A Multiscale, Statistically-Based Inversion Scheme for Linearized Inverse Scattering Problems*

Eric L. Miller[†]
Room 236E Forsyth
Northeastern University
Boston, MA 02115
and
Alan S. Willsky
Room 35-427
Laboratory for Information and Decision Systems
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts, 02139

September 13, 1994

Abstract

The application of multiscale and stochastic techniques to the solution of a linearized inverse scattering problem is presented. This approach allows for the explicit and easy handling of many difficulties associated with problems of this type. Regularization is accomplished via the use of a multiscale prior stochastic model which offers considerable flexibility for the incorporation of prior knowledge and constraints. We use the relative error covariance matrix (RECM), introduced in [20], as a tool for quantitatively evaluating the manner in which data contributes to the structure of a reconstruction. Given a set of scattering experiments, the RECM is used for understanding and analyzing the process of data fusion and allows us to define the space-varying optimal scale for reconstruction as a function of the nature (resolution, quality, and distribution of observation points) of the available measurement sets. Examples of our multiscale inversion algorithm are presented using the Born approximation of an inverse electrical conductivity problem formulated so as to illustrate many of the features associated with inverse scattering problems arising in fields such as geophysical prospecting and medical imaging.

---

1 Introduction

A common objective of many applied inverse problems is the desire to recover characteristics of a medium based upon observations arising from the interaction of transmitted energy with the unknown environment. As discussed extensively in [1], such problems are found in fields as diverse as medical imaging, nondestructive testing, oceanography, remote sensing, ultrasonic imaging, electrical impedance tomography, and geophysical prospecting. The model problem considered in this work is a two-dimensional inverse electrical conductivity problem illustrated in Figure 1. Here, electromagnetic sources (indicated by the black circles) emit time-harmonic waves into a lossy medium. These primary fields are scattered by conductivity inhomogeneities located in the darkly shaded rectangle and 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.

Inverse problems in general, and in particular problems of the type illustrated in Figure 1, present a number of significant and well-recognized challenges. The objective of this paper is to present and demonstrate a methodology for the use of multiresolution methods and concepts in dealing with these issues. In particular, to provide some perspective and motivation for our work, we begin with a brief look at several of these challenges.

A first major issue that must be confronted in solving inverse problems is computational complexity. Indeed, while nonlinearities in the relationship between observables and unknowns present a considerable computational challenge [3, 24], the solution of linearized inverse problems - arising either in the iterative solution of the nonlinear problem or from

a linearized approximation to the true physics of the problem - also represents a computationally intensive task. In particular, the so-called first Born linearization [14, 23] of the inverse conductivity problems considered in this paper leads to large, dense sets of linear equations due to the nonlocal nature of the relationship between conductivity and measurements. While the task of solving such systems of equations certainly is less intensive than obtaining a solution to the nonlinear problem, it can still be prohibitively expensive especially for problems involving large amounts of data and a very fine discretization of the conductivity profile. Additionally, the complexity of these equations not only makes them a challenge to solve efficiently, but it also places a major impediment in the way of other and equally important calculations. For example, the calculation of error statistics associated with the solution to these inverse problems is a considerably more complex problem than calculating the estimates themselves.

As we will see, the use of the wavelet transform to produce multiresolution representations for the unknown conductivity field, the measurements, and the relationship between these two quantities simplifies the analysis considerably, making complex calculations simpler and prohibitively complex ones possible. Roughly speaking, the wavelet transform provides a decomposition in scale so that a nonlocal integral measurement of conductivity more-or-less becomes a local measurement of a coarse feature of the underlying conductivity distribution. As a result, the dense equations relating conductivity to the observed data become sparse in the transform domain leading to efficient inversion algorithms [19, 21].

Moreover, the use of this multiscale representation facilitates the solution of the problem of multisensor fusion and in particular, the combined inversion of data from measurement sets with different resolutions and spatial coverages. In many inverse scattering problems, a

large quantity of data from a collection of scattering experiments may be used to generate the reconstruction; however, the information conveyed by each measurement process may be far from complete. Referring to Figure 1, in the case of the conductivity problem considered in this paper, the measurements from a single experiment consist of the data obtained over one receiver array in response to one of the three sensors operating at a particular frequency. Given the measurements from many such experiments, there is a need for understanding precisely how the different data vectors contribute unique information to a reconstruction and the manner in which measurements from different sources are merged by the inversion routine. For example, the resolution and coverage provided by measurements from the left array in Figure 1 (with sources on the left) are considerably different from those obtained from the right hand array. By using multiresolution representations, the information provided by these data sets are explicitly related to the corresponding scales in the representation of conductivity making fusion simpler to perform and transparent to understand.

A last major challenge concerning inverse problems is that they are frequently ill-posed. For example, the problem illustrated in Figure 1 is ill-posed to the extent that the restriction of collected data to the boundaries of the medium combined with the physics governing the propagation of energy through a lossy medium make exact inversion a mathematical impossibility or, at best, an unacceptably sensitive procedure in which slight measurement errors are greatly magnified by the inversion process. Traditionally, this difficulty is overcome via the use of a regularization procedure which serves to stabilize the original inverse problem so that a unique, physically plausible solution may be computed [10]. Indeed, even if the problem is not ill-conditioned, a regularizer may be incorporated as a means of constraining the reconstruction to reflect prior knowledge concerning the behavior of this function [18].

For example, it is common practice to regularize a problem so as to enforce a degree of smoothness in the reconstruction [10, 18]. As discussed in [20], and in the next subsection, regularization techniques involving such differential penalties have direct interpretations as specifying prior statistical models on the phenomenon to be imaged. In principle, this provides a basis for the calculations of error variances and for considering questions such as the tradeoff between the resolution of reconstruction and the accuracy of the generated image, the value of additional measurement sets, etc. However, as discussed in Section 2.3, performing such analysis using traditional regularization formulations is a formidable and often prohibitive task.

By using a wavelet-based multiresolution framework, we are directly led to an alternative method for statistical regularization in the wavelet transform domain that has a number of attractive properties. First, the class of multiresolution models available to us is extremely rich, allowing us to capture a wide range of characteristics and constraints in our prior models. In this paper, we consider a very special and at the same time highly useful class of multiscale prior models, the so-called fractal priors model for the conductivity field. As shown in [18], this model is related to the traditional smoothness-based statistical regularizers and, with appropriately chosen parameters, produces estimates with similar characteristics. Moreover, Wornell [26] has demonstrated that this model is useful for representing, self-similar stochastic processes possessing $1/f$-type power spectra of the type that is commonly used to describe natural phenomena [7, 25]. Such a model is especially appropriate for inverse conductivity problems particularly in contexts such as geophysical exploration where fractal models are frequently used [9, 25]. Indeed, the fractal priors are exceptionally easy to implement [26], and lead to scale-space algorithms which are orders of magnitude more

efficient than those estimation schemes operating in real-space using a regularizer based upon some differential operator [18].

The direct scale-space form of these models facilitates the explicit analysis of the tradeoff between the incorporation of fine scale detail in a reconstruction and the accuracy in the resulting estimate. In particular, the *relative error covariance matrix* (RECM) introduced in [20] provides a rational basis for dealing with resolution/accuracy tradeoffs and identifying the optimal scale to which the conductivity may be reconstructed as a function of spatial position, the physics of the problem, the prior model, and the spatial coverage and measurement quality of the data provided to the inversion algorithm. Moreover, the RECM provides an explicit description of the information provided by the various data sources both individually and collectively at each point in space and scale. Thus, we can determine those points at which active fusion takes place, i.e. those points where the information provided by several data sets together significantly exceeds that provided by any one set individually. Conversely, we can map the scale space regions in which inversion is dominated by a single data set. Finally, we can use the RECM to assess the incremental value of additional sources of information

In the next section, we provide an overview and notation for the specific inverse problem of interest. Section 3 is devoted to the development of the physical model relating the observables to the conductivity field. In Section 4.1, we present the details behind our wavelet-transform domain representation of the inverse problem. Section 4.2 contains the description of the fractal multiresolution statistical regularization formulation and the tools for RECM-based analysis. A set of examples based upon the inverse conductivity problem and which are illustrative of the different facets of our approach are presented in Section 5.

Finally, conclusions and directions for further investigation are given in Section 6.

2   Problem Overview

2.1   An Inverse Electrical Conductivity Problem

We use a two-dimensional inverse electrical conductivity problem similar in structure to that considered in [13,24] as the primary vehicle for illustrating the multiscale, statistically-based inversion algorithms developed in this paper. The form of the problem is illustrated in Figure 1. Here, we have a set of three electromagnetic line-sources oriented perpendicularly to the page emitting time-harmonic, cylindrical waves into a medium. The electrical properties of this environment are assumed to be decomposed into the sum of two parts: (1) an infinite, known, and constant background and (2) a conductivity anomaly, g, which varies as a function only of the two variable x and z and which is known to lie in a closed and bounded area of the plane, denoted as A and indicated by the darkly shaded region in Figure 1. Upon interaction with the medium, the electromagnetic energy is scattered and the resulting field is measured by one of two arrays of receivers located on either vertical edge of the conductivity perturbation. Each array is composed of a set of line receivers all of which extend infinitely in the direction perpendicular to the page.

We consider inversions based upon the data obtained from a number of scattering experiments. Each such experiment produces a vector of measurements comprised of the observed scattered field obtained over a single receiver array due to energy put into the medium from one of the three sources operating at a particular frequency. Of interest in Section 5 are

inverse problems corresponding to different subsets of the nine experiments defined in Table 1. Here each source is capable of operating at a high, middle and low frequency labeled f_{HI}, f_{MID} and f_{LO} respectively. In any of these problems we use K to denote the number of experimental data sets used, and we index these sets as $i = 1, \ldots, K$.

As is described in Section 3, the use of the first Born approximation yields a linear relationship between the vector of observation associated with the ith scattering experiment, y_i, and a discrete representation of the conductivity anomaly, g. Thus, the observation model used here is of the form

$$y_i = T_i g + n_i \qquad (1)$$

where T_i is a matrix encompassing the (linearized) physics and n_i is an additive, zero-mean, uncorrelated, random vector representing the noise in the data. That is, the ith noise is modeled as $n_i \sim (0, r_i I)$ where I is an appropriately sized identity matrix[1]. It is often useful to work with the "stacked" system of data

$$y = T g + n \qquad (2)$$

where y contains the information from *all* sensors, $i = 1, 2, \ldots, K$. Thus, we have $y = [y_1^T \; y_2^T \; \ldots y_K^T]^T$, with T and n defined accordingly.

2.2 Regularized Inversion and Its Probabilistic Interpretation

A commonly used technique [10, 15, 17] for solving linear inverse problems of the form in (2) is to choose the estimate of g according to

$$\hat{g} = \arg\min_g \|y - Tg\|_{\mathcal{R}^{-1}}^2 + \|Lg\|^2$$
$$= \arg\min_g \|y - Tg\|_{\mathcal{R}^{-1}}^2 + \|g\|_{L^T L}^2 \qquad (3)$$

---

[1]The notation $x \sim (m, P)$ indicates that the random vector x has mean m and covariance matrix P.

where $\|x\|_M^2 = x^T M x$. The first term in (3) enforces fidelity to the data where the weighting $\mathcal{R}^{-1}$ reflects the relative quality of each of the measurement sets as measured by the inverse of the noise covariance while the second term can alternatively be viewed as a regularization term (in case (2) is ill-conditioned or underdetermined) or as a prior statistical model for g. In particular, as discussed in [18], this penalty term is equivalent to a prior model of the form[2]

$$Lg = w \qquad w \sim (0, I). \tag{4}$$

Thus, the nature of the regularization or the prior knowledge is captured in the structure of the matrix L. Frequently in regularization problems, this matrix is chosen so that some degree of smoothness is present in $\hat{g}$ in which case $L = \lambda L_0$ where L_0 is a discrete form of an appropriate differential operator [18]. The scalar factor λ is then used to provide a tradeoff between the influence each of the two terms in (3) exerts on the reconstruction.

The optimization problem given by (3) admits a solution which defines $\hat{g}$ in terms of the normal equations

$$(T^T \mathcal{R}^{-1} T + L^T L)\hat{g} = T^T \mathcal{R}^{-1} y. \tag{5}$$

As discussed in [18], this solution, $\hat{g}$ can also be interpreted as the linear least squares estimate (LLSE) or minimum variance estimate of g given the noisy measurements in (2) with $n \sim (0, \mathcal{R})$ and given the prior statistics for g implied by (4), i.e. g is zero mean and has $L^T L$ as the *inverse* of its covariance. Furthermore, the estimation error covariance matrix, that is, the covariance of $g - \hat{g}$ is given by

$$E\left[(g - \hat{g})(g - \hat{g})^T\right] = (T^T \mathcal{R}^{-1} T + L^T L)^{-1}. \tag{6}$$

---

[2]Note that we assume zero-mean in the prior model for g only for notational simplicity. There is no complexity added if we incorporate a prior mean, e.g. in a penalty term of the form $\|L(g - g_0)\|^2$

2.3   A Multiscale Representation of the Problem

For the inverse problems of interest here, the T_i in (1) represent discretized integral operators corresponding to different scattering experiments. Furthermore if L is a discrete version of a differential operator, then obtaining $\hat{g}$ from (5) requires the solution of an elliptic partial integro-differential equation, or in discretized form, a dense set of linear equations. Moreover, computing the error-covariance matrix in (5) corresponds to the explicit calculation of the inverse of the large, dense (integro-differential) matrix $T^T \mathcal{R}^{-1} T + L^T L$, an even more formidable task. Among the objectives in introducing wavelets and wavelet-based prior models are to make the computations in both (5) and (6) far simpler to perform and to facilitate a deeper look at data fusion, reconstruction accuracy, and resolution.

In Section 4.1 we define orthonormal, discrete wavelet transform operators W_i and W_g which transform the measurements, y_i and the discretized conductivity field, g, into their respective wavelet decompositions

$$\eta_i = \mathcal{W}_i y_i$$

$$\gamma = \mathcal{W}_g g.$$

In our analysis of (1), we use W_i and W_g to move from physical to scale space in the following manner

$$\eta_i = \mathcal{W}_i y_i = (\mathcal{W}_i T_i \mathcal{W}_g^*)(\mathcal{W}_g g) + \mathcal{W}_i n_i$$

$$\equiv \Theta_i \gamma + \nu_i \tag{7}$$

where $\mathcal{W}_g^* \mathcal{W}_g = I$ follows from the orthonormality of the wavelet transformation.

Finally, analogously to the physical space case, we define the stacked systems

$$\eta = \Theta \gamma + \nu \tag{8}$$

with $\eta = [\eta_1^T \ \eta_2^T \ \dots \eta_K^T]^T$ and Θ and ν are defined accordingly.

In analogy to the discussion in Section 2.2, we can now define a corresponding regularized inverse problem, or equivalently, a linear-least-squares estimation problem, in the wavelet transform domain. Specifically, we wish to reconstruct γ based on a prior model for γ, namely $\gamma \sim (0, P_0)$, together with the noisy measurements (8). That is, the LLSE, $\hat{\gamma}$, is defined as

$$\hat{\gamma} = \arg\min_{\gamma} \|\eta - \Theta\gamma\|^2_{R^{-1}} + \|\gamma\|^2_{P_0^{-1}} \tag{9}$$

so that $\hat{\gamma}$ satisfies normal equations of the form

$$(\Theta^T R^{-1}\Theta + P_0^{-1})\hat{\gamma} = \Theta^T R^{-1}\eta \tag{10}$$

and the corresponding error-covariance matrix is given by

$$P = E[(\gamma - \hat{\gamma})^T(\gamma - \hat{\gamma})]$$
$$= (\Theta^T R^{-1}\Theta + P_0^{-1})^{-1}. \tag{11}$$

Comparing (3), (5), and (6) to eqs. (9), (10), and (11), we see that the wavelet transformation has left us with a formulation of exactly the same structure as we had originally. The advantages of this transformation come from two important facts. First, the measurements operator Θ in the wavelet domain is far more sparse than the operator T. Secondly, as we will see in Section 4.2, the inverse of the prior covariance P_0^{-1} can be taken to be diagonal. In particular, as we discuss in Section 4.2, a specific diagonal choice for P_0^{-1} implies a smoothness penalty (or equivalently a fractal prior model) analogous to that captured by $L^T L$ in (3), (5), and (6) when L is a differential operator while other choices for the diagonal matrix P_0^{-1} allow us additional flexibility to capture a rich variety of other regularization objectives or prior models. As a consequence of the sparsity of Θ and the diagonal nature of P_0^{-1}, not only can (10) be solved efficiently, but also the the matrix to be inverted in (11) is sparse with *diagonal* regularization provided by P_0^{-1} thereby facilitating the computation of

the inverse and in particular the diagonal elements of P corresponding to the error variances in each of the wavelet coefficients in out multiscale representation of g.

3 Formulation of the inverse conductivity problem

In general, problems of the type illustrated in Figure 1 are characterized by a nonlinear relationship between the data and the conductivity perturbation [24]. As a first step in exploring the utility of multiscale and statistical methods for the solution of inverse problems, we consider in this paper a linearized form of the problem obtained using the so-called first Born approximation to Maxwell's equations [14]. In general, the Born approximation is valid only for conductivity perturbations which are "small" both in spatial extent as well as magnitude relative to the background [14]. For the issues of interest in this paper, such assumptions are not restrictive and in Section 6 we discuss the manner in which our approach can be extended easily to other linearizations as well as to the consideration of the full, nonlinear inverse problem.

Under the geometric configuration shown in Figure 1 and discussed in Section 2, all field quantities and material parameters are functions only of the two space coordinates, $\mathbf{r} = (x, z)$. As discussed in [12], for the ith data set, the model linking the data to to the conductivity under the Born approximation takes the form of a first-kind Fredholm integral equation

$$y_i(\mathbf{r}_j) = \int_C T_i(\mathbf{r}_j, \mathbf{r}')g(\mathbf{r}')d\mathbf{r}' + n_i(\mathbf{r}_j) \tag{12}$$

where $y_i(\mathbf{r}_j)$ is the observation at the jth point in the receiver array associated with the ith experiment, $T_i(\mathbf{r}, \mathbf{r}')$ is derived from the constant-background Green's function for this

problem and is dependent upon the value of the background conductivity as well as the source/receiver geometry for this scattering experiments, $g(\mathbf{r})$ is the conductivity perturbation, C is the region over which g is nonzero, and $n_i(\mathbf{r}_j)$ represents the noise corrupting the data.

Reduction of the integral equation (12) to matrix-vector relationships require representation of g in terms a a finite number of parameters. Here, we use a method-of-moments approach [16] in which we first pixelate C, and then assume g to be constant over each of the N_g subregions. Thus we have:

$$g(\mathbf{r}) = \sum_{k=1}^{N_g} g(k)\chi_{C_k}(\mathbf{r}) \tag{13}$$

with $\chi_{C_k}(\mathbf{r})$ the characteristic function of the set C_k, $C = \cup_{k=1}^{N_g} C_k$, and $g(k)$ the value of g over C_k[3]. Substitution of (13) into (12) yields

$$y_i(\mathbf{r}_j) = \sum_{k=1}^{N_g} T_i(\mathbf{r}_j, k)g(k) + n_i(\mathbf{r}_j)$$

with

$$T_i(\mathbf{r}_j, k) = \int_{C_k} T_i(\mathbf{r}_j, \mathbf{r}')dr'.$$

Denoting $y_i(\mathbf{r}_j)$ as simply $y_i(j)$ with analogous notation for $n_i(\mathbf{r}_j)$ and $T_i(\mathbf{r}_j, i)$ yields the discrete form of (12)

$$y_i(j) = \sum_{k=1}^{N_g} T_i(j, k)g(k) + n_i(j) \tag{14}$$

which we shall write simply as the matrix-vector product in (1).

---

[3]Note that for 1D problems (see Section 5.1) the C_k will represent intervals of the real line while for in two-dimensions (Section 5.2), C_k will be rectangular regions in the plane.

4 A Multiscale Formulation of the Inverse Problem

4.1 A Wavelet Representation of g and y

The fundamental idea behind the discrete wavelet transform (DWT) is to decompose signal, here represented as a vector, into a sequence of increasingly "coarser" representations while at the same time retaining the information lost in moving from a fine to a coarse scale. In our case, we will be concerned both with one and two dimensional signals where for simplicity, we first describe the wavelet representation and notation for a 1D signal vector, a. Following the wavelet literature, the elements of this vector are termed the *finest scale scaling coefficients* associate with a, and the vector a is denoted by $a(M_a)$ indicating that these scaling coefficients represent a at scale M_a. The scale number reflects the number of elements in a. Typically, we consider vectors of length 2^m for which the scale number is the integer m.

Beginning with $a(M_a)$, a coarser representation (that is, a coarser set of scaling coefficients), $a(M_a - 1)$, is obtained by first passing $a(M_a)$ through a low pass, finite impulse response (FIR) filter, l, and then decimating the filtered output by a factor of two. Thus, $a(M_a - 1)$ is "coarser" than $a(M_a)$ in that the filtering/downsampling procedure has removed the high frequency structure from the original signal, and $a(M_a - 1)$ is half as long as $a(M_a)$. The detail lost in moving from $a(M_a)$ to $a(M_a - 1)$ is extracted separately by first high pass filtering $a(M_a)$ with the FIR filter h and then downsampling by two. This detail vector is denoted $\alpha(M_a - 1)$. The filtering and decimation procedure is successively applied to the coarsened versions of a resulting in a sequence of scaling coefficient vectors, $a(m)$, and a sequence of detail vectors, $\alpha(m)$, for $m = M_a - 1, \ldots, L_a$ where L_a is the coarsest scale at

which a is represented. Thus, at scale m, we have

$$a(m) = L(m)a(m+1) \tag{15}$$

$$\alpha(m) = H(m)a(m+1) \tag{16}$$

where $L(m)$ and $H(m)$ are operators (i.e. matrices) representing the filtering and decimation operations with the filters l and h respectively.

As described extensively in [4], the filters l and h are constructed so that $H(m)$ and $L(m)$ satisfy the so-called perfect-reconstruction properties

$$L(m)L^*(m) = I \quad H(m)H^*(m) = I$$

$$L^*(m)L(m) + H^*(m)H(m) = I \tag{17}$$

where $H^*(m)$ is the adjoint of $H(m)$. Using (15), (16) and (17), we see that $a(m+1)$ is obtained from its coarse scale representation and the detail at scale m via

$$a(m+1) = L^*(m)a(m) + H^*(m)\alpha(m). \tag{18}$$

Clearly, iterating (18) provides the mechanism for obtaining the scaling coefficients of a at scale m for $L_a < m \le M_a$ using the coarsest scale scaling coefficients $a(L_a)$ and intervening detail coefficients $\alpha(n)$ for $L_a \le n \le m-1$. In particular, we may construct an operator,[4] $\mathcal{W}_a$ from $H(m)$ and $L(m)$ which relates the finest scale scaling coefficients, $a \equiv a(M_a)$, to the coarsest scaling coefficients, $a(L_a)$, and the full set of detail coefficients $\alpha(m)$ for scales $m = L_a,\ L_a+1,\ \ldots,\ M_a$. That is, we may write

$$\alpha = \mathcal{W}_a a \tag{19}$$

where $\alpha = [\alpha(M_a-1)^T \ \ldots \ \alpha(L_a)^T \ a(L_a)^T]^T$ and $\mathcal{W}_a$ satisfies $\mathcal{W}_a^*\mathcal{W}_a = I$. We call the vector α the *wavelet transform* of a. The nth element of $\alpha(m)$ is denoted $\alpha(m,n)$ and is referred to as the nth *shift* of α at scale m. Similarly, $a(m,n)$ represents the nth element of the vector

---

[4]We choose to subscript the wavelet transform operator here as $\mathcal{W}_a$ to make explicit that this is the transform for a. We may (and in fact *will*) use different wavelet transforms for the different variables.

of scaling coefficients at scale m.

Given this implementation of the DWT, the relationships among the scale space compo-
nent in the decomposition of a are graphically represented in the form of a lattice as shown
in Figure 3 for the case of a wavelet decomposition with $l(n)$ and $h(n)$ of length 4 (such as
the so-called "D4" or Daubechies 4-tap wavelet decomposition described in [4].) The coeffi-
cients at any scale all lie on a common horizontal line with the finest scale coefficients at the
bottom of the lattice and the coarsest at the top. At the finest scale, the nodes represent
the finest set of scaling coefficients. Each node at all other scales contains one wavelet co-
efficient except for the top scale where the nodes contain the coarsest wavelet and coarsest
scaling coefficients. Finally, two nodes are connected by an arc if and only if there is a linear
relationship between the contents of these nodes as dictated by the structure of the wavelet
transform matrix $\mathcal{W}_a$.

For future reference we define some terminology related to the lattice structure in Figure
3. In particular, a coarse scale node is said to *impact* a finer scale if there exists a strictly
downward path on the lattice from the former to the latter. Thus, we define the downward
impact set, $\mathcal{D}(m,n)$ associated with node (m,n), as the set of finest scale nodes which (m,n)
impacts. Thus in Figure 3, $\mathcal{D}(\square)$ is comprised of all nodes marked with the symbol "■".

The wavelet decomposition of the scaling coefficients of a two dimension function is
obtained by considering a as a matrix and applying one wavelet transform, $\mathcal{W}_{a,z}$, to the
columns and a second wavelet transform, $\mathcal{W}_{a,x}$, to the rows. In general, $\mathcal{W}_{a,z}$ and $\mathcal{W}_{a,x}$
may each use different l and h filters and associated with each transform are a finest and
a coarsest scale of interest which we denote $M_{a,\beta}$ and $L_{a,\beta}$ for $\beta \in \{x, z\}$. We use $\mathcal{W}_a$ to

represent the composition of the operators $\mathcal{W}_{a,x}$ and $\mathcal{W}_{a,z}$ and write

$$\alpha \equiv \mathcal{W}_a a = \mathcal{W}_{a,z} a \mathcal{W}_{a,x}^*.$$

Furthermore, we note

$$(\mathcal{W}_a^* \mathcal{W}_a)a \equiv \mathcal{W}_{a,z}^* (\mathcal{W}_{a,z} a \mathcal{W}_{a,x}^*) \mathcal{W}_{a,x} = (I)a(I) = a$$

so that $\mathcal{W}_a^* \mathcal{W}_a = I$. As in the 1D case, we denote a particular element of α by $\alpha(m,n)$. Here, we understand m and n to be two-vectors indexing the scales and shifts in the x and z directions, i.e. $m = [m_x \ m_z]^T$ and $n = [n_x \ n_z]^T$ respectively. Also, we use the notation $\alpha(m)$ to indicate the collection of wavelet coefficients at all shifts and at scale $m = [m_x \ m_z]^T$. Unlike the 1D case, the two-dimensional wavelet transform induces a four dimensional lattice structure in scale space with two dimensions for scale and two for shift. Nonetheless, we define downward impact sets in the same manner as was the case in 1D. That is, $\mathcal{D}(m,n)$ is the set of nodes in a which are impacted by $\alpha(m,n)$.

The operators $\mathcal{W}_g$ and $\mathcal{W}_i$ introduced in Section 2.3 are then constructed using the procedures just described. In general, these operators may differ from one another. For example, in the problems considered here, all of the y_i are 1D signals, while g may either be a one or a two dimensional function. Furthermore, in principle, we can use different wavelets for different data sets and can use a different number of scales in the wavelet decomposition of each quantity (see the particular choices used in Section 5).

4.2 Multiscale Prior Models

As discussed in Sections 1 and 2, a key component in our formulation of the inverse problem is the use of a multiscale stochastic model for g to regularize the inversion and to capture prior information. To motivate the particular choice of prior model used here, consider the

case of a one dimensional function whose covariance matrix is $(L^T L)^{-1}$ with L representing first order differentiation. This implies that g is a Brownian motion satisfying $Lg = w$ with $w \sim (0, I)$. As discussed in [18], work by Wornell and others [8, 22] has demonstrated that Brownian motions and other related fractal processes can be closely approximated via a statistical model in which the wavelet coefficients of g are independent and distributed according to

$$\gamma(m, n) \sim (0, \kappa^2 2^{-\mu m}). \tag{20}$$

Here, κ^2 controls the overall magnitude of the process and the parameter μ determines the fractal structure of sample paths. The case $\mu = 0$, corresponds to g being white noise while as μ increases, the sample paths of g show greater long range correlation and smoothness.

In addition to defining the scale-varying probabilistic structure of the wavelet coefficients of g, we also must provide a statistical model for the coarsest scale scaling coefficients. Roughly speaking, these coarse scale coefficients describe the DC and low spatial frequency structure of g. In the applications we consider here, we assume that we have little *a priori* knowledge concerning the large-scale average value of g. Consequently, we take $g(L_g) \sim (0, p_{L_g} I)$ where p_{L_g} is some sufficiently large number. By choosing p_{L_g} in this manner, we avoid any bias in the estimator of the low frequency structure of g. Finally, we note that for these models, the resulting matrix P_0 in (10) is *diagonal* with nonzero entries corresponding to the variances associated with each component of the γ.

For the case where g is a two dimensional function, we consider the separable representation with

$$\gamma(m, n) \sim (0, \kappa_x^2 \kappa_z^2 2^{-(\mu_x m_x + \mu_z m_z)})$$

for $L_{g,x} \leq m \leq M_{g,x} - 1$ and $L_{g,z} \leq m \leq M_{g,z} - 1$. For $m = L_{g,x}$ we take $\gamma(m, n) \sim$

$(0, p_{L_{g,x}} \kappa_z^2 2^{-(\mu_z m_z)})$ with analogous results holding when $n = L_{g,z}$.

Clearly, other choices of statistics for the components of γ may be appropriate in specific applications, and our methodology can readily accommodate these. The specific choice we have made, leading to a $1/f$-like fractal model, is useful both in its ability to model natural phenomena and because the successively decreasing variances of the fine scale wavelet coefficients control the incorporation of high frequency information into the reconstruction. As will be seen in Section 5, this is precisely the type of regularization required for the inverse conductivity problem. Thus, the fractal model provides us with one physically meaningful way in which to specify the tradeoff which in turn determines the way in which the resulting estimation algorithm makes effective use of the data only over those scales where useful information is present.

4.3   The Relative Error Covariance Matrix

A key advantage of the use of statistical estimation techniques is the ability to produce not only the estimate of γ but also an indication as to the quality of this reconstruction in the form of the error-covariance matrix P defined in (11) for the problem of interest in this paper. While the information contained in P is certainly important for evaluating the absolute level of uncertainty associated with the estimator, in many cases, it is more useful to understand how data serves to reduce uncertainty relative to some reference level. That is, we have some prior level of confidence in our knowledge of γ and we seek to comprehend how the inclusion of additional data in our estimate of γ alters our uncertainty relative to this already established level. In this section we define the *relative error covariance matrix*

(RECM) and demonstrate its utility as a tool for capturing such changes in uncertainty. The analysis of the RECM in the wavelet domain is especially interesting because it allows for a localized characterization of the manner in which data impacts a reconstruction. Hence, we show how the RECM provides a natural means of evaluating the appropriate level of detail as a function of position which can be supported in a reconstruction based upon a given set of data and also leads directly to a quantitative, multiscale theory of sensor fusion.

The definition of the relative covariance matrix is motivated by the definition of the relative difference between two scalars a and b given by

$$1 - \frac{b}{a}. \tag{21}$$

The matrix analog to (21) to be considered in this paper is as follows. Let $\{1, \ldots, K\}$ denote the index set for the observations sets y_i. For any subset A of $\{1, \ldots, K\}$ let P_A denote the estimation error covariance as in (11) resulting from the estimation of γ based upon the data sets corresponding to A (i.e. $\{y_i \mid i \in A\}$ where for $A = \emptyset$, the empty set, $P_{\{\emptyset\}} = P_0$, the prior covariance. The RECM provides a measure of the relative quality of the estimate based upon data in two sets A and B and is given by

$$\Pi(A, B) = I - P_A^{-T/2} P_B P_A^{-1/2}. \tag{22}$$

The definition of $\Pi(A, B)$ in (22) possesses many useful properties. First, like an error covariance matrix, it is symmetric. Also if $\bar{\Pi}(A, B)$ represents the relative error covariance matrix for the estimation of g, i.e. the physical-space representation of the conductivity, then this is directly computable from $\Pi(A, B)$ using the wavelet transform

$$\bar{\Pi}(A, B) = \mathcal{W}^T \Pi(A, B) \mathcal{W}.$$

Moreover, it is not difficult to show that $\Pi(A, B)$ is normalized to the extent that for $A \subset B$,

$$0 \le \Pi(A, B) \le I.$$

We note that in this case $\Pi(A, B) = 0$ iff $P_B = P_A$ which indicates no additional reduction in uncertainty results from augmenting A with the data sets in in $B - A$. Also, $\Pi(A, B) = I$ if and only if $P_B = 0$ i.e. only when *all* uncertainty in γ has been removed when we use the additional information in B relative to A.

In the event P_A is diagonal, the diagonal components of $\Pi(A, B)$ are particularly easy to interpret. Let $\sigma_i^2(A)$ be the error-variance of the ith component of γ arising from an estimate based upon data from set A. Then, the ith component of the diagonal of $\Pi(A, B)$ is just

$$1 - \sigma_i^2(B)/\sigma_i^2(A) \tag{23}$$

which is nothing more than the relative size difference of the error-variance in the ith component of γ based upon data from sets A and B. Note that the diagonal condition of P_A is met in this paper when $P_A = P_0$, since the wavelet and scaling coefficients are uncorrelated for the fractal $1/f$ priors used here as well as for many other physically meaningful prior models. Thus, the diagonal elements of $\Pi(\{\emptyset\}, B)$ represent the decrease in uncertainty due to the data from set B relative to the prior model. Finally, as $\Pi(\{\emptyset\}, B)$ will be of interest frequently in the remainder of this work, we shall abuse notation and write $\Pi(\{\emptyset\}, B)$ as $\Pi(B)$ in cases when there will be be no confusion.

The quantity $\Pi(A, B)$ represents a useful tool for quantitatively analyzing the relationship between the characteristics of the data (as defined by Θ and R) and the structure of the estimate $\hat{\gamma}$. Consider, for example, the case in which we wish to assess the overall value of a collection of observation vectors. Let $\Pi_n^m(B)$ denote the diagonal element of the matrix $\Pi(B)$ corresponding to the wavelet coefficient at scale/shift (m, n)[5]. Because P_0 is diagonal, (23)

---

[5]At scale $m = L_g$, we are interested in both the wavelet and scaling coefficients of g. To avoid ambiguity, we use the notation $\Pi_n^{\overline{L}_g}$ to refer to the RECM information for the coarsest scaling coefficient of g at shift n. In the case of a two dimensional g, where there is confusion, we shall explicitly write $m = (m_x, m_z)$ placing

indicates that $\Pi_n^m(B)$ represents the relative decrease in the error variance associated with the component in the wavelet transform of g at scale/shift (m, n). Thus, $\Pi_n^m(B)$ provide us with a natural way in which to define $m^*(j)$, the appropriate level of detail which should be included in a reconstruction of $g(M_g)$ at shift j. Specifically, for each location j, we can examine the quality of the information present at this point and at all coarser scale "ancestors" of j. Using the terminology introduced in Section 4.1, we say that the data supports a reconstruction of $g(M_g, j)$ at scale m if there exists some node in the wavelet lattice of g at scale m which (1) impacts $g(M_g, j)$ (i.e. for some shift n, $g(M_g, j) \in \mathcal{D}(m, n)$) so that (m, n) is an ancestor of (M_g, j) and (2) for which the data provides a sufficiently large quantity of information regarding the structure of g at node (m, n) (i.e. $\Pi_n^m(B)$ is in some sense large). Clearly, $m^*(j)$ is the finest scale for which a node (m, n) may be found that satisfies the above two criteria. The precise quantification of "sufficiently large" will depend upon the particular application and on the structure of the particular inverse problems under investigation.

For the problems considered here, the diagonal structure of P_0 imply that $0 \leq \Pi_n^m(B) \leq 1$ so that determining whether $\Pi_n^m(B)$ is sufficiently large is accomplished by comparing this quantity to some threshold, τ, between zero and one. The procedure described in the preceding paragraph for determining the optimal scale of reconstruction suggests that the only elements of γ which need be estimated are those for which $\Pi_n^m(B) > \tau$. Hence, we are

---

a bar over that scale index (those scale indices) associated with coarsest scaling coefficient.

led to define $\hat{\gamma}_\tau$, a truncated version of $\hat{\gamma}$, as follows:

$$[\hat{\gamma}_\tau]_{(m,n)} = \begin{cases} 0 & \Pi_n^m(B) \leq \tau \\ \\ [\hat{\gamma}]_{(m,n)} & \text{otherwise} \end{cases} \tag{24}$$

where $[\hat{\gamma}]_{(m,n)}$ is the component in the vector $\hat{\gamma}$ at scale m and shift n. Defining $\hat{\gamma}_\tau$ in this

way ensures that $\hat{g}_\tau = \mathcal{W}^T \hat{\gamma}_\tau$ is in fact the reconstruction of g which at each shift n contains

detail information at scales no finer than $m^*(n)$.

In addition to its use in assessing the scale of reconstruction supported by the information

from a set of measurements, if we consider the case in which neither A nor B is empty, we

find that there are several ways in which $\Pi(A, B)$ may be of use in assessing the value of

fusing information from multiple data sets and in identifying how this fusion takes place.

For example, if $A \subset B$, then roughly speaking, if $\Pi(A, B)$ is significantly larger than 0,

there is a benefit in the additional information provided by the observations in $B - A$.

Moreover, $\Pi_n^m(A, B)$ can be used to pinpoint the scales and locations at which this fusion

has significant benefit[6] i.e., those scales and shifts at which *active* sensor fusion is taking

place. Furthermore, by varying the sets A and B, we can identify not only the optimal

scale for reconstruction at each point but can also identify which data sets are actively

used to obtain that estimate. That is, for each (m, n), we can in principal find the set

$A \subset \{1, \ldots, K\}$ so that $\Pi_n^m(A, \{1, \ldots, K\})$ is small (indicating that sensors not in A

provide little additional information to the reconstruction of wavelet coefficient (m, n)) and

so that for any $C \subset A$, $\Pi_n^m(C, A)$ is of significant size (so that all of the sensors actively

contribute to the reconstruction at this scale and shift.)

---

[6]In this case, because P_A is not in general diagonal, the diagonal elements of $\Pi(A, B)$ do not have the exact interpretation as the relative size difference of the error variance of γ based upon data from A and B; however the size of these diagonal components of $\Pi(A, B)$ still provides useful insight as to the scales and shifts where the observations from set B provide information not found in the data from set A.

5 Examples

5.1 A One Dimensional, Radial Profiling Problem

We begin by considering a radial profiling problem similar to that analyzed by Habashy *at. al* in [11, 13]. Here, g is assumed to vary only in the horizontal direction in Figure 1 with the specific true conductivity profile g to be used in this example shown as the solid line in Figure 5. The numerical values specifying the prior model and the parameters describing the background medium are given in Table 2. In this work, the signal to noise ratio of the vector $\eta_i = \Theta_i \gamma + \nu_i$ with $\nu_i \sim (0, r_i^2 I)$ and $\gamma \sim (0, P_0)$ is defined as

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

where N_g is the length of the vector γ and tr is the trace operation.

The objective of this example is to illustrate the utility of the RECM in analyzing the various ways in which the data available to a reconstruction impacts the estimate. Specifically, we explore inversions using data from the following three different combinations of the high and middle frequency scattering experiments described in Table 1:

D_{HI} Data collected at the left receiver array in response to all three sources operating at the the highest frequency (i.e. information from experiments 1–3 in Table 1).

D_{MID} Data collected at the left receiver array in response to the three sources operating at the middle frequency (i.e. information from experiments 4–6 in Table 1).

$D_{HI,MID}$ Data from $D_{HI} \cup D_{MID}$.

The information regarding the structure of g supplied by D_{HI} and D_{MID} is illustrated in

Figure 4. Recall from (1) that the jth observation point of the ith data set,

$$y_i(j) = \sum_{k=1}^{N_g} T_i(j,k)g(k) + n_i(j). \tag{25}$$

so that the jth row of T_i represents the map which takes conductivity, g, into the jth element of the ith observation vector. In Figure 4(a) (resp. 4(b)), a single row from kernels associated with high (resp. middle) frequency scattering experiments are shown. Specifically, we plot the maps associated with the observation point in the middle of the left receiver array for experiments whose source is the middle of the three line sources. From these illustrations, we see that the high frequency observations are most sensitive to variations in g close to $x = 0$ but provide essentially no information regarding the structure of g far from the origin. The data corresponding to middle-frequency sources better reflect the behavior of g away from $x = 0$ but still are comparatively insensitive to the conductivity far from the point of observation. Thus, in general we expect to obtain a relatively accurate reconstruction of g near $x = 0$ with decreasing fidelity as a function of radial position.

In Figure 5(a), the estimate obtained using data sets 1–12, $\hat{g}(D_{HI,MID})$, is compared with the true function. Clearly, we are able to resolve the left edge and to a lesser extent the magnitude of the conductivity anomaly located closest to the origin. However, the information provided by $D_{HI,MID}$ is insufficient to obtain an accurate estimate of the right edge of this structure or any but the coarsest information regarding the rightmost block. As a means of understanding how both D_{HI} and D_{MID} contribute information to this estimate, in Figures 5(b)–(c), $\hat{g}(D_{HI,MID})$ is graphed against $\hat{g}(D_{HI})$ and $\hat{g}(D_{MID})$ respectively. Again, we see that individually, the data from the high and middle frequency sources provide information about g close to $x = 0$. Further from the origin, $\hat{g}(D_{HI,MID})$ follows neither $\hat{g}(D_{HI})$ nor $\hat{g}(D_{MID})$ so that some level of data fusion must be taking place to the extent that the pres-

ence of both data sets together yields an estimate of g over this region which is substantially different from that obtained from either set alone.

A more accurate assessment of the manner in which this information is merged is obtained by analysis of the diagonal elements of the relative error covariance matrices, $\Pi(B)$ for $B \in \{D_{HI}, D_{MID}, D_{HI,MID}\}$. In Figure 6 these quantities are plotted for scales 3, 4, and 6. In each of these graphs, $\Pi_n^m(D_{HI})$ is marked with a $\circ$, $\Pi_n^m(D_{MID})$ with a $\times$, and $\Pi_n^m(D_{HI,MID})$ with a $+$. As there is strictly more information in $D_{HI,MID}$, than in either D_{HI} or D_{MID} alone, it is the case that all $+$'s must lie above the other two symbols. In those cases where $\Pi_n^m(D_{HI,MID})$ is significantly larger than both $\Pi_n^m(D_{HI})$ and $\Pi_n^m(D_{MID})$, we say that active sensor fusion is taking place. Indeed, in Figure 6(a), this is the case for the estimates of elements $5 - 8$ of $g(L_g)$. Moreover, examination of Figures 6(b)–(d) shows that active sensor fusion is occurring with respect to the estimates of the wavelet coefficients of g near the origin at scales 3, 4, and 6. We have omitted the RECM plot at scale 5 as no such fusion occurs at that scale in this example. Finally, Figure 6 is instructive to the extent that it demonstrates where the data do *not* support a reconstruction. The fact that Π_n^m is close to zero at all scales and for all wavelet coefficients corresponding to shifts far from $x = 0$ indicates that the information in D_{HI} and D_{MID} either alone or in combination is insufficient to reconstruct *any* detail in g over this domain.

This notion can be made more precise by considering the space-varying optimal scale of reconstruction, $m^*(j)$, defined in Section 4.3. In Figure 7(a) and (b), the optimal scale as a function of position is plotted for $\tau = 0.05$ and $\tau = 0.5$ respectively using data from $D_{HI,MID}$. For the smaller value of τ, we see that as the x grows large, the optimal scale drops from 6 to 3 in a manner quite consistent with the intuition developed by examination

of the estimates. That is, for a rather narrow region near the origin, the RECM information dictates that a fine scale reconstruction of g should be possible. As x increases, the scale of detail to be included in $\hat{g}(D_{HI,MID})$ decreases. For $\tau = 0.50$, Figure 7(b) shows similar characteristics to the $\tau = 0.05$ case; however, the more stringent threshold results in a more rapid decrease in scale as a function of distance. Finally, in Figures 7(c)-(d) the truncated estimates, $\hat{g}_\tau(D_{HI,MID})$, defined by (24), are compared against $\hat{g}(D_{HI,MID})$ for $\tau = 0.05$ and $\tau = 0.50$ respectively showing that there is little difference between the optimal estimate and its truncated versions.

The relative error covariance matrix also represents a useful tool for analyzing the incremental benefits associated with the addition of data to an already-formed estimate. In Figure 8, the diagonal elements of $\Pi(D_{HI}, D_{HI,MID})$ are displayed for the coarsest scaling coefficients and the finest wavelet coefficients. These plots illustrate that the middle-frequency data sets contribute new information to an estimate based upon the high-frequency observations at the coarsest scale away from the origin and at the finest scale, closest to the origin. For all other scales and shifts, $\Pi_n^m(D_{HI}, D_{HI,MID})$ is essentially zero. We note that the RECM information is in accord with the plots of the estimates in Figure 5 where we saw little difference in the actual estimates based upon the different data sets near the origin while farther from $x = 0$, the estimate generate from both D_{HI} and D_{MID} differed significantly in a very coarse scale manner from those obtained using either the high or the middle frequency data.

From this example, we see that the relative error covariance matrix provides new and useful insight into multisensor data fusion not obtainable by analysis of either the kernel functions or the estimates. The use of the RECM is significant because it allows for the for-

mulation of these conclusions before any data are obtained. For the radial profiling problem considered here, one would conclude that the data from the high and middle frequency data sets is useful for the recovery of the conductivity detail structure near the origin; however, additional observations are required to recover all but the coarsest scale information far from $x = 0$. Additionally, the RECM analysis suggests that the original parameterization of g involving 128 degrees of freedom is clearly excessive. Rather, at a threshold of $\tau = 0.50$, the data dictates that only 9 elements of γ (the nonzero elements of $\hat{\gamma}_{0.50}(D_{HI,MID})$) can be accurately recovered representing a 93% reduction in complexity of the inverse problem.

5.2 A Two-Dimensional, Cross-Well Tomography Problem

In this example we consider improving resolution near the right side of the conductivity anomaly by using observations obtained from sources located at the left and receivers on the right side of C. This observation configuration arises quite frequently in practice especially in the fields of medical imaging and geophysical prospecting [5, 6, 24] and we term inverse problems with this measurement geometry *cross-well tomography* problems as they model the case where the lines $x = 0$ and $x = 100$ are taken to be oil boreholes [24]. Additionally, we now assume a full 2D problem wherein g is varies both in the x and the z directions. The true conductivity anomaly to be reconstructed in this example is displayed in Figure 10(a) and the various parameter values needed for this experiment are given on Table 3.

From the radial profiling problem, to obtain information regarding the behavior of g far from $x = 0$ it is necessary to probe the medium with low frequency energy. Hence, for this problem, we augment $D_{HI,MID}$ with data sets 9–12 from Table 1. These data are generated

by low frequency sources located near the left side of the region of interest and measured by the receiver array located at right side. We denote this addition collection of observations D_{LO}. The structure of the kernels associated with this problem is seen in Figure 9. In particular recalling the configuration of sources and receivers in Figure 1, the plots in Figure 9 correspond to the maps taking g into the observation at the mid-point of the left (in (a) and (b)) or right (in (c)) receiver array in response to input energy from the middle source. Also, as g is a 2D function, so too are these maps; hence, each pixel in Figures 9(a)–(c) represents the weight placed on the corresponding element of g in the sum (25) with darker colors indicating larger magnitudes. The high and middle frequency scattering experiments are most sensitive to variations in g near the left side of the square. The structure of the low frequency kernel with areas of sensitivity near both the left and right vertical edges suggests that the addition of data from D_{LO} will improve the estimate of g near $x = 0$ and allow for the determination of at least some structure at the far side of the region.

In Figure 10, we see that the addition of the low-frequency, cross-well data does significantly improve the resolution on the right side of A. Figure 10(b) (resp. (c)) is a display of $\hat{g}(D_{HI,MID})$ (resp. $\hat{g}(D_{HI,MID,LO})$). Given only the high and medium frequency information, the anomaly near $x = 100$ is almost completely undetected; however, the addition of the low frequency data clearly improves the ability to resolve this second structure. We do note that while both conductivity perturbations are reflected in the estimates of g, the nature of the physics of the problem allows for only a comparatively coarse-scale or blurred reconstruction near the right vertical edge of the anomaly. In general, for inverse scattering problems of the type considered here, one requires data at more frequencies and/or from many source/receiver combinations in order to obtain significantly higher resolution

estimates of such anomalies.

As in the radial profiling problem, the relative error covariance matrix is a useful tool in understanding this sensor fusion problem. In the cross-well case however we have the additional ability to analyze the detail information in both the x and z directions. For this experiment, we have dense observations on either vertical edge and a rather sparse horizontal sampling. Thus, we anticipate that our ability to resolve detail in these two directions will be significantly different and this difference should be captured via the RECM analysis. In Figure 11, the finest scales supported in the reconstruction in both the x and z directions are plotted as a function of position for $\tau = 0.50$ for the two cases where data from $D_{HI,MID}$ and $D_{HI,MID,LO}$ respectively are available for the reconstruction. From Figure 11(a)-(b) we see that given only high and middle frequency information, detail in the reconstruction is limited to the region near $x = 0$ in both x and z which is consistent with the actual estimate in Figure 10(a). Figure 11(c)-(d) shows that the addition of the low-frequency measurements significantly raises the level of detail in the reconstruction over the right half of the region of interest which is in accord with the intuition provided by the structure of the kernel functions associated with these observations. Specifically, we note that the *minimum* level of z oriented detail increases from 2 in Figure 11(a) to 3 in Figure 11(c). Moreover, the finest scale of horizontal detail moves from 1 to 2 in the area near the right vertical edge.

Finally, $\hat{g}_{0.5}(D_{HI,MID,LO})$, the truncated estimate of g defined in (24), is plotted in Figure 10(d). In this case $\hat{g}_{0.5}(D_{HI,MID,LO})$, is composed of only 75 nonzero wavelet coefficients as opposed to the 256 in the original corresponding to a 70% reduction in inversion complexity. Visual comparison of this reconstruction with the full, untruncated estimate indicates that all of the features captured in the optimal estimate are in fact present in the truncated

version as well.

6   Conclusion and Future Work

In this paper, we have presented an approach to the solution of the inverse scattering problem in the Born approximation based upon techniques drawn from the fields of multiscale modeling, wavelet transforms, and statistical estimation. We begin with a system of noisy, linear integral equations describing the relationship between the scattered fields and the function to be estimated, g. After discretization, wavelet methods are used to transform the problem from real-space to scale-space. A linear least squares estimator serves as the inversion algorithm and produces a multiresolution estimate of g, i.e. an estimate of its wavelet transform γ. Regularization is achieved via a statistical model of γ which also provides a means of capturing any available prior information regarding the structure of g. The structure of this model allows us considerable flexibility in capturing the statistical structure of g, including the incorporation of scale-varying statistics. To illustrate our methods, we have used one of many possible statistical models, namely one that has the $1/f$-like fractal structure that is often posited as a meaningful model for natural phenomena.

Our approach makes extensive use of scale-space in the analysis of linear inverse problems. The *relative error covariance matrix* (RECM) represents a quantitative tool for understanding the various ways in which data from a multitude of sensors contribute to the final reconstruction of g. We demonstrate a method for determining the optimal level of detail to include in the estimate of g as a function of spatial location. The RECM explicitly provides a means of capturing the way in which this level is affected by changing the information used

in the inversion. For example, the incremental benefits associated with the addition of data from another sensor is readily explored using the RECM. Finally, we have shown the use of this quantity in describing the process of multisensor data fusion in a wavelet setting.

The RECM analysis can be of great use in the design of inversion experiments. Because the RECM is not a function of the data, one can evaluate and therefore alter the experimental configuration prior to actually collecting data. Moreover, having settled on the characteristics of the data sources, the RECM can be used to understand precisely where in a parameterization of g (i.e for which degrees of freedom) the data contributes useful and significant information thereby leading to a substantial reduction in inversion complexity.

We note that the general methodologies presented here are not restricted to the case of a Born linearization but in fact should prove to be useful in the analysis of other forms of the inverse scattering problem as well. Because the mathematical description of the problem in the less restrictive Rytov linearization is virtually identical to that obtained in the Born limit [5,6], our multiscale technique should find application in a variety of applications where the former approach is most appropriate. Additionally, in recent years, a variety of algorithms for solving the full nonlinear inverse scattering problem have been proposed [13,24]. In [13], it is necessary to expand the function g in an appropriate basis [13] which "is influenced by the *a priori* information available on the unknown profile." The use of a wavelet basis and fractal-type of regularizer would in this case be rather natural alternatives. Moreover, the algorithms in [13,24] employ variants of Newton's method to execute the nonlinear inversion. As discussed in [19], at each step of the algorithm one is faced with solving a linear systems which essentially has the same form as the least squares problem considered in this paper. Thus, by considering a multiscale form of the problem, a RECM analysis could be used to

determine which coefficients in the wavelet expansion should be determined at each iteration. Given that relatively few such elements were recoverable for the examples considered in this paper, the use of a multiscale formalism in connection with these nonlinear problems could provide significant computational savings.

Finally, although not considered extensively in this work, the multiscale, statistically based inversion algorithms admit highly efficient implementations. As discussed by Beylkin *et. al* in [2], wavelet transforms of many operator matrices, including those arising in the problem studied here, contain very few significant elements so that zeroing the remainder lead to sparse matrices Θ_i. The sparsity of Θ_i combined with the diagonal structure of P_0 imply that highly efficient, iterative algorithms such as LSQR [21] can be used to solve the normal equations. Indeed, in [19], we consider the development of a modified form of LSQR designed for the efficient and stable computation of $\hat{\gamma}$ as well as arbitrary elements in the error covariance and relative error covariance matrices.

References

[1] R.H.T. Bates, V.A. Smith, and R.D. Murch. Manageable multidimensional inverse scattering theory. *Physics Reports (Review section of Physics Letters)*, 201(4):185–277, 1991.

[2] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Communications on Pure and Applied Mathematics*, 44:141–183, 1991.

[3] D. L. Colton and P. B. Monk. The inverse scattering problem for time-harmonic acoustic waves in an inhomogeneous medium: Numerical experiments. *IMA Journal of Applied Mathematics*, 42:77–95, 1989.

[4] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–996, 1988.

[5] A. J. Devaney. A filtered backprjection algorithm for diffraction tomography. *Ultrasonic Imaging*, 4:336–350, 1982.

[6] A. J. Devaney. Geophysical diffraction tomography. *IEEE Trans. on Geoscience and Remote Sensing*, GE-22(1):3–13, January 1984.

[7] J. Feder. *Fractals*. Pergamon, New York, 1988.

[8] Patrick Flandrin. Wavelet analysis and synthesis of fractional brownian motion. *IEEE Trans. Information Theory*, 38(2):910–917, March 1992.

[9] M. E. Gregotski and O. Jensen. Fractal modeling techniques for spatial data. *IEEE Trans. Geoscience and Remote Sensing*, 31(5):980–988, September 1993.

[10] C. W. Groetsch. *The Theory of Tikhonov regularization for Fredholm equations of the first kind.* Pitman Publishing Limited, Boston, 1984.

[11] T. M. Habashy, W. C. Chew, and E. Y. Chow. Simultaneous reconstruction of permittivity and conductivity profiles in a radially inhomogenous slab. *Radio Science*, 21(4):635–645, July–August 1986.

[12] T. M. Habashy and R. Mittra. On some inverse methods in electromagnetics. *Journal of Electromagnetic Waves and Applications*, 1(1):25–58, 1987.

[13] 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.

[14] 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.

[15] Per Christian Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561–580, December 1992.

[16] R. F. Harrington. *Field Computations by Moment Methods.* Macmillan Publ. Co., 1968.

[17] Ranier Kress. *Linear Integral Equations.* Springer-verlag, Berlin, 1989.

[18] Mark R. Luettgen, W. Clem Karl, and Alan S. Willsky. Efficient multiscale regularization with applications to the compuation of optical flow. *IEEE Transactions on Image Processing*, 3(1):41–64, 1994.

[19] Eric L. Miller. *The Application of Multiscale and Statistical Technoques to the Solution of Inverse Problems.* PhD thesis, MIT, August 1994.

[20] Eric L. Miller and Alan S. Willsky. A multiscale approach to sensor fusion and the solution of linear inverse problems. submitted to *Applied and Computational Harmonic Analysis*, December 1993.

[21] C. C. Paige and M. A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, 8(1):43–71, March 1982.

[22] 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.

[23] Anton G. Tijhuis. Born-type reconstruction of material parameters of an inhomogeneous, lossy dielectric slab from reflected field data. *Wave Motion*, 11:151–173, 1989.

[24] Carlos Torres Verdín and Tarek M. Habashy. Rapid 2.5-D forward modelling and inversion via a new nonlinear scattering approximation. *To appear in Radio Science*, July-August 1994.

[25] Donald L. Turcotte. *Fractals and chaoes in geology and geophysics.* Cambridge University Press, 1992.

[26] G. W. Wornell. A Karhuenen-Loeve-like expansion for 1/f processes via wavelets. *I.E.E.E. Transactions on Information Theory*, 36:859–861, July 1990.

7   Figures and Tables
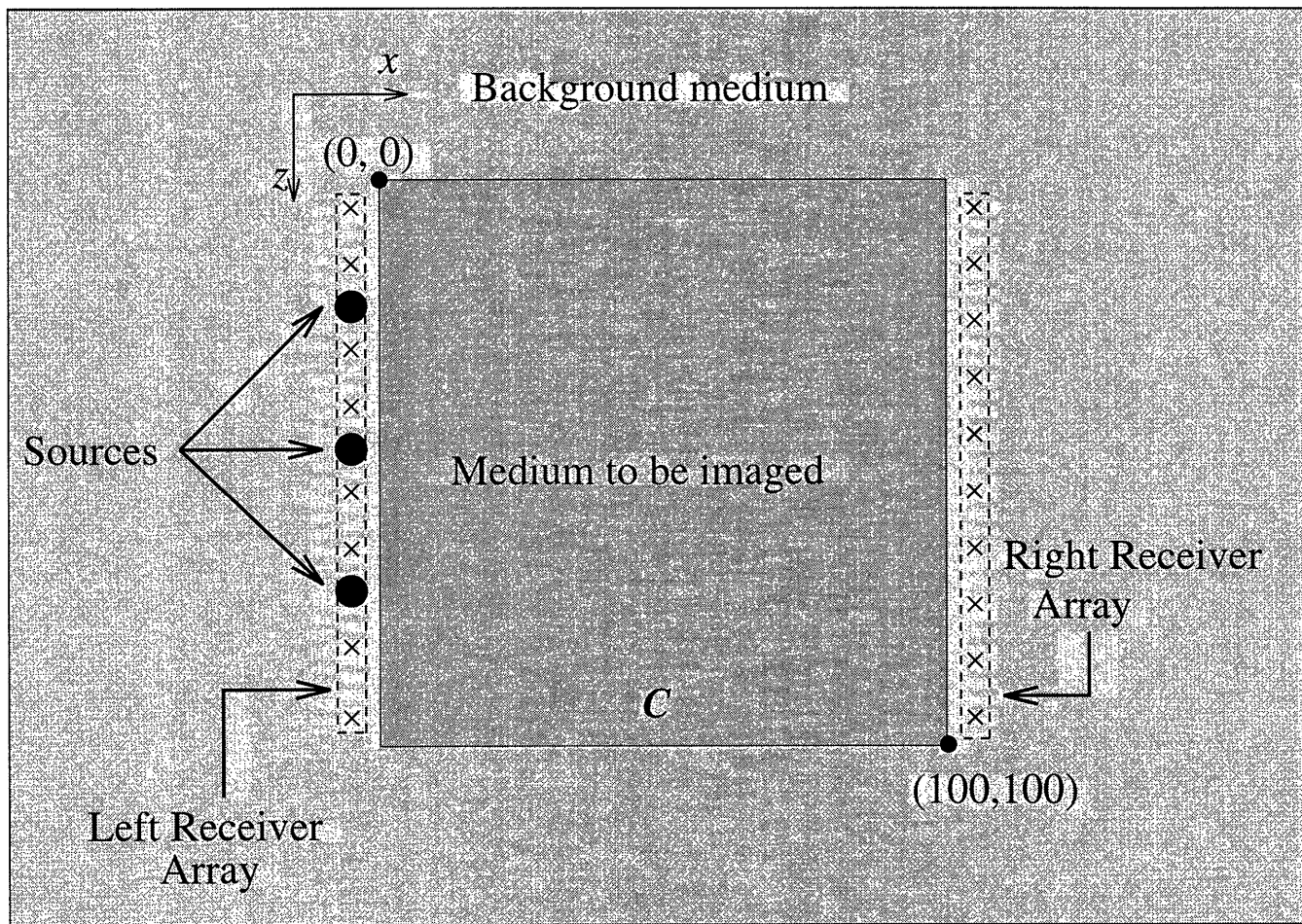
Figure 1: Configuration of inverse conductivity problem. The electromagnetic sources (indicated by the black circles) emit time-harmonic waves into a lossy medium which subsequently are scattered by conductivity inhomogeneities located in the darkly shaded rectangle, C. The secondary fields are observed 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.
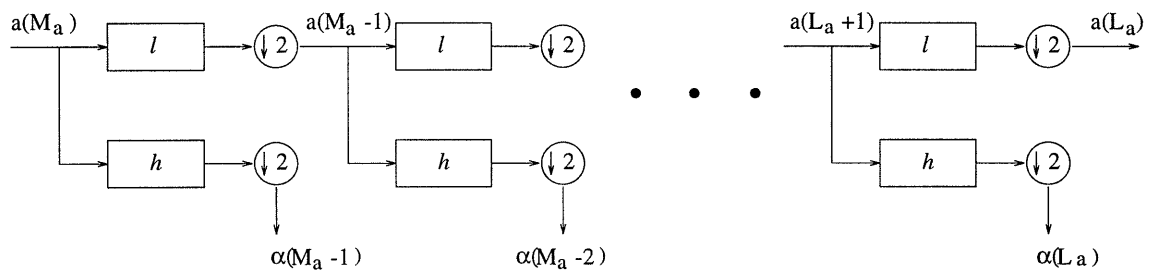
Figure 2: Wavelet transform pyramid.
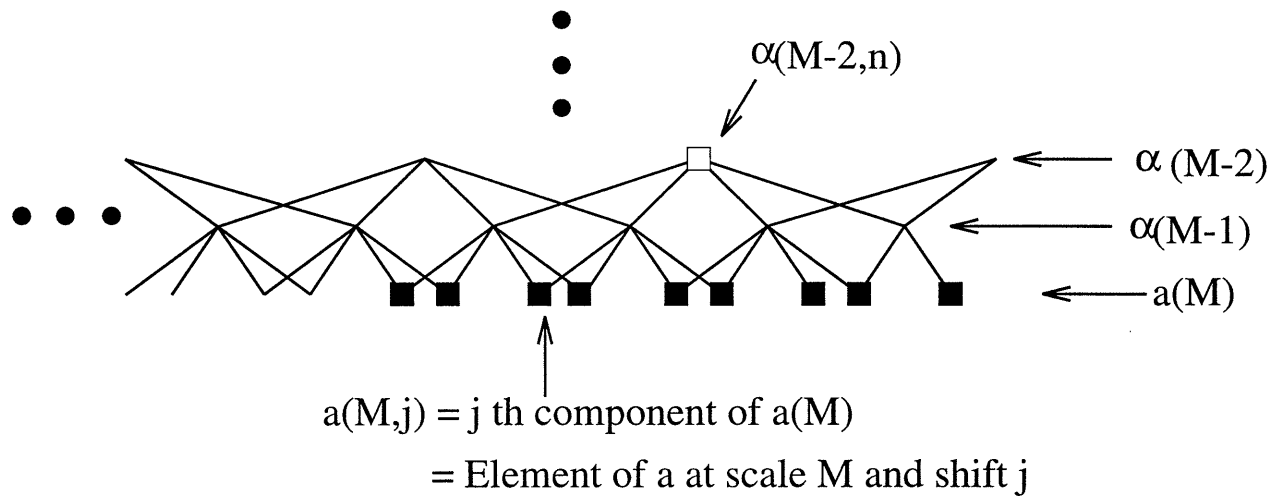
Figure 3: A sample lattice structure corresponding to a D4 wavelet transform. The finest scale is taken as M_a while the coarsest is L_a.
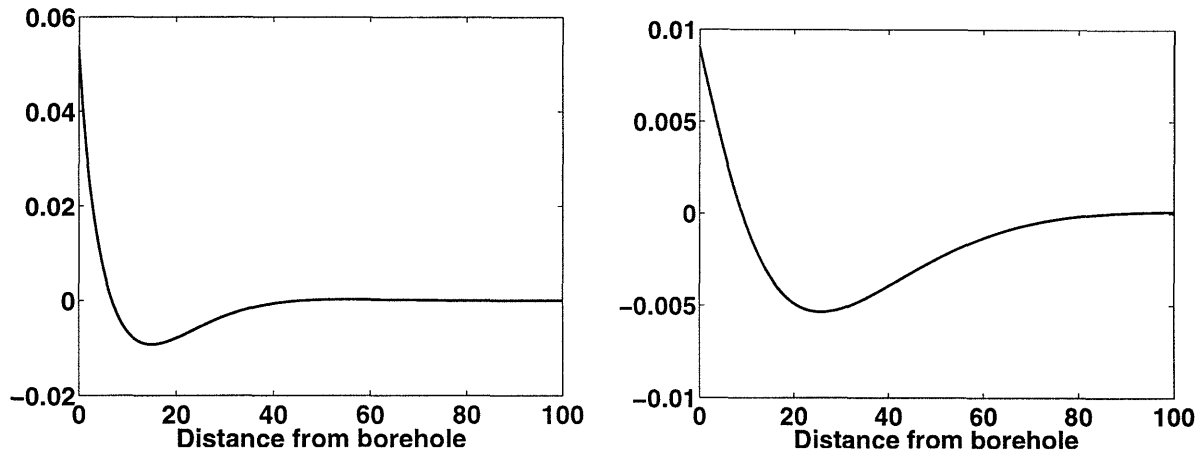
(a) Typical structure associated with high frequency kernel

(b) Typical structure associated with middle frequency kernel

Figure 4: Typical structure of kernel functions used in the reconstruction of g for the radial profiling example.

(a) g (solid line) versus $\hat{g}(D_{HI,MID})$ (dashed line)

(b) $\hat{g}(D_{HI,MID})$ (solid line) versus $\hat{g}(D_{HI})$ (dashed line)

(c) $\hat{g}(D_{HI,MID})$ (solid line) versus $\hat{g}(D_{MID})$ (dashed line)
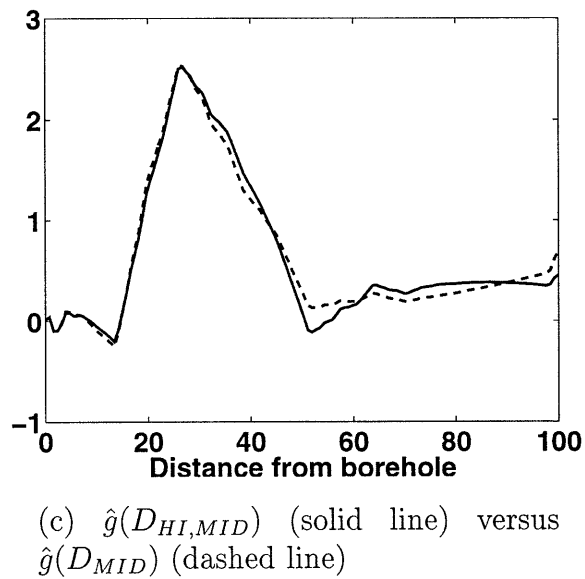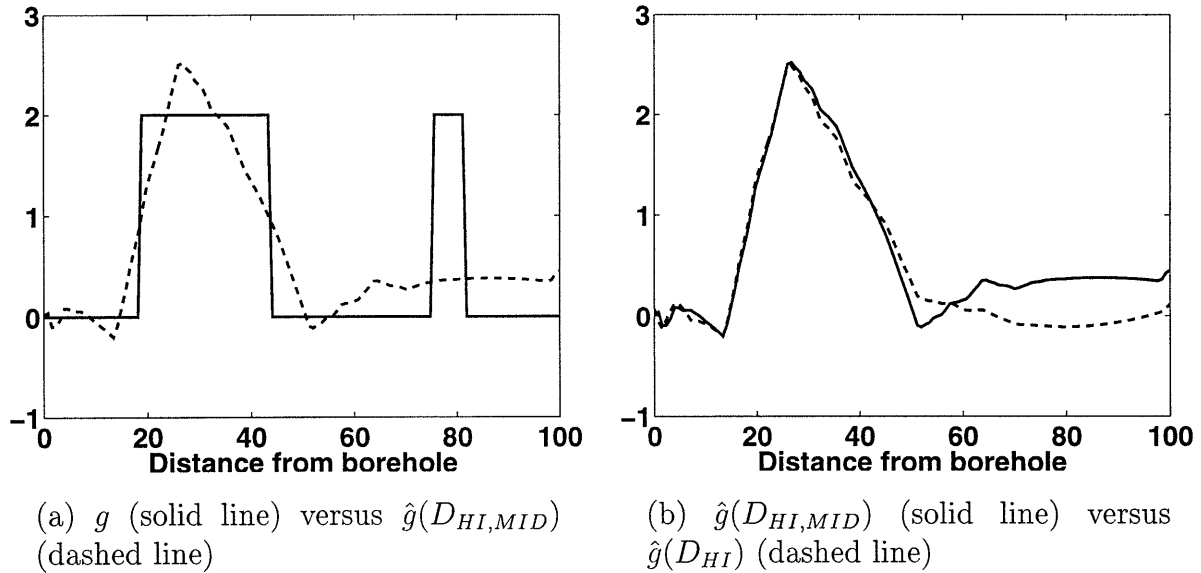
Figure 5: Estimates of g using various combinations of high and middle frequency data. We note that in all cases, the measurements provide sufficient information to reconstruct only those features of g near $x = 0$. At points further from the origin, only the coarsest scale characteristics of g are resolvable. Moreover, as $\hat{g}(D_{HI,MID})$ is significantly different from both $\hat{g}(D_{HI})$ and $\hat{g}(D_{MID})$ we conclude that some type of sensor fusion is occurring over the region far from $x = 0$.
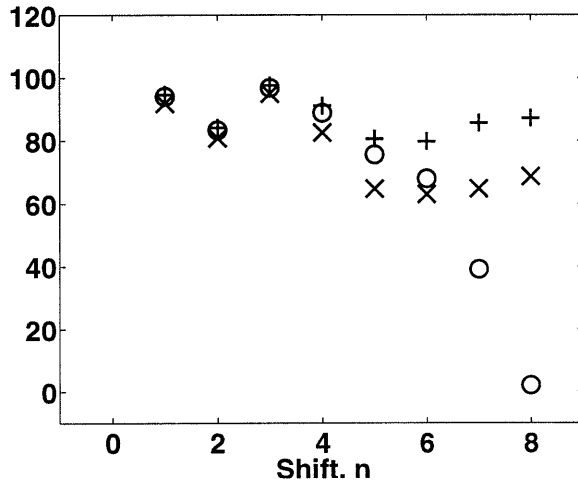
(a) RECM information for coarsest scaling coefficients (i.e. scale $\overline{3}$)
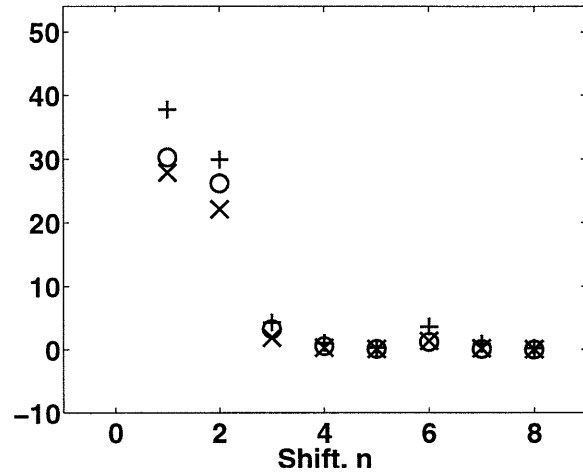
(b) RECM information for wavelet coefficients at scale 3
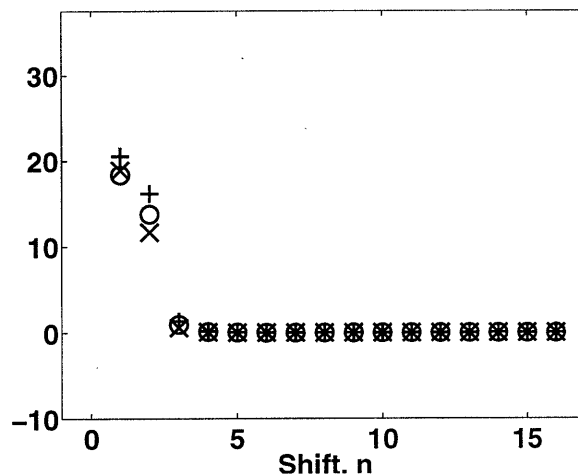
(c) RECM information for wavelet coefficients at scale 4
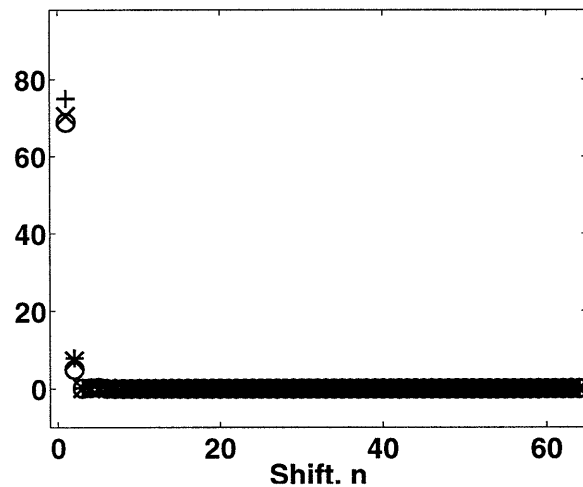
(d) RECM information for wavelet coefficients at scale 6

Figure 6: Diagonal elements of relative error covariances for three radial profiling experiments. In all cases, the symbol "+" corresponds to $\Pi(D_{HI,MID})$, "o" to $\Pi(D_{HI})$ and "×" to $\Pi(D_{MID})$. From (a) we see a significant level of sensor fusion taking place with respect to the estimates of the coarsest scale scaling coefficients far from the origin $x = 0$. From (b)–(d), we conclude that accurate reconstruction of the detail components of g is limited to shifts close to $x = 0$.
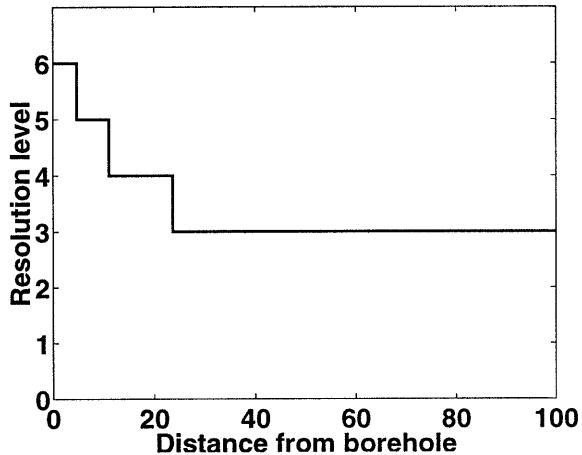
(a) The optimal scale of reconstruction as a function of position at scale $M_g = 7$ for a threshold value of $\tau = 0.05$.
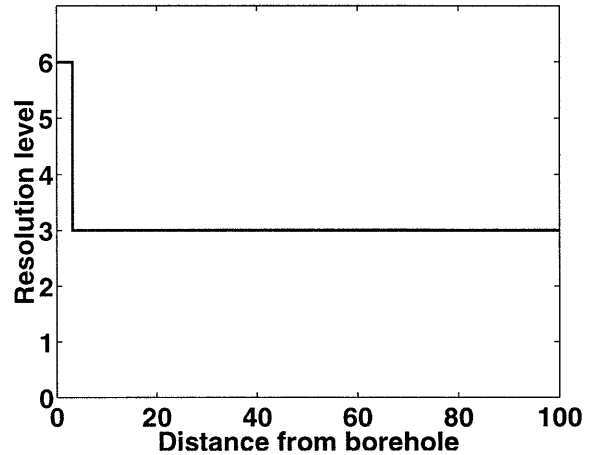
(b) The optimal scale of reconstruction as a function of position at scale $M_g = 7$ for a threshold value of $\tau = 0.50$.
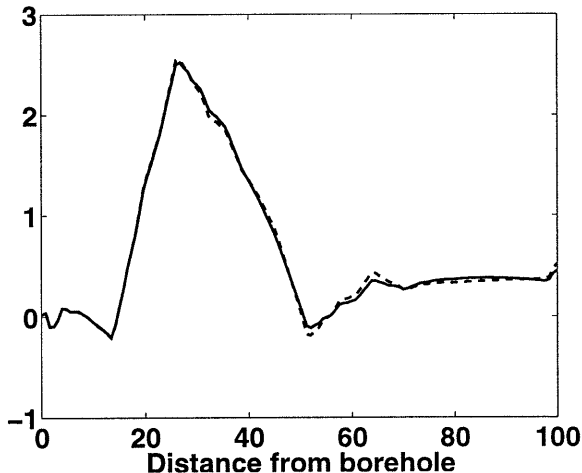
(c) $\hat{g}$ (solid line) vs $\hat{g}_{0.05}$ (dashed line)
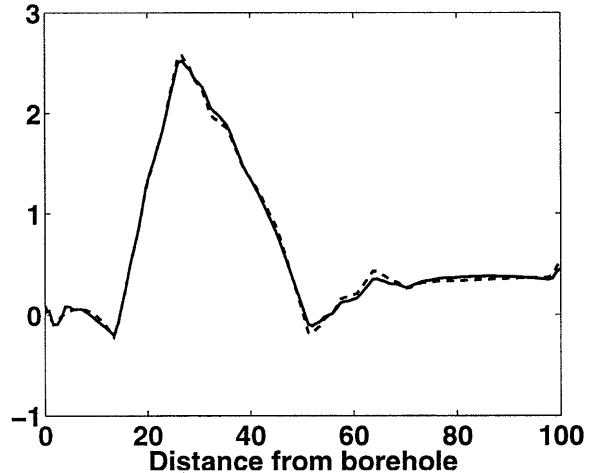
(d) $\hat{g}$ (solid line) vs $\hat{g}_{0.50}$ (dashed line)

Figure 7: Maps of the optimal scale of reconstruction and the associated estimates of g for threshold values $\tau \in \{0.05, 0.50\}$. These illustrations provide a quantitative verification of the intuition that resolution in the inversion should drop as a function of distance from the origin. In (c) and (d), the plots of $\hat{g}$ against $\hat{g}_{0.05}$ and $\hat{g}_{0.50}$ respectively show that little is lost in reducing the complexity of the model by eliminating degrees of freedom about which the data provides little or no information.

(a) $\Pi_n^{\bar{3}}(D_{HI}, D_{HI,MID})$

(b) $\Pi_n^6(D_{HI}, D_{HI,MID})$
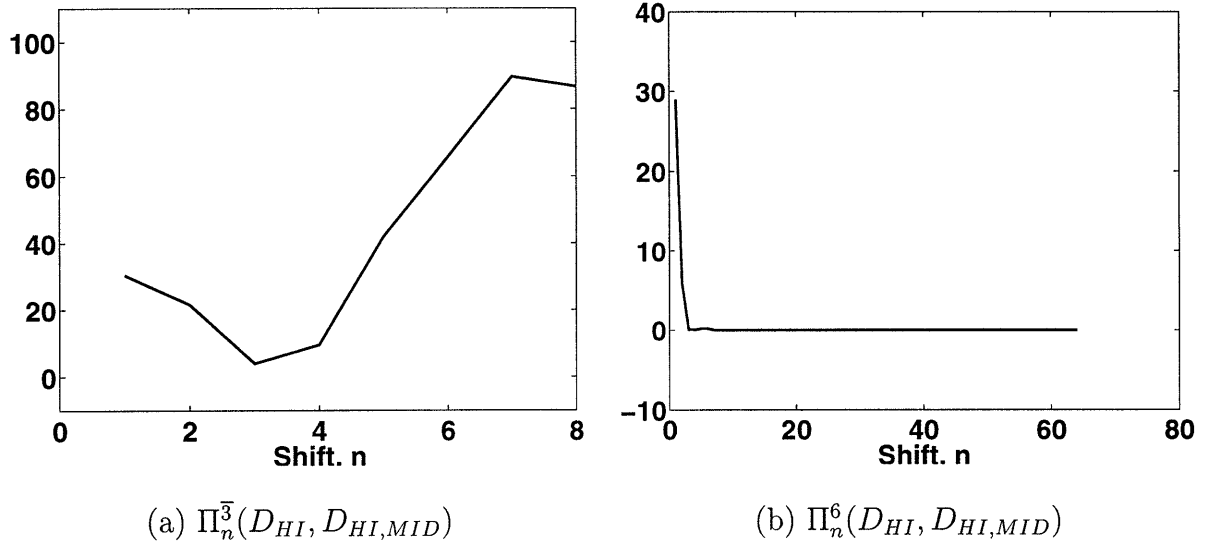
Figure 8: The incremental reduction in uncertainty obtained by adding data from the middle frequency observation to an estimate based upon the high frequency measurement sources. In accordance with Figure 6(a) we see significant benefits associated with determination of both the coarsest scale structure of g far from the origin as well as the finest scale structure closest to $x = 0$.
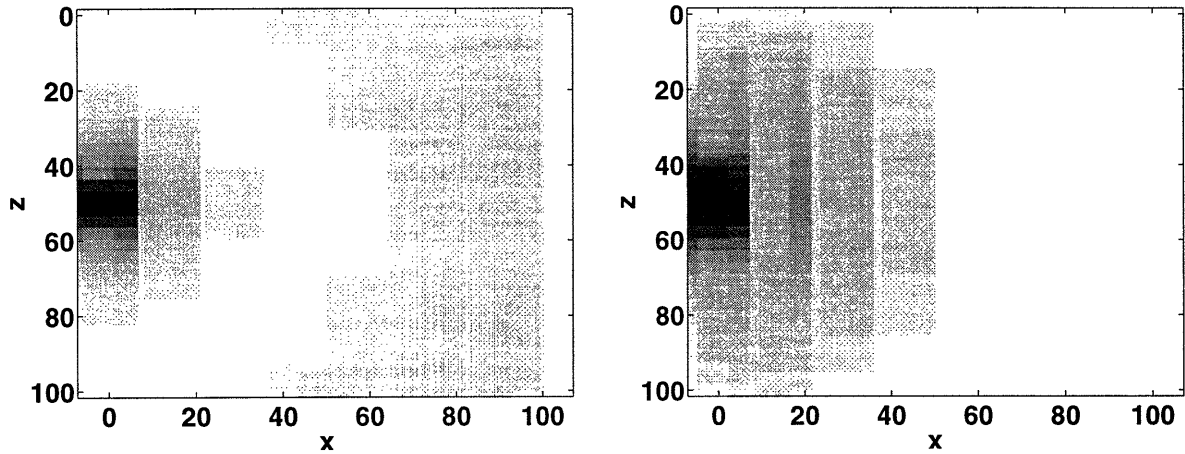
(a) Structure of typical high frequency kernel

(b) Structure of typical middle frequency kernel
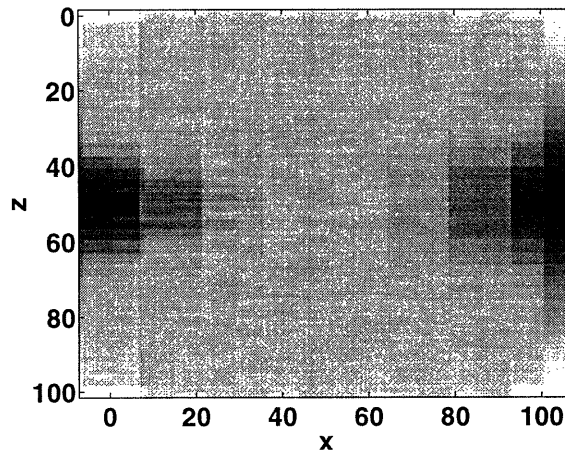
(c) Structure of typical low frequency kernel

Figure 9: Typical structure of kernel functions used in the reconstruction of g for the cross-well tomography example. Each image corresponds to map taking conductivity to the measurement obtained at the center of the left (in (a) and (b)) or right (in (c)) receiver array in response to excitation from the middle source with darker shades indicating larger values. As in the radial profiling example, the high and low frequency kernels are most sensitive to variations in g near the left edge of the square. The low frequency data should aid in the reconstruction of g near the either vertical edge.
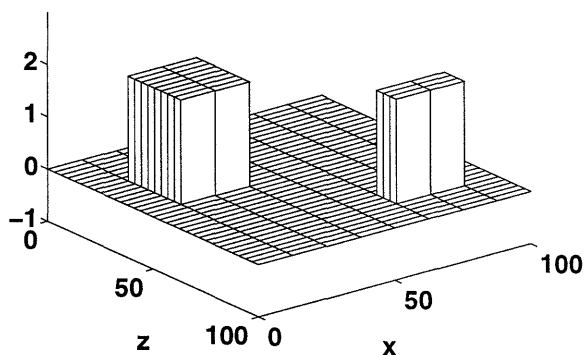
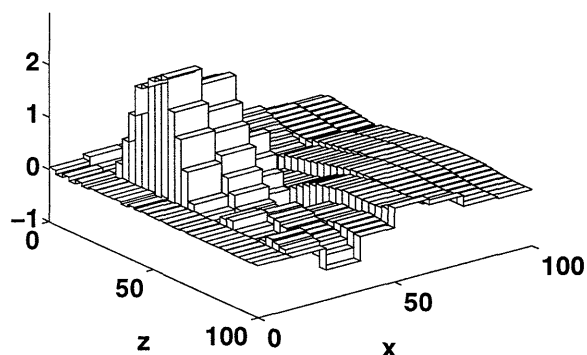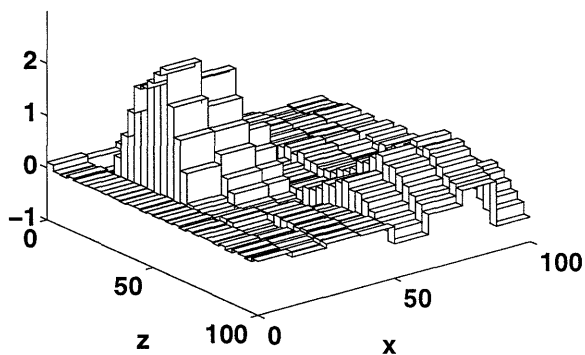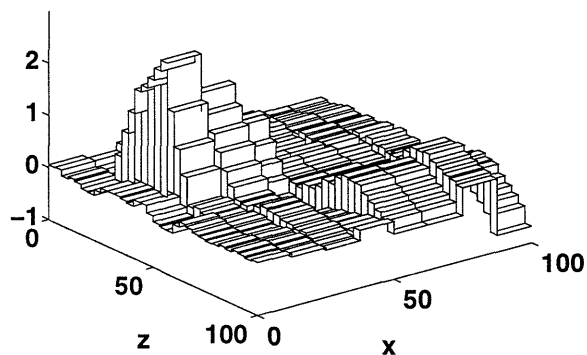(a) Conductivity structure to be reconstructed

(b) $\hat{g}(D_{HI,MID})$

(c) $\hat{g}(D_{HI,MID,LO})$

(d) $\hat{g}_{0.5}(D_{HI,MID,LO})$

Figure 10: In (b)–(d) the estimates of g in (a) are displayed using various combinations of high, middle and low frequency data. From (b), the high and medium frequency information provides insufficient information to reconstruct the anomaly near $x = 100$. As seen in (c), the addition of the low frequency, cross-well data sets clearly improves the ability to resolve this second structure. The truncated estimate $\hat{g}_{0.50}(D_{HI,MID,LO})$. Note that there is little difference between this function, composed of 75 non-zero elements in the wavelet transform domain and the optimal estimate $\hat{g}(D_{HI,MID,LO})$ in (c) which has 256 degrees of freedom.
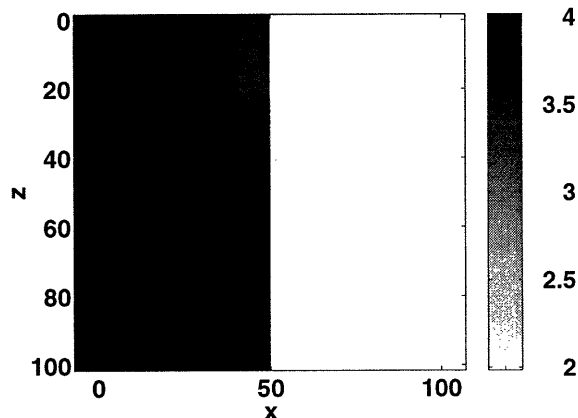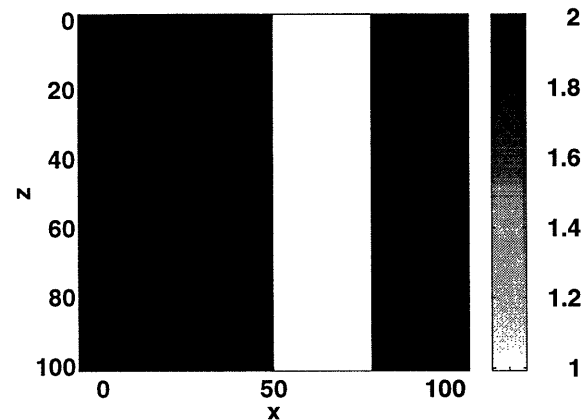
(a) The finest scale of to which z-oriented detail can be reconstructed at $\tau = 0.5$ given only high and middle frequency data.
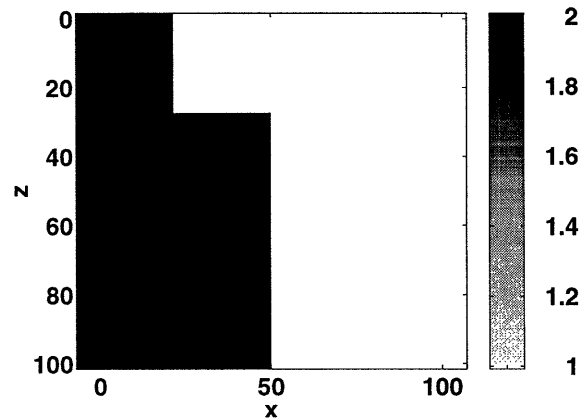
(b) The finest scale of to which x-oriented detail can be reconstructed at $\tau = 0.5$ given only high and middle frequency data.
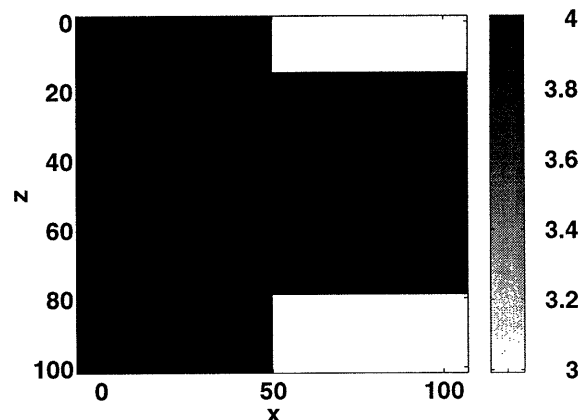
(c) The finest scale of to which z-oriented detail can be reconstructed at $\tau = 0.5$ given high, middle, and low frequency data.

(d) The finest scale of to which x-oriented detail can be reconstructed at $\tau = 0.5$ given high, middle, and low frequency data.

Figure 11: Maps of the optimal scale of reconstruction for the z and x components of detail for the threshold value $\tau = 0.5$. The maps verify of the intuition that the low-frequency, cross-well data provides improved resolution especially in the vicinity of the right vertical edge.

Experiment number	Source Position	Frequency of source (Hz)	Receiver Array
1,2,3	T,M,B	$f_{HI} = 398$	Left
4,5,6	T,M,B	$f_{MID} = 119$	Left
7,8,9	T,M,B	$f_{LO} = 6$	Right

Table 1: Data set definitions for observation processes of interest in the paper. The abbreviations in the column labeled "Source Position" correspond to the *T*op, *M*iddle, and *B*ottom line sources in Figure 1

Parameter	Value
Wavelet	Daubechies 6-tap
M_g	6
L_g	3
μ	1
σ^2	1
p_{L_g}	0.5
SNR^2 for D_{HI}	200
SNR^2 for D_{MID}	400
Background conductivity	1 S/m

Table 2: Parameters for radial profiling problem.

Parameter	Value	Parameter	Value
z Wavelet	Daubechies 6-tap	x Wavelet	Daubechies 2-tap
$M_{g,z}$	4	$M_{g,x}$	2
$L_{g,z}$	2	$L_{g,x}$	1
μ_z	1	μ_x	1
σ_z^2	1	σ_x^2	1
$p_{L_{g,z}}$	$\sqrt{2}$	$p_{L_{g,x}}$	$\sqrt{2}$
SNR^2 for D_{HI}	250	SNR^2 for D_{MID}	500
SNR^2 for D_{LO}	1000	Background conductivity	1 S/m

Table 3: Parameters for cross-well tomography problem