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

Eric L. Miller†
Room 235 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

December 13, 1995

**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 [28], as a tool for quantitatively evaluating the manner in which data contribute 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.

*This work was supported in part by the Office of Naval Research under Grant N00014-91-J-1004 and 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

†The work of this author was also supported in part by a US Air Force Laboratory Graduate Fellowship and summer work performed at Schlumberger-Doll Research

# 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. Indeed, these *inverse scattering* problems are found in fields such as medical imaging, nondestructive testing, oceanography, and remote sensing [2]. Here, we present and demonstrate a methodology for the use of multiresolution and statistical signal processing concepts which provides explicit and quantitative insight regarding many of the challenges associated with problems of this type. First, the use of these methods yields a new and interesting means of regularizing the inverse scattering problem. Additionally, the fact that the wavelet transform provides information regarding the local behavior of a function over a variety of spatial scales leads to the development of a powerful collection of tools for the analyses of sensor fusion and the tradeoff between reconstruction accuracy and resolution. As it is our desire to focus exclusively upon these issues rather than the complexities associated with the physical modeling of the problem, we limit our attention to problems where the first Born approximation [5] is a valid model for the relationship between the quantity to be reconstructed and the associated observations. In particular, we demonstrate our approach using a linearized inverse electrical conductivity problem arising in geophysical exploration [21,33].

A common difficulty associated with many inverse scattering problems, including those of interest in this paper, is that of ill-posedness [3,18] caused by the restriction of collected data to the boundaries of the medium and the physics governing the propagation of radiation through a lossy environment. Traditionally, this problem 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 [18]. Also, a regularizer may be incorporated as a means of constraining the reconstruction to reflect prior knowledge concerning the behavior of this function [18,26,36]. As discussed in [28], such regularization techniques have direct interpretations as specifying prior statistical models on the phenomenon to be imaged. This interpretation provides a basis for the calculations of error statistics which we find useful in the consideration of questions such as the tradeoff between the resolution of the reconstruction and the accuracy of the generated image, the value of additional measurement sets, etc. Moreover, the use of the Born approximation leads to a *linear* estimation problem where these statistics are dependent only upon the model structure and not the actual data values [35]. Thus, all of the analysis tools developed in this work can be implemented off-line so that we may examine the problem prior to the collection of data.

In this paper, the benefits associated with a statistical perspective are further enhanced through the use of a wavelet-based multiresolution framework. Here, we are led to an alternate method for statistical regularization specified directly 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 regularization scheme. In particular, we consider a highly useful class of multiscale prior models, the so-called fractal priors model. As shown in [26], this model is related to the traditional smoothness-based regularizers and, with appropriately chosen parameters, produces estimates with similar characteristics. Moreover, Wornell [38] 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 [14,34] and will therefore be appropriate for the conductivity inverse problem developed in Sections 2 and 4.

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 [28] 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. Moreover, the RECM allows for an

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

explicit description of the information provided by the various data sources both individually and collectively at each point in space and scale. Thus, we develop new, quantitative insight into the issue of multisensor data fusion in that (a) we are able to determine those regions in space where the information provided by several data sets together significantly exceeds that provided by any one set individually and (b) we use the RECM to assess the incremental value of additional sources of information.

In Section 2, we develop the physical model relating the observables to the conductivity field and review common inversion methods for this problem. In Sections 3.1 and 3.2, we present an overview of the discrete wavelet transform and a discussion of its application to the inverse scattering problem. Sections 3.3 and 3.4 respectively contain the description of the fractal multiresolution statistical regularization formulation and the tools for RECM-based analysis. A set of examples illustrative of the different facets of our approach are presented in Section 4. Conclusions and directions for further investigation are given in Section 5.

# 2 An Inverse Electrical Conductivity Problem

## 2.1 A Linear Forward Model

We consider an inverse scattering problem illustrated in Figure 1 and arising in the context of geophysical exploration [21, 33]. 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 decomposed into the sum of: (1) an infinite, known, and constant background and (2) a conductivity anomaly, $g$, which varies as a function of the two variable $x$ and $z$ and which is known to lie in region $C$ of the plane. The transmitted energy is scattered by $g$ and the resulting field is measured by one of the two arrays of point receivers. We consider inversions based upon the data obtained from a number of scattering experiments the details of which are provided in Table 1. Each experiment produces a vector of measurements comprised of the scattered field (both in-phase and quadrature components) observed over a single receiver array due to energy put into the medium from one of the three sources operating at a particular frequency. In particular, the left receiver array consists of 64 equally spaced elements extending from $z = 4.8$m to $z = 95.2$m all located at $x = -0.2$m. Similarly, the right array is composed of 48 equally spaced elements from $z = 14.9$m to $z = 82.2$m all located at $x = 100.2$m. The three sources are located along the line $x = -0.1$m at $z = 25$m, 50m, and 75m.

In this paper, we consider the situation in which the exact, nonlinear physical relationship between the observed scattered fields and the conductivity perturbation $g$ is represented accurately by the linear relationship obtained through the use of the first Born approximation [5]. The conditions under which such an approximation may be made are well documented in [5, 22] and are assumed to hold for the remainder of this work. For the $i^{th}$ 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 [20] which,

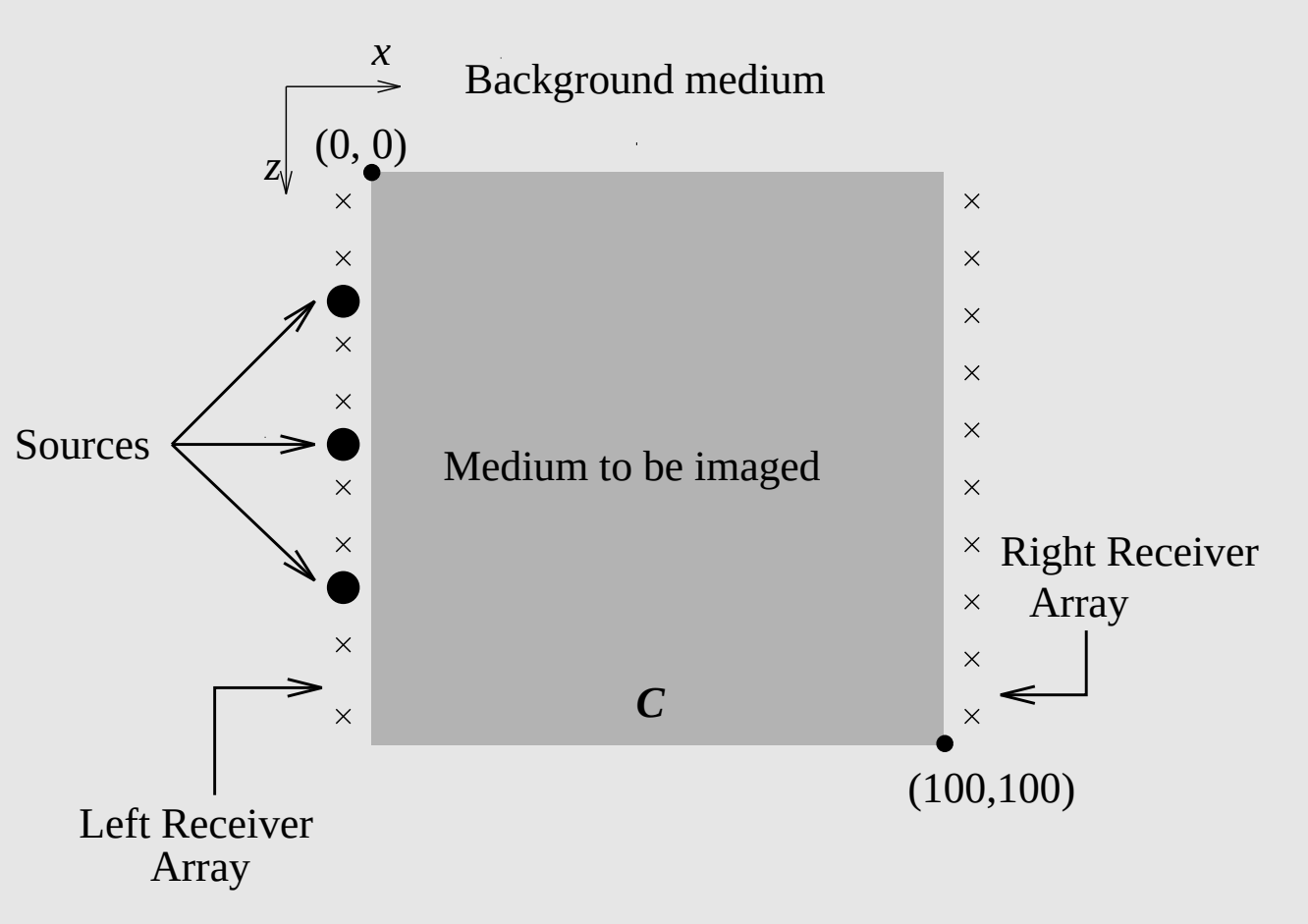



Figure 1: Configuration of inverse conductivity problem. The electromagnetic sources (indicated by the black circles) emit time-harmonic waves into a lossy medium which subsequently are scattered by conductivity inhomogeneities located in the darkly shaded rectangle, $C$. The secondary fields are observed 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.

upon discretization using a method of moments (MOM) approach with a pulse basis [23], yields the matrix-vector observation model

$$y_i = T_i g + n_i \tag{1}$$

where $y_i$ is the vector of scattered field observations obtained along the receiver array, $T_i$ is the MOM operator matrix associated with the Fredholm integral kernel, $g$ is the vector containing the MOM expansion coefficients of the conductivity perturbation and $n_i$ represents additive noise which is is taken to be uncorrelated and stationary so that $n_i \sim (0, r_i I)$ where $I$ is an appropriately sized identity matrix[1]. We define the "stacked" system of data as

$$y = Tg + n \tag{2}$$

where $y = [y_1^T \;\; y_2^T \;\; \ldots y_K^T]^T$, with $T$ and $n$ defined accordingly. Thus, the objective of the inverse

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

scattering problem is the determination of $g$ given the data and the model in (2).

### 2.2 Regularized Inversion and Its Probabilistic Interpretation

A commonly used technique [18,25,36] for solving linear inverse problems of the form in (2) is to choose the estimate of $g$ according to

$$\hat{g} = \arg\min_g \|y - Tg\|^2_{\mathcal{R}^{-1}} + \|g\|^2_{L^TL} \tag{3}$$

where $\|x\|^2_M = x^T M x$. 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 or as a prior statistical model for $g$. In particular, as discussed in [26], 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 $L$. Common choices for this term are discussed in [3,8,24,31].

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

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

As discussed in [26], this solution, $\hat{g}$ can be interpreted as the linear least squares estimate (LLSE) of $g$ given the noisy measurements in (2) with $n \sim (0, \mathcal{R})$ and the prior statistics for $g$ implied by (4), i.e. $g$ is zero mean and has $L^TL$ as the *inverse* of its covariance. Furthermore, the estimation error covariance matrix, i.e. the covariance of $g - \hat{g}$ is

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

## 3 A Multiscale Representation of the Problem

We are interested in the use of estimation-theoretic methods for solving the linearized inverse scattering problem in which the statistical quantities are examined and the inversion executed directly in the wavelet transform domain. Hence, we begin this section with a brief overview of the discrete wavelet transform (DWT). A more detailed description of this new signal processing tool including details associated with its implementation and examples of its application may be found in [1,4,6,7,10–12, 16,17,37]. Subsequent to its introduction, we describe in greater depth the manner in which the DWT can be employed in the context of the linearized inverse scattering problem.

### 3.1 A Wavelet Representation of $g$ and $y$

The fundamental idea behind the discrete wavelet transform 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 $M_a$ where the integer $M_a$ reflects the dimensionality of $a$.

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

[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 $\|L(g - g_0)\|^2$

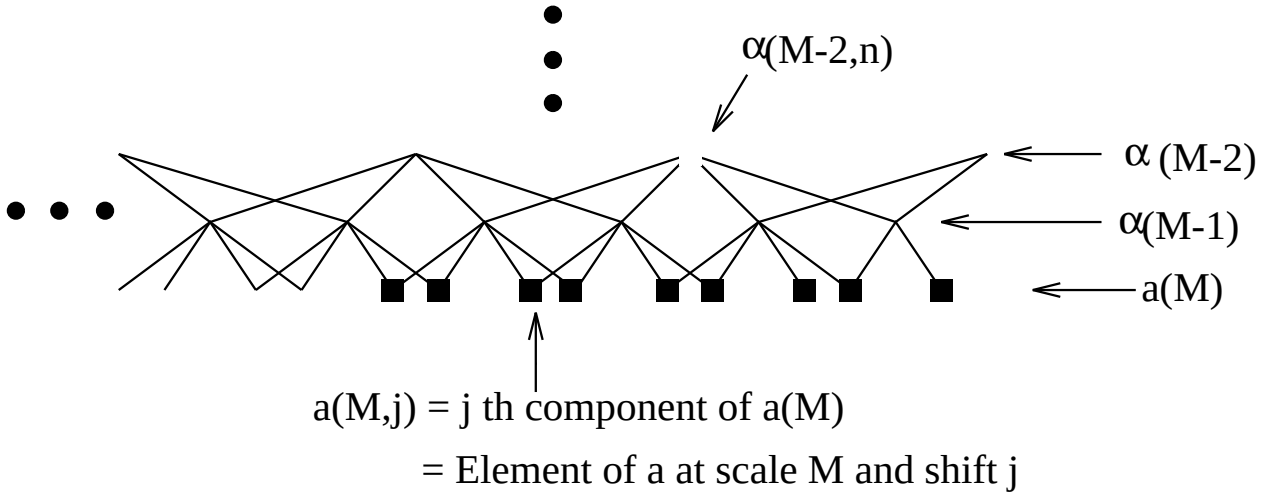



Figure 2: A sample lattice structure corresponding to a D4 wavelet transform. The finest scale is taken as $M_a$ while the coarsest is $L_a$. Also, the downward impact set associated with node matked by an open square, is comprised of all nodes marked with the closed square.

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.

As described extensively in [9,10,16], the filters $l$ and $h$ are designed so that we may construct an operator,[3] $\mathcal{W}_a$ 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

$$\alpha = \mathcal{W}_a a \tag{7}$$

where $\alpha = [\alpha(M_a - 1)^T \ \ldots \ \alpha(L_a)^T \ a(L_a)^T]^T$. Moreover, $\mathcal{W}_a$ can be made to be orthonormal so that the equality $\mathcal{W}_a^* \mathcal{W}_a = I$ holds. We call the vector $\alpha$ the *wavelet transform* of $a$. The $n^{th}$ element of $\alpha(m)$ is denoted $\alpha(m, n)$ and is termed the $n^{th}$ *shift* of $\alpha$ at scale $m$. Also, $a(m, n)$ represents the $n^{th}$ element of the vector of scaling coefficients at scale $m$. In general, we use Roman letters to represent vectors of approximation coefficients and their Greek counterparts for detail coefficients.

The relationships among the scale space component in the decomposition of $a$ are graphically represented in the form of a lattice as shown in Figure 2 for the case of a wavelet decomposition with $l(n)$ and $h(n)$ of length 4. The coefficients 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. Two nodes are connected by an arc if and only if there is a linear relationship between the two as dictated by the structure of the wavelet transform matrix $\mathcal{W}_a$. We say that a coarse scale node *impacts* a finer scale one if there exists a strictly downward path on the lattice from the former to the latter. Finally, as illustrated in Figure 2, 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.

[3] 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.

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. 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, it is easily shown that $\mathcal{W}_a^* \mathcal{W}_a = I$. As in the 1D case, we denote a particular element of $\alpha$ 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 and define downward impact sets in the same manner as was the case in 1D.

### 3.2 A Multiscale Inversion Problem

Using the methods in Section 3.1 we define orthonormal, discrete wavelet transform operators $W_i$ and $W_g$ which transform the measurement vectors, $y_i$ and the discretized conductivity field, $g$, into their respective wavelet decompositions:

$$\eta_i = \mathcal{W}_i y_i \quad \text{and} \quad \gamma = \mathcal{W}_g g.$$

In our analysis of (1), we use $W_i$ and $W_g$ to move from physical to scale space via

$$\eta_i = \mathcal{W}_i y_i = (\mathcal{W}_i T_i \mathcal{W}_g^*)(\mathcal{W}_g g) + \mathcal{W}_i n_i \equiv \Theta_i \gamma + \nu_i. \tag{8}$$

Finally, the wavelet-domain stacked system is

$$\eta = \Theta \gamma + \nu \tag{9}$$

with $\eta = [\eta_1^T \; \eta_2^T \; \ldots \eta_K^T]^T$ and $\Theta$ and $\nu$ are defined accordingly.

Analogously to Section 2.2, we define a linear least squares estimation problem in the wavelet transform domain. Specifically, we wish to reconstruct $\gamma$ based on the prior model $\gamma \sim (0, P_0)$, together with the noisy measurements (9). Thus, the LLSE, $\hat{\gamma}$, is

$$\hat{\gamma} = \arg\min_{\gamma} \|\eta - \Theta\gamma\|_{R^{-1}}^2 + \|\gamma\|_{P_0^{-1}}^2 \tag{10}$$

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

$$(\Theta^T R^{-1} \Theta + P_0^{-1})\hat{\gamma} = \Theta^T R^{-1} \eta \tag{11}$$

and the corresponding error-covariance matrix is given by

$$P = E[(\gamma - \hat{\gamma})(\gamma - \hat{\gamma})^T] = (\Theta^T R^{-1} \Theta + P_0^{-1})^{-1}. \tag{12}$$

Comparing (3), (5), and (6) to Eqs. (10), (11), and (12), 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, as we will see in Section 3.3, 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 different choices for the diagonal matrix $P_0^{-1}$ allow us additional flexibility to capture a rich variety of other regularization objectives or prior models. Second, as discussed in Section 3.4, the use of these prior models is instrumental in obtaining a great deal of insight into issues associated with the manner in which the information embedded in the data impacts the structure of the reconstructed conductivity field.

### 3.3 Multiscale Prior Models

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 $w \sim (0, I)$. Work by Wornell and others [15, 32] has demonstrated that Brownian motions and other related fractal processes can be closely approximated via a statistical model in which the wavelet and coarsest scale scaling coefficients of $g$ are independent and distributed according to

$$\gamma(m,n) \sim (0, \kappa^2 2^{-\mu m}) \tag{13a}$$

$$g(L_g,n) \sim (0, p_{L_g}). \tag{13b}$$

Here, $\kappa^2$ controls the overall magnitude of the process and the parameter $\mu$ determines the fractal structure of sample paths. The case $\mu = 0$, corresponds to $g$ being white noise while as $\mu$ increases, the sample paths of $g$ show greater long range correlation and smoothness. The scalar $p_{L_g}$ is chosen to be sufficiently large number so as to avoid any bias in the estimator of the low frequency structure of $g$. Finally, for these models, the matrix $P_0$ in (11) is *diagonal* with nonzero entries corresponding to the variances associated with each component of $\gamma$.

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 $L_{g,x} \leq m_x \leq M_{g,x} - 1$ and $L_{g,z} \leq m_z \leq M_{g,z} - 1$. For $m_x = L_{g,x}$ we take $\gamma(m,n) \sim (0, p_{L_{g,x}} \kappa_z^2 2^{-(\mu_z m_z)})$ with analogous results holding when $m_z = L_{g,z}$.

Clearly, other choices of statistics for the components of $\gamma$ may be appropriate in specific applications, and our methodology can readily accommodate these. The choice we have made, leading to a $1/f$-like fractal model, is useful both in its ability to model natural phenomena [14, 34, 38] 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 4, this is precisely the type of regularization required for the inverse conductivity problem. Additionally, we observe that the methods presented in the paper are not dependent upon a model with the variance structure of (13); rather, all that is required is the uncorrelated property of the wavelet coefficients (i.e. a diagonal $P_0$.) In fact, work performed in [6,7,13,16,26] indicates that a rich variety of stochastic phenomena can be described using more general wavelet-type models with uncorrelated coefficients thereby demonstrating that the methods presented in this paper are applicable in a wide range of circumstances.

### 3.4 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 $\gamma$ but also an indication as to the quality of this reconstruction in the form of the error-covariance matrix $P$ defined in (12). While the information contained in $P$ is certainly important for evaluating the absolute level of uncertainty associated with the estimator, in many cases, we have some prior level of confidence in our knowledge of $\gamma$ and we seek to comprehend how the inclusion of additional data in our estimate of $\gamma$ 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 impact a reconstruction.

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

$$1 - \frac{b}{a}. \tag{14}$$

The matrix analog to (14) 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 (12) resulting from the estimation of $\gamma$ 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

$$\Pi(A,B) = I - P_A^{-T/2} P_B P_A^{-1/2} \tag{15}$$

where $P_A^{T/2} \equiv (P_A^T)^{1/2}$.

The definition of $\Pi(A,B)$ in (15) possesses many useful properties. First, the structure of the RECM is dependent upon the model relating the data to the observations (i.e. $\Theta$, $R$, and $P_0$) but *not* upon the actual values of the data. Thus, $\Pi(A,B)$ may be computed off-line and the associated

analysis performed prior to the collection of data. Second, like the error-covariance matrix,the RECM 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

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

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

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

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 $\gamma$ 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 $i^{th}$ component of $\gamma$ arising from an estimate based upon data from set $A$. Then, the $i^{th}$ component of the diagonal of $\Pi(A,B)$ is just

$$1 - \sigma_i^2(B)/\sigma_i^2(A) \tag{16}$$

which is nothing more than the relative size difference of the error-variance in the $i^{th}$ component of $\gamma$ 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. 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. 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. Finally, the expression for $\Pi(B)$ simplifies from (15) to

$$\Pi(B) = I - P_0^{-T/2}(\Theta_B^T R_B^{-1} \Theta_B + P_0^{-1})^{-1} P_0^{-1/2} \tag{17}$$

where $\Theta_B$ and $R_B$ are the system matrix and noise covariance matrix corresponding to an inversion based upon data sets from $B$ and $P_B = (\Theta_B^T R_B^{-1} \Theta_B + P_0^{-1})^{-1}$ is the error-covariance matrix associated with this estimate.

The quantity $\Pi(A,B)$ represents a useful tool for quantitatively analyzing the relationship between the characteristics of the data 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. Letting $\Pi_n^m(B)$ denote the diagonal element of the matrix $\Pi(B)$ corresponding to the wavelet coefficient at scale/shift $(m,n)$[4] provides a natural way to define $m^*(j)$, the appropriate level of detail which should be included in a reconstruction of $g(M_g)$ at shift $j$. 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 3.1, we say that the data support 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 provide 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. 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, $\tau$, between zero and one. Additionally, we are led to define $\hat{\gamma}_\tau$, a truncated version of $\hat{\gamma}$, as follows:

$$[\hat{\gamma}_\tau]_{(m,n)} = \begin{cases} 0 & \Pi_n^m(B) \leq \tau \\ [\hat{\gamma}]_{(m,n)} & \text{otherwise} \end{cases} \tag{18}$$

where $[\hat{\gamma}]_{(m,n)}$ is the component in the vector $\hat{\gamma}$ at scale $m$ and shift $n$. Thus, $\hat{\gamma}_\tau$ is obtained by selecting a value for the threshold $\tau$ between zero and one, determining $\hat{\gamma}$ from (11), and then using (18) to zero all elements in $\hat{\gamma}$ for which the associated value of $\Pi_n^m$ is less than $\tau$. Defining $\hat{\gamma}_\tau$ in this way ensures

[4] 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 $n$.

| Parameter | Value |
|---|---|
| Wavelet | Daubechies 6-tap |
| $M_g$ | 6 |
| $L_g$ | 3 |
| $\mu$ | 1 |
| $\sigma^2$ | 1 |
| $p_{L_g}$ | 0.5 |
| $SNR^2$ for $D_{\mathrm{HI}}$ | 200 |
| $SNR^2$ for $D_{\mathrm{MID}}$ | 400 |
| Background conductivity | 1 S/m |

Table 2: Parameters for radial profiling problem.

that $\hat{g}_\tau = \mathcal{W}^T \hat{\gamma}_\tau$ is in fact a reconstruction of $g$ which at each shift $n$ contains detail information at scales no finer than $m^*(n)$.

# 4 Examples

## 4.1 A One Dimensional, Radial Profiling Problem

We begin by considering a radial profiling problem similar to that analyzed by Habashy *at. al* in [19,21]. 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 dotted line in Figure 3. 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 $\gamma$ and $tr$ is the trace operation. In this example we explore inversions using data from the following three different combinations of the high and middle frequency scattering experiments described in Table 1:

1. $D_{\mathrm{HI}}$: Data from experiments 1–3 in Table 1.
2. $D_{\mathrm{MID}}$: Data from experiments 4–6 in Table 1.
3. $D_{\mathrm{HI,MID}}$: Data from $D_{\mathrm{HI}} \cup D_{\mathrm{MID}}$.

In Figure 3, the estimate obtained using data sets 1–6, $\hat{g}(D_{\mathrm{HI,MID}})$ (solid line), is compared with the true function. We can 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 $D_{\mathrm{HI,MID}}$ is insufficient to obtain an accurate estimate of the right edge of this structure or any but the coarsest information regarding the rightmost block. This situation is best explained by appealing to the physics of the problem where the propagation and associated dissipation of energy through a lossy medium implies that the ability to resolve the structure of $g$ should decay radially from the borehole [5, 27]. To understanding how $D_{\mathrm{HI}}$ and $D_{\mathrm{MID}}$ individually contribute information to this estimate, $\hat{g}(D_{\mathrm{MID}})$ and $\hat{g}(D_{\mathrm{HI}})$ are also shown in Figure 3 as the dashed and dot-dashed lines respectively. Again, we see that the data from the high and middle frequency sources provide information about $g$ close to $x = 0$. Further from the origin, $\hat{g}(D_{\mathrm{HI,MID}})$ follows neither $\hat{g}(D_{\mathrm{HI}})$ nor $\hat{g}(D_{\mathrm{MID}})$ so that some level of data fusion must be taking place in that the presence of both data sets 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 in Figure 4 via the analysis of the diagonal elements of the relative error covariance matrices, $\Pi(B)$

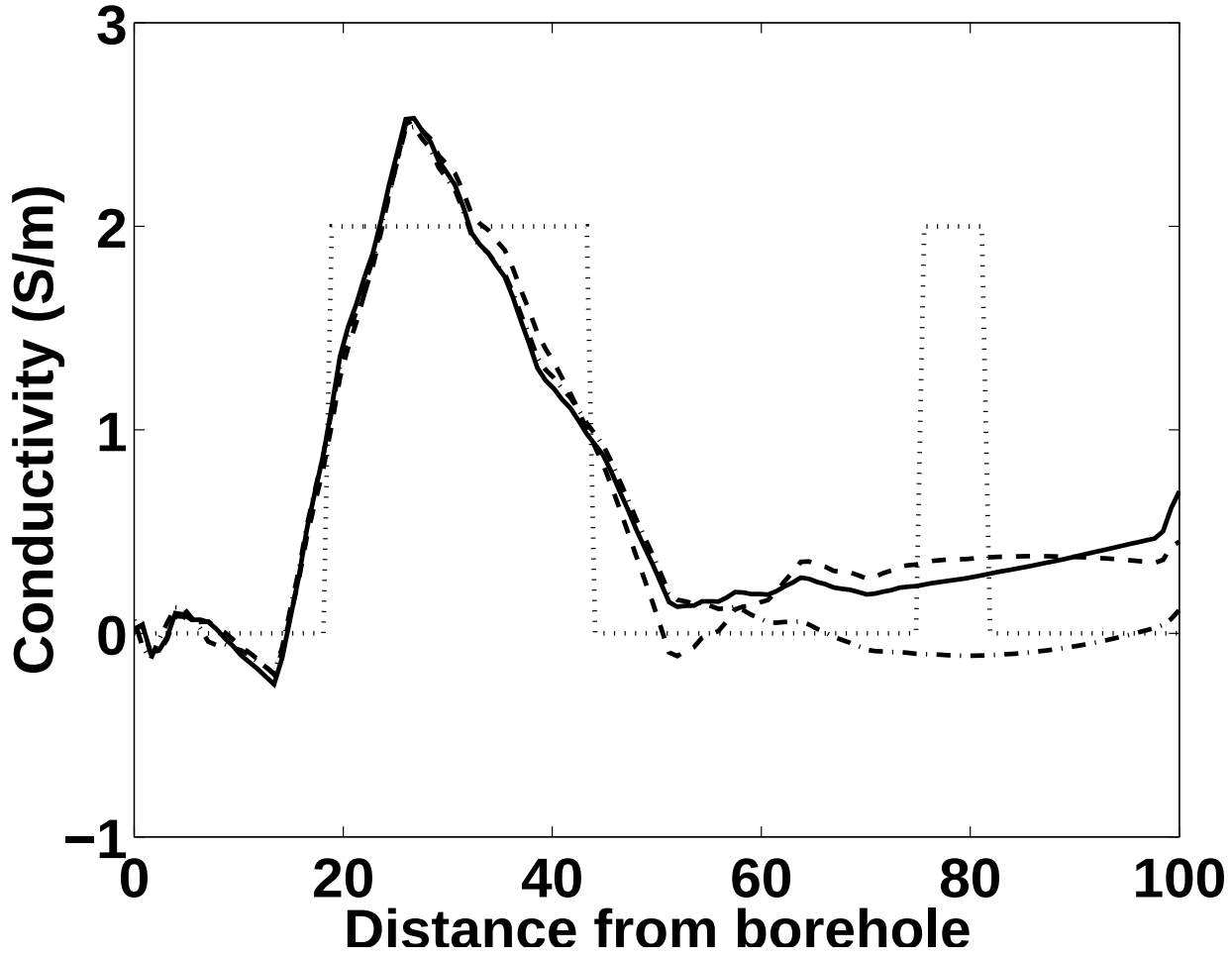
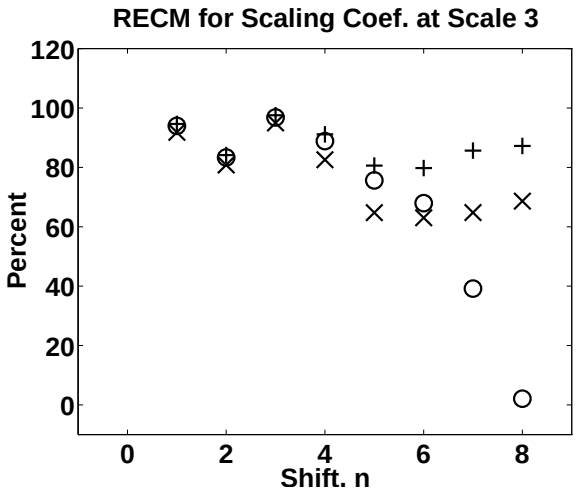


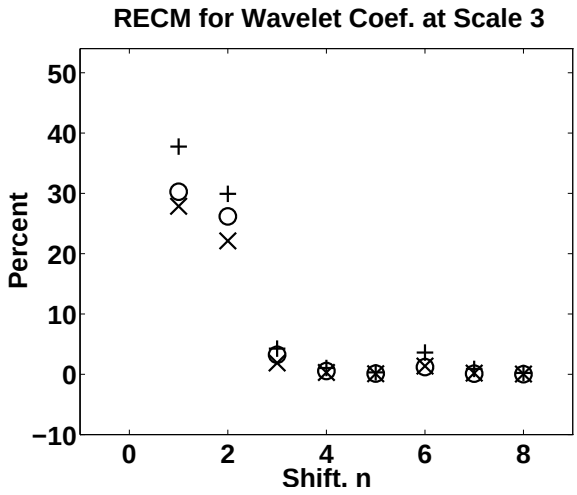

Figure 3: Estimates of $g$ using various combinations of high and middle frequency data. Dotted line = $g$, solid line = $\hat{g}(D_{\mathrm{HI,MID}})$, dashed line = $\hat{g}(D_{\mathrm{MID}})$, dot-dashed line = $\hat{g}(D_{\mathrm{HI}})$. In all cases, the measurements provide sufficient information 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 $\hat{g}(D_{\mathrm{HI,MID}})$ is significantly different from both $\hat{g}(D_{\mathrm{HI}})$ and $\hat{g}(D_{\mathrm{MID}})$ we conclude that some type of sensor fusion is occurring over the region far from $x = 0$.

for $B \in \{D_{\mathrm{HI}}, D_{\mathrm{MID}}, D_{\mathrm{HI,MID}}\}$. As there is strictly more information in $D_{\mathrm{HI,MID}}$, than in either $D_{\mathrm{HI}}$ or $D_{\mathrm{MID}}$ alone, the elements from $\Pi(D_{\mathrm{HI,MID}})$ must lie above the other two. In those cases where $\Pi_n^m(D_{\mathrm{HI,MID}})$ is significantly larger than both $\Pi_n^m(D_{\mathrm{HI}})$ and $\Pi_n^m(D_{\mathrm{MID}})$, we say that active sensor fusion is taking place. In Figure 4(a), this is the case for the estimates of elements $5-8$ of $g(L_g)$. Examination of Figures 4(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, the fact that $\Pi_n^m$ is close to zero at all scales and for all wavelet coefficients corresponding to shifts far from $x = 0$ indicates that the information in $D_{\mathrm{HI}}$ and $D_{\mathrm{MID}}$ either alone or in combination is insufficient to reconstruct *any* detail in $g$ over this domain.

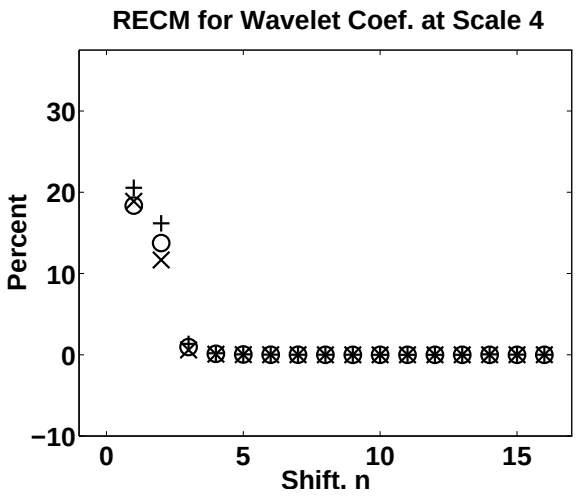
In Figure 5(a), $m^*(j)$, the optimal scale as defined in Section 3.4 is plotted for $\tau = 0.05$ (solid line) and $\tau = 0.5$ (dashed line) using data from $D_{\mathrm{HI,MID}}$. At $\tau = 0.05$, we see that as $x$ grows large, the optimal scale drops from 6 to 3 in a manner quite consistent with the intuition developed by examination of the estimates. The $\tau = 0.50$, case shows similar characteristics; however, the more



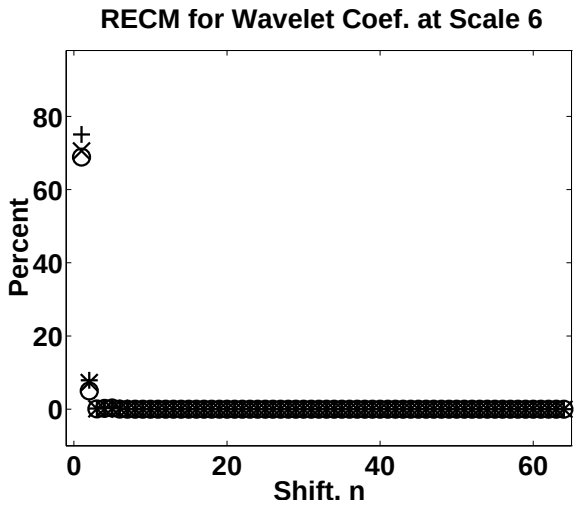

(a) RECM information for coarsest scaling coefficients (i.e. scale $\overline{3}$)




(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 4: Diagonal elements of relative error covariances for three radial profiling experiments. In all cases, the symbol "+" corresponds to $\Pi(D_{\mathrm{HI,MID}})$, "$\circ$" to $\Pi(D_{\mathrm{HI}})$ and "$\times$" to $\Pi(D_{\mathrm{MID}})$. 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 $x = 0$.

stringent threshold results in a more rapid decrease in scale as a function of distance. Finally, in Figure 5(b) the truncated estimates, $\hat{g}_\tau(D_{\mathrm{HI,MID}})$, defined by (18), are compared against $\hat{g}(D_{\mathrm{HI,MID}})$ for $\tau = 0.05$ and $\tau = 0.50$ respectively showing little difference among the three.

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 6, the diagonal elements of $\Pi(D_{\mathrm{HI}}, D_{\mathrm{HI,MID}})$ (i.e. the relative variance reduction information associated with the addition of middle frequency data to an estimate based upon high frequency information) are displayed

| 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 |
| $\mu_z$ | 1 | $\mu_x$ | 1 |
| $\sigma_z^2$ | 1 | $\sigma_x^2$ | 1 |
| $p_{L_g,z}$ | $\sqrt{2}$ | $p_{L_g,x}$ | $\sqrt{2}$ |
| $SNR^2$ for $D_{\mathrm{HI}}$ | 250 | $SNR^2$ for $D_{\mathrm{MID}}$ | 500 |
| $SNR^2$ for $D_{\mathrm{LO}}$ | 1000 | Background conductivity | 1 S/m |

Table 3: Parameters for cross-well tomography problem

for the coarsest scaling coefficients and the finest wavelet coefficients[5]. These plots illustrate that the middle frequency data sets contribute new information to a high frequency estimate at the coarsest scale away from the origin and at the finest scale closest to the origin which is in accord with the plots in Figure 3.

From this example, we see that the relative error covariance matrix provides new and useful insight into multisensor data fusion. Specifically, one would conclude that the data from the high and middle frequency data sets are 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 $x = 0$. The RECM analysis also suggests that the original parameterization of $g$ involving 128 degrees of freedom is excessive. Rather, at a threshold of $\tau = 0.50$, the data dictate that only 9 elements of $\gamma$ (the nonzero elements of $\hat{\gamma}_{0.50}(D_{\mathrm{HI,MID}})$) can be accurately recovered representing a 93% reduction in complexity of the inverse problem. Although not considered in this paper, the reduction in complexity might be realized in an inversion algorithm where we compute *only* those nine coefficients rather than following the procedure in (18) where we first computed all 128 elements of $\hat{\gamma}$ and then set all but nine to zero. In particular, we observe that just such an approach is considered in [27,29] in the context of a nonlinear inverse scattering problem.

## 4.2 A Two-Dimensional, Cross-Well Tomography Problem

In this example we consider improving resolution near the right side of the conductivity anomaly by augmenting $D_{\mathrm{HI,MID}}$ with data sets 9–12 from Table 1 which 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 $D_{\mathrm{LO}}$ and note that examination of the structure of the integral kernels for this problem leads to the observation that the low frequency observations are most sensitive to variations in $g$ near the left and right vertical edges with little resolution in the center of region $C$ [27]. Additionally, for this problem, $g$ varies both in the $x$ and the $z$ directions with the true conductivity anomaly to be reconstructed in this example displayed in Figure 7(a) and the parameter values needed for this experiment given in Table 3.

In Figure 7, we see that the addition of the low-frequency, cross-well data do significantly improve the resolution on the right side of $C$. Figure 7(b) (resp. (c)) is a display of $\hat{g}(D_{\mathrm{HI,MID}})$ (resp. $\hat{g}(D_{\mathrm{HI,MID,LO}})$). 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. While both conductivity perturbations are reflected in

[5] 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 $\gamma$ 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 $A$.

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.

Because this is a fully two dimensional example, we have the ability to use the RECM for the analysis of sensor fusion issues in both horizontal and vertical directions. In Figure 8, 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 $D_{\text{HI,MID}}$ and $D_{\text{HI,MID,LO}}$ respectively are available for the reconstruction. From Figure 8(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 7(a). Figure 8(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 8(a) to 3 in Figure 8(c). Moreover, the finest scale of horizontal detail moves from 1 to 2 in the area near the right vertical edge.

In Figure 7(d), $\hat{g}_{0.5}(D_{\text{HI,MID,LO}})$, the truncated estimate of $g$ defined in (18), is plotted. Here $\hat{g}_{0.5}(D_{\text{HI,MID,LO}})$, 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.

To demonstrate the flexibility of our multiscale prior modeling structure we observe that given the sets of data considered in this experiment, more detail can be obtained on the right side of the conductivity anomaly by changing a limited portion of the matrix $P_0$. Indeed, because each wavelet coefficient in $\gamma$ reflects our prior knowledge of $g$ over a limited area, we have extensive flexibility for choosing the spatial scale and physical location over which we modify the impact of the data on the reconstruction. To explore the possibility of improving the resolution near the rectangular structure located close to the right edge of $C$, we increase the variances in $P_0$ associated only with the finest scale wavelet coefficients that impact $g$ near the location of this perturbation. The result is shown in Figure 9 where we clearly have an improved picture as to the true nature of $g$ near the left side of the region of interest. Note that the prior model in this case is *not* strictly of the $1/f$ variety as we have manually altered the variances of a small number of the coefficients in $\gamma$. A natural extension of this exercise would be the development of automated methods for jointly determining the "appropriate" variance structure of the prior model and the estimate of the conductivity.

## 5 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 and statistical estimation. We pose the problem directly in the wavelet-transform domain and use a linear least squares estimator for the inversion algorithm and as the basis for the associated analysis methods. A prior statistical model of $\gamma$, the wavelet transform of the conductivity field, serves to regularize the problem. For much of the paper, we used a $1/f$-like fractal model that is often posited as a meaningful description of natural phenomena; however as discussed in [7,13,16,26] and demonstrated in Section 4.2, the utility of multiscale prior models extend beyond this particular class.

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 describing multisensor data fusion and identifying those degrees of freedom in $\gamma$ for the data contribute useful information. In this paper, we have made use of this information by first estimating all of $\gamma$ and then setting to zero those coefficients for which the RECM dictates there should be little information. In [27, 29], we consider an alternate approach in which the RECM-based information is used directly in the inversion routine to lower the computational complexity of the overall estimation procedure.

Although not considered extensively in this work, the inversion algorithms admit highly efficient implementations. As discussed in [1,4], 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 $\Theta_i$. The sparsity of $\Theta_i$ combined with the diagonal structure of $P_0$ (which is obtained using *any* uncorrelated multiscale model) imply that highly efficient, iterative algorithms such as LSQR [30] can be used to solve the normal equations. In [27], 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] 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] 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.

[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] Weng Cho Chew. *Waves and Fields in Inhomogeneous Media.* Van Nostrand Reinhold, New York, 1990.

[6] 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.

[7] 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.

[8] Jane Cullum. The effective choice of the smoothing norm in regularization. *Mathematics Of Computation*, 33(145):149–170, January 1979.

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

[10] Ingrid Daubechies. *Ten Lectures on Wavelets.* SIAM, 1992.

[11] R. A. DeVore, B. Jawerth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Trans. Info. Theory*, 38(2):719–746, March 1992.

[12] David L. Donoho. Nonlinear solution of linear inverse problems by Wavelet-Vaguelette decomposition. Technical Report Technical Report No. 403, Dept. of Statistics, Stanford University, 1992.

[13] M. Basseville *et. al.* Modeling and estimation of multiresolution stochastic processes. *IEEE Trans. Info. Theory*, 38(2):766–784, March 1992.

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

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

[16] Stuart A. Golden. Identifying multiscale statistical models using the wavelet transform. Master's thesis, Massachusetts Institute of Technology, April 1991.

[17] Steven J. Gortler, Perter Schroder, Micchael F. Cohen, and Pat Hanrahan. Wavelet radiosity. In *Computer Graphics, Annual Conference Series 1003.* Siggraph, August 1993.

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

[19] 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.

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

[21] 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.

[22] 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.

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

[24] Allen Q. Howard Jr., Weng Cho Chew, and Michael C. Moldoveanu. A new correction to the Born approximation. *IEEE Trans. on Geoscience and Remote Sensing*, 28(3):394–399, May 1990.

[25] Ranier Kress. *Linear Integral Equations.* Springer-Verlag, Berlin, 1989.

[26] 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.

[27] 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.

[28] 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.

[29] Eric L. Miller and Alan S. Willsky. Wavelet-based methods for the nonlinear inverse scattering problem using the Extended Born Approximation. submitted to *Radio Science*, March 1995.

[30] 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.

[31] John A. Scales and Adam Gersztenkorn. Robust methods in inverse theory. *Inverse Problems*, 4:1071–1091, 1988.

[32] 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.

[33] 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.

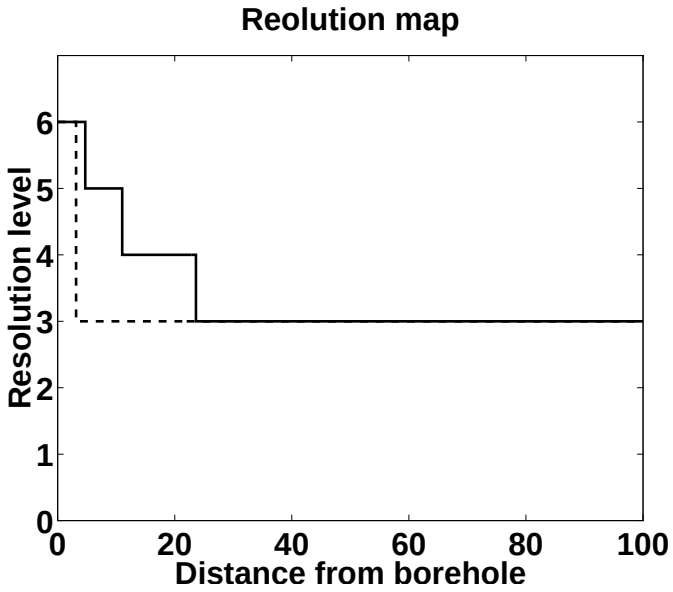
[34] Donald L. Turcotte. *Fractals and chaoes in geology and geophysics.* Cambridge University Press, 1992.

[35] Harry L. Van Trees. *Detection, Estimation and Modulation Theory: Part I.* John Wiley and Sons, New York, 1968.

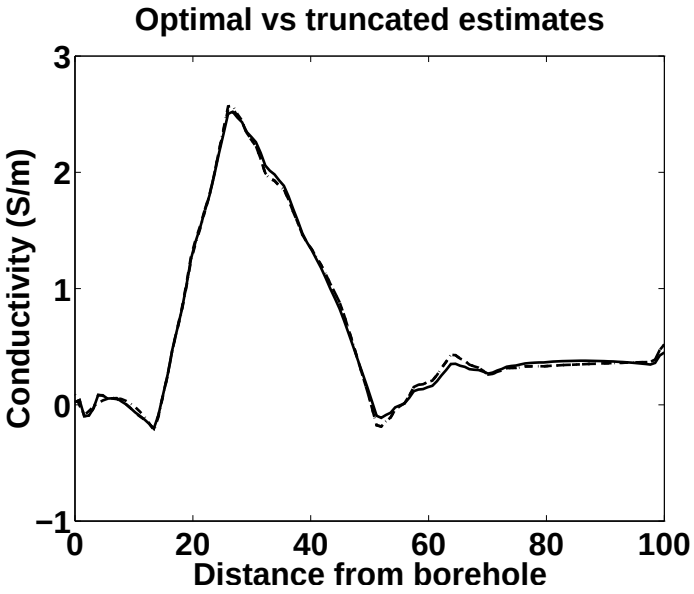
[36] Y. M. Wang and W. C. Chew. An iterative solution to the two-dimensional electromagnetic inverse scattering problem. *International Journal of Imaging Systems and Technology*, 1(1):100–108, 1989.

[37] Mladen Victor Wickerhauser. *Adapted wavelet analysis from theory to software.* A.K. Peters, Wellesley, MA, 1994.

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

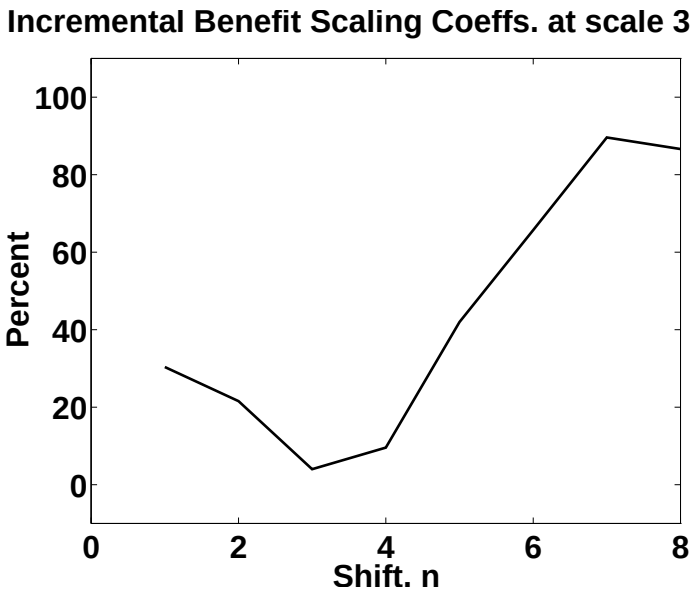



(a) The optimal scale of reconstruction as a function of position at scale $M_g = 7$ for a threshold value of $\tau = 0.05$ (solid line) and $\tau = 0.50$ (dashed line).

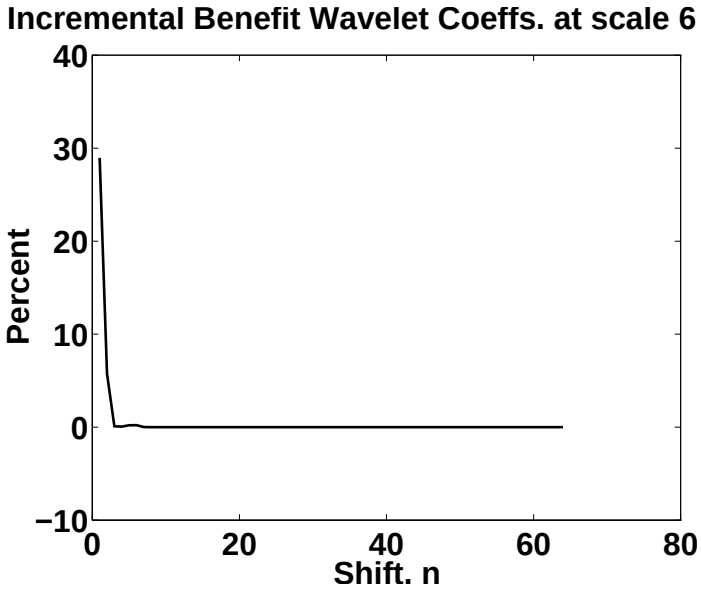



(b) $\hat{g}$ (solid line), $\hat{g}_{0.05}$ (dashed line), $\hat{g}_{0.50}$ (dot-dashed line)

Figure 5: 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. 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.

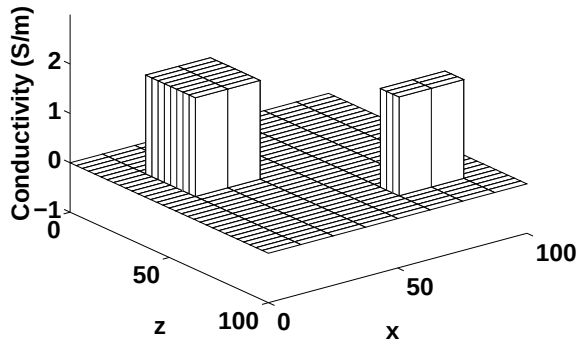



(a) $\Pi_n^{\overline{3}}(D_{\mathrm{HI}}, D_{\mathrm{HI,MID}})$

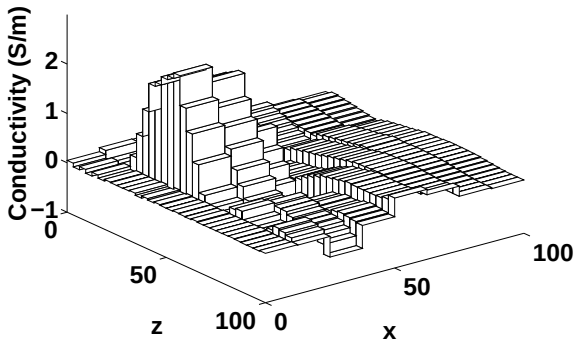



(b) $\Pi_n^{6}(D_{\mathrm{HI}}, D_{\mathrm{HI,MID}})$

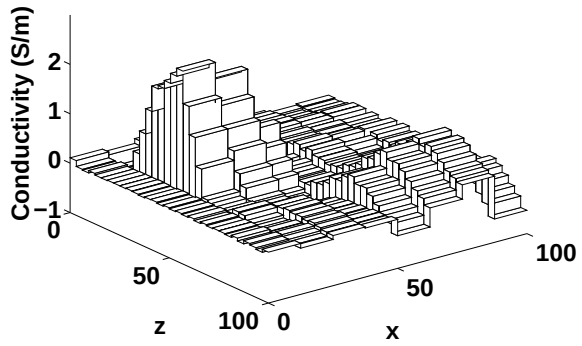
Figure 6: The incremental reduction in uncertainty obtained by adding data from the middle frequency observation to an estimate based upon the high frequency measurement sources. In accordance with Figure 4(a) we see significant benefits associated with determination of both the coarsest scale structure of $g$ far from the origin as well as the finest scale structure closest to $x = 0$.



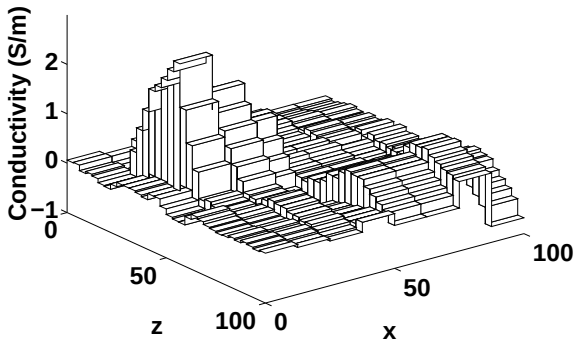

(a) Conductivity structure to be reconstructed




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

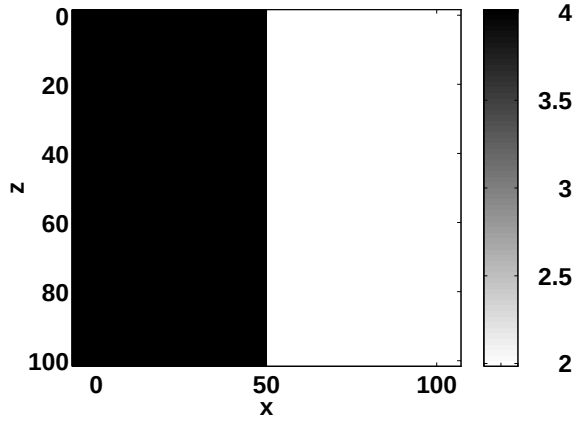



(c) $\hat{g}(D_{\mathrm{HI,MID,LO}})$

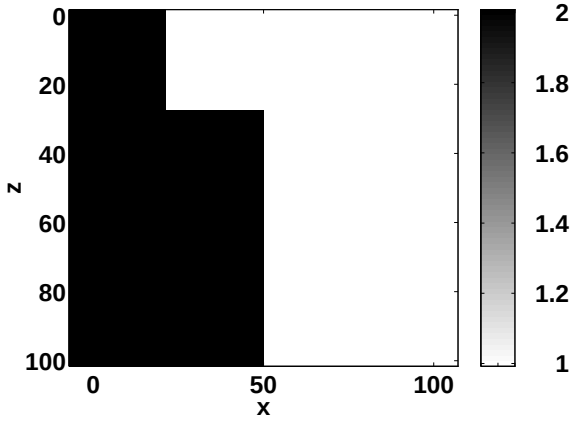



(d) $\hat{g}_{0.5}(D_{\mathrm{HI,MID,LO}})$

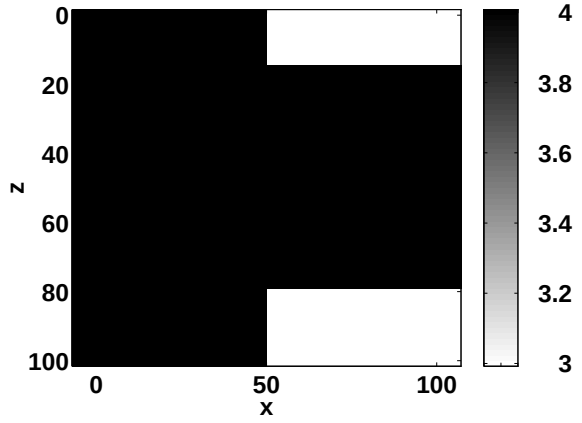
Figure 7: In (b)–(d) the estimates of $g$ in (a) are displayed using various combinations of high, middle and low frequency data. From (b), the high and medium frequency information provides insufficient information to reconstruct the anomaly near $x = 100$. As seen in (c), the addition of the low frequency, cross-well data sets clearly improves the ability to resolve this second structure. Note that there is little difference between truncated estimate $\hat{g}_{0.50}(D_{\mathrm{HI,MID,LO}})$ plotted in (d) which is composed of 75 non-zero elements in the wavelet transform domain and the optimal estimate, $\hat{g}(D_{\mathrm{HI,MID,LO}})$, in (c) which has 256 degrees of freedom.



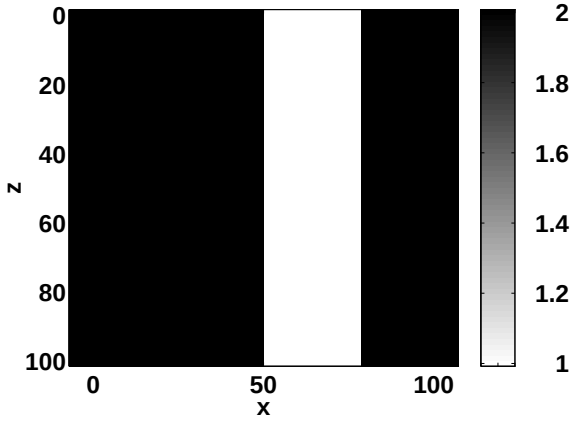

(a) The finest scale of to which $z$-oriented detail can be reconstructed at $\tau = 0.5$ given only high and middle frequency data.




(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 8: 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.

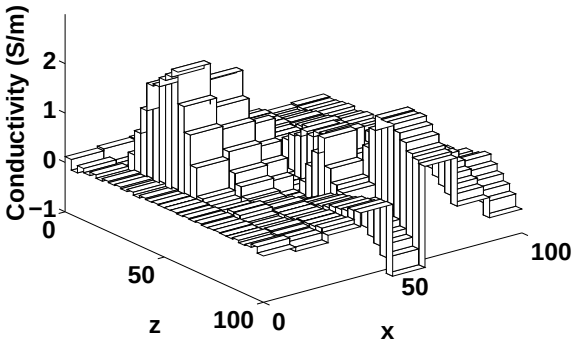



Figure 9: Estimates of $g$ using high, middle and low frequency data. Here, the variances associated with the fine scale wavelet coefficients governing the behavior of $g$ near the anomaly on the right side have been increased so as to allow more information from the data to be reflected in the estimate.