Massachusetts Institute of Technology
Woods Hole Oceanographic Institution

Joint Program in Oceanography/ Applied Ocean Science and Engineering

1930

DOCTORAL DISSERTATION

Geoacoustic Inversion by Mode Amplitude Perturbation

by

Travis L. Poole

February 2007

MIT/WHOI
2007-03

Geoacoustic Inversion by Mode Amplitude Perturbation

by

Travis L. Poole

Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

and

Woods Hole Oceanographic Institution
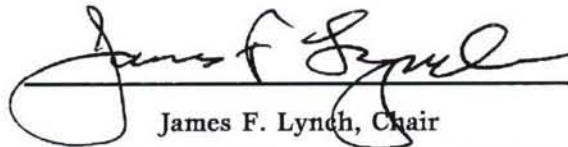Woods Hole, Massachusetts 02543

February 2007

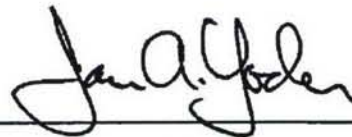DOCTORAL DISSERTATION

Approved for Distribution:

James F. Lynch, Chair
Department of Applied Ocean Physics and Engineering

Paola Malanotte-Rizzoli
MIT Director of Joint Program

James A. Yoder
WHOI Dean of Graduate Studies

Geoacoustic Inversion by Mode Amplitude Perturbation

by

Travis L. Poole

B.A. Physics, Mathematics

Luther College, 2000.

SUBMITTED TO THE DEPARTMENT OF MECHANICAL ENGINEERING IN
PARTIAL FULLFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN APPLIED OCEAN SCIENCES
at the
MASSACHUSETTES INSTITUTE OF TECHNOLOGY
and the
WOODS HOLE OCEANOGRAPHIC INSTITUTION
FEBRUARY 2007
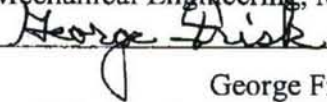
Signature of author_____

Department of Applied Ocean Physics and Engineering, WHOI
Department of Mechanical Engineering, MIT

Certified by_____

George Frisk
Scientist Emeritus, WHOI
Thesis supervisor

And_____

James Lynch
Senior Scientist, WHOI
Thesis supervisor

Accepted by_____

Henrik Schmidt
Chair, Joint Committee in Applied Ocean Science and Engineering, MIT/WHOI

And_____

Lallit Anand
Chairman, Department Committee on Graduate Students

1

Geoacoustic Inversion by Mode Amplitude Perturbation

by

Travis L. Poole

Submitted to the departments of Mechanical Engineering, MIT, and Applied
Ocean Physics and Engineering, WHOI, on December 12th, 2006 in partial
fulfillment of the requirements for the degree of Doctor of Philosophy in
Applied Ocean Sciences.
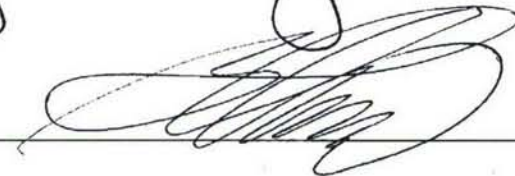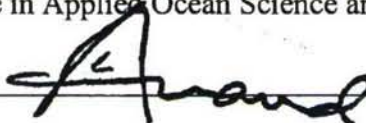
Abstract

This thesis introduces an algorithm for inverting for the geoacoustic
properties of the seafloor in shallow water. The input data required by the
algorithm are estimates of the amplitudes of the normal modes excited by a
low-frequency pure-tone sound source, and estimates of the water column
sound speed profiles at the source and receiver positions. The algorithm
makes use of perturbation results, and computes the small correction to an
estimated background profile that is necessary to reproduce the measured
mode amplitudes. Range-dependent waveguide properties can be inverted
for so long as they vary slowly enough in range that the adiabatic
approximation is valid. The thesis also presents an estimator which can be
used to obtain the input data for the inversion algorithm from pressure
measurements made on a vertical line array (VLA). The estimator is an
Extended Kalman Filter (EKF), which treats the mode amplitudes and
eigenvalues as state variables. Numerous synthetic and real-data examples
of both the inversion algorithm and the EKF estimator are provided. The
inversion algorithm is similar to eigenvalue perturbation methods, and the
thesis also presents a combination mode amplitude/eigenvalue inversion
algorithm, which combines the advantages of the two techniques.

Thesis Supervisor: George Frisk
Title: Scientist Emeritus, WHOI

Thesis Supervisor: James Lynch
Title: Senior Scientist, WHOI

Acknowledgements

This thesis would never have been completed if not for the input and support of many people. My advisors George Frisk and Jim Lynch deserve a great deal of thanks for their guidance and advice during both the research and writing of the thesis. Allan Pierce's ASA talk on mode amplitudes provided some of the early motivation for the thesis topic, and his comments on the derivations in the thesis have been very useful. Nick Makris also deserves thanks for serving on the committee and for the input he has provided.

I also owe thanks to a number of others who have not only been instrumental in expanding my knowledge of ocean acoustic, but have also made my time at WHOI a pleasure: Cindy Sellers, Jim Doutt, Art Newhall, Ying Tsong Lin, Luiz Souza, and Kyle Becker.

The moral support I received from my family and friends during my years in the program has also been invaluable. The support and encouragement from my parents, in particular, has been crucial during the process of earning my Ph.D. I also need to thank Emma Teuten for her support, patience, encouragement, and affection, all of which were critical for maintaining morale.

The funding that made this research possible came from the Office of Naval Research, and the WHOI Academic Programs Office. In addition to funding, the APO provided innumerable answers to questions I should have found myself, reminders of important tasks I had forgotten, and smiles and encouragements whenever I dropped in.

Contents

Chapter 1: Introduction, Methods, and Overview

Due to the rapid attenuation of electro-magnetic waves in the ocean, acoustic waves are the tool of choice (and necessity) for many underwater tasks. From the navigation of underwater robots, to the detection of submarines and buried mines, from the wireless transmission of data, to large-scale measurement of ocean temperatures, acoustics is the best, and often only, way to get the job done. Because the propagation of sound depends heavily on the environment through which it is propagating, the success or failure of an underwater task often depends critically on our ability to estimate the acoustic properties of that environment.

In cases where the acoustic energy is concentrated in the water column, such as in most deep ocean situations, we need only know the acoustic properties of the water itself. In low frequency (~500 Hz) cases in shallow water, however, the sound interacts strongly with the seafloor as well, and our ability to predict and interpret the propagation also depends on our knowledge of the ocean bottom. While sampling the water column directly has become fairly routine for the ocean science community, directly sampling the bottom is far more difficult. In many cases, it is impractical to take direct measurements of the acoustic properties below a meter or two into the seafloor.

For this reason, this thesis examines a new method for estimating the acoustic properties of the ocean floor. The method is similar to the eigenvalue perturbation method introduced by Rajan, Lynch, and Frisk [1], but obtains modal amplitude data using an Extended Kalman Filter (EKF), rather than obtaining eigenvalues via the Hankel

transform. This gives it the advantage of estimating the bottom parameters locally, rather than integrated over a large range. Further, it tolerates a greater amount of temporal variability in the water column over the course of the measurement. Both the method of Rajan *et al.* and the method described in this thesis are perturbative, meaning they both compute the small change, or "perturbation," to a background model of the bottom that would be necessary to cause the background model to produce the measured input quantities.

These methods stand in contrast to more commonly-used matched field methods (*e.g.,* [2]) which generate a large number of trial models, compare their predicted fields to the measured field, and declare the model that produces the best match to be the estimate of the bottom. These methods are considerably more computationally intensive than perturbative methods due to the large number of predicted fields, or "forward models," that must be computed. Further, because they can be thought of as repeated guess-and-check methods, they are unlikely to provide as much physical insight into the problem as methods which take greater advantage of our knowledge of acoustic propagation.

Another category of geoacoustic inversion methods is the exact methods (*e.g.,* [3], [4], [5]). These methods solve for the bottom parameters directly from some measurement of the acoustic field, without the need for any background model or the solving of any forward problems. However, these methods tend to require input data that are very difficult to measure, and often do not consider the possibility of measurement noise.

Whatever the method used, the seafloor parameter of greatest interest to the ocean acoustician is the sound speed, usually treated as a function of depth, and considered to

vary slowly in the horizontal direction. Also of importance are the density and attenuation of the sediment. Shear properties can also be considered, but are often of negligible effect because the rigidity of the sediments is usually far less than that of a solid [6], [7]. This thesis will ignore shear properties, and treat the seafloor as a fluid. This is not a necessary assumption of the method, however, and it should be entirely possible to add shear parameters to the method if so required.

Before we can consider the inverse problem, we must first examine the forward problem of calculating the field due to an acoustic point source in a waveguide. We will follow Frisk [8] and Jensen *et al.* [6] in deriving the normal mode equations, though Pekeris [9] is usually given credit for introducing the ocean acoustics community to the normal modes formulation. We start with the equation for the acoustic field due to a point source in a horizontally stratified, fluid medium, located at cylindrical coordinates $\vec{r}_s = (0,0,z_s)$, with the z axis positive downwards

$$\nabla^2 p(\vec{r}) + \rho(z)\nabla\left[\frac{1}{\rho(z)}\right] \bullet \nabla p(\vec{r}) + k^2(z)p(\vec{r}) = -4\pi \frac{\delta(r)}{r}\delta(\theta)\delta(z-z_s).$$

Here p is the acoustic pressure, ρ is the density, $k = \dfrac{\omega}{c}$ where ω is the angular frequency of the pure-tone source, c is the sound speed, and δ is the Dirac delta function. The cylindrical coordinates are r, θ, and z, and $\vec{r}$ is the position vector containing all three coordinates. Imposing cylindrical symmetry, and integrating from $\theta = 0$ to $\theta = 2\pi$ gives

$$\frac{1}{r}\frac{\partial}{\partial r}\left[r\frac{\partial p(r,z)}{\partial r}\right] + \frac{\partial^2 p(r,z)}{\partial z^2} + \rho(z)\frac{\partial}{\partial z}\left[\frac{1}{\rho(z)}\right]\frac{\partial p(r,z)}{\partial z} + k^2(z)p(r,z) = -2\frac{\delta(r)}{r}\delta(z-z_s).$$

We will solve the unforced version of this equation using the separation of variables, proposing a solution in terms of radial and depth functions:

$$p(r,z) = Z(z)R(r).$$

Substituting this into the equation, and dividing through by *Z(z)R(r)* gives:

$$\frac{1}{R}\left[\frac{1}{r}\frac{d}{dr}\left(r\frac{dR}{dr}\right)\right] + \frac{1}{Z}\left[\frac{d^2Z}{dz^2} + \rho(z)\frac{d}{dz}\left(\frac{1}{\rho(z)}\right)\frac{dZ}{dz} + k^2(z)Z\right] = 0$$

The first term is a function only of *r*, while the second is a function only of *z*. The only way for this to be true is for each term to be equal to a constant. We call this constant of separation k_n^2, and obtain the modal equation:

$$\frac{1}{\rho(z)}\frac{d^2Z_n}{dz^2} + \frac{d}{dz}\left(\frac{1}{\rho(z)}\right)\frac{dZ_n}{dz} + \frac{1}{\rho(z)}\left[k^2(z) - k_n^2\right]Z_n = 0$$

Though we have called the constant of separation k_n^2 in anticipation of a discrete set of ordered solutions, this does not become the case until we apply boundary conditions. If we apply the boundary conditions that *Z(0)=0*, implying a pressure-release surface at the air-water interface, and $\left.\frac{dZ}{dz}\right|_{z=D} = 0$, implying a rigid boundary at some depth, *D*, or the boundary condition *Z(D)=0*, implying a pressure release surface at some depth, *D*, (which is a non-physical, but potentially useful approximation) we have Sturm-Liouville problem[1]. We can make use of some of the well-known properties of Sturm-Liouville

---

[1] Technically, in order to achieve a proper Sturm-Liouville problem, we must break $\frac{1}{\rho(z)}\left[k^2(z) - k_n^2\right]$ into $\left[\frac{k^2(z) - \omega^2/c_{min}^2}{\rho(z)} + \frac{\omega^2/c_{min}^2 - k_n^2}{\rho(z)}\right]$, and the eigenvalues of the problem will be $\lambda_n = \frac{\omega^2}{c_{min}^2} - k_n^2$. However, it is traditional in ocean acoustics to refer to k_n as the eigenvalue, and that tradition will be continued in this thesis.

8

problems here, specifically that there is a discrete set of eigenfunctions which are orthonormal with respect to the weighting function $\dfrac{1}{\rho(z)}$, such that

$$\int_0^D \frac{Z_n(z)Z_m(z)}{\rho(z)} dz = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases}.$$

Further, the eigenfunctions constitute a complete basis set in that any function $f(z)$ can be expanded in terms of a weighted sum over them:

$$f(z) = \sum_m a_m Z_m(z).$$

If we think of the pressure field $p(r,z)$ at a given r, we have a function of z and can expand it in terms of the mode functions. The weights of the expansion at each range will be different, and we can define the function $R_m(r)$ to be the weight of mth mode function at each value of r. This way we can expand the pressure field as $p(r,z) = \sum_m R_m(r)Z_m(z)$. Substituting this into the original, forced equation we get

$$\sum_m \left\{ \frac{1}{r}\frac{d}{dr}\left(r\frac{dR_m(r)}{dr} \right) Z_m(z) + R_m(r)\left[\frac{d^2 Z_m(z)}{dz^2} + \rho(z)\frac{d}{dz}\left(\frac{1}{\rho(z)} \right)\frac{dZ_m(z)}{dz} + k^2(z)Z_m(z) \right] \right\}$$
$$= -2\frac{\delta(r)}{r}\delta(z - z_s)$$

We can use the modal equation to substitute for the term in square brackets:

$$\sum_m \left\{ \frac{1}{r}\frac{d}{dr}\left(r\frac{dR_m(r)}{dr} \right) Z_m(z) + k_m^2 R_m(r)Z_m(z) \right\} = -2\frac{\delta(r)}{r}\delta(z - z_s).$$

We can make use of the orthogonality property of the mode functions by applying the operator $\int_0^D (\cdot)\dfrac{Z_n(z)}{\rho(z)} dz$ to both sides of the equation, giving us

9

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dR_n(r)}{dr}\right) + k_n^2 R_n(r) = -2\frac{\delta(r)}{r}\frac{Z_n(z_s)}{\rho(z_s)} \, .$$

This is Bessel's equation with a line source driving term, the solution of which is

$$R_n(r) = i\pi H_0^{(1)}(k_n r)\frac{Z_n(z_s)}{\rho(z_s)} \, .$$

The Hankel function of the 2nd kind, $H_0^{(2)}(k_n r)$ would also solve the equation, but we keep just the 1st kind so as to have outwardly radiating energy. Putting everything together, we obtain the normal mode expression for the pressure field:

$$p(r,z) = \frac{i\pi}{\rho(z_s)}\sum_n Z_n(z)Z_n(z_s)H_0^{(1)}(k_n r) \, .$$

In the far field, where $k_n r \gg 1$ we can make use of the asymptotic expansion for the Hankel function, and express the pressure field as

$$p(r,z) \sim \frac{\sqrt{2\pi}e^{i\pi/4}}{\rho(z_s)}\sum_n Z_n(z)Z_n(z_s)\frac{e^{ik_n r}}{\sqrt{k_n r}} \, .$$

In reality, of course, there is not a perfectly rigid or pressure release boundary at the bottom of the waveguide (one might argue that there isn't a true pressure release surface at the top either, though the air-water interface actually approximates one extremely well [6],[8].) We can account for this by letting $D \rightarrow \bullet$, which results in a finite set of propagating modes and an infinite set of non-propagating modes which decay exponentially with range. While these non-propagating modes can be important at short ranges, it is typical to assume a receiver far enough from the source that they can be neglected. Just how far is far enough can be estimated by calculating the eigenvalue of the first non-propagating mode, since it will decay the slowest with range. Past ranges where this mode is deemed to have decayed sufficiently, all higher modes will be even

more decayed. This thesis will neglect these non-propagating modes when calculating pressure fields. However, it should be pointed out that in order to consider the normal modes a complete basis set, the non-propagating modes must be retained. When making use of the completeness of the normal mode set, it is often useful to employ a false bottom in the waveguide. A rigid or pressure-release boundary is added at some depth well below the point where all propagating modes are evanescent in depth, and thus quite small [10]. The advantage of this is that the non-propagating modes will have purely imaginary k_n values, making them and their mode functions much easier to calculate. The danger of doing this is that some of the modes that would have been non-propagating may become propagating when the false bottom is added. If care is taken to neglect these modes when calculating the field, very little error is added as a result of the false bottom. Some of the practical issues of this are discussed in the appendix.

A further concern is that in reality, the ocean is not horizontally stratified. However, if the waveguide properties change sufficiently slowly with horizontal position, which is often the case, we can make use of Pierce's adiabatic approximation [11]. The key assumption of this approximation is that energy is not transferred between modes. The result of the assumption is that we can solve for the eigenvalues and mode functions using the local waveguide properties, and compute the phase by integrating the range-varying eigenvalue:

$$p(r,z,z_s) = \frac{e^{i\pi/4}}{\rho(0,z_s)}\sqrt{2\pi}\sum_n Z_n(0,z_s)Z_n(r,z)\frac{e^{i\int_0^r k_n(r')dr'}}{\sqrt{\int_0^r k_n(r')dr'}}.$$

In order for this approximation to be valid, it must be true that little-or-no energy is transferred between the modes as they propagate in range. Using the coupled mode formulation of Evans [12], and the single-scatter approximation of Porter *et al.* [13] it can be shown that the field in a short, range-independent sub-section running from $r = r_{i-1}$ to $r = r_i$ is given by

$$p^{(i)}(r,z) = r^{-1/2} \sum_n A_n^{(i)} Z_n^{(i)}(z) e^{ik_n^{(i)}(r-r_{i-1})}$$

where the superscript *(i)* indicates the *i*th section. The values of $A_n^{(1)}$ for the first section come from the range-independent equation, and are given by

$$A_n^{(1)} = \frac{e^{i\pi/4}}{\rho(0,z_s)} \sqrt{2\pi} \frac{Z_n^{(1)}(z_s)}{\sqrt{k_n^{(1)}}}$$

and later values of $A_n^{(i)}$ are computed from earlier values via

$$A_n^{(i+1)} = \sum_m A_m^{(i)} e^{ik_m^{(i)}(r_i-r_{i-1})} \frac{1}{2} \left[\int_0^\infty \frac{Z_n^{(i+1)}(z)Z_m^{(i)}(z)}{\rho^{(i+1)}} dz + \frac{k_m^{(i)}}{k_n^{(i+1)}} \int_0^\infty \frac{Z_n^{(i+1)}(z)Z_m^{(i)}(z)}{\rho^{(i)}(z)} dz \right].$$

For the adiabatic condition to be valid, we need $A_n^{(i)} \approx A_n^{(j)}$ for all *n, i,* and *j*. When the bottom is unknown, it will not to be possible to compute these values, so estimations will have to be made of the validity of the adiabatic approximation. Also, after a range-dependent inversion has been obtained, we can check to see if the adiabatic approximation would hold if the inversion were correct.

Having covered the acoustic equations used in the rest of the thesis, we will now change topic a bit in order to cover another result of which we will make frequent use: the Moore-Penrose pseudo-inverse [14]. As presented in the thesis of Souza [3], the pseudo inverse allows us to solve the linear system

$$\vec{d} = [G]\vec{q}$$

where $\vec{d}$ is an $M \times 1$ vector of known values, $\vec{q}$ is an $N \times 1$ vector of unknown values that we wish to determine, and $[G]$ is an $M \times N$ matrix of known values. The pseudo-inverse makes use of singular value decomposition (SVD) to obtain the result

$$\vec{q}_{LS} = [\tilde{G}^{\#}]\vec{d} = [V_r][\Lambda_r]^{-1}[U_r]^T \vec{d} = \sum_{m=1}^{r} \lambda_m^{-1}(\vec{v}_m \vec{u}_m^T)\vec{d},$$

where $\vec{q}_{LS}$ is the least squares approximation to $\vec{q}$ (also minimum-norm when $M<N$), $[\tilde{G}^{\#}]$ is the pseudo inverse-matrix, $[V_r]$ is the $N \times r$ matrix whose columns are the right-singular vectors $\vec{v}_m$ of $[G]$, $[\Lambda_r]$ is the $r \times r$ matrix whose diagonal contains the singular values λ_m of $[G]$ in decreasing order, $[U_r]$ is the $M \times r$ matrix whose columns are the right singular vectors $\vec{u}_m$ of $[G]$, and r is the rank of $[G]$. The pseudo-inverse matrix has the advantage that it exists for non-square matrices, and those with less-than-full rank. In many cases, values of λ_m can become quite small, causing the solution to become unstable by being overly sensitive to errors in $\vec{d}$. In these cases, it is appropriate to neglect small singular values and truncate the sum given above. It should also be noted that when $[G]$ is full rank, the expression above reduces to $\vec{q}_{LS} = ([G][G]^T)^{-1}[G]^T \vec{d}$ when $M > N$, and $\vec{q}_{LS} = [G]^T ([G][G]^T)^{-1}\vec{d}$ when $M < N$, which are the standard least squares expressions. When $[G]$ is both full-rank and square, $[\tilde{G}^{\#}] = [G]^{-1}$.

With these tools in place, we are now ready to derive the inversion algorithm that is the focus of this thesis. The next chapter will carry us through this derivation, examining both the inversion algorithm itself, and the estimation problem that must be solved to obtain its necessary inputs. Chapter 3 of the thesis will be a presentation of the

method as applied to numerous synthetic and real-world experiments. In Chapter 4 we will analyze the reaction of the algorithm to input errors, and consider limitations of the method. Chapter 5 will compare and contrast the method with the more commonly-used method of eigenvalue perturbation, and introduce a method of combining the two methods for greater robustness to error. The thesis will conclude with Chapter 6, a summary of what has been presented, and suggestions for future research.

Chapter 2: Derivation of Methods

In this chapter we will derive the mode amplitude perturbation inversion, and the Extended Kalman Filter (EKF) method for obtaining the necessary input parameters for the inverse. The basic concept behind the mode amplitude perturbative inversion is that we compare the field (or parameters describing it) that we expect in a known "background" waveguide to the field (or parameters) observed in our experiment. We seek the small perturbation to our background model that will cause our predicted field to match the measured field. This is very similar to the eigenvalue perturbation method of Rajan *et al.* [1], except that in this case our input data will be the mode amplitudes rather than the modal eigenvalues. In chapter 5 we will compare the two methods, and attempt to combine them for greater error robustness.

We originally derived the mode amplitude perturbation result based on the work of Thode and Kim [15], but here we will follow the work of Tindle *et al.* [16], whose derivation is simpler. We start with the depth-separated normal mode equation

$$\frac{1}{\rho(z)}\frac{d^2 Z_n}{dz^2} + \frac{d}{dz}\left(\frac{1}{\rho(z)}\right)\frac{dZ_n}{dz} + \frac{1}{\rho(z)}\left[q(z) - k_n^2\right]Z_n = 0,$$

where $q(z) = k^2(z) = \dfrac{\omega^2}{c^2(z)}$. We propose a perturbation to the waveguide of $q(z) \to q(z) + \Delta q(z)$, which causes a perturbation in the other terms, $Z_n(z) \to Z_n(z) + \Delta Z_n(z)$, and $k_n \to k_n + \Delta k_n$.

15

Collecting the unperturbed terms we get the unperturbed equation. Collecting the terms of first order perturbations we get

$$\frac{1}{\rho(z)}\frac{d^2\Delta Z_n}{dz^2} + \frac{d}{dz}\left(\frac{1}{\rho(z)}\right)\frac{d\Delta Z_n}{dz} + \frac{1}{\rho(z)}\left[q(z) - k_n^2\right]\Delta Z_n + \frac{1}{\rho(z)}\left[\Delta q(z) - 2k_n\Delta k_n\right]Z_n = 0.$$

Because the unperturbed normal modes form a complete set, we can expand any function of z in terms of them. We propose an expansion of ΔZ_n of the form $\Delta Z_n(z) = \sum_j a_{nj}Z_j(z)$. Substituting this into the equation above we get

$$\sum_j a_{nj}\left[\frac{1}{\rho}\frac{d^2 Z_j}{dz^2} + \frac{d}{dz}\left(\frac{1}{\rho}\right)\frac{dZ_j}{dz}\right] + \frac{1}{\rho}\left(q(z) - k_n^2\right)\sum_j a_{nj}Z_j + \frac{1}{\rho}\left(\Delta q(z) - 2k_n\Delta k_n\right)Z_n = 0.$$

The term in square brackets can be replaced using the normal mode equation, giving

$$\sum_j a_{nj}\left[-\frac{1}{\rho}\left(q(z) - k_j^2\right)Z_j\right] + \frac{1}{\rho}\left[q(z) - k_n^2\right]\sum_j a_{nj}Z_j + \frac{1}{\rho}\left[\Delta q(z) - 2k_n\Delta k_n\right]Z_n = 0.$$

Some of the terms containing $q(z)$ cancel, leaving us with

$$\sum_j a_{nj}\left(k_j^2 - k_n^2\right)\frac{Z_j}{\rho} + \frac{1}{\rho}\left(\Delta q(z) - 2k_n\Delta k_n\right)Z_n = 0.$$

From this point we can make use of the orthonormality property of the normal modes. If we apply the operator $\int_0^D (\bullet)Z_n(z)dz$ to the equation, we are left with

$$\int_0^D \frac{\Delta q(z)Z_n^2(z)}{\rho(z)}dz - 2k_n\Delta k_n = 0 \Rightarrow \Delta k_n = \frac{1}{2k_n}\int_0^D \frac{\Delta q(z)Z_n^2(z)}{\rho(z)}dz,$$

which is the result of Rajan *et al* [1].

16

If we instead apply the operator $\int_0^D (\bullet) Z_m(z) dz$ (note the change of subscript from n to m),

we get

$$a_{nm}\left(k_m^2 - k_n^2\right) + \int_0^D \frac{\Delta q(z) Z_m(z) Z_n(z)}{\rho(z)} dz = 0 \Rightarrow a_{nm} = \frac{1}{k_n^2 - k_m^2} \int_0^D \frac{\Delta q(z) Z_n(z) Z_m(z)}{\rho(z)} dz,$$

which is valid for $m \neq n$. This result, along with the expansion used above allows us to

express the change in a mode function $\Delta Z_n(z)$ due to a change in the profile, $\Delta q(z)$.

Since the quantity that is actually changing in the bottom is the sound speed, $c(z)$, we can

use the perturbation result that $\Delta q(z) = -2\Delta c(z)\frac{\omega^2}{c^3(z)}$. It should be noted that Tindle et

$al.$ list the $m=n$ term as being $a_{nn} = -\frac{1}{2}\sum_{m \neq n} a_{nm}^2$ based on the idea that the mode function

must retain its normalization. However, this result is inconsistent with the earlier neglect

of terms of second or higher order. Pierce [17] has shown that the actual value for the

$n=m$ term must be zero, since the change of the mode function must be orthogonal to the

mode function itself:

$$\int_0^D \frac{Z_n^2(z)}{\rho(z)} dz = 1 \Rightarrow \frac{\partial}{\partial X}\int_0^D \frac{Z_n^2(z)}{\rho(z)} dz = \frac{\partial}{\partial X}1 \Rightarrow \int_0^D \frac{2Z_n(z)\frac{\partial Z_n(z)}{\partial X}}{\rho(z)} dz = 0$$

for any parameter X that would cause a change in the mode function. In order to properly

make use of the normalization as Tindle has done, one would have to retain higher order

terms in the original perturbation.

The sound speed, $c(z)$, is not the only bottom parameter that affects the mode

functions and eigenvalues. The density profile $\rho(z)$ is also a factor. Just as Tindle did

for sound speed, we can use perturbation theory to determine the effects of perturbations

17

to a density parameter. If we define the function $\beta(z) = \dfrac{1}{\rho(z)}$ and denote differentiation

with respect to z with primes, our modal equation becomes

$$\beta(z)Z_n'' + \beta'(z)Z_n' + \beta(z)\left[k^2(z) - k_n^2\right]Z_n = 0 .$$

We propose a perturbation $\beta(z) \to \beta(z) + \Delta\beta(z)$ which causes $Z_n(z) \to Z_n(z) + \Delta Z_n(z)$

and $k_n \to k_n + \Delta k_n$. Inserting these perturbations into the equation above, and keeping

only the terms of first order (zeroth order terms are the unperturbed equations, and for

small enough perturbations, higher order terms are negligible). The result is

$$\Delta\beta Z_n'' + \beta\Delta Z_n'' + \beta'\Delta Z_n' + \Delta\beta' Z_n' + \Delta\beta\left[k^2(z) - k_n^2\right]Z_n + \beta\left[k^2(z) - k_n^2\right]\Delta Z_n - \beta 2 k_n \Delta k_n Z_n = 0.$$

If we again make use of the expansion $\Delta Z_n = \sum_j a_{nj} Z_j$, we can write the equation as

$$\sum_j a_{nj}\left[\beta Z_j'' + \beta' Z_j' + \beta\left(k^2(z) - k_n^2\right)Z_j\right] + \Delta\beta Z_n'' + \Delta\beta' Z_n' + \Delta\beta\left(k^2(z) - k_n^2\right)Z_n - \beta 2 k_n \Delta k_n Z_n = 0.$$

We can use the fact that $-\beta k_n^2 Z_j = -\beta k_j^2 Z_j + \beta(k_j^2 - k_n^2)Z_j$ and use the modal equation

to simplify the term in square brackets:

$$\sum_j a_{nj}\beta\left(k_j^2 - k_n^2\right) + \Delta\beta\left[Z_n'' + \left(k^2(z) - k_n^2\right)Z_n\right] + \Delta\beta' Z_n' - \beta 2 k_n \Delta k_n Z_n = 0.$$

To further simplify things, we can again make use of the modal equation to replace the

term in the new square brackets:

$$\sum_j a_{nj}\beta\left(k_j^2 - k_n^2\right)Z_j + \Delta\beta\frac{\beta'}{\beta}Z_n' + \Delta\beta' Z_n' - \beta 2 k_n \Delta k_n Z_n = 0.$$

Now we are at a point at which we can make use of the orthonormality property of the

normal mode functions.

If we apply the operator $\int_0^D Z_n(z)(\bullet)dz$ to both sides of this equation we are left with

$$\int_0^D Z_n(z)Z_n'(z)\left(\Delta\beta\frac{\beta'}{\beta}+\Delta\beta'\right)dz - 2k_n\Delta k_n \Rightarrow \Delta k_n = \frac{1}{2k_n}\int_0^D Z_n(z)Z_n'(z)\left(\Delta\beta\frac{\beta'}{\beta}+\Delta\beta'\right)dz ,$$

which is our eigenvalue perturbation result for density. If we apply the operator

$\int_0^D Z_m(z)(\bullet)dz$ to both sides of the equation we get our mode function perturbation:

$$a_{nm}\left(k_m^2 - k_n^2\right) + \int_0^D Z_m(z)Z_n'(z)\left(\Delta\beta\frac{\beta'}{\beta}+\Delta\beta'\right)dz \Rightarrow a_{nm} = \frac{1}{k_n^2 - k_m^2}\int_0^D Z_m(z)Z_n'(z)\left(\Delta\beta\frac{\beta'}{\beta}+\Delta\beta'\right)dz.$$

Note that because both the equation for the eigenvalue perturbation, and the modal expansion of the mode function perturbation involve z derivatives, discontinuities must be handled with particular care.

Another bottom parameter of interest is the attenuation profile. Frisk [8] and other texts show that if we introduce an imaginary component to the sound speed profile by making $k(z)$ complex, the eigenvalue becomes complex as well. While the change in the mode function itself is usually negligible when attenuation is added, the complex portion of the eigenvalue creates the appearance that the entire mode function has been reduced. If we make the small change $k(z) \to k(z) + i\alpha(z)$, the modal eigenvalue will be changed as well: $k_n \to k_n + i\delta_n$, where the modal attenuation is given by

$$\delta_n = \frac{1}{k_n}\int_0^D \frac{\alpha(z)}{\rho(z)}k(z)Z_n^2(z)dz .$$

This leads to the apparent change in the mode function $Z_n \to Z_n e^{-\delta_n r}$, or $\Delta Z_n = -Z_n(1 - e^{-\delta_n r})$. For our purposes, it will often be best to look at the apparent change in the mode function over a small range step due to the attenuation within that

range step. In such a case, the apparent change in the mode function from the beginning of the step to the end will be $Z_n \rightarrow Z_n e^{-\delta_n \Delta r}$, where Δr is the length of the range step. If $\delta_n \Delta r \ll 1$ we can use the Taylor expansion for the exponential, and keep only the linear terms, giving $Z_n \rightarrow Z_n (1 - \delta_n \Delta r)$ or $\Delta Z_n \approx Z_n \delta_n \Delta r$.

Further bottom parameters, such as shear speeds and shear attenuations could also be sought, in which case perturbation results for those parameters would also be needed. However, shear parameters usually have a very small effect on acoustic propagation, especially when the source and receiver are not very close to the bottom. For the purposes of this thesis, the results for sound speed, density, and attenuation perturbations will be sufficient.

The next step in the derivation of the inversion algorithm is to discretize the representation of the bottom. We do this so that we can solve for a finite number of unknowns, rather than full functions of the continuous variable z. We start by making the assumption that the unknown functions $c(z)$, $\rho(z)$, and $\alpha(z)$ can be written in terms of a weighted expansion of some known depth functions with unknown coefficients. For example, $c(z)$ can be expanded as $c(z) = c_0(z) + \sum_i X_i c_i(z)$. Here $c_0(z)$ is a hypothesized, or background, model for the sound speed profile. The functions $c_i(z)$ are arbitrary functions that should capture the important properties of the sound speed profile. The unknown scalar coefficients X_i are what we seek, since once we have them, we can reconstruct the sound speed profile. The proper selection of the functions $c_i(z)$ is somewhat of an art, and there are an infinite number of possibilities. At one end of the spectrum, one can create a set of $c_i(z)$'s that are square pulses at every sample depth. In

this case, the coefficients represent the difference of the sound speed from the background at each depth. This is equivalent to the technique used by Rajan *et al.*, and can produce a detailed profile. However, since the number of propagating modes at the low frequencies which are being considered is usually far less than the number of depths at which the sound speed is sought, this results in an underdetermined problem, and further assumptions are needed to obtain a unique solution. Often one will make use of a smoothness constraint and solve a least-squares problem via the pseudo inverse, as will be discussed later in chapter 4.

Alternatively, one can attempt to describe the bottom using a small number of parameters, such as sound speed at the water-bottom interface (for which $c_i(z) = \begin{cases} 0 & z < h \\ 1 & z \geq h \end{cases}$, where h is the water depth), and linear gradient (for which $c_i(z) = \begin{cases} 0 & z < h \\ z - h & z \geq h \end{cases}$). In this case, the problem will be overdetermined, and more robust to measurement errors in the input data. However, such a model will not be able to correctly reproduce complicated bottom structure. The trade-off between the number and complexity of the $c_i(z)$ functions, and robustness to error and uniqueness of the solution will be addressed in chapter 4.

Once the unknown functions have been parameterized, we can write down the derivatives of the mode functions with respect to the unknown scalars a_i using the perturbation results we derived above. To do this, we simply make the substitution $\Delta c(z) = X_i c_i(z)$ for the sound speed equations, $\Delta \beta(z) = B_i \beta_i(z)$ for the density, and

21

$\alpha(z) = A_i \alpha_i(z)$ for the attenuation, and use the fact that $\dfrac{\Delta Z_n}{X_i} \approx \dfrac{\partial Z_n}{\partial X_i}$ for sufficiently

small values of X_i. The results of this are:

$$\frac{\partial Z_n(z)}{\partial X_i} = \sum_m a_{nm}^{X_i} Z_m(z) \text{ where } a_{nm}^{X_i} = \frac{2\omega^2}{k_m^2 - k_n^2} \int_0^D \frac{c_i(z) Z_n(z) Z_m(z)}{c_0^3(z)\rho(z)} dz$$

$$\frac{\partial k_n}{\partial X_i} = \frac{-1}{k_n} \int_0^D \frac{c_i(z) Z_n^2(z)}{c_0^3(z)\rho(z)} dz$$

$$\frac{\partial Z_n(z)}{\partial B_i} = \sum_m a_{nm}^{B_i} Z_m(z) \text{ where } a_{nm}^{B_i} = \frac{1}{k_n^2 - k_m^2} \int_0^D Z_m(z) Z_n'(z) \left(\beta_i \frac{\beta_0'}{\beta_0} + \beta_i' \right) dz$$

$$\frac{\partial k_n}{\partial B_i} = \frac{1}{2k_n} \int_0^D Z_n(z) Z_n'(z) \left(\beta_i \frac{\beta_0'}{\beta_0} + \beta_i' \right) dz$$

$$\text{and } \frac{\partial Z_n(z)}{\partial A_i} \approx \frac{Z_n(z)\Delta r}{k_n} \int_0^D \frac{\alpha_i(z)}{\rho(z)} k(z) Z_n^2(z) dz \,.$$

With these derivatives, we are nearly (but not quite!) ready to introduce the inversion. Given a measurement of the real-world mode functions, and an estimate of our background mode functions, we could use the difference and these derivatives to compute the necessary changes to the parameters in the background model needed to make the computed mode functions match those that were measured. However, in the real world, where source levels are often not known precisely, and where instrument calibration is at times questionable, using the mode functions themselves can be problematic. Further, since the expression for the field always contains the product of the mode functions at two depths, it is often not possible to measure the mode function by itself. Because of these issues, we will actually use the ratio of each mode function to the

first mode function. This eliminates source level and calibration concerns, and only slightly complicates our calculation. If we define the quantity

$$m_n(z, z_s) \int \frac{Z_n(z)Z_n(z_s)}{Z_1(z)Z_1(z_s)},$$

we can compute the derivative using the product and quotient rules, and our earlier results. Note that when using the adiabatic approximation, $Z_n(z)$ is computed at the receiver location, and $Z_n(z_s)$ at the source location. The derivative of the ratio with respect to some parameter γ can be written as

$$\frac{\partial m_n}{\partial \gamma} = \frac{\frac{\partial Z_n(z_s)}{\partial \gamma}Z_n(z)Z_1(z_s)Z_1(z) + \frac{\partial Z_n(z)}{\partial \gamma}Z_n(z_s)Z_1(z_s)Z_1(z) - \frac{\partial Z_1(z_s)}{\partial \gamma}Z_1(z)Z_n(z_s)Z_n(z) - \frac{\partial Z_1(z)}{\partial \gamma}Z_1(z_s)Z_n(z_s)Z_n(z)}{Z_1^2(z_s)Z_1^2(z)}.$$

With this, we are now ready to set up the inversion. We will set up the matrix $\left[\frac{\partial m}{\partial X}\right]$, each row of which is the partial derivatives of one mode function at one measurement depth with respect to each of the parameters sought. For example, if the parameters sought are X_1, X_2, B_1, A_1, the first line of $\left[\frac{\partial m}{\partial X}\right]$ will be

$$\left[\frac{m_2(z_1, z_s)}{\partial X_1} \quad \frac{m_2(z_1, z_s)}{\partial X_2} \quad \frac{m_2(z_1, z_s)}{\partial B_1} \quad \frac{m_2(z_1, z_s)}{\partial A_1}\right],$$

where z_1 is the first measurement depth. We start with m_2 because m_1 is always equal to 1. The next few rows of $\left[\frac{\partial m}{\partial X}\right]$ will be the same derivatives, but evaluated at different measurement depths. Once all measurement depths have been accounted for, the next line will start over with the derivatives of m_3 at each depth, and so on. The size of $\left[\frac{\partial m}{\partial X}\right]$ will therefore be $(M-1)J \times N$ where M is the number of modes used in the inversion, J is

the number of measurement depths used, and N is the number of parameters sought. The two other components of the equation are the column vector $\vec{X}$ which contains all the parameters we seek[2] (length N), and the column vector $\Delta\vec{m}$, which contains the difference between the mode ratios at each depth and mode number (length $(M\text{-}1)J$). We then have the equation

$$\left[\frac{\partial m}{\partial X}\right]\vec{X} = \Delta\vec{m}.$$

If the rank of $\left[\dfrac{\partial m}{\partial X}\right]$ is greater than or equal to N we will have an over- or fully-determined system. If it is less than N, the system will be underdetermined, and we will require a minimum norm assumption or other such constraint in order obtain a unique solution. In either case, we can solve the equation using the pseudo-inverse introduced in chapter 1:

$$\vec{X}_{LS} = \left[\frac{\partial \tilde{m}}{\partial X}^{\#}\right]\Delta\vec{m}.$$

It must be noted that we have assumed a linear relation between the variables when performing this inversion, when in reality the relationship is non-linear. Therefore we should not expect this single-step inversion to give us the correct bottom parameters. However, if our background model is not too far from reality, then our answer should at least give a correction to our background, which can be used as a new background, and the process can be repeated. After a few iterations, the algorithm should converge to the

---

[2] Some might find it more intuitive to think of the $\vec{X}$ vector as a $\Delta\vec{X}$ vector, which contains the differences between the desired parameters in the background model and the true case. This notation becomes a bit cumbersome, however, when partial derivatives become involved, so we have opted for calling the vector just $\vec{X}$.

values that give the best possible fit (in a least squares sense) to the input data, given our parameterization.

We now have a process by which to determine the bottom parameters from measurements of the mode functions. However, the mode functions cannot be measured directly, and must be extracted from measurements of the pressure field. This is a non-trivial process, and deserves attention on its own. There are a few ways to obtain the mode functions from the pressure field, depending on the amount of *a priori* knowledge one has about the bottom. While we will be focusing on an Extended Kalman Filter method that makes use of a Vertical Line Array (VLA) and very little *a priori* knowledge of the bottom, a few other possibilities will be mentioned first.

One option available if both the bottom and water column are range independent, is to use the Hankel transform method of Rajan *et al.* [1],[18], and implemented as well by Ohta *et al.* [19],[20]. Here a measurement of pressure as a function of range is Hankel transformed into the horizontal wavenumber domain. Peaks in the transform of the pressure data correspond to the normal modes. The k_r positions of the peaks indicate the eigenvalues, and the heights of the peaks indicate the magnitudes of the mode functions at the measurement depth. This method works well when the bottom and water column are range-independent over the aperture over which the data is transformed. Further, it is not necessary to know the water column sound speed profile to obtain the eigenvalues and amplitudes. However, to invert for the bottom properties, it is still necessary to measure the water column SSP. This method encounters difficulty when the bottom changes too rapidly to allow one to form a large enough range-independent sub-section of data, or "window," to obtain sufficient resolution to distinguish modes. Also, if the water

column is changing, either in range within a window, or in time while the window is being formed, the resulting transform can indicate modes with incorrect amplitudes, or at wavenumbers which do not match the true eigenvalues. An inversion using this method will be presented in Chapter three.

Another method of measuring the mode functions was proposed by Neilsen and Westwood [21], in which a matrix of range and depth measurements of the pressure field is subjected to a singular value decomposition. This method also requires no information about the water column or bottom to extract the mode functions, though again the water column information is necessary to do the subsequent inversion for bottom parameters. This method appears to work well when the VLA used to make the measurements spans the entire water column with small distance between the hydrophones. However, because it uses the approximation that the modes are orthogonal *in the water column*, it performs poorly when estimating the mode functions of modes with significant energy in the bottom. Unfortunately, these modes provide the most information about the bottom, and neglecting them (or worse yet using poor estimates of them) results in poor inversion results. Because of this, this method is not used to obtain input data for the mode amplitude inversion in this thesis.

If additional information about the propagation environment is available, further progress can be made using what is commonly called mode filtering. These methods generally assume the mode functions themselves are known *a priori*, and compute the amount of energy carried by each mode. Another option is to assume the eigenvalues and water column SSP are known, then use the shooting method to compute the un-normalized mode shapes, and then use the VLA measurement to estimate the amount of

26

energy carried by each mode. Buck *et al.*, [22] and the references within provide a good introduction to such methods. Candy and Sullivan [23] have developed a method that estimates both the eigenvalues and mode functions using an Extended Kalman Filter (EKF), and the shooting method, though its accuracy does not appear to be sufficient for the purposes of our inversion. On the whole, these matched-mode, or mode filtering methods may work well for the source-ranging problems they were designed to solve, but do not give sufficiently accurate estimates of the higher-order mode functions to be used in our inversion.

Because of this, we propose a new method for estimating the mode functions, which borrows from Candy and Sullivan's EKF technique, but uses successive range measurements as the steps of the filter, rather than successive depth measurements. This not only gives us the accuracy needed to conduct the inversion, but also gives us a range-dependent estimate of the mode functions. Further, because the method does not use a large range-window to transform the range data to the wavenumber domain, what happens between the source and receiver (*e.g.*, unknown variability) doesn't matter, so long as the adiabatic approximation holds. That means the unknown spatial and temporal variability in the water column doesn't degrade the measurement. By making use only of the endpoints, we avoid the requirement of a stable water column and range-independent bottom over the transform window. This is the key difference that separates this inversion method from the Hankel transform technique of Rajan *et al.*[1] In fact, the EKF method about to be described can also provide range-dependent measurements of the eigenvalues, which can be used to estimate the bottom properties via Rajan's method, as well as using the mode functions as described above.

The required inputs to the method are the pressure measurements on a VLA at a sequence of ranges from the source, as well as a measurement of the water column Sounds Speed Profile (SSP) at both the source and receiver location. The requirement of knowledge of the water column SSP at two locations is stronger than most of the methods used to estimate the mode functions. However, since it is necessary to have a good estimate of the water column SSP before one can carry out an inversion for the bottom parameters, this requirement isn't a new burden for our method. We are simply making use of the information earlier in the process.

Before we delve deeper into the specifics of the mode function estimator, it will be useful to consider the Extended Kalman Filter (EKF) in general. Kalman and Extended Kalman filters are well known, and used widely in many fields. We will not attempt an in-depth examination of them here, but rather will only mention what is necessary for our method, and point the reader to references if further information is required. In the ocean acoustics community, Candy and Sullivan [23] make frequent use of these filters for a variety of tasks. They typically cite Candy's book [24] as the best source for information on the subject. Souza [3] also made use of Kalman filters in his thesis, and listed Maciej Niedźwiecki [25] as a source. We have found the relatively short and straightforward write-up of Ribeiro [26] quite useful as well, and our brief coverage here will largely follow her presentation of the EKF.

The basic idea of the Extended Kalman filters is that of estimating a sequence of quantities based on the assumption that each quantity in the sequence is related to its state in the previous step. Symbolically, $x_{k+1} = f_k(x_k) + w_k$, where x_k is an element of the sequence we are trying to estimate, $f_k(x_k)$ is the function containing the dynamics of the

28

problem at hand, and w_k is a white Gaussian process with zero mean, and covariance Q_k. The measurement model is $y_k = h_k(x_k) + v_k$, where y_k is the sequence of measurements, $h_k(x_k)$ a function relating the desired quantities to the measurement, and v_k is a white, Gaussian process with zero mean and covariance R_k representing measurement errors. The goal of the filter is to provide a minimum least squares estimate of the sequence x_k given the sequence y_k. For the problems where the standard Kalman filter is applicable, the filter obtains this goal. The EKF, however, is a linear approximation to the optimal, non-linear filter that would actually give the least mean square solution. How well the filter does depends, in part, on how good an approximation the linearization is.

For each element in the sequence of measurements, the EKF performs the following steps:

1. Start with the last filtered estimate of the state variable: $\hat{x}(k \mid k)$, and its error covariance matrix $P(k \mid k)$. Note that standard notation for Kalman filters involves two indices separated by a | symbol. The meaning of the first is the step which is being estimated, and the meaning of the second is the step up to which the data is being used for the estimate. Thus $\hat{x}(k \mid k)$ is the estimate of the state variables at the kth step, based on information in the first k measurements, whereas $\hat{x}(k \mid k-1)$ is the estimate of the state variables at the kth step, based only on the first k-1 measurements.

2. Linearize the dynamics around this value of $\hat{x}(k \mid k)$. The function f maps the state variables at one step to their values at the next step. By linearizing this function, we make it easier to predict how errors in our current estimate of the state variables will

translate into errors in our next estimate of the state variables. In a normal (*i.e.,* non-extended) Kalman filter, this step will be unnecessary, since the dynamics will already be linear.

Let $F(k) = \nabla f_k |_{\hat{x}(k|k)}$.

3. Predict the next state to be measured, and its covariance. The f function gives us the state estimate, and the linearization form of it helps us translate our uncertainty from step to step. The covariance of state variable driving noise is also added to our estimate covariance, since we expect unpredictable changes in the state variables.

$$\hat{x}(k+1|k) = f_k(\hat{x}(k|k)) \text{ and}$$

$$P(k+1|k) = F(k)P(k|k)F^T(k) + Q(k) .$$

4. Linearize the measurement dynamics around $\hat{x}(k+1|k)$. The function h maps the state variables to a measurement. Linearizing this function allows us to predict how the measurement will change due to errors in our estimate of the state variables. Again, in a normal Kalman filter, this step will be unnecessary, since the measurement dynamics will already be linear.

Let $H(k+1) = \nabla h_{k+1} |_{\hat{x}(k+1|k)}$.

5. Compute the Kalman gain, and update the estimate based on the measurement and the predicted error:

$$K(k+1) = P(k+1|k)H^T(k+1)\left[H(k+1)P(k+1|k)H^T(k+1) + R(k+1)\right]^{-1}$$

$$P(k+1|k+1) = \left[I - K(k+1)H(k+1)\right]P(k+1|k) \text{ and}$$

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + K(k+1)\left[y_{k+1} - h_{k+1}(\hat{x}(k+1|k))\right].$$

After this the steps repeat themselves for the next value. So long as the linear approximations used in steps 2 and 5 are decent, the EKF should provide a good approximation to the optimal, non-linear filter that would produce the least mean square estimate. Beyond the steps listed above, we also need initial estimates of the state variable before any measurements are made, as well as an estimate of uncertainty in that estimate.

Having taken a brief look at the general EKF problem, we are now ready to examine the specific case of using it to estimate the mode functions for our inversion. We first remind ourselves of our adiabatic model for sound propagation:

$$p(r,z;z_s) = \frac{e^{i\pi/4}}{\rho(0,z_s)} \sqrt{2\pi} \sum_n Z_n(0,z_s) Z_n(r,z) \frac{e^{i\int_0^r k_n(r')dr'}}{\sqrt{\int_0^r k_n(r')dr'}}.$$

In order to deal only with unknown quantities, we define a new variable:

$$\breve{p}(r,z;z_s) = p(r,z;z_s)\rho(0,z_s)\sqrt{\frac{1}{2\pi}}e^{-i\pi/4} = \sum_n Z_n(0,z_s) Z_n(r,z) \frac{e^{i\int_0^r k_n(r')dr'}}{\sqrt{\int_0^r k_n(r')dr'}}.$$

The quantities that we require for our inversion are the products $Z_n(0,z_s)Z_n(r,z)$, from which we can form the ratios $m_n(z,z_s) \int \frac{Z_n(r,z)Z_n(0,z_s)}{Z_1(r,z)Z_1(0,z_s)}$. We will have to take a somewhat round-about path to obtain them, however, in order to meet the requirements of the EKF.

We are assuming that we have measured the SSP in the water column at both the source and receiver locations. Without the additional knowledge of the bottom SSP, we

are not able to predict the mode functions or eigenvalues. However, given the SSP of the

water column, and a horizontal wavenumber, we can use the shooting method (see

chapter 5 of [6]) to compute the function $\zeta(z;k_r)$, which is an unnormalized version of

the mode function that would exist if $k_n = k_r$. In other words, if $k_n = k_r$, then there

exists an unknown scalar μ such that $Z_n(z) = \mu\zeta(z,k_n)$, and $\zeta(z;k_r)$ can be computed.

Because there are an infinite number of such functions (each just a scaled version of all

the others), we define $\zeta(z;k_r)$ to be the one with the particular normalization

$\int_0^d \dfrac{\zeta^2(z,k_r)}{\rho(z)} dz = 1$. If we now define the scalar $A_n = \mu Z_n(0,z_s)$, we can re-write our

modified pressure as

$$\breve{p}(r,z) = \sum_n A_n \zeta(z,k_n(r)) \frac{e^{i\int_0^r k_n(r')dr'}}{\sqrt{\int_0^r k_n(r')dr'}}.$$

This will be our model for the EKF. Our state variables will be the eigenvalues at

the source, $k_n(0)$, the eigenvalues at the VLA $k_n(r)$, and what we will call the

amplitudes, A_n. All of these are real, scalar quantities, which we will collect into the

vector $\bar{x} = \begin{bmatrix} A_1 & A_2 & \cdots & k_1(0) & k_2(0) & \cdots & k_1(r) & k_2(r) & \cdots \end{bmatrix}^T$. We point out that if we

have estimates of the amplitudes and the eigenvalues at the VLA, we can compute our

desired mode ratios, since $m_n(z,z_s) = \dfrac{A_n\zeta(z,k_n(r))}{A_1\zeta(z,k_1(r))}$.

The very first step of the filtering process is to provide estimates of the quantities

in $\bar{x}$, and a covariance matrix, P, for that estimate. The performance of the estimator

depends, in part, on the quality of these initial estimates, and they may not be easy to

provide. In areas where the bottom properties have been estimated before, a set of background models taken from the literature can provide an estimate of the state variables, and provide clues to the possible variability of that initial estimate. In other cases we may need to depend more upon intuition and experience. However we produce these initial estimates, however, it is useful run the experiment multiple times with different reasonable initial guesses. If all initial estimates lead to the same final results, we can be confident that our estimate reflects reality. On the other hand, if the output of the filter varies greatly with our initial estimates, we must be skeptical of any result we produce. In some cases we will be able to toss out initial guesses based on the non-physical results they produce, such as eigenvalues in the wrong order, or mode amplitudes varying too rapidly for the adiabatic condition to hold. When different initial guesses produce differing-but-reasonable estimations, however, we should not trust any of them until we have a reason to believe that one of our initial guesses was superior to the others. While this issue is more problematic for the EKF estimator, it is also true of the mode amplitude perturbative inversion, which can depend on the initial background model. To some extent, this is simply a danger of dealing with non-linear problems. However, by testing multiple initial conditions, we can gain confidence (or avoid undeserved confidence!) that the output of our estimator is reflecting reality, rather than just our initial guesses.

Our next requirement is a model to predict how our state vector, $\tilde{x}$, will change with movement of the source to a new range. Normally we will not know the variables will change, and will just expect some small, random variation around their current values. Thus, our dynamics function for the problem is the very simple

$\hat{x}(k+1 \,|\, k) = f(\hat{x}(k \,|\, k)) = \hat{x}(k \,|\, k)$. In other words, before we make a measurement, our best prediction is that the state vector will remain the same. This is often referred to as persistence. This is already a linear dynamics model, so we run into no problems at step 2 of the process, and $F=I$. We must also supply a covariance matrix, Q, that describes how much random change we expect in each element of the state vector. Because they depend on the environment at the source, we will expect changes in the amplitudes and source-position eigenvalues. However, we should expect no change in the eigenvalues at the VLA, unless the water column has changed. Thus, when the water column at the VLA is stable, the bottom row and right-most column of Q will contain all zeros. With F and Q defined, we are able to predict $\hat{x}(k+1 \,|\, k)$ and $P(k+1 \,|\, k)$ in step 3.

Of course, in reality the water column will not be stable, and the quantities in question will change even at the stationary VLA. However, if we know the change in the water column from the last step, we can use our perturbation results to predict the changes in the values of the eigenvalues and mode amplitudes. This is an exception to our usual persistent model, but so long as the change in water column is benign enough for the adiabatic propagation model to hold, we should still have no problems at step 2. When the water column is changing we will also need to change the bottom row and right-most column of Q to indicate the uncertainty of our estimate of the new values of the state variables.

In step 4 we must linearize our measurement model, and compute the matrix H. We will use the modified pressure at each depth as our measurement, as discussed above. However, since the modified pressure is complex, and all the quantities we seek are purely real, we must take some care to keep all equations purely real. Otherwise, the

filter will likely generate complex values for the state variables. We therefore treat the real and imaginary parts of the modified pressure as separate, purely real variables. Our *h(x)* function becomes:

$$\bar{y}_k = h(\bar{x}_k) = \begin{bmatrix} \sum_n A_n \zeta(z_1, k_n(r)) \dfrac{\cos\left(\int_0^r k_n(r')dr'\right)}{\sqrt{\int_0^r k_n(r')dr'}} \\[2em] \sum_n A_n \zeta(z_1, k_n(r)) \dfrac{\sin\left(\int_0^r k_n(r')dr'\right)}{\sqrt{\int_0^r k_n(r')dr'}} \\[2em] \sum_n A_n \zeta(z_2, k_n(r)) \dfrac{\cos\left(\int_0^r k_n(r')dr'\right)}{\sqrt{\int_0^r k_n(r')dr'}} \\[2em] \sum_n A_n \zeta(z_2, k_n(r)) \dfrac{\sin\left(\int_0^r k_n(r')dr'\right)}{\sqrt{\int_0^r k_n(r')dr'}} \\[1em] \vdots \end{bmatrix}.$$

Because the integral over range will occur at all steps, but the only change occurs over the new range interval near the source, we define the quantity

$$\varphi_{n,k} = \int_0^{r_k} k_n(r')dr' \approx \sum_{j=1}^k k_n(r_j)\Delta r_j , \quad \text{and} \quad \text{note} \quad \text{that} \quad \varphi_{n,k+1} = \varphi_{n,k} + k_n(r_{k+1})\Delta r_{k+1}, \quad \text{and}$$

$$\dfrac{d\varphi_{n,k}}{dk_n(r_k)} = \Delta r_k .$$ So when we linearize our measurement equation we get

$$H = [H_1 \mid H_2 \mid H_3], \text{ with}$$

$$H_1 = \begin{bmatrix} \zeta(z_1,k_1(r))\dfrac{\cos(\varphi_{1,k+1})}{\sqrt{\varphi_{1,k+1}}} & \zeta(z_1,k_2(r))\dfrac{\cos(\varphi_{2,k+1})}{\sqrt{\varphi_{2,k+1}}} & \cdots \\[2ex] \zeta(z_1,k_1(r))\dfrac{\sin(\varphi_{1,k+1})}{\sqrt{\varphi_{1,k+1}}} & \zeta(z_1,k_2(r))\dfrac{\sin(\varphi_{2,k+1})}{\sqrt{\varphi_{2,k+1}}} & \cdots \\[2ex] \zeta(z_2,k_1(r))\dfrac{\cos(\varphi_{1,k+1})}{\sqrt{\varphi_{1,k+1}}} & \zeta(z_2,k_2(r))\dfrac{\cos(\varphi_{2,k+1})}{\sqrt{\varphi_{2,k+1}}} & \cdots \\[2ex] \zeta(z_2,k_1(r))\dfrac{\sin(\varphi_{1,k+1})}{\sqrt{\varphi_{1,k+1}}} & \zeta(z_2,k_2(r))\dfrac{\sin(\varphi_{2,k+1})}{\sqrt{\varphi_{2,k+1}}} & \cdots \\[2ex] \vdots & \vdots & \ddots \end{bmatrix},$$

$$H_2 = \begin{bmatrix} \dfrac{-\Delta r_{k+1}}{\sqrt{\varphi_{1,k+1}}}A_1\zeta(z_1,k_1(r))\left[\sin\varphi_{1,k+1}+\dfrac{\cos\varphi_{1,k+1}}{2\varphi_{1,k+1}}\right] & \dfrac{-\Delta r_{k+1}}{\sqrt{\varphi_{2,k+1}}}A_2\zeta(z_1,k_2(r))\left[\sin\varphi_{2,k+1}+\dfrac{\cos\varphi_{2,k+1}}{2\varphi_{2,k+1}}\right] & \cdots \\[3ex] \dfrac{-\Delta r_{k+1}}{\sqrt{\varphi_{1,k+1}}}A_1\zeta(z_1,k_1(r))\left[-\cos\varphi_{1,k+1}+\dfrac{\sin\varphi_{1,k+1}}{2\varphi_{1,k+1}}\right] & \dfrac{-\Delta r_{k+1}}{\sqrt{\varphi_{2,k+1}}}A_2\zeta(z_1,k_2(r))\left[-\cos\varphi_{2,k+1}+\dfrac{\sin\varphi_{2,k+1}}{2\varphi_{2,k+1}}\right] & \cdots \\[3ex] \dfrac{-\Delta r_{k+1}}{\sqrt{\varphi_{1,k+1}}}A_1\zeta(z_2,k_1(r))\left[\sin\varphi_{1,k+1}+\dfrac{\cos\varphi_{1,k+1}}{2\varphi_{1,k+1}}\right] & \dfrac{-\Delta r_{k+1}}{\sqrt{\varphi_{2,k+1}}}A_2\zeta(z_2,k_2(r))\left[\sin\varphi_{2,k+1}+\dfrac{\cos\varphi_{2,k+1}}{2\varphi_{2,k+1}}\right] & \cdots \\[3ex] \dfrac{-\Delta r_{k+1}}{\sqrt{\varphi_{1,k+1}}}A_1\zeta(z_2,k_1(r))\left[-\cos\varphi_{1,k+1}+\dfrac{\sin\varphi_{1,k+1}}{2\varphi_{1,k+1}}\right] & \dfrac{-\Delta r_{k+1}}{\sqrt{\varphi_{2,k+1}}}A_2\zeta(z_2,k_2(r))\left[-\cos\varphi_{2,k+1}+\dfrac{\sin\varphi_{2,k+1}}{2\varphi_{2,k+1}}\right] & \cdots \\[3ex] \vdots & \vdots & \ddots \end{bmatrix}$$

, and

$$H_3 = \begin{bmatrix} A_1\dfrac{\cos(\varphi_{1,k+1})}{\sqrt{\varphi_{1,k+1}}}\dfrac{\partial\zeta(z_1,k_1(r))}{\partial k_1(r)} & A_2\dfrac{\cos(\varphi_{2,k+1})}{\sqrt{\varphi_{2,k+1}}}\dfrac{\partial\zeta(z_1,k_2(r))}{\partial k_2(r)} & \cdots \\[3ex] A_1\dfrac{\sin(\varphi_{1,k+1})}{\sqrt{\varphi_{1,k+1}}}\dfrac{\partial\zeta(z_1,k_1(r))}{\partial k_1(r)} & A_2\dfrac{\sin(\varphi_{2,k+1})}{\sqrt{\varphi_{2,k+1}}}\dfrac{\partial\zeta(z_1,k_2(r))}{\partial k_2(r)} & \cdots \\[3ex] A_1\dfrac{\cos(\varphi_{1,k+1})}{\sqrt{\varphi_{1,k+1}}}\dfrac{\partial\zeta(z_2,k_1(r))}{\partial k_1(r)} & A_2\dfrac{\cos(\varphi_{2,k+1})}{\sqrt{\varphi_{2,k+1}}}\dfrac{\partial\zeta(z_2,k_2(r))}{\partial k_2(r)} & \cdots \\[3ex] A_1\dfrac{\sin(\varphi_{1,k+1})}{\sqrt{\varphi_{1,k+1}}}\dfrac{\partial\zeta(z_2,k_1(r))}{\partial k_1(r)} & A_2\dfrac{\sin(\varphi_{2,k+1})}{\sqrt{\varphi_{2,k+1}}}\dfrac{\partial\zeta(z_2,k_2(r))}{\partial k_2(r)} & \cdots \\[3ex] \vdots & \vdots & \ddots \end{bmatrix}$$

It is worth pointing out an odd issue of notation here. The wavenumbers at the VLA position have been denoted by $k_n(r)$, indicating that they are evaluated at the receiver position. This creates the impression that they should vary with range. However, since it is actually the source that is in motion, if the water column is stable,

$k_n(r)$ should remain the same for all values of r. This is the reason that r has not been labeled with k subscripts when it is the argument of the VLA eigenvalue.

It is within this step of creating the H matrix that care must be taken that our linearization is valid. Specifically, the H_2 portion of H deals with derivatives of the trigonometric functions, which are non-linear. For higher order terms not to dominate, we must have that $\Delta k_n \Delta r \ll 2\pi$. If we consider a linear approximation to the trigonometric functions to be valid within an eighth of a cycle, we must have $\Delta r < \dfrac{\pi}{4\Delta k_n}$.

This sets the maximum range steps we can use, given an estimate of the maximum change in an eigenvalue over distance. Changes in eigenvalues over any given range step should be expected to be quite small because if they weren't, the adiabatic approximation is likely to be invalid. The H_3 portion of H also involves a linear approximation to non-linear functions. However, for changes in the eigenvalues small enough to make the H_2 portion valid, the H_3 portion should be fine.

Given H, all we still need is the covariance matrix of the measurement noise, R, in order to compute the Kalman gain and filter our predicted values of the state vector and its covariance. After this, the process repeats itself, providing an estimate of the state vector and state covariance matrix at each range. The information in the state vector is then passed to the inversion algorithm described earlier in this chapter, and the bottom parameters extracted, providing us with our range-dependent inversion of the bottom. The R matrix involves a few factors, all which should be determined or estimated before the experiment. The first is the error due to the abilities of the instrument itself and a second is the errors that processing of the data from the receivers can cause. A third is

noise due to sound sources other than the experimental source, such as breaking waves, passing ships, etc. All of these potential sources of error should be estimated as part of the experiment to produce the R matrix.

One might wonder why the EKF estimator has been proposed for the task of estimating the range-dependent mode functions and eigenvalues, instead of other non-linear estimators. The main advantage of the EKF estimator is that it can, in theory, be run in real time during the experiment, whereas other methods make use of the entire data set from the start. Thus, it may be possible at some point to conduct an experiment using the EKF estimator in which the bottom properties are estimated as the acoustic measurements are being made, rather than waiting to process the data until after all measurements have been taken. Another advantage of the EKF is that it naturally deals with parameters that change from step to step. If we want to estimate the mode functions and the source- and receiver-position eigenvalues for seven modes at one thousand different ranges, that leaves us with 21,000 unknown parameters to solve for. Trying to do so all at once, with a nonlinear least squares estimator would likely involve prohibitive computational requirements. The nonlinear least squares algorithm is also very sensitive to outliers, whereas the EKF method is better at rejecting spurious measurements. Another option might be a maximum likelihood estimator, but this also suffers from trying to estimate a huge number of parameters at once. Further, it requires the user to specify a probability density for the unknowns, which is even more difficult to provide than the initial input for the EKF estimator. It is entirely possible that another non-linear estimator will be better suited to the task of determining the modal parameters from the

acoustic field than the EKF estimator, but the EKF estimator seemed a natural fit to the

task, and, as will be seen in chapter three, is able to perform the task well.

Chapter 3: Examples

In this chapter we will examine a number of synthetic and real-world experiments in order to test the validity and usefulness of the method described in the last chapter. We will start with very simple, idealized situations to test the basic concepts of the method, and work our way through increasingly complicated and difficult situations. By the end of the chapter we should be convinced that the method is capable of extracting geoacoustic data from the measurements of the pressure field in a weakly range-dependent environment.
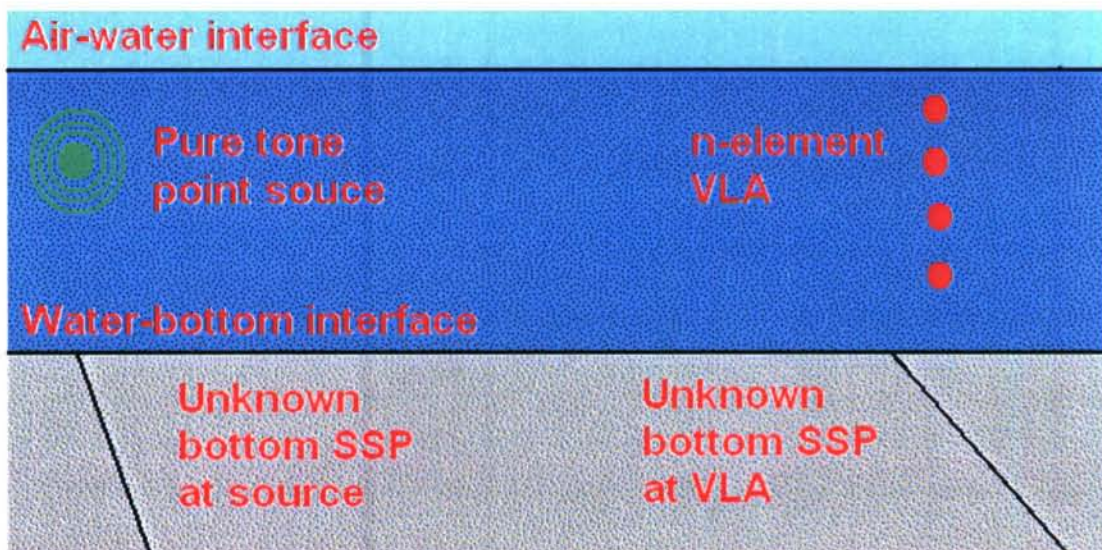
Figure 3.1 The generic setup of the Geoacoustic inversion experiments.

The basic set up of all the experiments in the chapter is shown in figure 3.1. The components needed are a pure-tone point source, a VLA to take the pressure measurements, and the unknown bottom profiles at the source and receiver locations. The acoustic properties of the water column are assumed to be measured at both the

source and receiver locations. This will require measurements of the water column at both the source and receiver positions. While such measurements are often made at one location, they have not typically been made at both locations during most ocean acoustics experiments. The number of the elements in the VLA will differ from experiment to experiment, and in some cases the array may only contain one element. When a range-dependent inversion is sought, either the source or the VLA must move, as the inversion only estimates local parameters at each location. Ideally, the VLA would be moved, so that the modes functions at the receiver position can be estimated better, but moving a VLA is far more difficult than moving a source, so usually it will be the source that is moved. In range-independent cases, nothing need move, as the SSP is the same everywhere.

Our first test of the method will be a very simple, two parameter, range-independent experiment, with perfect input data. This will be an idealized proof-of-concept test, which will let us know that the equations derived in chapter 2 are valid. We hypothesize a "true" waveguide of 75m of iso-velocity water over 25m of layered sediment. Below this we add a half space which continues to infinite depth. We excite a pressure field by adding an acoustic point source at 50m depth, emitting a pure tone at 50 Hz. We treat the density as constant (and known) in the bottom for simplicity. We use the normal mode program KRAKEN [27] to generate the mode functions that will be excited in this waveguide. For this first test, we will use the mode functions directly, rather than computing the field and trying to estimate them. This separates out any potential problems in the inversion method from problems in the mode function estimation algorithm. Since we are using the mode functions directly, and the problem is

range-independent, we do not need to specify a range for the VLA. We do, however, need to specify the depths of the receivers, and in this case we will use a 6-element VLA with 10m spacing[3]. The general set up is shown in figure 3.2.
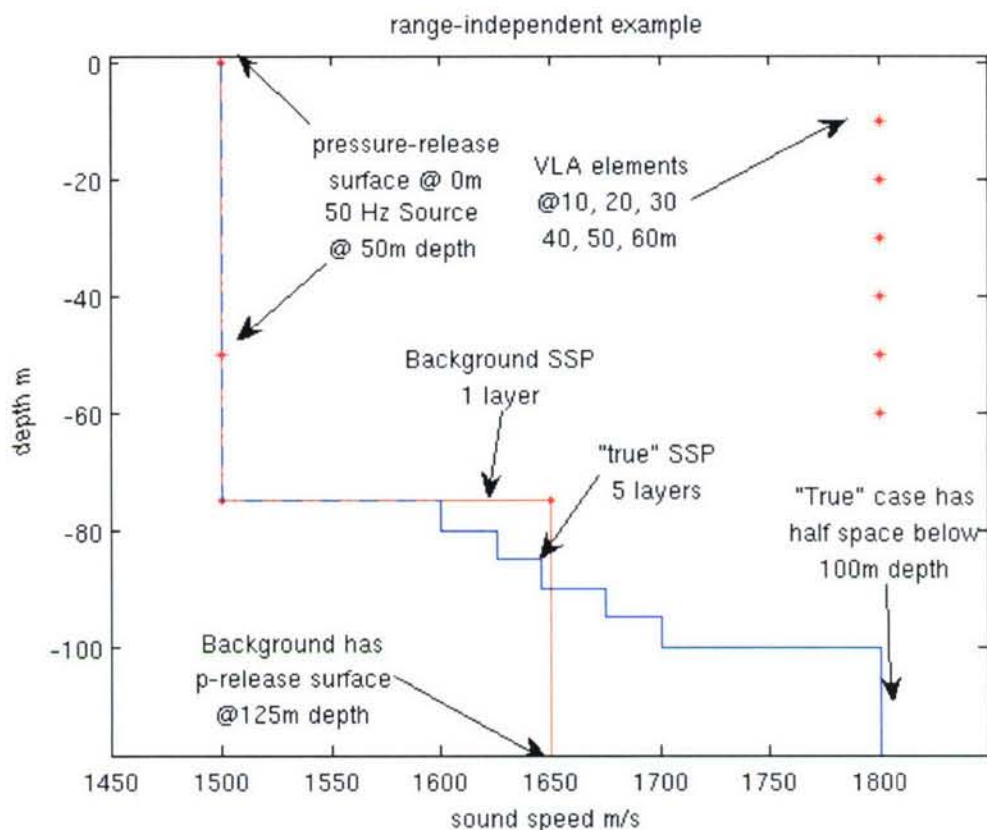

Figure 3.2 The SSP and experiment geometry for the first test of the method.

The information given so far is sufficient for solving the forward problem. However, to perform the inversion we also need to specify a background model, and parameterize the bottom. We choose a simple two-parameter bottom model which

---

[3] It should be noted here that the selection of the number of VLA elements used in the synthetic examples in this thesis is somewhat arbitrary. Generally the number of elements was selected based on what was considered realistically feasible, rather than on the requirements of the methods using them. In general, the more elements in the VLA, the better the results will be, especially with the EKF estimator. However, as a general rule of thumb, one element per propagating mode should be sufficient. Ideally, however, there will be at least 3 elements per mode, as this will provide as many measurements at each range as there are unknowns (these measurements will not be independent, however).

describes the bottom using only the sound speed at the water-bottom interface, and the sound speed gradient. This model is appealing because the low number of parameters allows us to solve a fully-determined problem (there are only 3 propagating modes in this waveguide), and because it is often capable of describing the true bottom as reasonable first approximation. While there are layers in the true bottom, the rate of increase of the sound speed of the layers is often nearly linear with depth, meaning a gradient can do well to approximate the SSP.

For the background model, we use a Pekeris-like waveguide, with a sound speed of 1650m/s throughout the bottom. However, unlike the standard Pekeris model, we include a pressure-release surface at 125m depth. The reason for this is that the expression for the derivative of the mode function includes a sum over all modes. In order to have a complete set, we must have an infinite number of modes, which is only possible if we include the leaky, non-propagating modes with complex eigenvalues. Solving for complex eigenvalues is a difficult task, and greatly increases the computation time needed for solving the problem. By adding a pressure-release false bottom, we can maintain a proper Sturm-Liouville problem, and the eigenvalues of the non-propagating modes become purely imaginary, and thus much easier to compute. The error introduced by our use of this false bottom should be small if it is placed at a depth below the turning point of the highest propagating mode [10]. This methodology will be used in all of our inversions, so it is important to test it in this early trial.

All the necessary components now being in place, we conduct the inversion. The mode functions are computed for the background model, and the mode ratios computed for each mode, and each receiver depth. These are compared to the mode ratios formed

43

using the output from KRAKEN. The derivatives of the ratios are computed following the results of chapter 2. The linear set of equations is solved using the pseudo-inverse, giving the necessary changes in the two parameters. These changes in the parameters are incorporated into the background model, and the process repeated again. After five such iterations, the method converges to the result shown in figure 3.3.
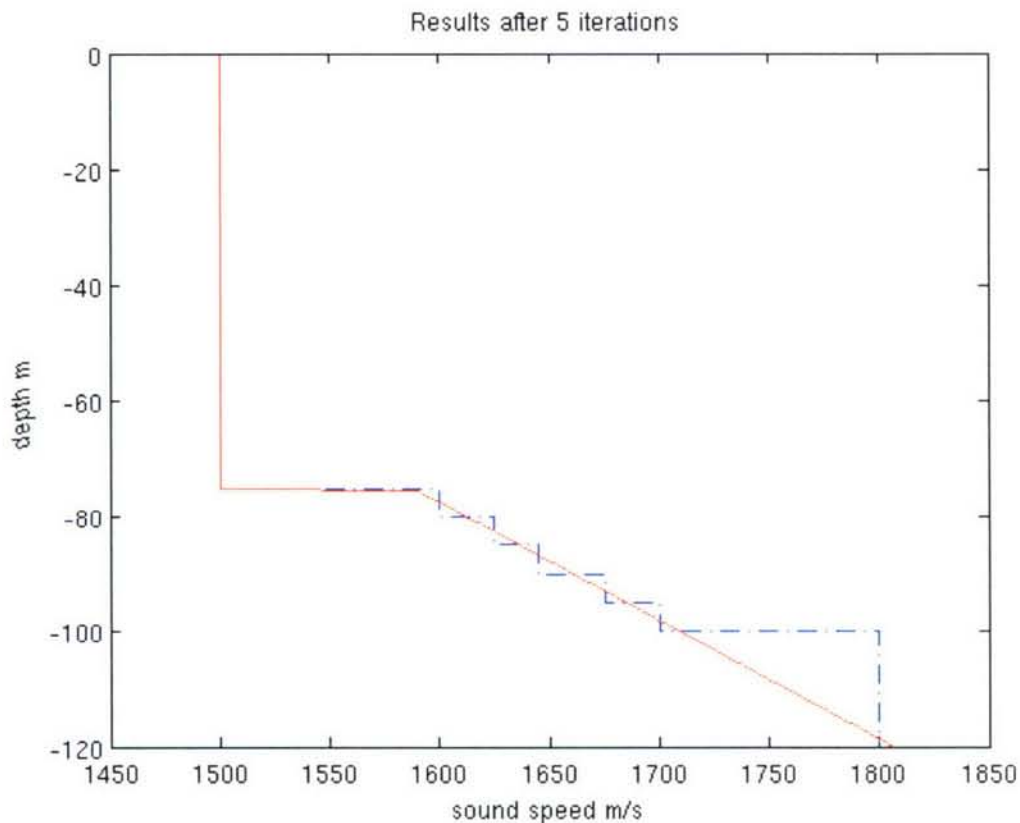

Figure 3.3 The results of the geoacoustic inversion after 5 iterations. The dotted blue curve shows the "true" SSP, while the solid red curve indicates the inversion result.

As the figure shows, the inversion result matches the "true" SSP quite well. There is non-trivial disagreement below 100m depth, but using our simple 2-parameter model, this is unavoidable. For the first 25m of the bottom, the fit is just about as good as our model would allow, showing that given perfect input data, the method will provide a good estimate of the bottom SSP. Also, since this good result was obtained using a

background model containing a pressure-release false bottom, we gain confidence that adding such false bottoms in a proper fashion will not corrupt the inversion results.

It clear that such an over-simplified and idealized test is insufficient to determine the merit of the inversion method. However, as a first result it is encouraging. By adding layers of complexity and realism to the simulated experiments, we can better learn how well the method is likely to perform in the real world. The next step in testing the model is to add an element of range-dependence. For the moment we will continue to treat the input data as given, as we are still trying to probe the feasibility of the inversion method itself, rather than deal with the separate-but-related issue of data collection.

Test two will investigate the method's ability to handle range dependence by using two waveguides. One waveguide will correspond to the environment near the source, and the other to the environment near the VLA. We will assume that sound can propagate adiabatically from the source to receiver, and again make use of the mode functions directly. However, since the mode functions will be different at the source and receiver, we will need two sets of them. Also, we will need two sets of parameters to describe the two bottoms.

The information needed to complete the forward problem in test 2 is shown in figure 3.4. The water column at the source is 100m deep, and a 125 Hz, pure tone source is placed at a depth of 50m. The bottom consists of 100m of sediment with an interface sound speed of 1495m/s, and a gradient of $3s^{-1}$. Below the sediment is an iso-velocity half-space. At the VLA position we have 90m of water, and 110m of sediment. The interface speed (1550m/s) and gradient ($2s^{-1}$) both differ from those at the source location. The bottom half-space is the same. The VLA itself consists of 4 elements with

20m spacing. Again we use KRAKEN to compute the mode functions for the 13 propagating modes, and use them to create the mode ratios that are the input to our method.
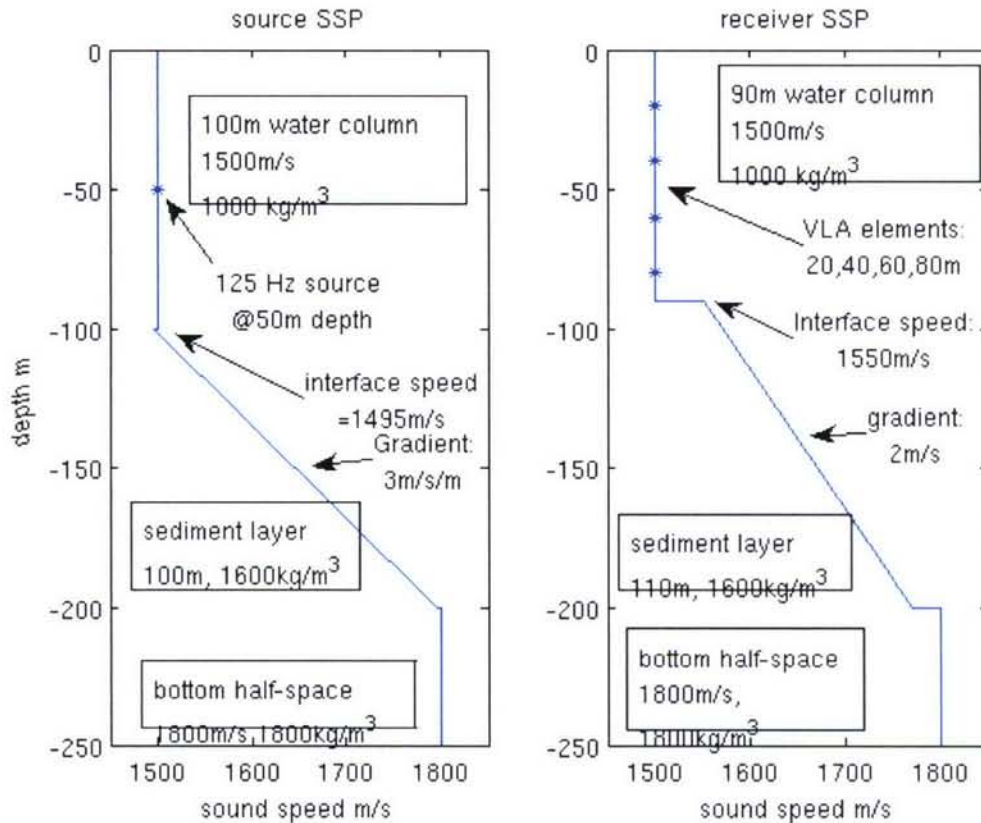

Figure 3.4 The two waveguides for the range-dependent problem of test 2.

As background models, we use Pekeris-like bottoms with pressure-release surfaces at 200m depth. We treat the iso-velocity water columns and the density profiles as known at both locations. We use the backgrounds to compute two sets of normal modes, and form the mode ratios for comparison with the KRAKEN results. We compute the derivatives, and solve our linear set equations, and repeat this process until convergence (in this case 5 iterations). The results of this inversion can be seen in figure 3.5.

As can be seen in the figure, the method does extremely well in estimating the SSP at both the source and receiver positions, showing that range-dependent inversion are definitely possible. Granted, our parameterization of the bottom matched the true bottom perfectly, and there was no measurement error in our input data, so this is still a far from realistic test case. Again, though, the result is very encouraging.
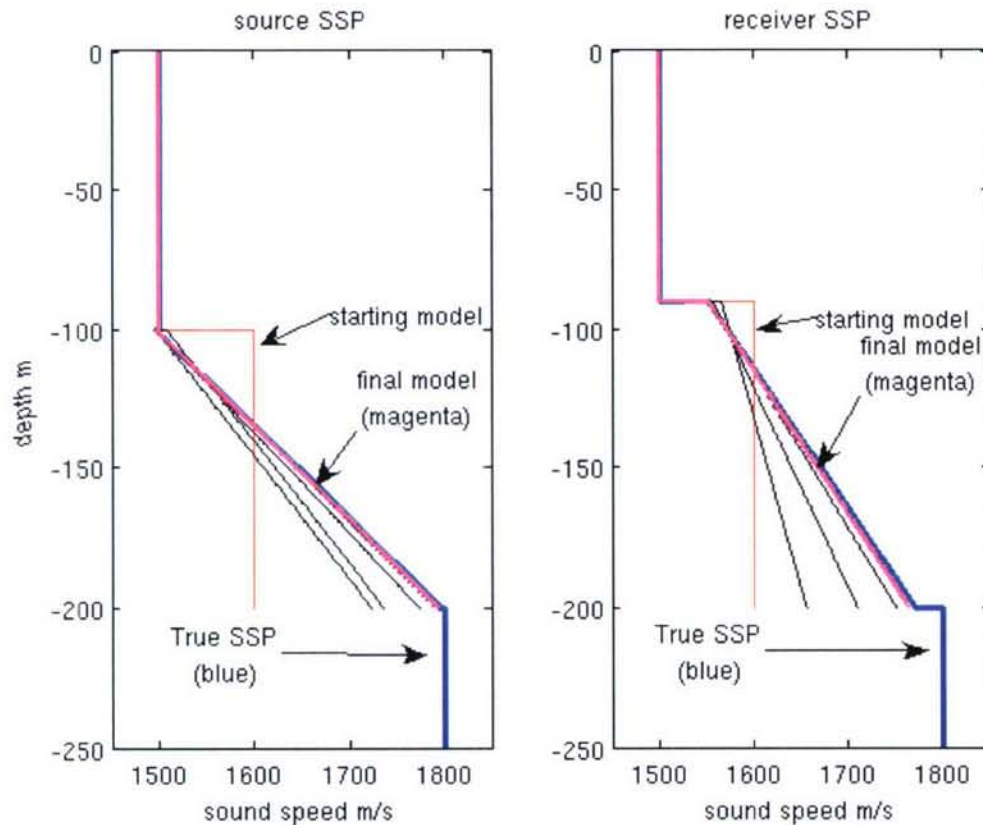

Figure 3.5 The results of the range-dependent inversion. The left panel shows the source location, and the right the VLA position. The red curves are the background models, the blue curves the true SSPs, the black curves pre-convergence estimates of the SSP, and the magenta curves the estimates after convergence.

At this point, we give up the assumption of perfect input data, and include in the problem the estimation of our input from the pressure field. For test three, we will use data from the MOMAX 2 experiment, which was carried out in the Gulf of Mexico in February of 1999 [28], [29]. This experiment is somewhat different from the generic

version shown in figure 3.1. Most strikingly, the pressure measurement is taken by only one receiver, so the "array" has only one element. Further, it was this receiver which moved during the experiment, not the source. A synthetic aperture horizontal line array (HLA) was formed as the receiver drifted, and the modal amplitudes could be obtained using a Hankel transform of the data, following the method of Rajan, *et al* [1]. This was possible because the waveguide was range-independent over the section of data used, and the 70m-deep water column was temporally stable during the time period over which the synthetic aperture was formed.
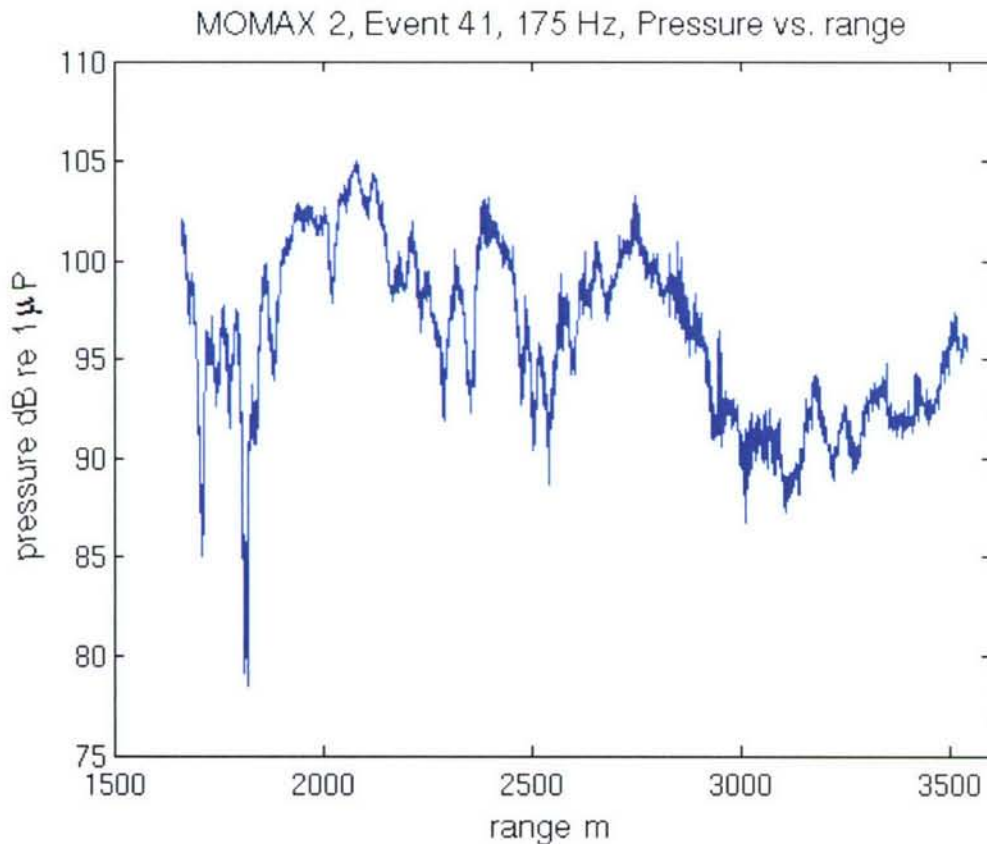
Figure 3.6 The log magnitude of the pressure as a function of range in MOMAX 2, event 41, at 175 Hz.

Figure 3.6 shows the pressure as a function of range measured by the drifting receiver in event 41 of MOMAX 2, after the raw pressure measurement was demodulated

for the source frequency of 175 Hz. This particular ~2 km section was selected from all the data from event 41 because it appeared that during the time it was taken, the water column was quite stable. Data at other, lower frequencies were also taken, but the transforms of these data did not have sufficient aperture to resolve all the modes, due to the fact that the longer wavelength of sound at the lower frequencies made the 2 km of acceptable data a shorter effective window.

Following the methodology of Rajan *et al.*[1], the data in figure 3.6 were Hankel transformed into the horizontal wavenumber domain. The result is plotted in figure 3.7. The 19 modal peaks used in the inversion method are indicated with pink asterisks. While it may seem unreasonable to assume so many peaks are modal peaks rather than side lobes of the finite-aperture transform, each peak lined up surprisingly well with the expected eigenvalue locations based on prior estimates [30] of the bottom profile. Further, the widths of the peaks also seem to indicate main lobes, rather than side lobes, as can be seen by observing the narrow side lobes to the right of the first (right-most) mode. It is possible that a few of the modes have been misidentified, but we believe the large majority of the modal peaks are correctly interpreted. The possibility of misidentifying modal peaks is a common problem when dealing with real-world data, and is something the method must be able to handle. In order to reduce the errors introduced by a few misinterpreted peaks, we again use a simple two-parameter model for the seafloor. By solving an overdetermined problem we tend to reject un-biased errors in our input, whereas if we solved an underdetermined problem, we would create bottom properties to fit the errors we introduced in the measurement. The cost of this, however, is that we are limited to inverting very simple SSPs in the bottom. Fortunately, our best

estimate of the bottom before the experiment [30], predicted a bottom that could be described reasonably with our 2-parameter model.
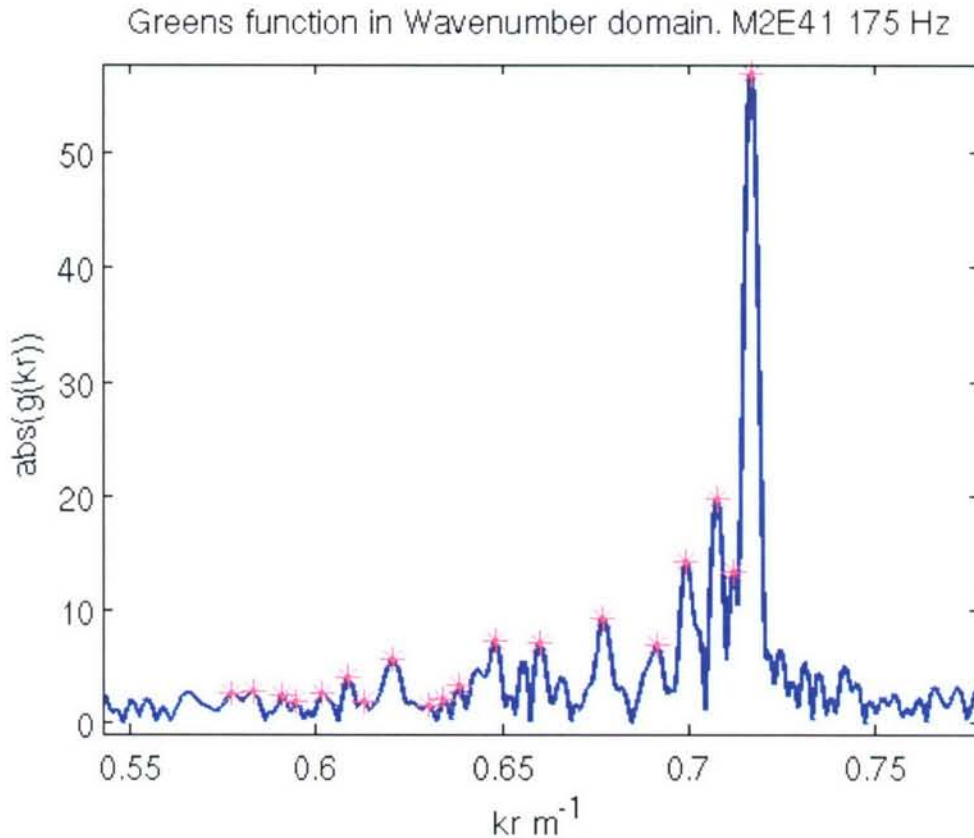

Figure 3.7 The Hankel transform of the pressure vs. range data show in figure 3.6. Mode peaks used in the inversion are indicated with pink asterisks.

For a background model, we again choose a simple Pekeris-like model. While we have a better *a priori* estimate of the bottom SSP in this case, we usually will not have such a good starting point, so it is a better test of the method to start with a poorer background model. We use our background model to compute the mode ratios at the receiver depth, and compare those to our measurements. We compute the derivatives of the ratios with respect to our two parameters, and solve the linear system for the necessary changes in the parameters. We repeat this process until convergence, and the results are shown in figure 3.8.

50

As can be seen in the figure, the final inversion result matches the expected profile of the bottom very well. That 25 iterations were required before convergence may be a sign that some modes were misidentified, or that our bottom parameterization failed to capture some of the features of the true SSP. We can gain further insight into these possibilities by looking at the mode function ratios. Figure 3.9 depicts the mode ratios of
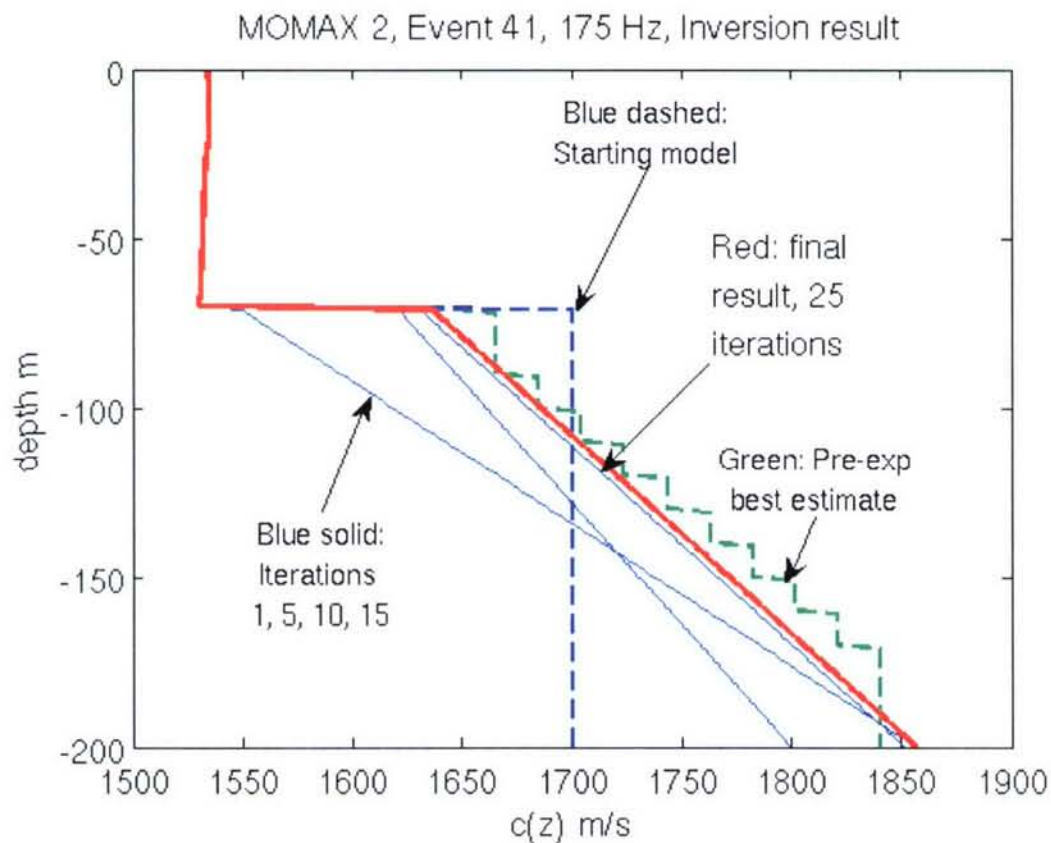

Figure 3.8 The results of the geo-acoustic inversion of the 175 Hz data from MOMAX 2, event 41. The dashed blue curve shows the starting model of the bottom, and the solid blue curves show iterations 1, 5, 10, and 15. The red curve is the final result after convergence at 25 iterations. It matches quite well the best *a priori* estimate of the bottom profile, shown in dashed green.

the final inversion result, and also those that were measured. About two thirds of the mode ratios measured match the modeled values well. Modes 5, 7, 8, 14, 15, and 16, however do not. With only two parameters, it is not possible to match all the mode ratios well, so a solution that minimizes the mean-square differences is accepted. If more

parameters were used, we would likely get a better match to the measured values. If more than 19 parameters were used, we would likely get a perfect match. However, since we expect some error in our input parameters, a perfect match to them would not necessarily mean a perfect match to the true bottom profile. More likely, it would mean that additional features were present in our inversion model that weren't in the true profile. These false features would be present due to their ability to "fit the error." Yet another possibility is that uncertainty in the water column or density profiles (which were treated as known) could be the source of disagreement. Errors in the input and inversion result will be discussed in greater detail in chapter 4.
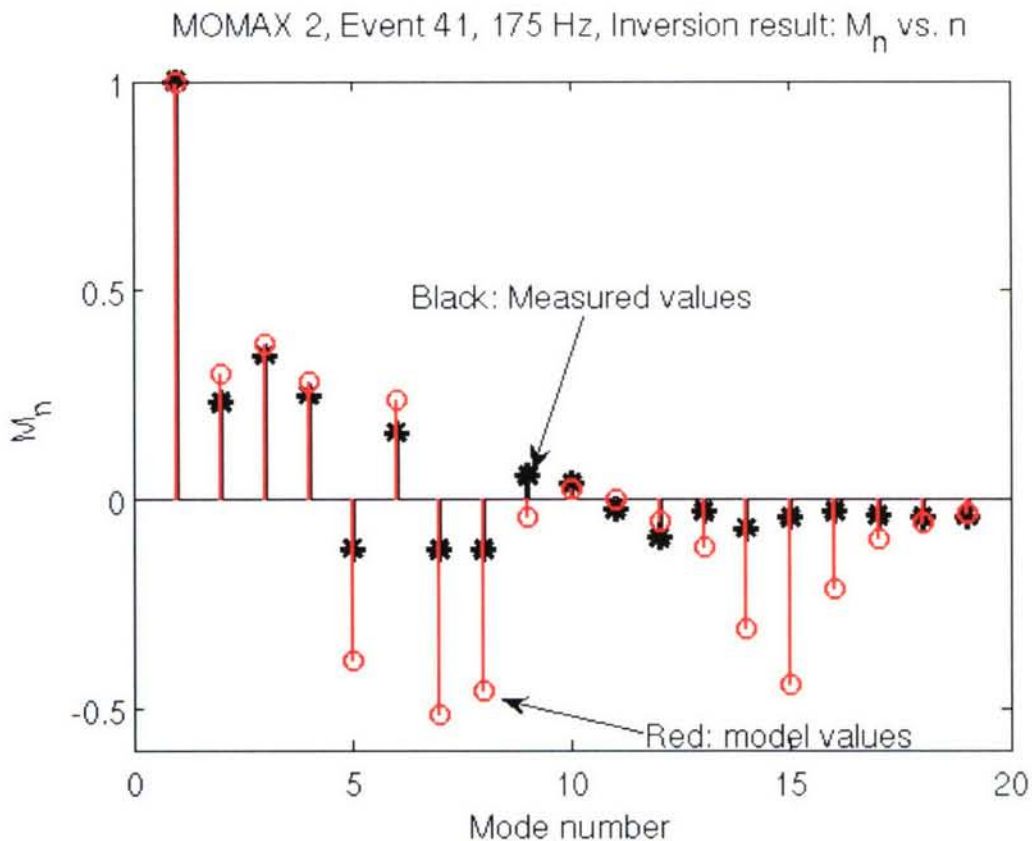

Figure 3.9 Mode function ratios at the receiver depth as a function of mode number in the 175 Hz data from MOMAX 2, event 41. Red circles indicate the mode ratios for the inverted bottom model, and black asterisks indicate values based on the Hankel transform of the measured pressure data.

Our next test of the method will in one sense be a step backwards, since we are returning to synthetic data, and less-than-realistic input. In another sense, however, it will be a step forward, as it will be our first inversion that inverts for parameters as a function of range. Our waveguide will consist of 50m or water above a two parameter bottom with constant density. However, the values of the parameters will be different at each range, each progressing as a random walk from initial values of 1550 m/s for the interface sound speed, and 2.1 s^{-1} for the gradient. The waveguide will be excited by a 100 Hz pure tone source at 18m depth. The measurement will be made at a 9-element VLA with 5m spacing. We simulate the source being towed away from the VLA, and the sound propagating adiabatically to the array. For the moment we take the estimation of the mode amplitudes as a given, leaving that problem to the next test, though we add noise to the estimates to simulate imperfect estimation of the values.

A representative sample of the synthetic input data is shown in figure 3.10. It depicts the modal amplitudes $Z_n(z;0)Z_n(z_s;r)$ at one of the receiver depths as a function of source range. Error has been added directly to the true quantity to simulate imperfect estimation of the amplitudes. In reality, the error for the various amplitudes would be more complicated. For instance, the large-amplitude first mode would likely be estimated quite well, whereas modes with lower amplitudes would be estimated more poorly. Further, errors in different modes would likely be somewhat correlated. However, despite these unrealistic simplifications, the problem is still non-trivial.
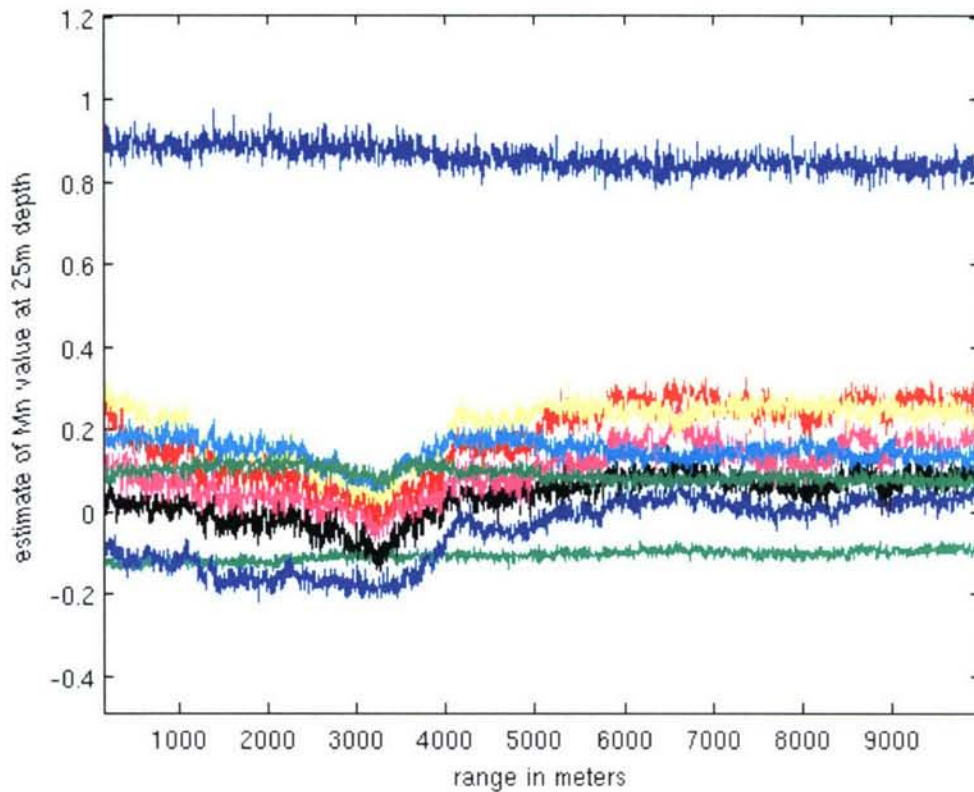
Figure 3.10 Mode amplitudes at 25m depth as a function of source range.

The fact that we will be inverting for the parameters at the source position warrants further mention. Since the environment at the location of the VLA is unchanging, the eigenvalues and mode functions there will be the same, regardless of the source position. What changes is the excitation of each mode at the source, and thus the amount of energy in each mode when it reaches the VLA. This is why we require that the sound propagate adiabatically. If energy transfers between the modes en route to the VLA, the modal amplitudes measured by the receivers will no longer represent the excitation of the modes at the source position. The excitation of the modes is based on the mode functions at the source position, and the mode functions are what allow us to extract the bottom parameters. Ideally, it would be the VLA that moves, since then the VLA would be observing changing mode functions, rather than changing excitations.

However it is far more difficult to move a VLA than to tow a source, so we simulate the more likely case of a moving source in our synthetic examples.
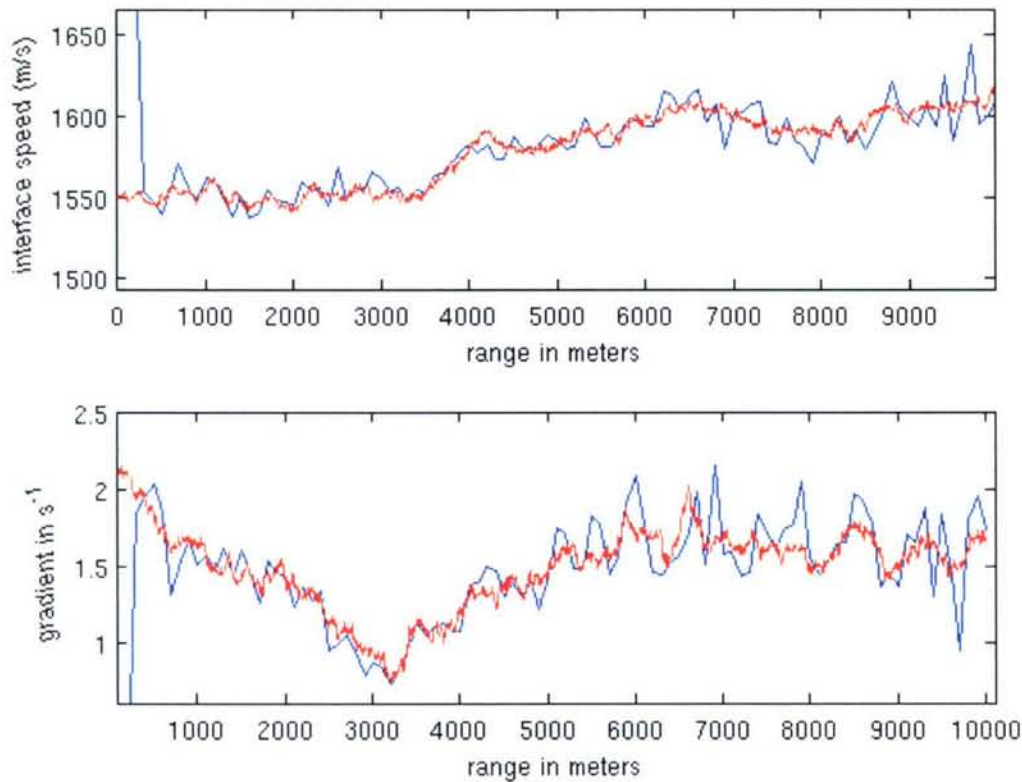

Figure 3.11 Range-dependent inversion results for interface sound speed (top) and gradient (bottom). In both plots, the true values are shown in red, and the inverted values shown in blue.

Figure 3.11 shows the result of the range-dependent inversion. The true values of the parameters are shown in red as a function of range, and the inverted values shown in blue. While there are errors visible in the estimate, this is to be expected when there are errors in the input data, and overall the inversion does a good job of capturing the range-dependent features of the bottom. The large errors in the estimate near the VLA (zero range) are due to the fact that only three iterations were used at each range step. Thus, for the first few ranges, the method had not yet converged. The last iteration of each range step was used as the starting model for the next range step. Since there was little

change in the bottom between range-steps, three iterations was normally sufficient to invert for the new bottom. Using more iterations would reduce the error at the shortest ranges, and might also slightly reduce errors at other ranges, but it would greatly increase computation time.

We have now seen evidence that we can invert for bottom parameters if we have good estimates of the modal amplitudes. But so far, with exception of the MOMAX 2 case, we have neglected the very important process of estimating the amplitudes from the pressure measurements. We were only able to get the amplitudes form the MOMAX 2 case because it had a range- and time-independent waveguide, allowing us to make use of the Hankel transform methodology of Rajan *et al* [1]. In many cases, the waveguide will not have these properties, and we will have to make use of other methods to extract the necessary information from the pressure field. Continuing towards our goal of a realistic range-dependent inversion, we will next drop our assumption that the input data are given, and include the estimation of the modal amplitudes as part of the problem.

In preparation for the Shallow Water '06 (SW06) experiment, an experiment was designed that will use the methods described in this thesis. The SW06 experiment took place off the coast of New Jersey near 73W lat and 39N long, in an area where buried river beds have been found. Figure 3.12 shows the region of the experiment, indicating the positions and depths of the buried river channels [31-33].

The experiment is intended both to test the inversion algorithm presented in this thesis, and to study sound propagation across these buried river channels. Thus, the design calls for placing a vertical line array (VLA) at 72°52' W lat, 39°16' N long. This location has the property that there are river channels in every direction from the VLA,

thus ensuring that a source drifting away from the VLA will cross at least one channel. The VLA itself will be cut for 80m water depth, and consist of 8 elements evenly spaced over 60m (approximately 10m to 70m depth) of its length. In addition to the hydrophones, the array will also contain six temperature sensors placed near the hydrophones so that the SSP at the receiver position is always well known.
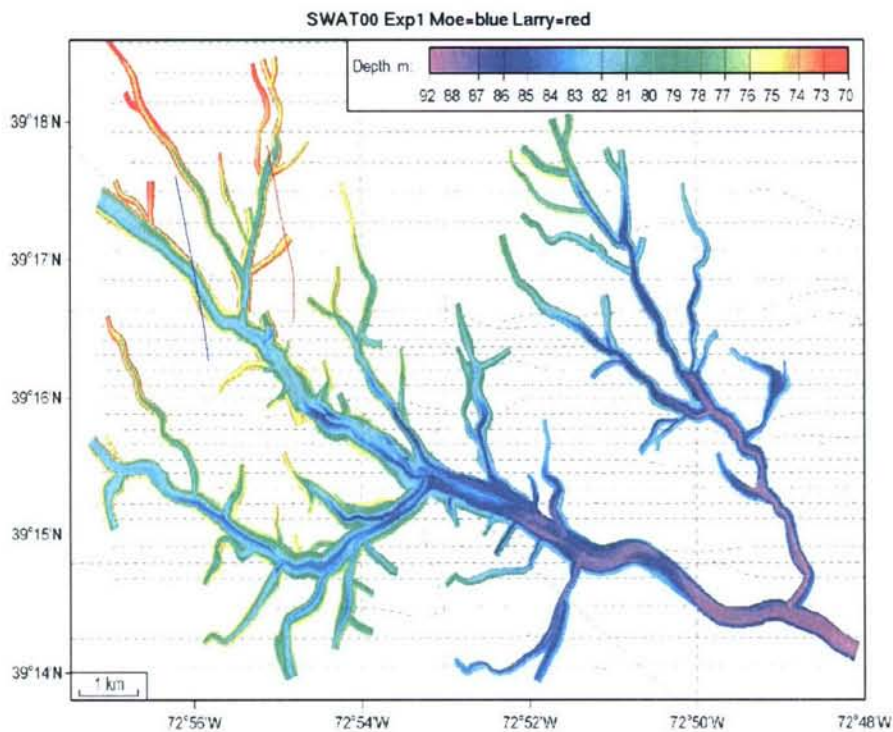

Figure 3.12 Area of the proposed experiment for SW06. The color bar shows the depth of buried river beds below the air-sea interface. The dotted lines are ship tracks during a previous experiment that located river beds using chirp sonar techniques. The blue and red lines in the top left of the figure are buoy drift tracks from the SWAT00 experiment.

To excite the field, a 125 Hz source will be suspended from a drifting surface ship, and lowered to a depth of 60m. Five temperature sensors will also be deployed along the cable, and a CTD will be attached to the source. This will ensure that the SSP

at the source position is also known. Because measurements of the SSP at this location at different times have shown variability at all depths, it is not possible to concentrate all temperature sensors at the expected thermocline depth and still expect a good estimate of the SSP. Of particular interest is a high-temperature "foot of the front" feature near the bottom which is often, but not always, seen. Lastly, CTD casts will be taken from the drifting ship at one-hour intervals to further improve our knowledge of the source-position SSP.

As the ship drifts away from it, the VLA will take measurements of the acoustic field which can be used as input to the EKF mode amplitude estimator described in the second half of chapter 2. These range-dependent estimates of the mode amplitudes can then be passed to the inversion algorithm described in the first part of chapter 2. It is hoped that the inversion will be able to detect the locations of the buried river channels by inverting for the sound speed in the bottom as a function of range and depth.

In order to test the feasibility of this design, a synthetic experiment was carried out. A waveguide was generated such that a 100 Hz source would start 1000m from the VLA, and drift to 3000m, crossing a buried river bed at 2000m range. The proper source frequency (125 Hz) was not used, because at the time the synthetic experiment was carried out, it was assumed the experiment would be using a 100 Hz source. This change is not expected to significantly changes results. The bottom model is depicted in Figure 3.13, and is based on a model provided by John Goff [34].
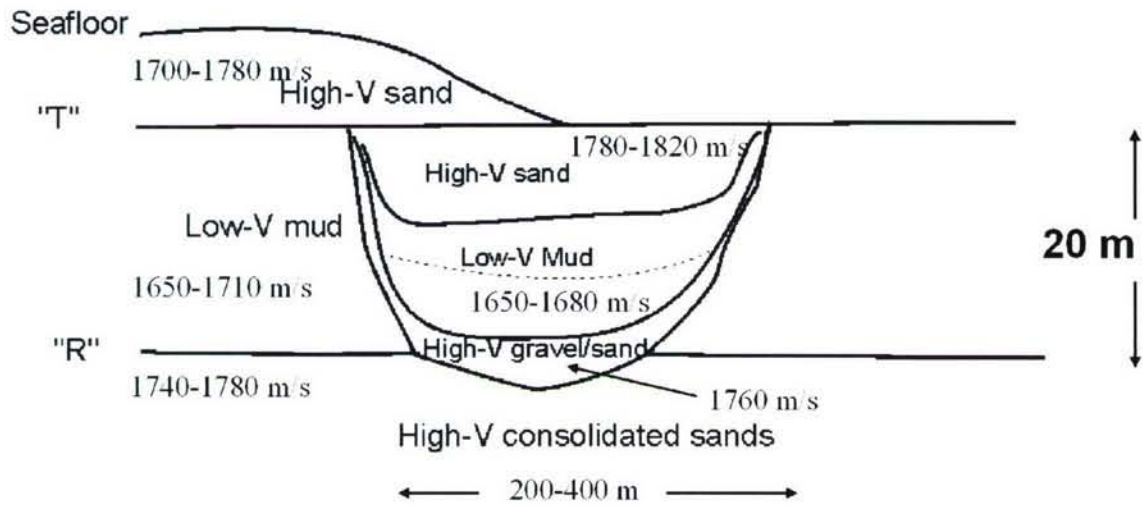
Figure 3.13 A conceptual drawing of a river cross section.

In the synthetic waveguide, the width of the river channel was set to 400m, symmetric about its center, which was placed at the 1000m point in the waveguide. The low-V mud layer was modeled with a linearly increasing sound speed (1625 m/s at interface and a 2.5 1/s gradient), and constant density (1600 kg/m^3). The high-V consolidated sands were modeled with a constant sound speed (1780 m/s) and density (1800 kg/m^3). The river channel itself was modeled with a two layer model, both layers having a linearly increasing sound speed and constant density, but the top layer (representing high-V sand) being much faster (1780m/s at the interface, with a 3 1/s gradient) than the lower layer (1650 m/s at the interface with a 1.5 1/s gradient) representing low V mud. The water column was modeled with two iso-velocity layers with a thermocline. The top layer had a sound speed of 1505 m/s from the free surface to 20m depth, and the lower layer had a sound speed of 1495 m/s from 30m depth and the seafloor. The thermocline was represented with a linear gradient connecting the sound

59

speeds of the layers over the 10 meters between them.   Figure 3.14 shows the waveguide
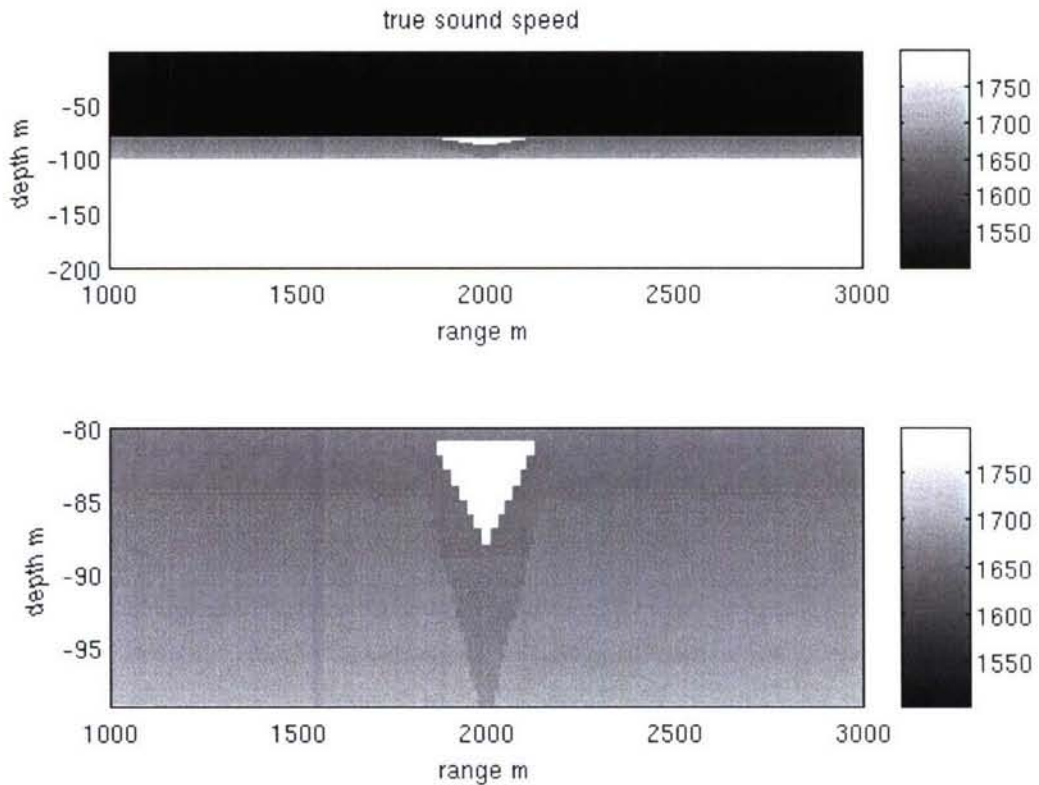
sound speeds, and that of just the sediment.

Figure 3.14: Sound speed in the synthetic waveguide, and sediment layer. The top figure shows the full 200m waveguide, including both the water column (blue) and the subbottom (red). The sediment layer can be seen between them, and the river channel in the center. The lower figure shows only the sediment layer, to more clearly display the river channel geometry.

An acoustic field was generated for the waveguide, assuming a fixed VLA and a

moving source passing through the waveguide. It was assumed that the source started

1000m from the VLA and drifted to 3000m away. The array was modeled with 12

elements, starting at 9m depth and having 6m spacing. This is more elements than the

array to be used in the experiment, and the difference was unintentional. However, the

results of other synthetic experiments did not vary significantly with the number of

elements in the VLA, so it is not thought using 8 elements would significantly alter the

results. The bottom between the VLA and the first source position is treated as range-independent, with profile equal to the profile at the first source position. The adiabatic approximation was used to generate the range-dependent field. Thirteen modes were used to generate the field, though those higher than mode 7 had very little energy in the water column, and would likely not be detected in a real experiment. Gaussian noise was added to the field to represent measurement errors. The standard deviation of this noise was set to 17% of the mean absolute value of the noise-free pressure measured along the VLA at the first range. Figure 3.15 shows an example of the noisy synthetic data at one depth.
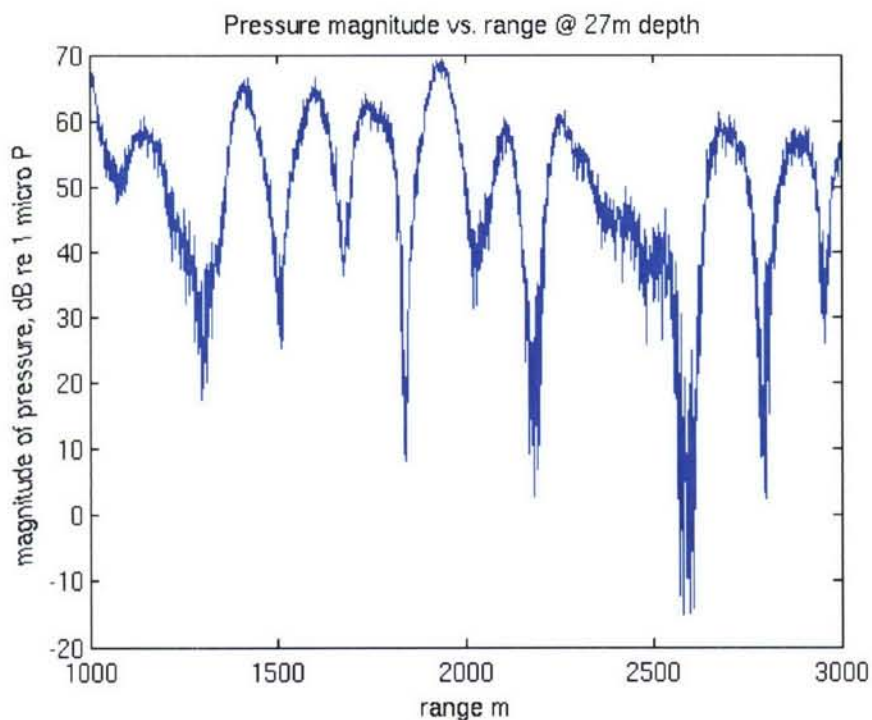

Figure 3.15: Log magnitude of the noisy pressure data as a function of range for the receiver at 27m depth, assuming a 170 dB re 1 micro P @ 1 m source. The noise levels visible are comparable to those seen in real world experiments such as MOMAX II.

The first step of the process is to pass the noisy pressure data to the EKF mode amplitude estimator. Using the known water column SSP at the VLA position, and initial

estimates of mode amplitudes and eigevalues, the EKF was able to track the amplitudes

and eigenvalues as the source moved over the river channel. The range step used was

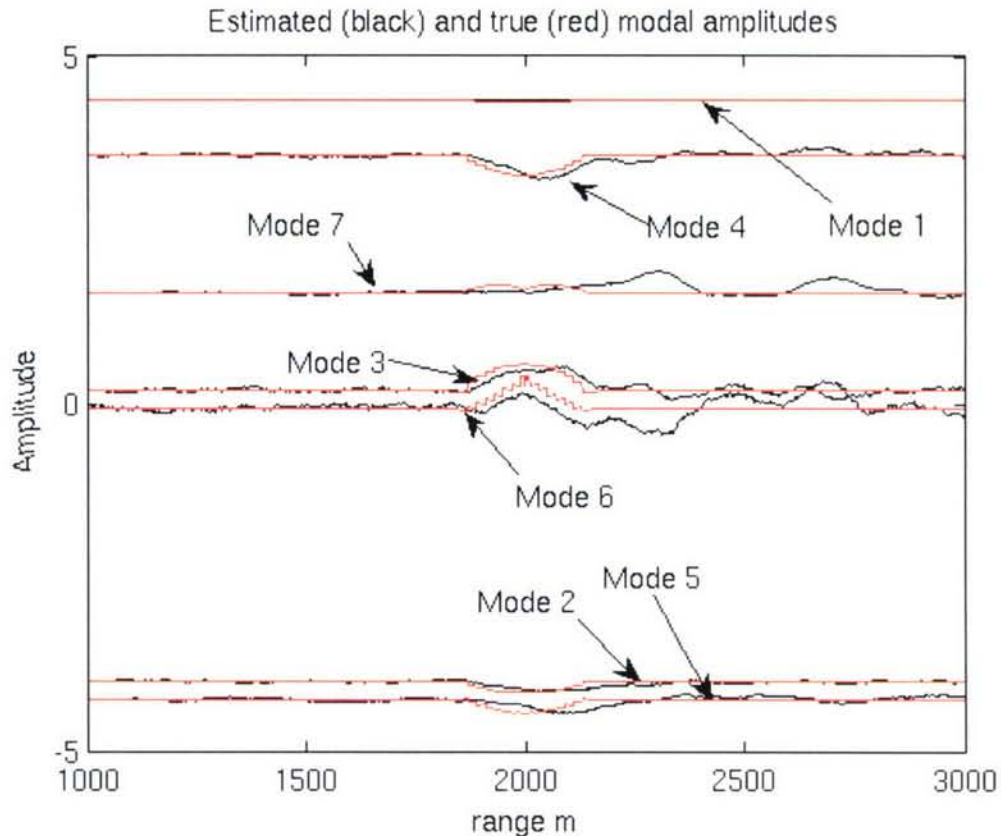1m. Figure 3.16 shows the output of the mode amplitude estimator.

Figure 3.16 EKF mode amplitude estimation of the amplitudes of modes one through 8. Estimates are in black, and the true values are shown in red. There is a clear change in the amplitudes as the source moves over the river channel at 2000m, which is detected and tracked by the EKF.

The mode amplitude estimator does an excellent job of tracking the amplitudes

while the bottom is unchanging. When the source reaches the river channel centered at

2000m, the amplitudes of many of the modes change. These changes are tracked by the

EKF. The estimates lag the true amplitudes slightly because the filter has to react to the

data in sequence, and at first treats the changes as being due to noise. After the source is

past the river channel, and back into a range-independent region, the filter does not do as good a job at tracking the low-amplitude modes, but mostly returns to the correct values.
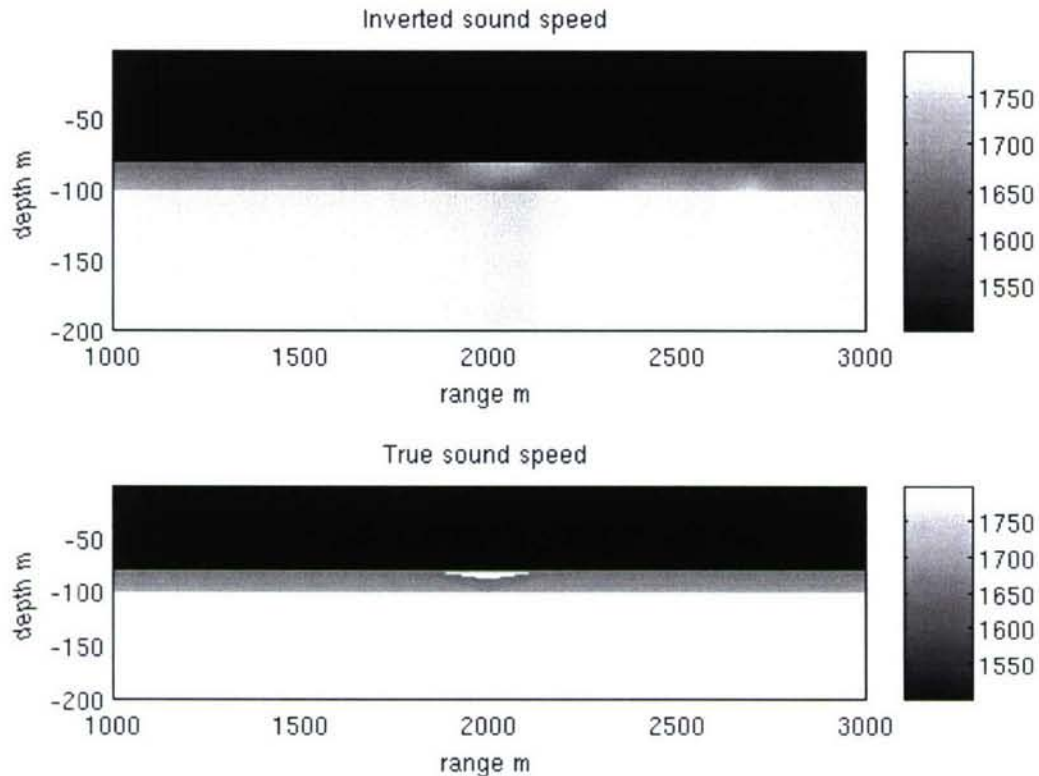
Inverted sound speed

True sound speed

Figure 3.17: The inverted sound speeds for the waveguide. The top portion shows the inverted sound velocity as a function of range and depth, while the bottom shows the true values. Note the high sound speed in the center of the top figure, indicating the detection of the buried river channel. The dark zones in the top figure are errors due to the errors in the estimator output, as can be seen in Figure 3.16, mode 7.

The estimates of the modal amplitudes were then passed to the inversion algorithm described in chapter 2. The inputs were the modal amplitudes and un-normalized mode functions given by the EKF estimator, and an estimate of the background waveguide. The bottom was parameterized with three parameters: an interface speed, a gradient in the mud layer, and a sound speed for the subbottom. It should be noted that this parameterization does not match the "true" waveguide that the algorithm is asked to invert. It works for the range-independent part of the waveguide,

but cannot correctly reproduce the SSP in the river channel. However, as can be seen in figures 3.17 and 3.18, the algorithm is still able to detect the river channel despite this short-coming.
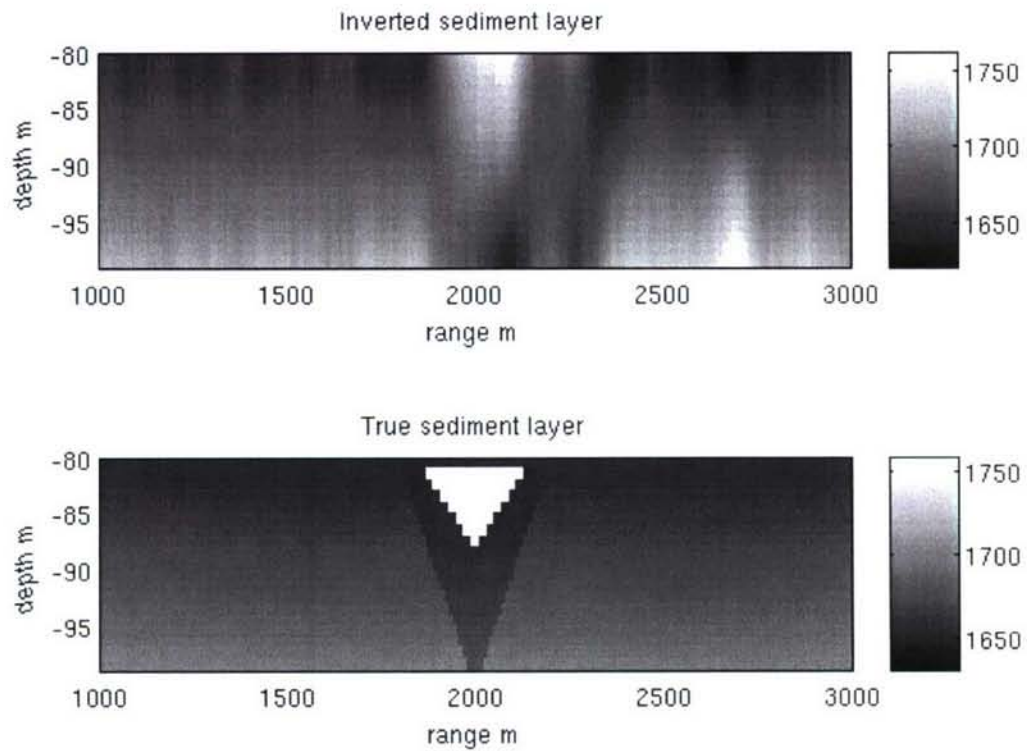

Figure 3.18: Inverted sound speeds in the sediment layer. The top portion shows the inversion result for the sediment layer only. The bottom portion shows the true values. Even though the parameterization is such that the inversions cannot accurately model the river channel, the presence of the channel is still clearly detected, and can be seen near the center of the top figure. The inversion also detects the low-velocity mud below the high velocity sand, indicated by the dark blue region near the bottom of the top figure. The high velocity zone in the top figure near 1700m range is an error in the inversion due to the error in the estimator output, as can be seen in mode 7 of figure 3.16.

These figures clearly show that the algorithm has detected the buried river channel. While the inverted sound speeds do not match the true values perfectly, especially for ranges after the river channel, there is very good agreement overall, and

values for the interface sound speed match the correct values quite well, as seen in figure 3.19.
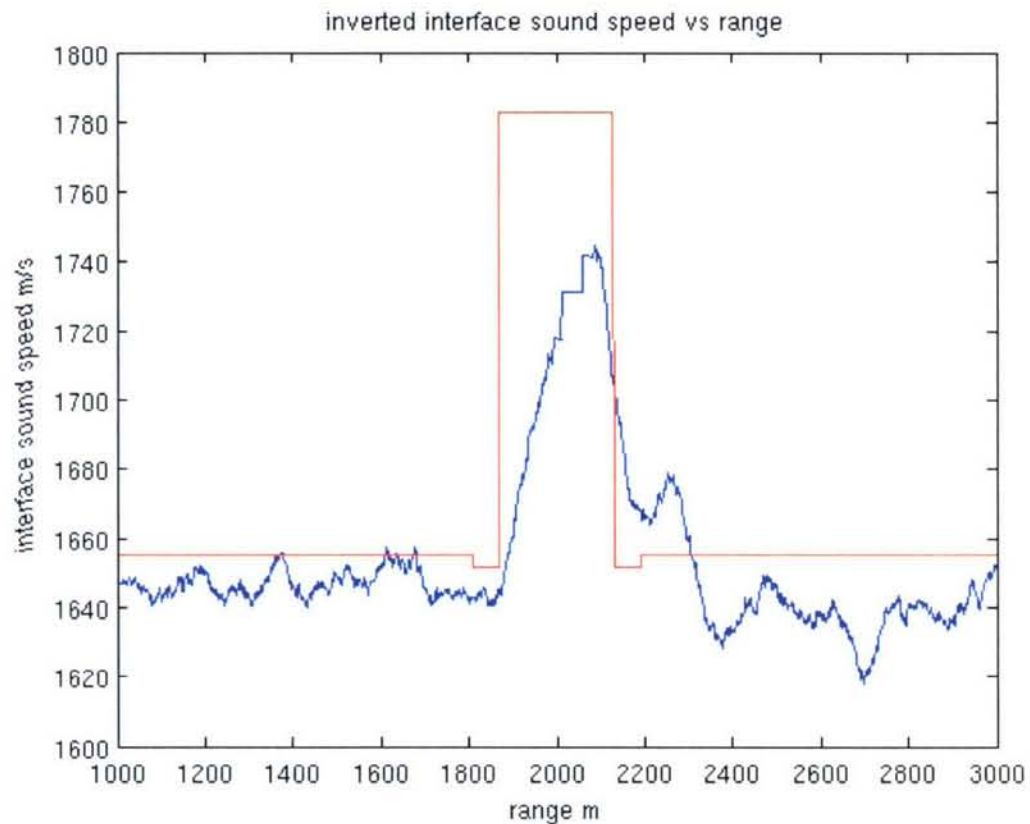

Figure 3.19: Inverted interface sound speed. The correct values of this parameter are shown in red. The river channel extends from 1800m range to 2200m.

We do notice a lag in the predicted position of the channel due to the lag in detection from the EKF, but it is not large compared to the size of the river channel. We also notice poorer inversion results for ranges past the channel. It is expected that given a longer range, the EKF would reach the correct values for the amplitudes, and the inversion would be as good as it was for shorter ranges.

This synthetic experiment demonstrates the feasibility of the mode amplitude perturbative inversion algorithm to produce a range-dependent inversion showing small-scale features such as buried riverbeds. In this case the size of the feature is 400m long,

and up to 20m deep, and the wavelength of sound in the water is about 15m. However, the high-velocity portion of the feature is only 300m wide and 0 to 10m deep. Using traditional Hankel transform methods, it would be difficult to resolve such small features due to the necessity of creating a wide enough window to resolve the individual modes. Lastly, significant noise was added to the pressure measurements in this synthetic experiment, showing that the EKF estimator is robust enough to handle real world data.

Chapter 4: Error Analysis

In this chapter we will look at the uncertainty in our estimate of the bottom parameters by analyzing the errors that propagate through the inversion process. Errors in the inversion output can arise from a number of sources, the most obvious source being errors in the input data. Another source of error is model mismatch, which occurs when our parameterization of the bottom does not match reality. Related to this is the problem of solving underdetermined problems and the assumptions needed to produce a unique solution.

We will start by investigating the simplest source of error in the inversion: errors caused by imperfect estimation of the input data. Because the inversion step is solved in using a linear approximation, errors in the input data propagate through the inversion in an easily predicted manner. Thus, if we have an estimate of the uncertainty in the input data, we can get an estimate of the uncertainty in the output. Conveniently, the EKF method for obtaining the input data estimates the variance of the estimate along with the estimate of the input data itself, so no additional measurements or estimates are needed to predict the first two moments of the error statistics in the inversion result.

Recall that in chapter 2 we discussed that the equation we are solving is of the form $\left[\dfrac{\partial m}{\partial X} \right] \bar{X} = \Delta \bar{m}$, where $\left[\dfrac{\partial m}{\partial X} \right]$ is a matrix containing the derivative of the mode ratios at all the depths with respect to all the parameters sought, $\bar{X}$ is a vector of the parameters sought, and $\Delta \bar{m}$ a vector that contains the difference between our predicted mode ratios

and the ratios we actually measure. We use the pseudo inverse to invert the derivative

matrix, and solve for the parameter vector: $\vec{X}_{LS} = \left[\dfrac{\partial \tilde{m}}{\partial X}^{\#}\right]\Delta\vec{m}$.

If our estimate of the mode ratios has errors, our equation becomes

$$\left[\frac{\partial m}{\partial X}\right]\vec{X} = \Delta\vec{m} + \delta\vec{m} ,$$

where the vector $\delta\vec{m}$ contains the error in our estimates of the mode ratios, the variance

of which can be determined from our output of the EKF. If we try to invert this

equation, we will end up with errors in our parameter vector:

$$\vec{X}_{LS} + \delta\vec{X}_{LS} = \left[\frac{\partial \tilde{m}}{\partial X}^{\#}\right]\Delta\vec{m} + \left[\frac{\partial \tilde{m}}{\partial X}^{\#}\right]\delta\vec{m} \Rightarrow \delta\vec{X}_{LS} = \left[\frac{\partial \tilde{m}}{\partial X}^{\#}\right]\delta\vec{m} .$$

It is clear that the error in our bottom estimate will be a linear function of the errors in our

input data, allowing us to easily compute the error statistics of our inversion.

We can confirm this with a simple test case. We make use of the test case in

chapter 3 which involved adding random errors to the estimates of the modal amplitudes

in a range-dependent waveguide. We recall that the waveguide contained a 50m water

column above a two-parameter bottom with constant, known density. The two

parameters in the bottom (interface sound speed and gradient) varied with range,

following a random walk. The mode functions were generated for each range, and the

true mode ratios computed. Random white, Gaussian noise was added to these mode

ratios, and the results used as input to the inversion algorithm. Because of the noise in

the inputs, the outputs of the algorithm did not match the true values of the bottom

parameters. The standard deviation of this difference from the true values as a function

of the standard deviation of the input is shown in figure 4.1. The results match our prediction of a linear relationship quite well.
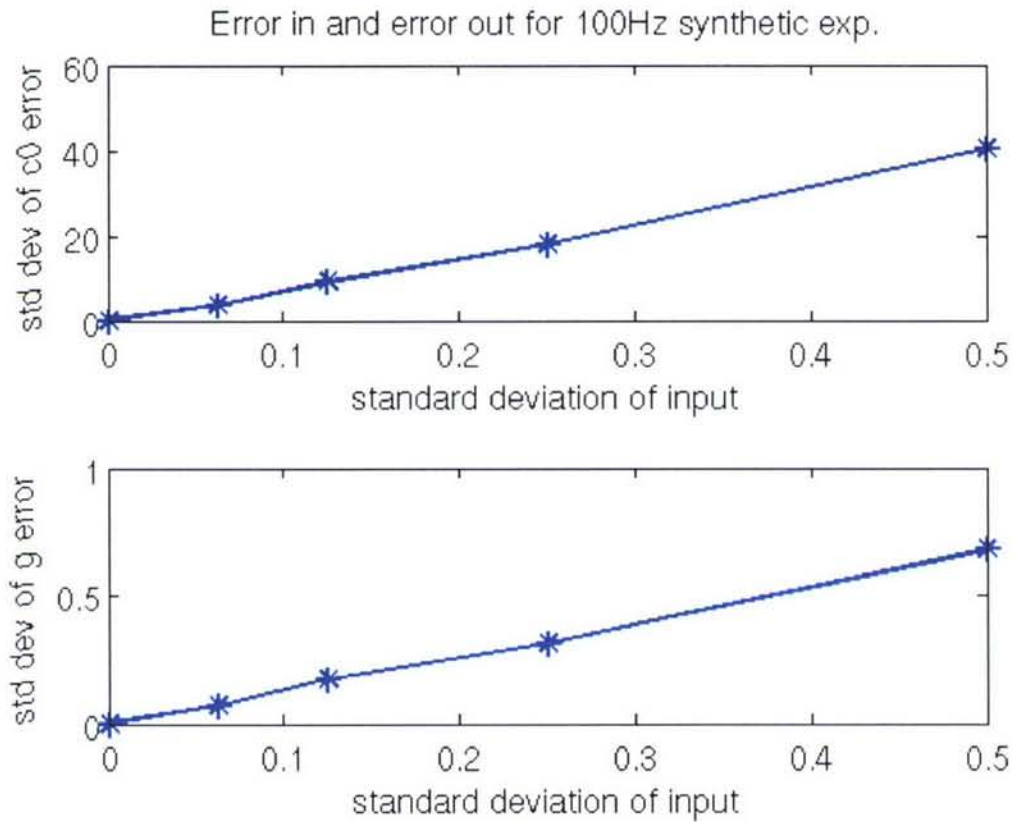

Figure 4.1 The standard deviation of the error in the estimate of the two parameters as a function of the standard deviation in the input mode ratios. The top plot shows the standard deviation of the error in interface sound speed, the lower plot shows the standard deviation of the error in the gradient estimate. In both cases error in input and error in output are linearly related.

It should be pointed out at this point that the assumption of zero-mean, Gaussian noise in the mode ratios is not strictly valid given an assumption of zero-mean, Gaussian noise in our mode amplitude parameters A_n. However, if the error in the estimate of the first mode is much smaller than the first mode itself, the approximation to a Gaussian distribution is good. This can be seen by examining the quantities in our EKF, and how

they relate to the quantities in our inversion. We form the mode ratios for our inversion using the mode amplitudes and unnormalized mode functions from the EKF:

$$m_n(z, z_s) = \frac{A_n \zeta(z, k_n)}{A_1 \zeta(z, k_1)} \, .$$

If we add errors in our estimates of the amplitudes and eigenvalues, we get the result that

$$m_n(z, z_s) + \delta m_n(z, z_s) \approx \frac{(A_n + \delta A_n)\left(\zeta(z, k_n) + \delta k_n \dfrac{\partial \zeta(z, k_n)}{\partial k_n}\right)}{(A_1 + \delta A_1)\left(\zeta(z, k_1) + \delta k_1 \dfrac{\partial \zeta(z, k_1)}{\partial k_1}\right)} \, .$$

We have assumed that the change in the unnormalized mode function due to the error in the estimate of the wavenumber can be computed linearly. Since we are assuming the changes due to the bottom properties are also linear, this should be a good assumption. If our errors are large enough to cause non-linear changes in our measured quantities, we will have no hope of inverting the linear (and thus much smaller) changes due to bottom properties. We can re-write the equation above as

$$m_n(z, z_s) + \delta m_n(z, z_s) \approx \frac{A_n \zeta(z, k_n) + \left(\delta A_n \zeta(z, k_n) + \delta k_n A_n \dfrac{\partial \zeta(z, k_n)}{\partial k_n} + \delta A_n \delta k_n \dfrac{\partial \zeta(z, k_n)}{\partial k_n}\right)}{A_1 \zeta(z, k_1) + \left(\delta A_1 \zeta(z, k_1) + \delta k_1 A_1 \dfrac{\partial \zeta(z, k_1)}{\partial k_1} + \delta A_1 \delta k_1 \dfrac{\partial \zeta(z, k_1)}{\partial k_1}\right)}$$

We can factor out the error-free term from the denominator as well:

$$m_n(z, z_s) + \delta m_n(z, z_s) \approx \frac{A_n \zeta(z, k_n) + \left(\delta A_n \zeta(z, k_n) + \delta k_n A_n \dfrac{\partial \zeta(z, k_n)}{\partial k_n} + \delta A_n \delta k_n \dfrac{\partial \zeta(z, k_n)}{\partial k_n}\right)}{A_1 \zeta(z, k_1)\left(1 + \dfrac{\delta A_1}{A_1} + \delta k_1 \dfrac{\partial \zeta(z, k_1)}{\partial k_1} \dfrac{1}{\zeta(z, k_1)} + \dfrac{\delta A_1}{A_1}\delta k_1 \dfrac{\partial \zeta(z, k_1)}{\partial k_1} \dfrac{1}{\zeta(z, k_1)}\right)}$$

We now wish to show that most of the terms in the denominator are much less than one, allowing us to make use of an approximation for the term in parentheses in the

denominator. The $\delta A_1 / A_1$ must certainly be much less than one, or we have little hope of inverting the bottom parameters at all. The first mode is the easiest to estimate, both in terms of eigenvalues and amplitude, so our error in that estimate should be much less than the estimate itself. We can also check that this is true using the output of the EKF. If this is not the case, we should not attempt the inversion. The second term is less clear, but if we keep in mind that the first mode function has no zero-crossings other than at the free surface, thus ensuring that the $1/\zeta$ does not blow up, we gain confidence that this term must be small. If the change in the unnormalized mode function due to error could be on the order of the mode function itself, our linear approximation is invalid. Thus, if the second term is not much less than one, we again should probably not be attempting the inversion. The error in the unnormalized mode function is due to the error in our estimate of the eigenvalue, so again we are able to check if that it is sufficiently small using the output of the EKF. If the first two terms are much less than one, then the third term, which is the product of the two, must also be much less than one. We can thus make use of the approximation $\dfrac{1}{1+x} \approx 1-x$, valid for $|x| \ll 1$. We then get

$$
\begin{aligned}
m_n(z,z_s) + \delta m_n(z,z_s) &\approx \frac{[A_n\zeta(z,k_n) + (termI)](1 - termII)}{A_1\zeta(z,k_1)} \\
&= \frac{A_n\zeta(z,k_n)}{A_1\zeta(z,k_1)}\left[1 + \frac{termI}{A_n\zeta(z,k_n)} - termII - \frac{(termI)(termII)}{A_n\zeta(z,k_n)}\right]
\end{aligned}
$$

where

$$
termI = \left(\delta A_n\zeta(z,k_n) + \delta k_n A_n \frac{\partial\zeta(z,k_n)}{\partial k_n} + \delta A_n \delta k_n \frac{\partial\zeta(z,k_n)}{\partial k_n}\right), \text{ and}
$$

$$
termII = \left(\frac{\delta A_1}{A_1} + \delta k_1 \frac{\partial\zeta(z,k_1)}{\partial k_1}\frac{1}{\zeta(z,k_1)} + \frac{\delta A_1}{A_1}\delta k_1 \frac{\partial\zeta(z,k_1)}{\partial k_1}\frac{1}{\zeta(z,k_1)}\right).
$$

Making use of the fact that $m_n(z, z_s) = \dfrac{A_n \zeta(z, k_n)}{A_1 \zeta(z, k_1)}$, we find that the error in the mode

ratio is

$$\delta m_n(z, z_s) \approx m_n(z, z_s) \left[\frac{termI}{A_n \zeta(z, k_n)} - termII - \frac{(termI)(termII)}{A_n \zeta(z, k_n)} \right].$$

We can further simplify the estimate of the mode ratio error by neglecting the product of

term I and term II because it involves products of quantities presumed to be small. The

result of which is:

$$\delta m_n(z, z_s) \approx m_n(z, z_s) \left(\frac{\delta A_n}{A_n} + \delta k_n \frac{\partial \zeta(z, k_n)}{\partial k_n} \frac{1}{\zeta(z, k_n)} - \frac{\delta A_1}{A_1} - \delta k_1 \frac{\partial \zeta(z, k_1)}{\partial k_1} \frac{1}{\zeta(z, k_1)} \right).$$

Thus, if all our assumptions have been valid, and the errors in the eigenvalue and

amplitude estimates are Gaussian distributed, then the error in our estimated mode ratio

will be approximately Gaussian distributed. We can then compute the covariance of our

bottom parameter estimates as described earlier:

$$\delta \vec{X}_{LS} = \left[\frac{\partial \tilde{m}^\#}{\partial X} \right] \delta \tilde{m} \Rightarrow \mathrm{cov}(\delta \vec{X}_{LS}) = \left[\frac{\partial \tilde{m}^\#}{\partial X} \right] \mathrm{cov}(\delta \tilde{m}) \left[\frac{\partial \tilde{m}^\#}{\partial X} \right]^T$$

Of course, all this assumes that our parameterization allows us to accurately

describe the bottom in the first place. It is quite possible that our parameterization will

not allow us to capture certain details of the bottom profile, leaving us with errors no

matter how good our input data are. Because of this, our selection of the

parameterization significantly affects the quality of our inversion. The ideal

parameterization is the one containing the least number of parameters that can capture all

desired features of the bottom and reproduces the measured acoustic field. Of course,

since we don't know all the features of the bottom *a priori*, we must do some guess work.

On the one hand, we can try a simple parameterization with very few parameters, and run the risk of not being able to invert important features of the bottom. On the other hand, we can treat the sound speed at each depth as a separate parameter, and guarantee our ability to model the true bottom, at the expense of having to solve an underdetermined problem.
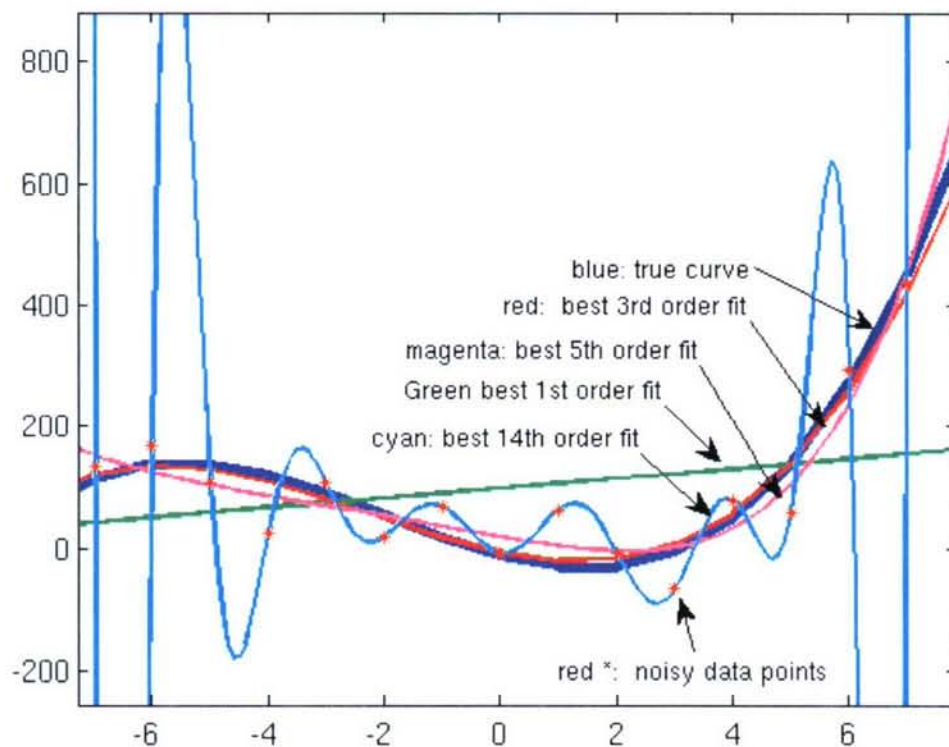

Figure 4.2 Fitting noisy data with a polynomials of various order. The fifteen data points (red asterisks) were generated by adding zero-mean Gaussian noise with a standard deviation of 10 to samples of a 3rd order polynomial (thick blue curve). The other curves are the best-fit polynomials of various orders for the data points. Notice that as the order increases past 3rd order, the fit to the points improves, but the fit to the "true" curve gets worse. With a 14th order polynomial, it is possible to perfectly fit the data points, but the resulting polynomial is a very poor fit to the true 3rd order curve.

The problem of parameterization is somewhat similar to the problem of fitting a curve to a number of data points. If we wish to fit a polynomial to the points, we have to decide which order of polynomial to use. The higher order polynomial we use, the better

our fit to the data will be, up to the point of using a polynomial of order equal to the number of points, at which we will have a perfect fit to all points. However, while high-order polynomials might fit the given points well, they will usually do a poor job of fitting the curve between the points if there is any measurement error. An example of this is shown in figure 4.2.

When we increase the number of parameters past the number of mode ratios, the problem becomes underdetermined, and we would be able to fit the input data exactly if it were a truly linear problem. However, since there will be errors in the input data, fitting them perfectly won't give the best inversion result. This is akin to using a polynomial of order equal to the number of data points in figure 4.2. The cyan curve fits the points perfectly, but it doesn't give a good approximation of the true curve. The analogy here is not perfect, but it is instructive. If we continue it a little further, we notice that using too low an order of polynomial is potentially just as bad as using too high an order. It is simply not possible to capture all the features of the true curve using too low an order of polynomial. In our analogy to estimating bottom parameters, this is akin to using too simple of a parameterization, and not being able to invert for all the features of the bottom.

Leaving the polynomial analogy for a more concrete example, we will now show an example of inversions under three possible parameterizations. The true bottom will consist of two layers and a sub-bottom. Five parameters are sufficient to exactly parameterize the bottom (an interface speed and gradient for each layer, and a speed for the sub-bottom). One of our three parameterizations will be the correct parameterization. Another will be under-parameterized, with only 3 parameters (interface speed and

gradient for the entire sediment thickness and a sub-bottom speed), and the last will be an over-parameterized c-at-every-depth model. Over the bottom we place a 100 m iso-velocity water column, and assume perfect measurement of the mode functions on a 17 element VLA within it. These mode functions are used to invert for the bottom in the three parameterizations given above. We assume perfect measurement of the mode functions so as to separate the errors due to parameterization from the errors due to measurement noise examined above. Figure 4.3 shows the results of the three inversions. As we would expect, the under-parameterized model is unable to capture all the relevant details of the profile. It effectively produces an averaging of the sediment layers. The five-parameter model is able to perfectly reproduce the bottom. This is not surprising as this is the ideal parameterization. Lastly, the over-parameterized model does a decent job of capturing the layered features, but also inserts oscillations into the profile that are not present in the true profile.

The second plot of Figure 4.3 illustrates the problems of using an under-parameterized bottom model. The inverted profile doesn't match the true SSP in the sediment layers because the parameterization used simply does not make this possible. Even with perfect input data, we cannot hope to detect all the features of the bottom in such a case. How much of a problem this is depends upon the motivation for the inversion. If the goal is merely to obtain a bottom model that allows for accurate predictions of acoustic propagation in the region, a simple bottom model that produces the correct modal parameters will suffice. On the other hand, if the inversion is being carried out to detect specific bottom features, such as buried river beds, a very simple model will not suffice. If this is the case a more complicated bottom model will be

required. It may be useful, however, to start with a simple bottom parameterization, and use the inversion result as the initial background for a more complicated model.
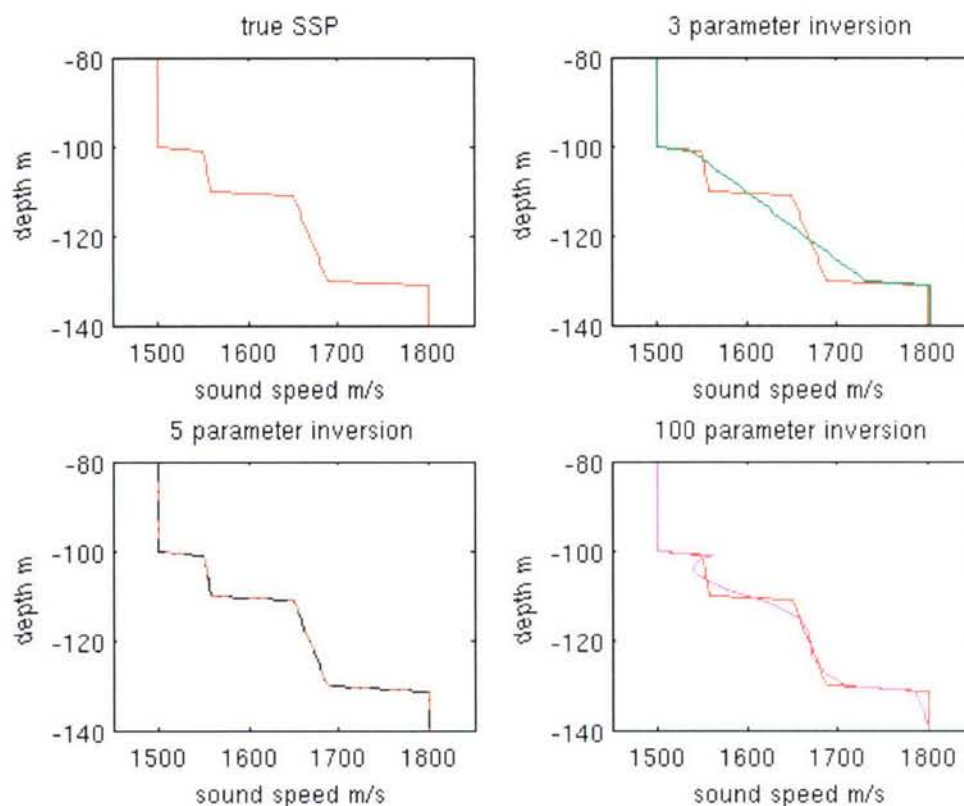

Figure 4.3 Inversions of a 2-layer bottom using different parameterizations. The top left figure contains the true profile. The top right figure contains the true profile, and the inversion result using a under-parameterized 3 parameter model. Note that the layered structure cannot be determined. The bottom left figure shows a perfectly parameterized 5 parameter model. The inversion matches the true background perfectly. Last, in the bottom right, an over-parameterized c-at-every-depth model has been used. The layered structure has been detected, but additional oscillations due to over-parameterization are visible as well.

The fourth plot of Figure 4.3 shows that even with perfect input data, we often cannot achieve a prefect match to the true profile with a highly over-parameterized bottom model. Part of the problem is that we are attempting to solve an underdetermined problem. Underdetermined problems in linear algebra generally have an infinite number of solutions. Our use of the pseudo-inverse selects one of the infinite number of

solutions, but does so using assumptions about the norm of our solution, which need not be true in reality. This makes it very difficult to estimate the error in the inversion, as our result might fit the data (*i.e.*, the m_n ratios) exactly, but still not match the true SSP.

Another way of thinking of this situation involves the idea of the null space of the matrix $\left[\dfrac{\partial m}{\partial X}\right]$. The null space is the set of vectors which give a zero-vector as the product when they are multiplied by the matrix. For example, in the problem $[A]\vec{x} = \vec{b}$ where $[A]$ is *NxM* and *M>N* there will exist vectors $\vec{y}$ such that $[A]\vec{y} = \vec{0}$. By adding a multiple of one of these vectors to $\vec{x}$ we don't change the product: $[A](\vec{x} + \vec{y}) = [A]\vec{x} = \vec{b}$. When we try to solve for the vector $\vec{x}$, we may instead end up with a result of $(\vec{x} + \vec{y})$. When we use the pseudo-inverse we make the assumption that the vector we are looking for is the vector with the minimum norm out of all the vectors that solve the equation. However, this assumption is made because it allows us to easily find a unique solution, not because it is likely to be true in the physical world. Because of this, when we solve the underdetermined problem our answer will often have extra information in it, beyond the true SSP. This accounts for the difference between the true SSP in the fourth plot of Figure 4.3 and the inverted SSP. These errors are difficult to quantify because any scalar multiple of any member of the null space could be added to our inversion result without changing the measured quantity (the m_n ratios). Or, more precisely, this would be the case if the problem were truly linear. At some point large changes in the SSP due to addition of null space vectors will cause the linear assumption to fail, and result in changes to the m_n ratios. However, if we think it possible that errors this large are present, we cannot trust our inversion results at all. The thing to keep in mind when

dealing with underdetermined problems is that true SSP may be the inversion result plus a weighted sum of the vectors from the null space of the $\left[\dfrac{\partial m}{\partial X}\right]$ matrix. Without making further assumptions about the bottom, it will not be possible to determine the weights in the sum, however.

Rajan *et al.* [1] and Souza [3] have addressed errors in such under-determined problems using the quantities resolution and variance. The variance is due to errors in the input data, as described above. The resolution is due to approximating an integral over depth with a weighted sum. Neither of these, however, addresses the problem of null space vectors and the underdeterminedness of the problem. Further, both of these measurements are only valid if the algorithm converges to the global minimum of the cost function, and not just a local minimum.

Convergence to the incorrect minimum of the least squares cost function is a problem faced by both over-determined and undetermined problems. In order for the process to converge to the best answer (in the least squares sense), the initial background must be near the true SSP. Just how close the background needs to resemble the true SSP depends on the parameterization and background used. In general, the fewer parameters we use, the less the background needs to resemble the true SSP. Additionally, when fewer parameters are used, the more likely a convergence to the wrong minimum is to be recognized as being non-physical. Figure 4.4 shows examples of convergence to the wrong minimum using the same parameterizations used in figure 4.3. When the various parameterizations are started with a Pekeris background model, only the 3-parameter model converges to the correct answer (or as correct a result as the 3-parameter model can give). Both the 5-parameter model and the c-at-each-depth model converge to non-

physical SSPs. However, if these parameterizations are used starting with the output of

the 3-parameter model as the background SSP, they converge to the correct result.
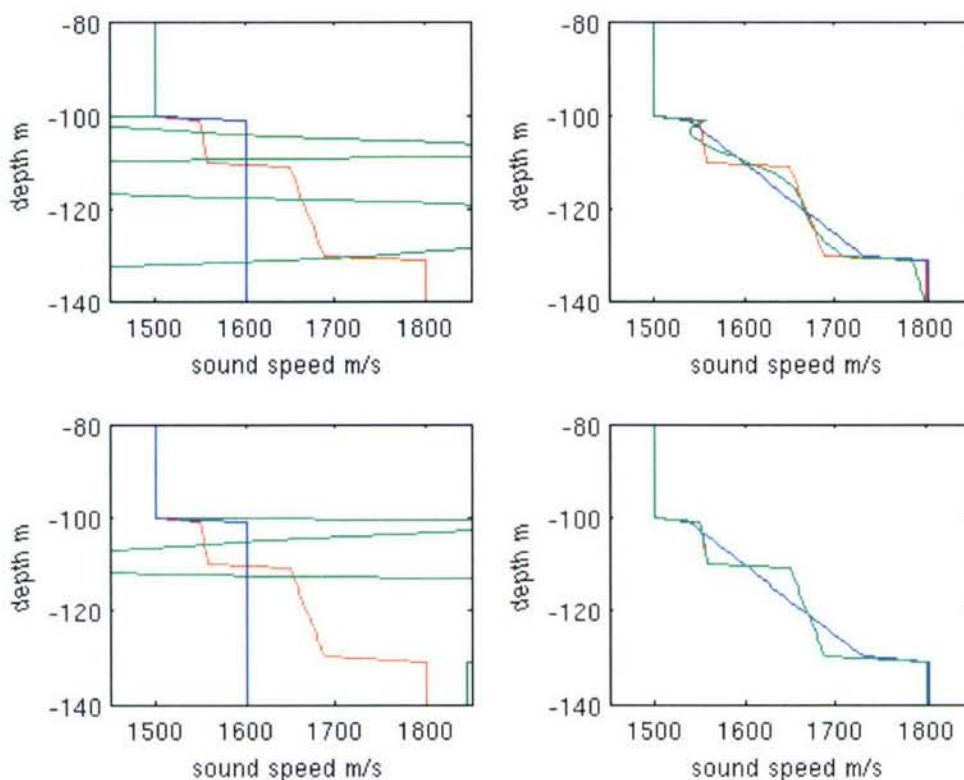

Figure 4.4 Converging to different answers when using different backgrounds. In each plot the red curve is the true SSP, the blue curve is the background used, and the green curve is the result of the inversion after convergence. The top left plot shows the result of the c-at-each-depth parameterization when starting with a Pekeris background. Note the non-physical result shown in green. The top left is the result of the same parameterization when the output of the 3-parameter inversions is used as the background model. In this case the algorithm converges to a result very close to the true profile. Similarly in the bottom two plots we see the results for the 5-parameter inversion using the Pekeris background and the 3-parameter background. Again using the Pekeris background produces a non-physical result, but using the 3-parameter background results in perfect agreement. Note that the 3-parameter inversion converged to its result after starting with the Pekeris background.

Figure 4.4 suggests a tactic of starting with a simple parameterization and adding

more complexity to it after it converges. Such a technique will make it possible to

capture more features of the SSP while reducing the risk of convergence to incorrect

minima. Another example of the usefulness of this concept is shown in figures 4.5-4.7.

The true bottom for which we are trying to invert is based on an inversion by Ohta *et al.* from the SWAT [35] experiment. It consists of a nearly linearly increasing sound speed with depth, but contains a sound speed minimum just below the water-bottom interface. The water column was modeled as 80m of isovelocity water. A 50 Hz sound source was used to excite the waveguide, and the field measured on an 8-element VLA. In order to separate noise effects from parameterization effects, it was assumed the mode functions were perfectly measured at the VLA. As can be seen in Figure 4.5, when the c-at-each-depth parameterization was used with a Pekeris background, the algorithm converged to an incorrect answer. However, when a 2-parameter model is used, the algorithm converges to good approximation of the true bottom, as seen in Figure 4.6. However, due to the simplicity of the parameterization, the inversion doesn't show the sound speed minimum in the bottom. Figure 4.7 shows the results of using the c-at-each-depth model starting with the result of the 2-parameter model as the background. Now that the background matches the true profile well, we get a very good inversion that captures the important detail of the sound speed minimum.

Another thing that must be kept in mind when parameterizing the bottom is the independence (or lack thereof) of the parameters. Zanolin *et al.* [36] have shown that in some cases certain parameterizations require an unreasonable number of data samples in order for the estimate to achieve the Cramer-Rao lower bound. The Cramer-Rao bound is the lower bound on the variance of an unbiased estimator, and is computed from the Fischer information matrix. No unbiased estimator can achieve a lower variance than the CRLB, and in some cases no estimator will exists which achieves the bound. Further, even when the Cramer-Rao lower bound is met, it may not be low enough to meet desired

levels of uncertainty. This is due both to the non-linear relation between the parameters and the data, and coupling of certain parameters. For example, it becomes far more difficult to estimate the interface speed of a layer efficiently when the gradient, or layer thickness are unknown as well, especially when the layer thickness is close to a single wavelength of sound in the layer.
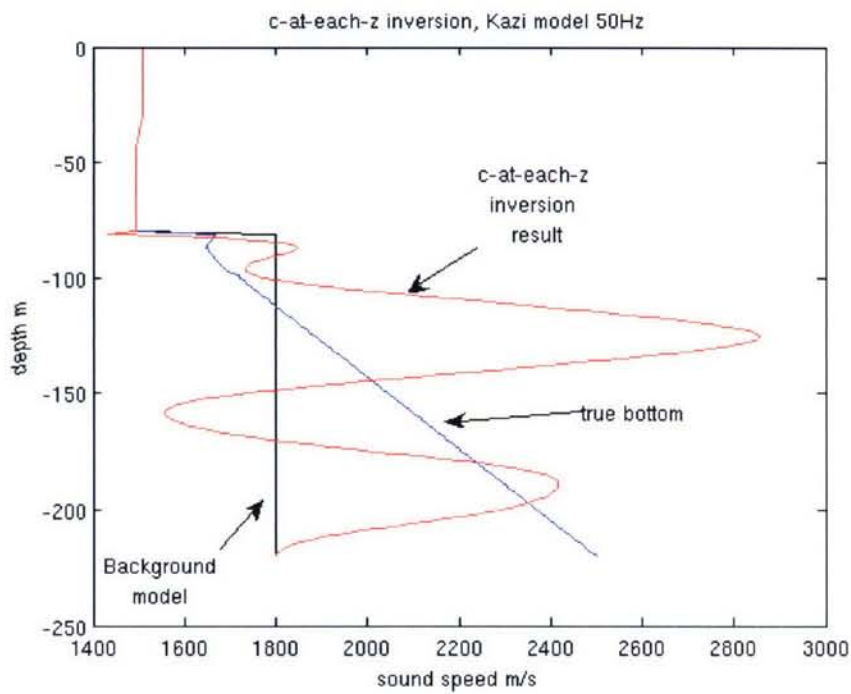

Figure 4.5 A c-at-each-depth inversion for the "Kazi" profile of Ohta, *et al.*. When the initial background is not close to the true profile, the algorithm often converges to an incorrect answer, as shown here. Highly over-parameterized bottom models are especially prone to this.
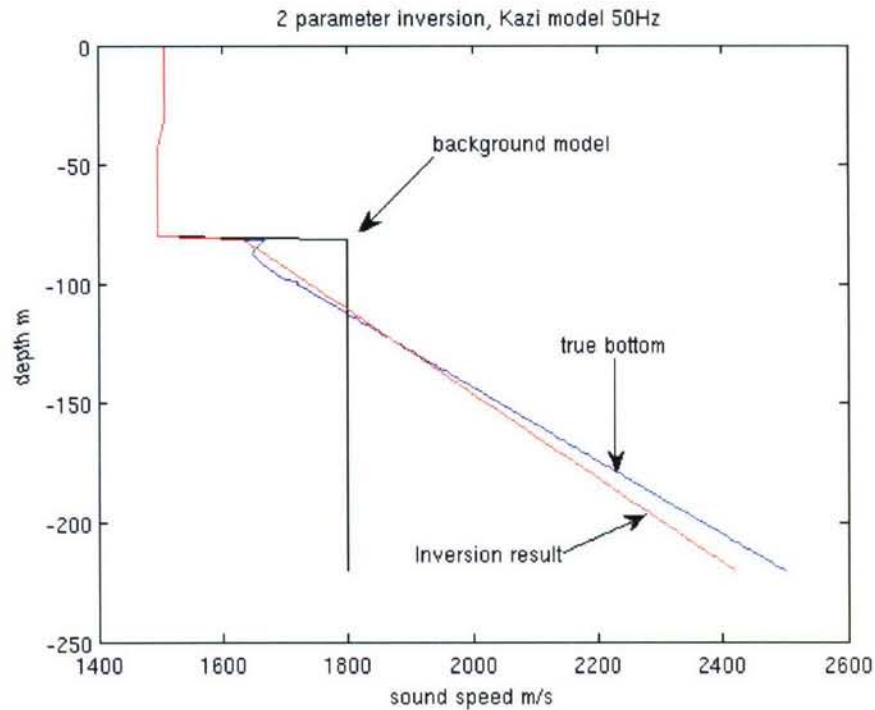
Figrue 4.6 A two-parameter inversion for the "Kazi" profile. Due to the simplicity of this parameterization, it is likely to converge to a good result even when the initial profile is far from the true profile. However, it is unable to capture all the details of the true profile, such as the sound speed minimum just below the water-bottom interface.

Yet another source of potential problems is the number of the singular values used when calculating the pseudo inverse matrix. When the number of parameters is small, the inversion process usually remains stable. When the number of parameters is large, however, it often becomes unstable if singular values smaller than 1% of the largest singular value are used. This effect is usually detectable, however, by varying the tolerance for the size of singular values accepted, and watching for very large changes in the inverted SSP. As a rule of thumb, we have usually started highly over-parameterized inversion at a tolerance level of 10% of the largest singular value, and decreased to 1% after the initial convergence. After each convergence the tolerated size is reduced until a

large, non-physical change in the inverted SSP is observed. This works best when combined with the tactic of starting with a simple parameterization.
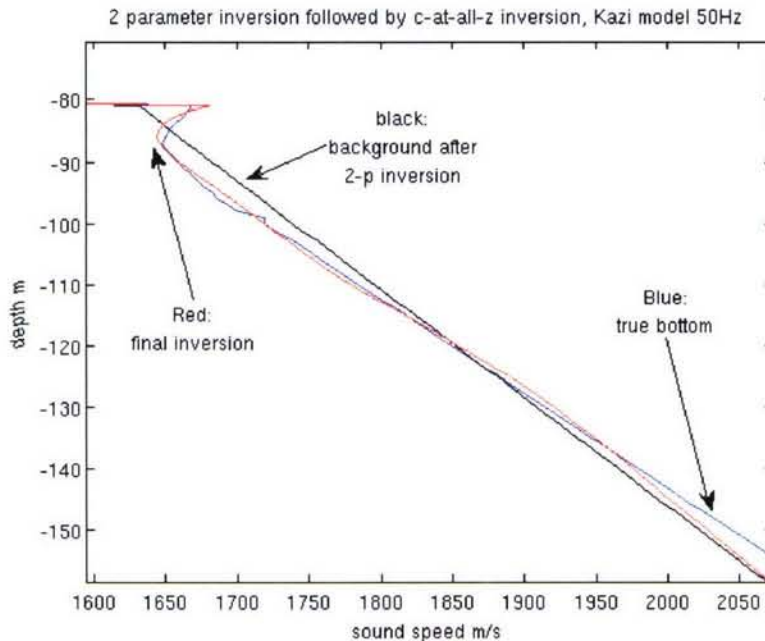

Figure 4.7 A c-at-each-depth inversion of the "Kazi" profile using a good initial background. In this case the output of the 2-paramter inversion is used as the background for the c-at-each-depth inversion, allowing the algorithm to converge to a very good approximation of the true profile Note that the sound speed minimum near the interface is visible in the inversion result.

Figures 4.8 and 4.9 show an example of this. Here we are trying to invert for a one-layer bottom over an iso-velocity sub-bottom. The water column is 100m deep and isovelocity, and the field is excited by a 125 Hz source. Again, in order to avoid confusion with errors due to noise, we treat the mode functions as perfectly measured. In figure 4.7 we see the results of the c-at-each-depth inversion starting with a Pekeris background, using two different tolerances for singular value size. Neither inversion is particularly good, but the fact that they are significantly different is telling. Figure 4.8 shows the result after having used the output of the 2-parameter inversion as the

background. In this case, the size of the singular values used is down to 10% of the largest singular value.
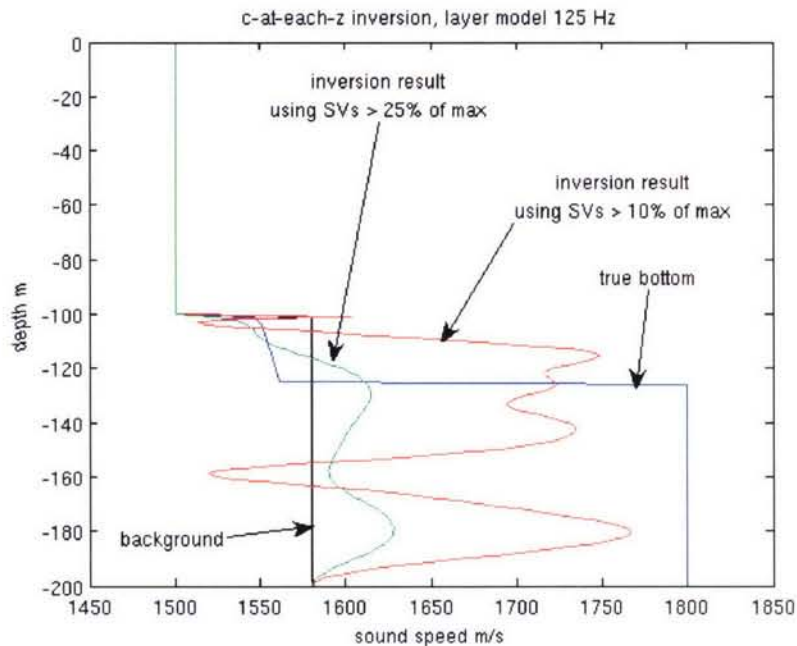

Figure 4.7 Differences in inversion due to the number of singular values used. Using more singular values changes the inversion result. If the background is a poor match the true profile, using more SVs usually results in more and larger oscillations in the inverted profile.

Because the results of the inversion depend on how well the bottom model is able to match the true profile, and on the initial background SSP, and on the number of singular values used, the inversion process often is more of an art than a science. This is especially true of over-parameterized bottom models, for which the initial background model and the number of singular values used are more likely to affect the outcome. This makes models with few parameters more attractive. However, such models are unable to capture all the features of most profiles, and very limited in the types of profiles they can produce. The user must balance these factors, and select a parameterization appropriate

for the task, and always keep in mind that the inversion process is far from a black box. Estimates of the error variance in the inversion can be computed based on estimates of the uncertainty in the input data, but this error variance does not take into account the possibilities of convergence to incorrect minima, nor problems with the stability of the inversion of the matrix.
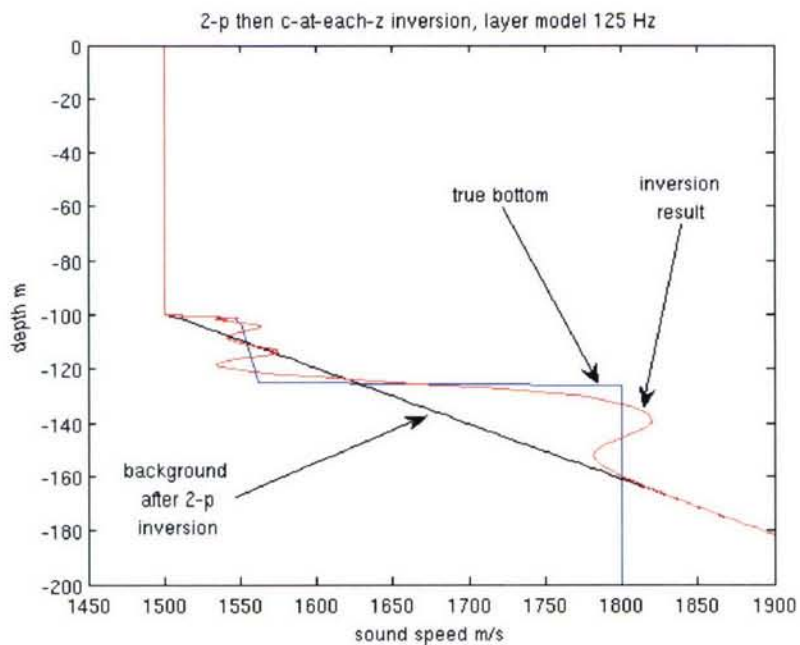

Figure 4.8 Good agreement with smaller singular values requires a better background. In this case the background is closer to the true profile, so a c-at-each-depth inversion is able to perform well. At first SV that are 25% of the largest SV or larger are tolerated. After an initial convergence, the tolerance is changed to 10% of the largest SV, and the result matches the true profile well.

Chapter 5: Combination with Eigenvalue Perturbation

When introducing a new inversion algorithm, one must address the issue of whether it is superior in anyway to existing algorithms. The mode amplitude perturbative inversion scheme introduced in chapter 2 is no exception, and in this chapter we will compare it to the Eigenvalue perturbation method of Rajan *et al.* [1]. The two methods are similar in many ways: both require a background sound speed profile, and use it to predict values for certain field parameters, and both compare these predicted values to measured values of the same quantities, and using a linear approximation, compute the correction to the background needed to bring the predicted values into agreement with the measurements. The main difference between the methods is the input data used. There is also a difference in the experiments typically used to obtain the necessary input data (*i.e.*, Hankel transform methods vs. the EKF estimator), but this is not actually a difference in the methods themselves. In fact, both the Hankel transform and the EKF estimator can provide the input data necessary to use either method. .

The two most common metrics of performance of an inversion algorithm are the bias and variance of the error in the inverted parameters. As shown in chapter 4, these quantities can be computed from the inversion matrix and an estimate of the measurement error statistics. This does require, however, that the algorithm converges to the correct minima so that errors in the output are due only to the errors in the input. Thus, in order to say whether an inversion generated by the mode amplitude perturbation method is superior to one generated by eigenvalue perturbation, we require an estimate of the error statistics of the input data to the two methods. The EKF estimator provides

estimates of the uncertainties in both the mode amplitudes and the eigenvalues, making it relatively simple to compute the uncertainties in the output parameters.

Also, regardless of whether or not a new method outperforms previous methods, one can consider the question of whether the new method can be combined with the old to provide a better answer than either method on its own. In the case of the mode amplitude and eigenvalue perturbation methods, the answer is yes. The similarity of the two methods provides a simple way of combining the two, but care must be taken that the contributions of each are weighted appropriately.

To begin our comparison of the two methods, it will be useful to remind ourselves of the equations at the heart of each. The eigenvalue perturbation method makes use of this expression for the change in an eigenvalue with respect to a change in the bottom:

$$\Delta k_n = \frac{1}{2k_n} \int_0^D \frac{\Delta q(z) Z_n^2(z)}{\rho(z)} dz .$$

Here k_n is the *nth* modal eigenvalue, D is the total depth of the waveguide (at times treated as infinity, but often set to some large, finite value as a practical approximation), $Z_n(z)$ is the *nth* mode function, $\rho(z)$ the density profile, and $q(z) = k^2(z) = \frac{\omega^2}{c^2(z)}$, where $\omega = 2\pi f$ is the angular frequency of the source, and $c(z)$ is the sound speed profile. It should be pointed out here that k_n and $Z_n(z)$ in this equation is computed for one value of $k(z)$, and that a given change in $k^2(z)$, $\Delta q(z)$ leads to the change in the eigenvalue, Δk_n. Similarly, the mode amplitude perturbation method is based upon the following equations:

$$a_{nm} = \frac{1}{k_n^2 - k_m^2} \int_0^D \frac{\Delta q(z) Z_n(z) Z_m(z)}{\rho(z)} dz \quad \text{and} \quad \Delta Z_n(z) = \sum_{j \neq n} a_{nj} Z_j(z),$$

which can be combined as:

$$\Delta Z_n(z) = \sum_{j \neq n} Z_j(z) \frac{1}{k_n^2 - k_j^2} \int_0^D \frac{\Delta q(z') Z_n(z') Z_j(z')}{\rho(z')} dz'.$$

These equations, and the eigenvalue perturbation equation given above, are derived in more detail in chapter 2 of this thesis. Also mentioned in chapter 2 is the fact that we can parameterize the bottom so as to invert for a finite number of unknowns. If we define $\Delta c(z) = X_i c_i(z)$, then we have

$$\frac{\partial k_n}{\partial X_i} \approx \frac{\Delta k_n}{X_i} = \frac{\partial k_n}{\partial X_i} = \frac{-1}{k_n} \int_0^D \frac{c_i(z) Z_n^2(z)}{c_o^3(z) \rho(z)} dz,$$

where $c_0(z)$ is the background profile with which k_n and $Z_n(z)$ are computed. Likewise, for the mode amplitude perturbation we get

$$\frac{\partial Z_n(z)}{\partial X_i} = \sum_m a_{nm}^{X_i} Z_m(z) \quad \text{where} \quad a_{nm}^{X_i} = \frac{2\omega^2}{k_m^2 - k_n^2} \int_0^D \frac{c_i(z) Z_n(z) Z_m(z)}{c_0^3(z) \rho(z)} dz.$$

With these equations we can compute the partial derivatives of a measurable quantity with respect to the unknown quantities we seek. Thus we can solve an equation of the general form

$$\left[\frac{\partial Q}{\partial X} \right] \vec{X} = \Delta \vec{Q},$$

where Q is a set of measurable quantities, such as eigenvalues or mode amplitude ratios, and X is a set of unknown parameters which we wish to determine. We solve for the unknown parameters by inverting the derivative matrix, usually using the pseudo inverse described in chapter 1:

$$\vec{X}_{LS} = \left[\frac{\partial \tilde{Q}}{\partial X}^{\#}\right] \Delta \vec{Q}.$$

This computed change in the bottom parameters is added to the background, and the process is repeated until convergence is achieved. Assuming the algorithm converges within the correct minima, and the measurement errors are small enough that they affect the estimate linearly, we can compute the bias and variance of our inversion from the bias and variance of the input data. The bias of the input data should be zero. If it is not, it should be subtracted from the input data before the inversion. For completeness, however, we state that if there is bias in the input, the bias in the parameter estimation would be equal to the input bias times the inversion matrix:

$$< \vec{X}_{LS} > = \left[\frac{\partial \tilde{Q}}{\partial X}^{\#}\right] < \Delta \vec{Q} >.$$

As pointed out in chapter 4, the variance of the estimate is also dependent on the variance of the input:

$$\text{cov}(\vec{X}_{LS}) = \left[\frac{\partial \tilde{Q}}{\partial X}^{\#}\right] \text{cov}(\Delta \tilde{Q}) \left[\frac{\partial \tilde{Q}}{\partial X}^{\#}\right]^{T}.$$

Thus, using the information in the P matrix of the EKF estimator, and the inversion matrix of whichever method one is using, it is possible to determine the variance of the error of the estimate. When using the mode amplitude perturbation method, one must remember that actual input data are the mode ratios, which must be computed from the quantities produced by the EKF estimator. Using the approximation derived in chapter 4, we have that

89

$$\delta m_n(z,z_s) \approx m_n(z,z_s)\left(\frac{\delta A_n}{A_n} + \delta k_n \frac{\partial \zeta(z,k_n)}{\partial k_n} \frac{1}{\zeta(z,k_n)} - \frac{\delta A_1}{A_1} - \delta k_1 \frac{\partial \zeta(z,k_1)}{\partial k_1} \frac{1}{\zeta(z,k_1)}\right),$$

which can be used to compute the covariance matrix of the input data using the P matrix. Note that errors in the mode ratio can be due to either errors in the estimate of the mode amplitudes, or the receiver-position eigenvalues.

With the expressions for error covariance in hand, it is possible to compare the performance of the two inversion methods. The "better" algorithm is the one with the lower error covariance. However, it is often the case that one method will perform better at estimating one parameter, while the other performs better at estimating another. For example, eigenvalue perturbation tends to provide a lower-variance estimate of sub-bottom features, whereas mode amplitude perturbation methods do better at estimating shallower parameters such as interface speeds. This is not unexpected, since the mode functions contain little energy at large depths, and thus are not much affected by the sub-bottom, whereas the eigenvalues are determined by the boundary conditions, and thus are highly affected by sub-bottom parameters. Thus, which method is preferred will depend, in part, on which parameters are considered the most critical.

Further, which method is better will depend on the quality of the input data. If the eigenvalues are estimated more precisely, then eigenvalue perturbation will be superior. On the other hand, if the mode amplitudes are estimated well, but the eigenvalues have significant error, mode amplitude perturbation should be used. Which inversion method wins will also depend somewhat on the environment, as this will change the inversion matrix. In short, in order to determine the preferred method for a given scenario, one must perform the inversions and compare the computed error variances.

In order to illustrate this, we can examine a synthetic example. In this case, we will use a waveguide 200m deep, and 4000m in length. The water column is 50m deep, and treated as isovelocity (1500 m/s, 1000 kg/m^3). Below this is a 70m deep, range-dependent sediment layer. Two parameters (interface speed and gradient) are sufficient to describe this layer at any given range. Density is treated as constant in this layer (1600 kg/m^3). Below the sediment layer is a range-independent sub-bottom (1800 m/s, 1800 kg/m^3). The waveguide is shown in figure 5.1.
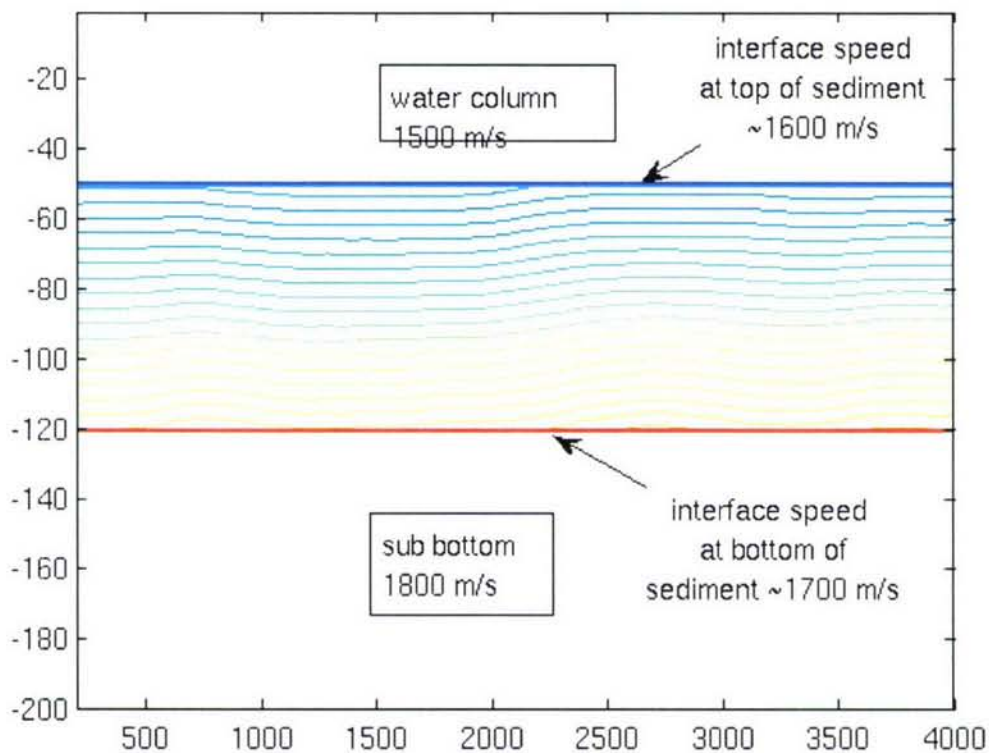

Figure 5.1 Sound speed contours of the synthetic waveguide. The contours indicate lines of constant sound speed. The water column and sub bottom are treated as isovelocity and range-independent, whereas the sediment layer varies both in depth and range.

A synthetic field was generated for the waveguide, based on a VLA at range 0, and a 100 Hz source moving from 200 m to 4000 m from the VLA, at a constant depth of 30 m. Seven propagating modes were used to generate the synthetic field. The VLA had 9 elements, with 5 m spacing, going from 5 m depth to 45 m. Zero mean, Gaussian random noise was then added to the field. Figure 5.2 shows the measured pressure magnitude as a function of range for a single depth, indicating the noise levels present.

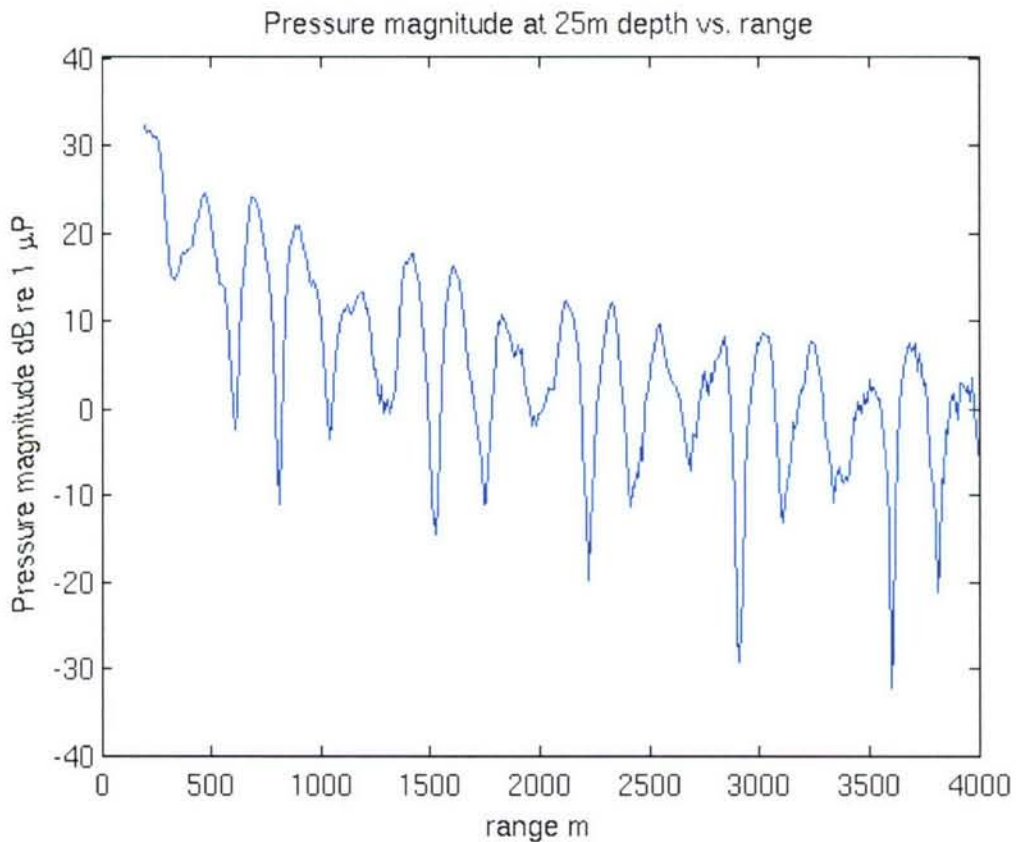Figure 5.2 Pressure magnitude as a function of range. This figure shows the pressure magnitude measured at the 25m VLA phone in the synthetic experiment.

This synthetic field measurement was then passed to the EKF estimator to determine the mode amplitudes and eigenvalues, as described in chapter 2. Figures 5.3 and 5.4 show the estimated and true values of these parameters as a function of range.

While there are errors in the estimates, the EKF estimator does a good job of tracking the true values throughout the synthetic experiment. Note that just as was seen in the synthetic experiment from chapter 3 involving buried river channels, the EKF estimate lags the true value slightly.
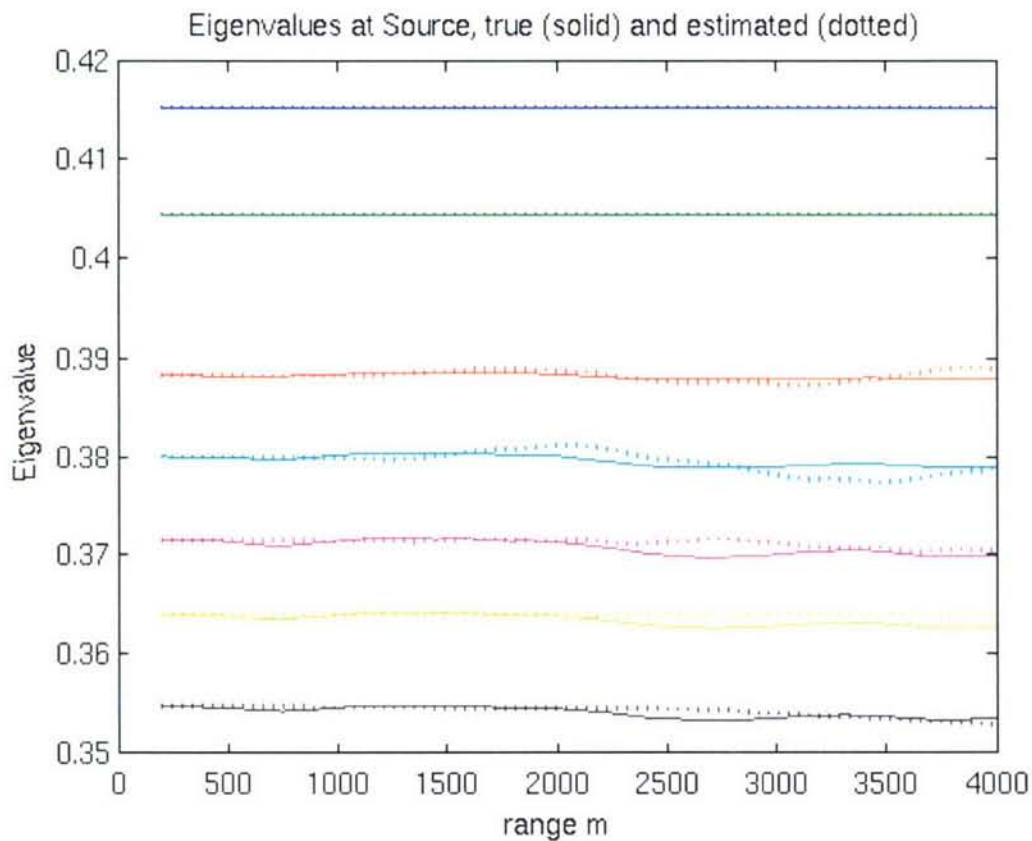
Figure 5.3 Eigenvalues estimates from the EKF estimator. The figure shows the values of the 7 eigenvalues at the source location as a function of range. The solid lines indicate the true values, while the estimates are shown with dotted lines.

With the estimates of the mode amplitudes and eigenvalues, we can estimate the bottom parameters. For the inversions, we treat the water and sub bottom and known, and invert for the interface speed and gradient of the sediment layer. In order to compare

the mode amplitude perturbation method with the eigenvalue perturbation method, we do

separate inversions using each, and compute the error variances as described earlier.
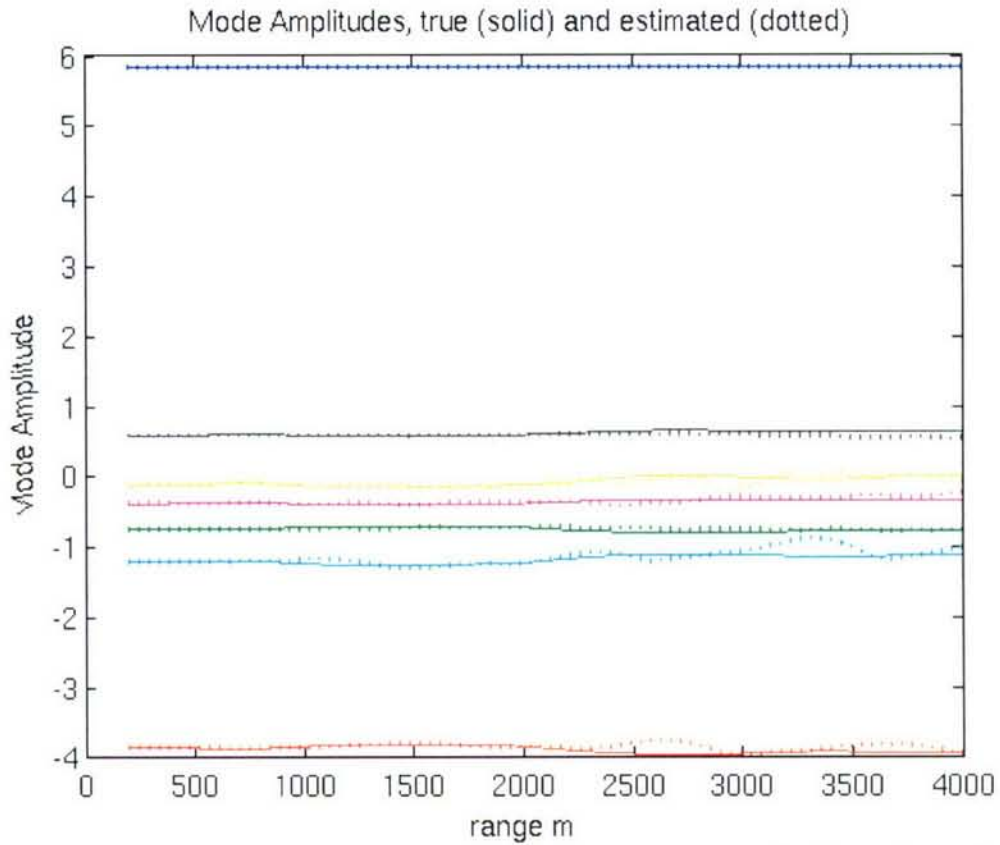

Figure 5.4 Estimates of the mode amplitudes from the EKF estimator. The figure shows the mode amplitudes at the VLA as a function of source-receiver range. The true values are depicted with solid lines, and estimated values are shown dotted.

The error variance of the 4 estimated parameters are shown in figure 5.5. In this

case, the error variances favor the mode amplitude perturbation for some of the

parameters, and eigenvalue perturbation for others. For example, the interface speed at

the source location as estimated by mode amplitude perturbation has a lower error

variance than when estimated by eigenvalue perturbation. But the opposite is true of the

receiver-position estimate of the interface speed. Error variance of the source-position

gradient is similar with both methods, but the VLA-position estimate of the gradient

appears to favor mode amplitude perturbation. This, however, is likely due to unjustified

overconfidence in the estimates coming out of the EKF estimator. We will see shortly

that the eigenvalue perturbation estimate of the VLA-position gradient is actually

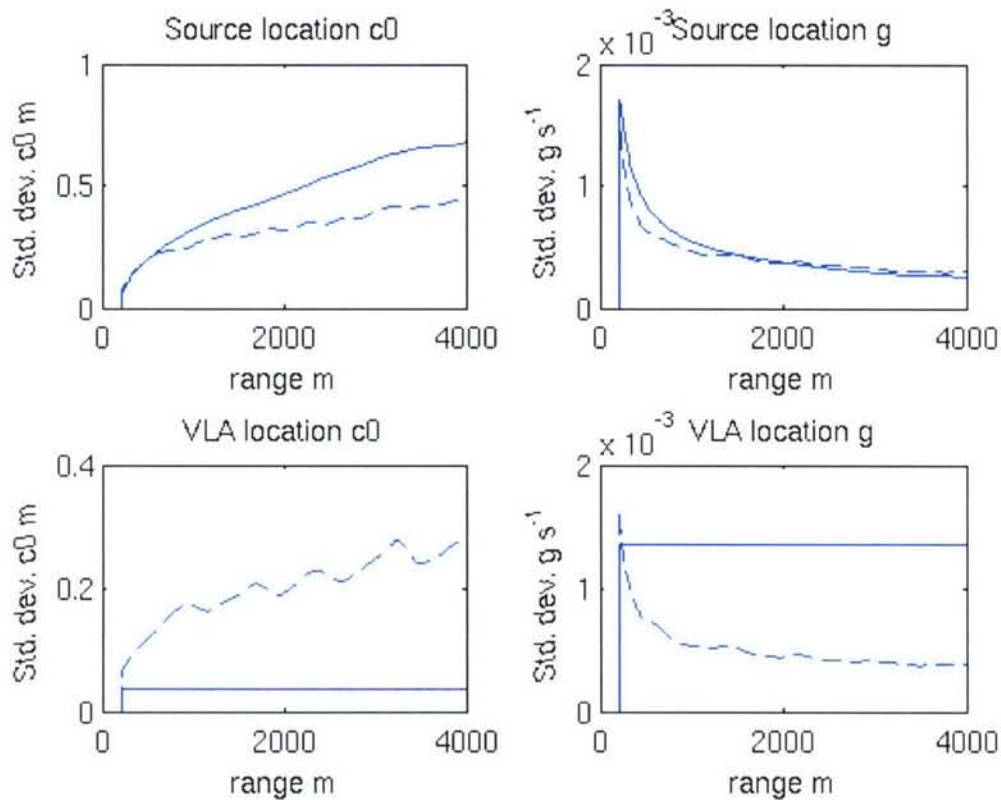superior to the mode-amplitude estimate.

Figure 5.5 Error variance of the 4 estimated parameters. In each sub plot, the error variance of the parameter as determined by eigenvalue perturbation (solid) and mode amplitude perturbation (dashed) are shown as a function of source range. For the interface sound speed, the mode amplitude perturbation method performs better at the source location, but the eigenvalue perturbation method performs better at the VLA position. The source position estimate of the gradient is a toss-up, with mode amplitude perturbation doing better at long ranges, and eigenvalue perturbation doing better at short ranges. Mode amplitude perturbation appears to out perform eigenvalue perturbation at estimating the VLA position gradient, but the level of confidence estimated for the mode amplitude method is likely due to unjustified overconfidence in the EKF estimator's values of the input parameters.

Since both methods appear to be superior for different parameters, it is logical to

ask if we can combine the two methods for even better estimates. At first it might seem

best to simply take the values of the different parameters from the two inversions with the

lowest error variance. However, the eigenvalue perturbation estimate of the VLA-position interface speed, for example, depends on the eigenvalue perturbation estimate of the source position interface speed. The estimates of the various parameters are not independent and mixing and matching them may lead to problems. A better solution is to use all the information available to both methods at once. To do so we can combine the two derivative matrices into one, and the two data vectors into one, forming a new equation:

$$
\begin{bmatrix} \dfrac{\partial Q_{MA}}{\partial X} \\ - \\ \dfrac{\partial Q_E}{\partial X} \end{bmatrix} \bar{X} = \begin{vmatrix} \Delta \vec{Q}_{MA} \\ - \\ \Delta \vec{Q}_E \end{vmatrix}.
$$

This equation is essentially the same as the one used by either of the methods, but we cannot solve it in quite the same way, because the two types of input data differ significantly in size. Eigenvalue perturbations are typically on the order of .001, whereas mode ratio amplitudes are closer to order .1 or even 1. Using the pseudo inverse to find a least-squares solution to the above equation would essentially result in the exact same answer as if we just used the mode amplitude equation. The eigenvalue perturbations are just too small to affect the answer. Thus, we must take into account the relative sizes of the perturbations, rather than the absolute sizes. One way of doing this is to weight each equation (*i.e.,* each row of the above matrix and the corresponding entry in the data vector) by the standard deviation of the perturbation, and then compute the pseudo inverse.

An even better way is to make use of the Stochastic Inverse [3], [37]. This method for creating the inverse matrix not only accounts for the relative sizes of the

different perturbations, but also for the relative sizes of the changes in the parameters. For example, an interface speed can easily change by 10 m/s over an experiment, whereas changes in a gradient are likely to be much, much smaller. We will leave the derivation of the stochastic inverse to the references, and state here only the expression for the inverse. For the equation

$$\vec{d} = [G]\vec{q} + \vec{e}$$

where d is the measurement vector, e a zero-mean noise vector with covariance R_e, and q the unknown, zero-mean parameter vector with covariance R_q, the stochastic inverse solution is

$$\hat{q}_{SI} = \left[\tilde{G}^{SI}\right]\vec{d} = [R_q][G]^T \left[[G][R_q][G]^T + [R_e]\right]^{-1}\vec{d}.$$

This inversion minimizes the errors squared, but normalizes by the uncertainty in the quantity with the error. For example, an error in a quantity with variance .0001 should be much less than an error in a quantity with a variance of .1. The inversion also tries to minimize the q vector, again weighted by variance of the elements in the vector.
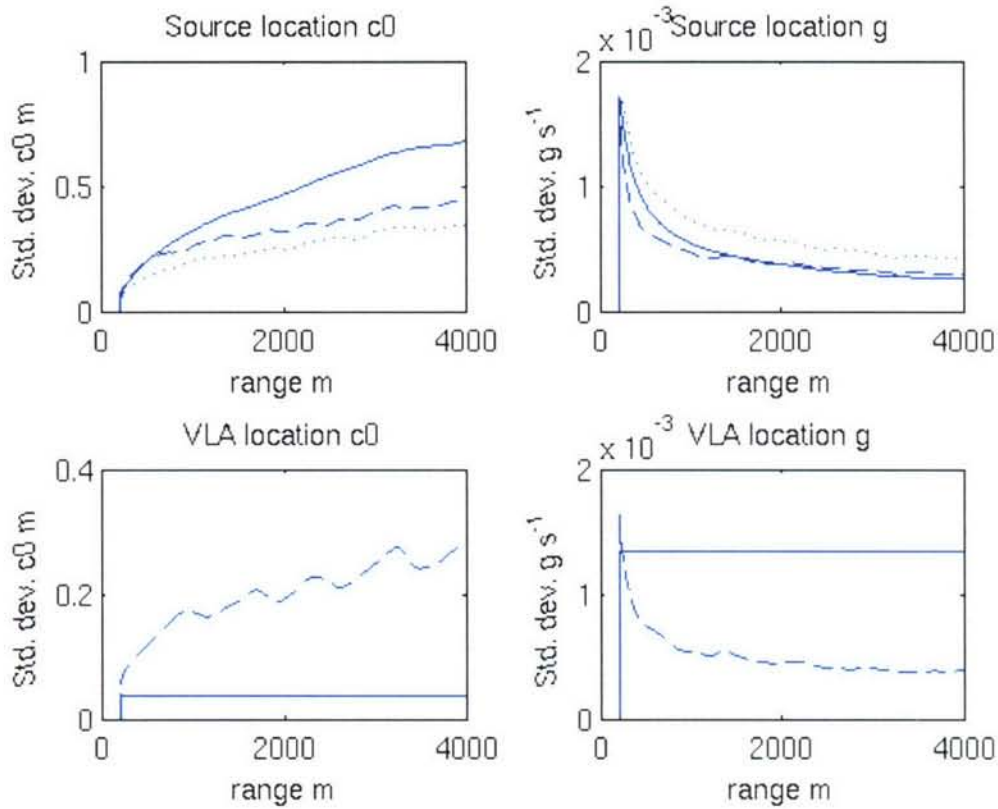
Figure 5.6 Error variance of the 4 inverted parameters. Eigenvalue perturbation is shown in solid, mode amplitude in dashed, and the combined method shown in dots. For the bottom two plots, the curve for the combined method overlays that of the eigenvalue method.

Using the stochastic inverse, the P matrix from the EKF estimator, and an estimated matrix of the unknown parameter variance, it is possible to solve the combined mode amplitude/eigenvalue perturbation equation. And just as with the pseudo inverse, we can use the stochastic inverse to compute the error variance of our inversion. If we do so for our synthetic example, we can get the best of both methods, as shown in figure 5.6. As can be seen in the figure, the estimate of the source location interface speed is improved over either method, and the variance of the estimate of the VLA position parameters matches that of the eigenvalue method. On the other hand, there is a slight decrease in the quality of the source-location estimate of the gradient.

Now that we have looked at the predicted accuracy of our inversion, let us examine the actual accuracy. Figures 5.7 through 5.10 show the true values of the parameters, along with the inverted values from the three methods. Figure 5.7 shows the inversion results for the interface speed at the source position. While all three methods are fairly comparable in quality, for most of the experiment the combination result tracks the better of the two other inversion. The differences are not particularly dramatic in this case, but over the course of the experiment, the combination result does appear to be the best of the three.
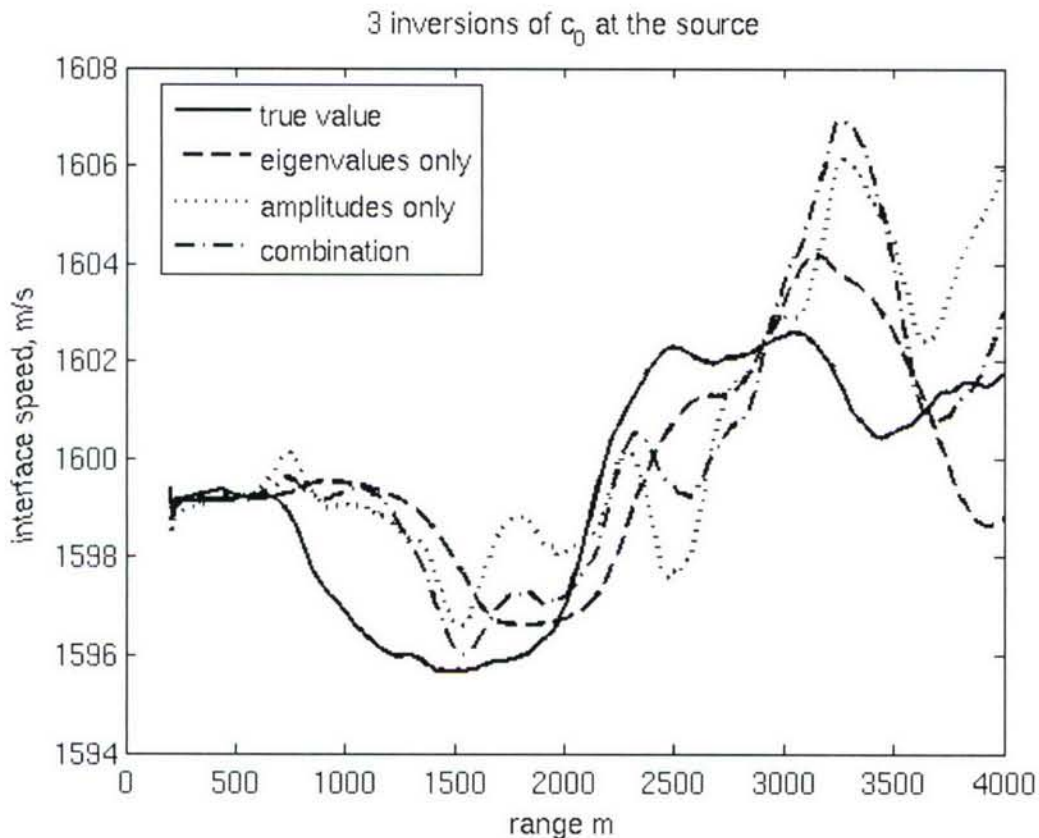
Figure 5.7 Interface sound speed at the source, real and inverted. The figure shows the true interface speed (solid) as a function of range for the synthetic experiment. The three inverted values are also shown. The eigenvalue perturbation result is shown dashed, the mode amplitude inversion dotted, and the combination method result in dash-dot. For most of the experiment, the combination method tends to track whichever of the two inversion results is closer to the true value.

Figure 5.8 Interface speed at the VLA, real and inverted. The eigenvalue perturbation result lies atop the constant true value of the VLA-position interface speed. As the estimate of the VLA position eigenvalue changes very little, the inverted value of the parameter changes very little as well. The mode amplitude, however, can change due to changes at the source, or at the VLA, and thus t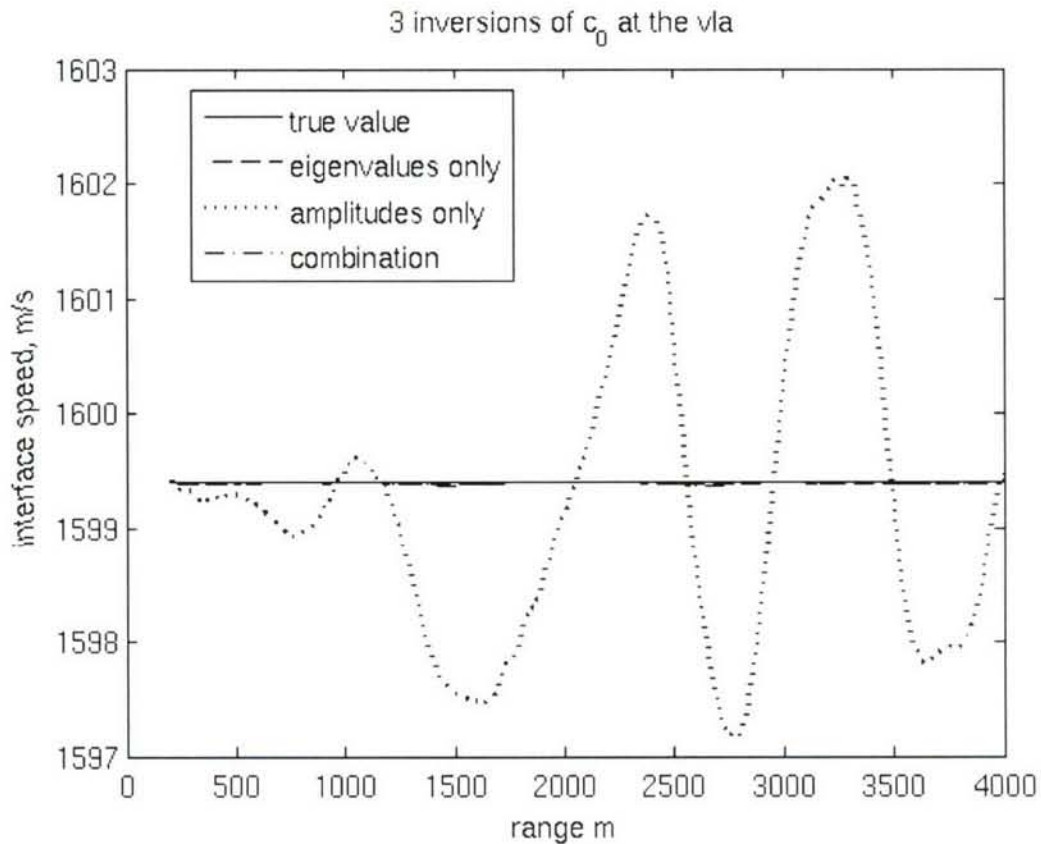he mode amplitude perturbation result varies with range, and has much larger error. The combination inversion method tracks the result of the eigenvalue perturbation well, and thus remains at the true value.

Figure 5.8 shows the inverted values of the interface speed at the VLA position. Here the improvements of the combination method are clear. The eigenvalue perturbation method, using an unchanging estimate of the local eigenvalue does a good job of tracking the true value of the parameter. The mode amplitude method, however, uses mode ratios which are affected by changes at both the source and receiver position, and thus struggles to distinguish between changes at the source, and changes at the VLA. Thus, its estimate of the parameter is non-constant, and has significantly larger errors

than the eigenvalue perturbation method. The combination method is able to use the local eigenvalue estimate, and thus tracks the true value well.
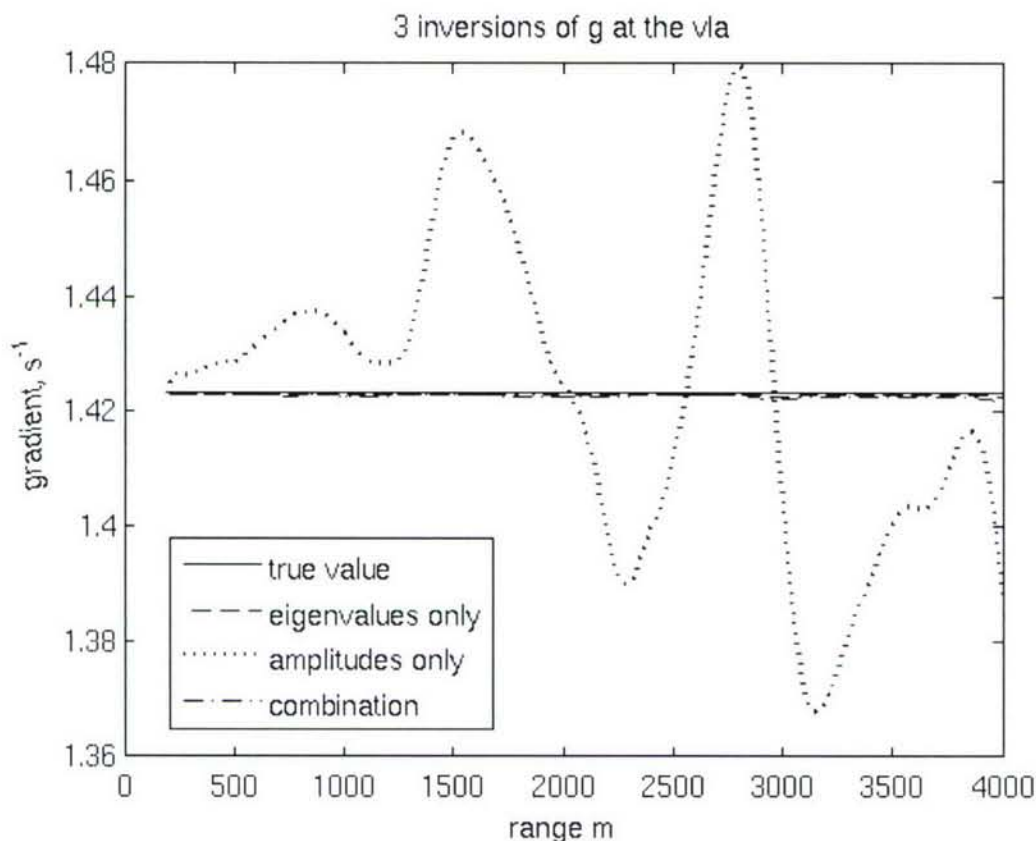

Figure 5.9 Gradient at the VLA, real and inverted. As in figure 5.8, the true value is constant, and the unchanging local eigenvalue estimate gives a good result for the eigenvalue perturbation method. The mode amplitude perturbation method, however, has trouble distinguishing between local and source-position changes, and thus provides a poorer estimate of the parameter. The combination result tracks the eigenvalue perturbation result, and the true value, well. Note that the level of error in the mode amplitude perturbation result is at odds with the predicted error variance for this parameter shown in figures 5.5 and 5.6. This is likely due to and overconfidence in some of the estimates coming out of the EKF estimator.

Figure 5.9 is similar. Again the mode amplitude perturbation method struggles to distinguish between local and source-position changes, and thus gives a non-constant estimate with larger errors. The methods that make use of the estimate of the local eigenvalue, however, do a good job at tracking the parameter.

Figure 5.10 shows the estimates of the gradient at the source position. Similar to the source-position interface speed estimates, all three methods have similar levels of

bias. However, the combination method does significantly better than the other two methods at tracking the features of the parameter in range. The maxima and minima of the parameter are clearly visible in the combination method estimate, though the values of the estimate are slightly off.
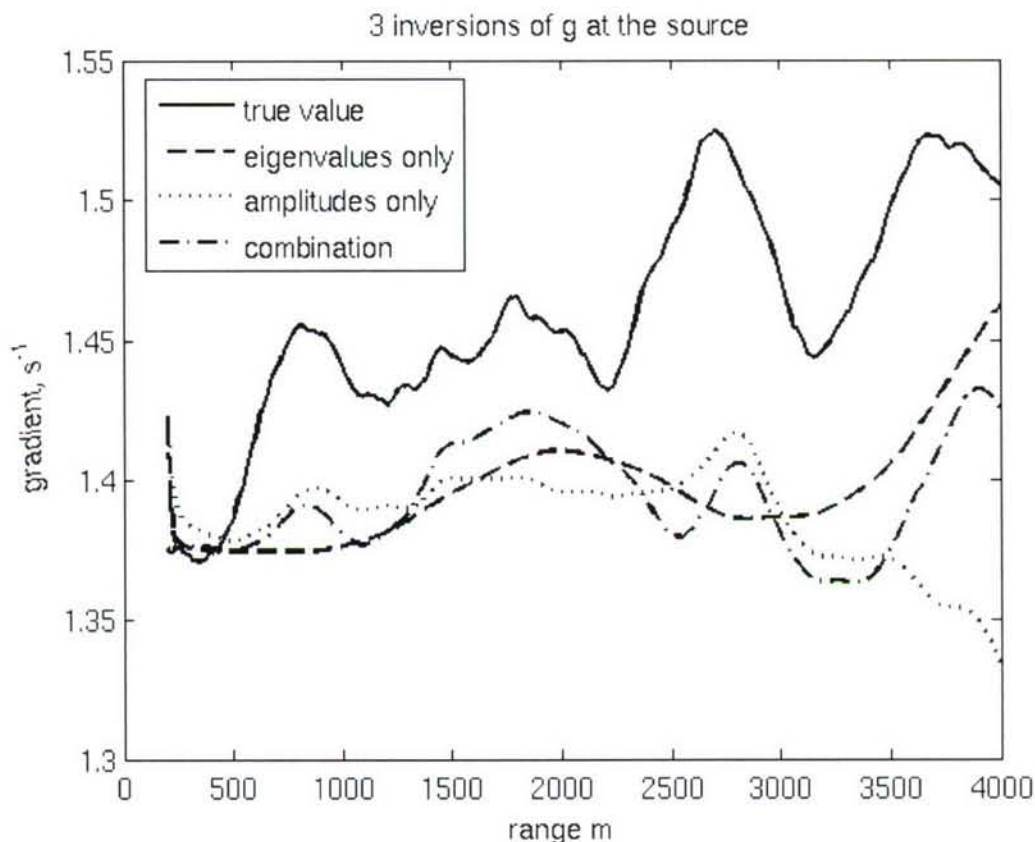
Figure 5.10 Gradient at the source, real and inverted. The figure shows the source-position gradient as a function of source-receiver range. The true value is shown with a solid line. All three estimates are comparable, and show similar amounts of bias. However, the combination method does the best at tracking the features of the parameter with range.

All of this shows that when the methods converge to the correct profile, small but non-trivial improvements can be made to the estimates by combining the two methods by using the stochastic inverse. Additionally, there is reason to believe that using the combination method will also improve the chances of converging to the correct profile. Recall the problem mentioned in chapter 4 of converging to the wrong minima. By using

more input data (*i.e.*, both the amplitudes and the eigenvalues), we constrain the problem further, making it less likely that we will converge to an incorrect profile. A profile that would have been a minimum of the cost function when using just one set of data may not be a minimum when both sets are used. To illustrate this, we will examine some real-world data from the MOMAX 3 experiment 4.

Similar in most respects to the MOMAX experiment discussed in chapter 3, this one (MOMAX 3, experiment 4) involved a 75 Hz, pure-tone source suspended from a moving ship. The pressure field excited by this source was measured by hydrophones suspended from drifting buoys. Over the course of the experiment, the source moved several kilometers away from the buoys, which remained relatively stationary. By invoking reciprocity, we can create a synthetic aperture horizontal array along the source path. Using this synthetic aperture, it is possible to estimate the eigenvalues of the modes propagating across the aperture. We can then use eigenvalue perturbation to invert for the bottom sound speed profile.

The exact details of this experiment are secondary to our discussion here, as we are using it only to illustrate a point. What matters is that during one point in the experiment the 75 Hz source was towed from 3 km away from a hydrophone buoy, to 5 km away, and the hydrophone at the buoy recorded the pressure field. The bottom was considered range-independent, but the 80m water column had both range- and time-dependence due to internal waves. Because only a single receiver was used, rather than a VLA, the EKF estimator could not be applied. However, the Hankel transform method of Rajan, *et. al* [1] could, with the usual range-dependent errors. This allowed us to estimate the eigenvalues of the propagating modes. But due to the fluctuations in the

field due to internal waves and a possibly changing source depth, the modal amplitudes could not be accurately estimated.
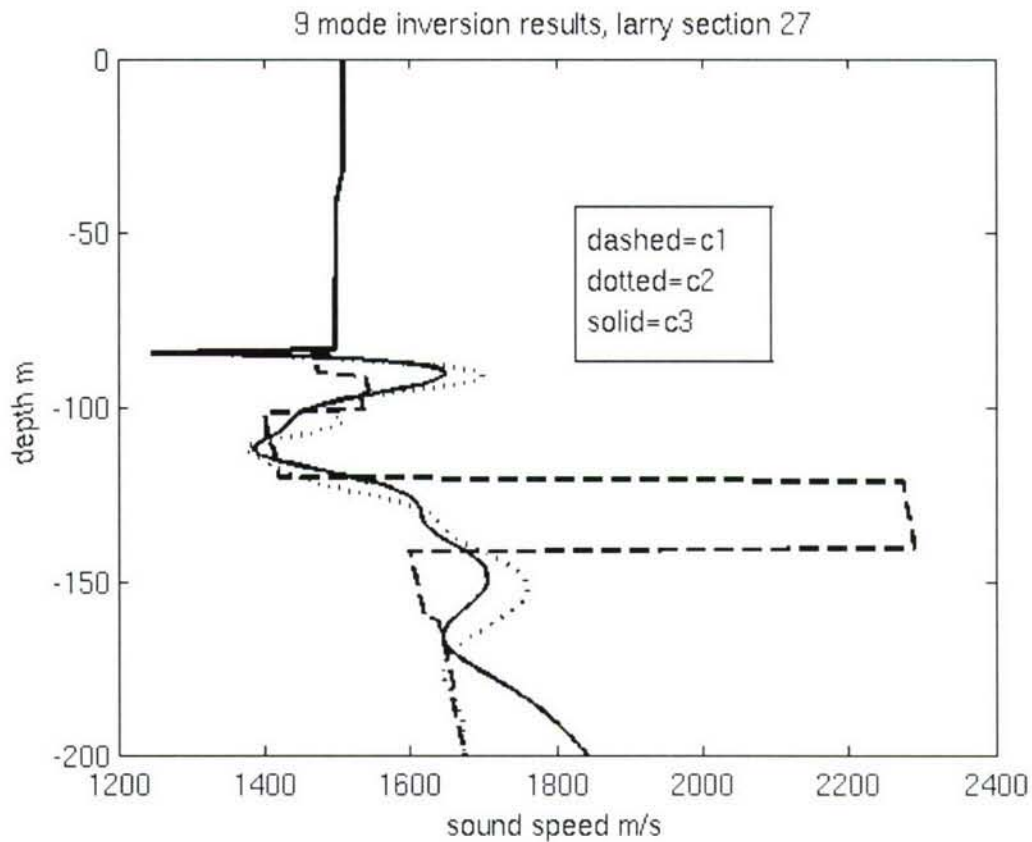
Figure 5.11 Three profiles producing the same eigenvalues. The three sound speed profiles produce almost the same set of eigenvalues when excited by a 75 Hz source, despite being significantly different from each other.

Using the values of the eigenvalues obtained using the Hankel transform of the pressure data, three different bottom SSPs were obtained, each of which provided an excellent match to the measured eigenvalues. Figure 5.11 shows the three bottom profiles, and figure 5.12 shows how well they match the measured eigenvalues. All three inverted profiles in figure 5.11 appear rather non-physical, and it is likely that none of them match the true profile well. Additionally, there is a question of whether the first mode in figure 5.12 is actually a mode, or an artifact of the transform. These concerns,

however, are slightly beside the point. The issue of importance is that we have found three different bottoms, all of which match the measurement well. Using the eigenvalues only, we are unable to determine which of them (if any!) is the closest to the true profile. Even using the pressure field generated by the three bottoms does not help us significantly, as can be seen in figure 5.13. All three bottoms provide similar matches to the measured pressure, and any differences from the measurement are as likely due to internal waves as to the bottom.
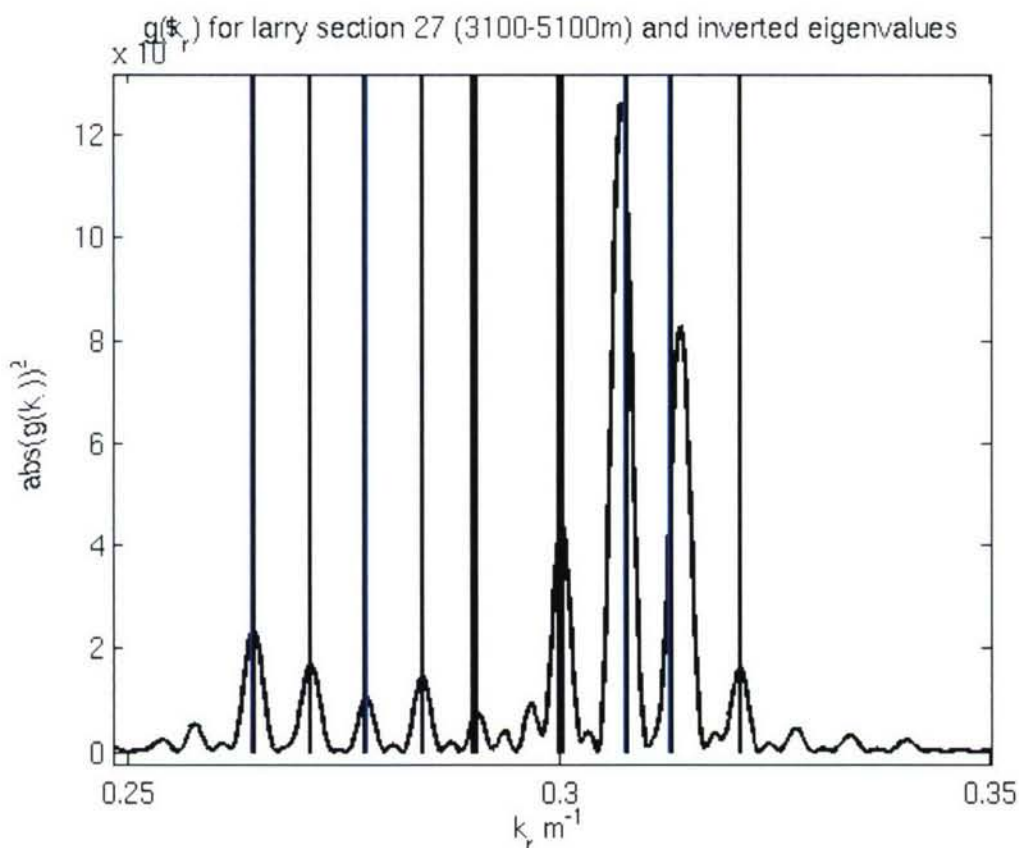
Figure 5.12 Eigenvalues of the profiles in figure 5.11. The vertical lines represent the eigenvalues of the three profiles shown in figure 5.12. The solid, non-vertical curve is the transform of the pressure data from the MOMAX 3, experiment 4 data.

If we had access to estimates of the mode amplitudes, however, we would be able to determine which of the three is correct (or perhaps that none of the three is correct). Figure 5.14 shows the 7th mode for each of the three bottoms. Note that the distribution

of energy in this mode, between the water and the bottom, varies significantly between

the three profiles. Thus, an estimate of the mode amplitudes would provide us additional

information with which to select the best profile.
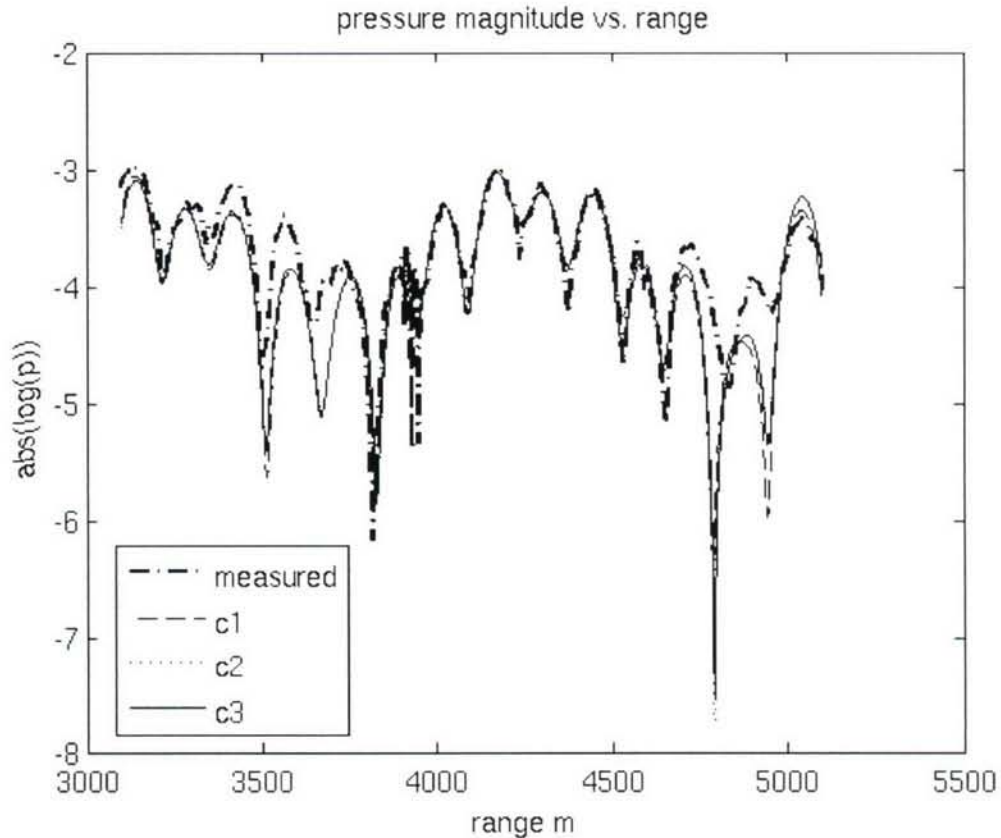
Figure 5.13 Pressure magnitude vs. range for the three profiles in figure 5.11. The dash-dot curve represents the pressure measured in MOMAX 3, experiment 4. The other curves correspond to the three sound speed profiles shown in figure 5.11.

As was stated already, we do not believe that any of the three profiles shown is a

good match to the true SSP, as all are rather unphysical (though others have found

evidence of such "double-ducting" profiles before [38]). Probably each of these

inversions is a result of convergence into an incorrect minimum. What is important,

though, is that if we had accurate mode amplitude information, there would no longer be

minima at these locations, since at most one of them would match the amplitude

information as well. Thus using the combined amplitude/eigenvalue perturbation

method, we would expect to be less likely to fall into an incorrect minima as often as we would with either the mode amplitude or eigenvalue method on their own.
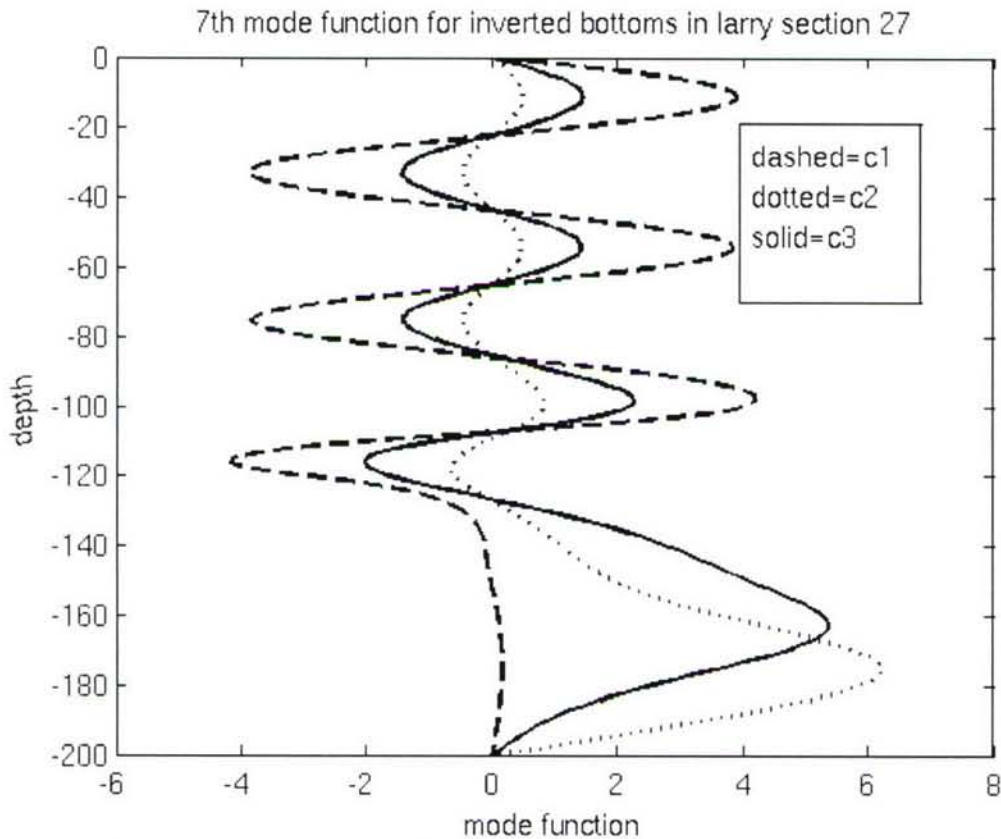
Figure 5.14 Mode 7 for the profiles in figure 5.11. The three curves represent the mode function for seventh mode for each of the profiles shown in figure 5.11. Note that the distribution of energy between the water and the sediment varies significantly between the three cases.

The combination inversion method will not solve the problem of convergence to incorrect minima entirely. There will still be incorrect minima to fall into. Presumably, however, there will be less of them near the true profile, making the inversion less sensitive to the selection of the initial background model. The problem will remain undetermined, however, unless only a small number of parameters is used. Thus, even if we converge to the correct minimum, there is still the chance that the true profile contains some component from the null space of the inversion matrix. After what we have seen in this chapter, though, we can be confident that using both eigenvalue and mode amplitude

107

data should improve our inversion results. Because of this, we say that the mode amplitude perturbation method has truly added to our ability to perform geo-acoustic inversions, and made possible a level of performance that was not available using the earlier method of eigenvalue perturbation alone.

Chapter 6: Summary, conclusions, and future work

In this thesis we have introduced a mode amplitude perturbation algorithm, and an EKF estimator capable of estimating range-dependent parameters of the pressure field using a VLA and a moving source. Additionally, we have described a method of combining the mode amplitude perturbation method with the eigenvalue perturbation method of Rajan *et al* [1]. to achieve superior inversions. These contributions will hopefully prove useful to other members of the ocean acoustics community. In this chapter we will summarize the findings of this thesis, and consider possible future work.

The mode amplitude perturbation method is derived in chapter 2 of the thesis, and numerous examples of its application are given in chapter 3. In many ways similar to eigenvalue perturbation, it is based on the idea of determining the small change to a background model that would be needed to bring the mode amplitudes predicted by the background model into agreement with the amplitudes actually measured. The first derivative of the mode function with respect to sound speed perturbation is expressed in terms of a weighted sum over all the modes in the background model. Since the mode functions themselves cannot be measured directly, the expressions for the derivatives are used to determine the derivative of the mode ratios, defined as $m_n(z,z_s) \int \dfrac{Z_n(z)Z_n(z_s)}{Z_1(z)Z_1(z_s)}$, with respect to the bottom parameters. With these derivatives and a measurement of the mode ratios, a linear set of equations can be solved to determine the necessary corrections to the bottom model.

The algorithm treats the relationship between the bottom parameters and the acoustic parameters as linear, an approximation that is only valid if the background model is very similar to the true SSP. Either the pseudo-inverse matrix described in chapter 1, or the stochastic inverse used in chapter 5 can be employed for this. In either case, we minimize a cost function (the sum of squared differences in predicted and measured mode ratios in one case or those squared differences weighted by their variance and the variance of the bottom parameters in the other) based on the assumption of a linear dependence on bottom parameters. Multiple iterations, in which the inverted bottom from one step is used as the background for the next, can be used to overcome the non-linearity of the problem. In effect we use a number of small, linear steps rather than one non-linear step. If the original background model was close enough to the true bottom, we should get convergence after a few iterations.

The errors in this process were discussed in chapter 4. When our background is close enough to the true bottom that we get convergence to the correct answer, the only errors will be due to imperfect estimations of the input parameters. These will propagate through the algorithm linearly, making it a simple task to estimate the error variance of the inversion based on the error variance of the input. More serious errors can occur when the original background is not sufficiently close to the true bottom, and the algorithm converges to an incorrect minimum of the cost function. In such cases there is no obvious way to estimate the error in the inversion. Often, but not always, using a small number of parameters will make it obvious that the algorithm has converged to an incorrect minimum, as the resulting SSP will appear non-physical. However, using a small number of parameters makes it impossible to correctly model some bottoms, and

causes us to miss important features of the SSP. To balance these issues, it can be useful to start with a simple parameterization, and then increase the number of parameters after initial convergence. This limits the likelihood of convergence to an incorrect minimum early on, but still allows us to capture more complicated bottom features in our final inversion. When large numbers of parameters are included, however, it becomes necessary to keep an eye on the stability of the inversion matrix. Often it is necessary to limit the number of singular values used in creating the pseudo inverse to avoid instabilities in the inversion.

These general properties of the mode amplitude perturbation method are very similar to those of the eigenvalue perturbation method. The only real difference between the two is in the input data used. This makes it straightforward to combine the two algorithms into a single inversion, taking advantage of both the redundant information in the two data sets for greater robustness to error, and the independent information to reduce the chance of convergence into incorrect minima of cost function. The only obstacle to this combination is the vastly different size of the perturbations to the input data in each case. The equations used must be properly weighted, or one set of data will completely dominate the inversion. As described in chapter 5, the stochastic inverse matrix provides the proper weighting, based on the covariance of the input data and that of the unknown bottom parameters. The resulting inversion method can provide improved error variance, and generally combines the advantages of the two methods.

A separate but related contribution in the thesis is the EKF estimator of the field parameters. This estimator is derived in the second half of chapter 2, and provides the input data needed for the mode amplitude perturbation inversion method, as well as that

for the eigenvalue perturbation method, and the combined eigenvalue-mode amplitude method. Unlike the Hankel transform method used by Rajan *et al* [1]., the EKF estimator does not make use of range-windowed data, and thus should be able to detect very small scale changes in the bottom. Further, because the range window in normally formed over time, the Hankel transform method can struggle with temporal variation in the water column. The EKF estimator, on the other hand, makes a separate estimate of the field parameters at each measurement time. If simultaneous sound speed measurements of the source- and receiver-position water columns are made along with the acoustic measurements, these temporal variations can be accounted for. All that is required is that the fluctuations, both in time and space, are sufficiently small that the adiabatic approximation is valid.

The basic idea behind the EKF estimator is that of guessing, checking, and updating. The estimator tracks three sets of parameters as state variables: the mode amplitudes and eigenvalues at the VLA position, and the eigenvalue at the source position. At each step of the process, the estimator predicts the pressure field at the VLA based on its current estimate of the state variables. It then compares its prediction to the actual measurements made on the VLA. Based on its levels of confidence in its prediction and the measurement, it adjusts its prediction to better match the measurement. When confidence in the prediction is high, and the measurements are expected to have large errors, the estimator will weight the prediction more than the measurement. Conversely, if there are low levels of measurement noise, but large uncertainty in the current prediction of the state variables, the measurement carries more weight, and the prediction is altered more significantly.

The EKF estimator actually bears a number of similarities to the perturbative inversion methods. It treats the pressure field as linearly dependent upon the state variables, an approximation that is only valid if the estimates of the parameters are close to being true. This is analogous to the requirement that the background model be sufficiently close to the true SSP for the perturbative inverse methods (which are also based on a linear approximation) to work. Further, when computing the correction to the initial prediction of the state variables, the EKF estimator solves an inverse problem very similar to the one solved by the perturbative inversion methods. Both involve a matrix equation involving a matrix of the first derivatives of the measured quantities with respect to the unknown parameters, and both must compute an effective inversion matrix (almost never actually a true inverse matrix, since the matrices usually aren't square) in order to determine the corrections to the unknown parameters.

It should be pointed out that neither the perturbative inverse methods, nor the EKF estimators are black boxes. Both depend upon the initial estimate of the parameters being sought (*i.e.,* the SSP for the inversion, and the state variables of the EKF estimator), and poor initial estimates can lead to poor final estimates. Further, the EKF estimator requires estimates not only of the initial values of the state parameters, but also of their covariance matrix, and the covariance matrix of the measurement errors. These covariance matrices are usually not measured directly, but rather estimated, and poor estimates can lead to poor performance by the EKF. Adding even further complication, if the stochastic inverse is used to perform the perturbative inversion, one requires an estimate of the covariance matrix of the unknown parameters, which is largely guess work. Furthermore, if a large number of parameters is used in the bottom model, one

113

must decide how many singular values to use when creating the inversion matrix in order avoid instability issues. Oftentimes, obtaining a believable result from the EKF estimator for one of the perturbation algorithms requires numerous attempts, using slightly different initial estimates of the unknowns and variances. Similarly, obtaining a believable result from the pertubative inversions often requires numerous attempts using different initial backgrounds, different bottom parameterizations, or different tolerances for the number of singular values used when forming the inversion matrix. At times it may seem that the act of inverting for a bottom SSP involves as much 'black art' as 'black box!' This is due less to the methods themselves, however, than due to the complexity of the problem. The task we are trying to perform is quite complicated and difficult: to use a finite sampling of a pressure field (which varies non-linearly with bottom properties) to determine a profile which is in reality a function of a continuous variable (or of four continuous variables to be completely accurate!). We are attempting a non-linear, underdetermined inversion, which only become possible after numerous simplifying assumptions, and even it then remains a very non-trivial task!

The task is an important one, though, since bottom has such an important effect on sound propagation in shallow water. Anyone using low-frequency sound as a tool in shallow water will benefit from improved estimates of the geoacoustic properties of the bottom. It is hoped that the methods described in this thesis will help provide these improved estimates.

It is also hoped that the work presented in this thesis will be used as a starting point for future studies. One example of this would be using the algorithms presented on real world data, such as that taken during the Shallow Water '06 experiment. Another

would be to combine the EKF estimator directly with the inversion methods to create an estimator that treated the bottom parameters as the state variables. We will examine both of these possibilities briefly.
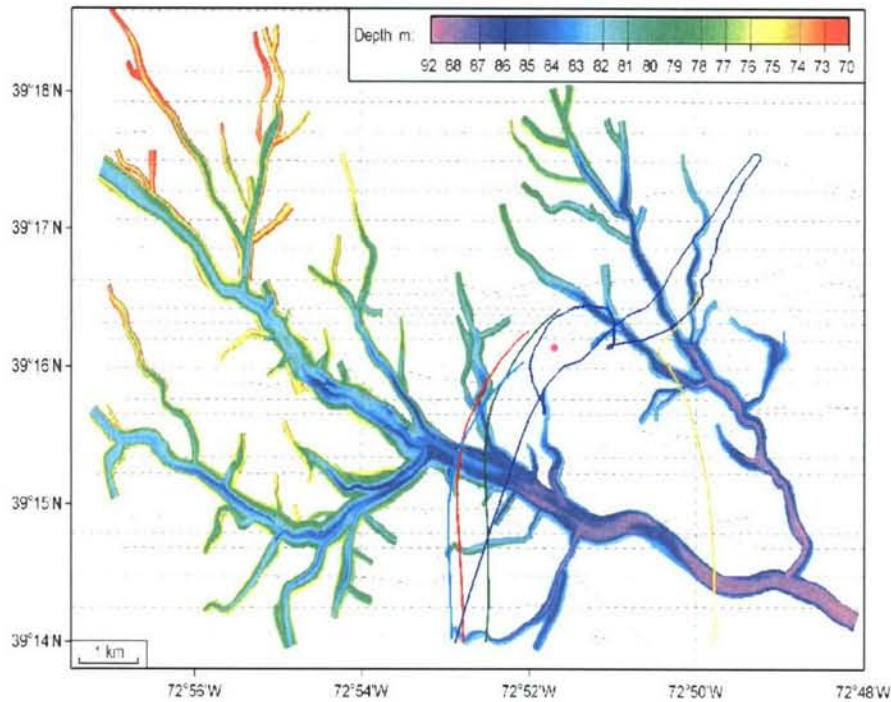
Figure 6.1: Source and receiver tracks from the SW06 experiment overlaid on Goff's estimate of buried river channel locations. The dot at approximately 75°52' W, 39°18'N is the Webb VLA location. The curves indicate paths of the source and four drifting MOMAX buoys during the course of the experiment. The estimated positions of buried river channels are shown as well.

The SW06 experiment is discussed near the end of chapter 3, and this discussion would be the best place to start for anyone wanting to apply the algorithms presented in this thesis to that data set. The first step would be the initial processing of the raw data to get them to the point where they are in the $p(r,z)$ form. In other words, the binary data in the measurement devices must be converted into meaningful pressure measurements matched to the range and depth at which they were made. This step of the process is a

115

separate issue from those addressed in this thesis, and for the moment we will treat it as granted. Figure 6.1 shows the positions of the VLA, source, and drifting MOMAX buoys during the experiment [39]. Once we have $p(r,z)$, the next step is to estimate a background model for the VLA. One way to do this would be to use the Hankel-transform and eigenvalue perturbation to get an estimate of the average, range-independent bottom model for the whole experiment. This too would require an initial background model, however. Such a background could be generated based on John Goff's model of the bottom in the area, shown in Figure 3.13.

Once an initial background model is in hand, it can be used to generate initial estimates of the field parameters. We also need to predict the covariance matrix of the field parameters, which is, unfortunately, a somewhat arbitrary decision. Essentially this matrix will reflect our level of uncertainty about the bottom, and our expectation of the variation in the bottom. It is entirely possible that numerous guesses at this matrix will be necessary before a good inversion is produced. We will also need an estimate of the measurement noise covariance matrix. However, in this case it at least will be possible to generate this in a more rigorous fashion. Measurement noise is usually high frequency variation on top of the lower frequency signal. By filtering out the signal portion of the acoustic data, and treating what is left over as noise, it should be straightforward to generate a noise covariance matrix.

With these estimates it will be possible to initialize the EKF estimator, and set it running. Provided the data are good, and none of our assumptions has been violated, the result should be a range dependent estimate of the field parameters. The next step is to use these estimates to invert for the bottom profile. This will require a parameterization

of the bottom. It is suggested that the initial parameterization be as simple as is practical. Two or three parameters is probably best. Use the mode amplitude perturbative inversion method described in this thesis to convert the field parameter estimates into bottom parameter estimates. Using the inverted bottom at one range as the background model for the next range should work well. After this, if a more detailed bottom profile is required, a more complicated parameterization can be used. Take care, however, to use only singular values large enough to avoid instability when the number of parameters becomes as large, or larger, than the number of input data.

It should also be possible to use the combined mode amplitude/eigenvalue method described in chapter 5. In addition to the requirements given above, this will also necessitate estimating a bottom parameter covariance matrix. Again, this is a somewhat arbitrary estimation. However, if the mode amplitude and/or eigenvalue methods are used first, their results could be used to estimate bottom covariance. The EKF estimator step will be exactly the same as when the mode amplitude perturbation method was used. The inversion process will be very similar as well, but will include the additional information of the eigenvalues, and the necessary covariance matrices. Whichever method is used to do the inversion, it is hoped that the methods presented in this thesis will prove useful.

As mentioned earlier, another avenue for future work would be to eliminate the field parameter estimation step, and design an EKF estimator that predicted the geo-acoustic parameters directly. Using the expressions for the derivatives of the mode functions and eigenvalues with respect to bottom parameters, it would be possible to determine the derivative of the pressure field with respect to bottom parameters. The

geoacoustic parameters could therefore be used as the state variables of the EKF. This natural combination of the two processes would eliminate the intermediate step of estimating the acoustic field parameters. Care would need to be taken to ensure the validity of the linear approximations involved. For example, the phase of the field at long ranges varies very non-linearly with respect to the eigenvalues, thus the measurement would likely need to be started at a short range. Another issue would be the number of receivers necessary. This would affect both the robustness to measurement noise, and the number of parameters which could be inverted without additional assumptions. There is also the issue that this estimator would require estimates of the covariance of the bottom parameters, which could be problematic. However, the combined eigenvalue-mode amplitude perturbation method also requires this, so this would not be unique to the one-step estimator. As a trade off, however, we would no longer require an estimate of the field-parameter covariance matrix.

Yet another area that deserves further investigation is inversions for density and attenuation profiles. While the necessary perturbation results for these quantities were derived in chapter 2, the thesis has focused on sound speed perturbations. While inverting for density parameters should not add significant complexity to the inversion, doing so for attenuation will be slightly harder. This is because attenuation is intrinsically tied to cumulative range. An error in the estimate of a parameter at one range could throw off estimates of attenuation parameters at all subsequent ranges. This does not make inverting for attenuation parameters impossible, but it will make doing so more difficult, and potentially introduce stability issues.

Surely there are many other possible ways to expand on the research presented in this thesis. Those mentioned are merely the most obvious next steps. It is hoped that the material presented here will be useful to the ocean acoustic community both as a useful tool, and as a starting point for further advances.

Appendix: Numerical Calculation of Mode Functions

This appendix covers the process of numerically determining the mode functions. The methods used are based on the analytical solution to the proper Sturm-Liouville problem given in chapter 2, but here we will focus on the practical issues involved with solving the problem computationally. The first step of the process is to represent the waveguide by dividing it into a stack of thin iso-velocity layers, bounded above and below by pressure-release surfaces. The pressure-release boundary at the top of the waveguide represents the air-water interface. The lower pressure-release surface is an artificial boundary used to make calculation of the eigenvalues of non-propagating modes easier. While the non-propagating modes are usually neglected because they decay exponentially with range, we require them for calculating the derivatives of the mode functions. By adding the pressure-release boundary at depth, the eigenvalues of the non-propagating modes will be purely imaginary, and thus much easier to compute than the complex eigenvalues that would exists if we used a half space as the bottom of the waveguide. Because this boundary is artificial, care must be taken that it be placed much deeper than the turning points of the propagating modes. This will ensure that true propagating mode functions will be quite small at the interface depth, and that approximating them as zero there will not lead to significant errors in the rest of the mode function. Typically this false bottom is placed about 3 water column depths below the water-bottom interface, though depending on the frequencies and sound speed profiles used, this may need to be increased.

The thickness of the layers themselves is usually set to 1 meter, however if finer resolution is needed, or if high frequencies are used, thinner layers may be appropriate.

Within each layer, sound speed and density are considered constant, allowing an exact solution of the depth-dependent mode equation within them, consisting of up- and down-going plane waves.

Once the waveguide is described this way, it is possible to solve for the eigenvalues. This is the most difficult and computationally-intensive part of the process. The user specifies the number of eigenvalues sought, and the program proceeds to determine subintervals of the k_r^2 axis containing these eigenvalues. The program makes use of k_r^2 rather than k_r because the former will always be purely real, even when the latter becomes imaginary. This allows us to search the real-axis only, instead of the complex plane, and is the main motivation for including the false bottom (which makes the k_r values of non-propagating modes purely imaginary). The process for separating the real axis into subintervals containing only one eigenvalue makes use of the shooting method, and a process similar to bisection. Given any value of k_r^2, a shooting method solution to the mode equation can be obtained by specifying a value for the mode function at the free surface (zero), and it's first depth derivative. This is sufficient the determine the exact solution in that layer, and thus the value of the mode function and it's slope at the bottom of the layer. Continuity of pressure and vertical displacement at the lower interface provides the information needed to solve for the exact solution in the next layer. This process can be continued all the way to the bottom of the waveguide. However, the solution will only match the boundary condition at the false bottom when the given value of k_r^2 corresponds to an eigenvalue.

By counting the number of zero-crossings of a shooting method solution, it is possible to determine which two eigenvalues the given k_r^2 is between. A solution

containing n zero-crossings (including the zero at the free surface) implies that k_r^2 lies

between the n-1th and nth eigenvalues. This allows us to divide the real line into

subintervals containing only one eigenvalue, and to assign an ordinal rank to each

subsection (*e.g.*, "this subsection contains the 3rd eigenvalue").

Once subintervals are determined for each eigenvalue sought, determination of

the eigenvalues proceeds using the shooting method and bisection. The shooting method

is tried at the mid point of the subinterval in the shooting method, and the number of

zero-crossings computed. This tells us if the eigenvalue is to the left or right of the

midpoint, and effectively gives us a new subinterval containing the eigenvalue that is one

half the size of the previous subinterval. This process is repeated until a very small

subinterval remains (usually $10^\wedge$-14 m$^\wedge$-2 wide). We call the midpoint of this subinterval

the eigenvalue.

Once the eigenvalues are obtained, it is time to compute the mode functions.

Again we make use of the shooting method, but now we shoot upwards from the bottom.

We do this because mode functions that are evanescent in the bottom will have errors due

to the numerical error in our estimation of the eigenvalue, and these errors will grow

exponentially in the bottom. However, if we start at the bottom with no error, the error at

the upper interface, where the modes are not evanescent, will be very small. In other

words, the shooting method is unstable when shooting from propagating to evanescent

portions of the waveguide, but stable when shooting from evanescent to propagating

portions. The shooting method provides us only with an unnormalized mode function,

however. The last step of the process is to normalize the mode function by dividing it by

the square root of the numerical integration over depth of the mode function squared

divided by density. The output of this numerical algorithm has been compared to

Porter's KRAKEN normal mode code [27], and the two methods agree to well within the

acceptable levels of error for the methods described in this thesis.

References

[1]   S. D. Rajan, J. F. Lynch, and G. V. Frisk, "Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water," J. Acoust. Soc. Am. **82**, 998-1017 (1987).

[2]   M. D. Collins and W. A. Kuperman, "Nonlinear inversion for ocean-bottom properties," J. Acoust. Soc. Am. **92** (5), 2770-2783, (1992).

[3]   L. L. Souza, *Inversion for Subbottom Sound Velocity Profiles in the Deep and Shallow Ocean.* PhD thesis, MIT and WHOI, Cambridge and Woods Hole, MA 2005.

[4]   I. M. Gelfand and B. M. Levitan, "On the Determination of a Differential Equation from its Spectral Function," Amer. Math. Soc. Trans. Series 2, vol 1, 253-304 (1955).

[5]   K. M. Becker. *Geoacoustic Inversion in laterally varying shallow-water environments using high-resolution wavenumber estimation.* PhD thesis, MIT and WHOI, Cambridge and Woods Hole, MA 2001.

[6]   F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics.* (AIP Press, New York, 1994

[7]   E. L., Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**(5), Nov. 1980.

[8]   G. V. Frisk, *Ocean and Seabed Acoustics: A Theory of Wave Propagation.* (P T R Prentice Hall, Englewood Cliffs, New Jersey, 1994).

[9]   C. L. Pekeris, "Theory of propagation of explosive sound in shallow water," Geol. Soc. Am. Mem. **27** (1948).

[10]   A. O. Williams, Jr., "Hidden depths: Acceptable ignorance about ocean bottoms," J. Acoust. Soc. Am. **59**(5), May 1976.

[11]   A. D. Pierce, "Extension of the method of normal modes to sound propagation in an almost-stratified medium," J. Acoust. Soc. Am. **37**, 19-27 (1965).

[12]   R. B. Evans, "A couple mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetratable bottom," J. Acoust. Soc. Am. **74**, 188-195 (1983).

[13]   M. B. Porter, F. B. Jensen, and C. M. Ferla, "The problem of energy conservation in one-way models," J. Acoust. Soc. Am. **89**, 1058-1067 (1991).

[14]   S. L. Marple Jr. *Digital Spectral Analysis with Applications.*   Prentice Hall, Englewood Cliffs, NJ, 1987.

[15]   A. Thode and K. Kim, "Multiple-order derivatives of a waveguide acoustic field with respect to sound speed, density, and frequency," J. Acoust. Soc. Am. **116**, 3370-3383 (2004).

[16]   C. T. Tindle, L. M. O'Driscoll, and C. J. Higham, "Coupled mode perturbation theory of range dependence," J. Acoust. Soc. Am. **108**, 76-83 (2000).

[17]   A. Pierce, private communication (2005).

[18]   S. D. Rajan, G. V. Frisk, J. F. Lynch, "On the determination of modal attenuation coefficients and compressional wave attenuation profiles in a range-dependent environment in Nantucket sound," IEEE J. Ocean. Eng. **17**, 118-128 (1992).

[19]   K. Ohta, K. Okabe, I. Morishita, S. Ozaki, and G. V. Frisk, "Inversion for seabed geoacoustic properties in shallow water experiments," Acoust. Sci. & Tech. **26**, 326-337 (2005).

[20]   K. Ohta and G. V. Frisk, "Modal Evolution and inversion for seabed properties in weakly range-dependent shallow-water waveguides," IEEE Journal of Oceanic Engineering, 22(3):501-521, July 1997.

[21]   T. B. Neilsen and E. K. Westwood, "Extraction of acoustic normal mode depth functions using vertical line array data," J. Acoust. Soc. Am. **111**, 748-756 (2002).

[22]   J. R. Buck, J. C. Preisig, and K. E. Wage, "A unified framework for mode filtering and the maximum *a posteriori* mode filter," J. Acoust. Soc. Am. **103**, 1813-1824 (1998).

[23]   J. V. Candy and E. J. Sullivan, "Model-based identification: An adaptive approach to ocean-acoustic processing," IEEE J. Ocean. Eng. **21**, 273-289 (1996).

[24]   J. V. Candy, *Signal Processing: The Model-Based Approach*. New York: Academic, 1970.

[25]   Maciej Niedźwiecki, *Identification of Time-varying Process*. Wiley, Chichester, England, 2000.

[26]   Maria Isabel Ribeiro, *Kalman and Extended Kalman Filters: Concept,*

[27]   M. B. Porter, The KRAKEN normal mode program. Technical report, SACLANT Undersea Research Center, La Spezia, Italy, 1991.

[28]   G. V. Frisk, "LWAD 99-1 Modal Mapping Experiment II (MOMAX II)," Woods Hole, MA (2000).

[29]   G. V. Frisk, K. M. Becker, and J. A. Doutt. "Modal mapping in shallow water using synthetic aperture horizontal arrays." *OCEANS 2000 MTS/IEEE Conference and Exhibition, 11-14 Sept. 2000, Providence, RI,* volume 1, pages 185-188, September 2000.

[30]   B .R. Gomes, J. K. Fulford, and R. Nero,"Preliminary Pre-Experiment Environmental Characterization for the Littoral Warfare Advanced Development (LWAD) 99-1 SE Gulf of Mexico." Naval Research Laboratory, (1998).

[31]   J. A. Goff and S. Nordfjord, "Interpolation of Fluvial Morphology Using Channel-Oriented Coordinate Transformation: A Case Study From the New Jersey Shelf," Mathematical Geology, Vol. 36, No. 6, August 2004.

[32]   S, Nordfjord, J. A. Goff, J. A. Austin Jr., and C. K. Sommerfield, "Seismic geomorphology of buried channel systems on the New Jersey outer shelf: assessing past environmental conditions," Marine Geology 214 (2005) 339-364.

[33]   J. A. Goff, J. A. Austin Jr., S. Gulick, S. Nordfjord, B. Chistensen, C. K. Sommerfield, H. Olson, and C. Alexander, "Recent and modern marine erosion on the New Jersey outer shelf," Marine Geology 216 (2005) 275-296.

[34]   J. A. Goff, private communication, Feb. 2006.

[35]   K. Ohta, K. Okabe, I. Morishita, G.V. Frisk, and A. Turgut, "Modal Inversion Analysis for Geoacoustic Properties of the New Jersey Continental Shelf in the SWAT Experiments," under revision for publication in *IEEE J. Oceanic Engineering* (2006).

[36]   M. Zanolin, I. Ingram, A. Thode, and N. C. Makris, "Asymptotic accuracy of geoacoustic inversions," J. Acoust. Soc. Am. **116**(4), (2004).

[37]   Joel N. Franklin. "Well-posed stochastic extensions of ill-posed linear problems." *Journal of Mathematical Analysis and Applications,* 31:682-716, (1970).

[38]   J. A. Doutt, R. J. Cederberg, R. B. Evans, S. D. Rajan, W. M. Carey, and W. L. Siegmann, "Experimental horizontal wavenumber spectra and implications for full field processing," from O. Diachok *et al.*'s *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, 341-346. Kluwer Academic Publishers, Netherlands, 1995.

[39]   Cynthia Sellers, private communication, 2006.

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. MIT/WHOI 2007-03	2.	3. Recipient's Accession No.

4. Title and Subtitle
Geoacoustic Inversion by Mode Amplitude Perturbation

5. Report Date
February 2007

6.

7. Author(s)
Travis L. Poole

8. Performing Organization Rept. No.

9. Performing Organization Name and Address

MIT/WHOI Joint Program in Oceanography/Applied Ocean Science & Engineering

10. Project/Task/Work Unit No.
MIT/WHOI 2007-03

11. Contract(C) or Grant(G) No.
(C) N00014-02-1-0337

(G)

12. Sponsoring Organization Name and Address
Office of Naval Research
Woods Hole Oceanographic Institution Academic Programs Office

13. Type of Report & Period Covered
Ph.D. Thesis

14.

15. Supplementary Notes
This thesis should be cited as: Travis L. Poole, 2007. Geoacoustic Inversion by Mode Amplitude Perturbation. Ph.D. Thesis. MIT/WHOI, 2007-03.

16. Abstract (Limit: 200 words)
This thesis introduces an algorithm for inverting for the geoacoustic properties of the seafloor in shallow water. The input data required by the algorithm are estimates of the amplitudes of the normal modes excited by a low-frequency pure-tone sound source, and estimates of the water column sound speed profiles at the source and receiver positions. The algorithm makes use of perturbation results, and computes the small correction to an estimated background profile that is necessary to reproduce the measured mode amplitudes. Range-dependent waveguide properties can be inverted for so long as they vary slowly enough in range that the adiabatic approximation is valid. The thesis also presents an estimator which can be used to obtain the input data for the inversion algorithm from pressure measurements made on a vertical line array (VLA). The estimator is an Extended Kalman Filter (EKF), which treats the mode amplitudes and eigenvalues as state variables. Numerous synthetic and real-data examples of both the inversion algorithm and the EKF estimator are provided. The inversion algorithm is similar to eigenvalue perturbation methods, and the thesis also presents a combination mode amplitude/eigenvalue inversion algorithm, which combines the advantages of the two techniques.

17. Document Analysis   a. Descriptors
Inversion
Perturbation
acoustics

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement
Approved for publication; distribution unlimited.

19. Security Class (This Report)
UNCLASSIFIED

20. Security Class (This Page)

21. No. of Pages
126

22. Price

(See ANSI-Z39.18)   *See Instructions on Reverse*   OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce