

Applications of a Fast Helmholtz Solver in Exploration Seismology

by

Gregory Tsiang Ely

Submitted to the Department of Earth, Atmospheric, and Planetary
Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Degree Month Degree Year

© Massachusetts Institute of Technology Degree Year. All rights
reserved.

Author
.....

Department of Earth, Atmospheric, and Planetary Sciences
.....
.....

Thesis date

Certified by
.....

Alison Malcolm
Associate Professor
Thesis Supervisor

Accepted by
.....

Chair name here
Chairman, Department Committee on Graduate Theses

**Applications of a Fast Helmholtz Solver in Exploration
Seismology**

by

Gregory Tsiang Ely

Submitted to the Department of Earth, Atmospheric, and Planetary Sciences
on Thesis date, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Seismic imaging techniques rely on a velocity model inverted from noisy data via a non-linear inverse problem. This inferred velocity model may be inaccurate and lead to incorrect interpretations of the subsurface. In this thesis, I combine a fast Helmholtz solver, the field expansion method, with a reduced velocity model parameterization to address the impact of an uncertain or inaccurate velocity model. I modify the field expansion framework to accurately simulate the acoustic field for velocity models that commonly occur in seismic imaging. The field expansion method describes the acoustic field in a periodic medium in which the velocity model and source repeat infinitely in the horizontal direction, much like a diffraction grating. This Helmholtz solver achieves significant computational speed by restricting the velocity model to consists of a number of non-overlapping piecewise layers. I modify this restricted framework to allow for the modeling of more complex velocity models with dozens of parameters instead of the thousands or millions of parameters used to characterize pixelized velocity models. This parameterization, combined with the speed of the forward solver allow me to examine two problems in seismic imaging: uncertainty quantification and benchmarking global optimization methods. With the rapid speed of the forward solver, I use Markov Chain Monte Carlo methods to estimate the non-linear probability distribution of a 2D seismic velocity model given noisy data. Although global optimization methods have recently been applied to inversion of seismic velocity model using raw waveform data, it has been impossible to compare various types of algorithms and impacts of parameters on convergence. The reduced forward model presented in this paper allows me to benchmark these algorithms and objectively compare their performance to one another. I also explore the application of these and other geophysical methods to a medical ultrasound dataset that is well approximated by a layered model.

Thesis Supervisor: Alison Malcolm
Title: Associate Professor

Acknowledgments

acknowledgments here.

Contents

1	Introduction	23
1.1	Uncertainty Quantification Challenges	24
1.1.1	Large dimensionality and visualization challenges	25
1.1.2	Computational Challenges & Expensive Forward Solves	25
1.1.3	Interpretation & Human bottleneck	27
1.2	Thesis Overview	28
1.2.1	Global Optimization	29
1.2.2	Uncertainty Quantification	29
1.2.3	Ultrasound and Bone	30
2	global optimization algorithms for building velocity models	31
2.1	Overview	31
2.2	Introduction	32
2.3	Forward Model	35
2.3.1	Reduced parameterization	35
2.3.2	Field Expansion	36
2.4	Global Optimization Methods	39
2.4.1	Particle Swarm Optimization	40
2.4.2	Simulated Annealing	42
2.5	Results	43
2.5.1	Simple models	44
2.5.2	Convergence and noise	50
2.5.3	Complex Examples	51

2.5.4	Marmousi Inversion	53
2.6	Discussion	56
2.7	Conclusion	59
3	Uncertainty Quantification of Seismic Velocity Models and Seismic Images	61
3.1	Overview	61
3.2	Introduction	62
3.3	Bayesian velocity model inversion	66
3.4	Field expansion forward solver	67
3.5	Efficient posterior calculation	70
3.6	Results	78
3.6.1	Simple Anticline	78
3.6.2	Gradient Migration	82
3.6.3	Convergence	86
3.6.4	Starting Model Sensitivity	86
3.6.5	Degrees of Freedom	89
3.6.6	Computational cost	90
3.7	Discussion	90
4	Imaging the Interior of Bone	93
4.1	Overview	93
4.2	Introduction	93
4.3	Acquisition Geometry	96
4.4	Applications of PSO & MCMC to Ultrasound	96
4.5	Axial: Methods & Synthetic	100
4.6	Methods: Multiple Suppression	105
4.6.1	Head wave velocity analysis	105
4.6.2	NMO semblance velocity analysis	106
4.7	Axial: <i>in vivo</i>	115
4.8	Transverse: Methods & Synthetic	118

4.9	Transverse: in vivo	126
4.10	Discussion	128
4.11	Conclusion	129
4.12	Acknowledgments	130
4.13	Appendix: Seismic techniques	130
4.13.1	Normal move-out and semblance	130
4.13.2	Kirchoff Migration	131
4.13.3	Reverse Time Migration	132
5	Conclusion	133
5.1	Future Work	133
5.1.1	Testing assumptions and validity of assumptions	134
5.1.2	Establishing test problems & reproducibility	134

List of Figures

1-1	Diagram illustrating the seismic uncertainty framework from raw data to probability distribution over a quantity of interest and final interpretation.	26
1-2	Left: Log-Likelihood distribution for a fictional function. Right: Likelihood for the same fictional problem. The local minima present in the right panel would create a problem for local optimization methods but does not contribute significantly to the likelihood function and does not impact the distribution.	29
2-1	sketch of field expansion velocity model and repeating boundary conditions.	36
2-2	A: Reduced parameterization diagram. B: Interpolated layer interfaces based on the reduced parameterization. C: Pixelized velocity model derived from reduced parameterization consisting of piecewise constant layers.	37
2-3	Real and imaginary components of the PySIT and field expansions fields at several frequencies. We generate these results with a dispersion term of 2.5×10^{-2}	39
2-4	Dispersion as a function of misfit for several source frequencies. . . .	40
2-5	Marmousi constant density velocity model.	44
2-6	A: All initial models and the true velocity profile. B: Initial best fitting and final best velocity profile after 250 iterations.	45

2-7 Particle Swarm Optimization Convergence as a function of number of layer.	46
2-8 Simulated Annealing convergence results.	47
2-9 Convergence as a function of source frequency and number of master layers. Median values are shown with a solid line and top/bottom 20 th percentile solutions are shown with dashed lines.	48
2-10 Convergence as a function of number of agents. Median values are shown with a solid line and top/bottom 20 th percentile solutions are shown with dashed lines.	49
2-11 Convergence as a function of maximum iteration. Median values are shown with a solid line and top/bottom 20 th percentile solutions are shown with dashed lines.	49
2-12 Model misfit as a function of initial guess quality. Median values are shown with a solid line and top/bottom 20 th percentile solutions are shown with dashed lines.	51
2-13 Illustrations of several signal to noise ratios at 5Hz. The real component of the field are shown in blue and the imaginary in red.	52
2-14 Model misfit as a function of Signal to Noise Ratio. Median values are shown with a solid line and top/bottom 20 th percentile solutions are shown with dashed line.	52
2-15 Left: True perturbed velocity models consisting of 2-5 master layers. Middle: the best fitting, lowest model NMSE. Right: The 5 th best fitting (20th percentile) velocity model.	54
2-16 Convergence results for Marmousi like inversion as function of number of master layers.	54

2-17 A: Data misfit between PSO inversions and the field generated from the Marmousi model at 3Hz and 5Hz as function of number of master layers. B: Model misfit for the PSO inversion between the smoothed Marmousi model shown in Figure 2-18A left and the smoothed inverted velocity models. For both the plots the median values are shown with solid lines and the top and bottom 20th percentile solutions are shown with dashed lines.	55
2-18 Left: Initial models used for FWI. Right: Final models after FWI. A: Inversion using the true smoothed Marmousi Velocity model. BCD: The initial models based on the best fitting, lowest data measurement error, PSO inversion at 3Hz and the resulting final models from FWI.	56
2-19 Left: Initial models used for FWI. Right: Final models after FWI. A: Inversion using the true smoothed Marmousi Velocity model. B,C,D: The initial models based on the best fitting, lowest data measurement error, PSO inversion at 5Hz and the resulting final models from FWI.	57
3-1 A: The base velocity model consisting of 3 perturbed layers. B: The reflector model is the same as the base model with the addition of an anticline reflector, indicated by the red arrow, in the deepest layer. The velocities within the layers are 1500, 2000, and 2500 m/s from shallow to deepest.	65
3-2 Left: Parameterization for gradient field-expansion. Each layer has an up and down velocity describing the velocity gradient and each boundary is described by N_q points (here 7). Center: The interpolated sub-layer boundaries between the master layers boundaries. Right: The resulting 2D velocity model. A small number of parameters (here 32) describe a complex smoothly varying velocity model that would ordinarily require over 30,000 parameters.	70

3-3	Six representative velocity models drawn from the results of the MCMC algorithm run on the model shown in Figure 3-1. Each velocity model extends to a depth of 3,000 m and 2,500 m horizontal distance. The resulting position of the migrated anticline reflector is also shown. . .	79
3-4	The mean velocity model from the MCMC algorithm for which six random models are shown in Figure 3-3. The solid black lines show the true position of the layer interfaces and reflector. The standard deviations of the layer boundaries are shown with the red dashed lines and the velocity uncertainties for each layer are shown. The mean position of the migrated reflector (at about 2200 m depth) and its standard deviation are also displayed.	80
3-5	Left: Histogram of migrated anticline relative height. Right: Histogram of migrated anticline absolute depth. Both histograms are shown with the same horizontal scale and bin size. The red vertical line denotes true values of the quantities of interest in Figure 3-1B. .	81
3-6	A: 122x384 pixel Marmousi velocity model. B: Smoothed Marmousi model generated by applying a Gaussian low pass filtered to the original Marmousi model. C: Field expansion model based on the smoothed Marmousi model.	84
3-7	Real and imaginary components of the measured field (blue) true simulated field (red) at 3Hz.	84
3-8	6 Velocity models drawn from the Markov-Chain; the true model is shown in Figure 3-6C. Note that deeper sections of the velocity vary more than shallow parts.	84
3-9	Zero-offset migration using the velocity models shown in Figure 3-8 using the true Marmousi reflectivity. Two potential traps are highlighted in blue and red. We characterize the uncertainty of the cross sectional area and depth of these two traps.	85

3-10 Histogram of the shallow (red) and deep (blue) anticlines depth and area. The Metropolis-Hastings runs were initialized with the true starting model. The red vertical line denotes true values of the quantities of interest in Figure 3-10.	85
3-11 Convergence plot for Metropolis-Hastings inversions using 2-5 master layers using 7 nodes per interface. Each plot shows \hat{R} for the 4 quantities of interest as a function of iterations. the 4 models shown had the following number of degrees of freedom: 2-layers 11, 3-layers 21, 4-layers 31, and 5 layers 41.	87
3-12 Histogram of anticline depth and area using a linear gradient initial model. The histograms using the true velocity as the starting model from Figure 3-10 are overlaid in red. The red vertical line denotes true values of the quantities of interest in Figure 3-10.	88
3-13 Left: The final \hat{R} for the 4 quantities of interest as a function of bottom velocity perturbations. Center: Histograms of anticline depth for the a zero perturbation linear gradient starting model with overlays of velocity perturbations of -25%, -15%, +15%, +25%. Right: Histograms of anticline Area for the a zero perturbation linear gradient starting model with overlays.	89
4-1 Geometry of the experiment. Each element acts sequentially as a source with the resulting wavefield recorded on all the elements. A: Transverse acquisition. B: Axial acquisition.	96

4-2	A: True velocity model used for synthetic experiment. B: Smoothed velocity model containing only the upper bone used for Kirchoff migration. C: Kirchoff migration using the smoothed velocity model with true bone interfaces highlighted in red and the silicone tissue interface highlighted in black at roughly 1.5 mm. The artifacts at the source are a normal part of any imaging method that can be suppressed if necessary by muting the direct arrival. Numerous multiple reflections are present in the migrated image within the marrow; these are the lines in the image that don't correspond to true reflectors.	97
4-3	Red: Field for an in vivo measurement across the 96 transducer at a frequency of 2MHz. Blue: Simulated field using the field expansion of a synthetic model based on the in vivo imaging. The upper panel shows the real components of the field and the lower panel shows the imaginary components.	98
4-4	Dashed: The true velocity profile for the synthetic bone experiment. solid: The best fitting velocity model for the 20 PSO inversion. . . .	98
4-5	Blue: Synthetic field across the 96 transducers at a frequency of 2 MHz without noise. Red: Simulated field with small amount of noise used for MCMC inversion.	99
4-6	Top: Three different synthetic velocity models. Middle: Synthetic common source gathers. The middle panel labeled ‘marrow flood - bone flood’ shows the difference between source gathers for the bone flood and marrow flood velocity models. Bottom: NMO semblance values of synthetic data.	103
4-7	A: Cartoon illustrating the three types of internal multiples seen when imaging in the axial geometry. B: predicted peaks for primaries and multiple in NMO semblance domain. Primaries are shown with blue circles; the potential marrow return should fall along the blue line. The asterisks indicate the predicted move out of the multiples.	104

4-8 Illustration of the dictionary construction and column-wise sparsity of the weighting matrix. The figure shows propagators across several receiver indices for a single hyperbolic reflection nmo time velocity pair (t_j, V_k) . For a single reflection all of the non-zero dictionary coefficients are contained in a single column of the weighting matrix. 108

4-9 **A:** Raw synthetic trace with multiples present. **B:** denoised data with mute applied. **C:** ℓ_2 norm of dictionary coefficients along the receiver index direction. Red lines indicate the boundaries of the mute in the NMO time-velocity space. 111

4-10 Migrated images with processed and unprocessed data. True interfaces are highlighted in red. The colormap is the same for all three panels with a dynamic range of 45dB. **A:** Migrated image using raw data. Numerous multiple returns are present that do not correspond to the true model. **B:** Kirchoff migrated image with only predictive deconvolution applied. Predictive deconvolution alone does not mitigate all of the multiples present in the data. **C:** Migrated image with sparse denoising and predictive deconvolution. All interfaces are estimated correctly with the exception of the lower endosteum due to using an incorrect velocity model with an infinite marrow half space shown in Figure 4-2B. The signal and clutter strengths given in Table 4.3 are calculated by averaging over the dashed white line and the pixels within the white box respectively. 113

4-11 In vivo migrated image of the radius. Upper images are generated with anisotropic migration. Lower images are generated with isotropic migration. (A) Migration using raw data. (B) Migration using the sparse dictionary after mute has been applied. (C) migration with hyperbolic sparse denoising and predictive deconvolution data. The Periosteum and Endosteum are labeled ‘P’ and ‘E.’ The signal and clutter intensities are reported in Tables 4.4 & 4.5 by averaging the pixel intensity over the dashed white lines (endosteum) and clutter (white box), respectively. The contrast between clutter and signal strength are given in Tables 4.6 & 4.7.	115
4-12 In vivo common source gather of the radius for a single source (A) raw data. (B) the sparse dictionary after mute is applied. (C) hyperbolic sparse denoising and predictive deconvolution. (D) The difference between A and B. E the difference between A and C. (F) The difference between D and E. The predictive deconvolution suppresses a peg leg multiple indicated by the arrows.	117
4-13 True velocity model for transverse synthetic experiments.	118
4-14 Left: snapshots of wave field propagation in transverse geometry. The white arrow indicates the return from the bottom of the marrow and the black arrow highlights the headwave. Right: common source gather for the simulated gather with the various arrivals and reflected signals labeled. See supplemental video synthMovie.avi	119
4-15 Transverse velocity models used for migrations. From left to right: tissue half space, bone half space, horseshoe.	119

- 4-16 Kirchoff migration results using the tissue half space (A), bone half space (B), and horseshoe models (C). To highlight potential reflectors below the upper endosteum the scale has been saturated for the horseshoe migration. The signal intensity for the upper periosteum, upper endosteum, and lower endosteum were calculated by averaging along the black, white, and green lines. The clutter strength was calculated by averaging over the pixels within the white box. The contrast between the clutter and pixel strengths are given in Table 4.8. 121
- 4-17 RTM results using the tissue half space (A), bone half space (B), and horseshoe models (C). To highlight potential reflectors below the upper endosteum the scale has been saturated for the horseshoe migration. The signal intensity for the upper periosteum, upper endosteum, and lower endosteum are calculated by averaging along the black, white, and green lines. The clutter strength is calculated by averaging over the pixels within the white box. The contrast between the clutter and pixel strengths are given in Table 4.9. 122
- 4-18 Transverse Kirchoff (A) migration and reverse time migration (B) using the smoothed true velocity model. The signal intensity for the upper periosteum, upper endosteum, and lower endosteum are calculated by averaging along the black, white, and green lines. The clutter strength is calculated by averaging over the pixels within the white box. The contrast between the clutter and pixel strengths are given in Table 4.10. 125
- 4-19 Velocity models for the in vivo data in the transverse configuration.
A: tissue flood. **B:** bone flood. **C:** marrow flood. 127
- 4-20 Transverse Kirchhoff migrations using the velocity models in Figure 4-19. **A:** tissue flood. **B:** bone flood. **C:** marrow flood. The signal intensity for the upper periosteum and endosteum were calculated by averaging along the red lines. The clutter strength was calculated by averaging over the pixels within the white box. The contrast between the clutter and pixel strengths are given in Table 4.11. 127

4-21 Transverse reverse time migrations using the velocity models in Figure 4-19. **A:** tissue flood. **B:** bone flood. **C:** marrow flood. Some structure may be visible in the horseshoe model that corresponds to the interior of marrow. The signal intensity for the upper periosteum and endosteum are calculated by averaging along the red lines. The clutter strength is calculated by averaging over the pixels within the white box. The contrast between the clutter and pixel strengths are given in Table 4.12. 128

List of Tables

1.1	Performance and tradeoffs of several types of solvers.	28
3.1	Table showing the \hat{R} values for the 4 quantities of interest initiating the Metropolis-Hastings with starting models close to the true starting model and a linear gradient. An \hat{R} value of less than 1.1 indicates convergence.	86
3.2	Computation time for the field expansion forward models.	90
4.1	MCMC inversion of bone model velocities.	100
4.2	MCMC inversion of depths of bone model.	100
4.3	Synthetic signal and clutter intensity. The proposed denoising scheme improves contrast by roughly 10.8 dB.	114
4.4	In vivo isotropic migration signal and clutter strength from Figure 4-11.114	
4.5	Invivo anisotropic migration signal and clutter strength from Figure 4-11.	114
4.6	Invivo Isotropic contrast between clutter and endosteum signal from Figure4-11.	114
4.7	Invivo anisotropic contrast between clutter and endosteum signal from Figure4-11.	114
4.8	Contrast between marrow clutter and interfaces for synthetic Kirchhoff migration examples using shown in Figure 4-16.	121
4.9	Contrast between marrow clutter and interfaces for synthetic reverse time migration examples using shown in Figure 4-17.	121

4.10 Contrast between marrow clutter and interfaces for the synthetic migrations shown in Figure 4-18.	126
4.11 Contrast between marrow clutter and interfaces for the in vivo Kirchhoff migrations shown in Figure 4-20.	126
4.12 Contrast between marrow clutter and interfaces for the in vivo reverse time migrations shown in Figure 4-21.	126

Chapter 1

Introduction

Uncertainty quantification (UQ) can help answer a fundamental and often overlooked question in solid Earth geoscience, ‘do we know what we think we know?’ The error bars found in other fields are frequently not found in geosciences, particularly seismic imaging. Geophysical imaging (particularly global scale) is unique in that the results are often not verifiable by other imaging modalities or by physically verifying the subsurface through drilling or trenching, leaving room for over interpretation of results. Without UQ we could be arguing in the noise and conflicting choices could be equally likely to be right or wrong. Significant resources are spent based on theoretical explanations of Earth’s interior that could be based on erroneous or flimsy interpretations of noisy subsurface measurements. Providing error bars on the estimates could decrease the allocation of these resources from unknowable problems or aspects of the subsurface and better focus these resources on explaining aspects of the Earth that are actually constrained by observable data.

If a region of the subsurface is in fact poorly resolved with existing measurements and surveys, UQ allows us to put rigorous limits on potential survey designs. On the global scale we characterize the probability distribution of potential sources relative to a seismometer network and use UQ to justify and develop future locations of seismometers that will improve the ability to resolve specific features of the Earth. Furthermore, UQ allows us to know if a future survey is well designed and will be able to resolve a region or question of interest.

In order for UQ to be valuable, there needs to be a cost of being wrong or a potential decision made on the results of an answer. In exploration seismology it is possible to assign a dollar value to a potential oil trap and the cost to drill it. Drilling an erroneous or bad reservoir can have an extremely high financial cost. Seismic UQ can estimate a range of values for the volume of a potential trap and guide the selection of well sites. For example, one may choose to drill a more certain but lower yielding trap vs a less certain but potentially higher return well. See [51] for a study of the importance of UQ in exploration geophysics.

Other fields related to geoscience, such as climate science, weather, and hydrology, have a much more developed sub discipline of UQ. This is likely because these fields can impact policy and planning decisions, such as development of water reservoirs, dams, or additional pumping stations. In the more regional or global scale geoscience there are fewer practical impacts the inferred results can have on day to day life and geohazards that tend to effect localized regions. In addition, it is unclear how much seismic imaging can aid in the prediction or risk mitigation.

1.1 Uncertainty Quantification Challenges

Uncertainty quantification is invaluable for building knowledge but is it rarely done. There are numerous computational, visualization, and interpretation challenges that prevent UQ from widely being adopted and the perceived reward may be too small to overcome these challenges. In addition, telling a story with UQ is difficult and can be less compelling, instead of a single objective history of what happened we have a population of unreliable narrators describing their versions of the past. In the context of seismic imaging and velocity model building, UQ describes the distribution of velocity models and seismic images given recorded data. In the fullest extent this would mean characterizing a non-linear distribution consisting of thousands or millions of variables, one for every pixel or voxel in a velocity model. Clearly analyzing this uncertainty in high dimensional space would be challenging.

1.1.1 Large dimensionality and visualization challenges

Even if it is possible to calculate uncertainty on such a large variable space, it is extremely challenging to display this uncertainty in meaningful way and it is unclear how to use this uncertainty estimate in downstream processing. Putting error bars on an image is not a straight forward task. If one assumes that uncertainty of a seismic velocity model is uncorrelated from pixel to pixel and thus can be described completely by the diagonal entries of a covariance matrix, it would be possible to display this as an image similar to an illumination map. However, these visualizations would not show any of the correlated errors that can be extremely significant, especially in height. For example, as shown in chapter 3 an erroneously high velocity or low velocity region can causes areas below to shift up and down but relative height or shape may be far more consistent. Alternatively as proposed by [70] and shown in chapter 3 we can instead view a series or movie of representative samples drawn from distribution of subsurface imaging to get a general sense of uncertainty.

Conveying a general sense of uncertainty in an image is extremely challenging and the UQ framework is likely the most useful when asking a specific question about the subsurface such as possible existence of hotspot, the angle of subduction slab, or the volume of potential trap where the UQ can be reduced to a low dimension. Figure 1-1 summarizes this workflow from raw data to defining uncertainty over a quantity of interest. However, asking a specific question about the subsurface requires a general sense of what is in the Earth and global optimization methods described in chapter 2 may be a better way to get an initial maximum Likelihood estimate or best guess solution of the subsurface. Once this ‘best’ image is initially analyzed and interpreted, specific questions can be formulated and we can develop a definitive UQ framework to answer those questions.

1.1.2 Computational Challenges & Expensive Forward Solves

Due to the non-linear nature of the wave equation and expensive forward solvers it is extremely computationally challenging to calculate an accurate distribution of

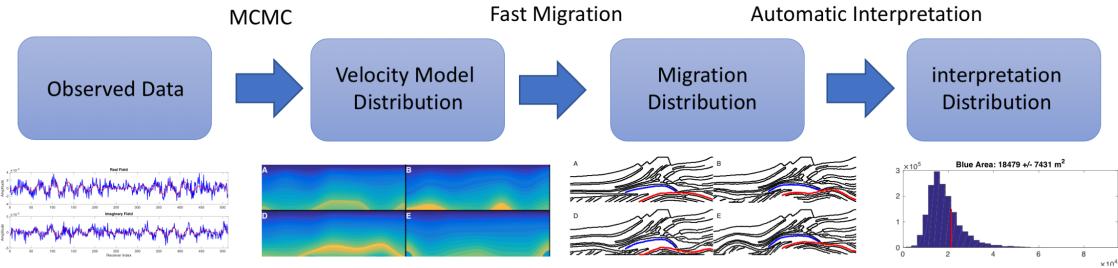


Figure 1-1: Diagram illustrating the seismic uncertainty framework from raw data to probability distribution over a quantity of interest and final interpretation.

velocity models given data. Markov Chain Monte Carlo (MCMC) methods and other non-linear UQ methods typically require far fewer degrees of freedom, 10s to 100s at most, than those used to describe even a modest 2D or 3D velocity model [30]. Table 1.1 summarizes the limitations and advantages of several types of solvers used in seismic imaging and Full Waveform Inversion (FWI). In addition, these UQ methods require tens to hundreds of thousands of forward solves, when typically less than a hundred forward solves are used for conventional FWI.

In order to provide an estimate of the velocity model distribution typically one of two assumptions are made: Gaussian distribution or reduced parameter space. First, one can assume that the velocity model distribution is well approximated by a Gaussian. This assumption essentially linearizes the wave equation about some point estimate and the covariance of the Gaussian matrix can be described through the Hessian [90]. One problem with this assumption is that it is likely impossible to store the full covariance matrix as this is of size n_{pixels}^2 where n_{pixels} is the number of pixels or voxels in the velocity model. However, there are methods that can draw representative solutions from a Gaussian distribution without explicitly generating a covariance matrix [19]. Although this approximation is computationally feasible, the assumption of a Gaussian distribution does not appear accurate as demonstrated in chapter 3. However, this assumption may be valid in low-noise scenarios in which the minima are well behaved or in scenarios where you have an excellent initial model, like time lapse seismic.

In order to allow for non-linear and non-Gaussian descriptions of the velocity model distribution, the number of degrees of freedom must be reduced significantly. Once the degrees of freedom are reduced, this parameterizations can be combined with fast forward models to run the tens or hundreds of thousands of forward solves needed to converge to the true velocity model distribution using MCMC. Depending on the particular problem, numerous methods exist to decrease the degrees of freedom [91, 12, 11, 66]. However, the challenge in reducing the degrees of freedom is choosing a parameterization that accurately models the physics of the inverse problem with fewer variables. The main drawback of this approach is that by design the fidelity of the model is significantly lower than the pixelized or voxelized models typically used in seismic imaging, and the scale of uncertainty attempting to be characterized may be significantly smaller than the reduced parameterization. Although numerous researchers have adopted this approach, few have attempted to find reduced parameterizations that allow for an accelerated forward solver such as the one explored in this thesis.

1.1.3 Interpretation & Human bottleneck

Even if you are able to generate an UQ estimate for the velocity model and seismic image, most seismic imaging questions require human interpretation of a final image. If we adopt an ensemble approach in which the distribution is characterized by a collection of samples, we typically need to interpret thousands of models to characterize uncertainty for a single velocity model. This approach is likely too onerous and some form of automation is needed to develop a full framework for seismic imaging UQ. In chapter 3 we use a very simple form of automatic interpretation that is relatively easy to implement and execute on thousands of models. Alternatively Machine Learning (ML) has recently been used to automate several interpretation tasks [29, 83]. Instead of simply using ML as a means of speeding up interpretation, ML could enable interpretation on tens of thousands of images that would be infeasible to do manually. This could allow for performing UQ on highly subjective interpretations that would be impossible to automate in a conventional manner.

	Gradient Methods	Global Optimization	MCMC
Number of variables:	$10^3 - 10^6$	$10^2 - 10^3$	$1 - 10^2$
Number of iterations:	≤ 50	$10^2 - 10^4$	$10^4 - 10^6$
Robust to Local minima:	No	Yes	Yes
Uncertainty Quantification:	No	No	Yes

Table 1.1: Performance and tradeoffs of several types of solvers.

1.2 Thesis Overview

This thesis presents a possible method for performing UQ for seismic imaging. It does so by combining an extremely fast method of solving the Helmholtz equation [48] with a reduced parameterization that can accurately characterize the velocity model with a fraction of the number of variables typically used to describe a 2D velocity model. Unlike other UQ methods, this thesis exploits the structure of the parameterization to generate very fast forward solves of the Helmholtz equation. This speed of forward model and parameterization allows us to compare several types of algorithms against each other that would be impossible to compare without the ability to solve the forward model rapidly.

In some cases performing a rigorous non-linear UQ analysis may be unnecessary or impossible to perform due to slowness of the forward solver or requiring a model have more degrees of freedom than possible for UQ. In addition, if there is little noise in the observed data, having a maximum likelihood estimate or best fitting model may be sufficient as illustrated in Figure 1-2. For example, once the maximum likelihood estimate is found, a local Gaussian approximation of the distribution in Figure 1-1 would likely be sufficient and MCMC methods be unnecessary. However, the wave and Helmholtz equations are non-linear and even in low noise environments local minima will trap gradient based solvers and lead to premature convergence to an incorrect answer.

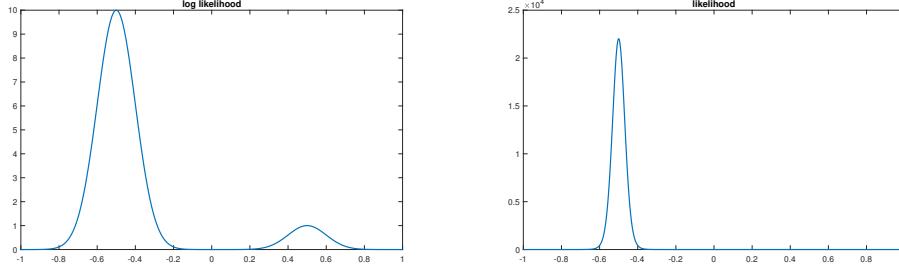


Figure 1-2: **Left:** Log-Likelihood distribution for a fictional function. **Right:** Likelihood for the same fictional problem. The local minima present in the right panel would create a problem for local optimization methods but does not contribute significantly to the likelihood function and does not impact the distribution.

1.2.1 Global Optimization

In the second chapter, we explore the application of global optimization methods to estimate an initial starting model without an initial guess. Furthermore, the speed of the forward solver allows us to benchmark several existing global optimization algorithms and study the effects of various parameters on convergence. Due to the cost of forward modeling steps typically used in seismic imaging, numerical studies of parameter impacts on global optimization methods have not been explored. In addition, we demonstrate how these inverted starting models can impact the final FWI results. In this chapter we also discuss in detail how to modify the field expansion method so that it can generate data that are comparable to finite difference simulations.

1.2.2 Uncertainty Quantification

In the third chapter of the thesis we use the field expansion method combined with MCMC methods to estimate the distribution of velocity models given recorded data. In this chapter we demonstrate the full framework from raw data to quantifying the specific question of interest, such as the cross sectional area and depth to a potential trap. Using this distribution of velocity models we are then able to generate probability distributions of seismic images and form probabilistic answers to questions about the seismic images. In addition, we explore the limitation on the number of

degrees of freedom we can realistically use for UQ.

1.2.3 Ultrasound and Bone

In the fourth chapter I attempt to apply the methods presented in the first two chapters to real data from medical ultrasound. This data was originally desirable as its layered geometry, lack of direct arrival, and lack of surface multiples made it an attractive data set to apply my techniques. However, the acquisition geometry (under sampled in the offset direction) and unmodeled physics (strong anisotropy), and high noise made these techniques difficult to apply and generate good results. While attempting to apply the more exotic techniques developed in the first two chapters of the paper, I explored the application of a number of conventional seismic techniques to the data set in order to understand what was actually contained in the datasets. In an attempt to understand the data and difficulty of imaging geometry we developed several novel techniques to image the interior of bone that are presented in this chapter.

Chapter 2

global optimization algorithms for building velocity models

2.1 Overview

In this chapter, we use a fast Helmholtz solver with global optimization methods to estimate an initial velocity model for full-waveform inversion based on raw recorded waveforms. More specifically, we combine the field expansion method for solving the Helmholtz equation with a reduced parameterization of the velocity model allowing for extremely fast forward solves of realistic velocity models. Unlike conventional Full Waveform Inversion that uses gradient based inversion techniques, global optimization methods are less sensitive to local minima and initial starting models. However, these global optimization methods are stochastic and are not guaranteed to converge to the global minima. Because our adaptation of the field expansion method is so fast, we can study the convergence of these algorithms across parameter choice and starting model. In this chapter we compare two commonly used global optimization methods, particle swarm optimization, and simulated annealing, and examine the limitations of these algorithms and their dependence on choice of parameters. We find that PSO out-performs SA and that PSO converges reliably for noisy data provided the number of parameters remains small. In addition, using this methodology we are able to estimate reasonable FWI starting models with higher frequency data.

2.2 Introduction

Most seismic techniques such as imaging, rely on a velocity model inverted from noisy data through a non-linear inverse problem. These inverted velocity models may be inaccurate and lead to incorrect interpretations of the subsurface. For example, an erroneously fast or slow section of a velocity model could cause a syncline structure to appear as an anticline and be incorrectly interpreted as a potential trap. Full Waveform Inversion (FWI) and other gradient based methods will converge to local minima [81] and will lead to inaccurate velocity models if initiated with an inaccurate starting model [79]. In recent years UQ methods have been applied to the FWI problem in order to mitigate the impact of local minima and provide a more probabilistic interpretation of the sub surface [66, 90, 55]. See [51] for a discussion of importance of risk assessment and Uncertainty Quantification (UQ) in seismic imaging. These methods allow for an informed decision about the reliability of the subsurface image and aid in risk estimates for drilling a potential well. However, these methods either require hundreds of thousands of wave equation forward solves to converge and give accurate error estimates of subsurface parameters [17] or use local approximations of a Gaussian to approximate the uncertainty [19]. The number of forward solves required for MCMC is likely infeasible for many seismic imaging problems and forward models. In addition in low noise scenarios there may only be a single region of high probability and MCMC techniques may be unnecessary. Instead, global optimization methods are an attractive alternative as they require far fewer forward solves and provide an estimate of the best fitting model (maximum Likelihood) at a fraction of the number of forward solves needed for UQ and yet can still avoid being trapped by local minima.

In recent years global optimization methods such as Simulated Annealing (SA) [12, 21, 61] and Particle Swarm Optimization (PSO) [16, 64] have become popular methods of calculating an initial velocity model. These methods can be an alternative or to supplement conventional methods of building velocity models such as tomography, NMO velocity analysis or very low frequency FWI [82]. Conventional

initial velocity building methods typically rely on gradient based minimization and often require the manual picking of arrival times or picks in semblance space. By contrast, global optimization methods, although not guaranteed to converge to the global minima, are far more robust to local minima and frequently do not require the calculation of a gradient, reducing memory constraints and algorithmic complexity. However, starting models, frequency, and other parameter choice can have a significant impact on the convergence of these algorithms. Although these algorithms require significantly fewer iterations than UQ algorithms, rigorous comparison requires numerous global optimization runs and thousands or tens of thousands of forward solves. In addition due to the stochastic nature of these algorithms, two different runs with the same starting model can converge to different final models, requiring even more simulations to understand performance. Due to the computational cost of most forward solvers, previous work has been unable to accurately characterize these impacts and benchmark these algorithms as significant simplifications were necessary to control this computational cost. For example, [60] use analytical solutions to test problems as well as the 1D wave equation to compare global optimization algorithms. In another attempt to make the problem computationally tractable, previous work has typically shown performance with one or only a few starting models when tens or hundreds of runs are necessary to generate a statistical model of convergence and accurately compare one algorithm to another [57]. Studies in the global optimization literature use test functions that are extremely fast to evaluate but are not representative of local minima found in FWI and seismic imaging. Other work combining global optimization methods and FWI have lacked a sufficiently fast test function that accurately accounts for the physics of seismic inversion in 2D.

In this chapter, we use the reduced model parameterization of velocity models introduced in [20, 88] and [12] combined with two global optimization methods to estimate an initial velocity model. Due to the rapid speed of our forward model and our approach we can quantitatively characterize the limits on the number of degrees of freedom, importance of initial guesses, choice of tuning parameters, comparison of different global optimization methods and impact of source frequency. These com-

parisons, would be impossible with conventional slower finite difference solvers.

From the numerical studies presented in this chapter we are able to draw several conclusions about the performance of global optimization methods for FWI and what may impact the likelihood of finding a good initial model. Although the FWI problem is known to be more prone to local minima at higher frequency [81], we find that source frequency does not significantly decrease the likelihood of convergence and a good starting model can be found with a relatively high frequency of 5 or 8 Hz from raw waveform data. This suggests we could build initial background velocity models with much higher frequencies than are typically used. We also find that the accuracy of the initial starting models did not have a significant effect on the likelihood of converging to the correct model. From our results we find that only having an extremely accurate initial model improved the likelihood of convergence and that if the initial model(s) are outside this range the likelihood of convergence is very weakly dependent of starting model. This suggests that only if the initial model is in the basin of attraction of the global minima are we guaranteed to converge to the true solution and outside of this range the global optimization methods are equally likely to find an accurate solution independent of starting model. In addition, we find that our implementation of Particle Swarm Optimization (PSO) out performs Simulated Annealing (SA) and is able to find a more accurate initial model with the same number of forward solves.

The remainder of the chapter is organized as follows. We first describe a reduced parameterization of the velocity model that is compatible with the field expansion forward solver. We then demonstrate that this first forward model and parametrization can yield comparable results as finite difference simulations if modifications are made to the forward solver. Second, we describe the two global optimization methods used in the chapter: Particle Swarm Optimization (PSO) and Simulated Annealing (SA). Finally, we demonstrate our inversion on several synthetic models of varying complexity to determine the limitation on number of degrees of freedom, choice of initial model, and iterations needed. The field expansion method achieves significant computational saving through restricting the velocity model to consist of series of

non-overlapping layers. Although this parameterization severely restricts the velocity models we can simulate, we find that the model is sufficiently accurate to estimate initial models for FWI as demonstrated later in the chapter.

2.3 Forward Model

Unlike gradient based minimization methods, global optimization and uncertainty quantification methods require far fewer degrees of freedom, tens vs thousands. In this section we briefly describe the reduced parametrization previously used by the authors in [17] for uncertainty quantification. The reduced parameterization is compatible with the field expansion method [48], allowing for extremely rapid forward modeling of the Helmholtz equation. We briefly describe the field expansion method and the modifications we make to the forward solver to mitigate several of the artifacts inherent to the field expansion method.

2.3.1 Reduced parameterization

The field expansion method is able to achieve rapid forward solves by heavily restricting the velocity model to consists of a number of non-overlapping piece-wise constant layers as shown in Figure 2-1. Although the Earth has often been approximated by series of layers, this parameterization of the velocity model is likely too restrictive for most applications. In addition, parameterizing each layer individually as a variable would require many more degrees of freedom than is compatible with global optimization methods. A more realistic parameterization introduced in [20] is to approximate the Earth as a series of curved interfaces with a velocity gradient between the many interfaces. This parameterization, illustrated in Figure 2-2 allows for complex velocity models with only a few degrees of freedom and has successfully been used for global optimization [12], tomography [88], and for uncertainty quantification [17]. [12] use a similar parameterization for global optimization with finite difference methods and thus are significantly limited in computation time. Through combining this parameterization with the field expansion, we can invert the velocity models in a fraction of

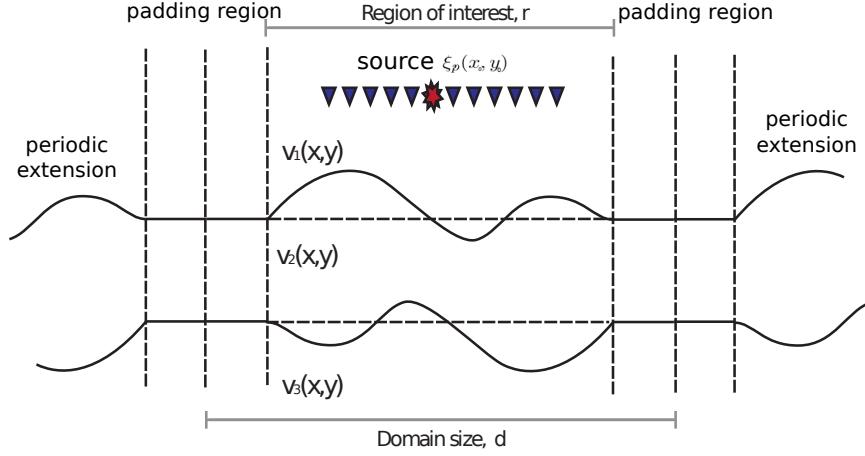


Figure 2-1: sketch of field expansion velocity model and repeating boundary conditions.

the time. [17] argues that the field expansion forward solver is roughly 30 to 40 times faster than finite difference methods depending on the model complexity.

We build a perturbed layered velocity model following from [17] and illustrated in Figure 2-2A. The velocity model consists of M master interfaces each with N_q control points which control the relative height of interface across horizontal position. The full position of interfaces are interpolated with cubic splines between the control points. The velocity inside layer i from interface $i - 1$ to i is a linear gradient from \mathbf{V}_i^u to \mathbf{V}_i^d increasing with depth. Although we cannot achieve a true gradient with the piece-wise constant constraint of the field expansion, we can approximate this gradient by dividing each layer into numerous constant velocity sub-layers as shown in Figure 2-2B. This parameterization allows us to generate fairly complex velocity models with a limited number of parameters. For a model with 4 layers and 7 control points per interface we require only 32 degrees of freedom compared to the tens of thousands needed for a pixelized velocity model as shown in Figure 2-2C.

2.3.2 Field Expansion

In this section we briefly describe the field expansion method for solving the Helmholtz equation, see [48, 16] for a more detailed description of the method. The field ex-

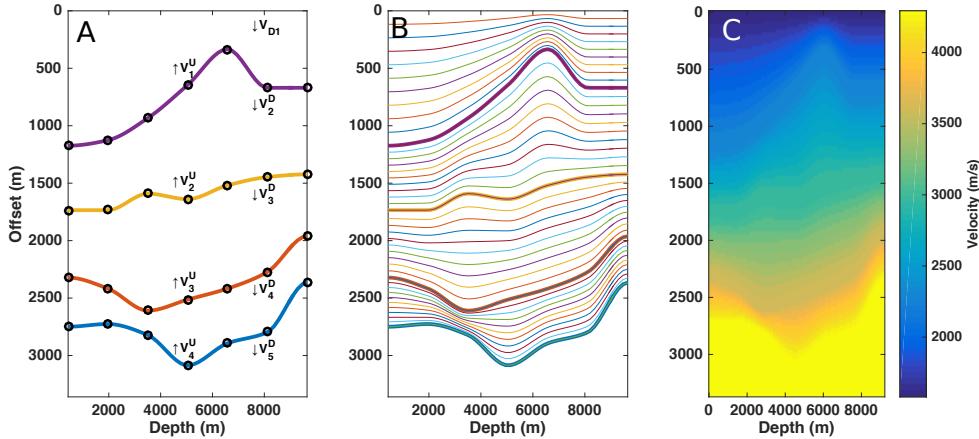


Figure 2-2: **A:** Reduced parameterization diagram. **B:** Interpolated layer interfaces based on the reduced parameterization. **C:** Pixelized velocity model derived from reduced parameterization consisting of piecewise constant layers.

pansion method solves the Helmholtz equation for periodic boundary conditions as illustrated in Figure 2-1. The method achieves significant computational saving by taking the analytic solution to the scattered field for a series of flat horizontal layers and then using series expansions to calculate the scattered field for non-flat interfaces. The source and model repeat infinitely in the horizontal direction. These boundary conditions and source configurations vary significantly from the boundary conditions typically used in seismic imaging. In the following section we discuss how we modify the velocity model and field expansion method to make the results comparable to finite difference methods with perfectly matching layers frequently used in seismology.

In the field expansion the field within a $v_m(x, y)$ layer subject to a periodic point source is given by

$$\xi_p(x, y) = \frac{1}{2id} \frac{e^{i(\alpha_p(x-x_0)+\beta_p|y-y_0|)}}{\beta_p} \quad (2.1)$$

where,

$$\alpha_p = \alpha + (2\pi/d)p \quad \beta_{j,p} = \begin{cases} \sqrt{k_m^2 - \alpha_p^2} & \alpha_p^2 < k_m^2 \\ i\sqrt{\alpha_p^2 - k_m^2} & \alpha_p^2 > k_m^2. \end{cases} \quad (2.2)$$

where d is inter domain spacing, p is the spatial mode number and k is the wavenumber of the source, $k = \frac{2\pi f}{v_1}$. From Equations 2.1 & 2.2 we see that the source becomes singular when β_p approaches infinity which occurs at particular combinations of wavenumber and spatial mode. This singularity corresponds to a physical resonance of periodic boundary condition known as Rayleigh or woods anomalies, see e.g. [49]. These Woods anomalies, generate high energy artifacts at specific frequencies that do not resemble anything in seismic imaging. These anomalies are resonances excited by the periodic source and velocity model.

To mitigate these artifacts we increases the domain spacing to be several times the size of interest. For example, if our velocity model is 3 km in offset, we set the domain size to be 15km. This padding of the domain decreases the amount of energy leaked from one repeating domain to the other. In addition, we make the velocity model slightly absorptive and dispersive by adding a complex term to the velocity. This prevents the singularity in Equations 2.1 & 2.2 and the Wood's anomalies are no longer present.

To demonstrate the effectiveness of this strategy we generate a synthetic velocity model consisting of flat layers and synthetic data at 3,5,8, and 10 Hz as shown in Figure 2-3. We then use the field expansion method to generate synthetic data in which we add a dispersive term from 10^{-4} to 10^{-1} and compare the field expansion simulation to data generated with PySIT's Helmholtz solver [32]. Figure 2-4 shows the mean squared error between the finite difference and field expansion data as a function of the dispersive term. From Figure 2-4 we see that a large misfit is apparent if no dispersive term is added and the misfit is minimized with a dispersive term of approximately $10^{-2.5}$ for a frequency of 3 Hz. From this figure we also see that this misfit tends to increase as a function of frequency and therefore it may be more useful to use the field expansion method at lower frequencies. These discrepancies appear to be comparable or less than the differences reported between several time domain finite difference solvers [69]. In addition, although the Normalized Mean Square Error (NMSE) increases as a function of frequency, from Figure 2-3 the field expansion phase agrees well with the PySIT forward solver and only varies in amplitude. Instead of

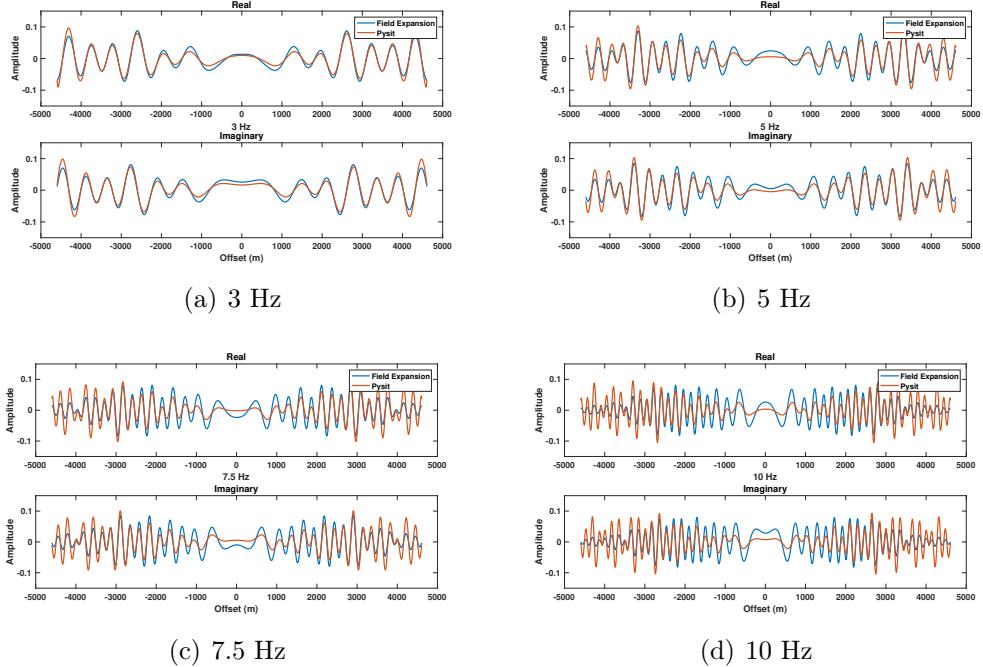


Figure 2-3: Real and imaginary components of the PySIT and field expansions fields at several frequencies. We generate these results with a dispersion term of 2.5×10^{-2} .

using a naive NMSE as an objective function, using an objective function that only measures difference in phase would likely provide more accurate inversions such as the zero-lag cross correlation objective function used in [46].

2.4 Global Optimization Methods

Although there are numerous varieties of global optimization methods [63], variants of Simulated Annealing (SA) and Particle Swarm Optimization (PSO) have gained popularity for seismic velocity model inversion. In their most basic form, PSO and SA solve a global optimization problem with a fixed number of continuous variables and do not attempt to optimize over degrees of freedom or solve a combinatoric problem. PSO is a parallel algorithm in contrast to simulated annealing which is a serial or sequential process. However, parallel variants of SA exist [54]. In this section we briefly describe the two algorithms.

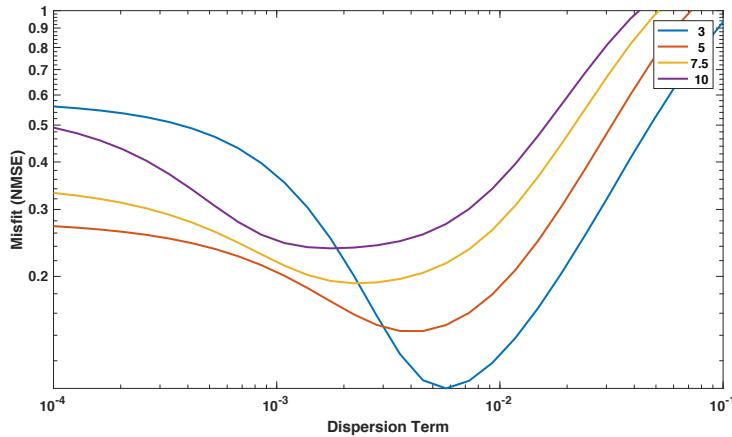


Figure 2-4: Dispersion as a function of misfit for several source frequencies.

2.4.1 Particle Swarm Optimization

In this section we briefly describe the global particle swarm optimization method for estimating a velocity model. The PSO algorithm is based on the behavior of flocks of animals [14]. In the algorithm a collection of agents move through search space and communicate their results to one another to ideally guide the search towards a global minima.

In PSO, the agents are given a random starting position in parameter space with a random agent velocity at which they move through the search space. At every iteration, each agent evaluates the forward model with parameters according to its current position and calculates a cost function between the measured and simulated field. In our case, we use the normalized mean squared error between the measured and observed field but any cost function could be used. Each agent as it moves through parameter space keeps track of the lowest scoring or best solution (personal best) it has visited and the best solution of visited by all agents (global best). Once the score of the new position is updated, the agent's velocity is calculated based on the location of the updated personal and global best such that the agent is accelerated towards to the local and global bests. Algorithm 1 summarizes the PSO algorithm for a swarm consisting of N_s agents with their position vector \mathbf{x} and agent velocity vector \mathbf{v} . Note that in this algorithm the evaluation of each agent's cost function

is independent from one another and the forward solves can be trivially parallelized across all agents. To apply the global optimization algorithms to the reduced forward model parameterization we must convert the field expansion velocity model to a vector of values ranging from -1 to 1. This is achieved by putting mapping the minimum and maximum permissible velocities and layer depths to the range of -1 to 1. In the expression for the agent velocity update in Algorithm 1, γ is the inertia term controlling how much the previous agent velocity is maintained from the prior iteration and a_g and a_p are the acceleration terms that determine how much the personal best and neighborhood best alter each agent's velocity at each iteration. For all of the simulations performed in this chapter the following parameters are used: $\gamma = .9$, $a_g = 1.49$, and $a_p = 1.49$. If the agent velocity exceeds .05 of the total search space then it is limited to .05 of the search space. This hard limiting of the speed of exploration prevents the search space from being explored too granularly and prevents the global optimization algorithm from missing a potential minima.

Algorithm 1 PSO

```

Initialization
for  $i = 1$  to  $N_s$  do
    Initiate personal best score to infinity:  $p_i^s = \infty$ 
    Initiate swarm position from a distribution across all of model space:  $\mathbf{x}_i = \mathbf{U}[-1, 1]$ 
    Initiate swarm velocity from a uniform distribution in some range:  $\mathbf{v}_i = \mathbf{U}[-vel_{max}, vel_{max}]$ 
end for
for  $i = 1$  to maxiterations do
    for  $i = 1$  to  $N_s$  do
        Evaluate forward solve and calculate agent's score:  $\mathbf{G}(\mathbf{X}_i)$ 
        Record personal best score and location if better than current best: if  $p_i^s < \mathbf{G}(\mathbf{X}_i)$  then  $p_i^s < \mathbf{G}(\mathbf{X}_i)$  and  $\mathbf{p}_i = \mathbf{x}_i$ 
        Determine neighborhood best location.  $\mathbf{g}_i \leftarrow \min(p_1^s, \dots, p_{N_s}^s)$ .
        Calculate new position:  $\mathbf{x}_i = \mathbf{x}_i + \mathbf{v}_i$ 
        Clamp velocities: If any entries of  $v_i$  exceed  $vel_{max}$  or  $-vel_{max}$  then set the
        entries to  $vel_{max}$  or  $-vel_{max}$ 
    end for
end for

```

2.4.2 Simulated Annealing

Simulated Annealing is a sequential global optimization method that is based on MH algorithms used for MCMC estimates for posterior distributions [75]. Unlike MCMC methods, simulated annealing does not strictly provide uncertainty estimates and only provides a maximum likelihood estimate or global minima from a series of proposed velocity models m_0, \dots, m_N [70]. Much like the standard MH sampler, an initial model m_{cur} is chosen and its likelihood (or misfit) is calculated, a proposal model m_P is selected and its likelihood is also calculated by,

$$L(\mathbf{m}) \equiv p(\mathbf{d} | \mathbf{m}) \propto \exp \left[-\frac{1}{2} (\mathbf{f}(\mathbf{m}) - \mathbf{d})^\top \Sigma^{-1} (\mathbf{f}(\mathbf{m}) - \mathbf{d}) \right]. \quad (2.3)$$

Where f is the forward solver, \mathbf{d} , is the measured data, \mathbf{m} is the proposed model, and Σ is the measured noise covariance. In the case of a noiseless observation we can set Σ to be the identity matrix. If the proposed model's likelihood, l_* , is greater than the current model likelihood, l_{i-1} , then the model is accepted. Otherwise the less likely proposed model is accepted with a probability of,

$$\frac{l_* / l_{i-1}}{t_i} \quad (2.4)$$

where t_i is the annealing temperature that follows an exponential cooling schedule,

$$t_i = t_1 * T_r. \quad (2.5)$$

where t_1 is the initial temperature that controls the initial chance of a bad solution being accepted and T_r is the cooling that controls how quickly bad solutions are rejected as a function of iteration. From the acceptance criteria, we see that when the temperature is higher the sampler is more likely to accept proposals that poorly fit the data and potentially climb out of local minima. As the temperature decreases the sampler is less and less likely to accept a proposal that is worse than the current one. In Algorithm 2 unlike MH, the temperature changes as a function of iteration under an exponential cooling schedule. The standard SA algorithm given in Algorithm 2

is identical to the MH procedure except that the acceptance probability changes as a function of iteration and the proposal distribution. Numerous variations exist for cooling schedules and proposal distributions [34] and several of the variations have been used to estimate initial velocity models in seismic imaging [63, 12, 85]. In this chapter we choose to use simulated annealing with a exponential cooling schedule and Gaussian proposal distribution.

Algorithm 2 Simulated Annealing

```

Initialize starting model:  $\mathbf{m}_0$ 
Initialize starting temperature:  $t_0$ 
Initialize cooling ratio:  $T_r$ 
for  $i = 1$  to maximum iterations do
     $\mathbf{n} \leftarrow \text{Normal}(0, \mathbf{C}_i)$  proposed model perturbation
     $\mathbf{m}_* \leftarrow \mathbf{m}_{i-1} + \mathbf{n}$ 
     $l_* \leftarrow L(\mathbf{m}_*)$  proposed model score
     $\alpha_i \leftarrow \frac{l_*/l_{i-1}}{t_i}$  acceptance probability
     $\text{coin} \leftarrow \text{Bernoulli}(\alpha_i)$  “unfair” coin flip
     $t_{i+1} = t_I \times T_r$  Decrease temperature.
    if  $\text{coin} = 1$  Accept/reject proposed model based on flip result then
         $\mathbf{m}_i \leftarrow \mathbf{m}_*$ 
    else
         $\mathbf{m}_i \leftarrow \mathbf{m}_{i-1}$ 
    end if
end for

```

2.5 Results

In this section we present the results of using PSO and SA to invert for an unknown seismic velocity model. We first test the inversion methods on a variety of simple flat layer velocity models to test the limitation on the number of degrees of freedom, reliability of inversion, importance of source frequency and noise. Based on the values gleaned from these simplified runs, we test the limitations of these algorithms on more complex velocity models and perform an inversion on data generated with a finite difference forward solver from a gridded Marmousi velocity model shown in Figure 2-5.

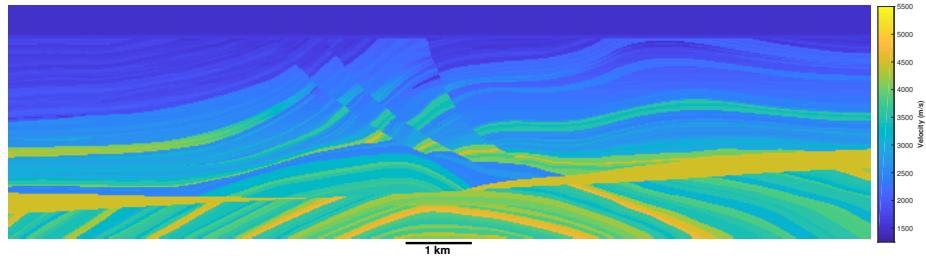


Figure 2-5: Marmousi constant density velocity model.

2.5.1 Simple models

To test the performance and limitations of the global optimization methods we generate a set of flat layered velocity models using 2 to 10 master layers with 10 sub layers per gradient. This velocity model is a rough approximation of 1D gradient of the Marmousi velocity model shown in Figure 2-5. Using, a flat layer model allowed us to rapidly test the dependence of our approach on numerous factors such as: total degrees of freedom, accuracy of initial models, and source frequency. Using a flat layered velocity model allows us to compute forward solves extremely quickly as the Taylor series order can be set to zero. This accelerated forward solve allows us to more easily test and tune parameters.

Number of layers

To test the limitations on the number of layers, we generate a set of true velocity models consisting of flat layers roughly based on the Marmousi velocity model consisting of 2-10 master layers with 9,200 meters of offset and at a frequency of 5 Hz. The dashed line in Figure 2-6AB top shows the true velocity profile for a 4 layer example.

For each number of master layers, we initiate PSO with 48 agents from a uniform distribution centered around the true velocity model. The velocity gradients, \mathbf{v}_n^u & \mathbf{v}_n^d , in the distribution are limited to +/- 2,500 m/s around the true velocity model and master interfaces, \mathbf{d}_n , +/- 1,000 m of the true velocity model. Once initialized,

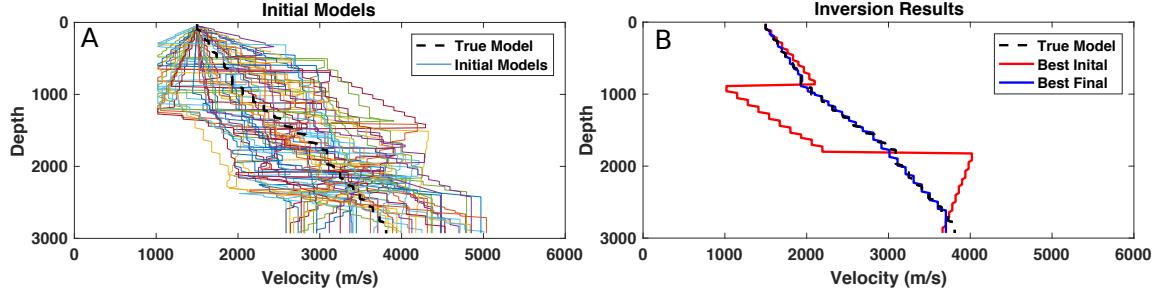


Figure 2-6: Single run of PSO inversion with 4 master layers. **A:** All initial models and the true velocity profile. **B:** Initial best fitting and final best velocity profile after 250 iterations.

we run the inversion for 250 iterations resulting in $48 \times 250 = 12,000$ evaluations of the forward model. The velocity at the surface for each agent is set to the true value of 1,500 m/s to reflect the fact that the surface velocity is generally known or can easily be picked from the direct arrival. Figure 2-6 top left shows the initial velocity models and true velocity model as well as the initial best fitting and final best fitting velocity model for a single run. From this figure, we see that although the best initial velocity model is far from the true velocity after 250 iterations the model has converged to the true model. PSO is a stochastic algorithm and performance can vary from run to run and initial starting models. We therefore repeated this inversion procedure 25 times for each number of master layers drawing different initial models from the same uniform distribution, keeping the number of layers fixed for each set of 25 runs at each number of master layers. Figure 2-7 shows the median error for each of the 25 runs as a function of number of master layer and the two red lines show the misfit of the 20 top/bottom percentile solutions, the 5th best and 5th worst fitting out of a possible 25 models according to NMSE of the velocity model. We find that for up to 4 master layers nearly all PSO inversions converge to the true model independent of the initial model. However, beyond 4 master layers the likelihood of convergence to the true velocity significantly degrades.

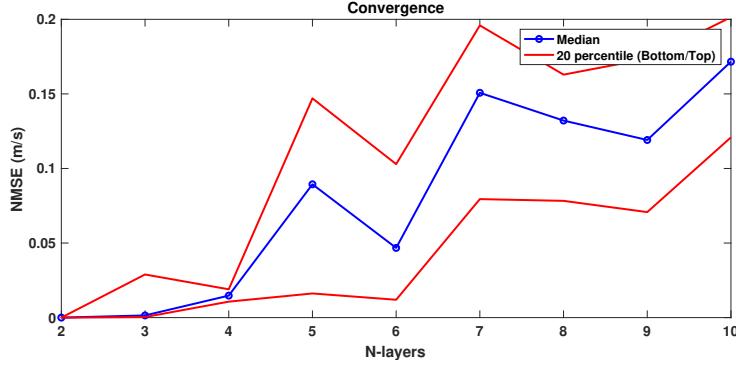


Figure 2-7: Particle Swarm Optimization Convergence as a function of number of layer.

Simulated Annealing

To fairly compare SA to PSO we took all of the 1,200 (48 agents \times 25 runs) initial models for each number of master layers and ran SA for the same number of iterations, 250 iterations. For the inversion we choose a cooling schedule of $.99^t$, an initial temperature of $1 * 10^{-7}$ and a Gaussian step-size of $.5 * 10^{-3}$ of the total search space of each parameter. The initial temperature was chosen based on ensuring that nearly all initial proposed models are accepted as suggested in [12]. Although, a more rigorous comparison between SA and PSO involving optimization of all possible tuning parameters is beyond the scope of this chapter, this comparison does give insight into the performance differences between the two types of global optimization algorithms. Figure 2-8 shows the median and top/bottom 20 percentile error for SA inversions. The SA inversions appear to result in a much larger overall error compared to PSO with only a slight dependence on the number of layers. There appears to be a slight decrease in accuracy with 3 or more master layers. However, the NMSE for all of the inversions is significantly larger than PSO and for the remainder of the results we focus on the performance of PSO.

Source Frequency

To test the impact of source frequency, we run 25 inversions of the velocity model with 2-8 master layers with source frequencies of 3, 5, and 8Hz. For this example we

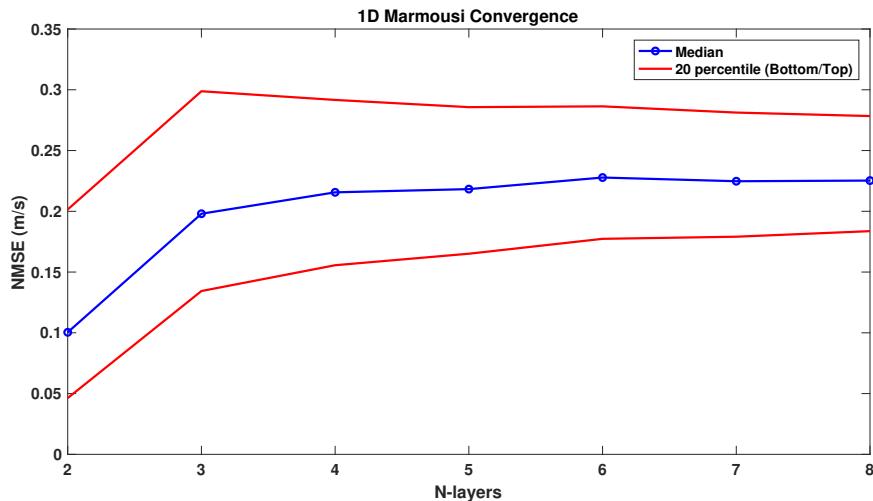


Figure 2-8: Simulated Annealing convergence results.

use a swarm size of 40 agents and run the inversion for 400 iterations to accommodate computational constraints of our cluster. The initial models are drawn from a uniform random distribution centered about the true velocity model with a range of +/- 1250 m/s for the velocity gradients and +/- 500 m for the depth of the master interfaces. Figure 2-9 shows the convergence rates as a function of the number of master layers for the three source frequencies. Much like the convergence plot shown in the previous section the likelihood of convergence significantly decreases for all models with more than 4 master layers. However, across frequency there appears to be very little difference in performance. This suggests, that if low frequency information is unavailable due to high noise, our approach could be used to invert an initial velocity model from higher frequency data.

Number of Agents

A major factor driving the computational cost of particle swarm optimization is the size of the swarm or number of agents. To test this limitation we run an inversion with swarm sizes of between 10 and 100 for 400 iterations for a 5 master layer velocity model at 5 Hz. As before, the initial models are drawn from a uniform random distribution centered about the true velocity model with a range of +/- 1250 m/s for

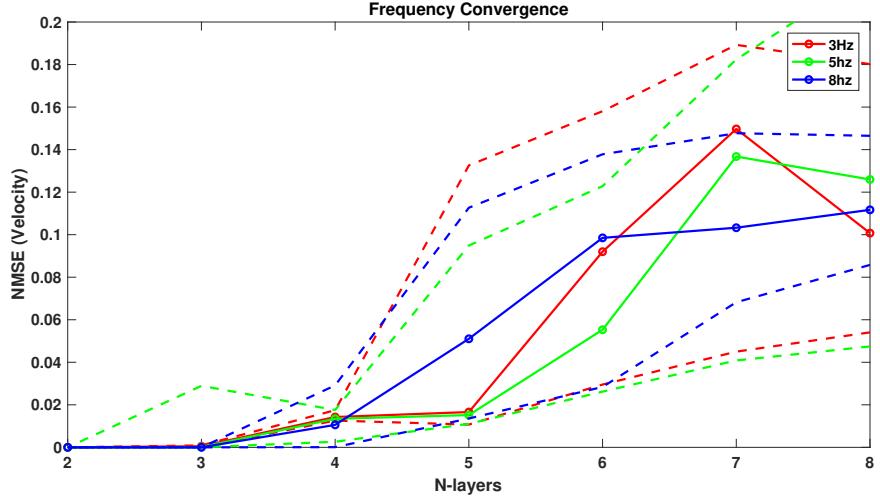


Figure 2-9: Convergence as a function of source frequency and number of master layers. Median values are shown with a solid line and top/bottom 20th percentile solutions are shown with dashed lines.

the velocity gradients and +/- 500 m for the depth of the master interfaces. Figure 2-10 summarizes the median and top/bottom 20 percentile convergence performance. From Figure 2-10 we see that the number of agents does not have a large effect on the convergence and so we choose to use a relatively modest 20-40 agents.

Number of Iterations

For the majority of the above simulation we run the simulations for 400 iterations. However, this number of iterations may not be optimal and the inversion may require more iterations to reach convergence. To test the impact of number of iterations on convergence, we run 25 inversions for a 5 master layer model for 50-1,000 iterations with a swarm size of 20 agents and source frequency of 5Hz. We initiate each agent position with a velocity model of +/- 1,250 m/s for the velocity gradient and +/- 500 m for master interface depth of the true model. Figure 2-11 shows the closeness of the forward model as a function of iteration. From this figure we see that there is a slight increase in the likelihood of convergence when 400 iterations are used. However, beyond 400 iterations there is little impact to the likelihood of convergence and additional iterations may be unnecessary.

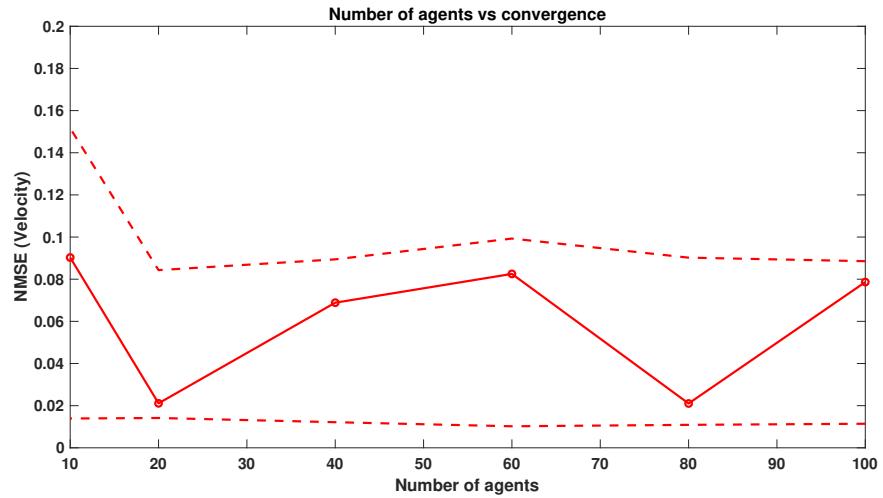


Figure 2-10: Convergence as a function of number of agents. Median values are shown with a solid line and top/bottom 20th percentile solutions are shown with dashed lines.

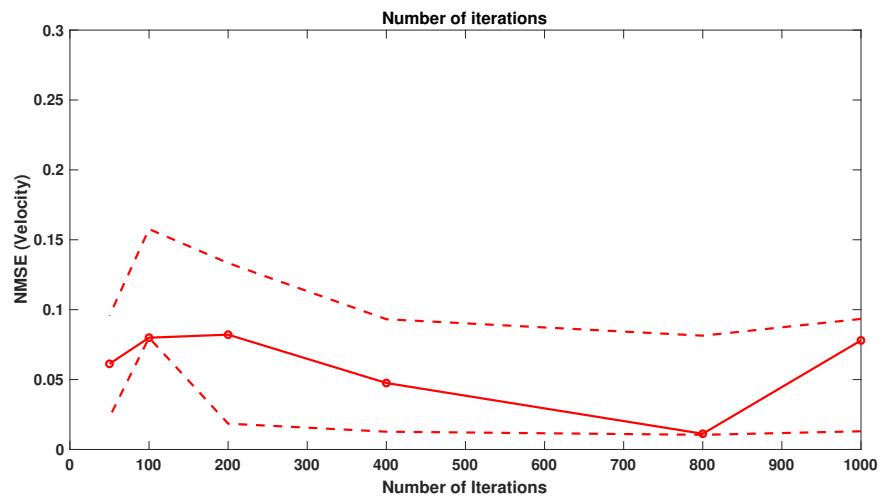


Figure 2-11: Convergence as a function of maximum iteration. Median values are shown with a solid line and top/bottom 20th percentile solutions are shown with dashed lines.

Quality of Initial Guess

In the previous examples we initiated each swarm with a velocity model of +/- 500 m/s and +/- 1,250 m/s of the true velocity model and initiated the inversion. For all inversions the agents' master layer depths are initialized from a uniform distribution +/- 125 m centered about the true interface depth and the inversion is run for 400 iterations. For real inversions, some prior information about the geology of the region or prior velocity analysis could influence the initial guess and its accuracy. From previous studies [17], choice of initial guess has significantly impacted inversion quality. To test the impact on initial guess quality we generate 25 inversions, varying the range of initial guesses from +/- 125 to +/- 1,000 m/s of the true velocity model. Figure 2-12 summarizes the convergence results for this experiment. From this figure we see only when initial guess quality is very good, less than +/- 500 m/s, do the inversions have a high likelihood of converging to the true model. However, if the initial guess quality is poor, +/- 500 m/s or more, then the likelihood of convergence appears to not significantly depend on the initial guess quality. Although the median of the 25 inversions increases steadily as function of initial guess, with the exception of the outliers at +/- 750 m/s, the 20th percentile results decrease dramatically beyond +/- 500 m/s. Because the range of values is so large after +/- 500 m/s the median outlier at +/-750 m/s is likely due to the large range of misfits observed at these wider initial guesses. These results suggest that unless the initial guess is within the base of attraction to the local minima, +/- 500 m/s, the PSO inversion quality appears to independent of initial guess.

2.5.2 Convergence and noise

In order to test the impact of noise on convergence of the PSO algorithm, we generate a synthetic trace using a 4 master layer velocity model. We choose this model because the previous inversions reliably converge to the true model without any noise. We then use the velocity model to generate a synthetic trace at 5 Hz and add Gaussian white noise at 10 different magnitudes such that the Signal to Noise Ratio (SNR),

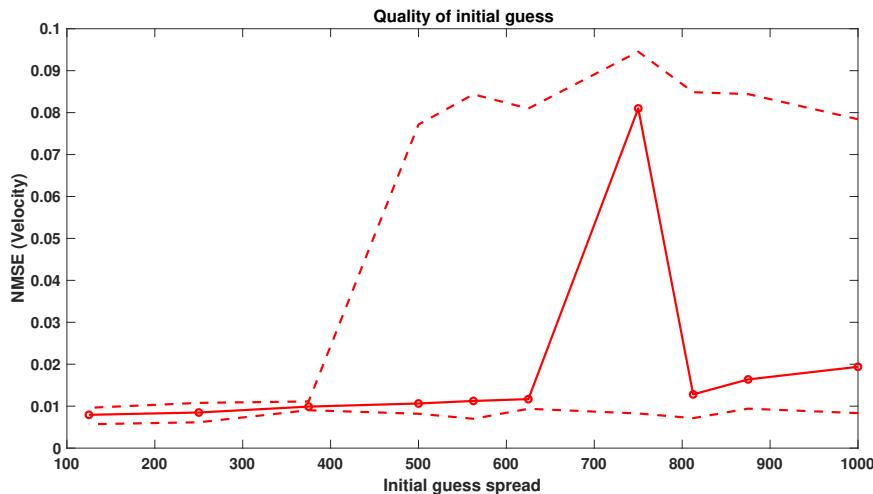


Figure 2-12: Model misfit as a function of initial guess quality. Median values are shown with a solid line and top/bottom 20th percentile solutions are shown with dashed lines.

where SNR is the ℓ_2 norm of the signal divided by the ℓ_2 norm of the noise, ranges from .11 to 53. Figure 2-13 illustrates the observed trace over a range of SNRs. At each of the SNRs we perform the PSO inversion initiating the starting models from a uniform distribution $+\/- 1250$ m/s the true velocity gradient and $+\/- 500$ m of the true master layer interface and run our inversion scheme for 400 iterations. We repeat this procedure 25 times for each noise level, generating a new noise realization at each run. Figure 2-14 shows the model misfit as a function of SNR. From the figure we see that the inversion accuracy decreases as we increase the noise level and for SNRs below 0.5 the inversion is extremely unreliable. However, for even fairly low SNRs the inversion scheme is able to produce reliable solutions. This is likely due to the fact that our reduced parameterization highly regularizes the inversion scheme, and likely improves the robustness to noise compared to conventional FWI schemes that have many more degrees of freedom and can overfit noise.

2.5.3 Complex Examples

In the previous examples we used an extremely simple forward model consisting of flat layers that does not fully capture the Earth's complex subsurface. In this section

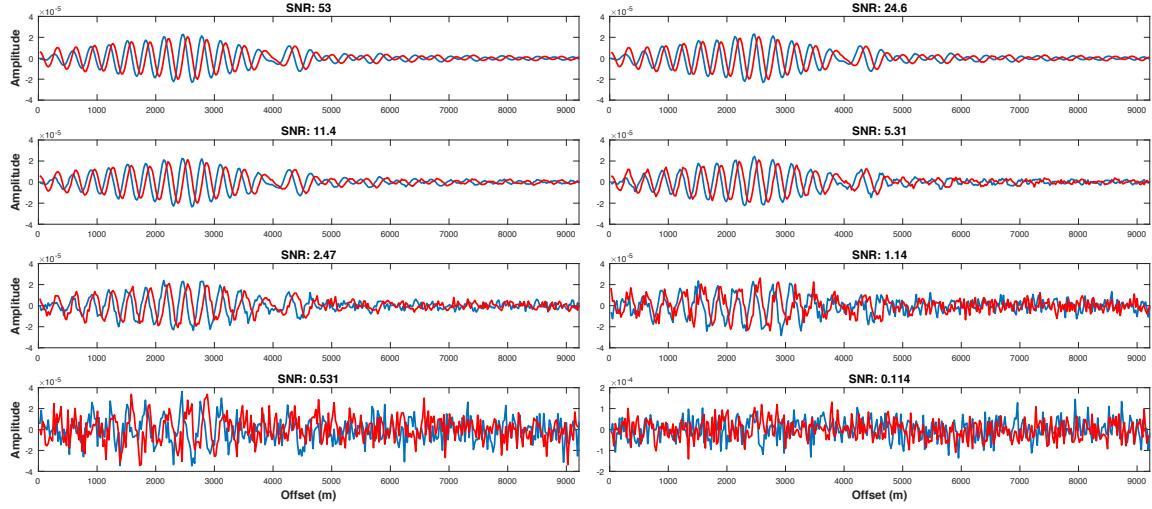


Figure 2-13: Illustrations of several signal to noise ratios at 5Hz. The real component of the field are shown in blue and the imaginary in red.

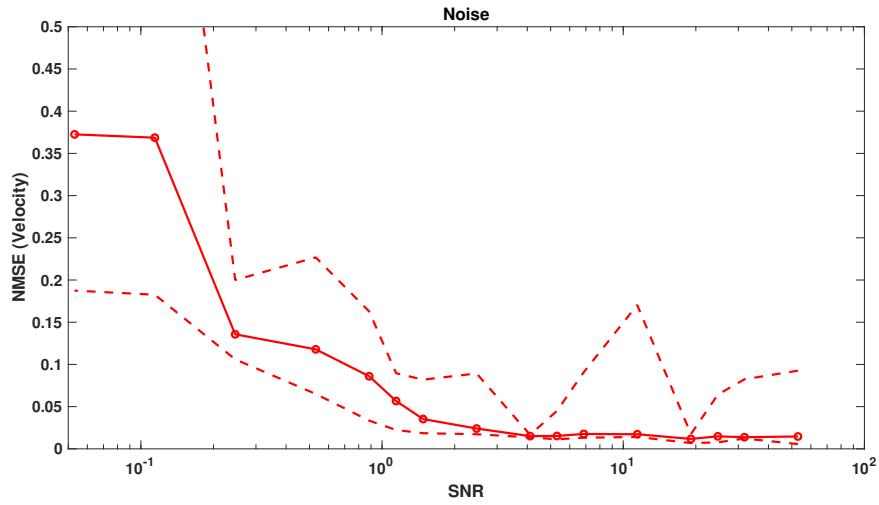


Figure 2-14: Model misfit as a function of Signal to Noise Ratio. Median values are shown with a solid line and top/bottom 20th percentile solutions are shown with dashed line.

we demonstrate our inversion algorithm on two complex velocity models that more accurately represent real world geology.

Perturbed Field Expansion Model

In the first example we generate several perturbed layered velocity models, shown in the left most panel of Figure 2-15, consisting of between 2-5 master layers and 7 control points per interface. These velocity models are roughly based on the Marmousi velocity model shown in Figure 2-5. We then initiate a swarm of 20 agents from a uniform distribution about the true model with $+/- 1250$ m/s of the velocity gradient, $+/- 1000$ m for depths of the master layer interfaces. The locations of the control points are randomly chosen from a uniform distribution centered about 0 with $+/- 100$ m. We then run 25 inversions for 250 iterations at 3 Hz for each of the four velocity models. As with the flat layered velocity models used in the previous section, convergence likelihood decreases as the number of layers increases, becoming unreliable after 5 or more layers, see Figure 2-16. Figure 2-15 shows the best fitting model and the 5th best model, or 20th percentile solution, compared to the true velocity model. Although, the inversions tend to match the upper structure of the velocity model accurately, the bottom sections are more inconsistent particularly as the number of master layers increases.

2.5.4 Marmousi Inversion

To test the effectiveness of our inversion scheme on more realistic data, we generate 3 and 5 Hz traces using the Marmousi velocity model for a single source shown in Figure 2-5 with Pysit's Helmholtz solver [32]. We then take the 1D vertical approximation of the pixelized velocity model and generate a profile similar to the profile shown in Figure 2-6. We then initiate the swarm with a velocity gradient $+/- 1,250$ m/s about the true profile and set the master layers to be equally spaced as a function of depth. The perturbation on each layer are initiated randomly from a uniform distribution from ± 50 m. We then run this PSO algorithm for 250 iterations. As with the prior

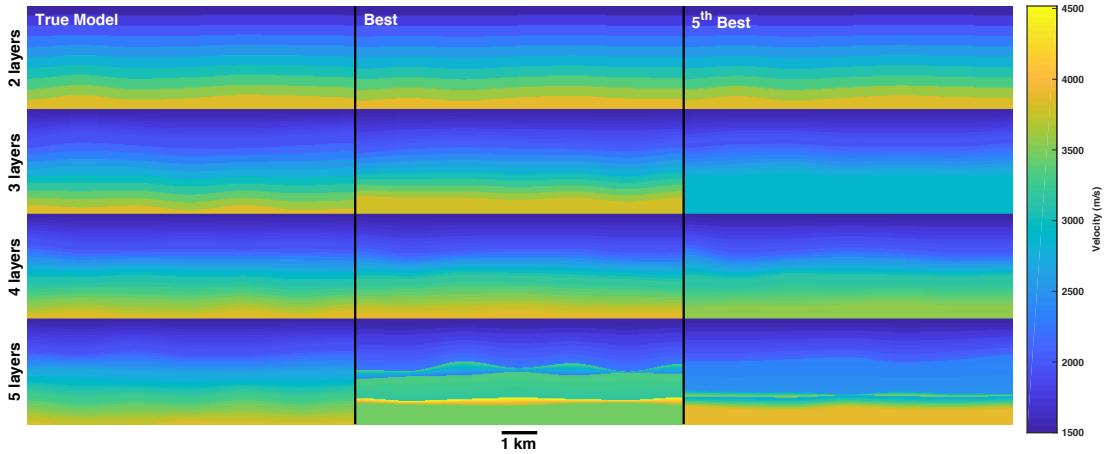


Figure 2-15: **Left:** True perturbed velocity models consisting of 2-5 master layers. **Middle:** the best fitting, lowest model NMSE. **Right:** The 5th best fitting (20th percentile) velocity model.

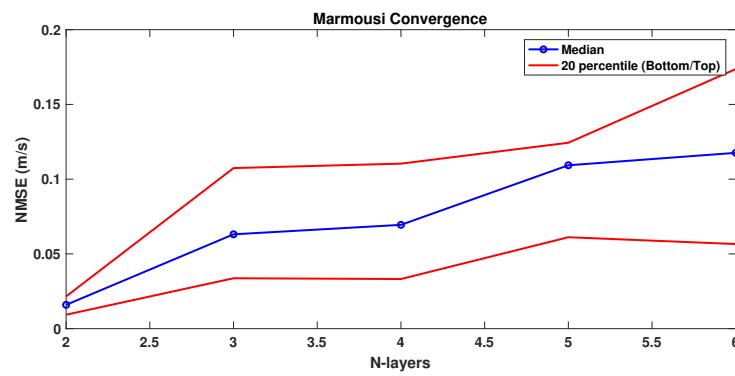


Figure 2-16: Convergence results for Marmousi like inversion as function of number of master layers.

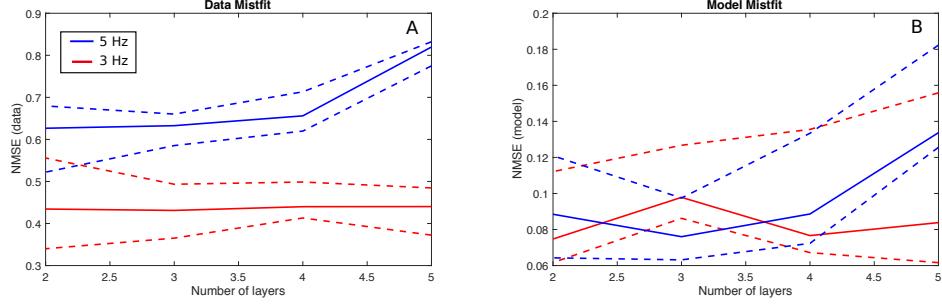


Figure 2-17: **A:** Data misfit between PSO inversions and the field generated from the Marmousi model at 3Hz and 5Hz as function of number of master layers. **B:** Model misfit for the PSO inversion between the smoothed Marmousi model shown in Figure 2-18A left and the smoothed inverted velocity models. For both the plots the median values are shown with solid lines and the top and bottom 20th percentile solutions are shown with dashed lines.

experiments we repeat this procedure 25 times for each number of master layers from 2 to 5. Figure 2-17 summarizes the data and model misfit between the Pysit simulated data and the inverted data using the field expansion. Although the final measurement error is significantly higher for the 5Hz data, the error between the inverted models is comparable for both 3 and 5 Hz, suggesting that both inversion frequencies could be used for finding an initial starting model. Unlike, the results in the previous section, the convergence does not monotonically increase as with the field expansion generated truth data. The difference between the data is likely due to the fact that the field expansion agrees more closely with the finite difference data at lower frequencies as shown in Figure 2-4.

To further test the effectiveness of the starting models we take the 3 inversions that have the lowest data misfit for the 3Hz and 5Hz inversions and use these models as an initial starting model for conventional FWI. We apply a Gaussian blur to each of the field expansion velocity models and then insert the known water level into the initial model. Without the application of the Gaussian blur the FWI inversions fail with a number of high frequency artifacts. Figure 2-18BCD left show the 3 initial starting models at 3Hz and 2-19BCD show the 3 initial models at 5Hz. For comparison we also ran FWI using the true Gaussian blurred velocity model shown in Figures 2-18A & 2-19A and the resulting inversion is shown on the right. We run all the FWIs with

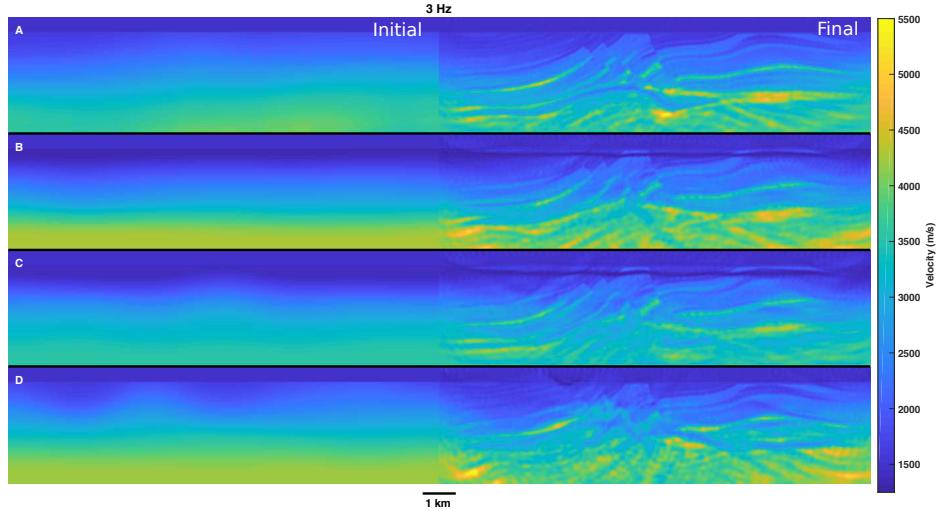


Figure 2-18: **Left:** Initial models used for FWI. **Right:** Final models after FWI. **A:** Inversion using the true smoothed Marmousi Velocity model. **BCD:** The initial models based on the best fitting, lowest data measurement error, PSO inversion at 3Hz and the resulting final models from FWI.

64 shots with 10 iterations at each frequency 3Hz, 4Hz, 5Hz, 6.5Hz, 8Hz and 10Hz using Pysit's Helmholtz solver. From the final models shown in the right of Figures 2-18 & 2-19 we see that the majority of PSO inverted initial models result in a good final velocity model. Although there are artifacts present in some of the inversion results, the upper structure of the velocity model is well recovered and for many of the models the deeper structure agrees well with the inversion from the true baseline, Figure 2-18A right.

2.6 Discussion

From these results we find that it possible to estimate a good starting model from medium frequency data, 3-8Hz, and the ultra low frequencies used to estimate a starting model based on raw data may be unnecessary. The difference between conventional Helmholtz solvers and the field expansion appears to increase as a function

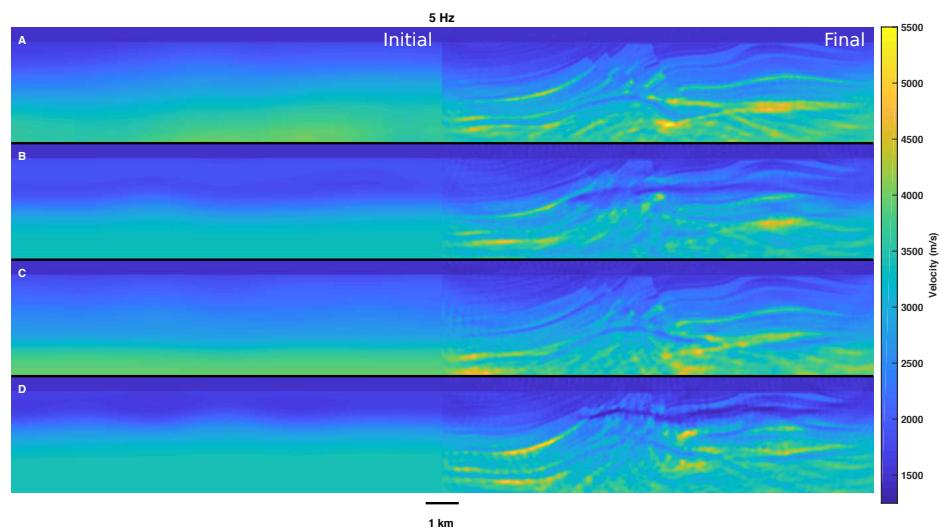


Figure 2-19: **Left:** Initial models used for FWI. **Right:** Final models after FWI. **A:** Inversion using the true smoothed Marmousi Velocity model. **B,C,D:** The initial models based on the best fitting, lowest data measurement error, PSO inversion at 5Hz and the resulting final models from FWI.

of frequency, modifying the minimized cost function to measure phase differences could significantly decrease these discrepancies. Additional work is needed to study the performance of the field expansion based inversions at higher frequencies and with alternative cost functions. Although we are able to estimate good starting models at 5Hz, we still need the 3Hz frequency data to generate a good inversion. Running, the inversion starting at 5 Hz, even for the smoothed true velocity model shown in Figure 2-18A left, resulted in an unusable model with numerous artifacts. Other approaches, such as low frequency interpolation [44, 36], could be used to fill in the lower (3-5 Hz) frequencies needed for FWI once a starting model is built.

From the results shown in Figures 2-18 & 2-19 we see that FWI can result in models that vary significantly for quite similar starting models. Although some of the artifacts are easily identifiable in the final image, some structures in the final image are significantly distorted or less pronounced. Little numerical work has been done on the impact of erroneous starting models and it would likely be fruitful to examine the performance of FWI across a population of initial models. However, it is difficult to determine what should be considered a likely or plausible starting model. One potential approach would be use the uncertainty quantification framework presented in [17] to generate a distribution of velocity models and select a few representative models and perform FWI with these initial models. This set of initial models could then be used to benchmark various FWI algorithms against one another and give more realistic bounds of performance than using a single initial model. Although we only looked at few varieties of global optimization algorithms and impacts of parameters, the approach in this chapter could easily be tested with a wide variety of global optimization algorithms currently being studied in the geophysics community. Using this reduced parameterization and the field expansion method would allow for objective comparisons among algorithms that would be impossible with conventional slower solvers.

2.7 Conclusion

In this chapter we developed a platform for comparing global optimization methods for seismic velocity estimation. Although we did not exhaustively test all possible combinations of tuning parameters and cost functions, this study can guide parameter choice for other forward models where it would be too expensive to explore these trade offs in solution space. In addition, the framework presented in this chapter could easily be extended to test and benchmark a variety of global optimization methods currently used by the seismic community. This study can guide parameter choice for other forward models where it would be too expensive to explore trade offs and the extent of the solution space. This testing framework could be extended to other local optimization procedures such as conventional FWI, whereby we could quantitatively compare the performance of algorithms by testing their performance over a population of reasonable starting models instead of single velocity model.

Chapter 3

Uncertainty Quantification of Seismic Velocity Models and Seismic Images

3.1 Overview

Seismic imaging is conventionally performed using noisy data and a presumably inexact velocity model. Uncertainties in the input parameters propagate directly into the final image and therefore into any quantity of interest, or qualitative interpretation, obtained from the image. We consider the problem of uncertainty quantification in velocity building and seismic imaging using Bayesian inference. Using a reduced velocity model, a fast field expansion method for simulating recorded wavefields, and the adaptive Metropolis-Hastings algorithm, we efficiently quantify velocity model uncertainty by generating multiple models consistent with low-frequency full-waveform data. A second application of Bayesian inversion to any seismic reflections present in the recorded data reconstructs the corresponding structures' position along with its associated uncertainty. Our analysis complements rather than replaces traditional imaging because it allows us to assess the reliability of visible image features, and to take that into account in subsequent interpretations.

3.2 Introduction

Uncertainty quantification of seismic images and velocity models allows us to know what parts of the image are believable, if our conclusions and interpretations are justifiable and reliable, or if more data need to be collected. It is an indispensable part of risk assessment and decision-making. [51] provides an overview of the current motivation and methods for uncertainty quantification in seismic imaging.

A naïve approach to uncertainty quantification would view the velocity at each grid point as a separate variable and attempt to describe the joint probability distribution of all velocity values. This amount of information quickly becomes overwhelming. Quantifying the uncertainty of every point in the model is also computationally challenging. For a very simple 128×128 model the velocity model covariance matrix, which describes the uncertainties under the simplest set of assumptions, would contain over 200 million entries. Even if we were to calculate this covariance matrix, displaying its content in a meaningful way, particularly the off-diagonal elements that describe important correlations and trade-offs between velocity values at different points, would be difficult. More realistic 3D models that contain hundreds, or even thousands, of points in each direction would make even the computation of the covariance impossible with today’s computing resources. Beyond the computational challenges, using a simple covariance to characterize uncertainty assumes that the errors are Gaussian. In reality, these Gaussian assumptions are not satisfied, the covariance matrix does not carry all of the information about uncertainty, and more general probability distributions need to be calculated. Characterizing uncertainty under these conditions using best-practice methods like Markov Chain Monte Carlo (MCMC) [6] or other sophisticated statistical estimation techniques is out of question with current resources.

More importantly, this naïve approach to uncertainty quantification does not lead to meaningful measures of uncertainty. Although a large and detailed subsurface image is often available to us, we are often particularly interested in relatively simple features that are a small subset of the full image. After forming a standard large and

detailed image, we are interested in additional information that quantifies the uncertainty of specific image features. For example, we may be interested in estimating the volume of a potential reservoir or the depth of a specific structure because our next steps depend on these quantities more than others. A single seismic image allows an interpreter to estimate these quantities but it provides no information about error bounds on them. If there is a cost associated with incorrect estimates then this cost cannot be quantified. A good uncertainty quantification method will show how reliable these measurements are, which will provide an interpreter with additional information that complements the information extracted from a migrated image using conventional tools.

For specificity, we consider in our theoretical discussion and all numerical experiments a surface seismic reflection survey with shot and receiver locations at the surface. From a workflow perspective, we assume that a traditional seismic image obtained using reflection data and a migration velocity model is available to us. We proceed towards our goal of quantifying the uncertainty of a feature of interest in a seismic image in two steps. First, we estimate the uncertainty in the migration velocity model from surface reflection data using Bayesian Full-Waveform Inversion. [51] assumed in their analysis that the model uncertainty was Gaussian. They explain, however, that the true uncertainty is often not Gaussian. In this chapter, we calculate a general (typically non-Gaussian) Bayesian posterior distribution that describes the velocity model uncertainty. [8] previously used MCMC methods to estimate uncertainty on velocity models but did so by picking arrival times rather than using the full waveform data. Second, we propagate this velocity uncertainty in reflector locations through simple time migration to calculate the uncertainty of any quantity of interest (e.g. location of a structure, trap volume etc), in a final image. Our full analysis starts with recorded waveforms and produces uncertainty estimates of the velocity model, and of any feature in the image. To highlight the importance of velocity model uncertainty in our synthetic examples we only quantify the error in the migrated image due uncertainty in the velocity model. Accounting for noise in the high frequency time picks would also affect the migrated image, but we do not

address this here as it is address elsewhere e.g. [53].

Due to the computational cost of wave equation based forward solvers previous method of uncertainty quantification have typically used ray and travel time based methods and Gaussian assumptions to handle the large dimensionality of seismic velocity models. [40] use the Dix equation combined with a layer stripping procedure to estimate velocity model uncertainty. [25] assume a Gaussian velocity model distribution and use reflectivity forward modeling to explore different methods of conveying uncertainty such as showing the distribution of quantities of interest. [37] propose a framework for using multiple data sets to characterize 3D slowness using well logs and check shot travel times. In addition, numerous methods have been proposed for approximating the resolution matrix for large scale tomographic inversions. [50] approximates the resolution matrix using Lanczos algorithms for marine seismic data and [89, 76] use similar methods to approximate the resolution matrix for global tomography.

Uncertainty quantification is useful only if it is mathematically sound and numerically feasible. Solving almost any inverse problem with uncertainty quantification, including the one we present here, relies on being able to perform numerous instances of forward modeling. Finite-difference simulation of surface reflection data for a detailed velocity model is too computationally expensive to run more than 10-20 iterations for each inversion, far less than the thousands needed to perform uncertainty quantification using MCMC methods [66, 90]. [38] use the neighborhood algorithm and a reduced parameterization based on regional geology to estimate the posterior distributions efficiently; this is related to our work in that it simplifies the model to allow us to answer more specific questions about the solution. One way to reduce these computational costs is to use a simpler velocity model. While at first glance this seems a large sacrifice, in many cases the velocity-induced uncertainty of quantities of interest does not depend on the fine-scale details of the velocity model. For example changing a few pixels of the velocity model will not generate a noticeable kinematic error, these must accumulate over large distances and thus large swaths of the velocity model. We follow this approach, specifically we adopt a fast Helmholtz solver for a

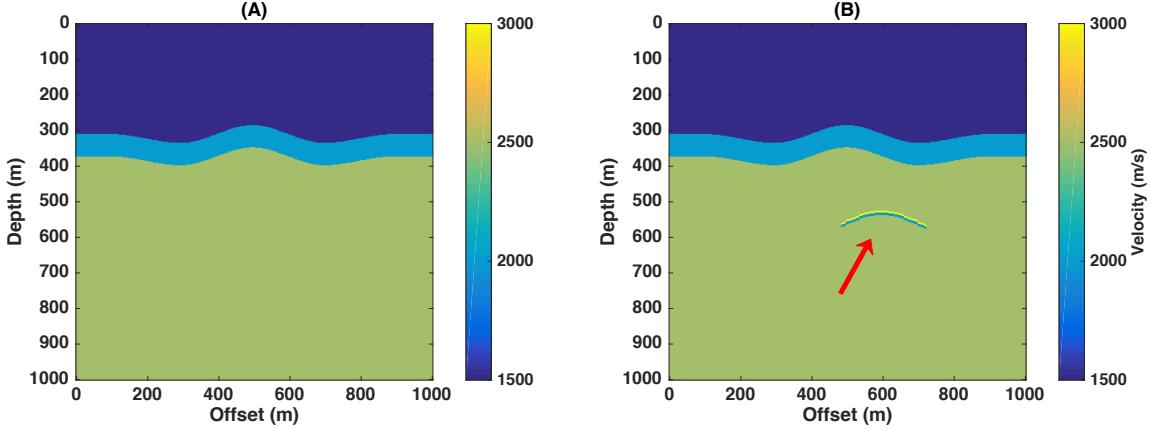


Figure 3-1: A: The base velocity model consisting of 3 perturbed layers. B: The reflector model is the same as the base model with the addition of an anticline reflector, indicated by the red arrow, in the deepest layer. The velocities within the layers are 1500, 2000, and 2500 m/s from shallow to deepest.

perturbed layered medium to reduce the computational cost of forward modeling. We parameterize the velocity model with a number of layers containing a linear gradient similar to the parameterization described in [12] where the authors used this parameterization to estimate initial velocity models for full waveform inversion. Although, this parameterization may not allow us to fully model all features in a hypothetical velocity model, it sufficiently captures the model's seismic response at low frequencies allowing us to characterize uncertainty for large-scale velocity model features. With an efficient forward modeling tool at hand, we perform Bayesian inversion using a tuned adaptive Metropolis-Hastings algorithm[30] to calculate the velocity model posterior distribution. This algorithm allows us to invert for the model quickly and without restrictive additional assumptions on the uncertainty distribution, such as assuming a Gaussian posterior.

In this article we demonstrate this uncertainty quantification framework on a simple three-layer velocity model containing an anticline reflector as Figure 3-1B and a complex velocity model based on the Marmousi velocity model [78]. One could imagine an interpreter viewing a similar image and velocity model and being asked to estimate the volume of potential trap contained in the anticline and the depth to the reflector. In the rest of the chapter, we show how we can characterize the uncertainty

of these physical quantities describing the trap directly from the raw data.

3.3 Bayesian velocity model inversion

Given a set of recorded traces, \mathbf{d} , our first step is to estimate the velocity model, , like the model shown in Figure 3-1A, and quantify its uncertainty. We assume that the observed data consist of reflections dependent on the true velocity model, , perturbed by additive Gaussian noise, i.e.,

$$\mathbf{d} = f(\mathbf{m}) + \mathbf{n}, \quad (3.1)$$

where $\mathbf{n} \sim N(\mathbf{0}, \Sigma)$ is zero-mean Gaussian noise with a covariance matrix Σ . The map f simulates wave propagation in a velocity model . Our choice of f will be discussed in detail in subsequent sections. The covariance matrix Σ can be estimated in practice by measuring receiver noise recorded when no coherent source signal is present, e.g., between event arrivals [26, 25]. Inaccurate estimates of the covariance would likely bias the results and inaccurately quantify uncertainty. If the noise is uncorrelated then this matrix is diagonal with the noise variance on each trace as the diagonal entries. Although in later simulations we generate examples in the presence of uncorrelated Gaussian noise, our framework can account for any form of correlated Gaussian measurement noise as long the covariance matrix Σ can be well estimated.

A likelihood function is the probability of observing data \mathbf{d} given a model . It follows from Equation 3.1 and the Gaussian assumptions on noise that the likelihood function is given by

$$L(\mathbf{m}) \equiv p(\mathbf{d} | \mathbf{m}) \propto \exp \left[-\frac{1}{2} (\mathbf{f}(\mathbf{m}) - \mathbf{d})^\top \Sigma^{-1} (\mathbf{f}(\mathbf{m}) - \mathbf{d}) \right]. \quad (3.2)$$

See [70] for a detailed discussion of this form of the likelihood function. Although we assume that the noise can be well approximated by a Gaussian distribution, we make no assumptions about the model distribution and this non-parametric framework allows for the velocity model to be described by any distribution, including multi-

modal and non-Gaussian distributions. Because the likelihood function $p(\mathbf{d} \mid \mathbf{m})$ includes $f()$, we must calculate the likelihood numerically by first simulating the data that would be recorded in a given velocity model, and then applying Equation 3.2. A posterior, $p(\mid \mathbf{d})$, is the result of a probabilistic inversion for the model from the data \mathbf{d} . In other words, it is the probability of the model given the observed data. The posterior is given by Bayes' theorem that states:

$$p(\mid \mathbf{d}) = \frac{p(\mathbf{d} \mid \mathbf{m}) p(\)}{p(\mathbf{d})}. \quad (3.3)$$

Here, $p()$ is the prior distribution of models, and can be used to input prior information about the velocity model from other forms of velocity analysis or even to simply impose limits on the range of acceptable velocities. The normalization constant $p(\mathbf{d})$ ensures that the left hand side integrates to unity and is thus a valid probability distribution.

While conceptually simple, Equation 3.3 provides no analytic expression for the posterior $p(\mid \mathbf{d})$, and thus it must be calculated numerically. A large number of likelihood functions $p(\mathbf{d} \mid \mathbf{m})$ need to be evaluated in order to compare different models and construct a posterior distribution $p(\mid \mathbf{d})$. In order to make this feasible we must use an efficient forward model f and we should also wisely choose the models that we test and compare with the observed data. In what follows we discuss a fast Helmholtz solver we employ to calculate $f()$. We then present an adaptive Metropolis-Hastings algorithm [30], a type of MCMC sampler, that calculates the posterior without requiring knowledge of the normalization constant $p(\mathbf{d})$, which is usually the hardest quantity to compute.

3.4 Field expansion forward solver

A forward model used for uncertainty quantification should ideally satisfy the following conditions. First, the forward solver, $L()$, should be sufficiently fast to be able to calculate synthetic datasets for thousands of models . Second, the number of degrees

of freedom in the model should allow us to explore the model space with an MCMC sampling method; these algorithms work well with model spaces up to about 200 degrees of freedom as demonstrated in the original adaptive Metropolis-Hastings paper [30]. Third, the model should be parameterized in a way that allows us to capture essential features of the model and define meaningful quantities of interest. The field expansion method, proposed in [48] and adapted to velocity model inversion in [16], satisfies all of these requirements and so we use this solver.

In the field expansion method, we first approximate the velocity model by a perturbed layered medium, and then use a fast Helmholtz solver that is well suited for this type of model. This solver achieves substantial computational savings by using a Taylor expansion of the analytic solution to the 2D Helmholtz equation for a set of *flat* layers to calculate a general solution for a *perturbed* layer model. The computational cost of the forward solver scales as $\mathcal{O}(MN^2N_x \log N_x)$ where M is the total number of layers, N is the order of the Taylor series (usually $N < 10$) chosen based on the amplitude of layer depth variations, and N_x is the number of horizontal spatial modes which depends on receiver spacing [48]. Because the computational cost scales linearly as a function of the number of layers, forward solves for relatively complex velocity models can be calculated extremely rapidly. In contrast, a frequency domain finite difference direct solver with N_p mesh points using an LU factorization would involve the construction of an $N_p \times N_p$ matrix and a right hand solve that scales as $\mathcal{O}(\frac{2}{3}N_p^3)$. If we were to double the number of layers via the field expansion, computational cost would increase by a factor of 2 whereas the cost for the frequency domain solver would increase by a factor of 8. Similarly if we doubled the horizontal number of pixels of our model and the number of horizontal spatial modes the field expansion costs would increase by $2 \log(2)$ whereas the pixel model would again have a factor of 8 increase in computation time. In addition, as frequency increases both the number of spatial modes N_x and number of pixels N_p would need to increase for both the field expansion and finite difference forward solvers.

We construct the perturbed layered approximation to the true velocity model as follows. The model has a total of M master (or physical) layers. Each master layer

boundary is a cubic spline with N_q nodes or control points. The velocity inside the i -th layer, where $i = 1, \dots, M$, is assumed to be a vertical linear gradient going from V_i^u to V_i^d as the depth increases. Because our method requires a piece-wise constant velocity model, each master layer i is further split into a fixed number of homogeneous sub-layers, whose velocities are uniformly gridded between V_i^u and V_i^d . Presence of these sub-layers does not increase the model size because all velocities inside each master layer are controlled by only two parameters: V_i^u and V_i^d .

The resulting model is just another discretization of a velocity gradient, but instead of having the gradient relative to a fixed grid, the gradient is relative to the boundaries of the two layers. This parameterization allows us to efficiently model complex velocity models with just tens of parameters instead of thousands. Figure 3-2 shows master layer boundaries for a 5-layer model (left), sub-layers within each master layer (center), and the resulting 122×384 2D velocity model (right). This velocity model is quite complex, and yet it is fully described with only 32 parameters rather than over 40,000 parameters needed for a regularly gridded velocity model of the same size. [12] successfully use this parameterization to estimate initial velocity models but then use an expensive finite-difference solver for full-waveform inversion. [20] use this formulation to generate synthetic seismograms and [88] use it for tomography. We use this parameterization, and fully exploit its numerical efficiency, throughout our work from building an initial model to calculating a posterior distribution over all plausible velocity models.

The reduced model space that we have presented cannot capture all of the features of a high resolution velocity model. For example, our perturbed layer model cannot describe discontinuous features, such as the deeply buried anticline reflector shown in Figure 3-1B. It should allow us, however, to accurately simulate the low frequency components of recorded data that are insensitive to the finer details of the model. In addition, it's unlikely that one pixel of the model is different enough to be resolved, but it's more geologically plausible that the velocity of a layer or its shape has been incorrectly estimated in a way that will have a discernible impact on the final results.

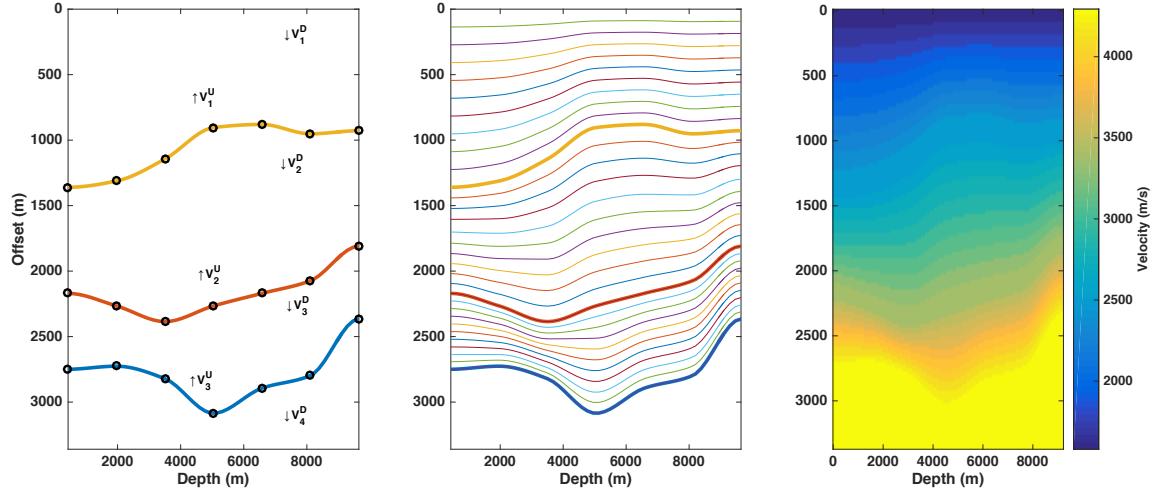


Figure 3-2: Left: Parameterization for gradient field-expansion. Each layer has an up and down velocity describing the velocity gradient and each boundary is described by N_q points (here 7). Center: The interpolated sub-layer boundaries between the master layers boundaries. Right: The resulting 2D velocity model. A small number of parameters (here 32) describe a complex smoothly varying velocity model that would ordinarily require over 30,000 parameters.

3.5 Efficient posterior calculation

Equation 3.3 provides a simple formula for the posterior probability distribution of velocity models given observed data and assuming the measurement noise is normally distributed. Although we have assumed in the problem statement that the data misfit is entirely due to Gaussian noise, we have not made any assumption about the distribution of velocity models. In particular, this framework easily deals with posterior distributions that are multi-modal due to cycle skipping. In the case of strongly non-Gaussian noise, a conservative estimate of uncertainty can be obtained by boosting the assumed noise energy, or by explicitly reformulating the likelihood function to account for a different model misfit.

For equation 3.3 to be useful, it must be accompanied by an efficient numerical scheme to calculate the posterior distribution. Markov Chain Monte-Carlo (MCMC) methods are algorithms that generate a sequence of samples — in our case a sequence of velocity models — in such a way that the sequence converges to the desired distribution independent of the initial starting model. In practice a large portion of the

samples are discarded to reduce the impact of the starting model, typically by removing the first half of the sequence. The Metropolis-Hastings algorithm is an MCMC method that is particularly well-suited for sampling from a Bayesian posterior because this method requires just an unnormalized probability density function. This means that the most arduous task of calculating the normalization constant $p(\mathbf{d})$ in Equation 3.3 can be avoided altogether [6]. Below we summarize the Metropolis-Hastings algorithm and comment on key details pertinent to our specific problem.

Generally speaking, the Metropolis-Hastings algorithm is an iterative procedure that generates a proposal for the next velocity model sample by randomly perturbing the current sample. The proposal is either accepted or rejected according to the ratio of the likelihood of the proposal to the likelihood of the current sample. If the proposal is accepted then it becomes the next sample; if the proposal is rejected then the present sample is reused as the next sample. As shown in [6], if the number of iterations is sufficiently large, such that the chain has converged and fully explored probability space, the velocity samples so obtained will be from the posterior distribution. Following the standard practice described in [6] section 6.5 we discard the first half of the generated samples and only use the second half to describe the posterior to remove the impact of the initial starting model. The choices of the initial model \mathbf{d}_0 and of the iteration-dependent perturbations affect the convergence of the empirical distribution of the generated samples to the posterior distribution, and therefore the overall computational cost. We now elaborate on both of these important factors.

We select a simple gradient as the starting model \mathbf{d}_0 . Our tests, over many different initial models find that our algorithm is rather stable to the choice of \mathbf{d}_0 . The likelihood of the initial model $L(\mathbf{d}_0) \equiv p(\mathbf{d} | \mathbf{d}_0)$ is calculated using Equation 3.2. At each step $i \geq 1$, a potential candidate for the next sample is generated by adding a zero-mean Gaussian perturbation to the last sample:

$$\mathbf{d}_* = \mathbf{d}_{i-1} + \mathbf{n}, \quad \mathbf{n} \sim N(\mathbf{0}, \mathbf{C}_i). \quad (3.4)$$

Here \mathbf{C}_i are iteration-dependent covariance matrices. We begin with an initial covari-

ance matrix \mathbf{C}_0 and progressively tune it at iteration $i = 1, \dots, k$ to ensure speedy convergence of the empirical distribution of generated samples to the true posterior distribution following [30]. Specifically, we set $\mathbf{C}_i = \mathbf{C}_0$ for a fixed number of iterations N_C , and then begin updating \mathbf{C}_i by setting

$$\mathbf{C}_i = S_d (\text{Cov}_{[0, \dots, i-1]} + \varepsilon \mathbf{I}_d), \quad (3.5)$$

where $d = \text{Length}(0)$ is the size (number of parameters) of the model, $S_d = 5.76/d$ is a heuristically defined constant [22, 30], \mathbf{I}_d is a $d \times d$ identity matrix, and $\varepsilon \ll 1$ is a regularization constant. The value N_C reflects our confidence in the initial guess of \mathbf{C}_0 . If \mathbf{C}_0 is believed to be close to $\text{Cov}_{[0, \dots, i-1]}$, then N_C can be set to a low value. If \mathbf{C}_0 is not close to the true covariance then N_C should be set to a large value so that a significant number of samples are collected before auto-tuning is used. In practice finding that the inversions are relatively robust to our choice of N_C and that a value of approximately 1,000 leads to good convergence.

A forward solve of the wave equation using the field expansion forward solver, described in the previous section, is performed, and the likelihood $L(*) \equiv p(\mathbf{d} | *)$ is calculated using Equation 3.2. If the proposed candidate is more likely than the last one, i.e. $L(*) > L(i-1)$, then the candidate is immediately accepted and we set $i = *$. If the proposed candidate is less or equally likely as the last model, $L(*) \leq L(i-1)$, then this candidate is accepted with probability $\alpha_i = L(*)/L(i-1)$, and it is rejected, i.e. $i = i-1$, with probability $1 - \alpha_i$. This process is then repeated for as many $i = 1, \dots, k$ as is feasible to yield a sequence of velocity models

$$_{0,1}, \dots, _k. \quad (3.6)$$

Because the algorithm does not immediately converge to the posterior distribution, we discard the first half of the samples $_{0,1}, \dots, _{k/2-1}$ and use only the remaining $k/2$ samples $_{k/2}, _{k/2+1}, \dots, _k$, to represent the posterior. This methodology is summarized in pseudo-code in Algorithm 3.

The performance of the Metropolis-Hastings algorithm as described is affected

Algorithm 3 Calculating posterior velocity models using the Metropolis-Hastings method

Require: \mathbf{m}_0 initial model
Require: $k \geq 1$ number of samples
Require: $1 \leq N_C \leq k$
Require: $\varepsilon \ll 1$ regularization constant

$d \leftarrow \text{Length}(\mathbf{m}_0)$
 $n_a \leftarrow 0$ number of proposal acceptances
 $S_d \leftarrow 5.76/d$ empirical constant

for $i = 1, \dots, k$ **do**

- $l_{i-1} \leftarrow L(\mathbf{m}_{i-1})$
- if** $i \leq N_C$ proposal covariance matrix **then**

 - $\mathbf{C}_i \leftarrow \mathbf{C}_0$ Use initial matrix

- else**

 - State $\mathbf{C}_i \leftarrow S_d(\text{Cov}[0, \dots, i-1] + \varepsilon \mathbf{I}_d)$ Use auto-tuned matrix

- end if**
- $\mathbf{n} \leftarrow \text{Normal}(\mathbf{0}, \mathbf{C}_i)$ proposed model perturbation
- $* \leftarrow \mathbf{m}_{i-1} + \mathbf{n}$ proposed model
- $l_* \leftarrow L(*)$ proposed model likelihood
- $\alpha_i \leftarrow l_*/l_{i-1}$ acceptance probability
- $\text{coin} \leftarrow \text{Bernoulli}(\alpha_i)$ “unfair” coin flip
- if** $\text{coin} = 1$ Accept/reject proposed model based on flip result **then**

 - $i \leftarrow *$
 - $n_a \leftarrow n_a + 1$

- else**

 - $i \leftarrow i-1$

- end if**

end for

by the choice of the initial covariance matrix \mathbf{C}_0 . In order to further speed up the convergence, we find a good covariance matrix \mathbf{C}_0 by test-running the Metropolis-Hastings algorithm with some fixed initial diagonal \mathbf{C}_0 discarding all samples, but keeping track of how many proposals have been accepted. A too low or too high acceptance rate is taken as a sign that the covariance matrix is either too large or too small, and it needs to be adjusted accordingly. Algorithm 4 spells out the exact procedure in pseudo-code. [6] and [70] provide further details about the Metropolis-Hastings algorithm along with convergence results.

Algorithm 4 Metropolis-Hastings \mathbf{C}_0 tuning procedure

Require: \mathbf{C}_0

Require: $n_{\text{tune}} \geq 1$

Require: $0 \leq R_a^{\min} < R_a^{\max} \leq 1$

```

1: repeat
2:   Run Algorithm 3 with  $qk = n_{\text{tune}}$ 
3:    $R_a \leftarrow n_a/n_{\text{tune}}$   $n_a$ : number of proposal acceptances
4:   if  $R_a > R_a^{\max}$  Covariance matrix too large then
5:      $\mathbf{C}_0 \leftarrow \mathbf{C}_0/2$ 
6:   else if  $R_a < R_a^{\min}$  Covariance matrix too small then
7:      $\mathbf{C}_0 \leftarrow 2\mathbf{C}_0$ 
8:   end if
9: until  $R_a^{\min} \leq R_a \leq R_a^{\max}$ 

```

It is clear from the above description that each iteration of the Metropolis-Hastings algorithm involves a forward wave solve in order to calculate the likelihood of the proposed velocity model. Under the best scenarios, the Metropolis-Hastings algorithm still requires thousands of iterations to converge, making the use of the finite-difference solvers, conventionally used in full waveform inversion, computational infeasible. The combined use of an efficient posterior sampler and a fast forward wave solver is key to practical and useful uncertainty quantification in velocity inversion.

Migrating reflectors in velocity models with uncertainty

Having established a method of estimating velocity uncertainties in low spatial frequency-velocity models from low-frequency data, we now turn our attention to the higher frequency part of the model. Specifically, we shift from estimating uncertainties in Figure 3-1A to estimating the uncertainties in the reflector shown in Figure 3-1B.

A reflector in the subsurface can be represented as a physical structure in the model space or as a reflection event in the data space. The two representations are related to one another through migration and demigration (modeling) operators, which use an assumed velocity model. When the velocity model carries uncertainty, this uncertainty propagates to the locations and depths of migrated structures. The same reflection data give rise to many plausible physical structures in the model space that are all consistent with the data.

These structures are not all equally likely however, and our goal in this section is to carefully quantify the uncertainty in the depths of migrated reflectors. We will use the same Bayesian framework to quantify the depth uncertainty of migrated reflectors that we used before to build velocity models with uncertainty from low-frequency recordings. As discussed above, the Bayesian framework requires fast forward modeling (in this case, demigration) so that data corresponding to a possible structure can be quickly compared to the recorded signals.

Following [53], we use map migration/demigration to move between the model-space and data-space representations of a reflector. Map migration is analytic, and hence very fast. [41] previously used map migration for uncertainty analysis. It maps an infinitesimal straight-line reflector in the model space to an infinitesimal linear event in the data space. If we represent an entire reflector as a collection of small line segments, then we can migrate and demigrate the entire structure all at once by migrating and demigrating the individual tiny pieces that make up this structure.

Map migration works with re-sorted common-offset gathers for all available offsets. For the sake of presentation clarity, we will use zero-offset migration to make the

formulas simple. The general formulas applicable to all offsets are given by [53]. In our case, the uncertain velocity model induces an uncertain depth on individual line segments, and we now proceed to quantify this uncertainty.

We represent a reflector in the model space as a collection of points $l_m = (x_m, z_m)$, where x_m is the horizontal dimensions, and z_m is the depth. In the time domain, the same reflector is represented as a collection of points $l_d = (x_d, t_d)$, where t_d is the two-way travel time from the source to the reflector, and V is the velocity along the ray path from the source to the reflector, which is a function of the velocity model \mathbf{m} . The migration and demigration are now reduced to simple relationships:

$$x_d = x_m \quad (3.7)$$

$$t_d = 2z_m/V.$$

Because the process of event picking in the data domain introduces errors, the complete forward model (or likelihood function), $p(l_d | l_m, V)$, for the migration problem has the form

$$x_d = x_m \quad (3.8)$$

$$t_d = 2z_m/V + n_p,$$

where n_p are random variables drawn from a distribution characterizing the pick errors. Observe that the likelihood function now incorporates two sources of migration error: the uncertainty in the assumed velocity model and the event picking errors. By applying Bayes' theorem in Equation 3.3 the posterior of the reflector location in

the depth domain given recorded reflection times is

$$\begin{aligned}
p(l_m | l_d) &= \int p(l_m, V | l_d) dV \\
&\propto \int p(l_d | l_m, V) p(l_m, V) dV \\
&= \int p(l_d | l_m, V) p(l_m) p(V) dV \\
&= \mathbb{E}_V [p(l_d | l_m, V) p(l_m)],
\end{aligned} \tag{3.9}$$

where $\mathbb{E}_V[X] = \int X p(V) dV$ denotes the average over all velocity models obtained in the previous section, and $p(l_m)$ is the prior distribution of the reflector location that is independent of the background velocity model and reflects our approximate understanding of possible reflector locations.

We now combine the method described in this section with the Metropolis-Hastings algorithm described in the previous section to quantify the uncertainty in a migrated image given raw data. First, we use the field expansion method with Markov Chain Monte Carlo sampling to estimate the distribution of velocity models given low frequency data. Second, using this distribution of velocity models we use the above framework to quantify uncertainty on migrated images.

Although we have presented a full framework for Bayesian uncertainty quantification which accounts for noise in migration time picks as well as uncertainty in the velocity model due to noise in the raw recorded traces, in the following results section we assume that reflected events are picked exactly because the effect of the velocity uncertainty in our experiments dwarfs the effect of other noise. When we assume that there is no noise, $p(l_d | l_m, V_i)$ becomes a Dirac delta function for a given velocity model V_i and determining the distribution of migrated images only requires MCMC sampling of the velocity model distribution. The impact of migration error picks is relatively benign and including its uncertainty would mask the effects of velocity model uncertainty.

3.6 Results

3.6.1 Simple Anticline

In this section we demonstrate the complete workflow of the uncertainty quantification framework on the anticline velocity model shown in Figure 3-1. For the purposes of illustration, in this example we have decreased the number of sub layers to 1 such that the velocity within each master layer is constant instead of a gradient. Using the field expansion method, we generate synthetic data at 5 Hz for 256 sources and a single shot on the upper surface of the model. We then add independent Gaussian noise to the observed data such that the SNR is 1.2, where SNR is defined as the ℓ_2 norm of the true signal divided by the ℓ_2 norm of the noise. For this inversion we set the prior distribution to be uniform with hard limits on the permissible size of the interface perturbation (1,000 m), layer velocities (250 to 4,000 m/s), and interface depths (250-3,000 m). Next we run the adaptive Metropolis-Hastings algorithm for 100,000 iterations initiating the algorithm with the known true velocity model to quantify its uncertainty as the algorithm should not depend on the initial starting model. See the later section ‘Starting Model Sensitivity’ for a discussion of the impact of using initial models that differ from the true velocity model. As explained in the section ‘Efficient Posterior Calculation’, we discard the first 50,000 iterations to reduce the impact of the initial starting model, which can bias the results. Figure 3-3 shows several of the velocity models drawn at random from the remaining 50,000 samples of the distribution.

Figure 3-4 shows the mean velocity model as well as the standard deviation of the layer velocities and boundaries. Although this figure gives some idea of the uncertainty, it cannot show the correlation between different errors and this may somewhat misrepresents and even overstates the magnitude of the uncertainty in the inverted model. By examining Figure 3-3, we see that although the midpoints and depths of layer boundaries vary significantly the shapes are relatively constant and velocities within a layer will increase or decrease to compensate for a change in layer thickness. In addition, we find that the velocity and layer boundary uncertainty

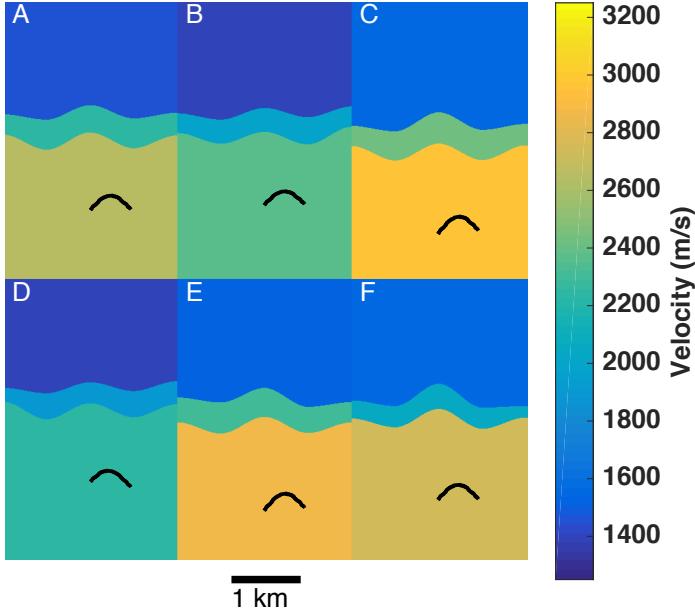


Figure 3-3: Six representative velocity models drawn from the results of the MCMC algorithm run on the model shown in Figure 3-1. Each velocity model extends to a depth of 3,000 m and 2,500 m horizontal distance. The resulting position of the migrated anticline reflector is also shown.

increases with depth as expected.

Using this empirical distribution of velocity models we perform zero-offset time migration on an anticline reflector embedded in the bottom-most layer of our velocity model as shown in Figure 3-1B. For simplicity and to more strongly illustrate the effects of velocity model uncertainty, we assume that the time picks have no associated errors and have been picked perfectly. Errors in the time picks and more sophisticated migration methods could be integrated into the uncertainty quantification framework using the method described in the previous section and in detail in [53], but we do not address these errors in this example. The resulting standard deviation of the migrated image is shown in Figure 3-4, where the black dashed lines show the one sigma interval. Due to the uncertainty of the layer velocities the migrated reflector depth is significantly less certain than the layer boundaries as shown in the same figure. Again, this representation of error does not fully capture the true uncertainty of the migration as one could imagine fitting a nearly flat reflector inside the migrated black error bars of the anticline reflector in Figure 3-4. However, if one looks at Figure

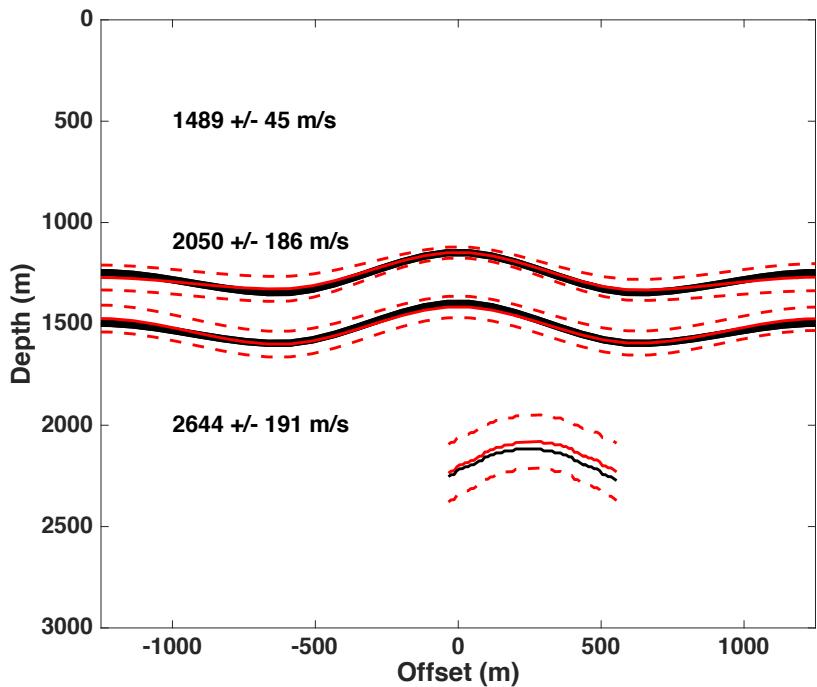


Figure 3-4: The mean velocity model from the MCMC algorithm for which six random models are shown in Figure 3-3. The solid black lines show the true position of the layer interfaces and reflector. The standard deviations of the layer boundaries are shown with the red dashed lines and the velocity uncertainties for each layer are shown. The mean position of the migrated reflector (at about 2200 m depth) and its standard deviation are also displayed.

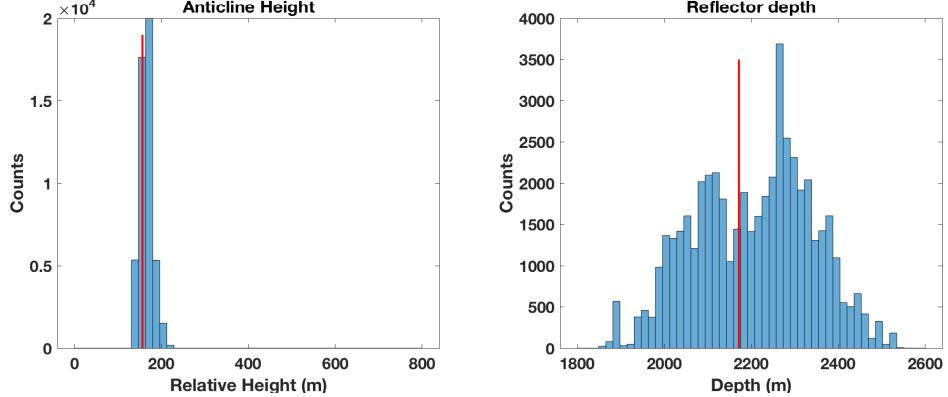


Figure 3-5: Left: Histogram of migrated anticline relative height. Right: Histogram of migrated anticline absolute depth. Both histograms are shown with the same horizontal scale and bin size. The red vertical line denotes true values of the quantities of interest in Figure 3-1B.

3-3 showing the migration for several realizations drawn from MCMC uncertainty quantification, the depth of the reflector varies widely while its shapes remains nearly unchanged. In order to fully understand the uncertainty of both the velocity model and migrated image, one needs to view a series of images of the distribution which is difficult and time consuming to view or present. In [39] the authors generated a movie consisting of snapshots of the migrated imageâŽs distribution to convey the uncertainty of the migration.

Instead of presenting the full model and image as shown in Figure 3-4, which can misrepresent the error, we are often interested in a small aspect of the migrated image such as the depth or dip of a given reflector and do not need an image describing the full uncertainty of the model. For example, Figure 3-5 shows the distributions of the depth to the anticline reflector and the height of anticline across the entire population of migrations. Although the depth of the anticline varies significantly 2200 ± 135 m the relative height is extremely stable 165 ± 15 m, illustrating that certain quantities of interest like the cross sectional area of a trap can be significantly more stable than the absolute depth.

3.6.2 Gradient Migration

We apply our uncertainty quantification framework to a complex gradient based model to demonstrate the wider applicability of our approach. In order to do so we smooth the Marmousi velocity model [78] shown in Figure 3-6A with a low pass Gaussian filter as shown in Figure 3-6B. Based on this smoothed model, we manually generate a layered gradient approximation suitable for our field expansion modeling method, shown in Figure 3-6C. The gradient layered approximation is parameterized by 4 layers divided into 10 sub layers with 7 nodes per layer describing the shape of each interface as shown in Figure 3-2. The gradient layer model well approximates the low pass pixelized model, but uses only 32 parameters instead of the 46,848 parameters used in the full smoothed model.

Using the gradient layered model and the field expansion method, we generate synthetic data at 3 Hz with a single shot and 512 equispaced receivers. We then add uniform Gaussian noise such that the SNR is 0.75, Figure 3-7 shows the true and measured field used in this example. For all of the inversions using the gradient model we set the prior distribution to be uniform with hard limits on the permissible size of interface perturbations (1,000 m), gradient minimum and maximum velocities (1,000 to 6,000 m/s), and interface depths (500-4,000 m). Next we run the adaptive Metropolis-Hastings inversion for 400,000 iterations with 8 initial starting models perturbed slightly about the true model. We perform 8 separate inversions with 8 different starting models in order to examine their convergence as described in subsequent sections. We then discard the first 200,000 samples to reduce the impact of the starting model and initial choice of \mathbf{C}_0 .

Because the velocity model is too complex and the bottom-most sections of the model are ill-determined, it is uninformative to show the mean and standard deviation of the layer boundaries as in Figure 3-4. If one were to show an analogous figure for the gradient Marmousi velocity model, the means of the bottom most layers would overlap. Instead we need other ways to visualize and make use of velocity model uncertainty such as viewing a collection of samples from the posterior distribution of

velocity models or quantities of interest. Figure 3-8 shows 6 velocity models randomly drawn from the chain. Note that as expected, the deeper velocities vary more than those in the upper sections of the model. We can further interpret the velocity model error through rapid forms of imaging. To do so we generate a reflectivity model from the Marmousi velocity model and then use the layered gradient model in Figure 3-6 Right to generate traveltimes. Using these traveltimes, we perform zero-offset migration with the entire distribution of velocity models, using the framework described above. Figure 3-9 shows the resulting migrations for the 6 velocity models shown. From this representation it is clear that some aspects of the image are significantly more stable than others and that this stability correlates with depth. For example in some realizations the salt flank in the bottom left of the image could be interpreted as a potential trap if migration is performed with a single velocity model. In Addition, the shallow blue anticline's shape appears to be significantly more stable than the deeper red anticline.

This becomes more apparent when, as in the previous section, we view a histogram of the quantities of interest as shown in Figure 3-10. This figure shows histograms of both anticlines' depths and cross sectional areas for the 8 Markov-chains with the first 200,000 samples removed to generate a total of 1.6 million samples from the posterior distribution along with the true values of each quantity of interest denoted by red vertical lines. The predicted means of the distributions do not perfectly align with the true values as the inversion is performed on the noisy data shown in Figure 3-7. However, all true quantities of interest are within 1.2 standard deviations of the estimated mean value. Although some quantities could be well approximated by Gaussian distributions, several have asymmetrical distributions. This illustrates the importance of careful uncertainty quantification as the deeper anticline, which has a higher average area, is significantly more uncertain than the shallow anticline. We also note that a significant fraction of realizations have near zero cross sectional area.

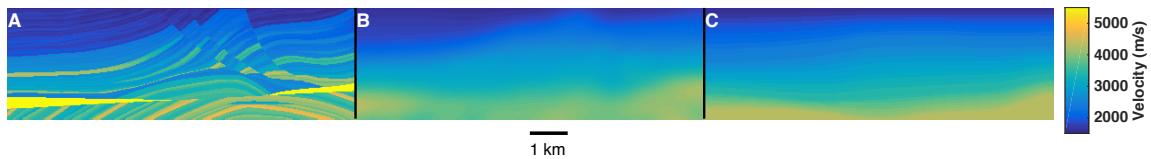


Figure 3-6: **A:** 122x384 pixel Marmousi velocity model. **B:** Smoothed Marmousi model generated by applying a Gaussian low pass filtered to the original Marmousi model. **C:** Field expansion model based on the smoothed Marmousi model.

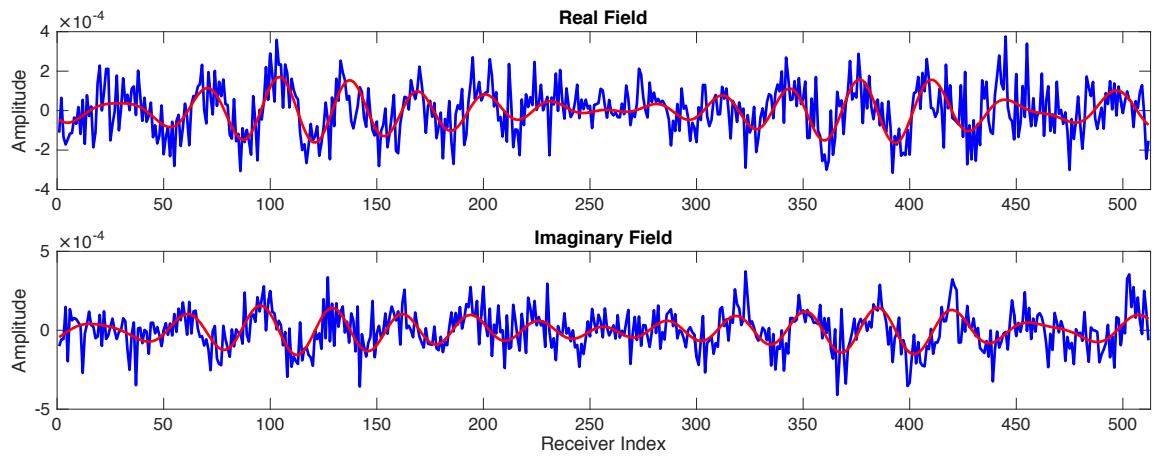


Figure 3-7: Real and imaginary components of the measured field (blue) true simulated field (red) at 3Hz.

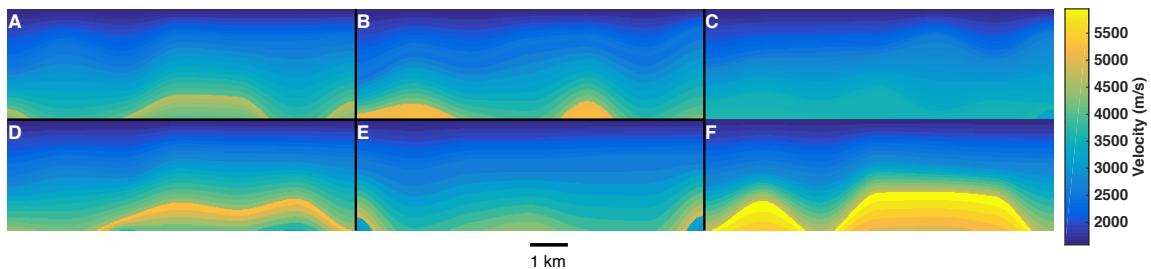


Figure 3-8: 6 Velocity models drawn from the Markov-Chain; the true model is shown in Figure 3-6C. Note that deeper sections of the velocity vary more than shallow parts.

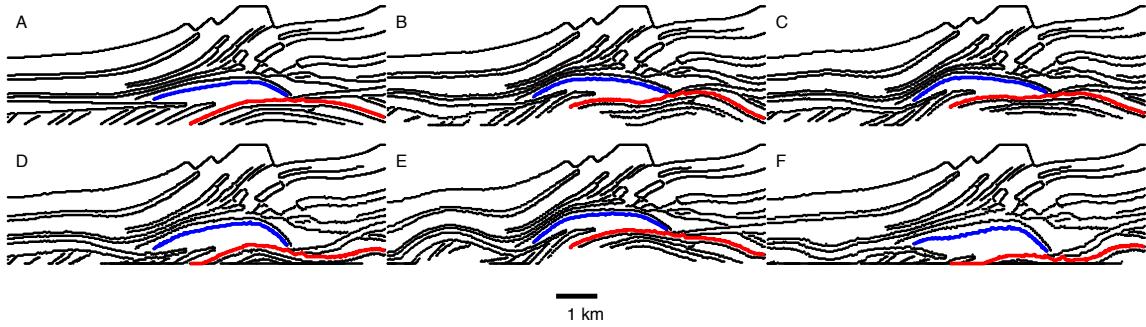


Figure 3-9: Zero-offset migration using the velocity models shown in Figure 3-8 using the true Marmousi reflectivity. Two potential traps are highlighted in blue and red. We characterize the uncertainty of the cross sectional area and depth of these two traps.

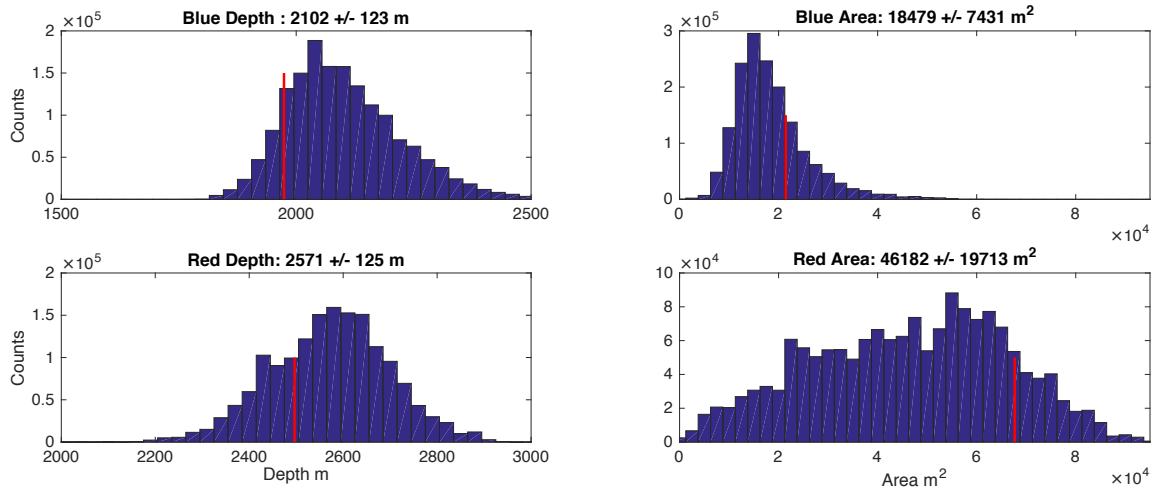


Figure 3-10: Histogram of the shallow (red) and deep (blue) anticlines depth and area. The Metropolis-Hastings runs were initialized with the true starting model. The red vertical line denotes true values of the quantities of interest in Figure 3-10.

Quantity of interest:	Blue depth	Blue area	Red depth	Red area
Initiated near true starting model	0.97	0.92	1.05	0.99
Initiated from linear gradient	0.98	0.92	1.01	1.02

Table 3.1: Table showing the \hat{R} values for the 4 quantities of interest initiating the Metropolis-Hastings with starting models close to the true starting model and a linear gradient. An \hat{R} value of less than 1.1 indicates convergence.

3.6.3 Convergence

In order to determine convergence of the Markov-chains we use the [23] \hat{R} criteria, where one compares the the average variance of a quantity of interest within each chain to the variance of mixture of all of the chains. For a collection of M Markov chains of length N samples we discard the first $N/2$ samples and compute the variance of each individual chain var_m for each quantity of interest and the variance of all the chain mixed together var_{mix} for the remaining $N/2$ samples for each chain. We then define \hat{R} as

$$\hat{R} = \sqrt{\frac{\text{mean}(var_m)}{var_{mix}}}. \quad (3.10)$$

Typically one views at \hat{R} as function of iteration and the chain is said to have converged for a given quantity of interest when \hat{R} is less than 1.1 [6]. Figure 3-11 shows \hat{R} as function of iteration for a collection of 8 chains for several uncertainty quantity inversion with 2-5 master layers approximating the velocity model shown in Figure 3-6B. For the examples used in ‘Gradient Migration’ section using starting models chosen as a perturbation about the true model. The final \hat{R} values are given in Table 3.1. All of the \hat{R} values are well below 1.1 indicating convergence.

3.6.4 Starting Model Sensitivity

In the previous section we unfairly initiated the starting model of the adaptive Metropolis-Hastings algorithm with the true model. Although this may be useful in the context of survey design, it is an unrealistic assumption for most real data contexts. In addition, MCMC algorithms should converge regardless of the initial

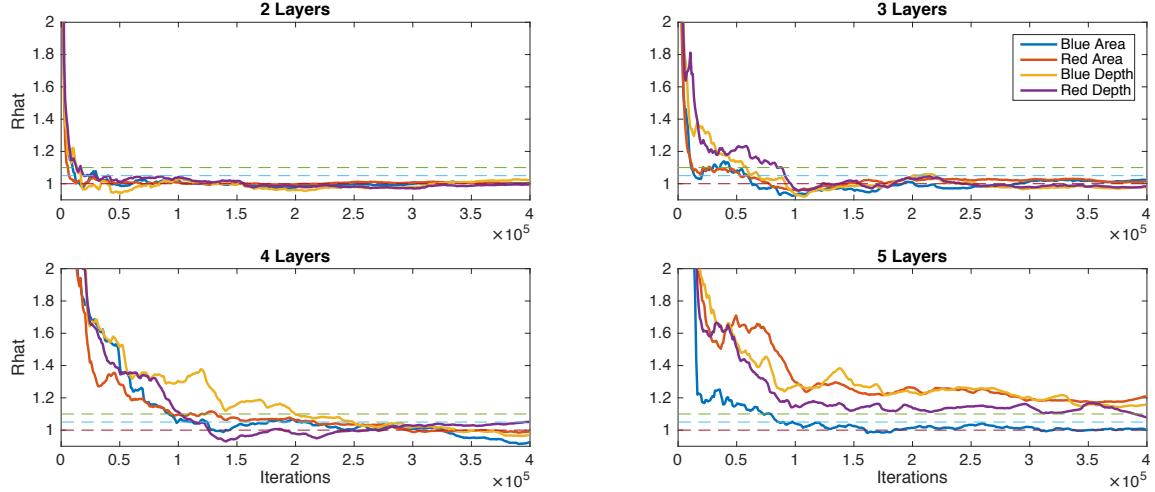


Figure 3-11: Convergence plot for Metropolis-Hastings inversions using 2-5 master layers using 7 nodes per interface. Each plot shows \hat{R} for the 4 quantities of interest as a function of iterations. the 4 models shown had the following number of degrees of freedom: 2-layers 11, 3-layers 21, 4-layers 31, and 5 layers 41.

starting model. We repeated the Metropolis-Hastings procedure as in the previous section 14 times but instead of using the true model as the starting model we used a linear gradient from 0 to 4300 m/s instead of the true model containing perturbations and increased the total number of iterations to 500,000 and discarded the first 250,000 samples. In addition, We removed the 6 chains with the lowest acceptance rate as initiating the chain with incorrect models often leads to premature convergence at local minima. Figure 3-12 shows the histograms for the quantities of interest for the 8 remaining chains using the linear gradient instead of the true velocity. The posterior distributions using the linear gradient closely match the distributions achieved with the true initial velocity model. In addition, the final \hat{R} for the 4 quantities of interest, given in Table 3.1, indicate convergence as they are well below the 1.1 criteria.

To further determine the robustness of our algorithm to incorrect starting models, we initiated the MCMC algorithm with several incorrect linear gradients in which the top velocity was fixed at 1500 m/s and the bottom velocity varied from 2,800 m/s to 5,800 m/s, which is equivalent to a -35% or +35% variation from the true linear gradient approximation, and ran the algorithm 14 times at each velocity gradient for 500,000 iterations and again removed the 6 chains with the lowest acceptance rate.

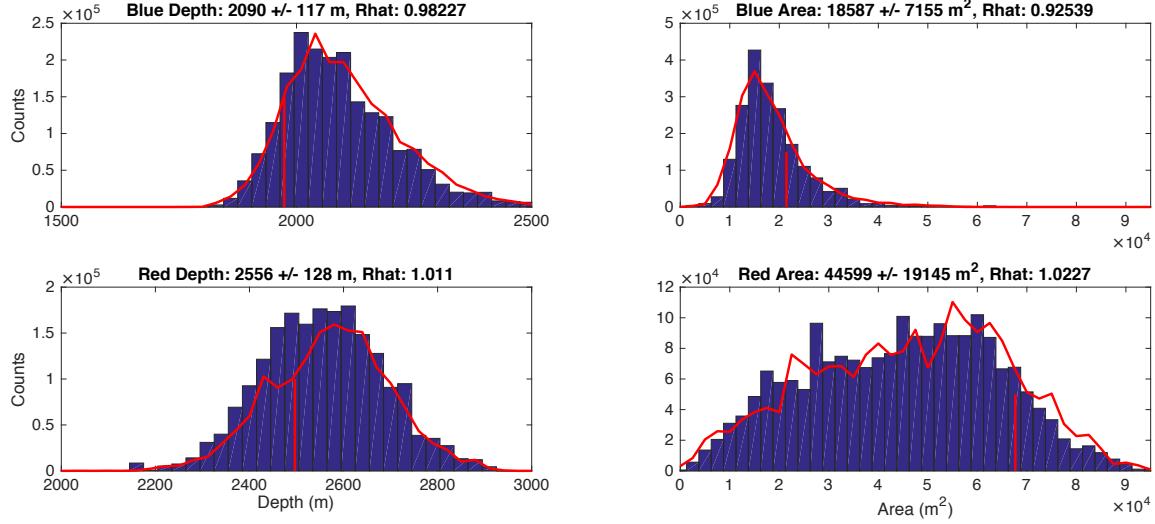


Figure 3-12: Histogram of anticline depth and area using a linear gradient initial model. The histograms using the true velocity as the starting model from Figure 3-10 are overlaid in red. The red vertical line denotes true values of the quantities of interest in Figure 3-10.

This procedure left 8 chains remaining and 2 million samples once the first 250,000 samples of each chain were discarded for burn in. Figure 3-13 left shows the final \hat{R} for the 4 quantities as a function of velocity perturbation to the bottom gradient. The right panels of Figure 3-13 show the zero velocity perturbation histogram with overlays of the distributions initiated with incorrect velocity perturbations of -25%, -15%, +15%, and +25%. Good convergence was achieved for all 4 quantities of interest with velocity perturbations from -25% to +35%. The area of the blue anticline appeared to converge extremely well for all starting models and this convergence is especially apparent in the top right panel of Figure 3-13 which shows that the overlays of the blue anticline area are nearly identical for all starting models. The uncertainty of the remaining quantities of interest appears to be well estimated for incorrect starting models. These results suggest that our framework is relatively robust to incorrect starting models and can tolerate incorrect estimates of roughly plus or minus 25%. Beyond this perturbation the parameter space appears to be too large and empty for the algorithm to find the maximum likelihood model in a reasonable number of iterations.

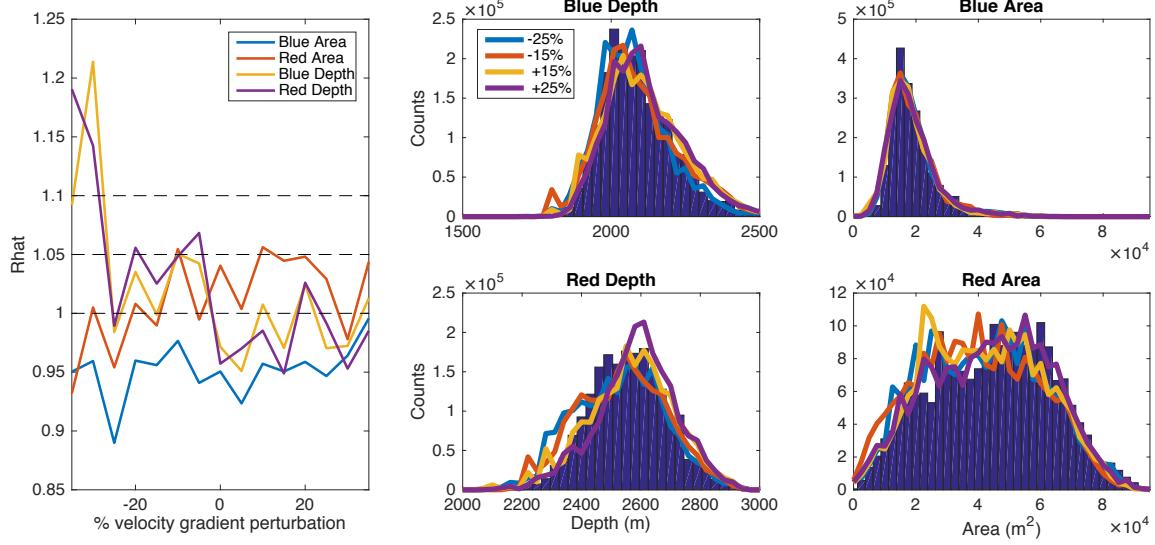


Figure 3-13: **Left:** The final \hat{R} for the 4 quantities of interest as a function of bottom velocity perturbations. **Center:** Histograms of anticline depth for the a zero perturbation linear gradient starting model with overlays of velocity perturbations of -25% , -15% , $+15\%$, $+25\%$. **Right:** Histograms of anticline Area for the a zero perturbation linear gradient starting model with overlays.

3.6.5 Degrees of Freedom

In the previous examples the synthetic true data and inversion data were all generated with 4 master layers and 7 nodes per interface describing the shape of the interface. To test the impact of degrees of freedom on convergence, we ran 8 Metropolis-Hastings runs for 400,000 iterations at each combination of 2 to 5 master layers and between 3 to 7 nodes describing each interface. All inversions were run with the initial models randomly perturbed about the true starting model. We then calculated \hat{R} for each quantity of interest as function of iteration and show these in Figure 3-11. Dashed lines indicate an \hat{R} of 1.1, 1.05, and 1. From Figure 3-11 it is apparent that the number of layers strongly influences the convergence rate, with the 2 layer model converging rapidly in well under 50,000 iterations and the 5 layer model not converging in the allotted iterations except for the blue anticline area. The 3 and 4 layer models converge at approximately 100,000 and 250,000 iterations respectively for all quantities of interest. Although not shown, the number of nodes similarly impacts the convergence rate with fewer nodes typically increasing convergence speed but the

Number of master layers	2	3	4	5
Total number of layers	11	21	31	41
Runtime (hours per 10^5 solves)	1.88	3.26	5.05	6.79

Table 3.2: Computation time for the field expansion forward models.

effect is significantly less pronounced than that of the total number of master layers.

3.6.6 Computational cost

The majority of the simulations shown in this section were run in parallel on a cluster where the specifications of each machine vary. Because there is no communication between each individual chain the algorithm can easily be distributed to a number of independent cores on a cluster. In order to more accurately capture the computational cost of the algorithm, we profiled the MATLAB code on a single core of a Macbook pro with a 2.6GHz intel Core i7 processor with 16 gigabytes of ram. The run time per 100,000 iterations is given in Table 3.2. The longest running chains we present could be run on the author’s laptop in roughly 27 hours. For a rough comparison we also used the PySIT [33] time domain finite difference solver to generate a 3 second shot for the 122x384 velocity model shown in Figure 3-6C on the same machine. This simulation took approximately 6.5 seconds to run and would take 180 hours to 100,000 simulations. The longest chains presented in this chapter would take roughly 5.4 weeks to run with time domain finite difference versus 27 hours with the field expansion.

3.7 Discussion

Uncertainty quantification is a useful additional step that should follow traditional velocity model building and/or imaging. Putting error bars on quantities of interest is indispensable for assessing the reliability of our conclusions and de-risking seismic processing-driven decisions. But characterizing and presenting the results of the uncertainty quantification of velocity models and migrated images creates conceptual and implementational challenges. A naïve approach to uncertainty quantification

that seeks to answer all possible questions will likely fail. Smart uncertainty quantification entails first asking the right questions about the right quantities, and then using extremely fast forward models and intelligent statistical sampling techniques.

In this chapter we demonstrate that our uncertainty quantification framework is robust to initial starting models and that the method can be initiated from a linear gradient that can deviate by up to 25 percent from the true velocity model. Initiating starting models far from the true starting model slows convergence and it may be more efficient to pursue a hybrid approach in which a global optimization method, like particle swarm optimization or simulated annealing, is used to find an initial starting model close to the true model.

We find that the method converges for complex velocity models with up to 31 parameters but failed to converge for models with 41 degrees of freedom in a reasonable number of iterations. In order to increase the number of degrees of freedom this approach could handle one could potentially adopt more advanced MCMC methods that accelerate mixing. We did not find that using a tighter velocity model prior distribution to improves convergence but we do find that applying a prior only accelerated convergence if the data was highly uninformative and the posterior becomes roughly equivalent to the prior.

Given these results and the reduced number of degrees of freedom our algorithm is able to handle, it is unlikely that we can fully sample the velocity model at the minimum resolvable wavelength. For future inversions, it may be judicious to either reduce the scale of velocity model and/or the forward model frequency to better capture the full velocity model distribution.

Due to our choice of forward solver for rapid computation, our approach is limited in the velocity structures we can model. Our framework cannot easily model discontinuous layers and shapes such as salt bodies. In future work different forward solvers could easily be incorporated into our framework as long as the solvers were both computational efficient for the tens or hundreds of thousands of iterations needed for uncertainty quantification and possible to describe through a reduced parameterization.

Conclusions

We presented a Bayesian uncertainty quantification framework together with a fast algorithm to quantify errors in a velocity model and in a seismic image in a meaningful and numerically efficient way. We successfully tested the framework on both a simple and complex synthetic example and found the framework to be robust to both the choice of the initial model and parameterization of the model.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant DGE-0806676 and MIT's Earth Resources Laboratory. It also incorporates work supported by Total in the past. Alison Malcolm is grateful for financial support from Chevron, NSERC and the Research and Development Corporation of Newfoundland and Labrador (now InnovateNL).

Chapter 4

Imaging the Interior of Bone

4.1 Overview

Imaging through bone is particularly challenging due to the large impedance and velocity contrast between soft tissue and bone that results in strong multiply reflected waves and other complex wave propagation. In this chapter we present several methods for imaging the interior of bone through a novel suppression of multiply scattered waves using a group sparse hyperbolic radon transform and applying techniques from seismic imaging such as time migration that can image in complex velocity models. In addition, we present several synthetic finite difference models to illustrate and understand the challenges posed by imaging through bone in two acquisition geometries. We demonstrate our proposed method on synthetic and *in vivo* data.

4.2 Introduction

As medical imaging becomes more prevalent and relied upon, reducing both radiation exposure and expense of imaging is highly desirable. Although ultrasound is not typically used to image bone, the decreased expense, compactness, portability, and lack of ionizing radiation make it ideal for image guided surgery and frequent monitoring of existing conditions where constant x-ray exposure may create a health risk and Magnetic Resonant Imaging (MRI) may be too costly.

Using ultrasound to image through and around bone introduces complicated physical phenomena that are poorly imaged with conventional modalities. Previously ultrasound has been used to detect surface defects and fractures in bone [3] but imaging beneath the surface remains challenging. Typical ultrasound imaging is based on a single scattering assumption in an isotropic uniform velocity model [65]. Bone has a significantly higher velocity (3,000-4,100 m/s) [27] than soft tissue (1,400-1,700 m/s) [65], causing acoustic ray paths to deviate significantly from the ray paths traveled in soft tissue. This high contrast in acoustic impedance leads to strong multiply scattered waves (also called multiples) that further complicate imaging. Due to the fine porous structure of bone, this velocity can be anisotropic in that the speed of acoustic propagation is directionally dependent with the fast axis being aligned with the long axis of the bone [27, 4] as shown in Figure 4-1. In addition, bone can act as a waveguide generating strong head and guided modes that are poorly modeled by ray based methods and yet can be more energetic than reflected arrivals. [74] exploited these guided waves to estimate long bone velocity and thickness. [56], accounted for the acoustic anisotropy and complex velocity models to image bone. However, [56] find that imaging beneath the bone is challenging due to strong multiples, complex wave paths, and high speckle noise created by sub-wavelength heterogeneities. In this chapter, we use imaging techniques from seismology that account for these more advanced physical phenomena to significantly improve ultrasound's ability to image through bone.

Seismologists frequently encounter complex and unknown velocity models in the Earth that generate similar phenomena as bone in medical ultrasound. The challenges generated by bone in the body are analogous to the complications generated by salt bodies in seismic imaging. Salt bodies can have irregular continuous shapes and the seismic velocity of salt is extremely fast compared to sedimentary rocks conventional imaged. The contrast between salt (4,500-5,500 m/s) & sedimentary rocks (2,000-3,500 m/s) is similar to the contrast between soft tissue (1,400-1,700 m/s) & bone (3,000-4,100 m/s) and thus generates similar challenges when attempting to image through them. Exploration seismology has developed a wide array of techniques

to image through salt such as Reverse Time Migration (RTM) [1, 87] that we can leverage to image through bone. RTM imaging is better able to resolve the bottom of salt structures compared to Kirchoff migration methods, which are based on ray travel times [7]. Reverse time migration has recently been used for ultrasonic imaging to diagnosis and detect breast cancer [59].

It is not only velocity contrast that has similarities to seismic. Because many rocks are deposited in layers, they tend to be anisotropic and seismologists have developed a wide array of techniques for imaging in anisotropic mediums [73]. [56], modified the typical delay and sum algorithm to account for varying velocities and anisotropy of the bone based on anisotropic propagation models used in seismic imaging. Using these techniques, the authors successfully imaged the near side periosteum (outer surface of the bone cortex) & endosteum (inner surface of the bone cortex) of the radius and tibia. However, multiply scattered waves, complex wave paths, and high speckle noise prevented them from imaging through the marrow and recovering the far side endosteum. Multiples are a commonly encountered problem in seismic imaging and a wide array of techniques have been developed to suppress them; see [77] for a relatively recent overview. To both estimate cortical shell thickness and potentially study the interior structure of the marrow these multiples must be suppressed.

In this chapter, we leverage numerous techniques from seismic and sub-salt imaging to suppress these multiples and improve imaging in two acquisition modalities: axial and transverse as shown in Figure 4-1. First we use finite difference simulations to demonstrate the challenges presented by this imaging problem in the axial geometry and develop techniques to image through the bone on synthetic data sets. In the axial modality, multiples present the largest challenge. We then demonstrate the proposed framework on in vivo data from the radius and tibia. Second, we examine finite difference simulations of the experiment in the transverse geometry in which the lack of illumination and complex velocity model makes it difficult to recover an accurate image of the interior with simple ray based imaging techniques. We then compare the results of reverse time migration and ray based migration techniques on the synthetic and in vivo data. Throughout the chapter we make use of several

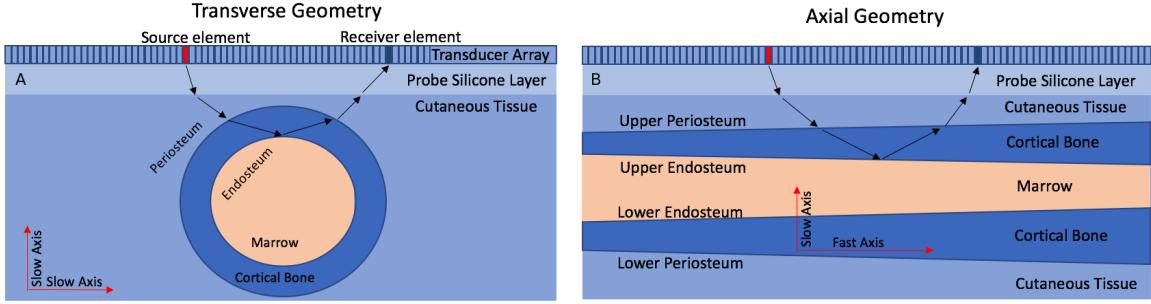


Figure 4-1: Geometry of the experiment. Each element acts sequentially as a source with the resulting wavefield recorded on all the elements. **A:** Transverse acquisition. **B:** Axial acquisition.

seismic techniques and terminology that are described in more detail in the appendix.

4.3 Acquisition Geometry

All of the results shown in this chapter were acquired with a synthetic aperture acquisition [35] in which each transducer element is activated one at a time and all transducers record the resulting reflections. This method of acquiring ultrasound data is analogous to conventional 2D exploration seismic acquisition in which an explosive source generates a spherical wave and a linear array of hydrophones or geophones record the subsurface reflections. We use a Philips P4-1 transducer with 96 elements spaced 0.295mm apart sampling $35 \mu\text{s}$ at a rate of $.1\mu\text{s}$. For each survey this generates a 3D data cube of 96 source x 96 received x 350 time samples.

4.4 Applications of PSO & MCMC to Ultrasound

Applying the methods presented in Chapters 2 & 3 proved difficult for the ultrasound datasets described in this chapter. As noted in the previous section, the bone is anisotropic and this propagation is not captured by the field expansion forward model. In addition, we found that the data is under sampled in the offset direction as shown in Figure 4-3 comparing a synthetic field to a measured *in vivo* field at 2 MHz. The measured data appears spiky and misses many of the features of the more finely

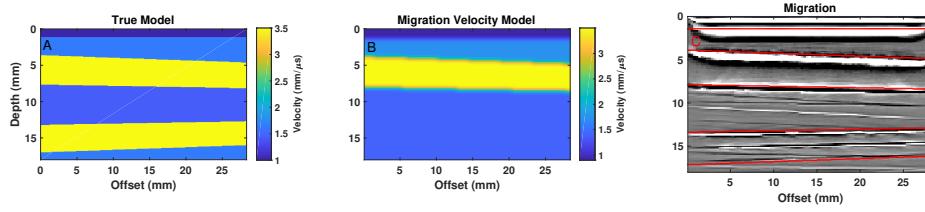


Figure 4-2: **A:** True velocity model used for synthetic experiment. **B:** Smoothed velocity model containing only the upper bone used for Kirchoff migration. **C:** Kirchoff migration using the smoothed velocity model with true bone interfaces highlighted in red and the silicone tissue interface highlighted in black at roughly 1.5 mm. The artifacts at the source are a normal part of any imaging method that can be suppressed if necessary by muting the direct arrival. Numerous multiple reflections are present in the migrated image within the marrow; these are the lines in the image that don't correspond to true reflectors.

sampled synthetic data.

To further explore the feasibility of using the field expansion to invert the ultrasound data we ran 20 PSO inversions on noiseless synthetic data at 2Mhz having the same acquisition geometry and velocity model as the in vivo dataset. Each inversion was run for 250 iterations with swarm size of 40 agents. For simplicity we only simulate the upper most part of the cortical shell and restrict the velocity model to consist of flat layers. Figure 4-4 shows the resulting inversions for the 20 runs and the true velocity profile. From this figure we see that the PSO inversion was able to recover the velocity above the top of the bone but was unable to recover the lower structure of the velocity model. The inverted bone velocity, its thickness and the marrow velocity vary significantly across inversions. These results suggest that this experimental setup is likely poorly suited to invert for bone velocity model even in synthetic data.

To more rigorously test the limits of the geometry and resolution of this setup, we take the velocity profile shown with the dashed line in Figure 4-4 and generate a

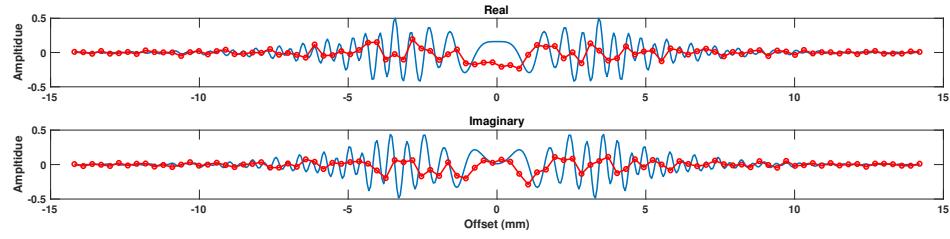


Figure 4-3: **Red:** Field for an in vivo measurement across the 96 transducer at a frequency of 2MHz. **Blue:** Simulated field using the field expansion of a synthetic model based on the in vivo imaging. The upper panel shows the real components of the field and the lower panel shows the imaginary components.

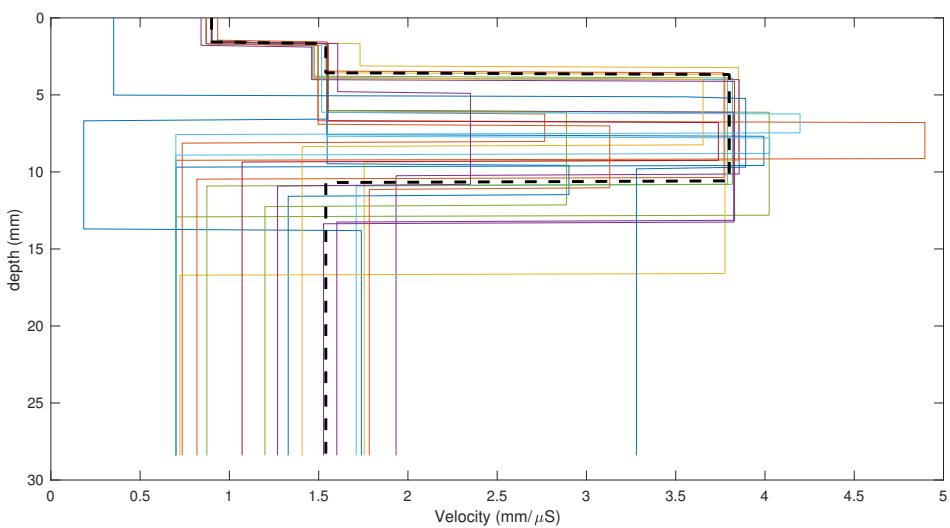


Figure 4-4: **Dashed:** The true velocity profile for the synthetic bone experiment. **solid:** The best fitting velocity model for the 20 PSO inversion.

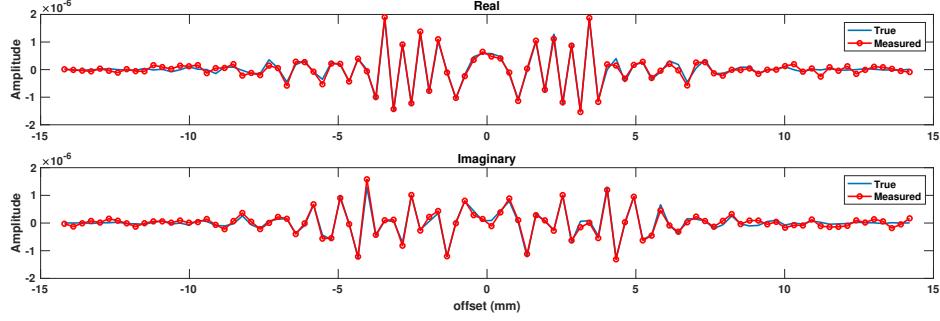


Figure 4-5: **Blue:** Synthetic field across the 96 transducers at a frequency of 2 MHz without noise. **Red:** Simulated field with small amount of noise used for MCMC inversion.

synthetic measurement using the field expansion method at 2 MHz. We then add a small amount of noise to the field as illustrated in Figure 4-5. This amount of noise is a very conservative estimate of the noise observed from *in vivo* data and is used to provide an estimate of the lower bounds of our ability to invert this dataset at a single frequency. Based on this noisy data we perform MCMC inversion as described in chapter 4, initiating the chain with the true velocity model and running the inversion for 50,000 iterations. Because our velocity profile has fewer than 10 parameters running the inversion for 50,000 iterations was likely sufficient as suggested by Figure 2-7. Tables 4.2 & 4.1 summarize the posterior distribution of velocity models given the data shown in red from Figure 4-5. These tables agree with the results shown for the PSO inversion and demonstrate that thickness of the bone and marrow velocity are extremely poorly determined, ± 7 mm and $\pm 1.2 \text{ mm}/\mu\text{s}$ respectively. This strongly suggest that an inversion based on a single frequency of this dataset with the experimental receiver spacing is unable to resolve the velocity model. Based on these finding, we pursue a variety of time domain methods to estimate the bone velocity model and image the interior of the marrow.

	Truth	MCMC
Silicone Velocity (mm/μs):	.9	.8889 \pm .0005
Tissue Velocity (mm/μs):	1.54	1.545 \pm .005
Bone Velocity (mm/μs):	3.8	3.65 \pm .16
Marrow Velocity (mm/μs):	1.54	3 \pm 1.2

Table 4.1: MCMC inversion of bone model velocities.

	Truth	MCMC
Silicone depth (mm):	1.6	1.599 \pm .001
Top of bone (mm):	3.6	3.60 \pm .001
Bottom of bone (mm):	10.6	17 \pm 7

Table 4.2: MCMC inversion of depths of bone model.

4.5 Axial: Methods & Synthetic

In the axial geometry the high impedance contrast between bone and soft tissue generates internal multiples which can corrupt imaging and velocity analysis. Because the reflectors being imaged are planar it is extremely difficult to separate the multiples from primary reflections (singly scattered waves).

To illustrate the difficulties created by these multiples, we generate a synthetic data set using the finite difference solver contained within the seismic imaging package, Pysit, [33] with the velocity model shown in Figure 4-2A. For these simulations we use perfectly matched layer boundary condition for all boundaries. However, the transducer array generates some surface related multiples that are not considered in this chapter and will be studied in future work. Because we use an acoustic model, mode-converted shear waves in the cortical bone are not modeled or used for imaging. We then migrate the data using Kirchoff migration with the velocity model shown in Figure 4-2B to generate an image. This is realistic because the upper sections of bone can generally be imaged without suppressing the multiples [56]. Kirchoff migration generates an image of the subsurface by summing observed data along predicted travel time curves. We give a more detailed description of Kirchoff migration in the appendix. From this migration we see that there are numerous multiples that

are difficult to separate from true tissue and bone interfaces highlighted in red. In the in vivo data sets these multiples are significantly stronger and are more difficult to separate than the synthetics shown. Although the multiples are problematic in this geometry, the primary from the marrow return is clearly visible, suggesting the interior of the bone could be well imaged if the multiples are successfully suppressed.

To further understand these multiples, we generate a series of finite difference synthetic data sets using an isotropic forward model of increasing complexity as shown in the top 3 panels of Figure 4-6. Figure 4-6 middle row shows the resulting source gathers in each model. A source gather is a collection of signals that all share a common transmit location. Although in this geometry the cortical bone tissue is anisotropic, an isotropic model sufficiently illustrates the challenges introduced by multiples. We use a flat layer model to reduce diffraction artifacts generated by the combination of high contrast sloping layers and rectangular grids, causing the sloping interface to generate diffractions as it changes in thickness. This simplification is also somewhat conservative as the multiples are more likely to exactly align with the primaries.

From the middle row of Figure 4-6 we see that the increase in velocity between the silicone and top-tissue interface generates internal multiples. In the bottom row of plots, we show Normal Move Out (NMO) semblance plots that measure the coherent energy (see Equation 4.13 and appendix) along hyperbolic travel time curves corresponding to flat horizontal reflectors embedded within an *average* or NMO velocity given by,

$$t_{NMO}(x) = \sqrt{t_0^2 + \frac{h^2}{4V_{NMO}^2}}, \quad (4.1)$$

where t_0 is the arrival time of the reflection at zero offset, V_{NMO} is the average velocity, and h is offset, the distance between the transmit and receive elements. These hyperbolas are completely characterized by V_{NMO} describing the curvature of the hyperbola and t_0 , the vertex of the hyperbola in the time domain. See [86] or the appendix for a description of how to estimate true material velocities from the NMO

velocity. The multiples mentioned above are also apparent in the NMO semblance plot shown in Figure 4-6 bottom in which the NMO velocity of the multiples asymptotically approaches the tissue layer velocity with each reflection appearing with a period of roughly $1 \mu\text{S}$.

When imaging the tibia, in which the bone cortex can be thick relative to the cutaneous tissue layer, these slower tissue internal multiples, waves that bounce between two or more interfaces before returning the surface, tend to fall on top of the bottom of the bone return complicating velocity analysis and imaging. However, when imaging the radius in which the bone is thinner, this tissue multiples tend to arrive well after the bottom of the bone return. The tissue multiple can also complicate interpretation of the marrow return for both the tibia and radius. Note in addition to the reflections, there is a strong head wave along the upper periosteum interface that can be exploited to estimate the axial velocity and approximate thickness of the cutaneous tissue as described in [56] and reviewed in the next section.

For the marrow flood model in which the bottom of the bone is added to velocity model. In the NMO semblance plot we see the top of the silicone return at ($V_{NMO} = 1 \frac{\text{mm}}{\mu\text{s}}$, $t_0 = 4 \mu\text{s}$) the tissue primary return at ($V_{NMO} = 1.25 \frac{\text{mm}}{\mu\text{s}}$, $t_0 = 5 \mu\text{s}$) and the top of the bone return at ($V_{NMO} = 2.1 \frac{\text{mm}}{\mu\text{s}}$, $t_0 = 5 \mu\text{s}$). The addition of the bottom of the bone not only introduces faster bone internal multiples highlighted here but an array of complex peg leg internal multiples in which energy bounces within the tissue and bone several times. These peg-leg multiples, illustrated in Figure (4-7A), appear in the NMO semblance plots as numerous peaks following the bone arrival. The normal move-out velocities of these multiples can be roughly approximated by assuming that the normal move-out velocity is equal to,

$$V_{M_{NMO}} = \sqrt{mV_{bone}^2 d_{bone} + nV_{tissue}^2 d_{tissue} + lV_{sil}^2 d_{sil}} \quad (4.2)$$

with a NMO time of,

$$t_{M_{NMO}} = md_{bone} + nd_{tissue} + ld_{sil}. \quad (4.3)$$

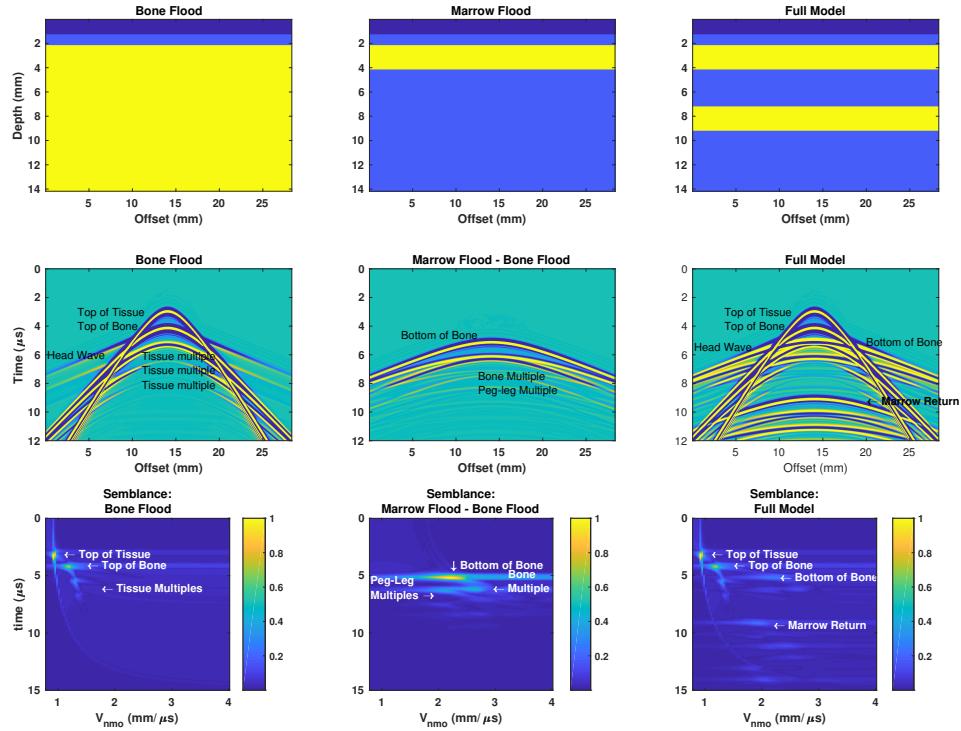


Figure 4-6: **Top:** Three different synthetic velocity models. **Middle:** Synthetic common source gathers. The middle panel labeled ‘marrow flood - bone flood’ shows the difference between source gathers for the bone flood and marrow flood velocity models. **Bottom:** NMO semblance values of synthetic data.

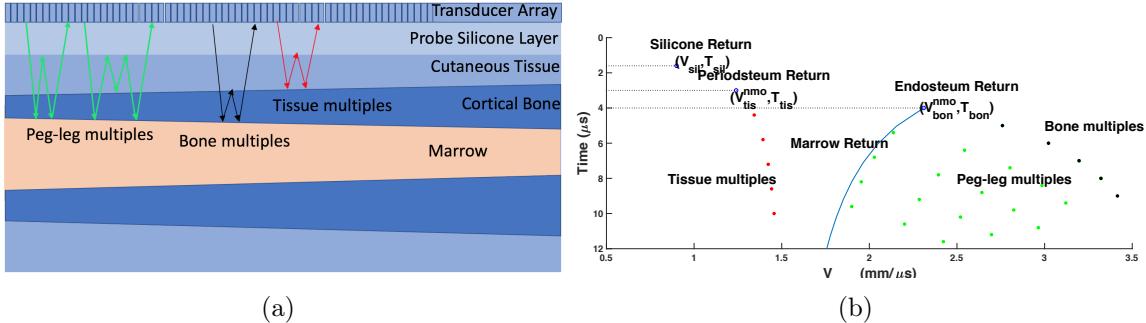


Figure 4-7: **A:** Cartoon illustrating the three types of internal multiples seen when imaging in the axial geometry. **B:** predicted peaks for primaries and multiple in NMO semblance domain. Primaries are shown with blue circles; the potential marrow return should fall along the blue line. The asterisks indicate the predicted move out of the multiples.

Where m, n, l are the number of bounces within the tissue, bone, and silicone layer respectively, $d_{bone}, d_{tissue}, d_{sil}$ are the two way travel times within each of the layers, and $V_{bone}, V_{tissue}, V_{sil}$ are the velocity of various materials.

Using this approximation, we can model the peaks of both the primaries and the multiples as shown in Figure 4-7B in which the primary reflections are denoted by circles and the blue line shows the velocity and time of the marrow return, and the multiples are shown with asterisks. Although the majority of these multiples are separable in the NMO velocity time domain, the peg-leg multiples in which the ray has the same path as the primary reflection with additional bounces in the cutaneous tissue layer have a NMO velocity nearly identical to the NMO velocity of the return from the bottom of the marrow. Because these multiples have extremely similar velocity move-out as the primaries, migration tends to focus these multiples. However, these multiples occur at regular period of d_{tissue} after the endosteum arrival and can be removed using predictive deconvolution [52] that suppresses periodic signals as described in the next section.

Thus far, we have demonstrated that multiples exist and are expected to cause problems in imaging the marrow and deeper layers of the bone. The challenge in suppressing these multiples is to remove them without changing the primaries that they may obscure. In the next section, we present a three step framework for suppressing

these multiples on synthetic data in which we first use conventional seismic velocity analysis tools to build a model of the upper section of bone. Second, we use a sparse hyperbolic radon transform to separate the majority of the multiples based on the velocity model we construct. Finally, we apply predictive deconvolution to remove the remaining periodic peg-leg multiples.

4.6 Methods: Multiple Suppression

From the synthetic Kirchhoff migration shown in Figure 4-2C, we see that there are numerous multiples present in the migrated image, complicating imaging of the marrow and lower endosteum. These multiples are particularly strong in the *in vivo* case [56] and it is difficult to identify the bottom the marrow without preprocessing the data to remove these multiples. In this section we present a framework for suppressing the multiples and denoising the data.

In order to suppress the multiples in the axial geometry, estimates of the tissue and bone NMO velocity and arrival times, t_0 s, at each trace are needed. To do so we first use the head wave to estimate the axial velocity of the bone, the inclination angle of the bone, and the approximate arrival in the time domain. Using this approximate arrival time we then use NMO semblance analysis to estimate stacking velocities and arrival times. Once we have this model of the upper layers we use it to help to suppress the internal multiples in the hyperbolic radon domain and using predictive deconvolution.

4.6.1 Head wave velocity analysis

The strong contrast in velocity between soft tissue and bone generates a head wave that we exploit to measure the axial bone velocity. Unlike reflected waves with curved arrival times, head waves follow a straight line in the common source domain and can be completely characterized by their $t = 0$ intercept and apparent velocity V or slope. To estimate the velocity we performed a slant stack semblance (see the Appendix and Equation 4.13) analysis in which we calculate the coherent energy along straight lines

in the common source gathers. We perform this analysis on the 10 left and right most source gathers and select the time intercept and V pair that corresponds to the largest semblance value as the estimated up slope and down slope velocities V_d, V_u and corresponding intercept values t_u and t_d . We then invert for the axial head wave velocity, V_h , and inclination angle, θ , of the top of the bone using,

$$\theta = \frac{1}{2} \left(\sin^{-1} \left(\frac{V_{tis}}{V_d} \right) - \sin^{-1} \left(\frac{V_{tis}}{V_u} \right) \right) \quad V_h = \frac{V_{tis}}{\sin \frac{1}{2} \left(\sin^{-1} \left(\frac{V_{tis}}{V_d} \right) - \sin^{-1} \left(\frac{V_{tis}}{V_u} \right) \right)}. \quad (4.4)$$

The t_u and t_d intercepts guide our search for the periosteum and endosteum interface and more precisely estimate the reflection t_0 s. In [56], the authors used a singular value decomposition analysis to obtain similar estimates of the axial velocity from the headwave.

4.6.2 NMO semblance velocity analysis

To perform the NMO semblance velocity analysis we first rearrange the data into common midpoint gathers where each gather is defined by the collection of signals or traces such that midpoint defined by, $m = \frac{x_{src} + x_{rec}}{2}$, is constant for the collection of traces and each trace has a different offset, $h = x_{src} - x_{rec}$. When the data are arranged in this manner both flat and sloping reflectors have a vertex at $h = 0$ as opposed to a common source gather in which the hyperbolas for sloping reflectors will be asymmetric and the vertex will be shifted to the right or left. Performing this analysis in the common midpoint domain simplifies and improves the accuracy of the velocity analysis.

In the common midpoint domain, we apply NMO velocity analysis by calculating the semblance, Equation 4.13 along hyperbolas given by Equation 4.1. Once the semblance is computed for each of the common midpoint gathers we more accurately estimate the tissue thickness and velocity by finding the true tissue velocity and depth that maximizes the sum of semblance across all midpoint gathers. For the purposes of maximizing semblance, we constrain this interface to follow a straight

line. Using the left and right t_0 s estimated from the head wave analysis we use linear interpolation to roughly estimate the t_0 of the tissue bone interface and window the semblance to the expected range of velocities of $.9\text{-}2 \text{ mm}/\mu\text{s}$ and $t_0 +/ - 2.5 \mu\text{s}$. We initiate a maximization scheme with the tissue thicknesses and t_0 s derived from the head wave velocity analysis and a nominal value of $1.65 \text{ mm}/\mu\text{s}$ for a velocity. To estimate the bone velocity, we then repeat this process with a search window of $.1\text{-}10 \mu\text{s}$ after the upper periosteum arrival and an initial velocity of $3.25 \text{ mm}/\mu\text{s}$. The slope of the layers is sufficiently small that its impacts on the recovered velocities are negligible [42]. These estimates of stacking velocity and arrival times are then used for filtering the data in the next section. The delay between the upper periosteum and the silicone cutaneous tissue interface can be used for the delay parameter of the predictive deconvolution filter applied below. Alternatively, the bone velocity can be estimated using the migration based scheme described in [56].

Sparse Hyperbolic Radon

From Figures 4-6 and 4-7 showing the semblance plots, it is clear that a significant number of the multiples can be distinguished from primaries in the NMO domain. In addition because the signal of interest is dominated by a few relatively planar reflectors, we expect only a few peaks in the NMO velocity-time space or equivalently the signal should be compact in this space. In this section we formulate a method to denoise and suppress these multiples by exploiting this compactness or sparsity in the NMO velocity-time domain. This method was first introduced in [18] and is summarized below.

To exploit the sparsity of the signal, we derive a set, Γ , of hyperbolic basis functions such that an observed source or midpoint gather D is a linear combination of several basis functions. In the axial geometry the reflectors of interest are well approximated by several gently sloping planar reflections. Therefore the travel time of a given reflector can be approximated by Equation 4.1. We then let each entry of the dictionary Γ describe the hypothetical signal at a single receiver at index, i , and offset x_i for a planar reflector with hyperbolic travel time curve characterized by NMO time

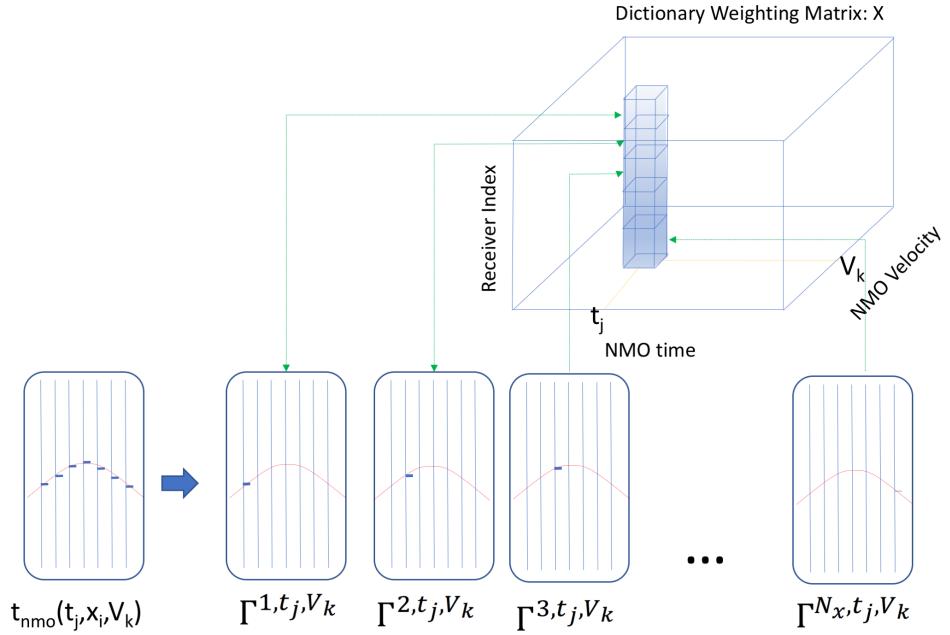


Figure 4-8: Illustration of the dictionary construction and column-wise sparsity of the weighting matrix. The figure shows propagators across several receiver indices for a single hyperbolic reflection nmo time velocity pair (t_j, V_k) . For a single reflection all of the non-zero dictionary coefficients are contained in a single column of the weighting matrix.

t_j and NMO velocity V_k from an impulsive source,

$$\Gamma^{i,j,k} = \begin{cases} \delta(t_{NMO}(t_j, x_i, V_k)) & \text{if } i' = i \\ \vec{0} & \text{if } i' \neq i. \end{cases} \quad (4.5)$$

Each entry in Γ describes a single receiver response and is all zeros except for the i^{th} index for all receiver indices i' from 1 to N_x as illustrated in Figure 4-8. The reflection from an impulsive source $D_\delta(t_j, V_k)$ with NMO time t_j and NMO velocity V_k can be written as the weighted sum across offset index,

$$D_\delta(t_j, V_k) = \sum_{i=0}^{i=N_x} w_i \Gamma^{i,t_j,V_k} \quad (4.6)$$

where w_i is the amplitude of the reflection at each offset index. Using Equations 4.1 and 4.5 we build a dictionary of receiver responses,

$$\Phi = [\Gamma^{i,1,k}(:), \Gamma^{i,2,k}(:), \dots, \Gamma^{N_x,N_t,N_v}(:)], \quad (4.7)$$

indexed along three different dimension: NMO time, NMO velocity, and offset index. Following from Equation 4.6, the observed signal can be described as the weighted sum of the dictionary elements,

$$D = \Phi X(:) + N. \quad (4.8)$$

where X is a dictionary weight matrix with 3 dimensions, N is Gaussian noise, and $(:)$ is the MATLAB like vectorization operator that converts an N-dimensional array to a vector. From Equation 4.6 the signal for an impulsive source written in terms of Equation 4.8 will have only non-zero entries of w_i along a single column of X with a constant velocity and time index , as shown in Figure 4-8. Extending this compactness to a gather with n_D reflections, X can be written with n_D non-zero columns that each correspond to an NMO semblance peaks like those shown in Figure 4-6. To exploit this column-wise sparsity we pose the denoising and multiple removal problem as a

convex minimization problem [72],

$$\hat{X} = \min_{\hat{x}} |\Phi X(:, \cdot) - B|_2^2 + \lambda |X|_{1,2}. \quad (4.9)$$

In this regularized expression we minimize the misfit between the observed and predicted data within our dictionary and impose $\ell_{1,2}$ norm penalty along the receiver axis of this dictionary to encourage our inverted \hat{X} to be column-wise sparse and allows the amplitude of the signal to vary as a function of offset. This constraint can be thought of as penalizing the number of non-zero columns of \hat{X} . In Equation 4.9 λ is the regularization tuning parameter that controls the relative importance of minimizing mean square error versus maximizing the group sparsity of the solution. In this chapter we manually tuned λ for a single source gather and used the value for all experiments. However, λ could be chosen using semi-automatic methods such as the L-curve method [9] or automated Kolmogorov-Smirnov (KS) test based method [15].

This dictionary can be thought of as the transpose of a generalized radon transform across the offset and time dimensions of the data. When viewed as matrix, the dictionary is primarily zeros with very few non-zero entries. This makes the application of Φ and Φ^t to a vector computationally inexpensive. Variations of sparse radon domain parabolic, linear and hyperbolic have been used throughout geophysics to denoise, interpolate data, and suppress multiples, [45, 62, 47, 71, 67]. The Φ dictionary does not have any linear algebraic structure, such as the linear radon that can be efficiently computed in the Fourier domain, and it would be computationally challenging to store a QR or LU factorization as these factored matrices would be dense. Methods that rely on LU or partial factorizations to solve Equation 4.9, such as ADMM [24], are computational infeasible and we instead make use of the first order method FISTA [2], Algorithm 5, that only requires the application of Φ and Φ^t to a vector.

To remove the multiples from the data we solve Equation 4.9 for each common shot gather by applying Algorithm 5. Figure 4-9C shows the ℓ_2 norm of dictionary

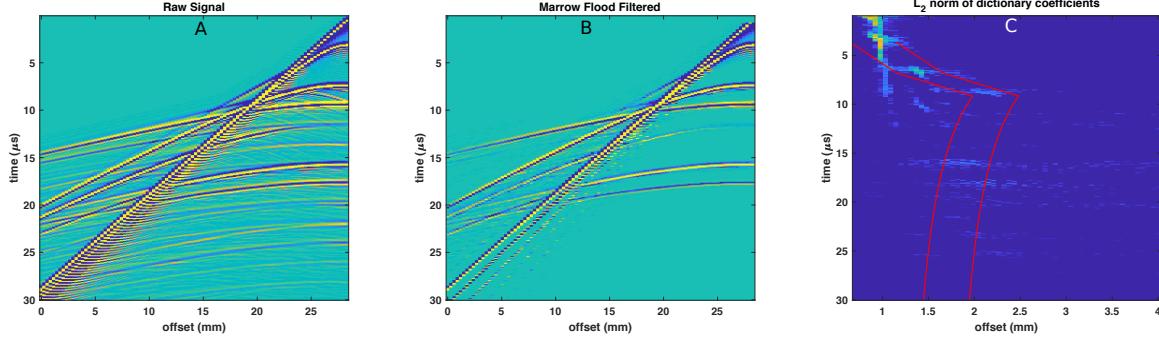


Figure 4-9: **A:** Raw synthetic trace with multiples present. **B:** denoised data with mute applied. **C:** ℓ_2 norm of dictionary coefficients along the receiver index direction. Red lines indicate the boundaries of the mute in the NMO time-velocity space.

coefficients summed across the receiver index of the estimated dictionary coefficients for a synthetic trace. From this figure we see that there are numerous peaks of energy some of which fall within the expected NMO time-velocity range (red dashed lines) and several peaks corresponding to multiples that fall outside the range. Then using the estimated velocities from our velocity analysis we generate a mute in the NMO time-velocity space and set the dictionary coefficients to zero if they are significantly faster or slower than expected and reconstruct the signal using Equation 4.8. The boundaries of the mute are outlined in red in Figure 4-9C. With the application of the sparse decisioning the multiples' energy is reduced significantly and the primary reflections are clearly visible in Figure 4-9AB. Once the mute has been applied the only peg-leg multiples remaining should occur after the bottom of the bone arrival with a period of d_{tissue} which can now be removed with predictive deconvolution.

Algorithm 5 FISTA: solves $\min_{\hat{X}} \|\Phi \hat{X} - B\|_2^2 + \lambda \mathbf{F}(X)$ (Eq. 4.9)

- 1: $\rho = \text{step size.}$
 - 2: $\hat{X} = \mathbf{0}$ //Initialize variables.
 - 3: $k = 1$
 - 4: **while** true **do**
 - 5: $\hat{X}_{old} = \hat{X}; k = k + 1;$
 - 6: $U = \hat{X} + \frac{k-1}{k+2}(\hat{X} - \hat{X}_{old})$ //Interpolation
 - 7: $Z = U - \frac{1}{\rho} \Phi^T (\Phi U - B)$ //Gradient calculation
 - 8: $\hat{X} = \min_{\hat{X}} \frac{1}{2} \|\hat{X} - Z\|_2^2 + \frac{\lambda}{\rho} \sum_{N_x}^{i=1} \|Z(:, i)\|_2$ //Apply shrinkage operator
 - 9: **end while**
-

Predictive Deconvolution

Before applying the predictive deconvolution we apply an NMO correction [86] in which the data is stretched according to the Equation 4.1 using the NMO velocities recovered from the velocity analysis. This aligns the data such that the primary reflections are flattened. This procedure can also be computed in the common midpoint domain which would allow for better suppression of multiples due to strongly sloped reflectors. Once the data are aligned, we apply a tapered mute to remove the primaries preceding the upper endosteum return and apply predictive deconvolution [52] with a period of d_{tissue} inverted from the velocity analysis to remove the remaining peg-leg multiples in the source domain. We apply the mute first because predictive deconvolution is only able to remove data at a single repetition interval, thus including all of the primary reflections would result in subtracting multiples that are not present, complicating the interpretation. We then apply the inverse NMO transform to return the data back to its original domain and stitch the data back together with the hyperbolic denoised data prior to the mute such that the inverted data contain all of the primaries. Figure 4-10 shows migration of the synthetic survey with and without the denoising and multiple removal with 45 dB of dynamic range. The figure is shown with envelope detection applied in the time domain where the displayed signal $y(t)$ is given by,

$$y(t) = 20 \log_{10} |(x(t) + jH(x(t)))| \quad (4.10)$$

where H is the Hilbert transform and $x(t)$ is the signal at each receiver.

If one were to apply predictive deconvolution to the data without first reducing the other multiples, the predictive deconvolution would erroneously subtract arrivals from the data and generate artificial reflectors. Figure 4-10C shows a migrated image with only predictive deconvolution applied. Although the method is able to suppress the peg-leg multiples as their moveout is similar to the expected lower endosteum return, the method generates additional artifacts and is unable to suppress the bone and tissue multiples. If however, we apply the sparse denoising then predictive deconvolution,

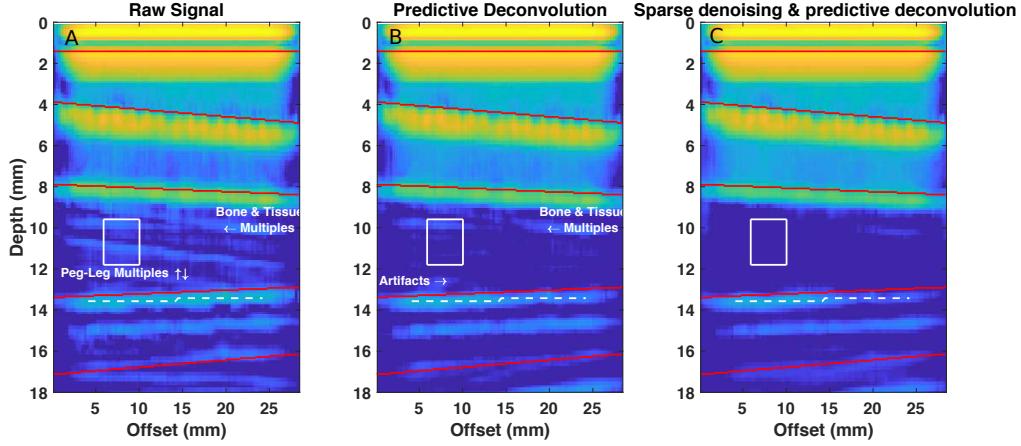


Figure 4-10: Migrated images with processed and unprocessed data. True interfaces are highlighted in red. The colormap is the same for all three panels with a dynamic range of 45dB. **A:** Migrated image using raw data. Numerous multiple returns are present that do not correspond to the true model. **B:** Kirchoff migrated image with only predictive deconvolution applied. Predictive deconvolution alone does not mitigate all of the multiples present in the data. **C:** Migrated image with sparse denoising and predictive deconvolution. All interfaces are estimated correctly with the exception of the lower endosteum due to using an incorrect velocity model with an infinite marrow half space shown in Figure 4-2B. The signal and clutter strengths given in Table 4.3 are calculated by averaging over the dashed white line and the pixels within the white box respectively.

both the peg-leg and bone and tissue internal multiples are significantly reduced in the final image as shown in Figure 4-10A. The lower endosteum signal was calculated by averaging over the white dashed line and the clutter region by averaging over the region indicated by white line, see Figure 4-10 & Table 4.3. Predictive deconvolution alone decreases the interface intensity by 3.7 dB and sparse denosing with predictive deconvolution decreases tissue intensity by 5.6 dB. Predictive deconvolution alone increased contrast moderately by 2.7 dB between clutter and the interface from 13.1 dB to 15.4 dB. However, the clutter was significantly reduced by the the sparse denoising and predictive deconvolution resulting in an interface clutter contrast of 23.8 dB, a 10.8 dB improvement over the raw image.

Interface	Raw	predictive	Sparse & Predictive
Clutter:	-36.6dB	-42.9dB	-53dB
Signal:	-23.5dB	-27.2dB	-29.1dB
Contrast:	13.1dB	15.7dB	23.8dB

Table 4.3: Synthetic signal and clutter intensity. The proposed denoising scheme improves contrast by roughly 10.8 dB.

Interface	Raw	Sparse	sparse & predictive
Clutter:	-27.1dB	-36.1dB	-38.3dB
Upper endosteum:	-6.9dB	-6.1dB	-8.3dB
Lower endosteum:	-25.8dB	-30dB	-32.2dB

Table 4.4: In vivo isotropic migration signal and clutter strength from Figure 4-11.

Interface	Raw	Sparse	Sparse & predictive
Clutter:	-24.5dB	-33.4dB	-35.6dB
Upper endosteum:	-10.1dB	-9.5dB	-11.2dB
Lower endosteum:	-15.3dB	-18dB	-19.6dB

Table 4.5: Invivo anisotropic migration signal and clutter strength from Figure 4-11.

Narrow-offset migration

In order to image the interior of the bone we use both the anisotropic migration described in [56] and a narrow-offset method using an isotropic forward model. Although the bone is anisotropic in the axial geometry it may be advantageous to use a simpler and potentially faster isotropic model for imaging. Because the anisotropic effects are most pronounced at far offsets if we include these components in the final image artifacts complicate the interpretation of the resulting image. In addition, estimating the thickness of the marrow channel doesn't require far offset data. If one instead limits the isotropic migration to only use data and update the image for a narrow range of offsets the image quality is significantly improved. We use the narrow offset data to perform Kirchoff migration with an isotropic eikonal solver on anisotropic data. See Figure 4-11 for a comparison of the isotropic and anisotropic migrations on in vivo data.

Interface	Raw	Sparse	sparse & predictive
Upper endosteum:	20dB	29.9dB	30dB
Lower endosteum:	1.3dB	6.1dB	6.1dB

Table 4.6: Invivo Isotropic contrast between clutter and endosteum signal from Figure4-11.

Interface	Raw	Sparse	sparse & predictive
Upper endosteum:	14.3dB	23.8dB	24.4dB
Lower endosteum:	9.2dB	15.4dB	16dB

Table 4.7: Invivo anisotropic contrast between clutter and endosteum signal from Figure4-11.

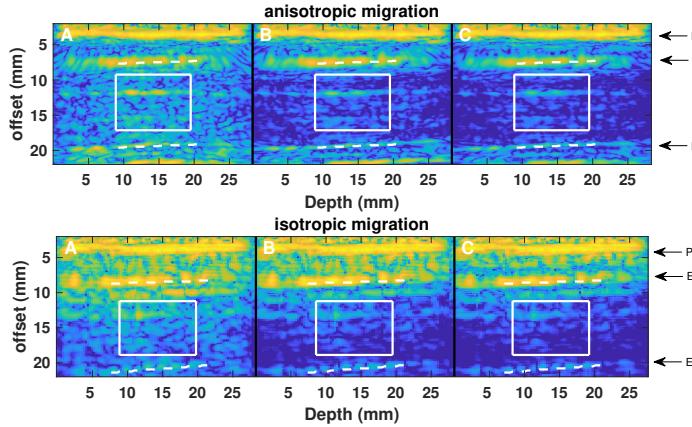


Figure 4-11: In vivo migrated image of the radius. Upper images are generated with anisotropic migration. Lower images are generated with isotropic migration. (A) Migration using raw data. (B) Migration using the sparse dictionary after mute has been applied. (C) migration with hyperbolic sparse denoising and predictive deconvolution data. The Periosteum and Endosteum are labeled ‘P’ and ‘E.’ The signal and clutter intensities are reported in Tables 4.4 & 4.5 by averaging the pixel intensity over the dashed white lines (endosteum) and clutter (white box), respectively. The contrast between clutter and signal strength are given in Tables 4.6 & 4.7.

4.7 Axial: in vivo

In this section we demonstrate the results of our denoising and multiple removal procedure on several in vivo datasets. To denoise the in vivo datasets we first apply the semblance velocity analysis previously described to determine the approximate stacking velocities, and arrival times corresponding to the upper bone interfaces. We then denoise the data by solving Equation 4.9 using the FISTA algorithm for 100 iterations with a λ of 250 for each common source gather and mute the dictionary coefficients that have their NMO velocity outside of $+/- .25 \text{ mm}/\mu\text{s}$ of the expected returns. Next we apply predictive deconvolution with the delays estimated from the NMO semblance procedure to remove the remaining peg-leg multiples. Figure 4-12 shows a single source gather with and without multiple removal.

We then migrate the data using both the narrow-offset method and the anisotropic migration scheme. For the narrow-offset method, we use only 5 mm of offset to perform Kirchoff migration for the raw and processed data; we use data from all

offsets for the anisotropic migration. All migrations are shown with log envelope detected with 45 dB of dynamic range. Figure 4-11 upper shows the anisotropic migration of the denoised dataset shown in Figure 4-12. The denoising procedure significantly decreases the speckle in the migrated image and increases the contrast of all of the interfaces. In the denoised migrated image Figure 4-11C the lower endosteum interface is clearly visible at a depth of about 20mm. Although, the effect of the predictive deconvolution is minimal in the data domain, the migrated image with predictive deconvolution significantly reduces the internal multiple visible at a depth of 11 mm, making it much easier to correctly determine the lower endosteum interface. Figure 4-11 lower shows the same data set migrated using the narrow-offset migration. Despite using a much simpler isotropic forward model, we are still able to image the lower endosteum visible at an approximate depth of 20mm. For the majority of images generated with the raw unprocessed data the return for the lower marrow interface is difficult to distinguish.

Like the synthetic example, the denoising procedure decreases the intensity of the endosteum interfaces but significantly reduces the clutter strength, see Tables 4.4 & 4.5. However, contrast between the upper endosteum is improved by the denoising process by 10dB for the isotropic migration and anisotropic migrations, see Tables 4.6 & 4.7. The lower endosteum contrast is improved for both isotropic migration by 4.8dB and 6.8dB for the anisotropic migration. Although the isotropic migration gives yields better contrast between the upper endosteum and clutter (20dB vs 14.3 dB), the anisotropic migration better resolves the lower endosteum (1.3dB vs 9.2 dB) and yields a larger contrast improvement when combined with sparse hyperbolic denoising and predictive deconvolution. The combination of both the anisotropic migration and denoising process results in an almost 15dB improvement in contrast between the lower endosteum and clutter compared to isotropic migration without denoising. The sparse hyperbolic radon and predictive deconvolution significantly improves recovery of the marrow return as well as the bottom of the bone.

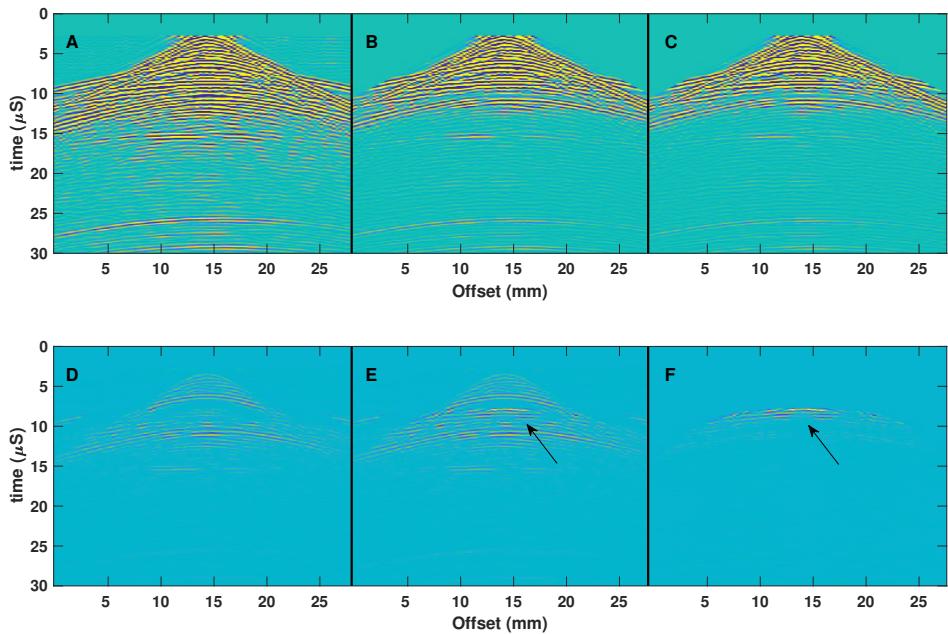


Figure 4-12: In vivo common source gather of the radius for a single source (A) raw data. (B) the sparse dictionary after mute is applied. (C) hyperbolic sparse denoising and predictive deconvolution. (D) The difference between A and B. E the difference between A and C. (F) The difference between D and E. The predictive deconvolution suppresses a peg leg multiple indicated by the arrows.

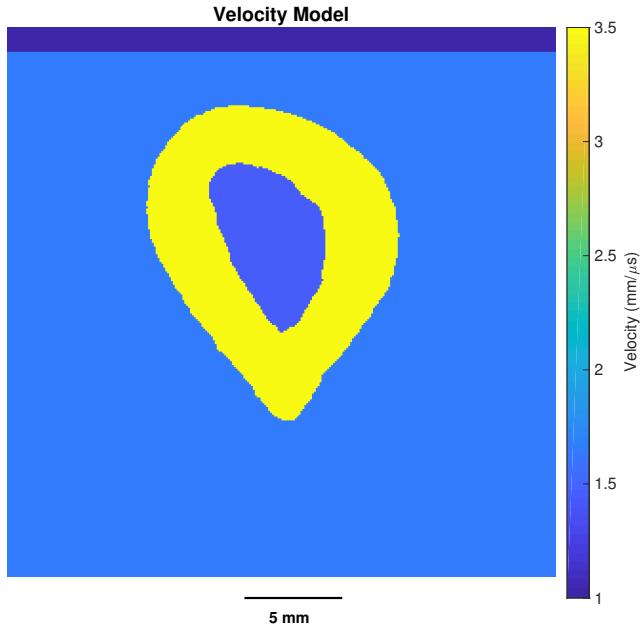


Figure 4-13: True velocity model for transverse synthetic experiments.

4.8 Transverse: Methods & Synthetic

In the transverse geometry the velocity model is significantly more complex than in the axial acquisition described in the previous section, making it ill suited for estimating bone and tissue velocities. Due to the orientation of the transverse isotropy (See Figure 4-1), one can ignore the anisotropic effects and use an isotropic assumption for imaging. This geometry provides a much better orientation for estimating cortical shell thickness and potentially its entire profile instead of an approximate diameter measured in the axial plane. However, the high contrast and complex shape of the bone makes imaging the interior extremely challenging as the high velocity of the bone causes waves to scatter at sharp angles, leading to little recorded energy at the surface. In addition, the complexity and contrast of the model could be challenging for ray based migration as significant amounts of energy could return along multiple paths not accounted for in Kirchoff migration using a shortest ray assumption.

To illustrate the challenges presented by this geometry, we used the cross section

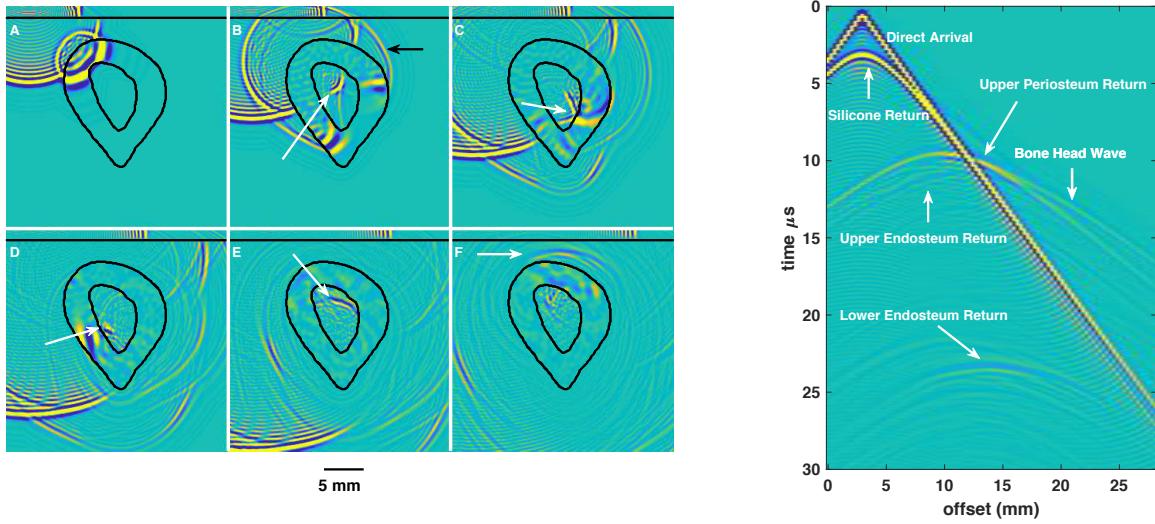


Figure 4-14: **Left:** snapshots of wave field propagation in transverse geometry. The white arrow indicates the return from the bottom of the marrow and the black arrow highlights the headwave. **Right:** common source gather for the simulated gather with the various arrivals and reflected signals labeled. See supplemental video synth-Movie.avi

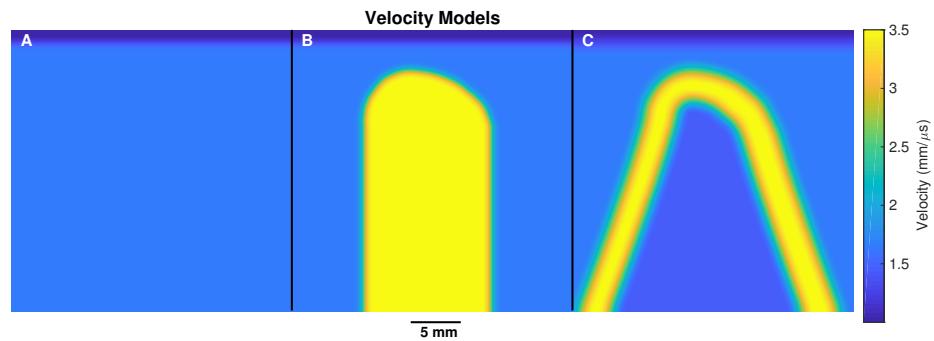


Figure 4-15: Transverse velocity models used for migrations. From left to right: tissue half space, bone half space, horseshoe.

of an in vivo radius generated from CT imagery [56] to produce a synthetic velocity model shown in Figure 4-13 in which the transverse bone velocity was assumed to be $3.250 \text{ mm}/\mu\text{s}$. We generate 96 finite difference runs using this velocity model to simulate the source/receiver geometry of the experiment. Figure 4-14 Right shows the recorded shot gather and Figure 4-14 Left shows several snapshots of the simulated wave-field propagating through the velocity model and demonstrates the significant complexities of the wave propagation in this geometry. Much like the axial case head and guided waves are present within the bone (black arrow in Figure 4-14 Left). However, the travel time of this wave as a function of offset is more complex as it is propagating along a curved surface and thus does not follow a straight line in the data domain as we see in Figure 4-14C. The head wave arrival can be seen in the shot gather, where there appears to be a split in the top of the bone return at long offsets. From the snapshots we also observe that the guided wave within the bone reaches the backside of the bone well before the direct arrival traveling through the slow marrow, with the guided wave leaking off energy into the tissue. A supplementary movie is included with the manuscript to further illustrate the complexities wave propagation in this geometry.

The direct arrival once within the marrow (white arrow in Figure 4-14 Left) has a complex path and reflects off of both sides of the interior of the bone before being reflected to the surface and becomes visible in the shot gather at roughly $23 \mu\text{s}$. Despite the complex wave propagation and presence of guided waves, the amplitude of the lower endosteum reflection remains the most energetic arrival in the later section of recorded source gathers. Notably in the in vivo data the speckle SNR is significantly lower than in the synthetic models and the return from the lower endosteum is not easily visible.

The fact that the majority of the energy from the marrow reflection is multiply scattered significantly complicates the imaging and velocity model estimation as imaging the left side of the endosteum is dependent on the position of the unknown right side of the endosteum and vice versa. Furthermore, if one were to invert for the bottom of the bone using a single scattering assumption then the interior of the

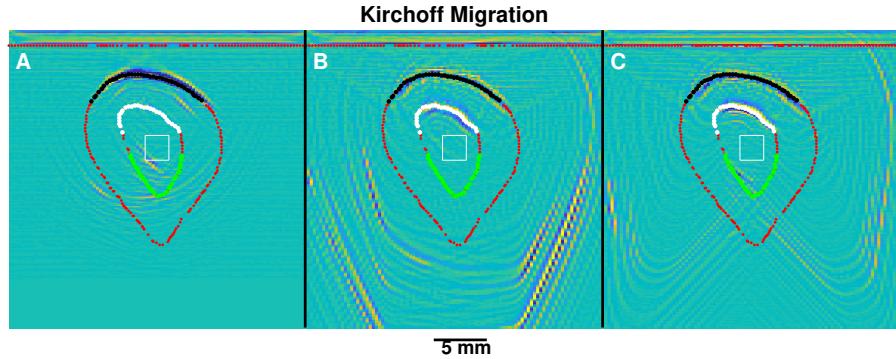


Figure 4-16: Kirchoff migration results using the tissue half space (A), bone half space (B), and horseshoe models (C). To highlight potential reflectors below the upper endosteum the scale has been saturated for the horseshoe migration. The signal intensity for the upper periosteum, upper endosteum, and lower endosteum were calculated by averaging along the black, white, and green lines. The clutter strength was calculated by averaging over the pixels within the white box. The contrast between the clutter and pixel strengths are given in Table 4.8.

Interface	Tissue	Bone	Horseshoe
Upper periosteum:	28.9dB	20.3dB	21.3dB
Upper endosteum:	1.2dB	16.6dB	22.2dB
Lower endosteum:	7.1dB	-5.8dB	12.3dB

Table 4.8: Contrast between marrow clutter and interfaces for synthetic Kirchhoff migration examples using shown in Figure 4-16.

Interface	Tissue	Bone	Horseshoe
Upper periosteum:	40.1dB	29dB	36.9dB
Upper endosteum:	8.6dB	25.35dB	29.8dB
Lower endosteum:	10.1dB	-8.5dB	9.6dB

Table 4.9: Contrast between marrow clutter and interfaces for synthetic reverse time migration examples using shown in Figure 4-17.

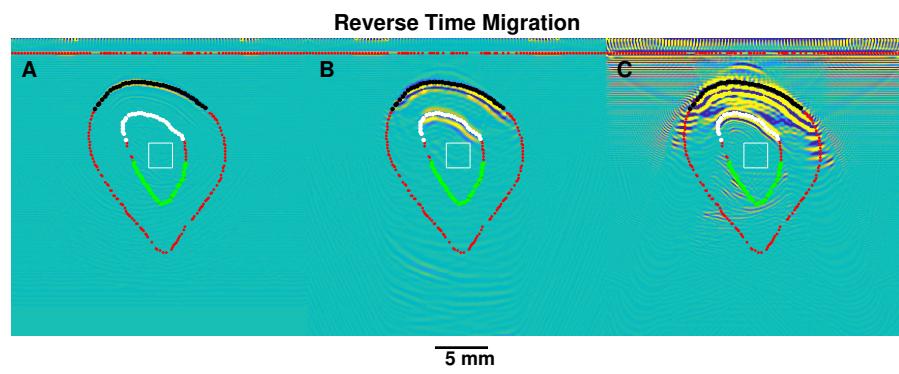


Figure 4-17: RTM results using the tissue half space (A), bone half space (B), and horseshoe models (C). To highlight potential reflectors below the upper endosteum the scale has been saturated for the horseshoe migration. The signal intensity for the upper periosteum, upper endosteum, and lower endosteum are calculated by averaging along the black, white, and green lines. The clutter strength is calculated by averaging over the pixels within the white box. The contrast between the clutter and pixel strengths are given in Table 4.9.

bone would be significantly distorted and appear deeper than it actually is because of the longer time taken for the multiply scattered waves to reach the receiver. The simulation shown in Figure 4-14 also implies that since the bottom of the bone is not parallel to the top of the bone that in the axial geometry the marrow return could be reflected off axis and may not even appear in the recorded data.

In addition to dealing with the complexities mentioned above, for in vivo data in which the background velocity model is unknown, one would need to build a velocity model in conjunction with imaging. Following standard seismic practice, to build a velocity model we start with a top down approach in which we apply reverse time migration to increasingly complex velocity models built from a previous migration. In the following section we follow this same procedure to image using in vivo data. This process illustrates which sections of the image are realistically resolvable.

To determine the location of the upper bone we generate both RTM and Kirchoff migrated images using a velocity model consisting of a thin layer of silicone rubber and an infinite half space of tissue as shown in Figure 4-15A. We then apply a Laplacian edge enhancement filter to the migrated images to reduce low frequency artifacts. Using this velocity model we are able to clearly resolve only the top bone as energy reflected off the sides of the bone is scattered downwards and does not return to the transducer. In addition, the return from the backside of the bone is also imaged yet its thickness is severely distorted because we are using too slow a velocity model.

Once we locate the uppermost section of bone we generate a bone half space velocity model in which we consider the velocity model below the top of the bone interface to entirely consist of bone. We apply RTM again using the velocity model shown in Figure 4-15B and obtain the resulting image in Figure 4-17B. The bottom of the uppermost section of the bone return is now visible in the image. Even with this more accurate velocity model the sides of the bone are not recoverable. For both the tissue and bone half space images reverse time migration and Kirchhoff migration obtain good estimates of the top and bottom of the bone, Figures 4-16B & 4-17B. RTM also yields a higher contrast between clutter and the upper periosteum & endosteum as shown in Tables 4.8 & 4.9 for all three velocity models. Using a half

space velocity model to locate the bottom interface is commonly used to image the bottoms of salt bodies in seismic imaging [43].

Now that we have an image of the top periosteum and endosteum we generate a horse shoe velocity model in which we downward continue the sides of the bone as shown in Figure 4-15C. With this velocity model we see that we are able to image some of the reflections off of the interior of the marrow using RTM. The Kirchhoff migration Figure 4-16C performs significantly worse than RTM, Figure 4-17C and the Kirchhoff migrated images contain only a small amount of energy from the marrow return. Although RTM is able to recover more of the marrow return, the position of this reflector is distorted and several artifacts are present. Due to this distortion, the lower endosteum contrast appears to be lower than that of the Kirchoff migration, Tables 4.8 & 4.9, although more of the interface is visible in the RTM image. Because our velocity model does not incorporate accurate sides of the bone, the migration is attempting to image the interface as if this return was from single scatterer and does not accurately consider reflections off the sides of the interior of the bone. Although this method does not perfectly recover the bottom of the bone interface, it does illustrate that a rough approximation of marrow thickness can be obtained and that one could use an RTM migration to image the interior of the marrow.

From Figure 4-16C we see that the Kirchoff migration is able to recover the uppermost section of the bone and only a small section of interior marrow/bone interface and fails to image the exterior sides of the bone. The RTM image however recovers a much larger extent of the bottom of the bone, Figure 4-17. The reason for this is that in the presence of strong velocity contrasts and multiple travel paths there are many arrivals at different times, and Kirchoff migration only accounts for the first arrival. These artifacts are not present in the reverse time migration as the imaging techniques accounts for all arrivals. Due to the acquisition geometry and the high velocity of the bone, waves steeply refract downward when they encounter the sides of the bone and very little energy is directed upwards toward the transducers making it impossible to resolve the sides of the bone. The improved image achieved by reverse time migration suggests that the single ray approximation used in the Kirchoff

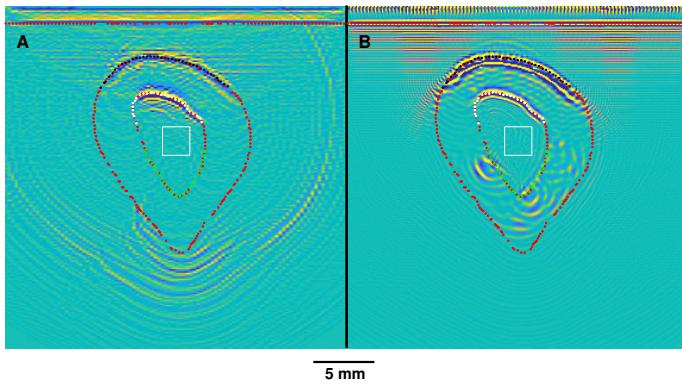


Figure 4-18: Transverse Kirchoff (A) migration and reverse time migration (B) using the smoothed true velocity model. The signal intensity for the upper periosteum, upper endosteum, and lower endosteum are calculated by averaging along the black, white, and green lines. The clutter strength is calculated by averaging over the pixels within the white box. The contrast between the clutter and pixel strengths are given in Table 4.10.

migration is unable to capture the multiply scattered energy.

To test the limits of this imaging geometry, we apply both Kirchoff migration and reverse time migration with a smoothed version of the true velocity model shown in Figure 4-13 and give the results in Figure 4-18. Although, it is unrealistic that one would have a full smoothed velocity model before knowing the measurements of the interior of the marrow, using the true smoothed velocity model illustrates the best case scenario of what could be recovered with a particular imaging modality. With the true velocity model, RTM results in a significantly improved contrast compared to Kirchoff migration for all three interfaces as shown in Table 4.10.

Even in the best case scenario we are only able to image a limited section of the interior of the bone with RTM and the Kirchhoff migration failed to image the interior of the marrow. When a realistic horse shoe velocity model is used the image of the interior of the bone the primary reflections from the bottom of the marrow are extremely weak and the resulting image is corrupted by multiples and limited aperture artifacts. These results suggest that it will be challenging to image the interior of the bone without using the additional energy from the multiples.

Interface	Kirchoff	RTM
Upper periosteum:	19.9dB	28.7dB
Upper endosteum:	22.2dB	25.1dB
Lower endosteum:	-3.3dB	14.27dB

Table 4.10: Contrast between marrow clutter and interfaces for the synthetic migrations shown in Figure 4-18.

Interface	Tissue	Bone	Horseshoe
Upper periosteum:	24dB	8dB	18dB
Upper endosteum:	3.2dB	12.5dB	15.6dB

Table 4.11: Contrast between marrow clutter and interfaces for the in vivo Kirchhoff migrations shown in Figure 4-20.

Interface	Tissue	Bone	Horseshoe
Upper periosteum:	33.2dB	21.7dB	24dB
Upper endosteum:	6.5dB	24.7dB	15.3dB

Table 4.12: Contrast between marrow clutter and interfaces for the in vivo reverse time migrations shown in Figure 4-21.

4.9 Transverse: in vivo

Following from the synthetic results shown in the previous section, we apply a top down approach in which we iteratively build a velocity model for in vivo data. We first apply Kirchhoff and reverse time migration to the in vivo data set using the transverse tissue flood model shown in Figure 4-19 left. The 10 left and right most common source gathers contain significant amounts of out of plane artifacts and were discarded for the migration. For the tissue flood model both Kirchhoff and reverse time migration are able to recover the top and bottom of the upper section of bone and as with the synthetics the sides of the bone are not recovered. In addition, RTM generally improved the contrast between the interfaces and clutter compared to Kirchhoff migration as shown in Tables 4.11 & 4.12. For the bone flood velocity model shown in Figure 4-19B RTM resulted in a 12dB gain in contrast over Kirchhoff migration. Unlike in the synthetic data, in the in vivo data for the transverse geometry, we are unable to see a return in the raw unprocessed data and the signal to speckle noise ratio may be too low to recover the backside of the marrow.

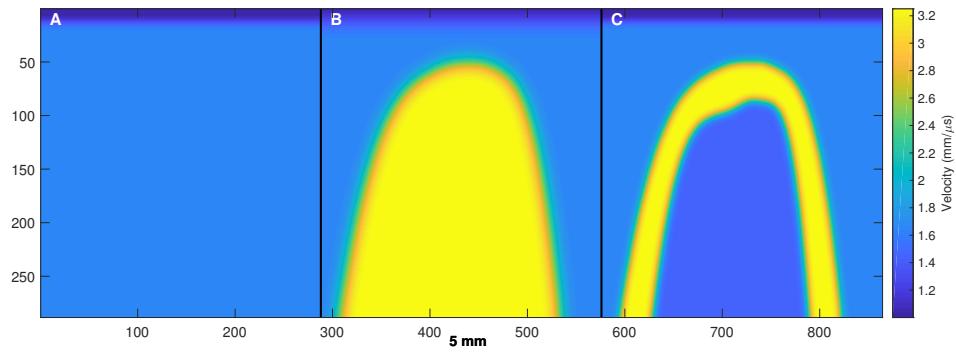


Figure 4-19: Velocity models for the in vivo data in the transverse configuration. **A:** tissue flood. **B:** bone flood. **C:** marrow flood.

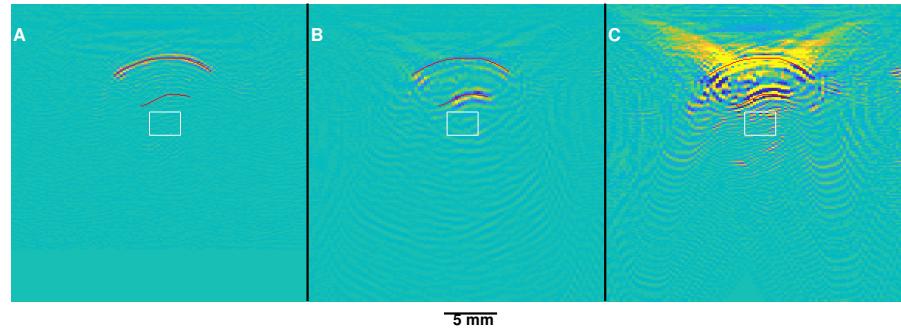


Figure 4-20: Transverse Kirchhoff migrations using the velocity models in Figure 4-19. **A:** tissue flood. **B:** bone flood. **C:** marrow flood. The signal intensity for the upper periosteum and endosteum were calculated by averaging along the red lines. The clutter strength was calculated by averaging over the pixels within the white box. The contrast between the clutter and pixel strengths are given in Table 4.11.

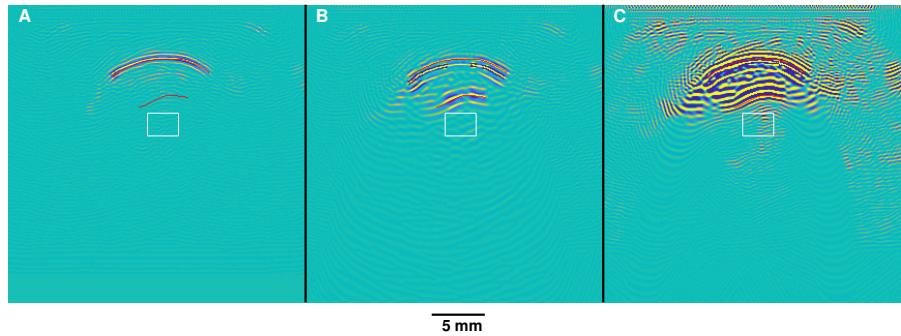


Figure 4-21: Transverse reverse time migrations using the velocity models in Figure 4-19. **A:** tissue flood. **B:** bone flood. **C:** marrow flood. Some structure may be visible in the horseshoe model that corresponds to the interior of marrow. The signal intensity for the upper periosteum and endosteum are calculated by averaging along the red lines. The clutter strength is calculated by averaging over the pixels within the white box. The contrast between the clutter and pixel strengths are given in Table 4.12.

4.10 Discussion

Our results suggest that although internal multiples make it challenging to image the interior of bone, imaging in the axial geometry poses significantly fewer challenges than the transverse geometry. These multiples can be well suppressed in *in vivo* data with the combination of sparse hyperbolic radon transform and predictive deconvolution. Due to the complexity of imaging in the transverse geometry, it might be advantageous to use a patch based ultrasound array and reconstruct the bone cross section through generating numerous axial slices.

Imaging in the interior of the marrow in the transverse geometry proves to be challenging and our ability to image the back side endosteum is inconclusive. In both the axial and transverse geometries we neglect to model elastic effects, specifically the presence of mode-converted shear waves. In the axial acquisition we observe mode conversion between compressional and shear waves within the bone that could be corrupting our images. These are present in the common midpoint gathers as reflections after the upper periosteum and may degrade the image quality between the

upper periosteum and endosteum. Exploiting information in the shear waves could provide better diagnostics of bone structure and health as part of a more careful velocity analysis. The presence of shear waves indicates a possible other avenue for estimating bone parameters. In this paper, we use a fairly primitive method of estimating bone velocities and [56] proposes a migration based method for inverting velocities and anisotropic parameters. To extract more information from the data we could use the anisotropic velocity analysis techniques proposed in [73] that use mode-converted shear waves within the bone to estimate additional material parameters. Accounting for the shear waves could also significantly improve our ability to image in the transverse geometry as the shear wave velocity is significantly lower than the compressional velocity and energy in the shear mode would refract at shallower angles allowing for more of the bottom of the bone to be illuminated.

The transverse geometry remains a challenging acquisition scheme due to the complex velocity model and lack of primary reflections. For in vivo datasets both RTM and Kirchoff fail to image the interior of bone and even the synthetic dataset proves difficult to image. To improve the imaging performance one could use Marchenko methods [80] to move virtual source & receivers to the underside of bone. This would remove the dependence on the complex velocity model and allow one to image the bottom of the bone using a simple velocity model similiar to Figure 4-15C.

4.11 Conclusion

In this paper we proposed several methods for imaging the interior of bone in two different acquistion geometries and used finite difference modeling to be better understand the challenges presented in each imaging modality. We demonstrated that in the axial geometry it is possible to image the interior of bone if multiples are significantly suppressed in the raw data and showed how to perform this suppression. Imaging in the transverse geometry remains challenging due to lack of received energy and complex velocity model.

4.12 Acknowledgments

Guillaume Renaud acknowledges the support from the French ANR program (17-CE19-0008-01) and Alison Malcolm acknowledges the support from the NSERC Discovery Grant program.

4.13 Appendix: Seismic techniques

In this section we give a brief overview of common seismic techniques and terminology used throughout this paper.

4.13.1 Normal move-out and semblance

In exploration seismology a velocity model is often approximated by a set of, N horizontal layers leading to the normal moveout (NMO) approximation [10]. Under this approximation the arrival times at each detector are hyperbolic as a function of offset and are given by Equation 4.1 where the NMO velocity, V_{NMO} is,

$$V_{(i)NMO} = \sqrt{\frac{\sum_{i=1}^{i=N} V_{(i)int}^2(t_i - t_{i-1})}{\sum_{i=1}^{i=N} t_i - t_{i-1}}} \quad (4.11)$$

where V_{int} is the true velocity of each layer and t_i is the zero offset travel time of the i_{th} interface or reflector. If a series of NMO velocities are observed one can invert for the true velocity of each layer using the Dix equation [13],

$$V_{(i)int} = \sqrt{\frac{t_i V_{(i)NMO}^2 - t_{i-1} V_{(i-1)NMO}^2}{t_i - t_{i-1}}}. \quad (4.12)$$

To estimate various parameters such as NMO velocity or head wave velocities seismologists measure coherent energy across multiple receivers [86]. These measures of coherence are typically made along theoretical travel time curves calculated using the NMO equation, Equation 4.1, for estimating vertical velocities, or Equation 4.4 for estimating horizontal and head-wave velocities. Methods based on coherence mea-

sures are typically more robust and can operate at higher noise levels than methods that require picking arrival times and then finding the best fit travel time curve. Throughout this paper make use of an unnormalized cross correlation,

$$s = \left| \sum_{t=t_0-\frac{l}{2}}^{t=t_0+\frac{l}{2}} \left(\sum_{i=0}^{N_x} d(t, x_i)^2 \right) - \sum_{t=t_0-\frac{l}{2}}^{t=t_0+\frac{l}{2}} \left(\sum_{i=0}^{N_x} d(t, x_i) \right)^2 \right| \quad (4.13)$$

to estimate coherent energy along a travel time curve. Where $d(t, x_i)$ is the observed data at the theoretical travel time t at receiver x_i and l is the time coherence window length. Often a taper is applied to the time axis of Equation 4.13 to reduce the edge effects caused by the application of a hard edged filter.

4.13.2 Kirchoff Migration

In order to transform raw seismic or ultrasound data, seismologists use a variety of imaging methods such as Kirchoff and reverse time migration that both rely on a background velocity model. Kirchoff migration is based on summing across predicted travel times and typically uses an Eikonal solver [28] to estimate these travel times and map raw data to the image domain, or methods based on Snell's law used in [56]. Because Kirchoff migration typically relies on ray tracers rather than finite difference forward modeling it is typically very fast to generate an image. The data are then summed along these modeled travel times to generate an image. Because Kirchoff migration assumes a single ray path for a single point in image space, the algorithm cannot model multiply scattered waves or even the possibility that energy may have traveled along more than one path between two points. This is a common phenomenon generated in complex velocity models. In addition, Kirchoff migration also breaks down when the velocity model contains large velocity contrasts as the method does not capture critically refracted or guided waves.

In this paper we used a modified version of the code published in [56] to generate anisotropic migrations and reconstruct the image until the lower endosteum. In [56], the code reconstructs the image only until the upper endomsteum and in this paper

we present the first images of the lower endosteum using ultrasound.

4.13.3 Reverse Time Migration

A more advanced imaging algorithm commonly used in exploration seismology is reverse time migration [1, 87]. Reverse time migration uses wave equation forward models such as finite difference solvers, to more accurately model the physics of wave propagation. In reverse time migration the recorded seismic data is back propagated with a wave equation solver and convolved with the simulated wave-field to generate an image. Because reverse time migration typically involves several finite difference simulations it is significantly slower to generate an image than Kirchoff migration. However, the method can image with multiply scattered waves and generate superior images in complex velocity models and is commonly used to image beneath salt in exploration seismology, a situation that is quite analgous to imaging the marrow beneath the bone.

Chapter 5

Conclusion

This thesis illustrates the power of simpler faster models neglected in seismic imaging. The use of reduced parameterizations combined with conventional methods for global optimization and UQ allows us to answer a number of questions about how well suited these classes of algorithms are for answering seismic imaging questions and their limitations. Our framework doesn't allow for the use of the high fidelity models frequently used in seismic imaging as we are restricted to layered velocity model. However, this parameterization is extremely powerful and allows us to explore questions that would be infeasible with high fidelity models. Despite the limitations of the forward model, it is possible to improve the uncertainty estimates by accounting for the error introduced due to a lower fidelity model. Several UQ frameworks try to estimate and account for this type of impact in UQ estimates [84].

5.1 Future Work

There are numerous algorithmic improvements that could be made to both the uncertainty quantification and global optimization algorithms presented in this thesis. However, there has not been a forward solver that accurately captures the physics of seismic imaging and is fast enough to allow for the comparison of different algorithms. As a result, for numerous seismic imaging and FWI problems there are very few standard test problems. Even if we made algorithmic improvements such as

faster methods of MCMC, due to the lack of fast forward solvers we could not benchmark them against each other and demonstrate that one algorithm is better than another. This thesis moves towards establishing a baseline and a set of test problems for seismic imaging.

5.1.1 Testing assumptions and validity of assumptions

The UQ framework allows for testing a number assumptions made in seismic imaging. Although a Gaussian velocity model distribution is often assumed for the purposes of UQ, this assumption is frequently not valid as shown in chapter 3. However, this and other assumptions may be valid, especially in low noise situations in which there should be only one minima. Other local approximations that could allow some non-linearity (i.e kernel methods) may also be valid. Our approach now allows for these and other assumptions to be rigorously tested and examine how and when they break down. One could even take the results shown in chapter 3 and empirically fit distributions to them to see how well various approximations could work. Once these assumptions are validated on a model with fewer degrees of freedom, they could be extended to high dimensional models that more accurately model the subsurface.

5.1.2 Establishing test problems & reproducibility

In chapter 2 we established that using the field expansion solver and a reduced parameterization is a suitable test problem for benchmarking global optimization algorithms for FWI. A next step would be to conduct a more rigorous comparison across a wider range of global optimization methods and algorithms.

FWI fidelity is frequently tested with just a single velocity model that is often just a severely smoothed version of the true velocity model. However this may be an extremely unrealistic starting model and, as shown in chapter 2, similar starting models can lead to drastically different qualities of inversion for FWI. From chapter 3 we see that in the presence of noise a wide range of velocity models are possible given a single observed data sets. It would be far more realistic to test FWI from

a population of starting models and then compare the performance across realistic initial models. The framework in chapter 2 could be used to generate a distribution of realistic starting models for the test problem. Then either a random drawing of a handful of samples or using a clustering algorithm to find a collection of models that best represent the inversion could be used to initiate conventional FWI.

One issue with comparing the results of FWI in a quantitative manner is that the traditional ℓ_2 norm does not accurately capture the misfit between inverted velocity models. The incorrect velocity model can cause structures to move up and down causing the difference between models to be exaggerated beyond what is visible to the eye and what may matter for seismic imaging or later processing. Much like seismic imaging, medical ultrasound relies on subjective imaging quality of experts to determine the quality or accuracy of an image. However, unlike seismic ultrasound establishes a number of test problems for contrast and resolution, despite its subjective nature. In the future we need to better define quality metrics and misfits between inverted velocity models. Some metrics exist in computer vision that could be applied to seismic to add quantitative measures of how good one inversion method is versus another [31, 68, 58, 92].

Once these test problems have been established there remains a major challenge in reproducing these and other methods. Very few seismic imaging source codes are open source and thus impossible to reproduce. In addition, generating results tends to involve huge computational resources and computing clusters that many other researchers may not have access to. Even if the source code is available, setting up a working environment can be huge barrier to entry for installing these codes as there are numerous poorly documented dependencies and little incentive to maintain or support code once its released. Seismic and geophysical codes tend to be complex and involve large datasets, making them difficult to understand and less portable. One possible solution I've been investigate is using lightweight virtual machines images (Docker) to wrap seismic inversion or other geophysics along with all of its dependencies into an image [5]. Once someone has created an image the code contained in it can be run on almost any machine without having to install additional libraries and exists in its own

environment such that updates to the hosts machine software doesn't affect the image environment. In addition, virtually all cloud computing environments (Amazon Web Services, Google Cloud, Microsoft Azure) support the use of containers such that it is fairly straight forward to scale up the results and run large experiments on publicly available clusters.

Bibliography

- [1] Edip Bayosal, Dan D Kosloff, and John WC Sherwood. Reverse time migration. *Geophysics*, 48(11):1514–1524, 1983.
- [2] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.
- [3] Valeria Beltrame, Roberto Stramare, Nicola Rebellato, Federico Angelini, Anna Chiara Frigo, and Leopoldo Rubaltelli. Sonographic evaluation of bone fractures: a reliable alternative in clinical practice? *Clinical imaging*, 36(3):203–208, 2012.
- [4] Simon Bernard, Joannes Schneider, Peter Varga, Pascal Laugier, Kay Raum, and Quentin Grimal. Elasticity–density and viscoelasticity–density relationships at the tibia mid-diaphysis assessed from resonant ultrasound spectroscopy measurements. *Biomechanics and modeling in mechanobiology*, 15(1):97–109, 2016.
- [5] Carl Boettiger. An introduction to docker for reproducible research. *ACM SIGOPS Operating Systems Review*, 49(1):71–79, 2015.
- [6] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo*. CRC Press, May 2011. Google-Books-ID: qfRsAIKZ4rIC.
- [7] Jacobus Buur and Thomas Kühnel. Salt interpretation enabled by reverse-time migration. *Geophysics*, 73(5):VE211–VE216, 2008.
- [8] Camila C. S. Caiado, Richard W. Hobbs, and Michael Goldstein. Bayesian Strategies to Assess Uncertainty in Velocity Models. *Bayesian Analysis*, 7(1):211–234, March 2012.
- [9] D Calvetti, S Morigi, L Reichel, and F Sgallari. Tikhonov regularization and the l-curve for large discrete ill-posed problems. *Journal of computational and applied mathematics*, 123(1-2):423–446, 2000.
- [10] Richard J Castle. A theory of normal moveout. *Geophysics*, 59(6):983–999, 1994.
- [11] Debanjan Datta, Mrinal Sen, Faqi Liu, and Scott Morton. Salt model building by shape-based parameterization and global fwi. In *SEG Technical Program*

Expanded Abstracts 2016, pages 1069–1073. Society of Exploration Geophysicists, 2016.

- [12] Debanjan Datta and Mrinal K Sen. Estimating a starting model for full-waveform inversion using a global optimization method. *Geophysics*, 81(4):R211–R223, 2016.
- [13] C Hewitt Dix. Seismic velocities from surface measurements. *Geophysics*, 20(1):68–86, 1955.
- [14] Russ C Eberhart and James Kennedy. A new optimizer using particle swarm theory. In *Proceedings of the sixth international symposium on micro machine and human science*, volume 1, pages 39–43. New York, NY, 1995.
- [15] Gregory Ely, Shuchin Aeron, and Eric L Miller. Exploiting structural complexity for robust and rapid hyperspectral imaging. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 2193–2197. IEEE, 2013.
- [16] Gregory Ely, Alison Malcolm, and David Nicholls. Combining Global Optimization and Boundary Integral Methods to Robustly Estimate Subsurface Velocity Models. Society of Exploration Geophysicists, December 2015.
- [17] Gregory Ely, Alison Malcolm, and Oleg V Poliannikov. Assessing uncertainties in velocity models and images with a fast nonlinear uncertainty quantification method. *Geophysics*, 83(2):R63–R75, 2018.
- [18] Gregory Ely, Alison Malcolm, and Guillaume Renaud. Suppression of internal multiples with a group sparse radon transform for imaging the interior of bone. In *2018 SEG International Exposition and Annual Meeting*. Society of Exploration Geophysicists, 2018.
- [19] Zhilong Fang, Curt Da Silva, Rachel Kuske, and Felix J Herrmann. Uncertainty quantification for inverse problems with weak pde-constraints. *Geophysics*, 83(6):1–107, 2018.
- [20] L Neil Frazer and Mrinal K Sen. Kirchhoff-; helmholtz reflection seismograms in a laterally inhomogeneous multi-layered elastic mediumâ†. theory. *Geophysical Journal International*, 80(1):121–147, 1985.
- [21] Bruno Galuzzi, Elena Zampieri, and Eusebio Stucchi. Global optimization procedure to estimate a starting velocity model for local full waveform inversion. In *International Conference on Optimization and Decision Science*, pages 171–179. Springer, 2017.
- [22] Andrew Gelman, Gareth O Roberts, Walter R Gilks, et al. Efficient metropolis jumping rules. *Bayesian statistics*, 5(599-608):42, 1996.

- [23] Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992.
- [24] Tom Goldstein, Brendan O’Donoghue, Simon Setzer, and Richard Baraniuk. Fast alternating direction optimization methods. *SIAM Journal on Imaging Sciences*, 7(3):1588–1623, 2014.
- [25] Wences P Gouveia and John A Scales. Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis. *Journal of Geophysical Research: Solid Earth*, 103(B2):2759–2779, 1998.
- [26] Wenceslau Peres Gouveia. *Bayesian seismic waveform data inversion: Parameter estimation and uncertainty analysis*. PhD thesis, advisor: John A. Scales, Dept. of Geophysics. Colorado School of Mines, 1996.
- [27] Mathilde Granke, Quentin Grimal, Amena Saïed, Pierre Nauleau, Françoise Peyrin, and Pascal Laugier. Change in porosity is the major determinant of the variation of cortical bone elasticity at the millimeter scale in aged women. *Bone*, 49(5):1020–1026, 2011.
- [28] Samuel H Gray and William P May. Kirchhoff migration using eikonal equation traveltimes. *Geophysics*, 59(5):810–817, 1994.
- [29] Pablo Guillen, German Larrazabal*, Gladys González, Dainis Boumber, and Ricardo Vilalta. Supervised learning to detect salt body. In *SEG Technical Program Expanded Abstracts 2015*, pages 1826–1829. Society of Exploration Geophysicists, 2015.
- [30] Heikki Haario, Eero Saksman, and Johanna Tamminen. An Adaptive Metropolis Algorithm. *Bernoulli*, 7(2):223–242, 2001.
- [31] Dave Hale. Dynamic warping of seismic images. *Geophysics*, 78(2):S105–S115, 2013.
- [32] R Hewett and L Demanet. the pysit team, 2013. *PySIT: Python seismic imaging toolbox v0.5*, 2013.
- [33] R.J. Hewett, L. Demanet, and the PySIT Team. PySIT: Python seismic imaging toolbox v0.5, 2013. Release 0.5.
- [34] Lester Ingber. Very fast simulated re-annealing. 1989.
- [35] Jørgen Arendt Jensen, Svetoslav Ivanov Nikolov, Kim Løkke Gammelmark, and Morten Høgholm Pedersen. Synthetic aperture ultrasound imaging. *Ultrasonics*, 44:e5–e15, 2006.
- [36] Yuchen Jin, Wenyi Hu, Xuqing Wu, and Jiefu Chen. Learn low wavenumber information in fwi via deep inception based convolutional networks. In *SEG Technical Program Expanded Abstracts 2018*, pages 2091–2095. Society of Exploration Geophysicists, 2018.

- [37] Jonathan A Kane, William Rodi, Kenneth P Bube, Tamas Nemeth, Don Medwedé, and Oleg Mikhailov. Structural uncertainty and bayesian inversion. In *SEG Technical Program Expanded Abstracts 2004*, pages 1511–1514. Society of Exploration Geophysicists, 2004.
- [38] Paul Käufl, Andreas Fichtner, and Heiner Igel. Probabilistic full waveform inversion based on tectonic regionalization—development and application to the australian upper mantle. *Geophysical Journal International*, 193(1):437–451, 2013.
- [39] Zvi Koren, Klaus Mosegaard, Evgeny Landa, Pierre Thore, and Albert Tarantola. Monte Carlo estimation and resolution analysis of seismic background velocities. *Journal of Geophysical Research: Solid Earth*, 96(B12):20289–20299, 1991.
- [40] E Landa, P Thore, V Sorin, and Z Koren. Interpretation of velocity estimates from coherency inversion. *Geophysics*, 56(9):1377–1383, 1991.
- [41] L. P. Letki, H. Ben-Hadj-Ali, and P. Desegaulx. Quantifying Uncertainty in Final Seismic Depth Image Using Structural Uncertainty Analysis-Case Study Offshore Nigeria. In *First EAGE West Africa Workshop 2013-Subsurface Challenges in West Africa*, 2013.
- [42] Franklyn K Levin. Apparent velocity from dipping interface reflections. *Geophysics*, 36(3):510–516, 1971.
- [43] Yunfeng Li, Yogesh Agnihotri, and Timmy Dy. Prismatic wave imaging with dual flood rtm. In *SEG Technical Program Expanded Abstracts 2011*, pages 3290–3294. Society of Exploration Geophysicists, 2011.
- [44] Yunyue Elita Li and Laurent Demanet. Full-waveform inversion with extrapolated low-frequency data. *Geophysics*, 81(6):R339–R348, 2016.
- [45] Zhongxiao Li and Zhenchun Li. Accelerated and automatic sparse parabolic radon transform in the mixed frequency-time domain with alternating split bregman algorithm. In *SEG Technical Program Expanded Abstracts 2016*, pages 4524–4528. Society of Exploration Geophysicists, 2016.
- [46] Youshan Liu, Jiwen Teng, Tao Xu, Yanghua Wang, Qinya Liu, and José Badal. Robust time-domain full waveform inversion with normalized zero-lag cross-correlation objective function. *Geophysical Journal International*, 209(1):106–122, 2016.
- [47] Wenkai Lu. An accelerated sparse time-invariant radon transform in the mixed frequency-time domain based on iterative 2d model shrinkage. *Geophysics*, 78(4):V147–V155, 2013.
- [48] Alison Malcolm and David P. Nicholls. A field expansions method for scattering by periodic multilayered media. *The Journal of the Acoustical Society of America*, 129(4):1783–1793, 2011.

- [49] Daniel Maystre. Theory of WoodâŽs Anomalies. In Stefan Enoch and Nicolas Bonod, editors, *Plasmonics*, number 167 in Springer Series in Optical Sciences, pages 39–83. Springer Berlin Heidelberg, 2012.
- [50] Susan E Minkoff. A computationally feasible approximate resolution matrix for seismic inverse problems. *Geophysical Journal International*, 126(2):345–359, 1996.
- [51] Konstantin Osypov, Yi Yang, AimÃƒl Fournier, Natalia Ivanova, Ran Bachrach, Can Evren Yarman, Yu You, Dave Nichols, and Marta Woodward. Model-uncertainty quantification in seismic tomography: method and applications. *Geophysical Prospecting*, 61(6):1114–1134, 2013.
- [52] KL Peacock and Sven Treitel. Predictive deconvolution: Theory and practice. *Geophysics*, 34(2):155–169, 1969.
- [53] Oleg V. Poliannikov and Alison E. Malcolm. The effect of velocity uncertainty on migrated reflectors: Improvements from relative-depth imaging. *Geophysics*, 81(1):S21–S29, 2015.
- [54] D Janaki Ram, TH Sreenivas, and K Ganapathy Subramaniam. Parallel simulated annealing algorithms. *Journal of parallel and distributed computing*, 37(2):207–212, 1996.
- [55] Anandaroop Ray, Anusha Sekar, G Michael Hoversten, and Uwe Albertin. Frequency domain full waveform elastic inversion of marine seismic data from the alba field using a bayesian trans-dimensional algorithm. *Geophysical Journal International*, 205(2):915–937, 2016.
- [56] Guillaume Renaud, Pieter Kruizinga, Didier Cassereau, and Pascal Laugier. In vivo ultrasound imaging of the bone cortex. *Physics in Medicine and Biology*, 2018.
- [57] Luis Miguel Rios and Nikolaos V Sahinidis. Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3):1247–1293, 2013.
- [58] Alexis Roche, Grégoire Malandain, Xavier Pennec, and Nicholas Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1115–1124. Springer, 1998.
- [59] O Roy, MAH Zuberi, RG Pratt, and N Duric. Ultrasound breast imaging using frequency domain reverse time migration. In *Medical Imaging 2016: Ultrasonic Imaging and Tomography*, volume 9790, page 97900B. International Society for Optics and Photonics, 2016.

- [60] Angelo Sajeva, Mattia Aleardi, Bruno Galuzzi, Eusebio Stucchi, Emmanuel Spadavecchia, and Alfredo Mazzotti. Comparing the performances of four stochastic optimisation methods using analytic objective functions, 1d elastic full-waveform inversion, and residual static computation. *Geophysical Prospecting*, 65:322–346, 2017.
- [61] Angelo Sajeva, Mattia Aleardi, Eusebio Stucchi, Nicola Bienati, and Alfredo Mazzotti. Estimation of acoustic macro models using a genetic full-waveform inversion: Applications to the marmousi modelgenetic fwi for acoustic macro models. *Geophysics*, 81(4):R173–R184, 2016.
- [62] Michael Schonewille, Peter Aaron, and Terry Allen. Applications of time-domain high-resolution radon demultiple. *ASEG Extended Abstracts*, 2007(1):1–5, 2007.
- [63] Mrinal K. Sen and Paul L. Stoffa. *Global Optimization Methods in Geophysical Inversion*. Cambridge University Press, February 2013.
- [64] Ranjit Shaw and Shalivahan Srivastava. Particle swarm optimization: A new tool to invert geophysical data. *Geophysics*, 72(2):F75–F83, 2007.
- [65] K Kirk Shung. *Diagnostic ultrasound: Imaging and blood flow measurements*. CRC press, 2015.
- [66] Georgia Stuart, Weihua Yang, Susan Minkoff, and Felipe Pereira. A two-stage markov chain monte carlo method for velocity estimation and uncertainty quantification. In *SEG Technical Program Expanded Abstracts 2016*, pages 3682–3687. Society of Exploration Geophysicists, 2016.
- [67] Yimin Sun, Eric Verschuur, et al. 3d receiver deghosting for seismic-streamer data using l1 inversion in an extended radon space. In *2017 SEG International Exposition and Annual Meeting*. Society of Exploration Geophysicists, 2017.
- [68] Martin Svedlow, Clare D McGillem, and Paul E Anuta. Image registration: Similarity measure and preprocessing method comparisons. *IEEE Transactions on Aerospace and Electronic Systems*, (1):141–150, 1978.
- [69] William Symes, Igor S Terentyev, and Tetyana Vdovina. Getting it right without knowing the answer: Quality control in a large seismic modeling project. In *SEG Technical Program Expanded Abstracts 2009*, pages 2602–2606. Society of Exploration Geophysicists, 2009.
- [70] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. siam, 2005.
- [71] Daniel Trad, Tadeusz Ulrych, and Mauricio Sacchi. Latest views of the sparse radon transform. *Geophysics*, 68(1):386–399, 2003.
- [72] Joel A Tropp. Algorithms for simultaneous sparse approximation. part ii: Convex relaxation. *Signal Processing*, 86(3):589–602, 2006.

- [73] Ilya Tsvankin, James Gaiser, Vladimir Grechka, Mirko Van Der Baan, and Leon Thomsen. Seismic anisotropy in exploration and reservoir characterization: An overview. *Geophysics*, 75(5):75A15–75A29, 2010.
- [74] Quentin Vallet, Nicolas Bochud, Christine Chappard, Pascal Laugier, and Jean-Gabriel Minonzio. In vivo characterization of cortical bone using guided waves measured by axial transmission. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 63(9):1361–1371, 2016.
- [75] Peter JM Van Laarhoven and Emile HL Aarts. Simulated annealing. In *Simulated annealing: Theory and applications*, pages 7–15. Springer, 1987.
- [76] Don W Vasco, Lane R Johnson, and Osni Marques. Resolution, uncertainty, and whole earth tomography. *Journal of Geophysical Research: Solid Earth*, 108(B1), 2003.
- [77] Dirk Jacob Verschuur. *Seismic multiple removal techniques: Past, present and future*. EAGE publications Netherlands, 2006.
- [78] Roelof Versteeg. The marmousi experience: Velocity model determination on a synthetic complex data set. *The Leading Edge*, 13(9):927–936, 1994.
- [79] J. Virieux and S. Operto. An overview of full-waveform inversion in exploration. *Geophysics*, 74(6):WCC1–WCC26, November 2009.
- [80] Kees Wapenaar, Jan Thorbecke, Joost Van Der Neut, Filippo Broggini, Evert Slob, and Roel Snieder. Marchenko imaging. *Geophysics*, 79(3):WA39–WA57, 2014.
- [81] Michael Warner, Tenice Nangoo, Nikhil Shah, Adrian Umpleby, and Joanna Morgan. Full-waveform inversion of cycle-skipped seismic data by frequency down-shifting. In *SEG Technical Program Expanded Abstracts 2013*, pages 903–907. Society of Exploration Geophysicists, 2013.
- [82] M. Woodward, D. Nichols, O. Zdraveva, P. Whitfield, and T. Johns. A decade of tomography. *GEOPHYSICS*, 73(5):VE5–VE11, September 2008.
- [83] Wei Xiong, Xu Ji, Yue Ma, Yuxiang Wang, Nasher M BenHassan, Mustafa N Ali, and Yi Luo. Seismic fault detection with convolutional neural network. *Geophysics*, 83(5):1–28, 2018.
- [84] Tianfang Xu and Albert J Valocchi. A bayesian approach to improved calibration and prediction of groundwater models with structural error. *Water Resources Research*, 51(11):9290–9311, 2015.
- [85] Yang Xue, Mrinal K Sen, and Zhiwen Deng. A new stochastic inference method for inversion of pre-stack seismic data. In *SEG Technical Program Expanded Abstracts 2011*, pages 2757–2761. Society of Exploration Geophysicists, 2011.

- [86] Öz Yilmaz. *Seismic data analysis: Processing, inversion, and interpretation of seismic data*. Society of exploration geophysicists, 2001.
- [87] Kwangjin Yoon, Kurt J Marfurt, and William Starr. Challenges in reverse-time migration. In *SEG Technical Program Expanded Abstracts 2004*, pages 1057–1060. Society of Exploration Geophysicists, 2004.
- [88] CA Zelt and RB Smith. Seismic travelttime inversion for 2-d crustal velocity structure. *Geophysical journal international*, 108(1):16–34, 1992.
- [89] Jie Zhang and George A McMechan. Estimation of resolution and covariance for large matrix inversions. *Geophysical Journal International*, 121(2):409–426, 1995.
- [90] Dehan Zhu and Richard Gibson. Seismic inversion and uncertainty analysis using a transdimensional markov chain monte carlo method. In *SEG Technical Program Expanded Abstracts 2016*, pages 3666–3671. Society of Exploration Geophysicists, 2016.
- [91] Dehan Zhu and Richard Gibson. Seismic inversion and uncertainty quantification using transdimensional markov chain monte carlo method. *Geophysics*, 83(4):1–76, 2018.
- [92] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003.