i \gamma}{\text{Power per pixel in } \nu_i} = \frac{\text{tr}(\Theta_i P_0 \Theta_i^T)}{N_g r_i^2}. \quad (27)$$

Thus, each 1×1 pixel in Figure 4(a) corresponds to an anomaly in \mathcal{J}_1 with the intensity of that pixel proportional to a_j^* . In all four cases, we see that near the middle of the region, the magnitude required to obtain the desired level of performance in the binary hypothesis test is significantly larger than that required near the vertical edges i.e. where the sources and receivers are located. For vertical values roughly in the range $40 \leq z \leq 60$, this effect is somewhat smaller. Also, as the areas of the anomalies increase, the required magnitudes decrease. This coincides with the intuition that large scale structures should be easier to detect than their fine scale counterparts. Finally, the ABR values in Figure 4 are quite small with the median values all less than 0.9. This implies that our statistical approach toward anomaly detection should prove quite advantageous in detecting relatively small amplitude conductivity anomalies.

To explain the behavior of a_j^* , we note that as described in [40] the low and medium frequency

kernels are most sensitive to the conductivity structure over the horizontal range $0 \leq x \leq 50$ so that the required magnitude for an anomaly to be “seen” in this area should be relatively low. The smaller values of a_j^* in the region $40 \leq z \leq 60$ are due primarily to the combined coverage of this region by more observation kernels, T_i , than is the case for the top and bottom edges.

5 Distinguishability Analysis

In this section, we explore issues associated with our ability to successfully distinguish between pairs of candidate anomalies in order to obtain quantitative insight into the ambiguity which exists in attempting to differentiate between anomalous structures of differing sizes, locations, and magnitudes. The results of this work then are used both in the formulation as well as the analysis of the detection algorithm in Section 6.

Before proceeding with the analysis, we note that the issue of distinguishability has been considered previously in the context of electrical impedance tomography [18, 22, 29]. In that work, distinguishability was examined in a deterministic setting where observation perturbation was modeled as a bounded but otherwise unknown signal. Under such a model, two conductivity profiles were defined to be distinguishable if the norm of the difference between the data sets produced by each exceeded the noise level. The notion of distinguishability developed below is rather different as it rests upon a statistical model for both the additive measurement noise and background perturbations in the medium’s conductivity.

The mathematical formulation of the distinguishability problem of interest in this work follows directly from Section 3.3. We begin by considering the following binary hypothesis testing problem

$$H_j : \eta = \Theta \mathcal{B}_j a_j + \Theta \tilde{\gamma} + \nu \quad (28a)$$

$$H_i : \eta = \Theta \mathcal{B}_i a_i + \Theta \tilde{\gamma} + \nu. \quad (28b)$$

The primary tool for our distinguishability analysis is the quantity $a_{i,j}^{min}$ defined in Section 3.3 to be the smallest value of a_i such that the performance of the binary hypothesis test in (28a)–(28b) meets or exceeds that defined by $P_{d,i,j}$ and $P_{f,i,j}$ independent of the amplitude of a_j . Finally, for all experiments and for all i and j of interest in this section, $P_{d,i,j}$ is equal to 0.95 and $P_{f,i,j} = 0.05$.

In Figures 6, $a_{i,j}^{min}$ is shown as a function of $j \in \mathcal{J}$ in the case where the geometric structure of anomaly $\tilde{\gamma}_i$ is given in Figure 5(a). Similarly, $a_{i,j}^{min}$ is displayed for the anomaly geometry of Figure 5(b) in Figure 7. Essentially these two examples demonstrate the manner in which the ability to differentiate structures is dependent upon the spatial position of the anomalies in region A . In both cases, we see that the largest values of $a_{i,j}^{min}$ are associated with hypothesis tests in which $\tilde{\gamma}_i$ is compared to a second, relatively close-by anomaly structure; however, these amplitudes are roughly twice as large for the structure located toward the middle of the region than for the anomaly closer to the source/receiver arrays.

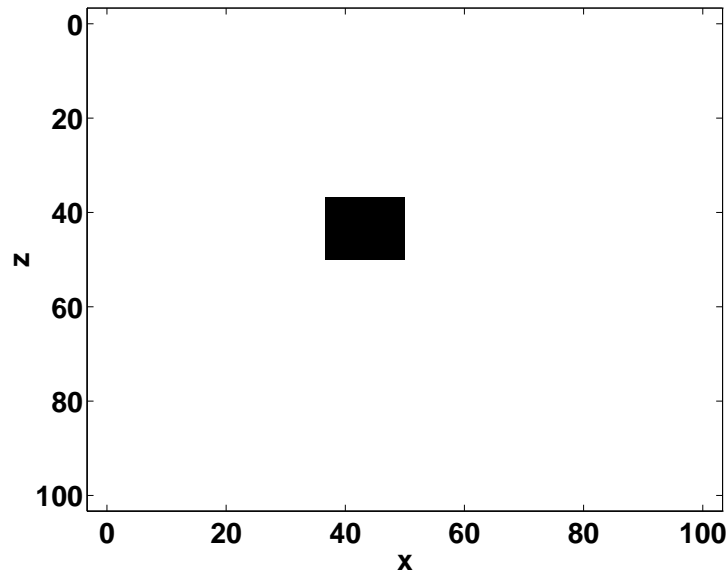
In Table 2, the ABRs corresponding to the largest and smallest values for $a_{i,j}^{min}$ in Figures 6 and 7 are shown. That is for i fixed, the entries in the first column of Table 2 are the anomaly-to-background ratios generated by

$$a_i^{max,min} = \max_j a_{i,j}^{min}$$

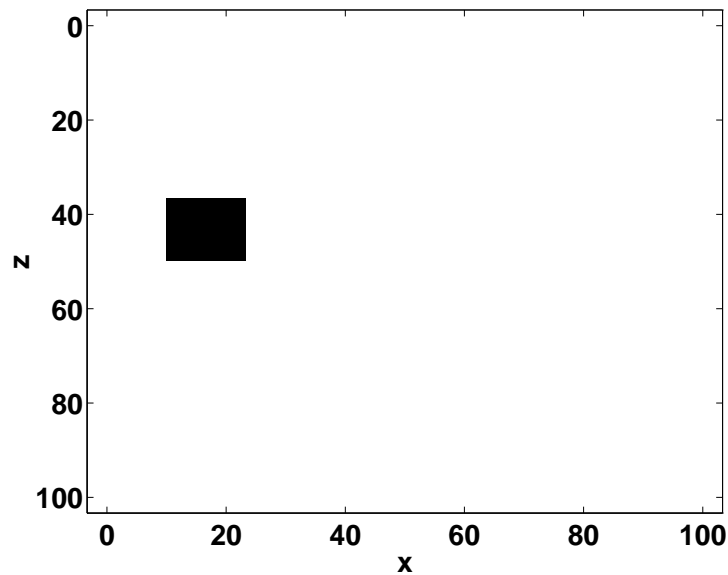
while those of the second column are associated with

$$a_i^{min,min} = \min_j a_{i,j}^{min}.$$

Note that if a_i is greater than $a_i^{max,min}$, a BHT with the anomaly $\tilde{\gamma}_i$ given by $\mathcal{B}_i a_i$ will meet the $P_{d,i,j}$ and $P_{f,i,j}$ specification regardless of both the amplitude as well as the location of γ_j , i.e. the performance will be independent of j . On the other hand if a_i is less than $a_i^{min,min}$ then for every j there will be some range of amplitudes a_j for which the performance specifications will



(a) First anomaly structure to be analyzed in distinguishability problems



(b) Second anomaly structure to be analyzed in distinguishability problems

Figure 5: Anomaly structures to be analyzed in distinguishability problems

not be achieved. Now, from the first row of Table 2, we see that for an anomaly with geometric structure in Figure 5(a), an ABR of 2.11 ensures that any binary hypothesis test in which this structure is compared to a member of \mathcal{J} will meet the performance specifications of $P_{d,i,j} = 0.95$ and $P_{f,i,j} = 0.05$. Alternatively, if the ABR falls below 0.56 then for all structures in \mathcal{J} , (i.e. all

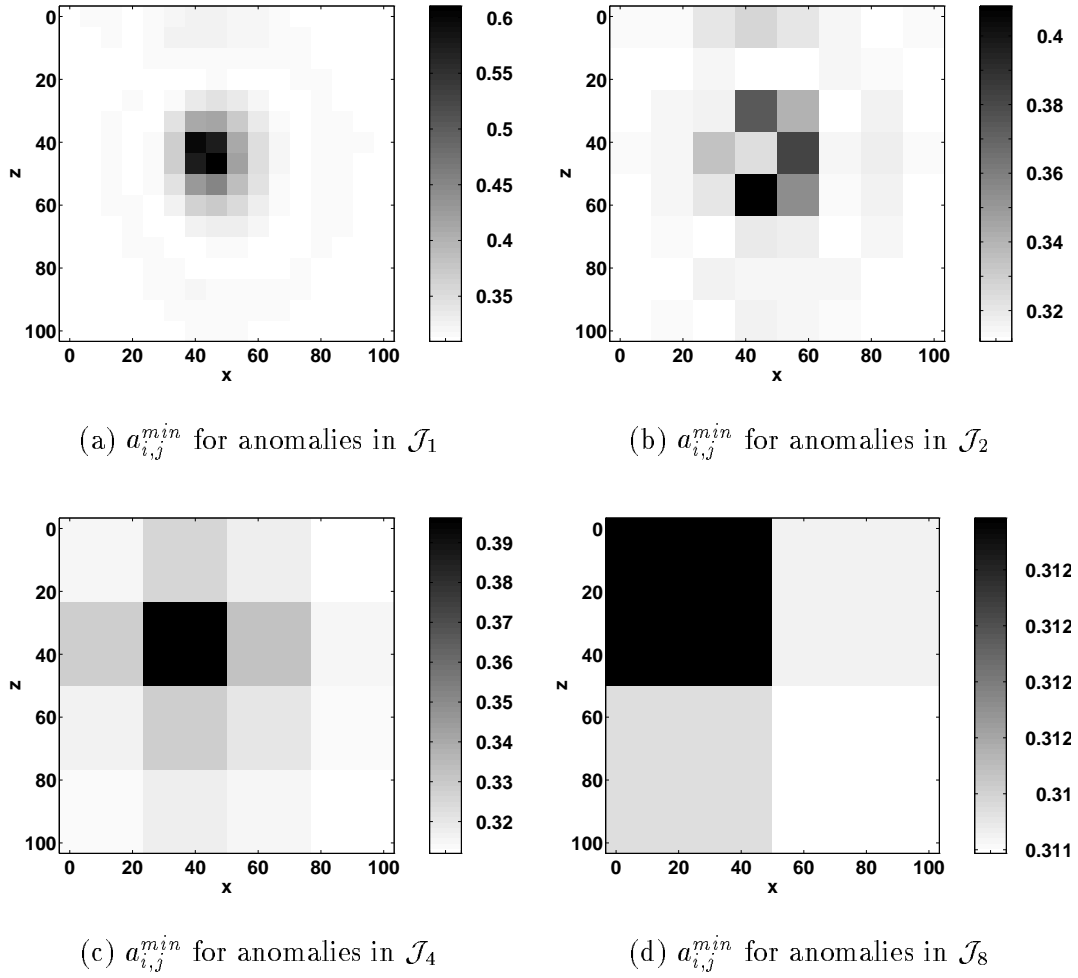


Figure 6: Images of the minimum magnitude of the anomaly in Figure 5(a) to guarantee a $P_d = 0.95$ and $P_f = 0.05$ in binary hypothesis tests involving this anomaly structure and elements of \mathcal{J} . Note that while the scales in these images are different the magnitudes are all less than 2.5.

\mathcal{B}_j) the performance of the BHT will fail to meet the $P_{d,i,j}$ and $P_{f,i,j}$ requirements for some range of a_j . Similar results hold for the second anomaly structure located closer to the left side except that in this case, the required values of the ABR are smaller.

Anomaly $\tilde{\gamma}_i$	Maximum ABR	Minimum ABR
Rightmost (Figure 5(a))	0.49	0.24
Leftmost (Figure 5(b))	2.11	0.56

Table 2: Minimum and maximum anomaly-to-background ratio associated with the smallest and largest values for $a_{i,j}^{min}$ for the anomaly structures in Figure 5(a) and 5(b).

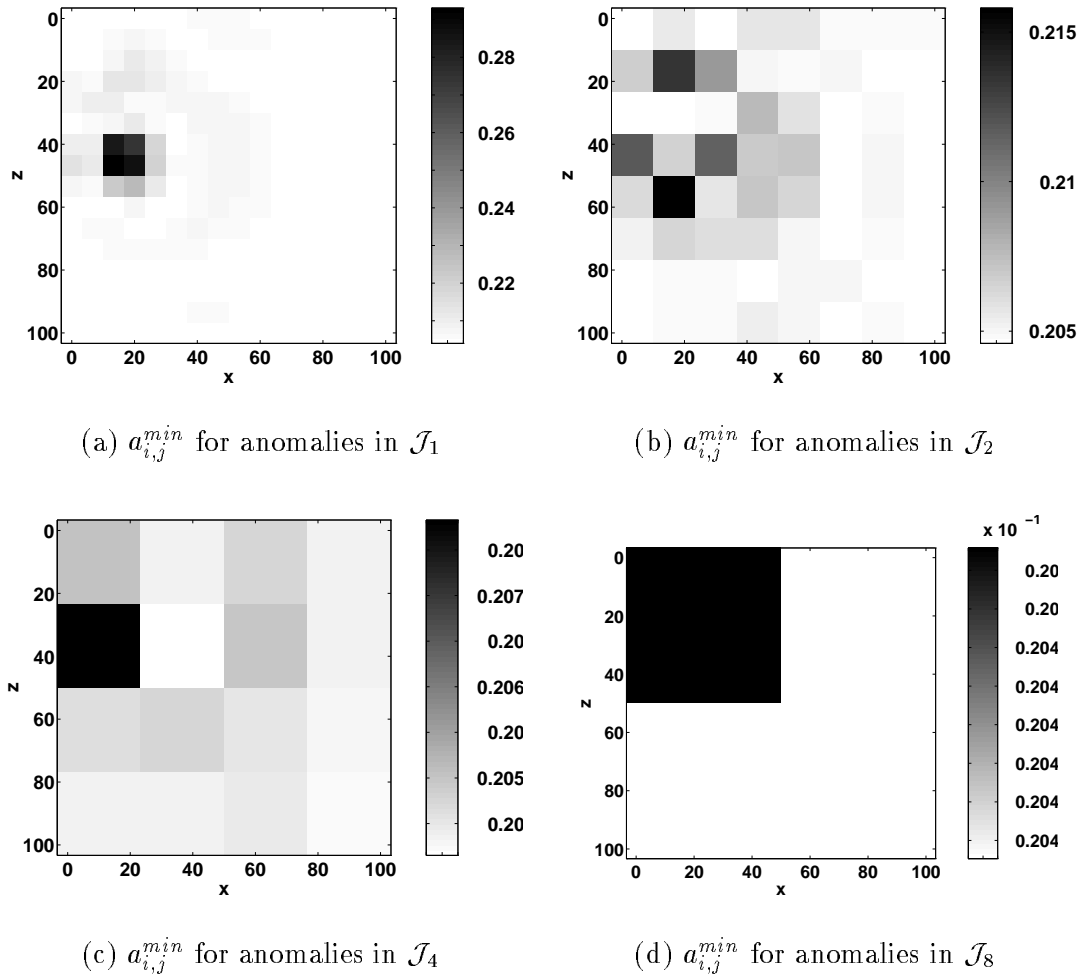


Figure 7: Images of the minimum magnitude of the anomaly in Figure 5(b) to guarantee a $P_d = 0.95$ and $P_f = 0.05$ in binary hypothesis tests involving this anomaly structure and elements of \mathcal{J} . Again, the scales in (a) through (d) are all different; however the overall range of values is between 0.9 and 1.3.

6 A Multiscale Algorithm for Anomaly Characterization

In this section we describe and analyze a multiscale, decision-theoretic algorithm to determine the positions, sizes and magnitudes of an unknown number of anomalous structures in region A . We begin with a small collection of relatively large rectangular areas in which anomalies *may* be located. Each region represents a top-level node in a tree of finer-scale subdivisions of A . We next use a decision-directed procedure for determining how best to move from one level of the tree, corresponding to a collection of coarse-scale hypotheses, to the next level in which anomalies are better localized using smaller-scale rectangles. The result of this procedure is a collection of rectangular areas of varying sizes and positions where we believe anomalies exist. To limit the number of targeted areas which contain no anomalies, the algorithm concludes with a pruning step where we also estimate the magnitudes of the final group of chosen anomaly structures.

6.1 A Scale Recursive, Decision Driven Detection Algorithm

The first step in our detection algorithm involves an M-ary Hypothesis test in which we consider 10 ways to subdivide A in order to better localize anomalous structures. As seen in Figure 8 the first configuration corresponds to the presence of a coarse scale anomaly with support over all of A . This particular structure indicates that *no* further decomposition is warranted. The next four possibilities each allows for a single anomaly localized to the top, bottom, left and right halves of A respectively. Because anomalies might lie both in the left/right as well as the top/bottom halves, the sixth and seventh structures in Figure 8 are included. Since multiple anomalies may be present in the region, the eighth configuration corresponds to the presence of one anomaly located in the left half and one in the right while the ninth presents the analogous situation but for the top and the bottom. Finally, for this initial decomposition only, we consider the last case where we conjecture that *no* anomalous regions exists in A .

Given the 10 choices in Figure 8, we formulate a 10-ary hypothesis testing problem the solution of which is obtained using the Generalized Likelihood Ratio Test (GLRT) discussed in Section 3.2. Using (17) we compute the values of the generalized log-likelihood function for each of the hypotheses under consideration. From Figure 8, if H_0 is chosen, no further decomposition occurs and we conclude that there is a single anomaly covering the entire region of interest. If H_9 is selected, the algorithm terminates with the conclusion that there is no anomaly in region A . Otherwise, we decompose that hypothesis with the largest generalized log-likelihood value.

Our scale-recursive decomposition of A continues by essentially repeating the hypothesis testing procedure for each of the subregions indicated by the initial 10-ary hypothesis test as being of interest. For example, consider the case where H_3 is chosen. Referring to Figure 8, this selection corresponds to an anomaly located in the left half of A . In an effort to better localize the anomalous activity in this region, we consider an M-ary hypothesis test similar to that described in the previous paragraph but where the underlying area involved in the decomposition is now the left half of A rather than all of A . While the subdivision is of a rectangular region as opposed to a square area, the form of the hypotheses fundamentally remains the same as in those displayed in Figure 8 in that we consider the possibilities of anomalies located in the top, bottom, left, and right halves, etc. of this long and thin structure. We note that the first of these nine hypotheses, H_0 , corresponds to the case where no further decomposition of the left half is warranted and thus serves as a means of terminating the scale recursive search over this region of A . Instead of ten, there are only nine hypotheses as we no longer include the possibility that no anomaly exists in the left half of A since the previous iteration indicated that *somewhere* in the left side there exists an anomaly.

This nine-hypothesis GLRT is repeated recursively beginning with the regions selected in the initial decomposition of A . This decision-theoretic localization process continues until no further subdivision in a particular region is warranted based upon the selection of the H_0 hypothesis at some stage of the process or because no addition refinement is possible because the structures under consideration are too small. Thus at the end of our scale-recursive decomposition of A we have a collection of rectangular regions where anomalous structures are likely to exist. We then collect the wavelet-domain representations of these rectangles as columns in a matrix labeled \mathcal{B}_{leaf} .

To limit the number of false alarms generated by our detection algorithm, we retain only those structures in \mathcal{B}_{leaf} corresponding to sufficiently “detectable” anomalies. Specifically, we begin computing \hat{a}_{leaf} , the amplitudes associated with \mathcal{B}_{leaf} , using (14) with \mathcal{B}_i replaced by \mathcal{B}_{leaf} . Next, for each column of \mathcal{B}_{leaf} , we calculate the minimum required amplitude to guarantee a set level of performance from a detectability-type hypothesis test developed in Section 4 (here we use $P_d = 0.80$ and $P_f = 0.10$). The final estimated anomaly structure generated by our algorithm is composed of

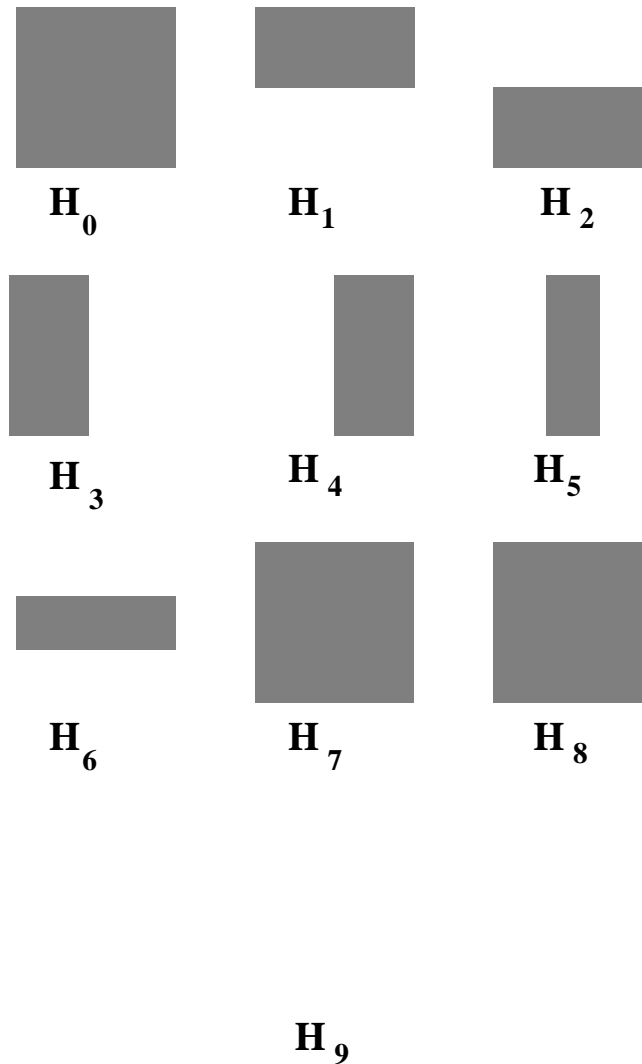


Figure 8: Geometric structures of nine possible decompositions used at each stage of our decomposition of A . The darkly shaded regions indicate the areas where anomalous structures are hypothesized to exist. While the figure illustrates the decomposition of a square region, analogous subdivision schemes are used for rectangular areas as well with the fundamental idea being the presence of anomalies in the top, bottom, left, right, etc.

those columns of \mathcal{B}_{leaf} and elements of a_{leaf} corresponding to anomalies whose amplitudes exceed this required minimum and we label these estimates $\hat{\mathcal{B}}$ and \hat{a} respectively.

6.2 Algorithm Analysis

The scale-recursive detection algorithm described in Section 6.1 requires that we be able to identify successfully large-scale structures covering the true, smaller-scale anomalies. The results of the distinguishability analysis suggest that the correct large-scale structures are likely to be selected. Indeed, Figures 6 and 7 showed that the largest values of $a_{i,j}^{min}$ corresponded to those j in \mathcal{J} which overlap anomaly i . From this, we conclude that small-scale anomalies “look” most like those large-scale counterparts located in the same region of A .

To further verify this intuition, we undertake a more detailed performance analysis of the GLRT used in the detection algorithm. Specifically, we consider the case where a single anomaly, \bar{g}^* , of unknown amplitude exists at some fine scale and we perform a generalized binary hypothesis in which the two hypotheses correspond to coarse scale structures one of which covers \bar{g}^* and one of which does not. We are interested in examining how the probability of correctly choosing the overlapping structure (which we call the probability of detection for these experiments) using the GLRT of Section 3.2 varies with the scale and position of the non-overlapping alternate as well as the amplitude of the true anomaly. High detection probabilities reflect favorably on the GLRT-based approach of the scale-recursive algorithm.

Following the notation of (17), let $l_1(\eta)$ be the statistic associated with the overlapping anomaly hypothesis and $l_0(\eta)$ be the statistic for the non-overlapping case. From (15) and (16), the probability of choosing the overlapping structure given knowledge of $\bar{\gamma}^* = \mathcal{W}_g \bar{g}^*$ is

$$\text{Prob}[L_1(\eta) > 0 | \bar{\gamma}^*] = \text{Prob}[l_1(\eta) - l_0(\eta) > 0 | \bar{\gamma}^*]. \quad (29)$$

Upon substituting (14) into (16) and using (17), straightforward linear algebra demonstrates that the random variable $L_1(\eta)$ may be written as

$$L_1(\eta) = x_1^2(\eta) - x_0^2(\eta) \quad (30)$$

where the two-vector $x(\eta) = [x_1(\eta) \ x_0(\eta)]^T$ is

$$x(\eta) = \mathcal{B}_{10}^T \Theta^T P_\eta^{-1} \eta \sim \mathcal{N}(\mathcal{B}_{10}^T \Theta^T P_\eta^{-1} \Theta \bar{\gamma}^*, \mathcal{B}_{10}^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_{10}) \quad (31)$$

and for $j = 0, 1$

$$\mathcal{B}_{10} = [s_1 \mathcal{B}_1 \ s_0 \mathcal{B}_0] \quad (32a)$$

$$s_j^2 = \frac{1}{2} [P_j(1 + \alpha P_j)] \quad (32b)$$

$$P_j = (\mathcal{B}_j^T \Theta^T P_\eta^{-1} \Theta \mathcal{B}_j + \alpha)^{-1} \quad (32c)$$

From (30), $\text{Prob}[L_1(\eta) > 0 | \bar{\gamma}^*] = \text{Prob}[|x_1(\eta)| > |x_0(\eta)| | \bar{\gamma}^*]$ which is the integral of the probability density function for $x(\eta)$ defined in (31) over the shaded region in Figure 9.

In Figure 10, detection probabilities are displayed for binary hypothesis tests where \bar{g}^* is the structure in Figure 5(a) and the hypotheses are pairs of structures from \mathcal{J} . For example, the shade of dark region in the lower left corner in Figure 10(a) is $\text{Prob}[L_1(\eta) > 0 | \bar{\gamma}^*]$ for the BHT where the first hypothesis is the large structure overlapping the true, smaller size anomaly (represented by the white region in Figure 10(a)) and the alternate hypothesis is the 8×8 pixel lower left corner of A . Similar interpretations hold for the other two dark areas in Figure 10(a) and for each of the smaller square areas in Figures 10(b)–(c). For all of these images, the ABR for the true, small anomaly is set to 1.5. Figures 10(a)–(c) indicate the manner in which the detection performance of the GLRT-based algorithm depends upon the scale of the hypotheses relative to that of the true anomaly. At the coarsest scale, detection probabilities are about 60%. However, for all finer scales, P_d rises sharply with the lowest values confined to structures which are close to the true anomaly.

In Figure 10(d), we display the minimum P_d at each scale as a function of true anomaly's ABR. For example, the points on each of the three curves at an ABR of 1.5 are the minimum P_d values in each of the three images in Figure 10(a)–(c). From these curves we see that at the coarsest scale, even at high ABRs, the detection probabilities reach about 80%. As expected, when the hypotheses are drawn from the finer scales, the minimum P_d rises quickly to close to 100%.

The results in Figures 10 indicate that if the scale-recursive anomaly detection algorithm developed in Section 6.1 correctly identifies the coarse scale structures overlapping the true anomalies, then the detection performance at finer scales should be quite good even at ABRs less than 1. Also, because the *lowest* detection probabilities at fine scales are associated with structures close to the true structure, it is anticipated that the scale-recursive detection algorithm should be very successful in producing estimates of anomalies which are “sufficiently close” to the truth if not exactly the truth. This idea will be made more precise in Section 6.3.

The analysis in this section indicates that the primary difficulty associated with the algorithm is

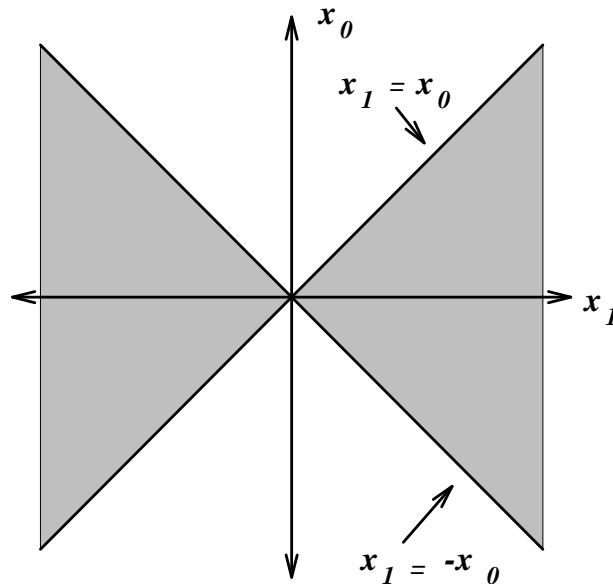


Figure 9: Integration region in $x_1 - x_0$ space for evaluation of $\text{Prob}[L_1(\eta) > 0 | \bar{\gamma}^*]$ in (29)

that coarse scale detection probabilities can be low. To overcome the potential problem of selecting the wrong area or areas of A for further refinement at coarse scales we modify the scale-recursive algorithm in the following manner. At the opening stage of the algorithm, rather than accepting the *single* hypothesis with the largest generalized log-likelihood value, we consider further refinement of A based upon those hypotheses corresponding to the *four* largest log-likelihood values (excluding H_0 and H_9). As will be seen in Section 6.3, despite the additional computational requirements of this approach, the overall complexity of the algorithm remains rather low. Finally, we note that one could extend this strategy of keeping additional structures for further refinement to more than just the first stage of the algorithm and could retain fewer or greater than four alternatives; however for the application of interest here, the choices described above were sufficient.

6.3 Examples

In this section, we examine the performance of the scale-recursive algorithm described in Sections 6.1 and 6.2. First, we use Monte Carlo studies to verify the ability of this approach to detect anomalous structures. The quantities of interest here are the sample probability of detection, \bar{P}_d , the sample average value of the number of false alarms per pixel \bar{P}_f , and the sample probability of error, \bar{P}_e . We say that a particular rectangular anomaly, $\bar{\gamma}^*$, has been detected if there exists a column in \hat{B} which is sufficiently close to $\bar{\gamma}^*$. Specifically, we define a “region of ambiguity” associated with the anomaly structure currently under investigation. This area is constructed such that anomaly structures identified in this region are “essentially indistinguishable” from the true anomaly. More formally, we compute the probability of successfully distinguishing $\bar{\gamma}^*$ from each member of \mathcal{J} in a binary hypothesis test of the form in (28a)–(28b). For each such test, the amplitudes of the two anomalies are chosen so that relative to the anomaly-free background, the two structures are equally detectable (i.e. they individually have the same d^2 value as defined with $P_d = 0.85$ and $P_f = 0.10$ in (18) and (26).) A pixel in A is said to be in the ambiguity region if (1) there exists a member of \mathcal{J} which is nonzero on that pixel and (2) the probability of distinguishing that element of \mathcal{J} from $\bar{\gamma}^*$ is below a given threshold, taken as 0.85 for all problems considered in

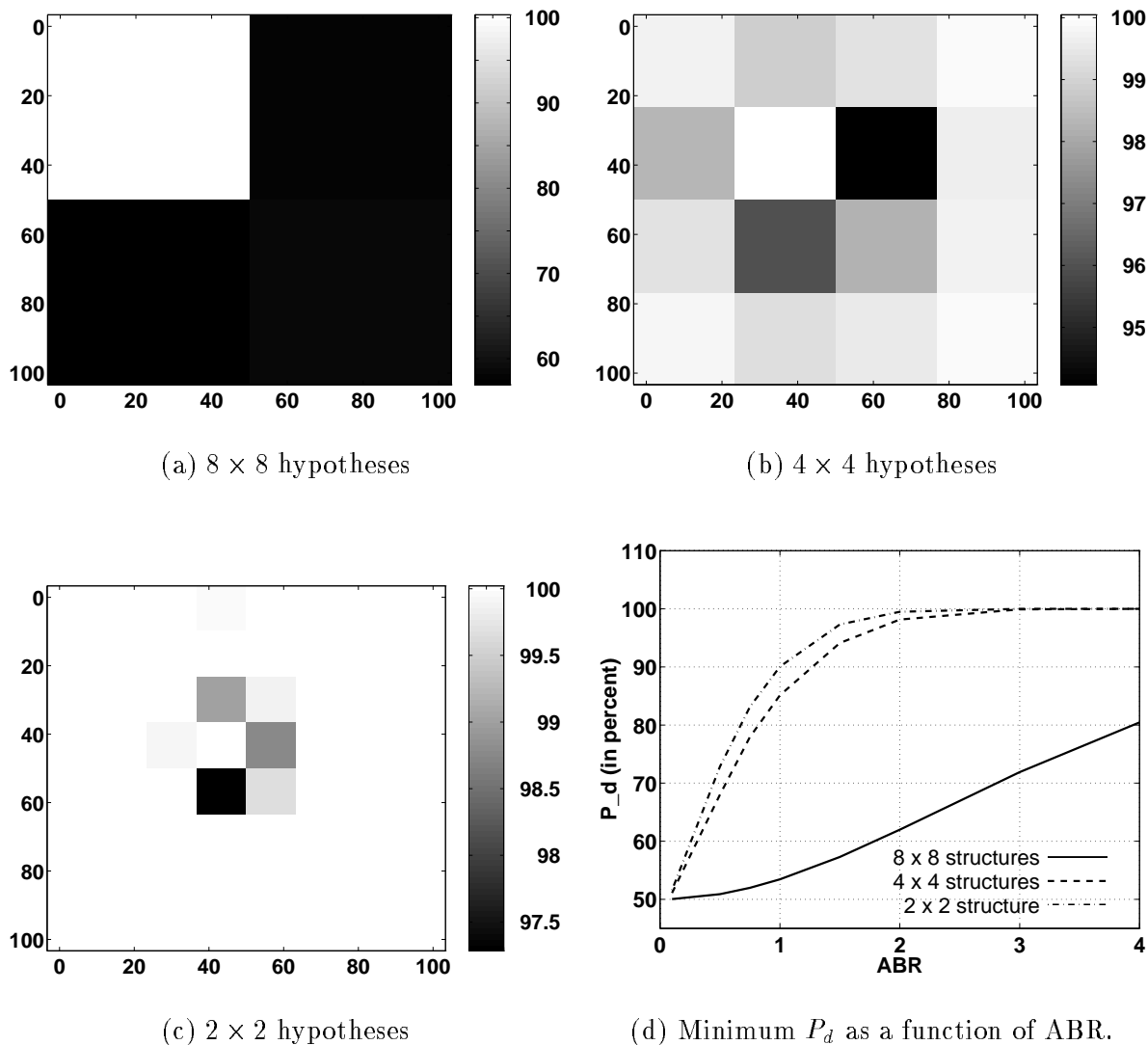


Figure 10: In (a)–(c) detection probabilities are displayed for binary hypothesis tests where \bar{g}^* is the structure in Figure 5(a) and the hypotheses are pairs of structures from \mathcal{J} . For each such test, one of the hypotheses is a larger scale structure overlapping \bar{g}^* while the second structures is from the same scale as the first but is disjoint from the pixels of \bar{g}^* . The shade of each square in (a)–(c) is the probability of correctly choosing the overlapping structure when the alternate is the anomaly occupying the square under investigation. The ABR for the true structure is 1.5. The minimum P_d at each scale as a function of true anomaly’s ABR is shown in (d).

this section. Finally, for an estimated structure to be called a detection the area of intersection between it and the region of ambiguity must be at least a quarter of the area of the estimated structure. Such a definition implies a constraint on the localization of an estimated anomaly in both space and scale before we will call it a detection. As an example, the region of ambiguity at $P_d = 0.85$ associated with the anomaly structure in Figure 5 is displayed in Figure 11. The elements of $\hat{\mathcal{B}}$ which do not correspond to detections are taken to be false alarms and the per-pixel false alarm rate, \bar{P}_f , is defined as the total number of false alarm pixels divided by the number of

pixels in region A . Finally, the sample probability of error is $\bar{P}_e = 1 - \bar{P}_d + \bar{P}_f$.

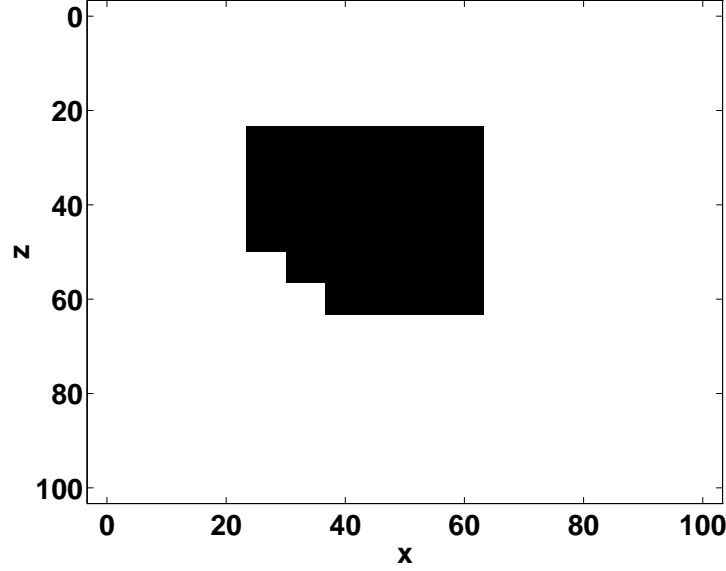


Figure 11: Region of ambiguity for structure shown in Figure 5 for $P_d = 0.85$.

We also examine the computational complexity of the scale-recursive algorithm. The complexity of the algorithm is quantified in terms of the number of Generalized Likelihood Ratio Tests (GLRTs) which must be performed in the processing of the data. As the spatial decomposition of region A is driven by the noisy data, the number of GLRTs will vary from one data set to the next. Thus, for a particular \bar{g}^* , the computational performance is based upon the average number of required GLRTs required per iteration of the corresponding Monte-Carlo.

Finally as discussed in Section 1, the detection algorithm results are used to improve the solution to the full reconstruction inverse problem. From our model for γ in (6), the estimate of the overall conductivity is the sum of the estimates of $\bar{\gamma}$ and $\tilde{\gamma}$, denoted $\hat{\gamma}$ and $\hat{\tilde{\gamma}}$ respectively, where $\hat{\gamma} = \hat{\mathcal{B}}\hat{a}$ is provided by our scale-recursive detection algorithm. Now, the linear least-squares estimate (LLSE) of $\tilde{\gamma}$ developed in [37,40] is based upon the assumption that *no* anomalies exist in the data; however, the output of the detection algorithm provides additional information through $\hat{\tilde{\gamma}}$ as to the structure of the conductivity field. To make use of the information in order to improve the estimate $\hat{\gamma}$, we define $\hat{\tilde{\gamma}}_c$ as the LLSE of $\tilde{\gamma}$ based upon a “corrected” data set in which the effects of $\hat{\tilde{\gamma}}$ have been removed. Mathematically this corrected estimate takes the form

$$\hat{\tilde{\gamma}}_c = P\Theta^T R^{-1} [\eta - \Theta\hat{\mathcal{B}}\hat{a}] \quad (33)$$

where $P = (\Theta^T R^{-1} \Theta + P_0^{-1})^{-1}$ is the error covariance matrix for nominal LLSE. Thus, the estimate of the overall conductivity field is

$$\hat{\gamma} = \hat{\tilde{\gamma}}_c + \hat{\tilde{\gamma}} = P\Theta^T R^{-1} \eta + [I - P\Theta^T R^{-1} \Theta] \hat{\mathcal{B}}\hat{a} \quad (34)$$

where we recognize the term $P\Theta^T R^{-1} \eta$ as the uncorrected LLSE estimate [49].

Unless otherwise stated, the data upon which the examples are based are generated using the Born-based measurements model in (2) for the scattering experiments described in Table 1. For all cases consider, the background conductivity, g_0 , is set to 1 S/m and at the highest ABRs of interest, the anomaly amplitudes are only 0.7 S/m. As discussed in [25], under these circumstances the Born approximation is known to be valid. In Section 7, we discuss issues associated with extending the work in this paper to account for the underlying non-linearity associated with the inverse conductivity problem. Finally, for all experiments the parameter α in (14) is set to 0.25.

6.3.1 The Single Anomaly Case

We begin by considering the case where it is known that there is a single anomaly of unknown amplitude and location in region A . Given that there is only one structure, the combinatorial complexity associated with an “exhaustive search” for the anomaly is sufficiently low that we shall compare both the detection/false-alarm performance as well as the complexity of the scale-recursive approach against an alternate algorithm akin to a multi-scale matched filter. This algorithm detects the single anomaly by computing the GLRT for each of the structures in family \mathcal{J} taking that element of \mathcal{J} associated with the largest GLRT statistic as the estimate. Because this method is multiscale in nature and has a fixed number of GLRTs per Monte-Carlo iteration (since there are a fixed number of structures in \mathcal{J}) it allows for a fair comparison against which we can judge the performance of the scale-recursive algorithm. For the scale-recursive method, we shall account for the knowledge that there is only a single anomaly in A by retaining only the column of \mathcal{B}_{leaf} associated with the most likely anomaly structure. Finally, for this example, the true anomaly structure is shown in Figure 5 and the SNR for all scattering experiments is 10.

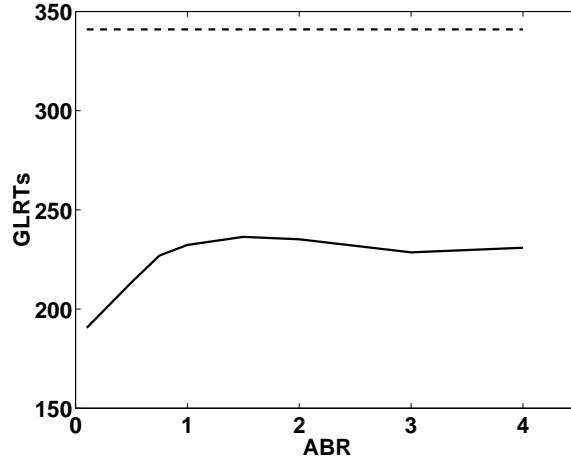
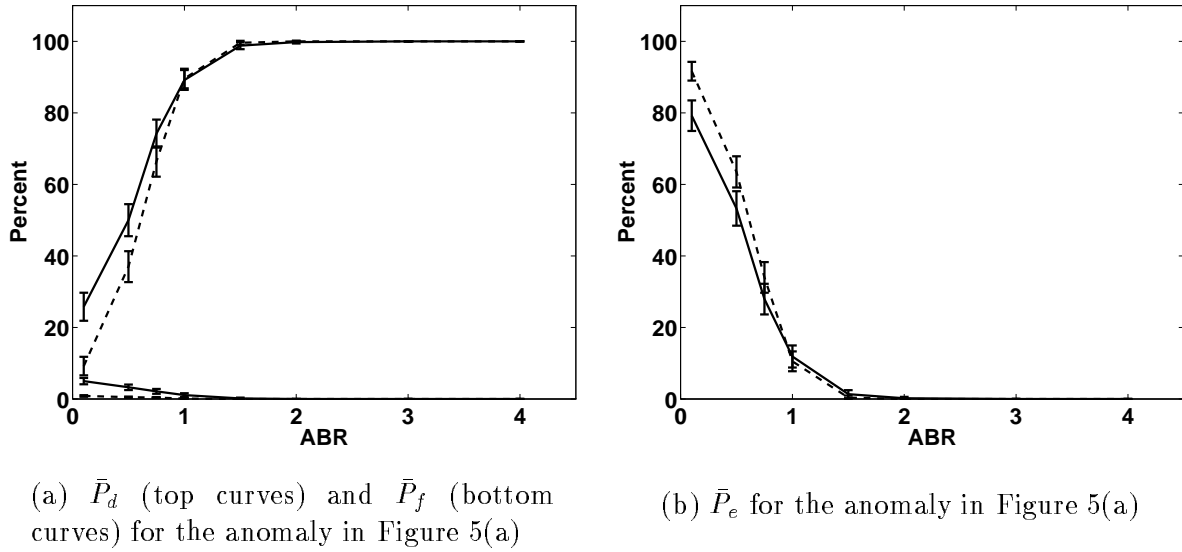
In Figure 12 we show \bar{P}_d , \bar{P}_f , \bar{P}_e and the average number of GLRTs per Monte-Carlo iteration as a function of anomaly-to-background ratio obtained after 500 Monte-Carlo iterations. The solid lines are the results for scale-recursive algorithm and the dashed lines indicate the performance of the multi-scale exhaustive search procedure. Figure 12(a) indicates that at low ABRs, the scale-recursive approach tends to have a higher detection probability than the exhaustive search with a slightly higher probability of false alarm. Even for the low ABR of 0.50, \bar{P}_d is well above 50% and rises to above 90% for ABR values greater than one. At high ABRs the performance of the two algorithms is about the same. Despite the slightly higher \bar{P}_f of the scale-recursive approach, the overall error probability is lower for the scale-recursive method at these small ABRs. Finally, from 12(c) the computational complexity of the scale-recursive characterization algorithm is seen to be roughly constant across the ABR range at 65% that of the exhaustive search.

In Figure 13(a) we display one realization of $g = \bar{g} + \tilde{g}$ obtained in our Monte Carlo process at an ABR of 1.5. Using the LLSE to perform the full reconstruction as in [40] results in the image in Figure 13(b). By incorporating the results of the scale-recursive detection algorithm into the inversion procedure through the use of (34), we obtain the estimate of the overall conductivity field shown in 13(c). Thus, successful identification of the highly parameterized anomaly structures can significantly improve localization both in space and scale and the GLRT procedure results in an accurate estimate of the structure’s amplitude. Also, the details in the remainder of the estimate do in fact reflect the coarse scale, fractal features of the conductivity profile in Figure 13(a).

6.3.2 The Multiple Anomaly Case

We now turn our attention to the case where multiple anomalies exist in region A .⁴ Lifting the single anomaly assumption causes the computational complexity of an exhaustive-search-type of approach to be prohibitive in that one would be required to examine the likelihood of all combinations of all non-overlapping, structures in a collection such as \mathcal{J} assuming separately $n = 1$ then $n = 2$ through $n = N_{max}$ anomalies exist in region A where N_{max} is a pre-determined maximum number of anomalies. Thus, here we present only the results of the scale-recursive detection algorithm. In particular, we explore the performance for the anomaly configuration in Figure 14.

⁴Note that in this multi-anomaly case, the ABR is used to determine the magnitude of each structure individually. For example at an ABR of one, the amplitude of the left anomaly is set so that if it were the only structure in the medium, the ABR would be one.

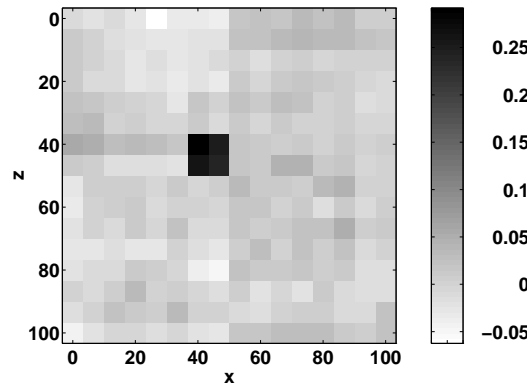


(c) Average number of GLRTs per Monte-Carlo iteration for anomaly detection algorithms for the anomaly in Figure 5(a).

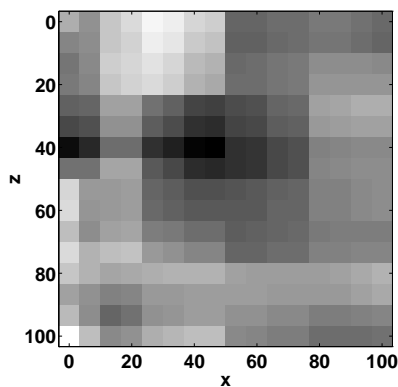
Figure 12: Performance curves as a function of ABR obtained after 500 Monte-Carlo iterations for the anomaly in Figure 5(a). Solid lines = results for scale-recursive algorithm. Dashed lines = results for multi-scale exhaustive search. The error bars are drawn at the plus/minus two standard deviation level.

The Monte-Carlo results for this experiment are displayed in Figure 15 where the top two curves of (a) correspond to the individual \bar{P}_d statistics for the two anomalies and the lowest of the three curves is a plot of \bar{P}_f . Here we see that both structures are quite easily detected with a \bar{P}_d of well over 90% even at the low ABR of one. As is expected, removing the single-anomaly assumption causes the algorithm to retain a greater number of candidate structures (including the true anomalies) thereby raising \bar{P}_f above that seen in Section 6.3.1.

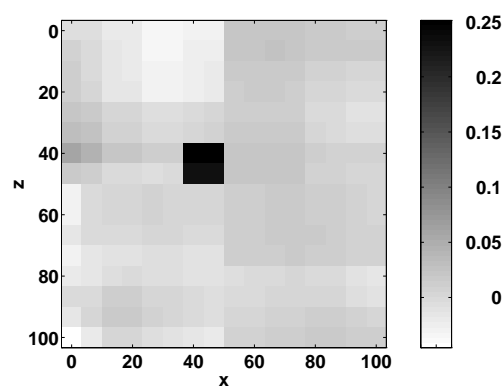
In Figure 15(c) we plot the average number of GLRTs as a function of ABR. Note that at



(a) Anomaly in fractal background
(ABR=1.5)



(b) Reconstruction of conductivity in
(a) using LLSE



(c) Reconstruction of conductivity
in (a) using (34)

Figure 13: Comparison of reconstructed conductivity profile using the LLSE of [40] and an estimate based upon the output of the scale-recursive anomaly detection algorithm. The true conductivity is shown in (a) and contains a single anomaly near the center of the region. The LLSE is shown in (b) and the estimate obtained from (34) is illustrated in (c). Here we see that the use of the information from the detection algorithm allows for the successful localization of the anomaly in space and scale without sacrificing our ability to resolve the fractal features of the conductivity profile in (a). Additionally, the GLRT procedure results in an accurate estimate of the anomaly's amplitude.

worst the complexity of this algorithm is still well below the complexity of the single-anomaly exhaustive search algorithm and only about 30% greater than the complexity of the single-anomaly scale-recursive algorithm. Thus, despite the fact that the multiple anomaly problem is, from a combinatorial viewpoint, significantly more complex than the single anomaly case, we see that the scale-recursive localization method represents a highly efficient and accurate means of localizing an unknown number of structures in the region of interest.

In Figure 16, we compare the full reconstruction results obtained from the LLSE to those where (34) is used to estimate the underlying conductivity for one run of the Monte-Carlo at an ABR of

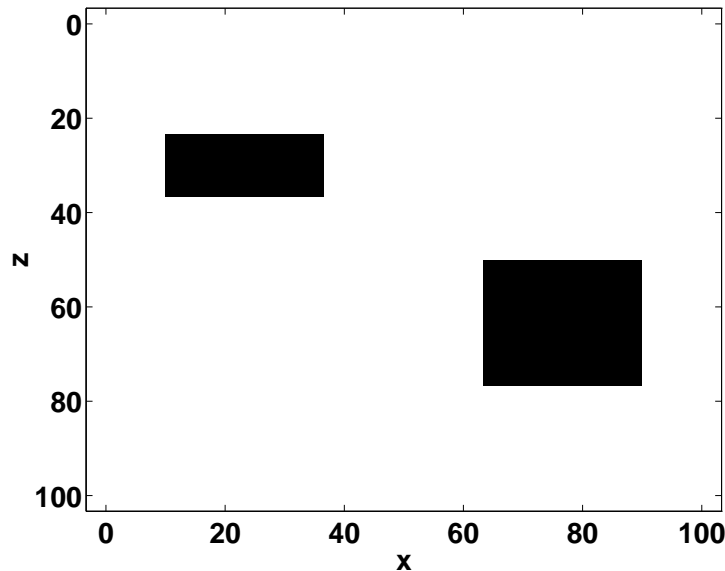


Figure 14: Two-region anomaly structure

1.5. From Figure 16(b) we see that the LLSE is successful in reconstructing the structure on the left; however, the lower amplitude/more pixel anomaly is almost completely undetected. Figure 16(c) indicates that the incorporation of the information from the anomaly detection algorithm significantly improves the localization in space as well as scale of both anomaly structures, especially the rightmost. Finally, the anomaly amplitudes are better estimated using the GLRT method.

7 Conclusion and Future Work

In this paper, we have presented a framework based upon techniques from the areas of multiscale modeling, wavelet transforms, and statistical decision and estimation theory for addressing a variety of issues arising in anomaly detection problems. Beginning with a linear model relating the data and the quantity to be reconstructed, we use the wavelet transform to take the problem from physical space to scale space where computational complexity is reduced for a wide variety of problems [1, 4, 41] and where we are able to take advantage of the rich and useful class of models recently developed for describing the structure of the medium in the absence of anomalous activity [21, 35, 47, 50]. The problems of characterizing the number, positions, and magnitudes of anomaly structures was formulated using the tools of statistical decision theory. To understand how the physics of the problem and the constraints on the geometry of the data collection process affect our ability to isolate anomalous regions, we defined and explored the issues of anomaly detectability and distinguishability. This analysis led to the development of a scale-recursive algorithm employing a sequence of Generalized Likelihood Ratio Tests for extracting anomaly information from data.

This work was presented in the context of a linearized inverse scattering problem arising in geophysical prospecting. The same scattering model is encountered in a variety of other fields where some form of energy is used to probe a lossy medium [18–20, 22, 29, 31]. More generally, the analysis and algorithmic methods developed in this work require only a measurements model of the form in (2) and are thus relevant for any linear inverse problem (eg. computed tomography) in which anomaly characterization is of interest.

An important extension of the work presented here is in development of algorithms and analysis

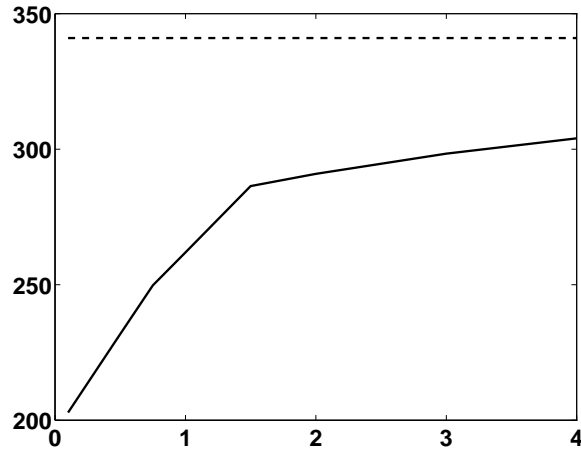
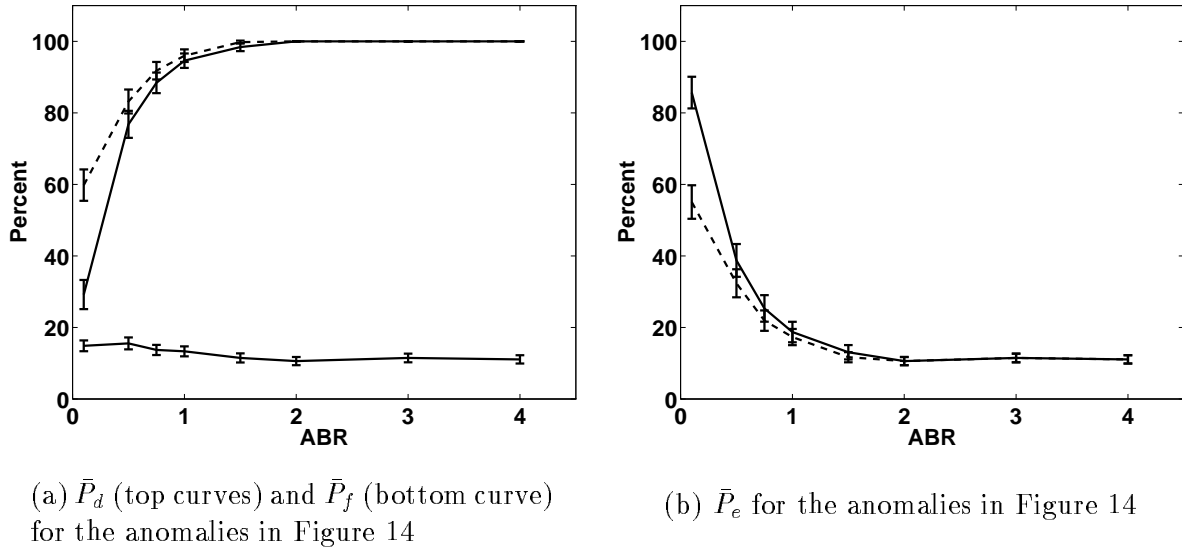


Figure 15: Performance curves obtained after 500 Monte-Carlo iterations of scale-recursive detection algorithm for the anomalies in Figure 14. Solid lines in (a) and (b) are detection and error probabilities for the upper left anomaly while dashed lines are for lower right anomaly. The error bars are drawn at the plus/minus two standard deviation level. In (c), the computational complexity associated with this scenario is shown by the solid line. For comparison, the dashed line is the complexity associated with the single anomaly exhaustive search.

methods for detecting anomaly structures using the *nonlinear* physics governing the relationship between the conductivity and the observed scattered electric field. The primary difficulty here is maintaining or improving the detection/false-alarm performance of the current method while retaining the low computation complexity in an algorithm based upon a significantly more complex scattering model. In [38] we present preliminary results for one form of a scale-recursive anomaly

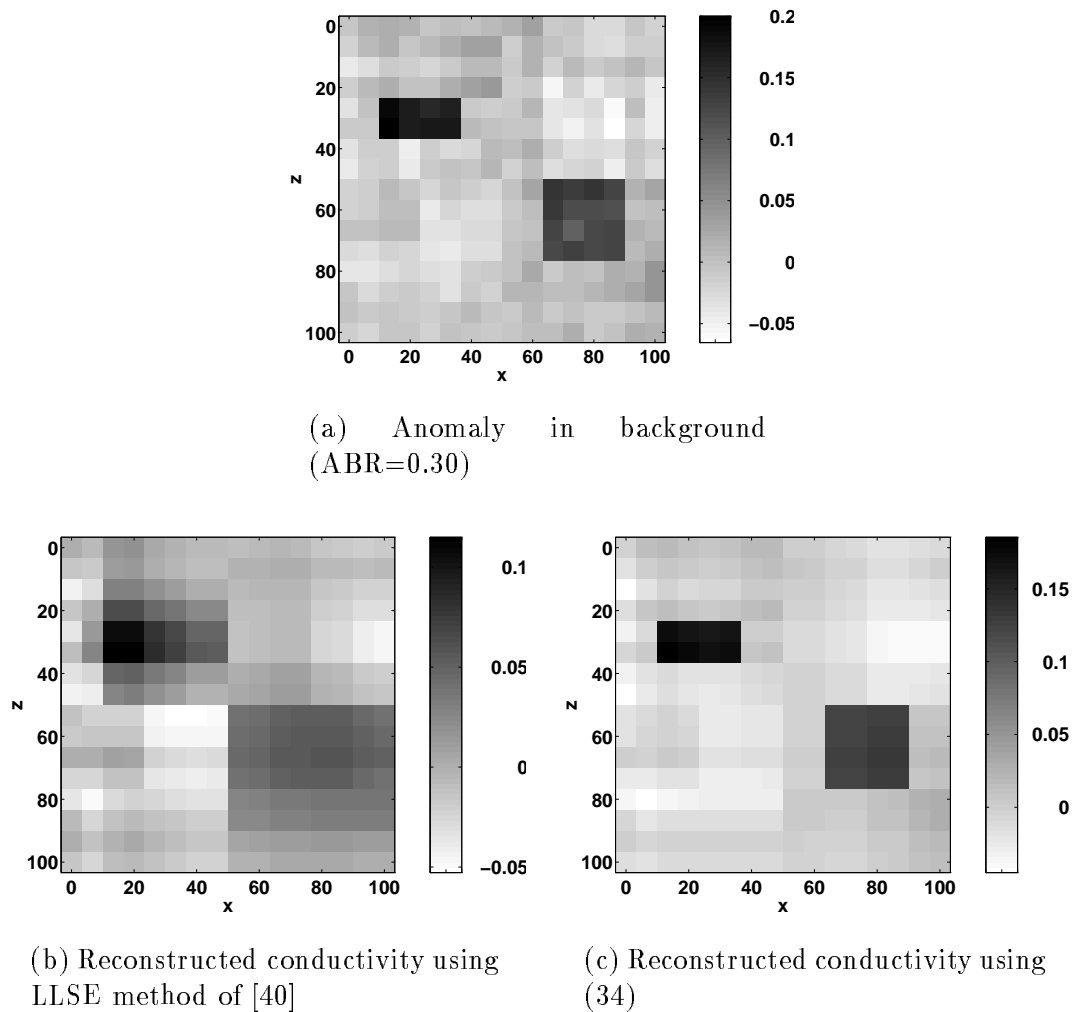


Figure 16: Comparison of reconstructed conductivity profile using the LLSE of [40] and an estimate based upon the output of the scale-recursive anomaly detection algorithm. The true conductivity is shown in (a) and contains a two anomalies. The LLSE is shown in (b) and the estimate obtained from (34) is illustrated in (c). Here we see that the use of the detection information allows for the successful localization of both anomaly structures and offers a significant improvement over the LLSE in localizing the anomaly in the lower right.

characterization algorithm using the computationally efficient, nonlinear Extended Born Approximation [48]. Further work remains in the exploration of detectability and distinguishability in the nonlinear context and the extension of this approach to higher frequency (eg. microwave) problems.

Another avenue of research is in the use of alternate methods for progressively dividing the region of interest. The problem of anomaly detection is similar to that of image segmentation in that the goal of both is to partition a two-dimensional grid of pixels into disjoint regions. The primary difference between these two problems is the data. In the segmentation case the data are the pixels in the image whereas we wish to do the anomaly localization given the significantly less informative observations of scattered radiation. For the segmentation problem, significant work has been performed in the use of hierarchical methods for performing this decomposition. For example,

segmentation techniques have been developed where (a) small structures are merged into larger regions [36, 42] and (b) both splitting as well as merging operations are used in the segmentation process [28]. Examining the utility of merge- and split/merge-algorithms for the anomaly detection problem would be of considerable interest especially as a means of overcoming the difficulties of detecting small-scale structures using large scale hypotheses.

References

- [1] B. Alpert, G. Beylkin, R. Coifman, and V. Rokhlin. Wavelets for the fast solution of second-kind integral equations. *SIAM J. on Scient. Comput.*, 14(1):159–184, 1993.
- [2] M. Bertero, C. De Mol, and E. R. Pike. Linear inverse problems with discrete data. I: General formulation and singular system analysis. *Inverse Problems*, 1:301–330, 1985.
- [3] M. Bertero, C. De Mol, and E. R. Pike. Linear inverse problems with discrete data. II: Stability and regularisation. *Inverse Problems*, 4:573–594, 1988.
- [4] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Communications on Pure and Applied Mathematics*, 44:141–183, 1991.
- [5] Yoram Bresler, Jeffrey A. Fessler, and Albert Macovski. A Bayesian approach to reconstruction from incomplete projections of a multiple object 3D domain. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(8):840–858, November 1989.
- [6] Kevin E. Brewer and Stephen W. Wheatcraft. Including multi-scale information in the characterization of hydraulic conductivity distributions. In Efi Foufoula-Georgiou and Praveen Kumar, editors, *Wavelets in Geophysics*, volume 4 of *Wavelet Analysis and its Applications*, pages 213–248. Academic Press, 1994.
- [7] Stephen R. Brown. Transport of fluid and electric current through a single fracture. *Journal of Geophysical Research*, 94(B7):9429–9438, July 1989.
- [8] W.C. Chew and Y. M. Wang. Reconstruction of two-dimensional permittivity distribution using the distorted born iterative method. *IEEE Trans. Medical Imaging*, 9(2):218–225, June 1990.
- [9] Weng Cho Chew. *Waves and Fields in Inhomogeneous Media*. Van Nostrand Reinhold, New York, 1990.
- [10] Kenneth C. Chou, S.A. Golden, and Alan S. Willsky. Multiresolution stochastic models, data fusion and wavelet transforms. Technical Report LIDS-P-2110, MIT Laboratory for Information and Decision Systems, May 1992.
- [11] Kenneth C. Chou, Alan S. Willsky, and Ramine Nikoukhah. Multiscale recursive estimation, data fusion, and — regularization. *IEEE Trans. Automatic Control*, 39(3):464–478, March 1994.
- [12] Kenneth C. Chou, Alan S. Willsky, and Ramine Nikoukhah. Multiscale systems, Kalman filters, and Riccati equations. *IEEE Trans. Automatic Control*, 39(3):479–492, March 1994.

- [13] David J. Crossley and Oliver G. Jensen. Fractal velocity models in refraction seismology. In Christopher H. Scholtz and Benoit B. Mandelbrot, editors, *Fractals in Geophysics*, pages 61–76. Birkhauser, 1989.
- [14] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–996, 1988.
- [15] A. J. Devaney. Geophysical diffraction tomography. *IEEE Trans. on Geoscience and Remote Sensing*, GE-22(1):3–13, January 1984.
- [16] A. J. Devaney and G. A. Tsihrintzis. Maximum likelihood estimation of object location in diffraction tomography. *IEEE Trans. ASSP*, 39(3):672–682, March 1991.
- [17] A. J. Devaney and G. A. Tsihrintzis. Maximum likelihood estimation of object location in diffraction tomography, part II: Strongly scattering objects. *IEEE Trans. ASSP*, 39(6):1466–1470, June 1991.
- [18] David C. Dobson. Estimates on resolution and stabilization for the linearized inverse conductivity problem. *Inverse Problems*, 8:71–81, 1992.
- [19] David C. Dobson and Fadil Santosa. An image-enhancement technique for electrical impedance tomography. *Inverse Problems*, 10:317–334, 1994.
- [20] David C. Dobson and Fadil Santosa. Resolution and stability analysis of an inverse problem in electrical impedance tomography: Dependence on the input current patterns. *SIAM J. Appl. Math.*, 54(6):1542–1560, December 1994.
- [21] Patrick Flandrin. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Information Theory*, 38(2):910–917, March 1992.
- [22] D. G. Gisser, D. Isaacson, and J. C. Newell. Electric current computed tomography and eigenvalues. *SIAM J. Appl. Math.*, 50(6):1623–1634, December 1990.
- [23] T. M. Habashy, W. C. Chew, and E. Y. Chow. Simultaneous reconstruction of permittivity and conductivity profiles in a radially inhomogeneous slab. *Radio Science*, 21(4):635–645, July–August 1986.
- [24] Tarek M. Habashy, Edward Y. Chow, and Donald G. Dudley. Profile inversion using the renormalized source-type integral equation approach. *IEEE Transactions on Antennas and Propagation*, 38(5):668–682, May 1990.
- [25] Tarek M. Habashy, Ross W. Groom, and Brian R. Spies. Beyond the Born and Rytov approximations: A nonlinear approach to electromagnetic scattering. *Journal of Geophysical Research*, 98(B2):1759–1775, February 1993.
- [26] R. F. Harrington. *Field Computations by Moment Methods*. Macmillan Publ. Co., 1968.
- [27] Jorg H. Hippler, Helmut Ermert, and Ludwig von Bernus. Broadband holography applied to eddy current imaging using signals with multiplied phases. *Journal of Nondestructive Evaluation*, 12(3):153–162, 1993.
- [28] Steven L. Horowitz and Theodosios Pavlidis. Picture segmentation by a tree traversal algorithm. *Journal of the ACM*, 23(2):368–388, April 1976.

- [29] David Isaacson. Distinguishability of conductivities by electrical current computed tomography. *IEEE Trans. on Medical Imaging*, MI-5(2):91–95, June 1986.
- [30] David Isaacson and Margaret Cheney. Current problems in impedance imaging. In David Colton, Richard Ewing, and William Rundell, editors, *Inverse Problems in Partial Differential Equations*, chapter Chapter 9, pages 141–149. SIAM, 1990.
- [31] David Isaacson and Margaret Cheney. Effects of measurement precision and finite numbers of electrodes on linear impedance imaging algorithms. *SIAM J. Appl. Math.*, 51(6):1705–1731, December 1991.
- [32] Dwight L. Jaggard. On fractal electrodynamics. In H. N. Kritikos and D. L. Jaggard, editors, *Recent Advances in Electromagnetic Theory*, pages 183–224. Springer-Verlag, 1990.
- [33] Jonathan M. Lees and Peter E. Malin. Tomographic images of p wave velocity variation at Parkfield, California. *Journal of Geophysical Research*, 95(B13):21,793–21,804, December 10 1990.
- [34] Valdis Liepa, Fadil Santosa, and Michael Vogelius. Crack determination from boundary measurements - Reconstruction using experimental data. *Journal of Nondestructive Evaluation*, 12(3):163–174, 1993.
- [35] Stephane G. Mallat. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Trans. PAMI*, 11(7):674–693, July 1989.
- [36] Jean Marie Beaulieu and Morris Goldberg. Hierarchy in picture segmentation: A stepwise optimization approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(2):150–163, February 1989.
- [37] Eric L. Miller. The application of multiscale and statistical techniques to the solution of inverse problems. Technical Report LIDS-TH-2258, MIT Laboratory for Information and Decision Systems, Cambridge, MA., August 1994.
- [38] Eric L. Miller. A scale-recursive, statistically-based method for anomaly characterization in images based upon observations of scattered radiation. In *1995 IEEE International Conference on Image Processing*, 1995. Washington D.C.
- [39] Eric L. Miller and Alan S. Willsky. A multiscale approach to sensor fusion and the solution of linear inverse problems. *Applied and Computational Harmonic Analysis*, 2:127–147, 1995.
- [40] Eric L. Miller and Alan S. Willsky. Multiscale, statistically-based inversion scheme for the linearized inverse scattering problem. to appear in *IEEE Trans. on Geoscience and Remote Sensing*, 1995.
- [41] Eric L. Miller and Alan S. Willsky. Wavelet-based, stochastic inverse scattering methods using the extended born approximation. In *Progress in Electromagnetics Research Symposium*, July 1995. Seattle, Washington.
- [42] Jacqueline Le Moigne and James C. Tilton. Refining image segmentation by integration of edge and region data. *IEEE Trans. on Geoscience and Remote Sensing*, 33(3):605–615, May 1995.

- [43] John E. Molyneux and Alan Witten. Impedance tomography: imaging algorithms for geophysical applications. *Inverse Problems*, 10:655–667, 1994.
- [44] David J. Rossi and Alan S. Willsky. Reconstruction from projections based on detection and estimation of objects—parts I and II: Performance analysis and robustness analysis. *IEEE Trans. on ASSP*, ASSP-32(4):886–906, August 1984.
- [45] Ken Sauer, James Sachs Jr., and Catherine Klifa. Bayesian estimation of D objects from few radiographs. *IEEE Trans. Nuclear Science*, 41(5):1780–1790, October 1994.
- [46] Alon Schatzberg, Anthony J. Devaney, and Alan J. Witten. Estimating target location from scattered field data. *Signal Processing*, 40:227–237, 1994.
- [47] A. H. Tewfick and M. Kim. Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Trans. Information Theory*, 38(2):904–909, 1992.
- [48] Carlos Torres-Verdín and Tarek M. Habashy. Rapid 2.5-D forward modeling and inversion via a new nonlinear scattering approximation. *Radio Science*, pages 1051–1079, July-August 1994.
- [49] Harry L. Van Trees. *Detection, Estimation and Modulation Theory: Part I*. John Wiley and Sons, New York, 1968.
- [50] G. W. Wornell. A Karhunen-Loeve-like expansion for $1/f$ processes via wavelets. *IEEE Transactions on Information Theory*, 36:859–861, July 1990.