Important Notice

UNIVERSITY OF CALGARY

Full waveform inversion and uncertainty quantification for robust subsurface monitoring in the energy transition

by

Jinji Li

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN GEOSCIENCE

CALGARY, ALBERTA

DECEMBER, 2025

Abstract

Full-waveform inversion (FWI) is a powerful method for estimating subsurface properties that govern seismic wave propagation, and it is increasingly recognized as a key technology in the global energy transition. Despite its potential, the practical deployment of FWI is hindered by several factors, for example, the limited subsurface illumination imposed by acquisition geometry, the influence of noise, and non-repeatability across repeated seismic surveys in geo-monitoring with FWI. These challenges reduce both the resolution and the reliability of FWI results, ultimately limiting confidence in model interpretation. In this thesis, I present approaches for addressing such issues and mitigating their impact, thereby improving the robustness and applicability of FWI for energy-transition-related monitoring and characterization tasks.

Seismic-while-drilling (SWD) is a well-established technique that is widely discussed in geothermal-related applications. This technique can offer a promising opportunity to enhance FWI by providing unique and transmissive ray paths during drilling, and thus can improve subsurface illumination and enrich FWI models. Conversely, FWI can supply SWD with auxiliary subsurface images that support real-time monitoring of the drilling process. An additional need during drilling is the continuous monitoring of drill-bit positions and drill-bit–rock interactions, which could be incorporated into the FWI framework as additional inversion variables. In this thesis, I propose novel approaches to integrate source-related unknowns into FWI and develop algorithms capable of jointly estimating source properties and subsurface physical models. Using this framework, I investigate how the additional ray paths from SWD can improve FWI results and how FWI can, in turn, be used to estimate drill-bit source characteristics in SWD applications.

Due to the inherent nonlinearity of FWI, the resulting models carry a substantial degree of uncertainty. This uncertainty becomes even more pronounced in time-lapse FWI, which is

increasingly recognized as a vital tool for geophysical monitoring. The amplified uncertainty poses a significant challenge for accurately quantifying temporal variations in subsurface models. I integrate two advanced yet computationally efficient uncertainty quantification approaches into the time-lapse FWI framework. One of them is from the sample-based family, and the other is from the variational inference realm. Using synthetic experiments with varying acquisition geometries and noise levels in the time-lapse seismic surveys, I evaluate and compare the performance of these two methods. Based on these comparisons, I provide recommendations regarding their practicality and applicability in realistic monitoring scenarios.

Finally, I develop a 3D FWI framework for delineating more spatial changes in the subsurface. Compared to 2D approaches, 3D FWI incorporates more realistic wave physics, enabling an improved representation of subsurface changes over time. I validate this framework through numerical experiments designed using a widely studied time-lapse dataset. In addition, I extend the targeted nullspace shuttle method to 3D FWI to suppress inversion artifacts arising from non-repeatability in time-lapse seismic surveys, and thus reduce the challenges in the interpretation. The results demonstrate the strong potential for achieving reliable 3D time-lapse FWI under realistic and relatively sparse acquisition constraints.

Preface

This thesis is written in a manuscript-style format, based on one published paper and two submitted papers. I was the lead investigator and manuscript composer for each of these papers. These papers have been republished in this thesis with permission from the co-authors and the Society of Exploration Geophysicists.

A version of Chapter 3 has been published as: Li, J. L., Keating, S., Innanen, K. A., Shor, R., and Kazemi, N., 2023, Simultaneous waveform inversion of seismic-while-drilling data for P-wave velocity, density, and source parameters: *Geophysics*, 88, no. 6., R751–R767.

A version of Chapter 4 has been submitted to *Geophysics* as: Li, J. L., Innanen, K. A., and Malcolm, A., Uncertainty quantification in time-lapse full waveform inversion of vertical seismic profiling data: contrasting Hamiltonian Monte Carlo and Stein Variational Gradient Descent methods.

A version of Chapter 5 has been submitted to *Geophysics* as: Li, J. L., Pike. K., Hall, K., and Innanen, K. A., 3D targeted nullspace shuttles in 4D FWI: synthetic time-lapse FWI experiments for CO_2 monitoring configured for the Snowflake dataset.

Acknowledgements

I would like to first express my deepest gratitude to my supervisor, Dr. Kristopher Innanen. Kris has consistently provided insightful guidance on every research project and embodies an inspiring attitude toward scientific work. In moments when I felt lost in my research or overwhelmed by competing research ideas, he was the steady presence who helped me regain clarity and focus. Beyond shaping my academic skills, Kris has encouraged me to grow as a better person rather than simply as a "project machine," offering support that has extended even beyond academic work. I am profoundly grateful for his mentorship, patience, and unwavering encouragement.

Dr. Daniel Trad has encouraged me to view my work from a practical perspective. Drawing on his extensive experience in both industry and academia, he has provided invaluable guidance in improving my codes and numerical methods. Our discussions, grounded in practical considerations, have greatly accelerated my experiments and tests, without which I could not have achieved nearly as much.

Meeting Dr. Alison Malcolm in the later stage of my study was a privilege. Her insightful advice and enthusiasm for geophysics have had a profound impact on me. I am sincerely grateful and feel very fortunate to have her as a member of my supervisory committee, and I hope to collaborate with her again in my future academic life.

Kevin Hall has provided invaluable support in all matters related to data and computers. His experience with field data has been particularly helpful for several of my projects, especially when real-world applications needed to be considered. Beyond his technical contributions, I have greatly enjoyed Kevin-style dry sense of humor, which has brought laughter to otherwise tedious days in the office.

I would like to thank Kevin Bertram for his technical support, particularly with his valuable advice on poster preparation.

I am especially grateful to Dr. Scott Keating for encouraging me through his example as he advances along his academic path. He was the one who first introduced me to this field, and I have been greatly inspired by his work throughout my research.

All the CREWES associates, staff, and students have made my time in the program truly enjoyable. I will always remember both the joy and the challenges we shared on that small boat with Paloma Lira Fontes, Kai Zhuang, Marcelo Guarido, and our captain, David Emery. Your companionship has meant a great deal to me. I am grateful to Kimberly Pike for her steady support and encouragement. I also want to acknowledge Anton Ziegon, whose intelligence and friendship were invaluable during the later years of my research. The responsibility, skills, and work ethic of Ivan Sanchez continually reminded me to stay humble and strive for improvement. I am fortunate to have the steady friendship of Shang Huang and her husband. I also extend my appreciation to Chioma Chineke, Xiaohui Cai, Carla Acosta, Christina Schumacher, Tianze Zhang, He Liu, Ziguang Su, Xin Fu, Qi Hu, Na Fan, Zhu Yang, Anqi Jia, and Arvin Karpia. I truly value the interactions we have shared.

I would like to thank the University of Calgary, the CSEG, and the SEG for financial support through scholarships.

I would also like to express my deepest gratitude to my parents for their unwavering love and encouragement. Miles away, their constant care has always been a source of strength.

Finally, and most sincerely, I would like to express my heartfelt appreciation to my wife, QIAO Xin, my steadfast companion on this rocky journey. Beginning our Ph.D. studies side by side in a foreign land, we have grown not only as researchers but as partners, supporting and uplifting each other through every challenge and triumph. Although miles still lie between us, her resilience and open heart have been a steady beacon, illuminating the path when days felt uncertain and reminding me that I am never alone in this pursuit. It is this shared journey, with its trials and joys day by day, that has made each achievement even more meaningful and has deepened my gratitude.

Table of Contents

List of Figures and Illustrations

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
2D	2-dimensional
3D	3-dimensional
ADVI	Automatic differential variation inference
BFGS	Broyden-Fletcher-Goldfarb-Shanno method
CCUS	Carbon capture, utilization, and storage
CMC	Carbon Management Canada
CO_2	Carbon dioxide
CPU	Central processing units
CREWES	Consortium for Research in Elastic Wave Exploration Seismology
CV	Coefficient of variation
ELBO	Evidence lower bound
FWI	Full waveform inversion
FRS	Field Research Station
HMC	Hamiltonian Monte Carlo
KL	The Kullback-Leibler divergence
L-BFGS	Limited memory Broyden-Fletcher-Goldfarb-Shanno method
LU	Denotes lower-upper
MC	Monte Carlo
MCMC	Monte Carlo Markov Chain
MUMPS	MUltifrontal massively parallel sparse
NRMSE	Normalized root mean squared error
OBN	Ocean-bottom node
PARDISO	Parallel direct sparse solver

PML	Perfect matched layer
RBF	The radial basis function
RMSE	Root mean squared error
SD	The gradient descent method
SNR	Signal-to-noise ratio
SWD	Seismic-while-drilling
SVGD	Stein variational gradient descent
TGN	The truncated Gauss Newton method
a_i	Lower boundary of model parameter m_i
b_i	Upper boundary of model parameter m_i
$\mathbf{B}$	The inverse of the approximated Hessian matrix
$\mathbf{B}_0^k$	The initial guess of the inverse of the $\mathbf{B}$ in the kth iteration
$\mathbf{B}_k$, $\mathbf{B}_{k-1}$	The inverse of the $\mathbf{B}$ in the kth and $k-1$th iteration
C	A satisfies the assumed wave equation
$\mathbf{C}_D^{-1}$	The inverse covariance matrix of the observed data
$\mathbf{C}_M^{-1}$	The inverse covariance matrix of the model
c	A phase velocity
c_1	The first Wolfe parameter
c_2	The second Wolfe parameter
c_{min}	Minimum velocity
$\mathbf{D}$	The updating direction
$\mathbf{d}$	Data vector
$\mathbf{d}_{obs}$	The observed data vector
$\mathbf{d}_{syn}$	The synthetic data vector
$data_{true}$	Data used in calculating the NRMSE
E	The expectation operator over a distribution

E_ε	The Huber norm value in a subsurface grid
e_i	The quantity to calculate the Huber norm
$\mathbf{f}$	Source term in wave equation
$\mathbf{f}_I$	Imaginary part of the source term in wave equation
$\mathbf{f}_{I_p}$	Imaginary part of the pth source unknown
$\mathbf{f}_R$	Real part of the source term in wave equation
$\mathbf{f}_{R_p}$	Real part of the pth source unknown
$\mathbf{f}_r$	Source term as the inversion variable
f_x	Source term in x-direction
f_z	Source term in z-direction
$\mathbf{g}$	The gradient term
h_m	A scale factor in the radial basis function
$\mathbf{H}$	Hessian matrix
$\mathbf{H_{GN}}$	Gauss-Newton Hessian
H	The value of the Hamiltonian equation
$\mathbf{I}$	The identity matrix
$\mathbf{J}$	The Jacobian matrix
K	The kinematic energy in the Hamiltonian equation
$k(.,.)$	A kernel function in the SVGD method
L	The length of the drilling trajectory
$L(\mathbf{m})$	The Lagranian of an optimization problem
M_0	Seismic moment scalar that denotes the relative energy
M_{ij}	The jth-derivative of the ith displacement component
M_{kl}	Moment tensor elements
$\mathbf{M}_{mass}$	The mass matrix in the HMC method
$\mathbf{M}_\varepsilon$	The model class that holds a small FWI objective function

$\mathbf{m}$	Model vector
$\hat{\mathbf{m}}$	Optimal model acquired by full waveform inversion
$\mathbf{m}^{'}$	Another model vector
$\mathbf{m}^{*}$	A model that holds a small FWI objective value
$\mathbf{m}_b$	Baseline model
$m_{b,i}$	The baseline model value in the ith grid
$\mathbf{m}_{cur}$	The current position (model) of a particle
$\mathbf{m}_m$	Monitor model
$m_{m,i}$	The monitor model value in the ith grid
$\mathbf{m}_k$, $\mathbf{m}_{k+1}$	The model in kth and $k+1$th iteration
$\mathbf{m}_{new}$	The new position (model) of a particle
$\mathbf{m}_p$	The model subset in $\mathbf{m}$
$\overline{\mathbf{m}_b}$	Baseline mean model
$\overline{\mathbf{m}_m}$	Monitor mean model
$\overline{\mathbf{m}_m - \overline{\mathbf{m}_b}}$	Mean model of $\mathbf{m}_m - \overline{\mathbf{m}_b}$
N	The total number of particles in the SVGD method
N_ω	Total number of angular frequencies
N_{grid}	The total number of subsurface grids
N_s	Total number of sources
$nPML$	PML boundary width
nx	Total grid points in x-direction
ny	Total grid points in y-direction
nz	Total grid points in z-direction
$P_{acceptance}$	The acceptance probability in the Metropolis algorithm
$P(\mathbf{d})$	The prior probability distribution of data
$P(\mathbf{d}\|\mathbf{m})$	The likelihood

$P(\mathbf{m}))$	The prior probability distribution of model parameters
$P(\mathbf{m}\|\mathbf{d})$	The posterior distribution
$\mathbf{p}$	The momentum vector of a particle in the HMC method
$\mathbf{p}_{cur}$	The current momentum vector of a particle
$\mathbf{p}_{new}$	The new momentum vector of a particle
Q	A family of distributions
Q_p	The P-wave quality factor
Q_s	The S-wave quality factor
$\mathbf{q}$	The position vector of a particle in the HMC method
$q^*(\mathbf{m})$	The optimal distribution in the SVGD method
$q_T^*(\mathbf{m})$	The optimal distribution acquired by a smooth transform
$q(\mathbf{m})$	A distribution q related to model $\mathbf{m}$
$q_0(\mathbf{m})$	The initial distribution q related to model $\mathbf{m}$
q_i, q_{i+1}	The distribution in the ith and $i+1$th iterations
$\mathbf{R}$	Receiver sampling matrix
$\mathbf{r}$	Position vector in Cartesian coordinates
$\mathbf{r}_s$	Position vector starts from arbitrary sources
$\mathbf{S}$	Impedance matrix
s	A slowness
T	A transform function
T_i	Transform function in the ith iteration
t	A time quantity
U	The potential energy in the Hamiltonian equation
$\mathbf{u}$	Wavefields
$\overline{\mathbf{u}}$	The wavefield satisfies $L(\overline{\mathbf{m}}) = \Phi$
u_x	The x-component of a wavefield

u_z	The z-component of a wavefield
V	The smooth transform direction
V^*	The optimal smooth transform direction
V_i^*	The optimal smooth transform direction in the ith iteration
V_P	P-wave velocity
$\mathbf{v}$	The arbitrary vector in Hessian-vector product
$v_{\mathbf{f}_I}$	The subvector of the imaginary part source subset in $\mathbf{v}$
$v_{\mathbf{m}_p}$	The subvector denoting the subset in $\mathbf{v}$
$v_{\mathbf{f}_R}$	The subvector of the real part source subset in $\mathbf{v}$
$\mathbf{w}$	The product of $\mathbf{R}^T\mathbf{RJv}$
$w_{dx}(\mathbf{r})$	Spatial weighting term in x-direction
$w_{dz}(\mathbf{r})$	Spatial weighting term in z-direction
$\mathbf{w}(\mathbf{r})$	Spatial weighting term
$\mathbf{x}$	A general inversion variable vector
$\mathbf{x}_p$	The pth variable in $\mathbf{x}$
x	Position coordinate that is orthogonal to y- and z-direction
y	Position coordinate that is orthogonal to x- and z-direction
z	Vertical coordinate
x_1, x_2, and x_3	Position coordinates also denote x, y, and z
α	The steplength in a gradient-based optimization problem
α_i	The steplength the ith iteration in the SVGD approach
α_k	The steplength the kth iteration in Line-search
$\Delta\mathbf{g}_k$, $\Delta\mathbf{g}_{k-1}$	The gradient increment in the kth and $k-1$th iteration
Δh	Grid interval
$\Delta\mathbf{m}$	Small increment in the model
$\Delta\mathbf{m}^{'}$	The target model increment

$\Delta\mathbf{m}_b^{'}$	Update direction of the baseline model as inversion variable
$\Delta\mathbf{m}_k$, $\Delta\mathbf{m}_{k-1}$	The model increment in the kth and $k-1$th iteration
$\Delta\mathbf{m}_m^{'}$	Update direction of the monitor model as inversion variable
ΔU	The gradient of the potential energy in the HMC method
ΔV_P	Time-lapse change of P-wave velocity
Δx	Small increment in x-direction
Δy	Small increment in y-direction
Δz	Small increment in z-direction
δ	Dirac function
$\delta\mathbf{m}$	Small increment in the model perturbation
δt	Time step in the HMC method
λ	Lagrange multiplier
$\overline{\lambda}$	The adjoint wavefield
$\tilde{\lambda}$	The first complex Lamé parameter
μ	Lagrange multiplier in the TGN method
$\overline{\mu}$	The adjoint wavefield in the TGN method
$\tilde{\mu}$	The second complex Lamé parameter
ω	Angular frequency
ω_0	Reference angular frequency
Φ	Objective function in full waveform inversion
$\mathbf{\Phi}_b$	Baseline objective function in full waveform inversion
$\mathbf{\Phi}_m$	Monitor objective function in full waveform inversion
Ψ	Objective function in the targeted nullspace shuttle
ρ	Density
σ	The standard deviation
τ	A time when a seismic source is activated

θ	The inclination angle of the drilling trajectory
ε	An acceptable misfit threshold
$\boldsymbol{\xi}$	A location where a seismic source is activated
$\mathcal{N}$	The normal distribution symbol
$\mathfrak{Im}$	Imaginary part
$\mathfrak{Re}$	Real part
$\sum$	Summation symbol
∇	The derivative operator
$\|\|.\|\|$	The L_2 norm operator
$< ., . >$	The inner product operator

Chapter 1

Introduction

1.1 Full waveform inversion in energy transition

The global energy landscape has undergone a profound transformation in recent decades, driven by the urgent need to mitigate climate change and accelerate the transition toward sustainable energy systems. This shift is characterized by a gradual reduction in reliance on fossil fuels and a growing emphasis on renewable and low-carbon alternatives. However, the path to a low-carbon future introduces significant technological challenges that require innovative solutions. Among the most promising approaches are geothermal energy and carbon capture, utilization, and storage (CCUS). Geothermal energy harnesses the internal heat of the Earth, providing a reliable, sustainable, and carbon-neutral resource with the potential to diversify energy portfolios and strengthen energy security (Barbier, 2002). In parallel, CCUS seeks to reduce greenhouse gas emissions from fossil fuel infrastructure by capturing carbon dioxide (CO_2) and storing it underground, thus mitigating the environmental impact of conventional energy systems and offering a critical tool for achieving climate goals (Lawton et al., 2019; Dziejarski et al., 2023).

Both geothermal energy and CCUS offer significant benefits but face distinct technical challenges. Geothermal energy extraction requires drilling deep into the interior Earth to harness heat from hot water or steam. Although the process resembles conventional resource extraction techniques, such as oil and gas drilling, it presents its own set of complications. The rock formations encountered are sometimes complicated, demanding advanced control methods to effectively manage the drilling processes. Moreover, deviated drilling trajectories are frequently employed, requiring careful optimization to avoid zones with natural fractures and faults in the rock (Archer, 2020). Effective monitoring during geothermal operations

is thus crucial, ensuring both safety and maximizing resource extraction efficiency (Auriol et al., 2019; Sharmin et al., 2023). In contrast, CCUS focuses on securely storing captured CO_2 in underground reservoirs. Ensuring containment and operational conformance requires advanced modeling, measurement, and monitoring technologies to delineate the subsurface accurately and guarantee the integrity of the storage site (Lawton et al., 2019). A key component of this process is time-lapse seismic monitoring, which verifies that stored CO_2 remains trapped underground and does not migrate into unintended areas (Lumley, 2001; Arts et al., 2004; Chadwick et al., 2010).

Seismic waves, generated either by natural events or artificial sources, contain valuable information about the subsurface properties. However, accurately predicting the internal structures of the interior Earth remains a complex challenge. While traditional seismic inversion techniques, such as travel-time tomography (Bregman et al., 1989), migration (Stolt, 1978; Nemeth et al., 1999), and AVO inversion (Ostrander, 1984; Innanen, 2014) have been useful for retrieving certain subsurface characteristics, they utilize a limited portion of the available seismic data, thus limits the comprehensive depiction of the resulting models. Full waveform inversion (FWI) offers a solution by leveraging all recorded seismic data types, including reflections, refractions, diving waves, and multiple-scattered waves (Virieux and Operto, 2009), to produce high-resolution subsurface models by iteratively minimizing the difference between the observed and synthetic seismic data (Tarantola, 1984, 2004; Fichtner, 2010), either in 2D or 3D spaces, and in time or frequency domains (Marfurt, 1984; Pratt, 1990; Brossier et al., 2009).

Despite the promise of FWI, its integration into geothermal energy and CCUS applications presents challenges that demand tailored strategies. In drilling environments, the seismic energy generated by the drill bits offers a unique opportunity to enhance subsurface illumination, yet accounting for the dynamic positions of the drilling-related sources and the evolving formation properties introduces significant inversion complexity and thus increases

the risk of data leakage, where physical properties may be misattributed. In CCUS monitoring, a central challenge lies in quantifying uncertainty in time-lapse FWI. Non-repeatability between baseline and monitor surveys, combined with acquisition noise, can propagate into the inversion results and undermine the confidence in delineating CO_2 plumes. Robust and efficient uncertainty quantification strategies are thus essential to ensure reliable and risk-aware monitoring. Finally, because CO_2 plumes evolve in three dimensions, extending FWI into the 3D domain is necessary. However, the added dimension intensifies non-uniqueness and inversion ambiguities, underscoring the need for advanced strategies to extract meaningful time-lapse variations.

1.2 Incorporation of seismic-while-drilling and FWI

The foundations of FWI were established through early acoustic formulations based on simplifying assumptions, such as isotropic density (Lailly, 1983; Tarantola, 1984), which provided the first framework for seismic imaging. However, these approaches often proved insufficient to reproduce the full complexity of observed seismic data. This limitation motivated the development of more advanced FWI formulations, incorporating elastic (Innanen, 2014; Pan et al., 2018), viscoacoustic (Métivier et al., 2015; Plessix et al., 2016), and viscoelastic physics (Keating and Innanen, 2020), which have progressively improved the reliability and resolution of subsurface models. In environments where a new class of drilling-related parameters is introduced, multiparameter FWI is particularly promising, as it can exploit the complex wavefields generated during the drilling process.

However, a persistent challenge in multiparameter FWI is the cross-talk, which arises when residual data are misattributed to parameters that only appear responsible for the observed discrepancies. This problem intensifies as the number of parameters in the inversion increases (Bunks et al., 1995; Operto et al., 2013; Innanen, 2014; Métivier et al., 2015; Pan et al., 2016; Keating and Innanen, 2020), especially when drilling-induced seismic sources are

themselves treated as inversion variables. Beyond parameter coupling, limitations in acquisition geometry further exacerbate cross-talk. Traditional seismic surveys often suffer from incomplete spatial sampling and restricted energy coverage of the subsurface (Jannane et al., 1989; Mothi et al., 2014; Kerrison et al., 2021). Because seismic waves preferentially refract toward high-velocity zones (e.g., salt bodies) and away from low-velocity anomalies (e.g., overpressure zones) (Cheadle et al., 1991; Kazemi et al., 2018, 2020), these acquisition gaps produce poorly constrained "shadow zones" where parameter estimates become unreliable. Although wide-aperture acquisition designs can improve ray coverage and mitigate some of these limitations, achieving a fully constrained subsurface model remains a fundamental challenge (Virieux and Operto, 2009; Vigh et al., 2021).

As mentioned, using dynamic sources generated by the drilling process offers a promising solution, a technique known as seismic-while-drilling (SWD). In SWD, the drill bit acts as a continuous seismic source, producing energy as it interacts with the surrounding rock. A key advantage of this approach is that it eliminates the need for downhole instrument installation, thereby reducing costs associated with deploying additional seismic sources. Furthermore, SWD operates simultaneously with drilling, enabling seismic data acquisition without interrupting operations and allowing for real-time monitoring or depth measurements after processing.

Although the idea of using drill-bit sources dates back to the 1930s, it was not until the 1980s that researchers began systematically studying such signals to produce reliable geophysical results (Klaveness, 1980; Katz, 1984; Staron et al., 1988). Since then, numerous studies have explored the potential of SWD for seismic imaging. For example, early work by James W. Rector and Hardage (1992) focused on the source signatures and wavefields generated by roller-cone bits, while Poletto and Miranda (2004) later addressed challenges like incorrect deconvolution due to drill-bit multiples. Other studies have examined the influence of drill-bit boundary conditions based on varying rock properties (Richard et al.,

2004; Germay et al., 2009).

In the field of seismic imaging, SWD has shown promise in both synthetic models and field tests. For example, Kazemi and Sacchi (2014) applied a multichannel sparse blind deconvolution (SMBD) algorithm to recover the drill-bit signature and integrate SWD data with surface seismic images, enhancing subsurface illumination. Further developments by Kazemi et al. (2018) demonstrated the feasibility of using SWD source variability to improve migration results, while Kazemi et al. (2021) developed a modeling framework for the drill-bit source signature that could be applied to successive FWI. These studies highlight that the transmissive ray paths generated by SWD can extend seismic coverage, especially in regions poorly illuminated by traditional surface acquisitions. As shown in Figure 1.1, such additional paths have the potential to improve FWI resolution and mitigate cross-talk by providing the missing data required to better constrain the subsurface.

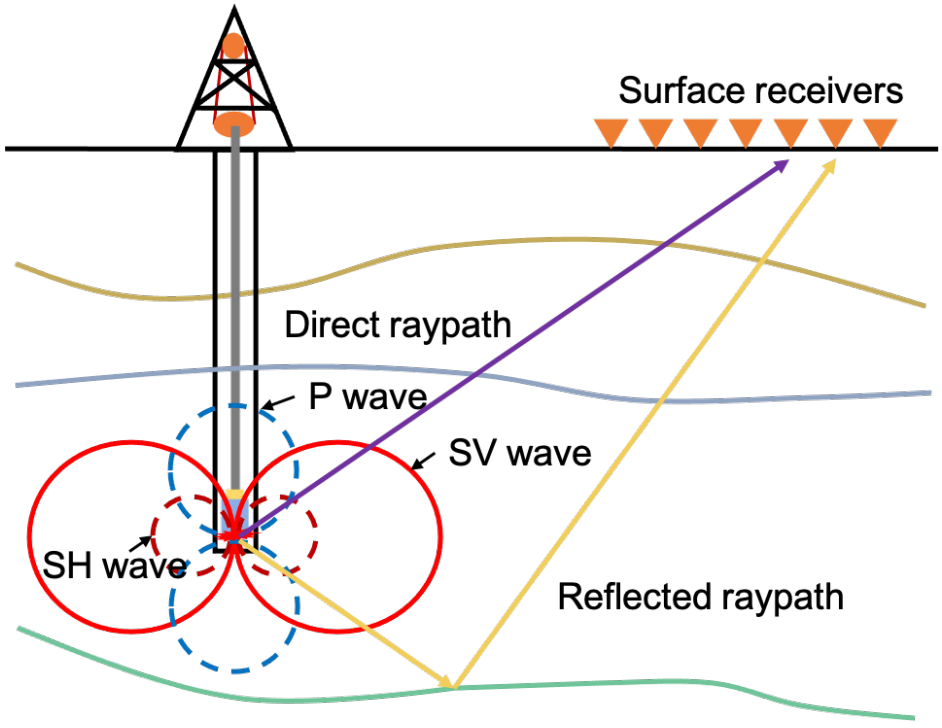

Figure 1.1: Schematic figure of the radiation patterns of SWD sources (modified after James W. Rector and Hardage (1992)).

Despite its potential, integrating SWD into FWI presents challenges. Accurately recovering both the subsurface properties and the seismic source characteristics generated by the drill bits is a complex task. To effectively integrate SWD data into the inversion, it is crucial to parameterize the moving source locations and radiation patterns. This process requires the simultaneous recovery of both the subsurface model and the source parameters, which is further complicated by the uncertainties inherent in the drilling environment. Although unique transmissive energy can enhance the subsurface illumination, cross-talk between the subsurface properties and the unknown source parameters could lead to errors and misinterpretations of the subsurface model if the inversion is not properly managed (Willacy et al., 2019).

1.3 Time-lapse FWI and its role in CCUS

The global CCUS sector has undergone remarkable expansion with a combined annual capacity exceeding millions of metric tons of CO_2. This growth is evidenced by the proliferation of over 100 projects at various stages of development, including the Shute Creek project in the USA (Parker et al., 2011), the Sleipner CCUS project in Norway (Furre et al., 2017), and the QUEST project in Canada (Duong et al., 2019). Another pilot-scale project is the Containment and Monitoring Institute (CaMI) project launched under Carbon Management Canada (CMC), and various monitoring research projects are ongoing in the Field Research Station (FRS), aiming at advancing monitoring technologies for CCUS. The FRS site is located 200 km southeast of Calgary in Newell County. Some details of this facility and recent research can be found in Lawton et al. (2019); Macquet and Lawton (2019); Wang and Lawton (2023).

Monitoring the injection and depletion of fluids in subsurface reservoirs is essential for understanding the dynamic processes that occur during fluid migration. These processes lead to changes in reservoir saturation, pressure, and temperature, which in turn affect seismic

velocities and densities in the subsurface. Time-lapse seismic is a well-established method that captures such changes over time by conducting repeated seismic surveys of the same area. This allows for continuous monitoring of reservoir evolution, with the assumption that geological structures remain static and any observed seismic variations are attributed to temporal changes within the reservoir—typically caused by fluid injection or depletion (Lumley, 2001).

Time-lapse seismic has long been a key tool in oil reservoir management, where it is used to monitor the changes in reservoirs caused by fluid injection and subsequent production. It plays a crucial role in identifying bypassed oil for infill drilling, helping to extend the economic life of oil fields (Lumley, 2001). Over time, its application has expanded to include monitoring long-term CO_2 sequestration in CCUS projects. Typically, a baseline survey is conducted before fluid injection begins, followed by a series of monitoring surveys over time to track the evolution of the reservoir. The insights from time-lapse seismic are often derived through qualitative analysis, which focuses on visualizing changes in amplitude maps and time shifts at specific horizons. For example, impedance contrasts and changes in seismic response, such as amplitude variations and tuning effects, have been used to characterize CO_2 accumulations in thin layers. Additionally, velocity pushdown effects, where seismic waves slow down as they pass through CO_2-saturated zones, have been identified (Arts et al., 2004). While useful, these qualitative approaches are often limited by their reliance on post-stack data and simplified convolution models. As a result, they may fail to capture more subtle changes, particularly when seismic variations visible at certain angles are obscured in stacked data. Furthermore, amplitude changes may not be well-preserved if the velocity model is not updated between surveys.

FWI, with its capacity to maximize data information and resolve detailed velocity models, is increasingly regarded as optimal for tracking CO_2 plumes (Nakata et al., 2022; Hu et al., 2022). This capability is critical for distinguishing velocity perturbations caused by

CO_2 migration from irrelevant background noise, a challenge magnified by the dynamic nature of storage sites. Different FWI schemes have been proposed for time-lapse monitoring. These include parallel FWI, where baseline and monitoring datasets are inverted separately and then compared (Plessix and Perkins, 2010); sequential FWI, where the baseline model is used as the initial model for monitor surveys to improve consistency (Asnaashari et al., 2015); and double-difference FWI, which focuses on inverting only the differences between baseline and monitor datasets, reducing the impact of artifacts and noise (Zhang and Huang, 2013). Additionally, joint inversion, which simultaneously inverts multiple datasets, ensures consistency and reduces uncertainty across time-lapse surveys (Rittgers et al., 2016). Several successful FWI applications have been demonstrated in the Sleipner project, including high-resolution velocity and thickness characterization of CO_2-bearing subsurface structures (Romdhane et al., 2014), the reconstruction of a finely resolved velocity model revealing at least seven distinct layers within the CO_2 distribution chimney (Mispel et al., 2019), and the effective use of sparse ocean-bottom node (OBN) data in FWI to obtain an accurate velocity model capturing plume evolution (Stock et al., 2024). Despite its advantages, time-lapse FWI faces challenges. Non-repeatability between baseline and monitor surveys, due to acquisition or environmental variations, can introduce inconsistencies that complicate data interpretation and may lead to false subsurface changes if not properly managed (Mosegaard and Tarantola, 1995; Kotsi et al., 2020; Keating and Innanen, 2021). Additionally, the inherent complexities of FWI, such as sparse data coverage, noise, and imperfections in forward modeling, contribute to uncertainty in the inversion process (Parker, 1977). Such uncertainty will get further amplified in a time-lapse context.

1.4 Uncertainty quantification in FWI: sample-based methods and variational inference

Uncertainty quantification has thus become an important aspect of time-lapse FWI, particularly in the context of CCUS projects where the complexity of FWI is increased because of the consecutive inversions. Quantifying uncertainty helps assess the reliability of the inversion results and ensures that decision-makers can confidently interpret the changes observed in the subsurface. Thus, addressing these uncertainties is vital for successfully applying FWI in CCUS projects, where accurate and reliable monitoring is crucial for operational safety and regulatory compliance.

One commonly employed set of approaches for assessing uncertainty in FWI problems is the realm of Monte Carlo (MC) based sampling methods within a Bayesian framework. These techniques involve proposing various combinations of model parameters and evaluating their acceptance probabilities, ultimately generating one or multiple Markov chains comprising numerous eligible models (Dosso et al., 2014; Fu and Innanen, 2022). This process aids in estimating crucial statistical information, such as the posterior distribution of model parameters (Brooks et al., 2011). MC-based methods have found widespread application in uncertainty quantification across various geophysical domains, including seismic tomography (Zhang et al., 2018), gravity and magnetic inversion (Zunino et al., 2022), and FWI (Gebraad et al., 2020; Fichtner et al., 2021), where more efficient MC variants are often employed. However, it is important to acknowledge that the sampling process may become computationally demanding due to the repetitive simulation of datasets, especially when dealing with higher-dimensional problems (Bellman, 2003). This high computational expense can hinder the practical application of MC approaches, particularly in scenarios where computational resources are limited. Additionally, sample-based methods may encounter difficulties when dealing with highly complex distributions. Hamiltonian Monte Carlo, or HMC, (Duane et al., 1987), initially conceptualized as a hybrid MC method, integrates el-

ements from two distinct methodologies: (1) gradient-based optimization, which efficiently identifies optima but lacks comprehensive uncertainty information, and (2) derivative-free MC methods, which may overlook potentially valuable derivative information. HMC endeavors to harness the strengths of both approaches by simulating a Hamiltonian dynamics system (Hamilton, 1834), wherein the Hamiltonian function encompasses the cumulative potential and kinetic energy of the system. Leveraging derivatives from both the potential and kinetic energy components, HMC facilitates targeted exploration of plausible regions within the model space, ultimately resulting in more efficient sampling in high-dimensional parameter spaces in complex Bayesian inference problems (Neal, 1993; Brooks et al., 2011). This approach has found widespread application in diverse fields, ranging from neural networks and machine learning (Neal, 1996; Bishop and Nasrabadi, 2006; Gutmann and Hyvärinen, 2012) to molecular simulations (e.g., (Dubbeldam et al., 2016)), nuclear physics (e.g., (Elhatisari et al., 2015)), genomics (e.g., (Honkela et al., 2015)), signal processing (e.g., (Wei et al., 2015)), and quantum mechanics (e.g. (Seah et al., 2018)). Discussions of HMC in FWI can be found in Fichtner et al. (2018); Gebraad et al. (2020); Kotsi et al. (2020); de Lima et al. (2023).

Variational inference methods (Jordan et al., 1998) present another alternative, recently gaining traction in various fields including statistical physics (Regier et al., 2015), data modeling (Tabouy et al., 2020), and machine learning (Kucukelbir et al., 2016). Essentially, this ensemble of methods can frame the problem as an optimization task aimed at acquiring an approximation that best represents the posterior probability density of model parameters (Blei et al., 2017). This aspect makes them particularly attractive for addressing larger-scale Bayesian problems where the computational effort of traditional sampling methods becomes a limiting factor. Several studies have sought to balance accuracy and computational tractability through approaches such as the evidence lower bound (ELBO) approximation and the mean-field variational family, which assumes independent latent variables

(Blei et al., 2017; Zhang and Curtis, 2020). Extensions that introduce dependencies among variables have also been explored (Jaakkola and Jordan, 1997). Additionally, automatic differentiation variational inference (ADVI) offers an efficient approximation using a sequence of invertible transformations, making it applicable to a wider range of inversion problems (Kucukelbir et al., 2016). The Stein variational gradient descent (SVGD) (Liu and Wang, 2019), which also performs transforms but does not require their analytical forms, provides insights into this area and has the potential to be more practically applied to tasks such as FWI. In the field of geophysics, researchers have pioneered in demonstrating the effectiveness of variational inference methods across diverse applications, spanning from seismic tomography (Zhang et al., 2018) to the intricate domain of FWI problems (Zhang and Curtis, 2020). Other applications can be found in Nawaz and Curtis (2018, 2019).

1.5 Targeted nullspace shuttles

While the methods discussed in the previous section characterize FWI uncertainty through global model-space sampling, an alternative and computationally efficient strategy targets the inversion nullspace, which is formed by a model subset that produces equally acceptable data misfits near convergence (Deal and Nolet, 2007). Instead of exhaustively exploring the entire solution space, this approach focuses on identifying model variations that preserve the data fit while revealing the structural ambiguity inherent in nonlinear inverse problems.

The concept of nullspace analysis is first introduced in seismic tomography by Deal and Nolet (2007), exploring how an inverted model could be altered without affecting the objective function. By projecting a predefined model perturbation onto the nullspace of the forward modeling operator, this approach is able to identify the closest update that leaves the misfit unchanged. This technique, then referred to as nullspace shuttling, is originally proposed as an alternative to conventional regularization. Building on this idea, more recent studies (e.g., Fichtner and Zunino (2019); Liu and Peter (2020)) have employed nullspace-

based methods to investigate uncertainty and non-uniqueness in nonlinear geophysical inversions.

Before the introduction of this technique in FWI, it is important to clarify the terminology here. In strict linear algebra, the nullspace refers to the exact kernel of a linear operator, containing directions that map to zero under that operator. In nonlinear FWI, however, the term is adopted in a more approximate sense. The "nullspace" commonly denotes directions in the model space that exhibit very weak sensitivity to the data misfit. Although this usage departs from the classical definition, the concept remains practically valuable. Modern implementations, such as those discussed later in this section, leverage these weakly constrained directions to refine model updates and improve structural consistency to enhance some desired features, even though the method is not intended as a probabilistic measure of uncertainty.

In the context of FWI, Keating and Innanen (2021) have formulated a structured nullspace shuttle method to perform targeted uncertainty quantification. Instead of attempting to fully characterize the high-dimensional posterior, this approach evaluates uncertainty with respect to specific hypotheses or features of the post-inversion model. This approach has also shown the potential in mitigating the cross-talk issues in multi-parameter FWI. In time-lapse FWI, Keating and Innanen (2024) have extended the concept of the targeted nullspace shuttle to actively address one of the core challenges in seismic monitoring, which is distinguishing genuine subsurface changes from artifacts introduced by non-repeatable factors such as acquisition, data processing, and noise. The method operates through an iterative optimization framework that identifies the optimal direction of model perturbation within the nullspace. This is done by minimizing a secondary objective function designed to enforce structural similarity between baseline and monitor models, while still preserving fidelity to the recorded data. In doing so, the algorithm produces time-lapse differences that are not only physically plausible but also more robust to acquisition-related inconsistencies and noise. This

approach serves as a post-inversion correction mechanism, refining inversion results without altering the data fit.

Subsequent validations by Pike et al. (2024a) and Pike et al. (2024b) have further confirmed the resilience of this method across a variety of challenging scenarios, including variations in acquisition geometry, signal-to-noise levels, and inversion parameterizations. These studies highlight the targeted nullspace shuttle as an effective and computationally efficient strategy for improving the interpretability and reliability of time-lapse FWI results, particularly in operational monitoring contexts such as CO_2 sequestration and fluid migration detection.

1.6 3D frequency-domain FWI

Many FWI applications and explorations rely on 2D wavefield approximations, which assume a simplified 2D subsurface model (e.g., Brossier et al. (2009); Groos et al. (2017); Keating and Innanen (2020)). This is largely due to the substantial computational demands that 3D FWI requires. However, 2D approximations may be inadequate for accurately capturing the complex scattering effects from 3D subsurface structures, resulting in a simplified representation of the actual 3D subsurface. In CCUS applications, azimuthal variations in the subsurface must be considered to ensure the success of accurate FWI monitoring. Achieving accurate seismic wavefield modeling in these cases necessitates full 3D simulations, which can provide a comprehensive view of the wavefield interactions and subsurface dynamics.

While research into 3D FWI has been expanding (e.g., Fichtner (2010); Castellanos et al. (2011); Guasch et al. (2012); Butzer et al. (2013); Butzer (2015)), it remains computationally expensive and is less commonly applied. While incorporating multiple physical parameters, such as density and shear wave velocities, makes the simulations more comprehensive, it also increases computational demands and can lead to data leakage between parameters, further complicating the inversion process. This added complexity makes 3D elastic FWI

challenging to implement on a large scale. By contrast, 3D acoustic FWI, which mainly captures the acoustic variations of the subsurface, is computationally less intensive, has seen wider adoption. Several notable examples of successful 3D acoustic FWI applications can be found in works such as Plessix (2009); Sirgue et al. (2010); Warner et al. (2013); Park and Oh (2023).

As mentioned, FWI can be performed in either the time or frequency domain. In contrast to time-domain FWI, which models the entire wavefield across all frequencies, frequency-domain FWI focuses on inverting a selected set of discrete frequencies. This targeted approach reduces the overall data volume and improves computational efficiency, particularly by excluding problematic data that may arise from acquisition errors in real-world applications. Frequency-domain FWI also offers advantages in managing frequency-dependent parameters (Toksöz et al., 1981) and helps mitigate the cycle-skipping issues often encountered in time-domain data matching (Bunks et al., 1995). The development of frequency-domain FWI began in the 1990s, spearheaded by Pratt (1999); Pratt and Shipp (1999). More recently, advancements in frequency-domain finite-difference modeling, as demonstrated by Operto et al. (2007), have proven the effectiveness of this method in addressing complex subsurface imaging challenges.

In FWI, forward modeling, namely solving the wave equation, is performed at each iteration of the optimization process. Various numerical approaches can be used for this purpose, including volumetric methods such as finite-element (Marfurt, 1984; Min et al., 2003), finite-difference (Virieux, 1986; Jo et al., 1996; Stekl and Pratt, 1998; Operto et al., 2007), finite-volume (Brossier et al., 2008), and pseudospectral methods (Danecek and Seriani, 2008). Among these, finite-difference methods are widely employed in FWI due to their computational efficiency and relative simplicity. In the frequency domain, finite-difference methods typically require solving large, sparse linear systems, which can be approached using either direct or iterative solvers. While direct solvers provide exact solutions and are often

more computationally efficient, their primary drawback compared to well-preconditioned iterative methods or time-domain formulations is the significant memory overhead associated with matrix factorization (e.g., L-U decomposition). This process requires substantial memory or disk space to store the lower and upper factors, making large-scale computations challenging (Operto et al., 2007; Virieux and Operto, 2009). In 3D applications, the storage and computational demands of frequency-domain finite-difference methods can thus become prohibitively expensive, limiting their practicality for large-scale inversions. Research has increasingly focused on developing preconditioners for iterative solvers in the 3D Helmholtz equation (e.g., Riyanti et al. (2006)). However, numerical stability remains a challenge, as the system matrix evolves throughout the inversion process, making effective preconditioning difficult to track and adapt. Advances in high-performance computing and computer hardware have significantly improved the feasibility of larger-scale 3D applications of the direct solvers. For example, the development of massively parallel direct solvers, such as the MUltifrontal Massively Parallel Sparse (MUMPS) solver (Amestoy et al., 2001) and the Intel Parallel Direct Sparse Solver (PARDISO) (Schenk and Gärtner, 2011), has enhanced the practicality of 3D frequency-domain FWI.

1.7 Thesis objectives

As discussed earlier, FWI has the potential to become a critical tool in the energy transitions. This thesis aims to advance FWI for subsurface imaging in these contexts through three primary research directions. The main objectives and contributions of the thesis are summarized as follows:

- Development of an model-source FWI framework.

  I develop an effective SWD-FWI methodology that incorporates the unique characteristics of drilling-generated seismic data. This includes modeling the drill-bit source signature, handling its non-impulsive nature, and designing an

inversion workflow that fully leverages SWD data for imaging while drilling.

- Uncertainty quantification for time-lapse FWI.

  I propose uncertainty quantification frameworks for time-lapse FWI using HMC and SVGD. The contribution includes a systematic approach to evaluating confidence in time-lapse changes and quantifying data-to-model uncertainty propagation that is relavent to acquisition geometries in dynamic subsurface monitoring.

- 3D acoustic frequency-domain FWI and the implementation of targeted nullspace shuttles.

  I extend the 2D frequency-dimain FWI to 3D computation that is tractable. In addition, I extend the targeted nullspace shuttle method to 3D acoustic frequency-domain FWI and demonstrate its application to time-lapse monitoring scenarios. This contribution improves the recoverability of subtle production- or injection-induced changes by systematically re-injecting nullspace energy in a controlled and computationally efficient manner.

Together, these contributions advance both the methodological and practical aspects of applying FWI to energy-transition problems, with particular emphasis on complex acquisition settings, uncertainty quantification, and time-lapse subsurface change detection.

The structure of the thesis is as follows:

In Chapter 2, I will introduce fundamental concepts utilized in this thesis, including modeling techniques for wave propagation (acoustic, elastic, viscoelastic, and 3D), drill-bit source simulation, inversion methodologies, essential terms from objective functions, numerical optimization techniques, Bayesian inversion, and the targeted nullspace shuttle.

In Chapter 3, I will detail the development of the SWD-FWI framework and demonstrate its effectiveness through synthetic examples. I will emphasize the advantages of integrating

SWD data with FWI, highlighting the improvements in subsurface imaging. Additionally, I will showcase how this integration enables the simultaneous retrieval of drill-bit source radiation patterns and locations while enhancing the subsurface image.

In Chapter 4, I will focus on uncertainty quantification in time-lapse FWI, specifically through the implementation of HMC and SVGD techniques. Using synthetic studies based on a simulated time-lapse scenario, I will demonstrate how these methods can effectively quantify uncertainties and their implications for inversion results. This framework will facilitate more reliable interpretations of plume migration over time by assessing the impact of uncertainties in subsurface properties.

In Chapter 5, I will showcase a 3D extension of the targeted nullspace shuttle within the framework of acoustic FWI. Discussions are based on a synthetic time-lapse experiment based on the Snowflake dataset from the CaMI project, which simulates CO_2 injection and enables assessment of the 3D targeted nullspace shuttle under various acquisition scenarios.

In Chapter 6, I will conclude the thesis with a summary of the novel contributions made throughout the research, addressing the challenges of integrating FWI techniques into the energy transition landscape. Additionally, I will outline potential future research directions that could further advance this field.

Chapter 2

Theories of frequency-domain full waveform inversion

Before detailing the novel contributions of this thesis, I will introduce several essential tools that are foundational to my work. In this chapter, I will cover key concepts, including the formulation of FWI as an optimization problem, the frequency-domain finite-difference concepts employed for wave propagation, the L_2 objective function I aim to minimize in FWI, and the corresponding derivatives, including the gradient and Hessian terms, the numerical optimization methods, the Bayesian theorem, and the targeted nullspace shuttle approach.

2.1 Viewing FWI as an optimization problem

FWI can be fundamentally viewed as an optimization problem (Lailly, 1983; Tarantola, 1984). The objective of FWI is to estimate a set of model parameters that can generate synthetic data that closely matches the observed data. This optimization problem is generally expressed as:

$$\hat{\mathbf{m}} = \underset{\mathbf{m}}{arg\,min} = \Phi\left(\mathbf{u}, \mathbf{m}, \mathbf{d}\right), \text{s.t. } C\left(\mathbf{u}, \mathbf{m}\right) = 0, \tag{2.1}$$

where $\hat{\mathbf{m}}$ represents the optimal model, Φ is the objective function quantifying the mismatch between the observed data $\mathbf{d}$ and the synthetic data generated by the current model $\mathbf{m}$, and $C\left(\mathbf{u}, \mathbf{m}\right) = 0$ denotes the constraint, which is typically the wave equation governing wave propagation physics. In this context, FWI strives to iteratively refine the initial model until an optimal solution is reached, where the synthetic data adequately matches the observed data within acceptable limits. The objective function value decreases as the model $\mathbf{m}$ produces synthetic data that better aligns with the observed data. Conversely, a greater mismatch between the two datasets results in a higher objective function value.

2.2 Forward modeling

In FWI, generating synthetic data to compare with real-world observations is essential. This process, called forward modeling, involves simulating data based on a proposed subsurface model and the acquisition setup. The choices made during forward modeling are critical, as they directly influence the accuracy of the inversion. One key decision is the complexity of the wave physics to be modeled, which dictates the range of data features that can be captured. For example, acoustic models fail to account for converted or shear wave arrivals. While incorporating more complex wave physics improves the fidelity of the data match, it also introduces additional parameters, increasing the risk of cross-talk during inversion. Additionally, more complex models raise computational demands, a factor that must be weighed for practical applications.

In this thesis, I employ several forward modeling approaches tailored to different application scenarios, with a focus on frequency-domain modeling due to its notable advantages. Specifically, it helps mitigate the cycle-skipping problem common in waveform matching (Bunks et al., 1995), facilitates the implementation of frequency-dependent effects such as attenuation and dispersion (Aki and Richards, 2002; Keating and Innanen, 2020), and typically requires a relatively small dataset. Although the computational cost of large-scale frequency-domain forward modeling remains a challenge, it is manageable within the experimental scope of this thesis when appropriate settings are selected.

2.2.1 Acoustic modeling

The frequency-domain acoustic wave propagation can be represented by:

$$\nabla^2 \mathbf{u}\left(\mathbf{r}, \mathbf{r}_s, \omega\right) + \frac{\omega^2}{\mathbf{c}^2\left(\mathbf{r}\right)} \mathbf{u}\left(\mathbf{r}, \mathbf{r}_s, \omega\right) = \mathbf{f}\left(\omega\right) \delta\left(\mathbf{r} - \mathbf{r}_s\right), \tag{2.2}$$

where $u\left(\mathbf{r}, \mathbf{r}_s, \omega\right)$ is the pressure field, ω is the angular frequency, $\mathbf{r} = (x, z)$ is an arbitrary location with 2D Cartesian coordinates, ∇^2 is the Laplacian operator, and c is the phase

velocity. On the right-hand side of equation 2.2, $\mathbf{f}(\omega)$ is the source signature of a certain frequency, and $\delta(\mathbf{r}-\mathbf{r}_s)$ is a Dirac function which is a spike at source location $\mathbf{r}_s$.

Replacing the velocity term with its reciprocal in equation 2.2, and slightly reorganizing the left-hand side leads to:

$$\left[(\nabla^2+\omega^2\mathbf{s}^2(\mathbf{r})\right]\mathbf{u}(\mathbf{r},\mathbf{r}_s,\omega)=\mathbf{f}(\omega)\,\delta(\mathbf{r}-\mathbf{r}_s), \tag{2.3}$$

where s is regarded as slowness. There is no analytical solution to equation 2.3 for an arbitrary medium. Generally, a finite-difference approach, where the spatial partial derivatives are replaced with finite-difference approximations is applied. Specifically, I use a centered second-order finite-difference approximation for the ∇ operator. This approach leads to a differential equation that approximates equation 2.3 at a grid point centered at the location with indices i in the x-direction and j in the z-direction. Here, the subscripts refer to the spatial grid indices in the x and z directions, while Δx and Δz represent the spatial increments in those respective directions. Applying this finite-difference scheme across the entire 2D spatial grid results in the following discretized form:

$$\omega^2 s_{i,j}u_{i,j}+\frac{u_{i-1,j}-2u_{i,j}+u_{i+1,j}}{\Delta x^2}+\frac{u_{i,j-1}-2u_{i,j}+u_{i,j+1}}{\Delta z^2}=f_{i,j}. \tag{2.4}$$

The general form of the above equation after applying it to the entire 2D space is:

$$\mathbf{S}(\mathbf{m},\omega)\,\mathbf{u}(\mathbf{m},\omega)-\mathbf{f}(\omega)=0, \tag{2.5}$$

where $\mathbf{u}$ represents the pressure wavefields at a specific frequency, with each column corresponding to a different seismic source and each row capturing the wavefield at various spatial locations. The matrix $\mathbf{f}$ contains the source terms using the same indexing scheme, while $\mathbf{S}$ operates as the matrix that applies the finite-difference approach outlined in equation 2.4. Known as the Helmholtz equation, this formulation enables a direct computation of $\mathbf{u}$. Even though it involves solving a large system of equations, the sparsity of $\mathbf{S}$ makes it computationally feasible, particularly for 2D models.

When approaching the boundaries of the spatial domain, applying equation 2.4 becomes problematic because values outside the grid are required. To overcome this, I employ a perfectly matched layer (PML) at the model edges following the method described by Zhang and Ballmann (1997), which simulates an infinite extension of the model by absorbing outgoing waves and preventing unwanted reflections. The same boundary treatment is applied at the top of the model to avoid surface reflections. Once the impedance matrix is set up, it exhibits a band-diagonal structure with some fringe elements. The values of the impedance matrix are complex numbers. Figure 2.1 illustrates an example of this impedance matrix, showing a grid of 50 by 50 points in the x- and z-directions, with a 3-point boundary layer. Note that the real part is displayed.

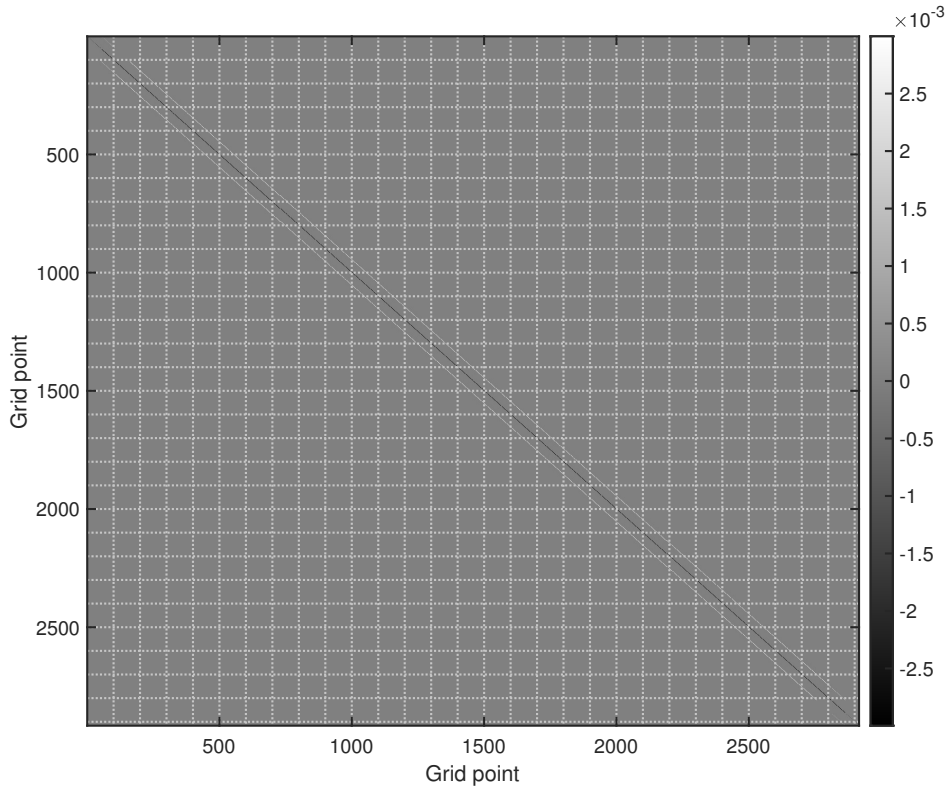

Figure 2.1: An example pattern of the acoustic impedance matrix from a 50×50 model with 3 grid cells as boundary width. The real part is displayed.

The size of the impedance matrix is $[(nx + 2 \times nPML) \times (nz + 2 \times nPML)]^2$, where

$nPML$ denotes PML boundary width. The non-zero entries depend on the specific grid implementation of the finite-difference discretization. In the finite-difference scheme I employ in this thesis, the non-zero entries consist of five diagonal components: one with $(nx + 2 \times nPML) \times (nz + 2 \times nPML)$, two with $(nx + 2 \times nPML) \times (nz + 2 \times nPML) - 2$, and two with $(nx + 2 \times nPML) \times (nz + 2 \times nPML) - 2 \times nz$ entries. These represent the main diagonal and the four sub-diagonals corresponding to the on-site, up/down, andleft-/right entries. Without boundary conditions, this matrix will be a symmetric pattern because of the reciprocity of Green's functions. However, in the example shown by Figure 2.1, this matrix is unsymmetric.

2.2.2 Viscoelastic modeling

In frequency-domain modeling, the complex structure of the impedance matrix provides a natural framework for incorporating frequency-dependent effects like attenuation (Toksöz et al., 1981), which plays a critical role in scenarios where amplitude information is essential. Viscoacoustic simulations, which account for attenuation, generally offer improvements over purely acoustic models by better matching measured seismic data. However, viscoacoustic modeling might not be sufficient in capturing the complete dynamics of the waveforms. To address this limitation, viscoelastic modeling is introduced as it more comprehensively accounts for the factors influencing seismic amplitudes.

The equations representing isotropic viscoelastic wave propagation in a 2D medium can be written as (Pratt, 1990; Brossier et al., 2010):

$$\begin{cases} \omega^2 \rho u_x + \frac{\partial}{\partial x}\left[\tilde{\lambda}\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z}\right) + 2\tilde{\mu}\frac{\partial u_x}{\partial x}\right] + \frac{\partial}{\partial z}\tilde{\mu}\left(\frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z}\right) + f_x = 0, \\ \omega^2 \rho u_z + \frac{\partial}{\partial z}\left[\tilde{\lambda}\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z}\right) + 2\tilde{\mu}\frac{\partial u_z}{\partial z}\right] + \frac{\partial}{\partial x}\tilde{\mu}\left(\frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z}\right) + f_z = 0, \end{cases} \tag{2.6}$$

where ρ is the density of the medium. The displacement components in the horizontal and vertical directions are denoted as u_x and u_z, while f_x and f_z refer to the source terms in

those respective directions. The parameters $\tilde{\lambda}$ and $\tilde{\mu}$ are the complex, frequency-dependent Lamé parameters, which account for attenuation and are linked to the quality factors Q_p and Q_s through the Kolsky-Futterman attenuation model (Kolsky, 1956; Futterman, 1962):

$$\begin{cases} \tilde{\lambda} = V_p^2 \left\{ 1 + \frac{1}{Q_p} \left[\frac{1}{\pi} \log \left(\frac{\omega}{\omega_0} \right) + \frac{i}{2} \right] \right\}^2 \rho - 2\tilde{\mu}, \\ \tilde{\mu} = V_s^2 \left\{ 1 + \frac{1}{Q_s} \left[\frac{1}{\pi} \log \left(\frac{\omega}{\omega_0} \right) + \frac{i}{2} \right] \right\}^2 \rho, \end{cases} \tag{2.7}$$

where ω_0 is the reference frequency. This system captures both the compressional and shear wave behaviors in a viscoelastic medium, providing a more detailed and accurate representation of wave propagation. By including shear deformations and frequency-dependent attenuation effects, viscoelastic modeling is especially valuable for scenarios that require accurate amplitude and phase information, making it a key tool in the analysis of seismic waveforms, especially those generated by sources with a more complex mechanism, which I will talk about in later sections.

Similar to the acoustic case, the viscoelastic wave equations are discretized using a finite-difference method, applying second-order centered finite-difference approximations for the spatial derivatives. The detailed formulation of these finite-difference equations follows the work of Pratt (1990). The key distinction in the viscoelastic case is the definition of the impedance matrix and source terms, where the dimensions of the impedance matrix are doubled to account for the two displacement components in x- and z-directions, and the first dimension of the source matrix is also doubled to represent both directional components, leading to a resultant wavefield with two directional components. Figure 2.2 provides an example of the viscoelastic impedance matrix, showing only the real part, based on a grid of 100 by 100 points in the x- and z-directions, with a 3-point boundary layer applied.

It is evident in Figure 2.2 that both dimensions of the impedance matrix are doubled because each interior grid point contributes two coupled equations to the overall system, leading to each element of the impedance matrix being represented as a 2×2 complex-

valued submatrix. The number of non-zero entries in this case is $9 \times (nx + 2 \times nPML) \times (nz + 2 \times nPML)$ because nine diagonal components are involved in the construction of the Helmholtz matrix. These components are similar to the acoustic case but include additional corner diagonal elements, reflecting the coupling between different displacement directions.

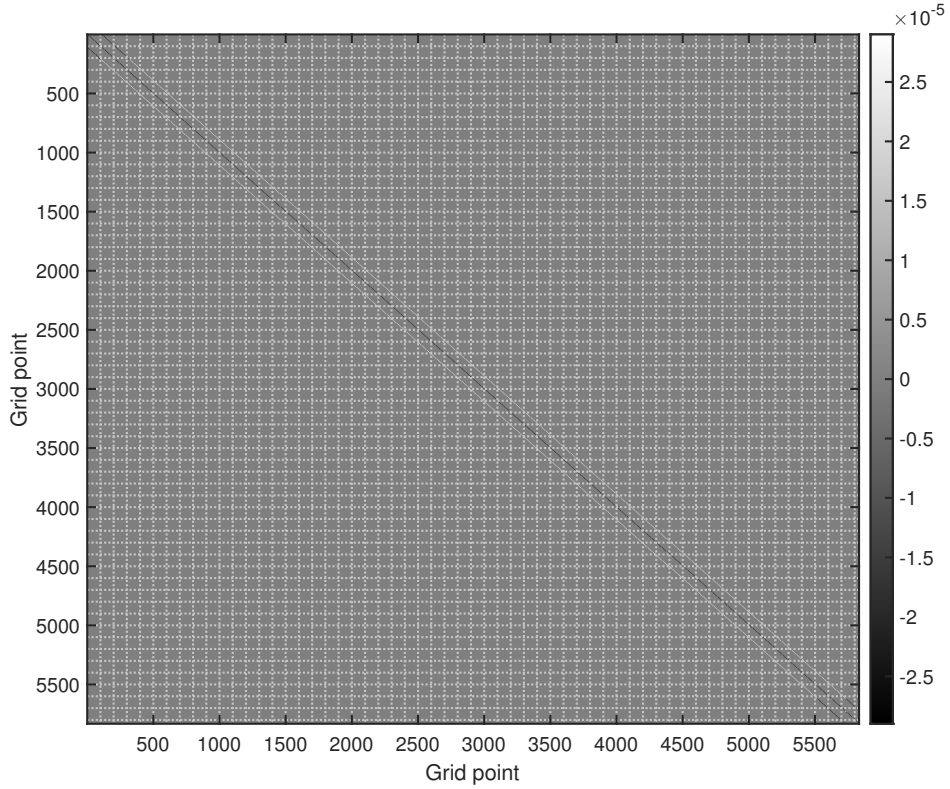

Figure 2.2: An example pattern of the viscoelastic impedance matrix from a 100×100 model with 3 grid cells as boundary width.

2.2.3 3D acoustic modeling

In the preceding sections, I have explained how the inclusion of more complex wave phenomena increases the dimensionality and complexity of the impedance matrix in frequency-domain modeling. This growing complexity poses considerable challenges in terms of computational resources, which I will discuss at the end of this chapter. Although a full 3D elastic modeling approach in the frequency domain would offer the most accurate representation of

wave propagation, it is prohibitively expensive for most practical applications. To conduct 3D experimental surveys within the scope of this thesis, a 3D acoustic modeling scheme is a more feasible choice.

In the frequency domain, simulating 3D acoustic wave propagation follows principles similar to the 2D case but with an additional extension in the y-direction. Here I discretize the Helmholtz equation using a second-order finite-difference approach on a staggered grid within a 3D Cartesian coordinate system as expressed in equation 2.8, where the indices $\frac{1}{2}$ indicate intermediate positions relative to the reference grid points i, j, k.

$$\begin{aligned}
\frac{\partial^2 u}{\partial x^2}|_{i,j,k} &= \frac{u_{i-\frac{1}{2},j,k} - 2u_{i,j,k} + u_{i+\frac{1}{2},j,k}}{\Delta x^2}, \\
\frac{\partial^2 u}{\partial y^2}|_{i,j,k} &= \frac{u_{i,j-\frac{1}{2},k} - 2u_{i,j,k} + u_{i,j+\frac{1}{2},k}}{\Delta y^2}, \\
\frac{\partial^2 u}{\partial z^2}|_{i,j,k} &= \frac{u_{i,j,k-\frac{1}{2}} - 2u_{i,j,k} + u_{i,j,k+\frac{1}{2}}}{\Delta z^2}.
\end{aligned} \tag{2.8}$$

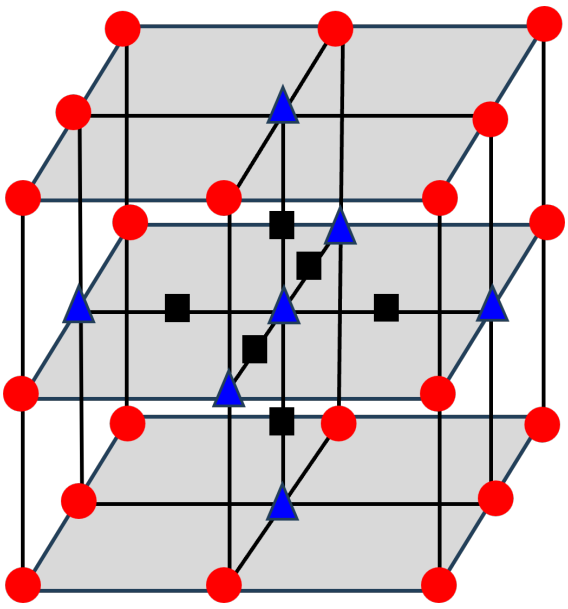

Figure 2.3: 3D finite difference stencils in staggered grid method. Red circles are pressure grid points, blue triangles are pressure grid points in the stencil, and black rectangles are the interpolated model parameters.

The staggered grid method places different variables at offset locations to improve accuracy and reduce numerical dispersion errors (Hustedt et al., 2004). While more advanced stencil definitions for 3D modeling exist (e.g., Stekl and Pratt (1998); Operto et al. (2007)), for this thesis, I employ a conventional approach by applying a 7-stencil on the classical Cartesian coordinate system. In such a setup, pressure values are positioned at the center of grid cells, and spatial derivatives are calculated using values at the cell edges, as illustrated in Figure 2.3.

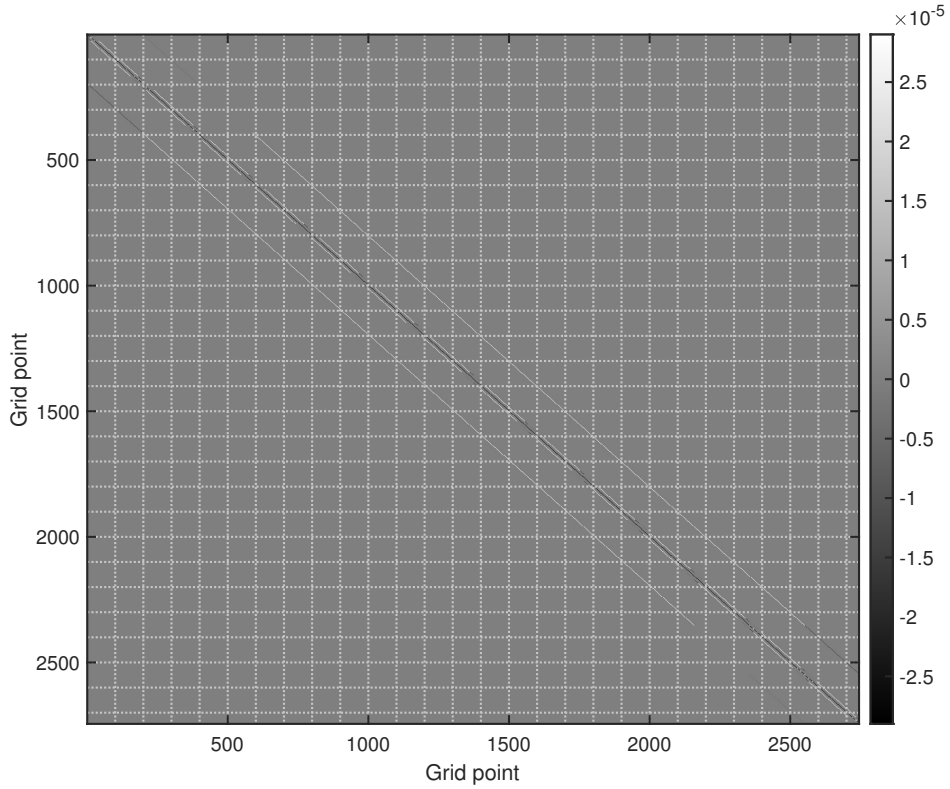

Figure 2.4: An example pattern of acoustic impedance matrix from a $10 \times 10 \times 10$ model with 2 grid cells as boundary width.

Again, the resulting matrix exhibits a band-diagonal structure with additional fringe elements. As shown in Figure 2.4, where a $10 \times 10 \times 10$ model with 2 grid cells as boundary width is adopted, each row and column of the matrix contains 6 non-zero entries. The coefficients are also of complex values, whereas in Figure 2.4 I show the real values.

2.2.4 Source modeling with moment tensors

To this point, various modeling techniques centered on constructing the Helmholtz matrix using model parameters are explored. However, when it comes to forward modeling for specific applications, such as drilling-related seismic modeling, accurately simulating the source mechanism is essential. The directional characteristics and complexities of seismic sources, such as those generated by drill-bit operations, require the incorporation of at least elastic wave equations within the modeling framework. Given the high computational demands of full 3D elastic wave propagation modeling, as previously mentioned, this thesis will focus on source modeling within a 2D space.

According to the representation theorem described by Aki and Richards (2002), which explains how elastic waves are generated by forces and influenced by medium responses, the sources can be represented as dipole forces along a fault whose dimension is much smaller than the dominant seismic wavelength and whose distance from the observation point is large enough for the far-field approximation to hold. Under these assumptions, it can be shown that

$$u_i(\mathbf{r},t) = \int_{-\infty}^{+\infty} \hat{M}_{kl}(\tau)\,\frac{\partial}{\partial \xi_l} G_{ik}(\mathbf{r},t;\boldsymbol{\xi},\tau)\,d\tau, \tag{2.9}$$

where $u_i = u_i(\mathbf{r},t)$ is the observed displacement field at position $\mathbf{r}$ at time t from a source at $\boldsymbol{\xi}$ and time τ, and $G_{ik} = G_{ik}(\mathbf{r},t;\boldsymbol{\xi},\tau)$ is Green's tensor, which is defined as the i th displacement component when a force is applied in the k th direction. M_{kl} is the moment tensor elements, which imply that a source is a summation of three force couples (see Figure 2.5). The above equation also shows that the recorded seismic waveforms can be regarded as a summation of the derivatives of Green's function weighted by entries in the moment tensor matrix:

$$M_0 \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix}, \tag{2.10}$$

where $\hat{M}_0$ is the seismic moment scalar that denotes the relative energy.

In this thesis, for modeling the source mechanisms, I use a finite-difference solution of the 2D isotropic viscoelastic wave equation in the frequency domain (equation 2.6) and 2D moment tensors (e.g., Aki and Richards, 2002; Vavrycuk, 2005; Tape and Tape, 2013) to represent source radiation. The source terms f_x and f_z in equation 2.6 are simulated with general 2D isotropic moment tensors, which have 3 independent components M_{11}, M_{12}, and M_{22} representing the derivative in the jth direction of the ith component of displacement. The moment tensor values lie between -1 to 1.

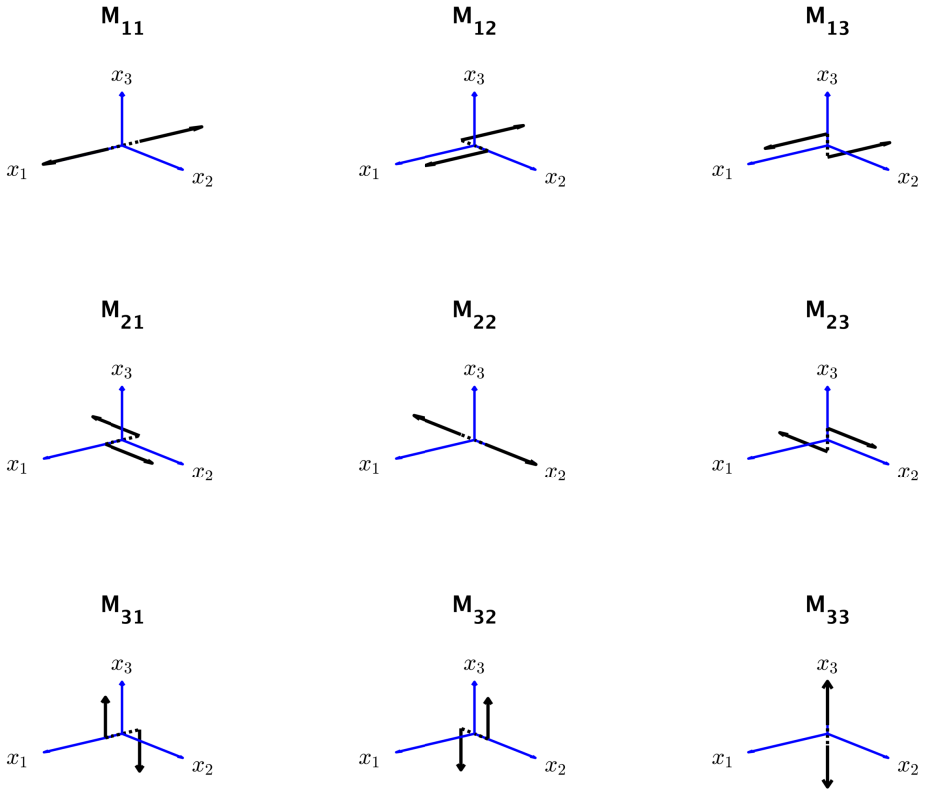

Figure 2.5: Moment tensor and the associated dipole forces. x_1, x_2, and x_3 denote the three directions on a Cartesian coordinate.

It is also necessary to approximate derivatives in the vicinity of the source in the finite-

difference model since source positions can deviate from the fine grid points. (Keating and Innanen, 2020). Figure 2.6 illustrates the first-order difference approach to the calculation of source-related derivatives. For a point source located midway between two finite-difference cell centers as Figure 2.6 (a), the finite difference approximation of $M_{11} = 1$ can be represented with differences formed between two adjacent cells by summing with weights $\frac{-1}{\Delta x}$ and $\frac{1}{\Delta x}$. If the source position is not equidistant between finite-difference cell centers, which is the general case shown by Figure 2.6 (b), I use three continuously-weighted members to approximate the first-order spatial derivatives. The finite-difference weights used for a source location between two finite-difference grid lines is a weighted average of the amplitudes used for a source at either of the bounding grid lines, as shown in Figure 2.6 (b).

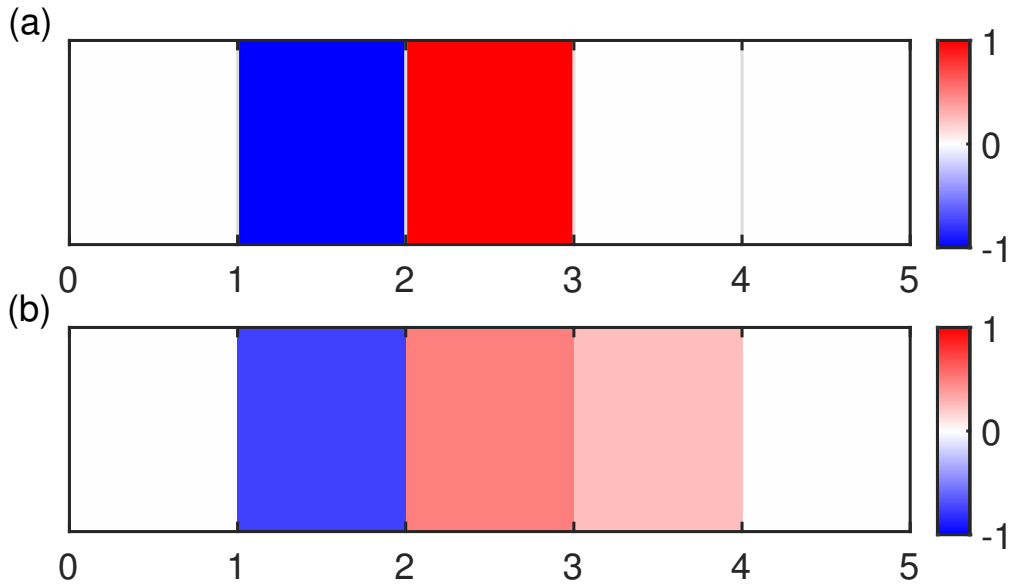

Figure 2.6: Scaled weights for approximating an example of $M_{11} = 1$ in one dimension. (a) weights for source location $x = 2$. (b) weights for source location $x = 2.25$.

Extending this concept to the 2D case, the finite difference approximation can be defined by weighting between intersections of nine grid cells in a small surrounding region. Specifically, for a point source shown by the red dot in Figure 2.7, I define stencil variables with indices of the four cells. The top-right cells $(0, -z, +x - z, +x)$ make the northeast stencil,

and the southeast, southwest, and northwest stencils can be represented by the bottom-right, bottom-left, and top-left four-cell sections following the same indexing rule.

With the stencils well defined, the finite difference approximation of moment tensors can be defined on each of the four stencils. In my formulation, f_x and f_z in equation 2.6 are defined by nine grid cells surrounding an arbitrary source position, considering a weighted summation of moment tensor components in each stencil based on the specific locations of a source within the center cell:

$$\begin{cases} f_x = \sum_{\mathbf{r}} \mathbf{w}(\mathbf{r}) \left[M_{11} w_{dx}(\mathbf{r}) + M_{12} w_{dz}(\mathbf{r}) \right], \\ f_z = \sum_{\mathbf{r}} \mathbf{w}(\mathbf{r}) \left[M_{12} w_{dx}(\mathbf{r}) + M_{22} w_{dz}(\mathbf{r}) \right], \end{cases} \tag{2.11}$$

where $\mathbf{w}(\mathbf{r})$ is the spatial weighting term considering the skewing of a point source to the middle of a grid cell. $\mathbf{r}$ is a position vector pointing in directions in a stencil which can be chosen from northeast-southwest-southeast-northwest. $w_{dx}(\mathbf{r}) = 1$ or -1 when $\mathbf{r}$ contains eastward (right) or westward (left) direction; $w_{dz}(\mathbf{r}) = 1$ or -1 when there is southward (down) or northward (up) in the stencils.

This treatment makes it possible to represent a point source in arbitrary locations precisely. Figure 2.8 demonstrates an approximation example based on the assumption of $M_{11} = 1$, while it is also applicable for a point source with a more general moment tensor combination. Small changes in the source locations simply correspond to a small re-weighting of the source term values. In such a case, the gradient of the objective function with respect to these variables is well defined, thus making the inversion more efficient. Additionally, the modeling of moment tensor sources provides the insight that wavefields generated by specific moment tensors can be simulated, if the mechanism is acquired, or can be approximated.

2.2.5 Solving the frequency-domain wave equations

Solving the frequency-domain wave equations is essentially inverting the impedance matrix on the left-hand side of equation 2.5 by a solver to get the wavefield $\mathbf{u}$. This inversion

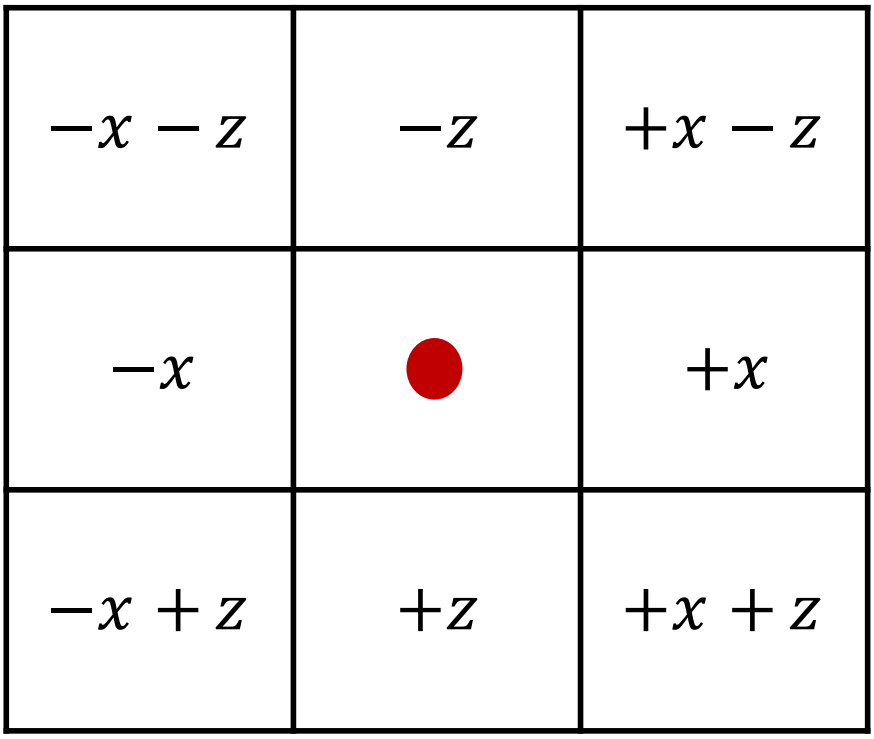

Figure 2.7: Stencil variables definition. The red dot in the center cell is the source location.

can be accomplished using either direct or iterative solvers. Direct solvers, such as those based on lower-upper (LU) factorization, are effective for 2D forward problems (Jo et al., 1996; Hustedt et al., 2004). However, their time and memory complexities, along with their limited scalability on large-scale distributed memory systems, render them discouraging for large-scale 3D problems (Virieux and Operto, 2009). Figure 2.9 illustrates the amount of LU factors as a function of the grid cell number for reference.

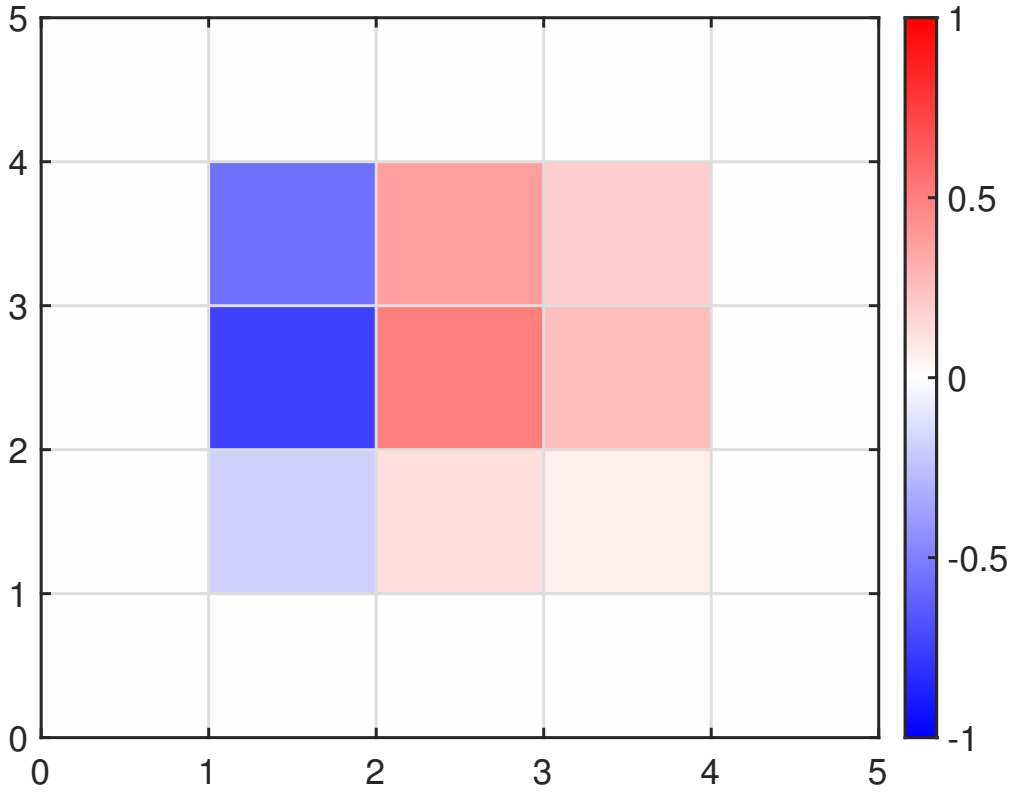

Figure 2.8: Scaled weights for approximating an example of $M_{11} = 1$ in two dimension with source location $x = 2.25$ and $z = 2.25$.

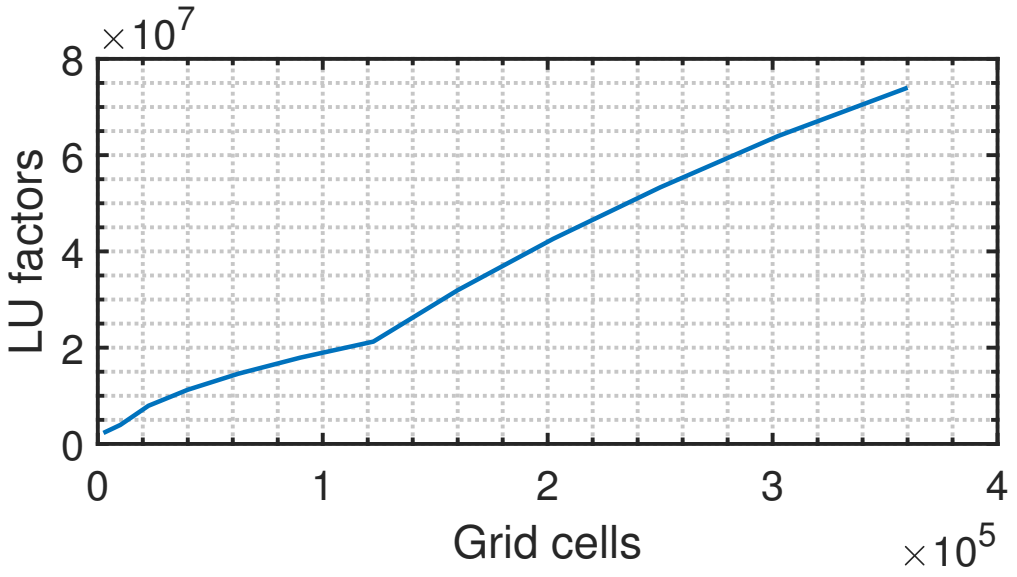

Figure 2.9: Amount of LU factors as a function of the grid cell number.

In contrast, iterative solvers offer a viable alternative for solving the time-harmonic wave equation (Riyanti et al., 2006; Plessix, 2009). These solvers typically employ Krylov subspace methods and are preconditioned using the dampened time-harmonic wave equation, which is solved with one multigrid cycle (Saad, 2003). Iterative methods are advantageous due to their lower memory requirements; however, their effectiveness is limited by the challenge of designing an efficient preconditioner for the indefinite impedance matrix. The extension of these methods to elastic wave equations remains an area for future research. Additionally, the time complexity of iterative solvers scales linearly with the number of sources, in contrast to direct solvers. A hybrid direct-iterative approach, utilizing domain decomposition methods (Saad, 2003), represents a middle ground between direct and iterative methods. However, this approach requires a more complex design and implementation, which is currently beyond the scope of this thesis.

After testing out cases between iterative and direct solvers, I decide to use a direct solver that is potentially more efficient and stable, as iterative solvers often require good preconditioners that are dependent on model patterns. Additionally, by using direct solvers, solving for multiple columns becomes more efficient as the computational cost scales sublinearly, approximately proportional to the log of the number of sources. The solver I use in the scope of this thesis is the Intel PARDISO solver (Schenk and Gärtner, 2011) incorporated with an Eigen (Guennebaud et al., 2010) environment. For solving an experimental-scale problem, such an incorporation is tested to be more computationally efficient.

Since numerical approximations are employed in the forward modeling process, ensuring numerical stability is crucial, as it directly affects both the accuracy and even the overall success of the simulation. Poor stability can lead to errors such as numerical dispersion, which distorts the wavefield and diminishes the precision of the results. To minimize numerical dispersion, careful attention must be paid to selecting appropriate grid spacing. In my

forward modeling approach, the criterion I follow for grid spacing is:

$$\Delta h < \frac{c_{min}}{6max(frequency)}, \tag{2.12}$$

where Δh is the grid interval in an arbitrary dimension, and c_{min} is the minimum wave velocity.

2.3 Inverse problem of FWI

In this section, I will outline the fundamental components of the inverse problem in FWI, focusing on the L_2 objective function, which is widely used and also adopted in this thesis. Additionally, I will discuss the gradient term, which is derived from the objective function and is essential for guiding the inversion process, as well as the Hessian term, which represents the second-order derivative of the objective function with respect to the inversion variables and plays a critical role in enhancing the efficiency of the inversion.

2.3.1 L_2 objective function

The objective function, or cost function in FWI, is a critical metric to quantify the difference between modeled and observed datasets, as introduced earlier around equation 2.1. It is pivotal in guiding the inversion process by driving model updates to minimize this mismatch. Numerous studies, such as those by Engquist and Froese (2014), Chi et al. (2015), and Yang and Engquist (2018), have explored alternative objective functions to address specific challenges in FWI, beyond the conventional formulation proposed by Tarantola (1984). These alternatives aim to improve robustness, account for noise, or mitigate cycle-skipping. However, each introduces its own complexities and trade-offs, often requiring careful customization for particular scenarios.

In this thesis, I adopt the original objective function because it captures both the amplitude and phase information in the seismic data, providing a comprehensive metric for model

evaluation. Given that the focus of this thesis lies elsewhere, such as in acquisition design and strategy assessment, designing specialized objective functions is not within the scope of this thesis. The classical objective function, widely used in FWI research, offers a reliable and well-established framework for my applications. It is expressed as:

$$\Phi = \sum_{j=1}^{N_\omega} \sum_{k=1}^{N_s} \frac{1}{2} \|\mathbf{R}\mathbf{u}_{j,k} - \mathbf{d}_{j,k}\|_2^2, \tag{2.13}$$

where s denotes the source index, and $\mathbf{R}$ is the sampling matrix representing the measurements of receivers.

Another part of the objective function may focus on the prior model fitting, as explored in Kamath et al. (2017); Esser et al. (2018); Wang et al. (2024). However, this part is highly case-dependent and can be modified to meet different models and objectives. In this thesis, I focus on the general FWI method rather than a specific application scenario, so this part will not be discussed in more detail.

Another component of the objective function often addresses the incorporation of prior model information, as discussed in works like Kamath et al. (2017), Esser et al. (2018), and Wang et al. (2024). This regularization term encourages the inversion to stay consistent with prior knowledge, which can help stabilize the solution and guide the model updates toward realistic results. However, the design of this term is highly case-specific, varying based on the nature of the problem, available prior models, and intended goals. Since this thesis focuses on evaluating the general FWI framework rather than tailoring it to a particular scenario, a detailed exploration of prior-model fitting is beyond the scope. Instead, I emphasize core elements of the FWI methodology, while leaving the adaptation of prior model regularization for future studies or context-specific implementations.

2.3.2 Gradient

The objective function of the FWI approach in this thesis is then:

$$\Phi(\mathbf{m}) = \sum_{j=1}^{N_\omega} \sum_{k=1}^{N_s} \frac{1}{2} \|\mathbf{R}\mathbf{u}_{j,k} - \mathbf{d}_{j,k}\|_2^2, \text{ s.t. } \mathbf{S}(\mathbf{m})\mathbf{u} = \mathbf{f}, \tag{2.14}$$

In the gradient derivation, I assume that the wavefield only has one source and frequency component for simplicity. In practice, contributions from multiple sources or frequencies can be obtained by summing their respective gradients.

The adjoint state method (e.g., Plessix, 2006) is used to determine the gradient through a combination of forward and backward propagated wavefields, the latter of which has the residual wavefield as the source. The Lagrangian of the current problem is

$$L(\mathbf{m}) = \frac{1}{2} \|\mathbf{R}\mathbf{u} - \mathbf{d}\|_2^2 + \langle \mathbf{S}(\mathbf{m})\,\mathbf{u} - \mathbf{f}, \lambda \rangle, \tag{2.15}$$

where λ is a vector of Lagrange multipliers, and $\langle \cdot, \cdot \rangle$ represents the inner product.

Assume that a wavefield $\overline{\mathbf{u}}$ satisfies $L(\overline{\mathbf{u}}) = \Phi$. The derivative of Φ with respect to the model parameter is then:

$$\frac{d\Phi}{d\mathbf{m}} = \frac{dL(\overline{\mathbf{u}})}{d\mathbf{m}} = \frac{\partial L}{\partial \overline{\mathbf{u}}} \frac{\partial \overline{\mathbf{u}}}{\partial \mathbf{m}} + \frac{\partial L}{\partial \mathbf{m}}. \tag{2.16}$$

By choosing $\overline{\lambda}$ such that $\frac{\partial L}{\partial \overline{\mathbf{u}}} = 0$, the otherwise expensive calculation of the Jacobian matrix $\frac{\partial \overline{\mathbf{u}}}{\partial \mathbf{m}}$, can be avoided. The $\overline{\lambda}$ is determined through backpropagation of data residuals, after which it can be derived that:

$$\frac{\partial L}{\partial \overline{\mathbf{u}}} = \mathbf{R}^T (\mathbf{R}\overline{\mathbf{u}} - \mathbf{d}) + \mathbf{S}^\dagger \overline{\lambda} = 0, \tag{2.17}$$

where $\dagger$ denotes the conjugate transpose. After calculating $\overline{\lambda}$ with the above equation, equation 2.18 can be reduced to:

$$\frac{dL}{d\mathbf{m}} = \frac{\partial L}{\partial \mathbf{m}} = \left\langle \frac{\partial \mathbf{S}}{\partial \mathbf{m}} \overline{\mathbf{u}}, \overline{\lambda} \right\rangle. \tag{2.18}$$

2.3.3 Gauss-Newton Hessian

The gradient in FWI is often poorly scaled due to geometric spreading and can be further distorted by spurious correlations caused by finite-frequency effects and doubly-scattered energy (Pratt et al., 1998). This makes it challenging to accurately update the model with just the gradient. The Hessian operator, defined as the second-order partial derivative of the objective function with respect to the model parameters, provides critical information about the curvature of the misfit surface, thereby offering more precise directions for model updates (Pratt et al., 1998).

Incorporating the inverse Hessian to precondition the gradient can significantly improve the scaling of the model update, leading to more efficient convergence and ensuring faster quadratic convergence rates. However, computing the exact Hessian matrix is computationally prohibitive, especially for large-scale problems, as it involves an enormous number of second-order derivatives. To address this limitation, approximate Hessian formulations, such as the Gauss-Newton Hessian, are often employed. The Gauss-Newton approximation ignores the second-order terms that arise from nonlinear interactions between wavefields, focusing instead on the terms related to the data residuals. This reduces the computational burden while still providing significant improvements over pure gradient-based updates. The resulting Gauss-Newton Hessian retains important curvature information, making it a practical choice for large-scale FWI applications.

In the Gauss-Newton approximation, it is assumed that the residual of the objective function, namely $\mathbf{Ru} - \mathbf{d}$, is sufficiently small. This approximation simplifies the Hessian by neglecting higher-order terms, resulting in a positive-definite matrix. A positive-definite Hessian ensures that the updates are well-scaled, promoting stable and efficient optimization. As suggested by Métivier et al. (2013), this property can accelerate the convergence of FWI. Applying the second-order derivative to equation 2.14 yields the following expression:

$$\frac{\partial^2 \Phi}{\partial \mathbf{m}^2} = \frac{\partial^2 \left(\mathbf{Ru} - \mathbf{d}\right)^T}{\partial \mathbf{m}^2} \left(\mathbf{Ru} - \mathbf{d}\right) + \frac{\partial \left(\mathbf{Ru} - \mathbf{d}\right)^T}{\partial \mathbf{m}} \frac{\partial \left(\mathbf{Ru} - \mathbf{d}\right)}{\partial \mathbf{m}}. \tag{2.19}$$

The Gauss-Newton approximation assumes the first term in equation 2.19 is small enough to be neglected. Thus, the above equation can be represented as:

$$\mathbf{H}_{GN} = \mathbf{J}^{\dagger}\mathbf{R}^{T}\mathbf{R}\mathbf{J}, \tag{2.20}$$

where $\mathbf{J} = \frac{\partial \mathbf{u}}{\partial \mathbf{m}}$ represents the Jacobian matrix, which describes the sensitivity of the wavefield $\mathbf{u}$ to changes in the model parameters $\mathbf{m}$. Similar to the gradient computation, the Jacobian can be efficiently evaluated using the adjoint state method, avoiding the need to explicitly form this matrix and thereby reducing computational costs.

Again making use of the adjoint state method, the Lagrangian can be set as:

$$\tilde{L} = \langle \mathbf{u}(\mathbf{m}), \mathbf{w} \rangle + \langle \mathbf{S}(\mathbf{m})\,\mathbf{u} - \mathbf{f}, \mu \rangle, \tag{2.21}$$

where $\mathbf{w} = \mathbf{R}^{T}\mathbf{R}\mathbf{J}\mathbf{v}$, and $\mathbf{v}$ is the vector to be multiplied by the Hessian. The Lagrange multiplier $\overline{\mu}$ represents the adjoint field, and satisfies:

$$\mathbf{S}^{\dagger}\overline{\mu} + \mathbf{w} = 0. \tag{2.22}$$

Directly solution for $\mathbf{w}$ still involves computing the Jacobian matrix J, but the product of J with the vector $\mathbf{v}$ is possible via the derivative of the forward problem with respect to variables m_p multiplied by vector elements v_{m_p} (Métivier et al., 2013; Keating and Innanen, 2020):

$$\frac{\partial (\mathbf{S}\mathbf{u} - \mathbf{f})\, v_{m_p}}{\partial m_p} = 0. \tag{2.23}$$

This allows us to acquire $\mathbf{w}$ and thus derive the Lagrange multiplier $\overline{\mu}$ in equation (2.22). The Gauss-Newton Hessian vector product $\mathbf{H}_{GN}\mathbf{v}$ is then $\frac{\partial \tilde{L}}{\partial \mathbf{m}}$.

As discussed in Section 2.2.5, the primary computational burden in evaluating the derivatives arises from solving wave-propagation equations. In the context of the FWI framework I present, this process involves solving linear systems to obtain the wavefield. Given

the complexity of wave propagation, these computations dominate the total cost, rendering other derivative-related operations relatively insignificant. For gradient evaluation, two wave-propagation problems must be solved: one for computing the forward wavefield and another for the adjoint field. Similarly, the Gauss-Newton Hessian-vector product also requires two such propagations.

2.4 Numerical optimization

The previously derived components, such as the gradient and the Gauss-Newton Hessian, are integrated into an optimization framework to iteratively update the model. In this section, I will provide a brief overview of key optimization strategies commonly employed in FWI and highlight the specific approaches used in this thesis.

2.4.1 Steepest descent

The steepest descent (SD) or gradient descent optimization strategy was the first proposed optimizer for the FWI problem (Tarantola, 1984). In this optimization strategy, the objective function is supposed to decrease at each iteration by updating the model in directions that are anti-parallel to the gradient with respect to the model parameters. In an arbitrary iteration k, such a process can be expressed as:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha \frac{\partial \Phi\left(\mathbf{m}_k\right)}{\partial \mathbf{m}_k}, \tag{2.24}$$

Here, $\mathbf{m}_k$ represents the current model, $\mathbf{m}_{k+1}$ is the updated model, and α is the step size that controls the magnitude of each update—details of which will be discussed in later sections. The SD approach offers simplicity in implementation and requires only a few evaluations of the objective function or its derivatives per iteration, making it computationally attractive. However, as noted earlier, the gradients in FWI are often poorly scaled, leading to slow convergence towards an optimal solution (Métivier et al., 2013). Moreover, the SD

method struggles with cross-talk between parameters in multiparameter FWI, limiting its effectiveness in more complex inversion scenarios.

2.4.2 L-BFGS optimization

The steepest descent method assumes that the second-order term in the misfit function, such as the Hessian or its approximation, is negligible. Newton methods offer an effective approach to include the second-order information of the objective function by:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha \mathbf{H}^{-1}\left(\mathbf{m}_k\right) \frac{\partial \Phi\left(\mathbf{m}_k\right)}{\partial \mathbf{m}_k}, \tag{2.25}$$

This approach effectively incorporates second-derivative information, which can dramatically improve the ability of the inversion to anticipate how changing one variable will alter the derivative of another. This allows for substantial cross-talk reduction (Innanen, 2014).

Computing the exact Hessian in large-scale or 3D FWI problems is often prohibitively expensive. To address this, descent directions that incorporate curvature information are typically obtained using an approximate Hessian. Quasi-Newton methods, such as the Broyden, Fletcher, Goldfarb, and Shanno (BFGS) optimization strategy (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970), provide approximations to the Hessian and its inverse (Nocedal and Wright, 2006). The BFGS method involves explicitly updating the approximation of the Hessian or its inverse at each iteration during the inversion process. I will briefly discuss this method in this section, while more details can be found in Nocedal and Wright (2006). I start rewriting equation 2.25 with an approximation of the hessian matrix:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha \mathbf{B}\left(\mathbf{m}_k\right) \frac{\partial \Phi\left(\mathbf{m}_k\right)}{\partial \mathbf{m}_k}, \tag{2.26}$$

where $\mathbf{B}$ is the inverse of the approximated Hessian matrix.

Consider two iterations $k-1$ and k, $\mathbf{B}_k$ should be consistent with the change in the gradient of the objective function, such that:

$$\mathbf{B}_k^{-1}\left(\mathbf{m}_k - \mathbf{m}_{k-1}\right) = \frac{\partial \Phi\left(\mathbf{m}_k\right)}{\partial \mathbf{m}_k} - \frac{\partial \Phi\left(\mathbf{m}_{k-1}\right)}{\partial \mathbf{m}_{k-1}}. \tag{2.27}$$

The equation above forms the core of the BFGS method, indicating that the inverse of the approximate Hessian can be iteratively estimated using information from both the current and previous gradient terms. Additionally, the estimated inverse Hessians from consecutive iterations must remain sufficiently similar to ensure stable updates. This closeness helps maintain convergence and prevents abrupt model changes during each step.

Applying the BFGS method efficiently in FWI is challenging due to memory limitations, as storing the full Hessian approximation, $\mathbf{B}_k$, becomes impractical. For reference, the size of $\mathbf{B}_k$ in 2D FWI is $(N_x \times N_y)$ by $(N_x \times N_y)$, and in 3D FWI it grows to $(N_x \times N_y \times N_z)$ by $(N_x \times N_y \times N_z)$, where N_x, N_y and N_z are grid point numbers in each dimension. Moreover, $\mathbf{B}_k$ is a complex-valued matrix, further increasing memory demands.

An effective alternative is the limited-memory BFGS (L-BFGS) approach, which bypasses the need to store or compute $\mathbf{B}_k$ explicitly (Nocedal and Wright, 2006). Instead, L-BFGS approximates the Hessian indirectly by retaining a limited number of past model updates and gradient vectors. These stored vectors are used to calculate the products of $\mathbf{B}_k$ with other vectors as needed, without forming the entire matrix. This feature makes L-BFGS specifically well-suited for FWI, as such vector products can be efficiently calculated by backpropagations, as discussed in Section 2.3.3. Thus, the L-BFGS in FWI drastically reduces memory usage while still maintaining efficient convergence properties.

The initial guess in the L-BFGS method is suggested by Nocedal and Wright (2006), that the model and gradient vectors in the last two iterations can be utilized, such that:

$$\mathbf{B}_0^k = \mathbf{I}\frac{\Delta \mathbf{m}_k^T \Delta \mathbf{g}_k}{\Delta \mathbf{g}_k^T \Delta \mathbf{g}_k} = \mathbf{I}\frac{\left(\mathbf{m}_k - \mathbf{m}_{k-1}\right)^T \left(\mathbf{g}_k - \mathbf{g}_{k-1}\right)}{\left(\mathbf{g}_k - \mathbf{g}_{k-1}\right)^T \left(\mathbf{g}_k - \mathbf{g}_{k-1}\right)}. \tag{2.28}$$

A key challenge with the L-BFGS method lies in storing intermediate vectors. However,

based on my tests, using practical iteration counts between 5 and 20 introduces minimal overhead in both runtime and storage, with the algorithm performing similarly to the full BFGS method. While Nocedal and Wright (2006) acknowledge that L-BFGS may converge more slowly than BFGS, research by Brossier et al. (2009, 2010) shows that L-BFGS delivers superior convergence rates and model resolution compared to the conjugate gradient method in FWI applications. In addition, research from Keating and Innanen (2020) also demonstrates that the L-BFGS method is very helpful in suppressing the cross-talk issues, which is a stubborn challenge in FWI. Its efficient balance between computational cost and scalability makes it well-suited for large-scale 3D FWI, where managing memory usage and processing time is crucial.

2.4.3 Truncated Gauss-Newton method

An alternative to the L-BFGS approach employed in this thesis is the truncated Gauss-Newton (TGN) method. Rather than directly estimating the full Hessian, the TGN method iteratively solves the Newton system over a limited number of iterations. In such a way, the TGN approach allows for a chosen level of approximation to the Hessian in each updating step. In each iteration k, the update direction is determined by:

$$\mathbf{D}_k = \underset{\mathbf{D}}{arg\,min} \frac{1}{2}\mathbf{D}^T\mathbf{H}\left(\mathbf{m}_k\right)\mathbf{D} + \mathbf{D}^T\mathbf{g}\left(\mathbf{m}_k\right), \tag{2.29}$$

where $\mathbf{D}$ is the update direction, and $\mathbf{g}$ represents the gradient of the objective function with respect to the model parameters ($\frac{\partial \Phi(\mathbf{m})}{\partial \mathbf{m}}$). The optimization process halts when either a predefined maximum number of iterations is reached or the solution falls below a specified tolerance threshold—justifying the "truncated" nature of the method.

The model is then updated in a Newton update fashion with the $\mathbf{D}$ in the above equation:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha\mathbf{D}_k, \tag{2.30}$$

Taking derivatives with respect to α from the Taylor expansion of the objective function gives an explicit solution of the step size in the TGN method:

$$\alpha = -\frac{\left[\mathbf{H}\left(\mathbf{m}_k\right)\mathbf{D} + \mathbf{g}\left(\mathbf{m}_k\right)\right]^T \mathbf{D}}{\mathbf{D}^T \mathbf{H}\left(\mathbf{m}_k\right)\mathbf{D}} \tag{2.31}$$

Equation 2.29 can be solved by simply taking first derivatives with respect to $\mathbf{D}$, leading to a linear system:

$$\mathbf{H}\left(\mathbf{m}_k\right)\mathbf{D} = -\mathbf{g}\left(\mathbf{m}_k\right), \tag{2.32}$$

which can be solved by the conjugate gradient method (e.g., Métivier et al. (2013)) or other iterative linear solvers. In this thesis, I use the L-BFGS method, which typically provides faster convergence, and I employ the well-tested implementation from Keating and Innanen (2020).

2.4.4 Line-search

In the non-TGN methods employed in this thesis, the update step size is determined after obtaining the update direction. For this purpose, I utilize a line search method, which is relatively efficient as outlined in detail by Nocedal and Wright (2006). Here, I will provide a brief overview of the approach.

In the line-search approach, refer to the update step size as α, the following optimization problem in each model updating process is solved:

$$\alpha_k = \underset{\alpha}{min}\Phi\left(\mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k\right). \tag{2.33}$$

However, in practice, the solution is often obtained within the required accuracy because it is too costly to solve this problem exactly. In this work, I use the following conditions:

$$c_1\Phi\left(\mathbf{m}_k\right) \leq \Phi\left(\mathbf{m}_k + \alpha_k\Delta\mathbf{m}_k\right) \leq \Phi\left(\mathbf{m}_k\right) + c_2\alpha_k\Delta\mathbf{m}_k^T\mathbf{g}\left(\mathbf{m}_k\right), \tag{2.34}$$

where c_1 and c_2 are constant factors which satisfy $0 \leq c_2 \leq c1 \leq 1$. In this thesis, c_1 is chosen to be 0.9, and c_2 is 0.001. The first inequality in equation 2.34 is the Armijo

condition constrains the increment of model parameters is not too small (Armijo, 1966), while the second one is known as the weak Wolfe condition that guarantees the objective function can be decreased sufficiently (Wolfe, 1969).

2.5 Bayesian theorem

In the Bayesian framework, the solution of inverse problems is a probability density function of plausible models that reasonably represent the observations. Such probability density is called the posterior distribution $P(\mathbf{m}|\mathbf{d})$. According to Bayes' theorem:

$$P(\mathbf{m}|\mathbf{d}) = \frac{P(\mathbf{m})\,P(\mathbf{d}|\mathbf{m})}{P(\mathbf{d})}, \tag{2.35}$$

where $P(\mathbf{m})$ is the prior probability distribution of model parameters, $P(\mathbf{d})$ is the probability distribution of the observed data, which is often regarded as a normalization factor and can be omitted in most cases. Additionally, $P(\mathbf{d}|\mathbf{m})$ in equation 2.35 is called the likelihood function that represents the conditional probability of a dataset generated by a model following some physics rules, for example, acoustic/elastic wave equations in FWI problems. The likelihood and the prior can be shown by (Tarantola, 2004):

$$P(\mathbf{d}|\mathbf{m}) \propto exp\left[-\frac{1}{2}\left(\mathbf{d}_{syn}-\mathbf{d}_{obs}\right)^T \mathbf{C}_D^{-1}\left(\mathbf{d}_{syn}-\mathbf{d}_{obs}\right)\right], \tag{2.36}$$

$$P(\mathbf{m}) \propto exp\left(-\frac{1}{2}\mathbf{m}^T\mathbf{C}_M^{-1}\mathbf{m}\right), \tag{2.37}$$

where $\mathbf{d}_{syn}$ is the synthetic data, $\mathbf{C}_D^{-1}$ in equation 2.36 is the inverse covariance matrix of the observed data, and $\mathbf{C}_M^{-1}$ in equation 2.37 is the inverse covariance matrix of the model.

Conventionally, the posterior distribution outlined in equation 2.35 is typically estimated through the exploitation of numerous representative models generated via MC sampling processes. Notable instances of such approaches in geophysical inverse problems can be observed in recent studies by Dosso et al. (2014); Gebraad et al. (2020); Kotsi et al. (2020); Fichtner et al. (2021); Fu and Innanen (2022). However, the reliance on frequent sampling

attempts may lead to inefficiencies and excessive utilization of computational resources. Variational inference offers an alternative approach to approximating the aforementioned posterior distribution by formulating the sampling process as an optimization problem that is potentially more efficient than the sampling approaches. In this section, I will provide a brief overview of two key approaches used in this thesis for uncertainty quantification: HMC from the MC family and SVGD from the variational inference methods. These methods represent significant components of my research and contribute to the overall understanding of uncertainty quantification in the context of FWI.

2.5.1 Hamiltonian Monte Carlo

HMC takes inspiration from the Hamiltonian dynamics (Hamilton, 1834), which can be conceptually understood by fictitiously visualizing a frictionless particle moving along a 2D surface with varying heights. In this scenario, the state of this system is described by the position of this particle $\mathbf{q}$ (2D vector) and its generalized momentum $\mathbf{p}$ (also a 2D vector). While moving, the potential energy of this particle can be represented by $U(\mathbf{q})$, and its kinetic energy is given by $\mathbf{p}^T\mathbf{M}_{mass}^{-1}\mathbf{p}/2$, where $\mathbf{M}_{mass}$ is the mass matrix of this particle. A generalization of this matrix is making the off-diagonal entries zeros while keeping the diagonal members equal to some preset values, but there are more constrained settings in some more complex problems (Fichtner et al., 2021). In any situation, the mass matrix is positive definite. When the surface is flat, the particle moves at a constant velocity equal to $\mathbf{M}_{mass}^{-1}\mathbf{p}$. However, if the surface inclines, the momentum allows the particle to continue, resulting in decreased kinetic energy and increased potential energy. Eventually, when the kinetic energy reaches zero, this particle will slide back down the slope, increasing kinetic energy and decreasing potential energy. During the whole process, this particle is governed by a "gravitational" force that is parallel to $-\nabla U(\mathbf{q})$, which is regarded as the steepest descent direction. Such mechanics can be described by a Hamiltonian equation $H(\mathbf{p}, \mathbf{q})$,

such that

$$H(\mathbf{p},\mathbf{q}) = U(\mathbf{q}) + K(\mathbf{p}). \tag{2.38}$$

This equation describes the total energy of a mechanic system. It holds for any coordinate systems as long as their link to a Cartesian set of coordinates is independent of time or velocity (Morin, 2008).

The partial derivatives of the Hamiltonian determine how $\mathbf{q}$ and $\mathbf{p}$ change over time, t, according to Hamilton's equations:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \tag{2.39}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}, \tag{2.40}$$

where $i = 1, 2, ..., n$ denotes the index in the n-dimensional vector.

The derivatives can be further represented as:

$$\frac{dq_i}{dt} = \left[\mathbf{M}_{mass}^{-1}\mathbf{p}\right]_i, \tag{2.41}$$

$$\frac{dp_i}{dt} = -\frac{\partial U}{\partial q_i}. \tag{2.42}$$

The fundamental idea behind HMC involves sampling from an auxiliary distribution in a phase space with twice the original dimensions. The auxiliary distribution is known as the canonical distribution (Davey, 2009), and this extended phase space is denoted as $(\mathbf{p},\mathbf{m})$, where $\mathbf{p}$ represents momentum variables and $\mathbf{m}$ represents the variables of interest (e.g., model parameters). For convenience, in the following discussions, I will replace the vector $\mathbf{q}$ in equations 2.38, 2.39, and 2.41 by the model vector $\mathbf{m}$ as I call it in forming the FWI problem. Additionally, artificial momentum variables are usually introduced to align with the quantity of position variables. In a more intuitive sense, the concept of HMC views a model $\mathbf{m}$ as analogous to a particle in mechanics. This particle moves from its present position to a new position following a Hamiltonian trajectory (Brooks et al., 2011; Betancourt, 2017). The

geometry of the trajectory is controlled by the misfit that is interpreted as potential energy U, as well as the kinetic energy K and mass $\mathbf{M}_{mass}$ of the particle, namely the artificially introduced auxiliary quantities (Fichtner et al., 2018).

The potential energy is derived by negating the logarithm of the probability density linked to the inversion variables:

$$U(\mathbf{m}) = -logP(\mathbf{m}|\mathbf{d}). \tag{2.43}$$

In many inversion problems, including FWI, a plausible assumption of the prior information is the Gaussian uncertainties (e.g., (Tarantola, 2004; Dosso et al., 2014; Fichtner et al., 2018; Fu and Innanen, 2022)). Assuming that both the model and data covariances are known, Gaussian, and uncorrelated, the posterior distribution (or potential energy in HMC) can be acquired by ignoring the practically expensive term $P(\mathbf{d})$ in equation 2.35, and combining equation 2.36 and equation 2.37:

$$U(\mathbf{m}) = -logP(\mathbf{m}|\mathbf{d}) = \frac{1}{2}(\mathbf{d}_{obs} - \mathbf{d}_{syn})^T \mathbf{C}_D^{-1}(\mathbf{d}_{obs} - \mathbf{d}_{syn}) + \frac{1}{2}\mathbf{m}^T\mathbf{C}_M^{-1}\mathbf{m} + const. \tag{2.44}$$

The HMC algorithm for an inversion problem can be summarized as follows. Initially, the momentum vector $\mathbf{p}$ is sampled from a multivariate normal distribution with zero mean and covariance matrix $\mathbf{M}_{mass}$, denoted $\mathcal{N}(0, \mathbf{M}_{mass})$. The algorithm then transitions from the current state $(\mathbf{p}_{cur}, \mathbf{m}_{cur})$ to a new state $(\mathbf{p}_{new}, \mathbf{m}_{new})$ using the Leapfrog method (Iserles, 1986). The new state is accepted with a probability determined by a variant of the Metropolis rule (Metropolis et al., 1953):

$$P_{acceptance} = min\left[1, exp\left(H(\mathbf{p}_{cur}, \mathbf{m}_{cur}) - H(\mathbf{p}_{new}, \mathbf{m}_{new})\right)\right]. \tag{2.45}$$

Figure 2.10 illustrates the trajectories of a sampler in a 2D problem at the beginning, middle, and end of the HMC process, showing how the gradient information facilitates a thorough exploration of the model space. Initially, particles are distributed according to an arbitrary initial distribution. Over successive iterations, the positions of these particles are updated

progressively toward the optimal distribution while ensuring that they remain sufficiently spaced from one another. The final arrangement of these particles provides an estimate of the posterior distribution.

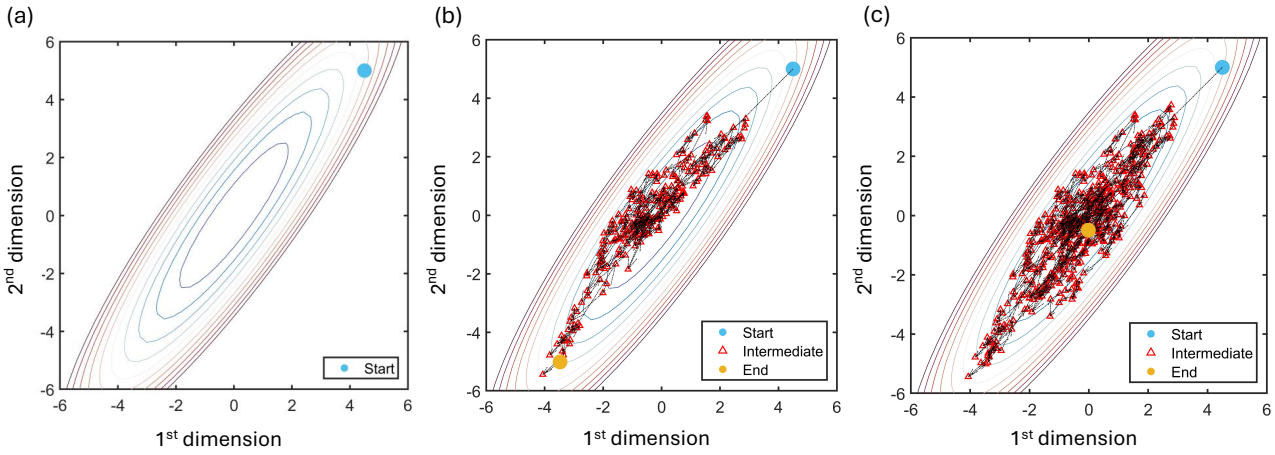

Figure 2.10: HMC mechanism. Dashed arrows show the gradient directions pointing to the next sampler position. (a) Model space with initial sampler position. (b) Model space with sampler trajectories in the middle of the sampling process. (c) Model space with sampler trajectories after the sampling process is finished.

2.5.2 Stein Variational Steepest Descent

Variational inference provides a powerful alternative to traditional sampling methods for approximating complex posterior distributions by framing the problem as an optimization task that seeks to minimize the KL divergence Kullback and Leibler (1951). The KL divergence, also referred to as relative entropy, quantifies the statistical distances between two posterior density functions: the approximate distribution generated by the inference process and the true posterior distribution (Blei et al., 2017). The KL divergence is explicitly expressed as:

$$KL\left[q\left(\mathbf{m}\right)||P\left(\mathbf{m}|\mathbf{d}\right)\right] = E_q\left[logq\left(\mathbf{m}\right)\right] - E_q\left[logP\left(\mathbf{m}|\mathbf{d}\right)\right], \tag{2.46}$$

where $q(\mathbf{m})$ is the current distribution, and E_q means taking expectations with respect to q. The objective in a variational Bayesian inference problem is:

$$q^*(\mathbf{m}) = argminKL\left[q(\mathbf{m})||P(\mathbf{m}|\mathbf{d})\right], q \in Q, \tag{2.47}$$

where Q is a computationally tractable family of distributions.

The SVGD method operates as an incremental-transform-based gradient descent technique aimed at minimizing the KL divergence and guiding a collection of particles representing the initial parameter distributions (Liu and Wang, 2016). Through performing smooth transforms, the derivative of the KL divergence concerning the magnitude of a small perturbation can be iteratively computed. Thus, the final posterior can be gradually constructed from an arbitrary initial distribution. One of the advantages of this approach is its capability to approximate various geometries of a posterior distribution (Blei et al., 2017). This versatility makes SVGD particularly promising in practical applications, as it can effectively handle a wide range of distributional complexities. The smooth transform can be shown as:

$$T(\mathbf{m}) = \mathbf{m} + \alpha V(\mathbf{m}), \tag{2.48}$$

where α is a magnitude that controls the magnitude of update with a smooth transform direction $V(\mathbf{m})$. Accordingly, the objective of the SVGD approach becomes as follows:

$$q_T^*(\mathbf{m}) = argminKL\left[q_T(\mathbf{m})||P(\mathbf{m}|\mathbf{d})\right], q_T \in Q. \tag{2.49}$$

where $q_T(\mathbf{m})$ denotes the transformed distribution from the initial distribution $q(\mathbf{m})$. The gradient of the KL divergence with respect to the perturbation magnitude α can be computed with Stein's operator (Liu and Wang, 2016):

$$\nabla_\alpha KL\left[q_T(\mathbf{m})||P(\mathbf{m}|\mathbf{d})\right]|_{\alpha=0} = -E_q\left[trace\left(A_P V(\mathbf{m})\right)\right], \tag{2.50}$$

where

$$A_P V(\mathbf{m}) = \nabla_{\mathbf{m}} logP(\mathbf{m}|\mathbf{d})\, V^T(\mathbf{m}) + \nabla_{\mathbf{m}} V(\mathbf{m}), \tag{2.51}$$

means applying Stein's operator (Stein, 1972) to the transform function.

Equation 2.50 suggests that by maximizing the right-hand side, the steepest descent direction of the KL divergence can be obtained. The choice of the optimal function set V^* can be simplified by using Stein's identity (Liu et al., 2016; Chwialkowski et al., 2016). This optimal function can be formulated as:

$$V^*\left(\mathbf{m}\right) \propto E_q\left[A_P k\left(\mathbf{m}, \mathbf{m}^{'}\right)\right], \tag{2.52}$$

where $k\left(\mathbf{m}, \mathbf{m}^{'}\right)$ is the kernel function controlling the interaction between two sets of model parameters $\mathbf{m}$ and $\mathbf{m}^{'}$. A commonly used kernel is the radial basis function (RBF) kernel:

$$k\left(\mathbf{m}, \mathbf{m}^{'}\right) = exp\left(-\frac{1}{h_m}||\mathbf{m} - \mathbf{m}^{'}||^2\right), \tag{2.53}$$

where h_m is a scale factor that intuitively controls the interaction intensity among distributions of different model parameters.

Combining the aforementioned elements with the optimization problem, an iterative process can be built to approximate the initial distribution to the target distribution. In the ith iteration for a set of model parameter $\mathbf{m}$, this process can be briefly illustrated below:

$$\begin{aligned} T_i\left(\mathbf{m}\right) &= \mathbf{m} + \alpha_i V_i^*\left(\mathbf{m}\right), \\ q_{i+1} &= q_{i[T_i]}. \end{aligned} \tag{2.54}$$

In Figure 2.11, I demonstrate the schematic mechanism of the SVGD method. Initially, particles are distributed according to an arbitrary initial distribution. Over successive iterations, the positions of these particles are updated progressively toward the optimal distribution while ensuring that they remain sufficiently spaced from one another. The final arrangement of these particles provides an estimate of the posterior distribution.

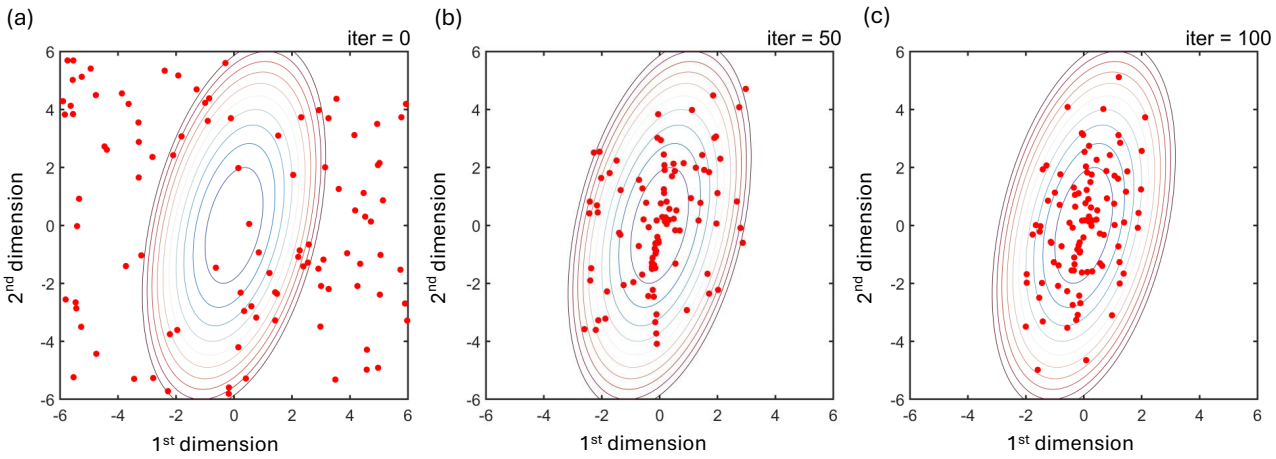

Figure 2.11: SVGD mechanism. (a) Model space with initial particle positions. (b) Model space with particle positions in the middle of the optimization process. (c) Model space with particle positions after the optimization process.

2.6 Targeted nullspace shuttles

In this section, I have a brief review of the targeted nullspace shuttle approach that is first invented and described in more detail by Keating and Innanen (2021). The optimal model from equation 2.1 in FWI generally takes the best waveform match, and can thus best approximate the true subsurface properties. In fact, such a definition of the ideal model is ambiguous in real data applications, but the assumption here is that this is the best solution to this inverse problem. In practice, several fundamental limitations prevent attainment of the best solution, such as the computational tractability in wavefield simulations, the data limitations, and the nonlinearity of the inverse problem. Assume that FWI typically converges to a $\hat{\mathbf{m}}$ satisfying:

$$\Phi\left(\hat{\mathbf{m}}\right) \leq \varepsilon, \tag{2.55}$$

where ε represents an acceptable misfit threshold. The set of all such models forms the ε-equivalence class:

$$\mathbf{M}_{\varepsilon} = \mathbf{m}|_{\Phi(\mathbf{m}) \leq \varepsilon}, \tag{2.56}$$

Such a class defines a computationally tractable subspace containing significantly fewer models than the full solution space. However, even within this reduced subspace, systematic exploration remains impractical due to the high dimensionality of typical FWI problems (often over 10^6 model parameters for 3D surveys). The nullspace shuttle method addresses this challenge by operating on a carefully selected subset of models that maintain data fit consistency and preserve a priori structural constraints. The approach is regarded as a post-inversion process, where the model $\mathbf{m}^*$ is defined as lying near a minimum and taking $\Phi(\mathbf{m}^*) = \sigma$. Taking this point, probing the model space along a direction $\Delta\mathbf{m}$ such that

$$\Phi(\mathbf{m}^* + \Delta\mathbf{m}) \approx \Phi(\mathbf{m}^*) + \mathbf{g}(\mathbf{m}^*)\Delta\mathbf{m} + \frac{1}{2}\Delta\mathbf{m}^T\mathbf{H}_{GN}(\mathbf{m}^*)\Delta\mathbf{m}, \tag{2.57}$$

where $\mathbf{g}(\mathbf{m}^*)$ is the gradient at $\mathbf{m}^*$, and $\mathbf{H}_{GN}(\mathbf{m}^*)$ is the Gauss-Newton approximation of the Hessian matrix at $\mathbf{m}^*$.

The model perturbation $\Delta\mathbf{m}$ is a direction vector, with the basis vector being:

$$\delta\mathbf{m} = \frac{\Delta\mathbf{m}}{||\Delta\mathbf{m}||}. \tag{2.58}$$

The target is now finding the possible updating direction of model updates without violating the objective function for which $\Phi(\mathbf{m}^* + \Delta\mathbf{m}) \approx \Phi(\mathbf{m}^*)$. The direction can be defined as $\Delta\mathbf{m}'$, and it is formed by scaling the basis vector in equation 2.58 such that

$$\Delta\mathbf{m}' = \alpha\delta\mathbf{m}, \tag{2.59}$$

with the scalar being a stepsize such that

$$\alpha = \frac{-2\mathbf{g}(\mathbf{m}^*)\delta\mathbf{m}}{\delta\mathbf{m}^T\mathbf{H}_{GN}(\mathbf{m}^*)\delta\mathbf{m}}. \tag{2.60}$$

The targeted nullspace shuttle method can then be formed to test specific hypotheses about features of the inversion results by constructing a hypothetical objective function over possible updates that can preserve the FWI objective, such that

$$\Delta \mathbf{m}^{'} = \underset{\Delta \mathbf{m}^{'}}{argmin} \Psi \left(\mathbf{m}^{*} + \Delta \mathbf{m}^{'} \right), s.t. \Phi \left(\mathbf{m}^{*} + \Delta \mathbf{m}^{'} \right) \approx \Phi \left(\mathbf{m}^{*} \right). \tag{2.61}$$

The hypothetical scalar function Ψ can be minimized through gradient-based approaches, where the gradients can be derived and calculated following the chain rules. Readers can refer to Keating and Innanen (2021) for more detailed derivations of these terms.

2.7 Conclusions

In this chapter, I provide an overview of the general FWI framework that underpins most implementations, including those adopted in this thesis, and outline key methodological choices tailored to my research. For forward modeling, I employ a viscoelastic frequency-domain finite-difference approach for computationally manageable tasks. For more computationally intensive problems, including 3D FWI, sampling-based methods, and variational inference, I use isotropic acoustic frequency-domain modeling. While some forward-modeling strategies could benefit from advanced spatial gridding techniques, the conclusions in subsequent chapters are not strongly dependent on the specific implementation described here. Although the results presented are based on these methodologies, variations in parameterization, objective functions, or forward-modeling techniques would not fundamentally alter the nature of the inversion problem or the primary findings, particularly for numerical experiments. The objective function throughout this thesis follows the standard L_2 data-fitting formulation, and optimization approaches are selected according to the practical demands of each problem, as illustrated in the following chapters. While the methodologies discussed here are not entirely novel, they provide the foundation for the original contributions presented in later chapters.

Chapter 3

Simultaneous waveform inversion of seismic-while-drilling data for P-wave velocity, density, and source parameters

3.1 Summary

FWI as an optimization-based approach to estimating subsurface models, is limited by incomplete acquisition and illumination of the subsurface. The incorporation of additional data from new and independent ray paths should be expected to result in significant increase in the accuracy of FWI models. In principle, the SWD technology can supply these additional ray paths; however, it introduces a new suite of unknowns, namely precise source locations (i.e., drilling path), source signature, and radiation characteristics. Here I formulate a new FWI algorithm in which source radiation patterns and positions join the velocity and density values of the grid cells as unknowns to be determined. I then conduct several numerical inversion experiments with different source settings, using a synthetic model. The SWD sources are supplemented by explosive sources and multi-component receivers at the surface, simulating a conventional surface acquisition geometry. The subsurface model and SWD source properties are recovered and analyzed. The analysis is suggestive that SWD involvement can enhance the accuracy of FWI models, with varying degrees of enhancement depending on factors such as trajectory inclination, source density, and drill path extension. The impact of SWD-FWI over standard FWI is reduced when low-frequency data are missing, but improvements over the models constructed with no subsurface sources remain. This formulation permits general source information, such as position and moment tensor components, to be independently obtained. This inversion scheme may lead to a range of

potential applications where both medium properties and source information are required.

3.2 Introduction

In the application of FWI to field data, insufficient acquisition geometry has become a major challenge. In particular, the limited spatial sampling and aperture of the seismic experiment (Jannane et al., 1989; Mothi et al., 2014). In general, the complicated and non-uniform ray paths associated with incomplete acquisition and complex heterogeneous media introduce shadow zones, regions of the geological medium that are poorly constrained. Physical properties in those regions are inaccurately recovered unless additional a priori subsurface information is used to fill in the gaps in ray path coverage (Schuster et al., 2004). The unavailability of low and intermediate model wavelengths from datasets associated with short-offset surface acquisitions is perhaps the best-known issue arising from acquisition limitations −− wide-aperture acquisition, which increases the number of overlapping ray paths, is essential in these cases (Virieux and Operto, 2009; Vigh et al., 2021).

A clear, but often practically difficult, solution is to include additional, e.g., subsurface, sources or receivers in the seismic survey, which introduce unique ray paths supporting estimation of underground media. Acquisition geometries which include sources or receivers in the subsurface, for instance, cross-well (Pratt et al., 1996; Pratt, 1999; Pratt and Shipp, 1999) and VSP (Pan et al., 2018; Podgornova et al., 2018) naturally supply these, and lead to different, and in some respects better, FWI model estimates. Generally, improvements are tied to the appearance of ray paths which interact with the medium through transmission-like geometries; in reflection surveys, these are for the most part limited to diving waves, which explains the need for long offsets (Virieux and Operto, 2009; Brittan and Jones, 2019).

SWD is a longstanding auxiliary tool in exploration and monitoring seismology (e.g., James W. Rector and Hardage, 1992; Miranda et al., 1996; Naville et al., 2004; Rocca et al., 2005; Poletto et al., 2020; Auriol et al., 2021; Poletto et al., 2022; Silvestrov et al., 2023). In

SWD settings, the drill bit acts as a seismic source, and sensors arrayed in a range of possible geometries produce an experiment similar to a reverse VSP (Khaled et al., 1996; Miranda et al., 1996). The approach is attractive in that it generates a range of important extra seismic information without interrupting drilling processes (Poletto and Miranda, 2022). SWD has attracted interest in recent years, taking advantage of advances in the understanding of drill-string dynamics and the use of neural networks for data analysis (e.g., Auriol et al., 2021). It has been pointed out that the introduction of SWD ray paths, which begin at the drill bit and end at surface or borehole receivers, can in principle constrain velocity models (Bertelli and di Cesare, 1999; Auriol et al., 2019), and that this could further impact imaging (Vasconcelos and Snieder, 2008; Kazemi et al., 2018; Goertz et al., 2020) and FWI (Kazemi et al., 2021; Li et al., 2022) by partially addressing the issues discussed above.

The use of SWD data in conjunction with other seismic data assists model construction via FWI, which means it will contribute to the optimization of drilling. A drilling project can in principle be de-risked when better estimates of physical parameters near the drill bit are available (Poletto et al., 2004; Martinez et al., 2020). Reducing uncertainty in, for instance, locations of formation tops, and heterogeneities in general within a complex stratigraphy, is helpful for de-risking in SWD projects (e.g., the avoidance of zones of high pore pressure). Further, near-real-time updating of drilling parameters requires accurate seismic velocity estimates within the formations interacting with the drill bit (Auriol et al., 2021). Models deriving from FWI would in principle address both of these issues.

Bertelli and di Cesare (1999) discussed the potential of SWD for seismic imaging and inversion. This was accomplished by employing near-real-time reprocessing techniques on the SWD dataset, which allowed for the continuous modification and refinement of the velocity model used for migration. Rocca et al. (2005) developed a migration approach wherein a circular line of receivers is deployed to apply 3D migration in the angular frequency domain, and this method successfully recovered the position of reflectors in a horizontally layered

medium. Vasconcelos and Snieder (2008) applied deconvolutional interferometry in SWD field data to generate an interferometric image of a fault zone at depth within the study area, and Poletto et al. (2012) tested SWD with seismic interferometry and migration to map a geothermal field around wells. Kazemi et al. (2018) formulated an illumination compensation for imaging of surface data using SWD data processed with a sparse multichannel blind deconvolution (SMBD) algorithm, originally built for other applications (Kazemi and Sacchi, 2014). Similar arguments led to the formulation of a two-stage sequential SWD-FWI algorithm, in which inversion of data from SWD sources was followed by inversion from conventionally-acquired surface data (Kazemi et al., 2021). These studies provide strong evidence that the ray paths associated with SWD do have a marked impact on model estimation when used in combination with other data sets. However, the assumption of acoustic wave physics, the estimation of a single unknown model parameter class (i.e., P-wave velocity), and the assumption that the radiation characteristics of the drill bit are known (or can be accommodated through preprocessing), are all idealizations that make proper feasibility analysis difficult.

The vertical and horizontal components of the SWD signal are affected by changes in borehole inclination and azimuth, in part because of drill bit radiation patterns (Figure 1.1 in Chapter 1), which makes the SWD signal strongly mult-component in nature. It follows that explanations of the SWD signal requires an elastic as opposed to acoustic theory. Further, the complicated interaction of the continuously-radiating source drill bit and the formation rock will induce strong modeling errors in an FWI scheme formulated with a static, known, and simply-radiating point source.

In this chapter, I introduce and examine a 2D elastic multi-parameter FWI approach that aims at recovering both source-related variables (positions and moment tensor components) and multiple unknowns associated with medium properties (P-wave velocity and density). In the theory section, I first show the formulation of seismic sources defined with moment tensors

and the treatment of arbitrary source locations. The inversion algorithm is then introduced, including the derivations of inversion terms used for subsurface models and unknown source parameters. Finally, I show several numerical experiments with various acquisitions, with subsequent analysis of the results.

3.3 Theory

3.3.1 General FWI objective function

For incorporation of source unknowns in the FWI problem, I present the reformed objective function as a more general form than equation 2.16:

$$\Phi\left(\mathbf{x}\right)=\sum_{j=1}^{N_{\omega}}\sum_{k=1}^{N_{s}}\frac{1}{2}\|\mathbf{R}\mathbf{u}_{j,k}-\mathbf{d}_{j,k}\|_{2}^{2},\ \text{s.t. }\mathbf{S}(\mathbf{x})\mathbf{u}=\mathbf{f}, \tag{3.1}$$

where $\mathbf{x}$ is a vector of inversion variables (which will ultimately comprise both subsurface and source unknowns), and other terms remain the same as equation 2.16. For simplicity, in the following derivations, I also assume that the wavefield only has one source and frequency component.

The adjoint state method (e.g., Plessix, 2006) is used to determine the gradient through a combination of forward and backward propagated wavefields, the latter of which has the residual wavefield as the source. The Lagrangian of the current problem is

$$L\left(\mathbf{x}\right)=\frac{1}{2}\|\mathbf{R}\mathbf{u}-\mathbf{d}\|_{2}^{2}+\left\langle \mathbf{S}\left(\mathbf{x}\right)\mathbf{u}-\left(\mathbf{f}_{R}+i\mathbf{f}_{I}\right),\lambda\right\rangle , \tag{3.2}$$

where λ is a vector of Lagrange multipliers, $\mathbf{f}_R$ and $\mathbf{f}_I$ are the real and imaginary parts of the source term $\mathbf{f}$, and $\langle\cdot,\cdot\rangle$ represents the inner product. The $\overline{\lambda}$ is determined through backpropagation of data residuals, after which we have:

$$\frac{\partial\Phi}{\partial\mathbf{x}}=\frac{\partial L\left(\overline{\mathbf{u}},\overline{\lambda}\right)}{\partial\mathbf{x}}. \tag{3.3}$$

The incorporation of the Hessian matrix in the optimization procedure accelerates convergence and improves the resolution of the inversion. However, as discussed earlier, the full

Newton approach is generally too computationally expensive to be practical (Pratt et al., 1998), especially in this case I have enlarged the subset of known parameters. In the methodology of this chapter, I make use of a TGN Hessian calculation with several fixed iterations. Hessian-vector products are derivable similarly to the gradient, where details can be found in Section 2.4.3 from equation 2.22 to equation 2.25. Differently, the Lagrangian can be set to a more general form to make both the derivation for model and source parameters available: the Lagrangian

$$\tilde{L} = \langle \mathbf{u}(\mathbf{x}), \mathbf{w} \rangle + \langle \mathbf{S}(\mathbf{x})\,\mathbf{u} - (\mathbf{f}_R + i\mathbf{f}_I), \mu \rangle, \tag{3.4}$$

where $\mathbf{w} = \mathbf{R}^T\mathbf{R}\mathbf{J}\mathbf{v}$, and $\mathbf{v}$ is the vector to be multiplied by the Hessian. The Lagrange multiplier $\overline{\mu}$ is defined in the same manner as in Chapter 2 and satisfies the same adjoint relation (see equations 2.21 and 2.22). The Lagrange multiplier $\overline{\mu}$ satisfies the same with equation 2.24. The product of the Jacobian $\mathbf{J}$ with the vector $\mathbf{v}$ is possible via the derivative of the forward problem with respect to variables $\mathbf{x}_p$ multiplied by vector elements $\mathbf{v}_{\mathbf{x}_p}$ (Métivier et al., 2013; Keating and Innanen, 2020):

$$\frac{\partial\left(\mathbf{S}\mathbf{u} - \mathbf{f}\right)\mathbf{v}_{\mathbf{x}_p}}{\partial \mathbf{x}_p} = 0. \tag{3.5}$$

This allows to derive the Lagrange multiplier $\overline{\mu}$ in equation (3.4). The Gauss-Newton Hessian vector product $\mathbf{H}_{GN}\mathbf{v}$ is then $\frac{\partial \tilde{L}}{\partial \mathbf{x}}$.

3.3.2 Incorporation of source unknowns

Now suppose that the model unknown vector $\mathbf{x}$ includes both medium property unknowns, contained in the vector $\mathbf{m}$, and source unknowns, contained in the real vectors $\mathbf{f}_R$ and $\mathbf{f}_I$, such that the full source vector in equation (3.1) is $\mathbf{f} = \mathbf{f}_R + i\mathbf{f}_I$, as seen in equations (3.2) and (3.4). Equation (3.3) then contains terms of the form

$$\frac{\partial L}{\partial \mathbf{m}_p} = \left\langle \frac{\partial \mathbf{S}}{\partial \mathbf{m}_p}\overline{\mathbf{u}}, \overline{\lambda} \right\rangle, \tag{3.6}$$

where p labels the pth component of model parameters, and of the form

$$\frac{\partial L}{\partial \mathbf{f}_{R_p}} = -\mathfrak{Re}\left(\overline{\lambda_p}\right), \tag{3.7}$$

and

$$\frac{\partial L}{\partial \mathbf{f}_{I_p}} = \mathfrak{Im}\left(\overline{\lambda_p}\right). \tag{3.8}$$

Both $\mathbf{f}_R$ and $\mathbf{f}_I$ are in general vectors of very high dimension, with numbers of elements equal to the number of wavefield grid cells multiplied by the number of the unknown sources considered. In practice it is necessary therefore to formulate the inverse problem with a restricted set of source-related variables being solved for. I use the variable $\mathbf{f}_r$ to parameterize the moment tensor or the position of a point source, in which case the derivative of the objective function is

$$\frac{\partial \Phi}{\partial \mathbf{f}_r} = \sum_p \left[-\mathfrak{Re}\left(\overline{\lambda}_p\right) \frac{\partial \mathbf{f}_{R_p}}{\partial \mathbf{f}_r} + \mathfrak{Im}\left(\overline{\lambda}_p\right) \frac{\partial \mathbf{f}_{I_p}}{\partial \mathbf{f}_r} \right], \tag{3.9}$$

where the first part within the summation on the right-hand side is the derivative of the real component of the source term with respect to $\mathbf{f}_r$, and the second part is the derivative of the imaginary component of the source term with respect to $\mathbf{f}_r$.

This concept also holds in the Hessian-vector products. Using the relation in equation 3.5, it can be shown that

$$\mathbf{S}\left(\mathbf{J}_{\mathbf{m}}\mathbf{v}_{\mathbf{m}}\right) = -\mathbf{u}\sum_p \left(\frac{\partial \mathbf{S}}{\partial \mathbf{m}_p} \mathbf{v}_{\mathbf{m}_p} \right), \tag{3.10}$$

$$\mathbf{S}\left(\mathbf{J}_{\mathbf{f}_R}\mathbf{v}_{\mathbf{f}_R}\right) = \sum_p \left(\mathbf{v}_{\mathbf{f}_{R_p}} \right), \tag{3.11}$$

and

$$\mathbf{S}\left(\mathbf{J}_{\mathbf{f}_I}\mathbf{v}_{\mathbf{f}_I}\right) = \sum_p \left(\mathbf{v}_{\mathbf{f}_{I_p}} \right), \tag{3.12}$$

where $\mathbf{v}_{\mathbf{m}_p}$ is the element of $\mathbf{v}$ corresponding to the pth element of $\mathbf{m}$, $\mathbf{v}_{\mathbf{f}_{R_p}}$ is the element corresponding to the pth element of $\mathbf{f}_R$ and $\mathbf{v}_{\mathbf{f}_{I_p}}$ is the element corresponding to the pth element of $\mathbf{f}_I$. This completes the source-medium waveform inversion formulation.

This method for inverting source parameters differs in several ways from other wave-based methods used in standard seismological applications (e.g., Kim et al., 2011). Firstly, I focus on determining effective moment tensor combinations that describe general motions caused by external mechanisms such as SWD, whereas in seismology one typically seeks a more comprehensive description of centroid moment tensors. I also assume that any partially or fully unsettled sources occur instantaneously, and thus, I do not consider the source-time function (moment rate function) that is used in the centroid moment tensor inversion. Instead of simplifying the problem by ignoring variations in structural properties, this inversion framework explicitly considers how the moment tensor and source location jointly affect the wavefield.

3.4 Numerical experiments

The numerical testing summarized in this section, involving simulated models, drilling paths, signatures, and wavefield data, is designed to validate the inverse formulation and to explore the potential benefits and limitations of combining SWD and surface seismic datasets in FWI. I adopt a single, fixed pair of true and initial models, plotted in Figure 3.1. The models are 300 by 150 grid points in the x and z directions, respectively, with each grid cell 20 meters on a side. Although this multi-parameter FWI has the capacity to recover other parameters, I would like to focus on the simultaneous inversion for P-wave velocity, density, and the source terms in the test sections. This is because the estimation of all physical parameters at the same time can become complex and beyond the scope of this medium-source FWI. Moreover, P-wave velocity and density are especially important in many application scenarios as they provide significant insights into subsurface features (Sheriff and Geldart, 1995). To

accomplish this, I used 1D background models for S-wave velocity based on the general relationship between P- and S-wave velocities, and homogeneous models for quality factors in the viscoelastic modeling.

In this section, I present examples based on three main assumptions. Firstly, the drill trajectory can be treated as if it occupies a discrete sequence of quasi-static positions along it. Secondly, the radiation patterns of the SWD sources are represented by independent and general moment tensors, which means the drill-bit-rock interaction can be approximated by force couples. Thirdly, the SWD source signature can be decomposed into individual frequencies which overlap with the frequencies used in surface FWI. The plausibility and the problem of transformation of SWD signals into useable seismic data have been discussed by Kazemi et al. (2021). Nevertheless, our inversion scheme can be customized to suit datasets with varying spectral content of diverse kinds of sources. This can be achieved by modifying the elements within the wavefield matrix. Specifically, I can assign weights to the wavefields generated by different sources based on the frequency spectrum of their respective signatures. I can then combine these weighted wavefields to form a new wavefield that can be used in the inversion process.

The multi-scale approach (Bunks et al., 1995) is utilized to avoid local minima during the inversion process throughout all the tests. I simplify the treatment of frequencies for synthetic data by assuming a constant and known frequency spectrum for each source, extended to the frequency bandwidth typically used in FWI. The frequency range is divided into 10 total frequency bands, with 12 equally-spaced sub-frequencies in each band. The lowest frequency in each band is set at 1 Hz, while the highest frequency increases linearly from 3 Hz to 15 Hz. The final model for each frequency band is generated using two levels of iteration: 3 outer iterations for updating models, and 20 inner iterations for approximating the inverse Hessian matrix by the L-BFGS algorithm (Nocedal and Wright, 2006).

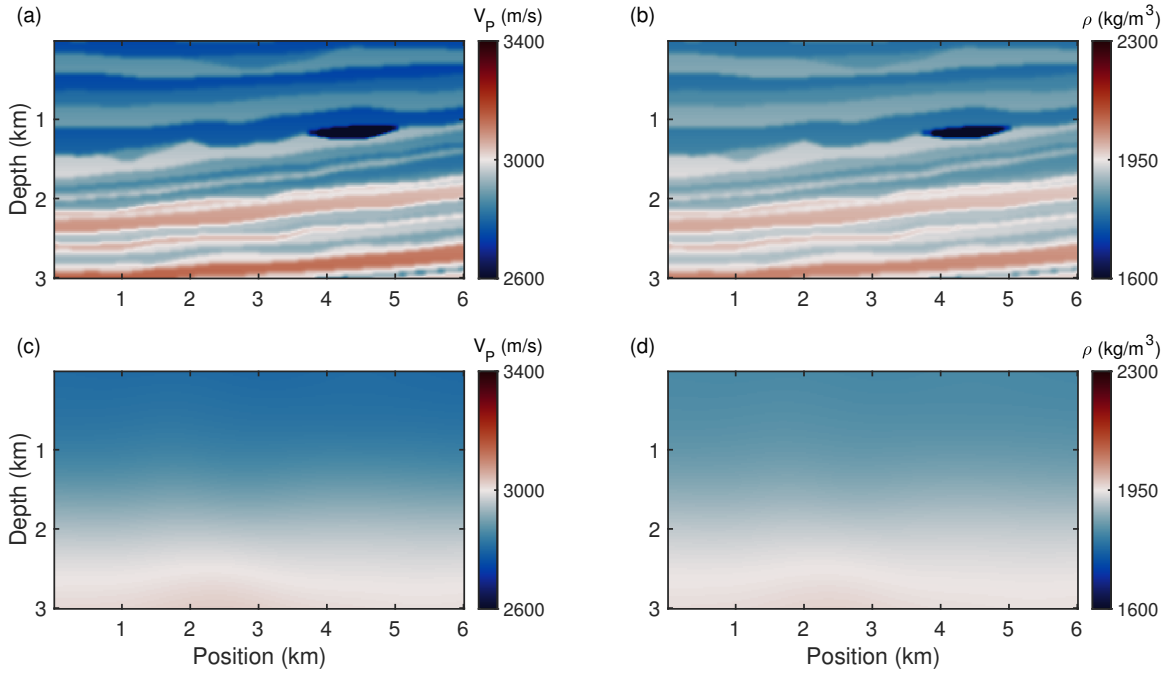

Figure 3.1: The true and initial models for synthetic tests. (a) True P-wave velocity model. (b) True density model. (c) Initial P-wave velocity model. (d) Initial density model.

3.4.1 Simultaneous inversion for subsurface parameters and moment tensors

In this section, I examine the simultaneous estimation of V_P models, density models, and borehole (drill) source moment tensors, assuming known surface source positions and radiation patterns, and borehole source positions. On the surface, 72 explosive sources, separated by 2 grid points, are placed along a line also occupied by 144 multi-component receivers separated by 1 grid point. A drill path is occupied by sources of known positions and unknown moment tensors. The drill path, the source numbers, and the positions are varied, to test the inversion as the effective aperture of the new sources changes. The simulated data I use in the inversions include those from the surface sources and those from the borehole sources. The surface sources are assumed to be stronger and to produce less noisy data. To mimic this, the borehole source amplitudes are scaled by 1/2, and Gaussian noise is added to the data such that the surface data has a signal-to-noise ratio (SNR) of 10 dB and the

drill-generated data has SNR = 5 dB. In this set of experiments, I first demonstrate with numerical examples that the problem as a whole is basically well-posed. A key feature of our approach is that data from a more or less continuously-radiating drill bit are parsed and treated as if the drill occupied a discrete set of fixed points along the path. Some number of such positions, which are not fixed by any theoretical or prior information, must be selected, and investigation is needed to understand whether or not this degree of freedom has a serious impact on the approach.

Figure 3.2 shows the schematic acquisition geometry of the source number test. Our main inversion focus is the area indicated by the black dashed rectangle in the model's lower-right part. The surface acquisition, as schematically shown by the red and blue markers, covers the right-hand side surface of the model. A drilling trajectory is then added in the deeper section of the left-hand side of our model, and unknown sources are indicated by the black stars. The drilling site on the ground has around 1.5 kilometers offset from the ground receivers, and various SWD sources are arranged along a deviated trajectory of 900 meters. The large distance between the drill site and the surface acquisition is based on the following two considerations: (1) in practical applications, there is usually a distance between the drilling sites and the surface sensors; (2) the far-field assumption should be satisfied in representing force-couple-defined sources with a point approximation (Aki and Richards, 2002). I have tested the cases where $N_S = 10, 15, 20, 25$, and 30. The initial and true radiation patterns for each unknown independent point source are set by random moment tensors in a range of -1 to 1.

I first show the results of subsurface properties with the surface-only acquisition in Figure 3.3 as a baseline inversion for further comparison with the cases where SWD sources are included. This baseline inversion is noise-free. The estimation of both P-wave velocity and density is crude, as shown by the indistinct structures in the recovered models. The right-hand side portions of the models are slightly updated, but the other parts where no extra

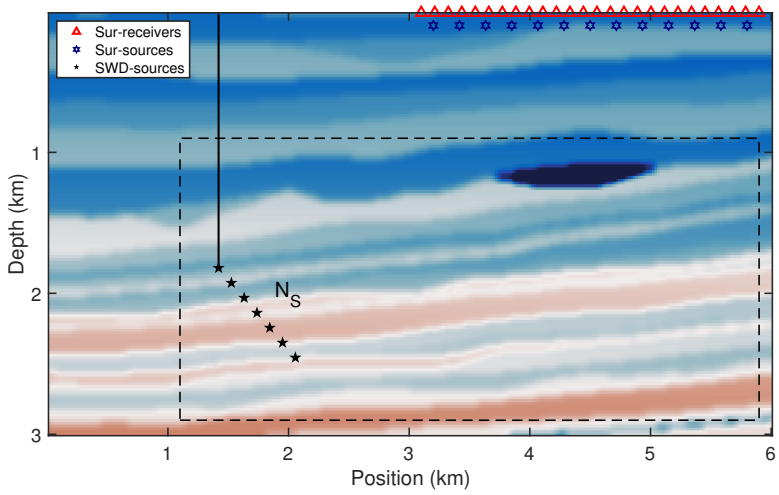

Figure 3.2: Schematic acquisition system in source number test. Surface sources and receivers are displayed by blue and red markers, respectively. Drilling trajectory and discrete sources are schematically shown by the black line and markers. The major inversion focus is indicated by the black dashed rectangle.

source exists are poorly recovered.

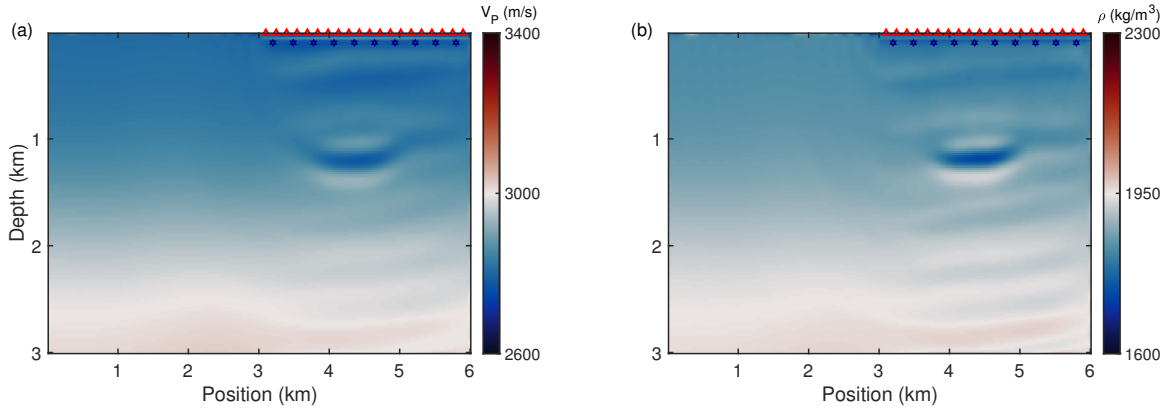

Figure 3.3: Inversion of V_P and ρ with surface acquisition. (a) V_P inversion with surface acquisition. (b) ρ inversion with surface acquisition.

The inversion results in Figure 3.4 undergo a significant positive change upon inclusion of the SWD data, even though there are only a few drilling sources involved (Figure 3.4

(a) and (b)). Not only is the structure below the surface acquisition well-resolved, but the intermediate medium between the extra sources to the receivers is also more accurately estimated. Although visually subtle, the continuity of the lower high velocity and density layers indicates small variations in each case.

For the misfit calculation, I use the root mean square error normalized by the difference between the maximum and minimum values of the true data in the focused area (shown by the portion in the dashed rectangle in Figure 3.2):

$$NRMSE = \frac{RMSE}{|max\,(data_{true}) - min\,(data_{true})|}. \tag{3.13}$$

In the inversion tests presented in this chapter, a 1% variation in the error term means approximately 40 m/s change in P-wave velocity, and 30 kg/m^3 in density within the focused area. The analysis of the NRMSE values reveals that the accuracy of the model recovery for V_P and the number of sources is not necessarily linked. However, for the density estimation, there is a positive correlation between the accuracy of the recovery and the number of sources.

I use 3 vertical profiles in the x-position of 3.6, 4.6, and 5.6 kilometers to quantify the improvement when combining the SWD data. The comparisons are shown in Figure 3.5, where the black lines are from true models, the dark green lines are from the baseline models, and N_{10}-N_{30} denote five cases in which N_S increases from 10 to 30. The NRMSE is also calculated in each profile. The profiles of all the SWD-FWI cases are much closer to the true cross-sections. In the first profile, the optimal velocity inversion occurs with a source number of 20, while the optimal density inversion occurs with a source number of 30. However, this trend does not hold in all profiles, indicating a weak correlation between source number and inversion results in local profiles.

The moment tensor inversion results, however, do not necessarily change in accordance with the number of unknown sources. The cross plots in Figure 3.6 show generally good convergence in all cases. Each black circle in the plot corresponds to a source with different initial moment tensor values, which are adjusted during the inversion process to approach the

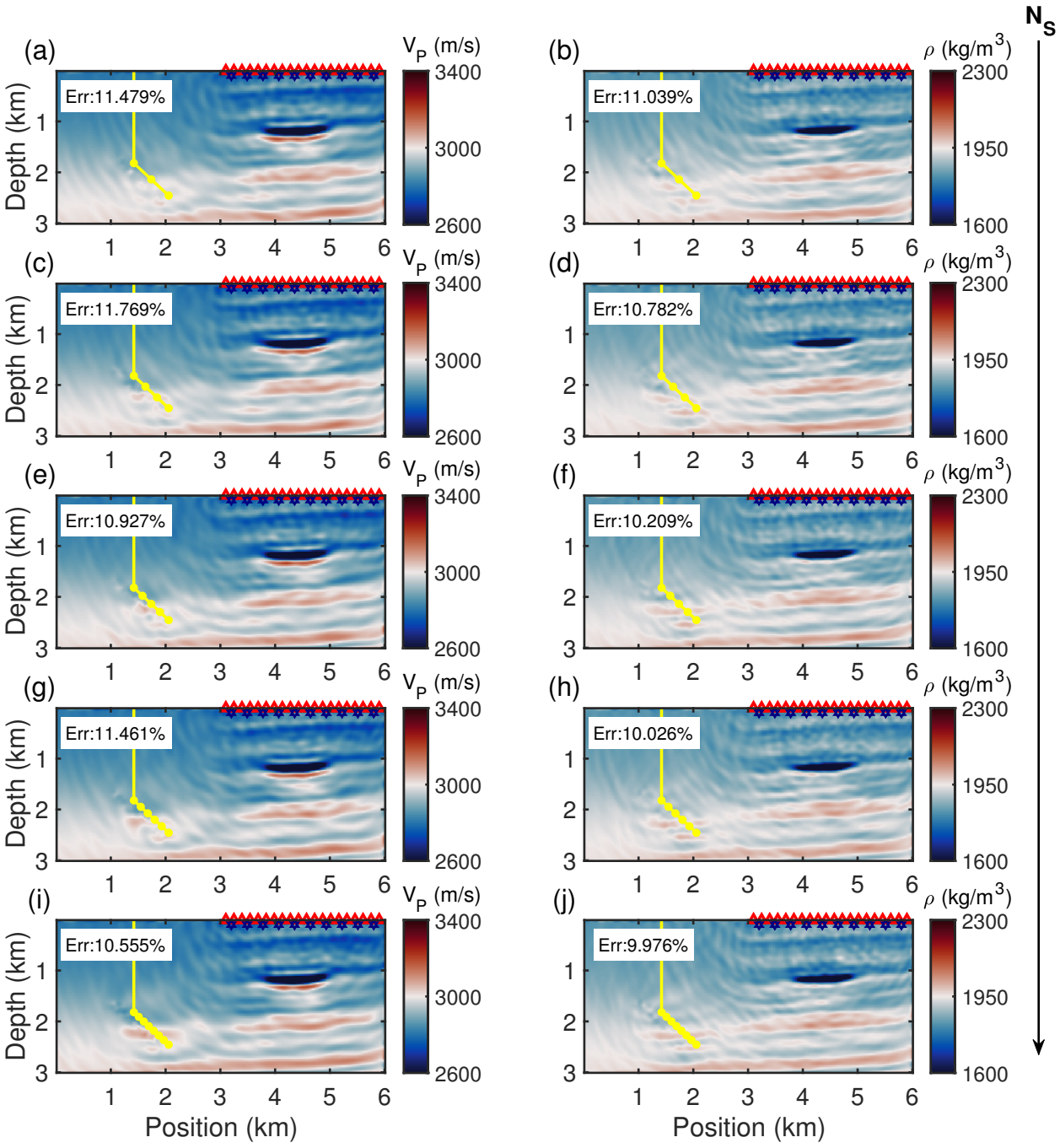

Figure 3.4: Inversion of V_P and ρ with surface and SWD acquisitions with different source number. Drilling trajectory and discrete sources are schematically shown by the yellow line and markers. The black arrow denotes the gradual increment of N_S. (a), (b) inversion with $N_S = 10$. (c), (d) inversion with $N_S = 15$. (e), (f) inversion with $N_S = 20$. (g), (h) inversion with $N_S = 25$. (i), (j), inversion with $N_S = 30$.

true values, as indicated by their proximity to the red line denoting a high level of agreement. A change of 1% in the error term results in an approximate variation of 0.02. Combining the inversions of models and source terms, it is reasonable to infer that the contribution from the increase of SWD sources outweighs the risk of larger nonlinearity introduced by more unknown radiation patterns.

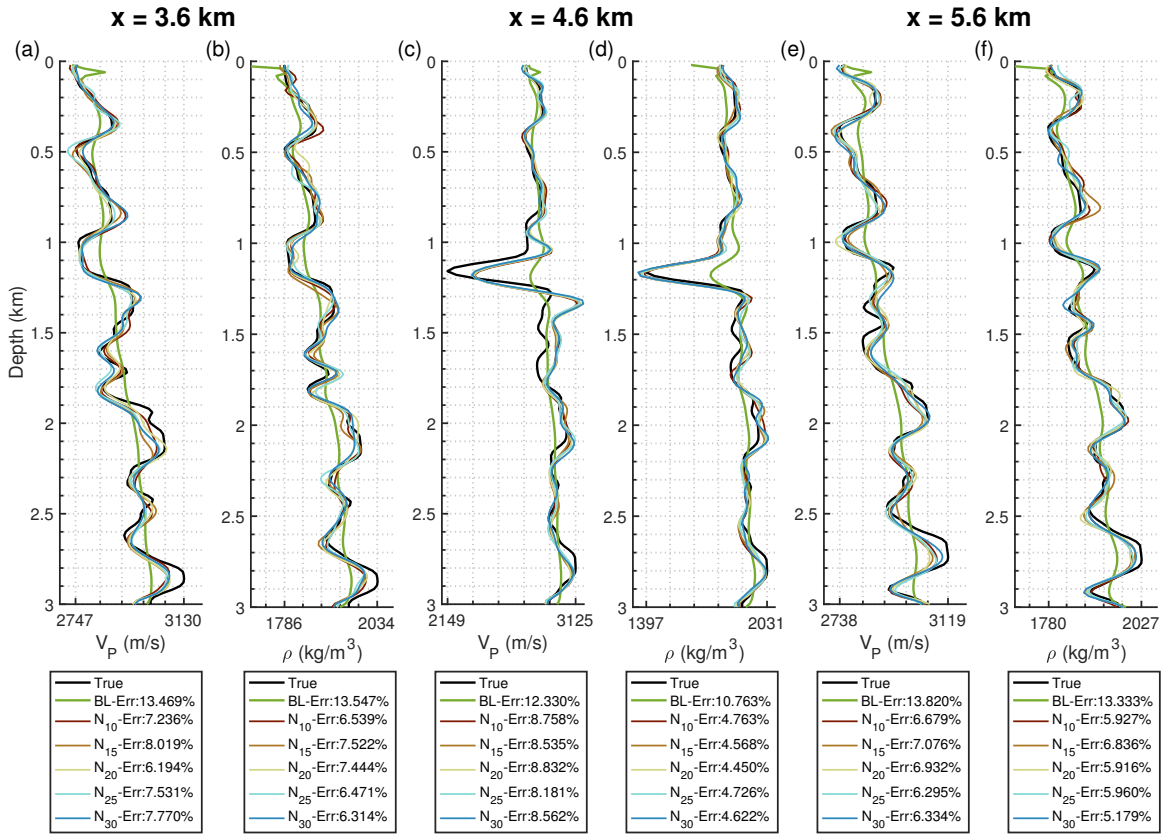

Figure 3.5: Vertical profiles in source number test. (a), (c), and (e) are P-wave velocity cross-sections, while (b), (d), and (f) are density cross-sections. BL denotes baseline.

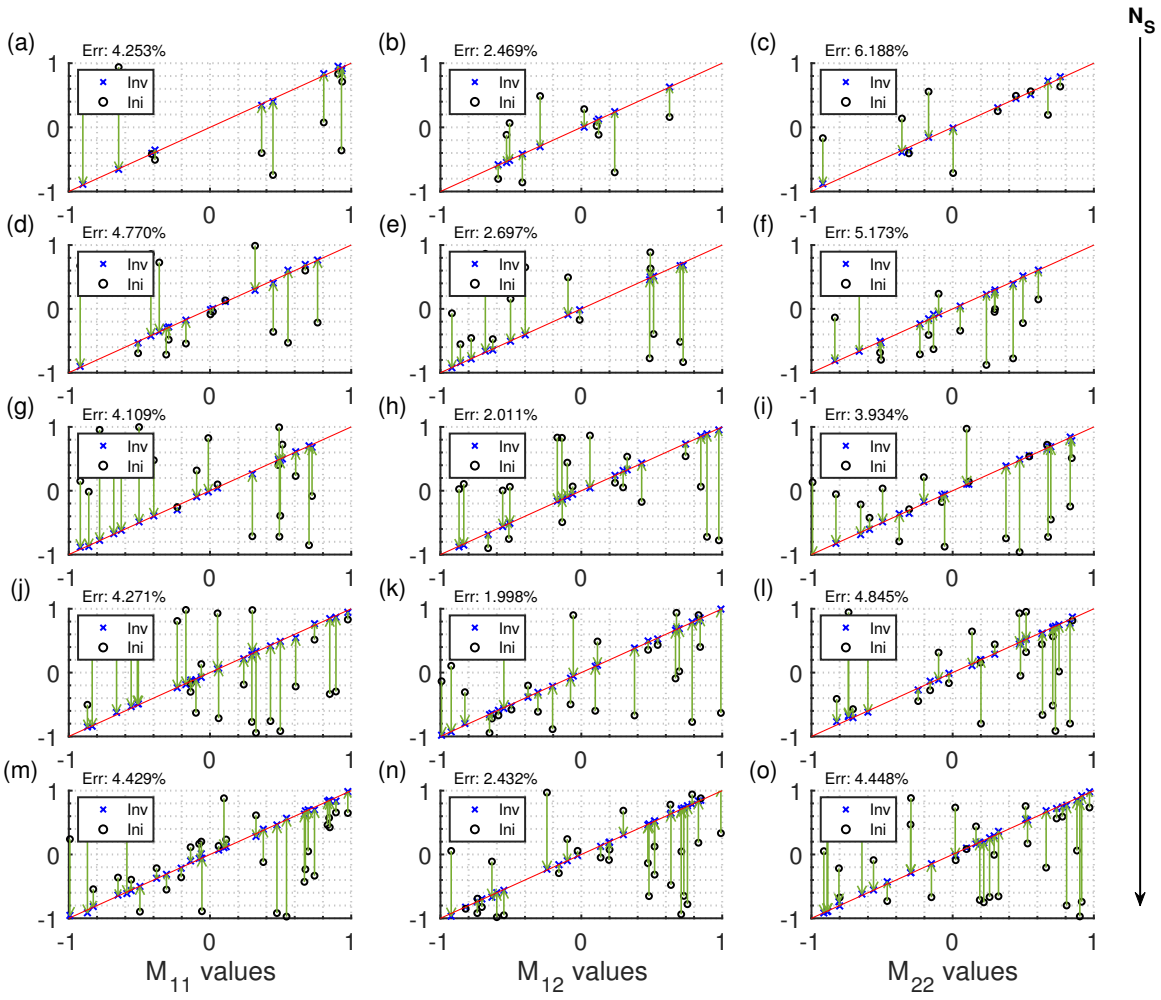

Figure 3.6: Inversion of moment tensors. The x-axis in each subplot represents the initial values, while the y-axis in each subplot represents the true values. The black arrow denotes the gradual increment of N_S, and the green arrows are from initial data points (black circles) to final data points (blue circles). The red line in cross plots represents the line of perfect agreement. (a)-(c) $N_S = 10$. (d)-(f) $N_S = 15$. (g)-(i) $N_S = 20$. (j)-(l) $N_S = 25$. (m)-(o) $N_S = 30$.

The next test I present here examines the influence of deviation of drilling trajectory from vertical. I implement a drilling path with a fixed length of 900 m same as the previous test, with 30 sources with unknown moment tensor settings. The angle θ in Figure 3.7 varies

from 0° (vertical) to 90° (horizontal), with an increment of 22.5°.

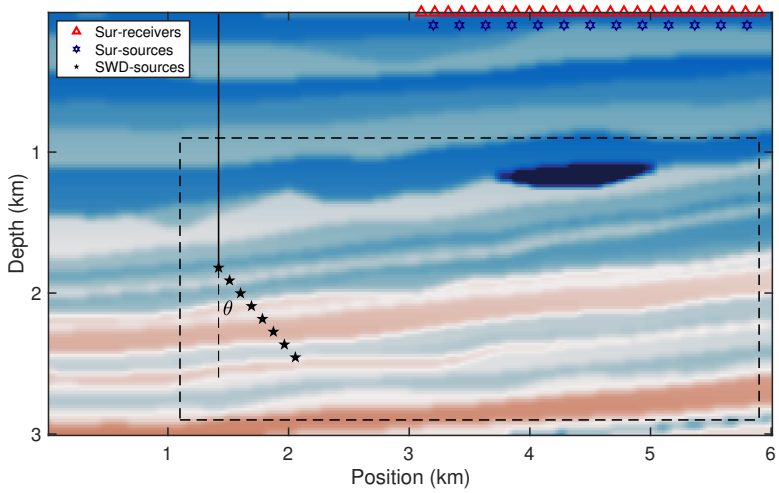

Figure 3.7: Schematic acquisition system in trajectory inclination test. Surface sources and receivers are displayed by blue and red markers, respectively. Drilling trajectory and discrete sources are schematically shown by the black line and markers. The major inversion focus is indicated by the black dashed rectangle.

As shown in Figure 3.8, the layered structures and the low-value region of both the P-wave velocity and density models are depicted after the FWI, emphasizing the help from additional sources when compared to the baseline inversion. The error is relatively lower in inversions with a horizontal drilling path, and the deeper layers with higher V_P and ρ values are better reconstructed because of a higher horizontal resolution.

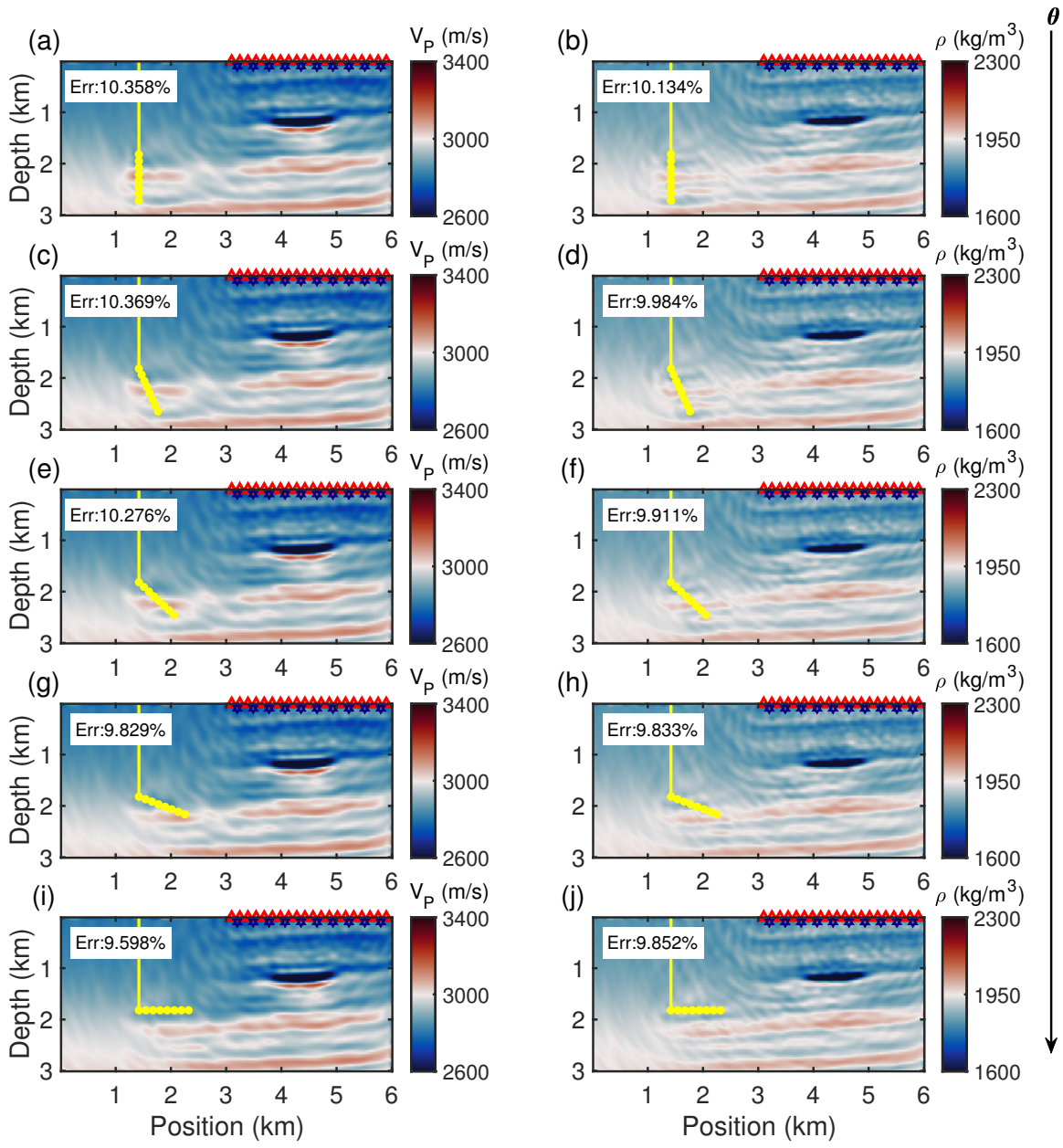

Figure 3.8: Inversion of V_P and ρ with surface and SWD acquisitions with different trajectory inclinations. Drilling trajectory and discrete sources are schematically shown by the yellow line and markers. The black arrow denotes the gradual increment of θ. (a), (b) inversion with $\theta = 0°$. (c), (d) inversion with $\theta = 22.5°$. (e), (f), inversion with $\theta = 45°$. (g), (h) inversion with $\theta = 67.5°$. (i), (j) inversion with $\theta = 90°$.

The profiles in Figure 3.9 also display steady FWI results with different deviations, and all cases with extra SWD sources are better consistent with the true values. The positive correlation between the inversion results and the increase in θ is generally maintained.

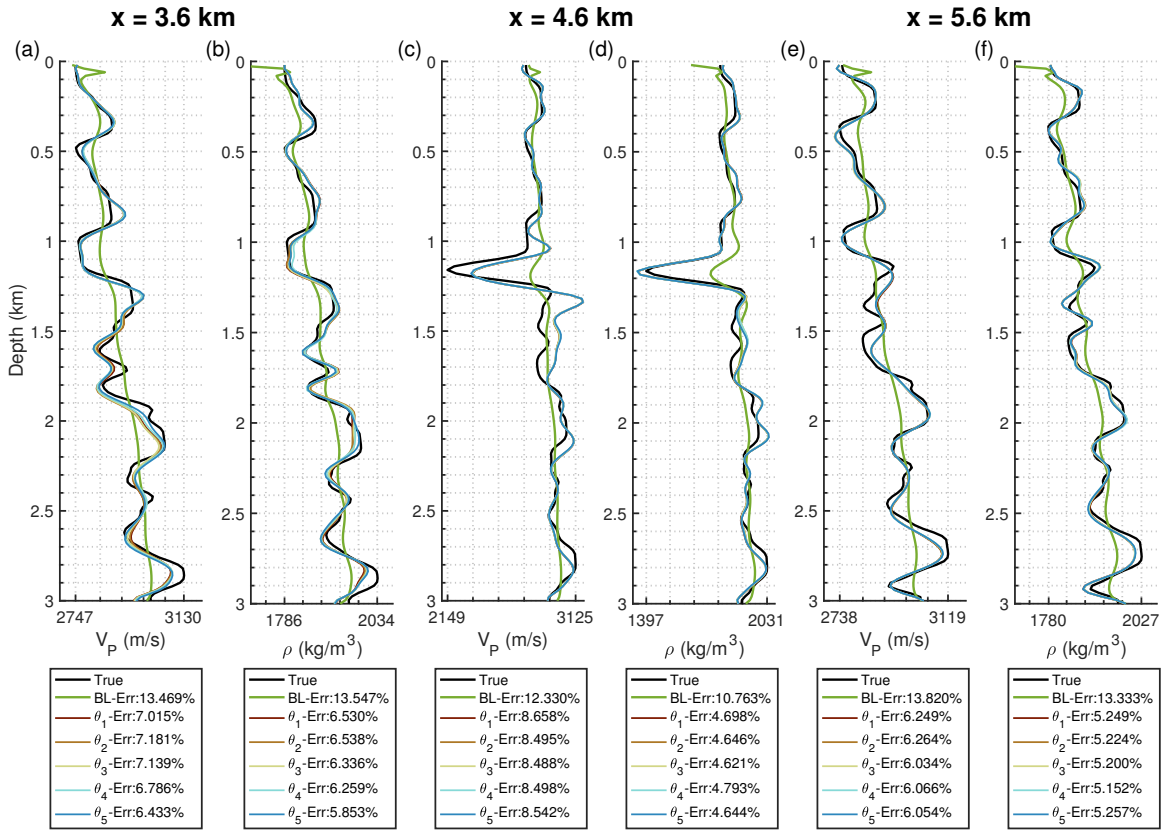

Figure 3.9: Vertical profiles in trajectory inclination test. (a), (c), and (e) are P-wave velocity cross-sections, while (b), (d), and (f) are density cross-sections. BL denotes baseline.

In the cross plots of moment tensor inversion, the overall NRMSE gets smaller when the trajectory deviates, and reaches its lowest in the case where $\theta = 22.5°$, but enlarged again as it becomes closer to horizontal. Generally, the inversion is consistent and accurate.

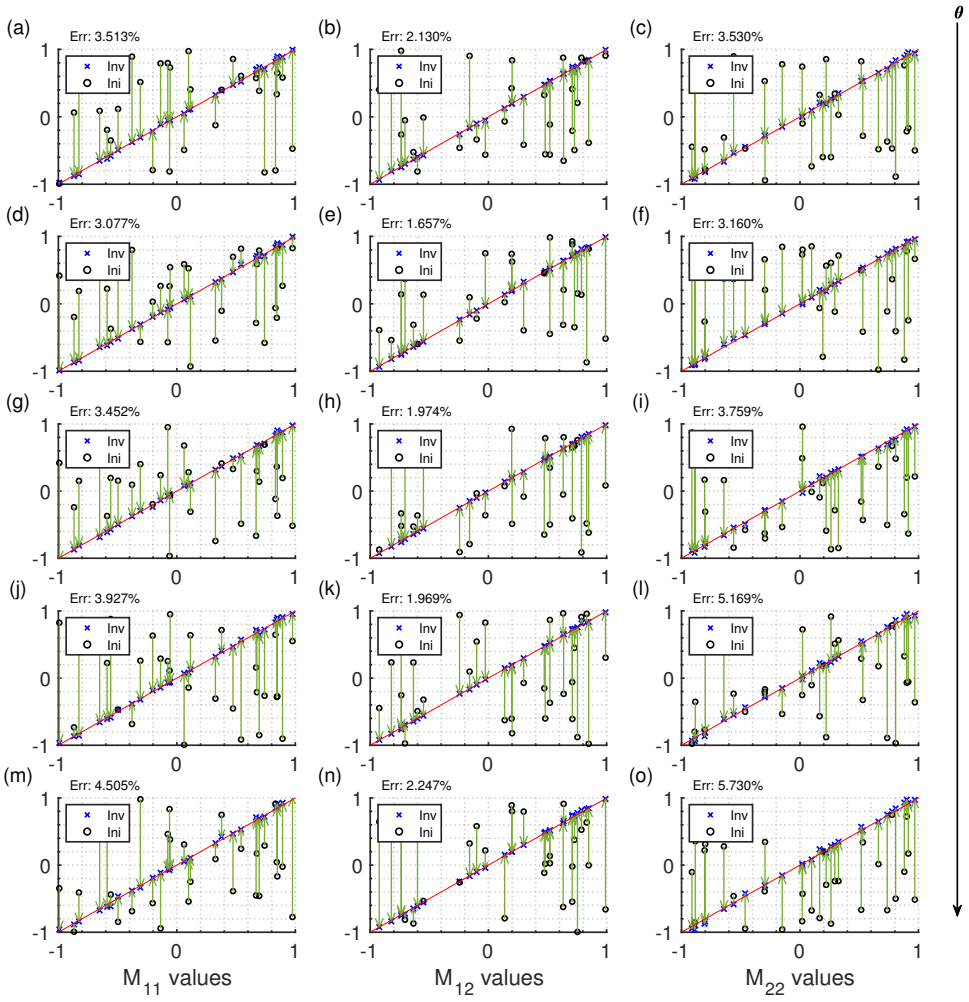

Figure 3.10: Inversion of moment tensors. The x-axis in each subplot represents the initial values, while the y-axis in each subplot represents the true values. The black arrow denotes the gradual increment of θ, and the green arrows are from the initial data points (black circles) to the final data points (blue circles). The red line in cross plots represents the line of perfect agreement. (a)-(c) $\theta = 0°$. (d)-(f) $\theta = 22.5°$. (g)-(i) $\theta = 45°$. (j)-(l) $\theta = 67.5°$. (m)-(o) $\theta = 90°$.

The experiments in the next phase aim at the possible effect of the change of drilling path extension. Figure 3.11 shows the schematic acquisition used in this test. I keep the inclination to be 45° and N_S to be 30, while going through five cases with different trajectory

lengths. The extension L in the first case is 450 meters with clustering sources, while the drilling length in the last case is 1500 meters with a relatively sparse source distribution.

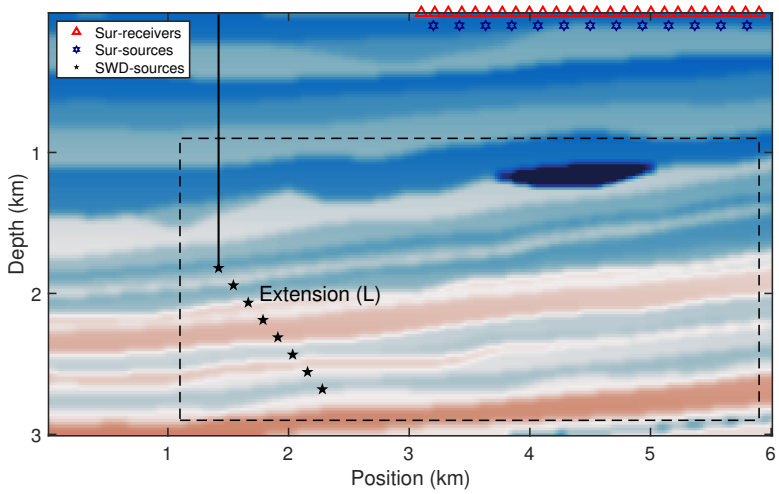

Figure 3.11: Schematic acquisition system in trajectory extension test. Surface sources and receivers are displayed by blue and red markers, respectively. Drilling trajectory and discrete sources are schematically shown by the black line and markers. The major inversion focus is indicated by the black dashed rectangle.

From the overall view in Figure 3.12, we can see a downgoing trend of the misfit in the ρ models, but this trend does not apply to the V_P estimation. In Figure 3.13, the profile that goes through the shallower body with low velocity and density values shows the best result when I use the largest extension, as it gets more comprehensive illumination when I enlarge the extension. However, it is hard to see a clear misfit tendency for the other profiles. The moment tensor inversions (see Figure 3.14) are stable as the cross plots present similar consistency between estimated and actual values. There seems to be little clear correlation between the misfits and drilling path length.

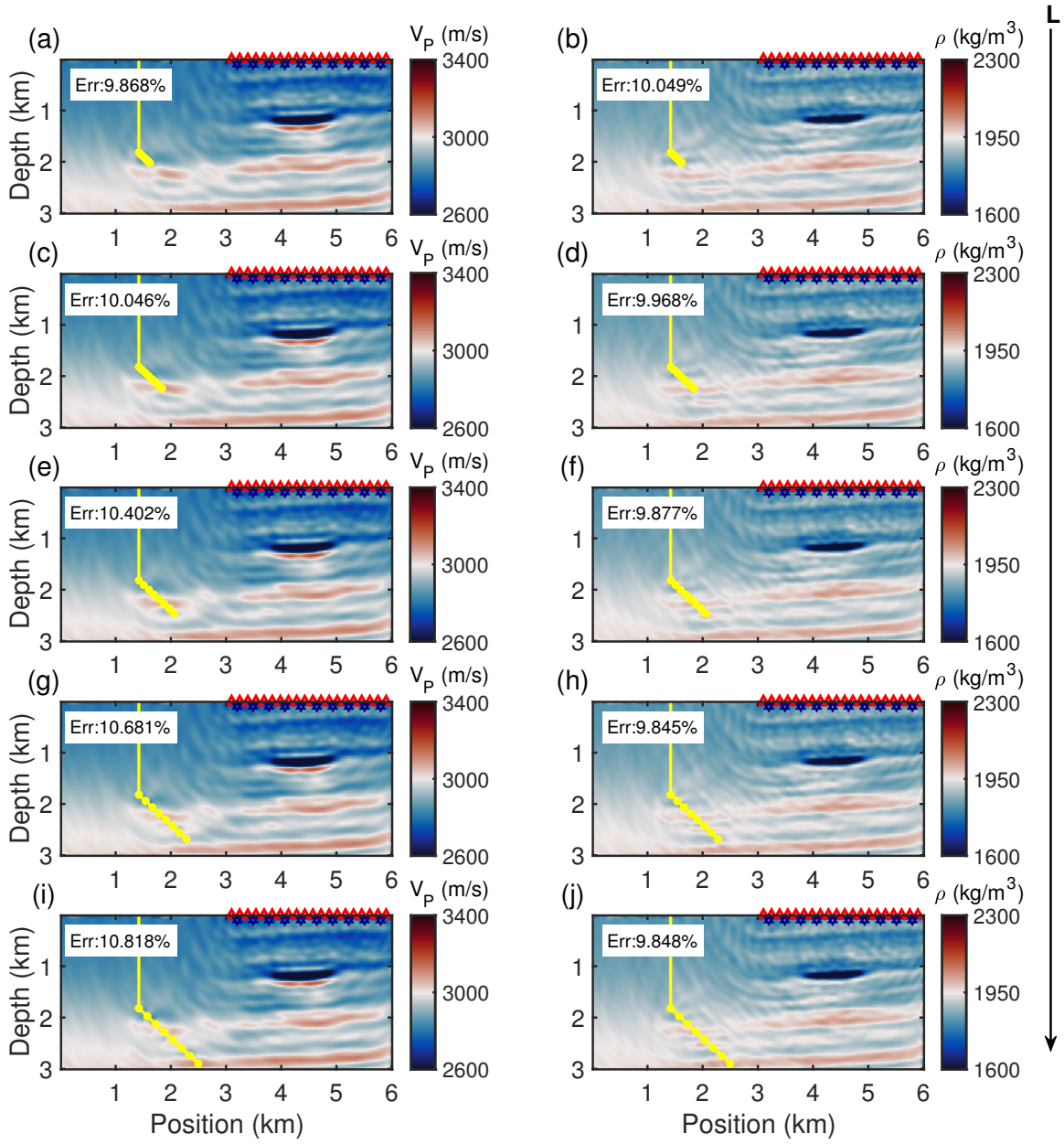

Figure 3.12: Inversion of V_P and ρ with surface and SWD acquisitions with different trajectory extensions. Drilling trajectory and discrete sources are schematically shown by the yellow line and markers. The black arrow denotes the gradual increment of L. (a), (b) inversion with $L = 450$ m. (c), (d) inversion with $L = 710$ m. (e), (f) inversion with $L = 975$ m. (g), (h) inversion with $L = 1250$ m. (i), (j) inversion with $L = 1500$ m.

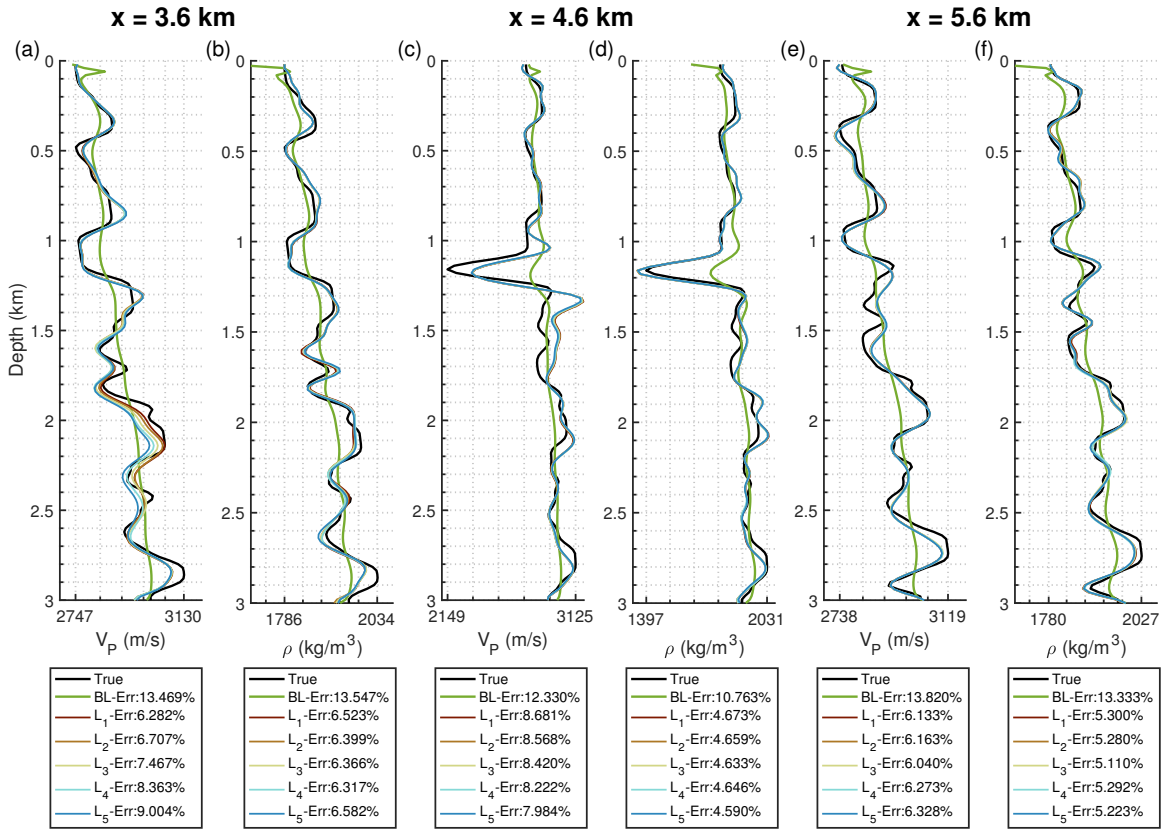

Figure 3.13: Profiles in trajectory extension test. (a), (c), and (e) are P-wave velocity cross-sections, while (b), (d), and (f) are density cross-sections. BL denotes baseline.

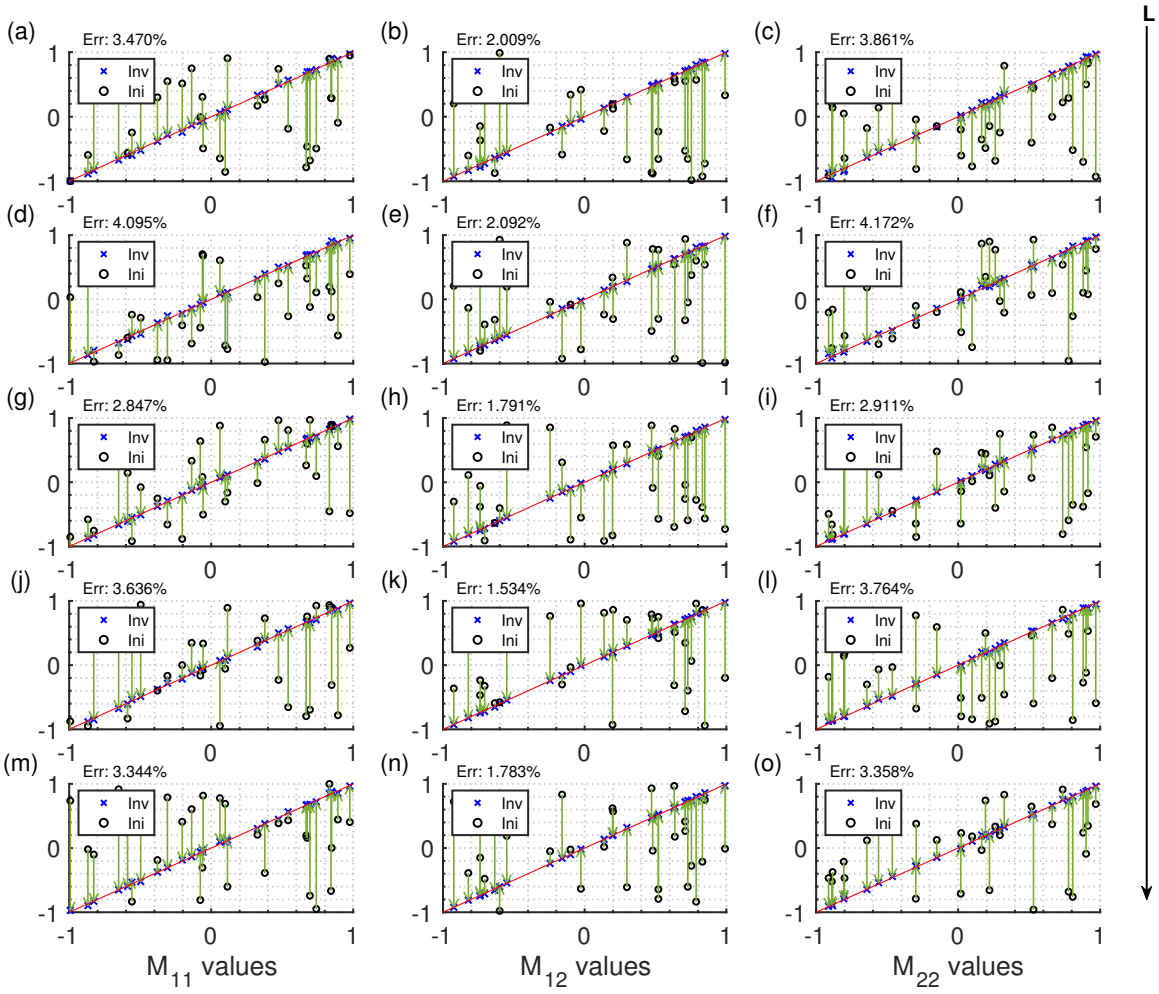

Figure 3.14: Inversion of moment tensors. The x-axis in each subplot represents the initial values, while the y-axis in each subplot represents the true values. The black arrow denotes the gradual increment of L, and the green arrows are from the initial data points (black circles) to the final data points (blue circles). The red line in cross plots represents the line of perfect agreement. (a)-(c) $L = 450$ m. (d)-(f) $L = 710$ m. (g)-(i) $L = 975$ m. (j)-(l) $L = 1250$ m. (m)-(o) $L = 1500$ m.

3.4.2 Simultaneous inversion for subsurface parameters, moment tensors and source positions

In some drilling programs, there are deviations between the planned and the actual wellbores because of the dip angle, hardness, and other formation or designing properties (Ma et al., 2016). In this section, I consider recovering P-wave velocity, density, source radiations, and positions with the above simultaneous inversion scheme in a more realistic situation. The estimated discrete source positions will be connected to depict the estimated drilling path.

The same synthetic models and major inversion target are kept in this experiment. I use a larger offset from the drilling site to surface sensors for the geometry to eliminate the potential mistakes from misupdated source positions. The initial well trajectory is shown by the purple line, while the true path deviates from it as the yellow line in Figure 3.15. This test assumes that the SWD sources are radiating independently from their positions. 30 discrete sources with unknown moment tensors and positions are used in this experiment.

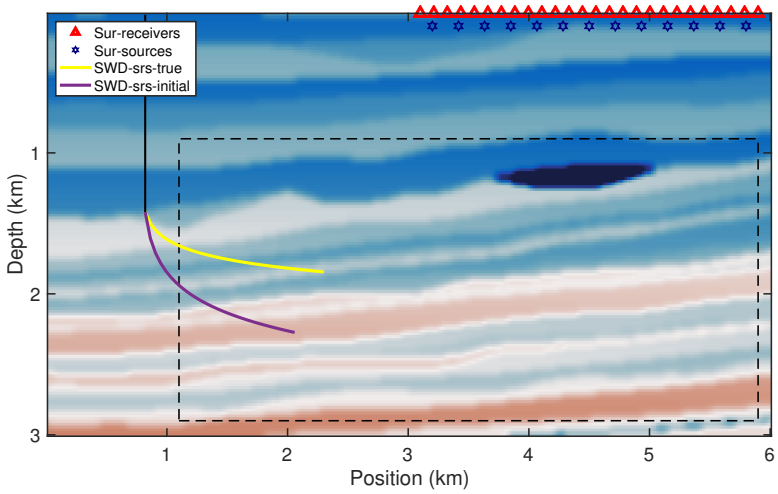

Figure 3.15: Schematic acquisition system considering source positions. Surface sources and receivers are displayed by blue and red markers, respectively. The true and initial trajectories are shown by the yellow and purple lines, respectively. The major inversion focus is indicated by the black dashed rectangle.

As shown in Figure 3.16, the inversion results for P-velocity and density models are also significantly enhanced with the intervention of SWD sources, suggesting the robustness of the FWI is kept despite the more nonlinear inverse problem with the involvement of unknown positions. Reliable recovery of moment tensors is shown in Figure 3.17. Figure 3.18 illustrates that the real drilling trajectories can also be depicted.

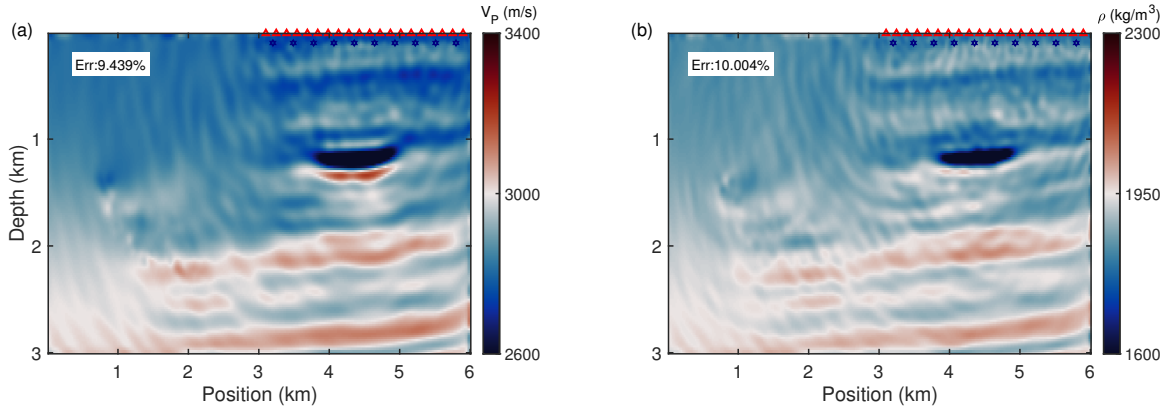

Figure 3.16: Inversion of V_P and ρ with surface and SWD acquisitions with SWD radiation and position unknowns. (a) V_P model. (b) ρ model.

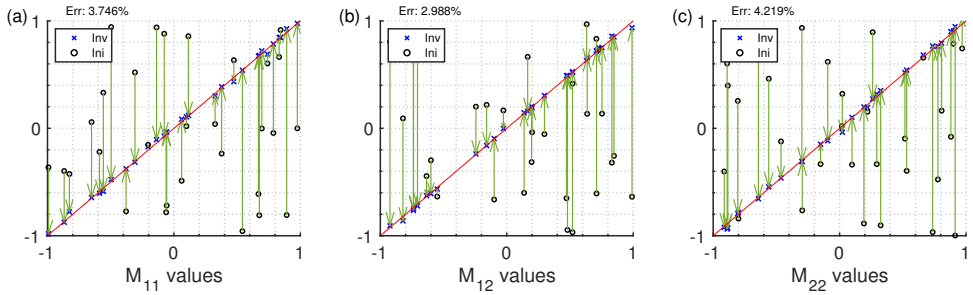

Figure 3.17: Inversion of moment tensors when positions join as source unknowns. The x-axis in each subplot represents the initial values, while the y-axis in each subplot represents the true values. The green arrows are from the initial data points (black circles) to the final data points (blue circles). The red line in cross plots represents the line of perfect agreement. (a) M_{11} results. (b) M_{12} results. (c) M_{22} results.

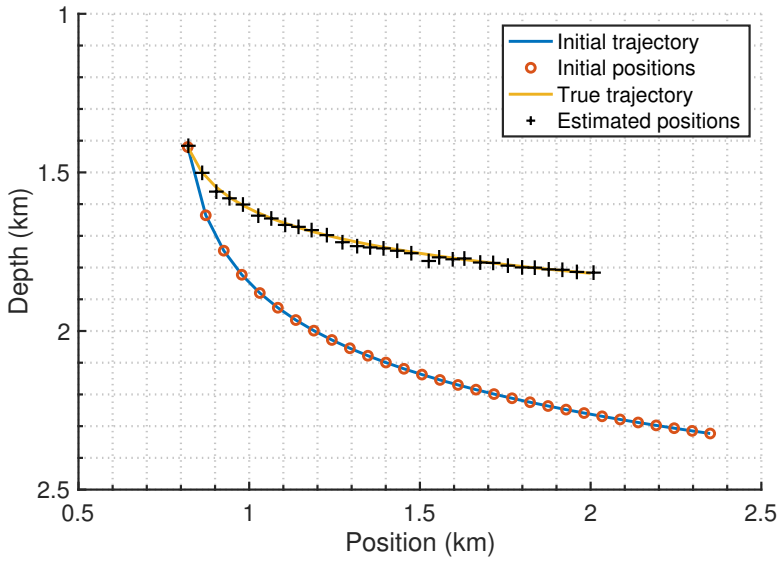

Figure 3.18: Inversion for SWD source positions. The blue line denotes the initial trajectory, the yellow line shows the true trajectory, and the red and black markers are the discrete initial and estimated source positions, respectively.

3.4.3 Inversion with frequency variations

The multi-band inversion investigated thus far involves random moment tensors with constant impulsive energy at each frequency, and in general the treatment of frequencies is necessarily somewhat idealized. The experiments in this section are designated to test the FWI performance in several scenarios where lower frequencies are missing. I use the most consistent acquisition in which there is a deviated trajectory with fixed inclination, extension, and source numbers (case 3 in the extension and inclination test, last case in the source number test). The varying parameter is the frequency spectrum. Figure 3.19 shows four frequency settings, where F_1 - F_4 denote weaker energy of the lower frequencies. I conduct the simultaneous inversion for P-wave velocity, density, and moment tensor values, while excluding source positions in the inversion variables to simplify this problem.

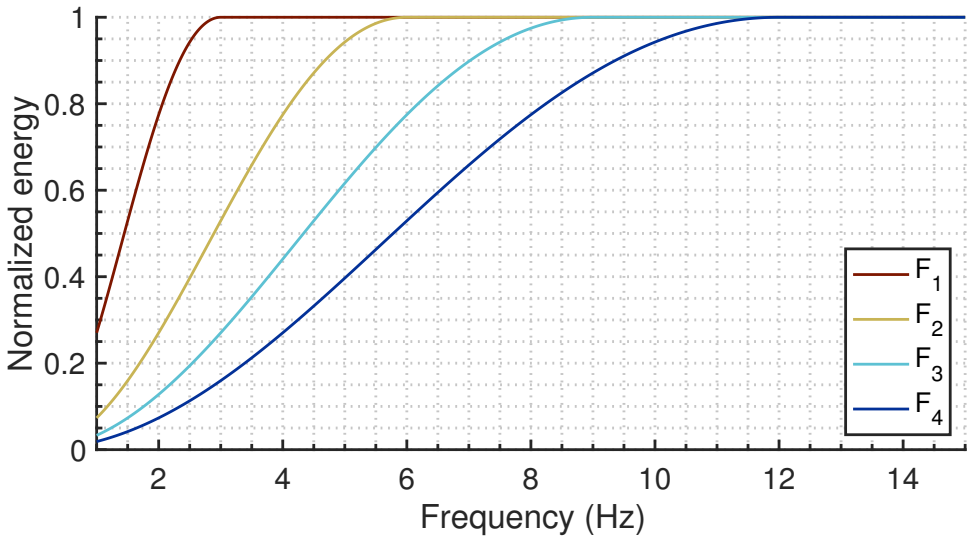

Figure 3.19: Frequency components in this section's tests.

As shown in Figure 3.20, poorer recovered models are acquired when the inversion is conducted with mainly relatively higher frequencies, but the shallower small region with low velocity and density is still characterized because of the increased illumination. Compared with the baseline inversion, which uses a sufficiently broad frequency range, the contribution from the SWD-generated datasets is self-evident. The results are gradually refined when more low-frequency components are included, although major errors cluster around unknown sources. The radiation inversions in Figure 3.21 show a similar correlation with the intervention of lower frequency components. This section suggests that, in practical applications, the SWD is also supportive even with a lack of low frequencies.

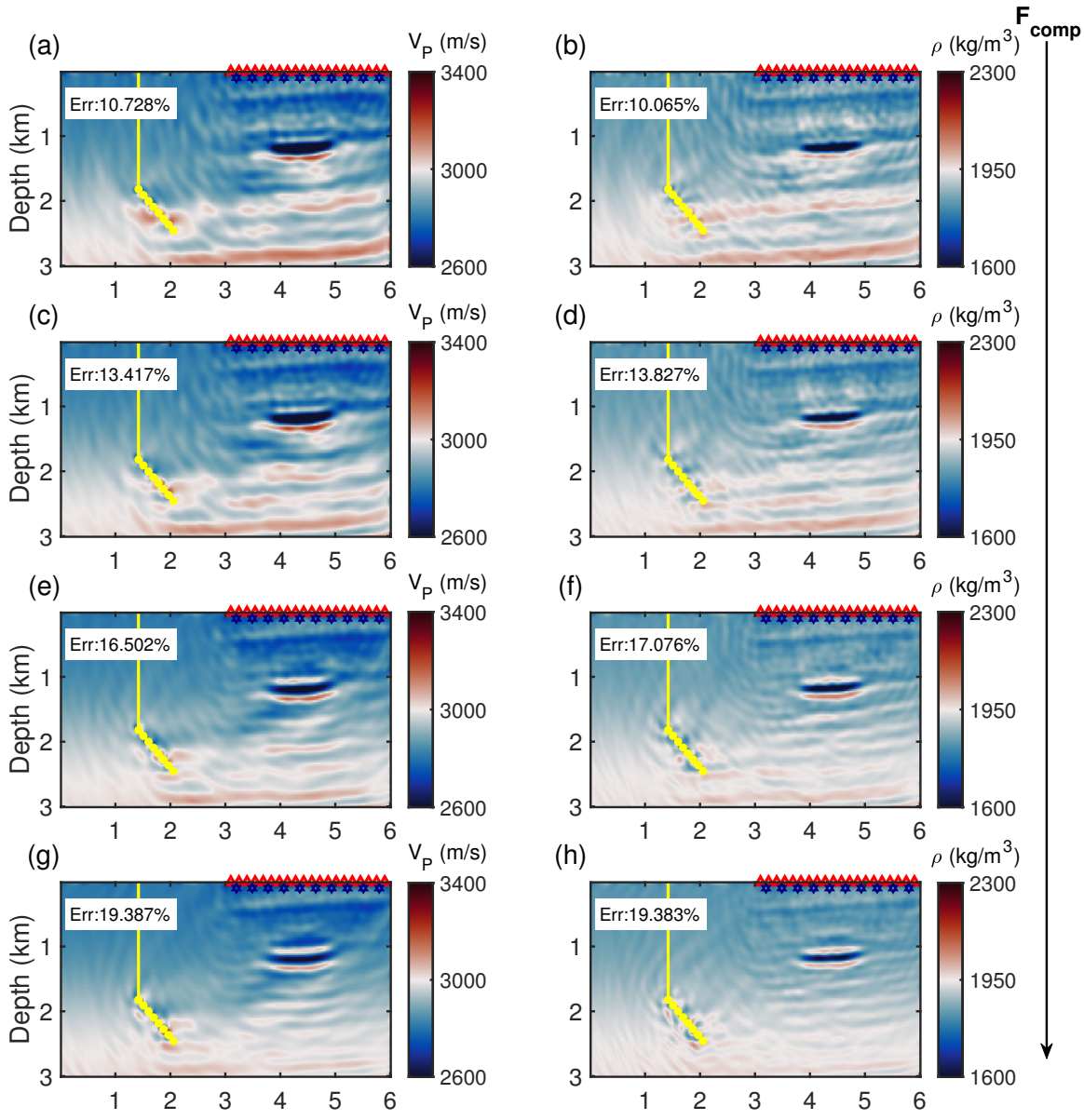

Figure 3.20: Inversion of V_P and ρ with surface and SWD acquisitions with different source spectra. Drilling trajectory and discrete sources are schematically shown by the yellow line and markers. The black arrow denotes spectra with less comprehensive low frequencies. (a) and (b) correspond to F_1 in Figure 3.19. (c) and (d) correspond to F_2 in Figure 3.19. (e) and (f) correspond to F_3 in Figure 3.19. (g) and (h) correspond to F_4 in Figure 3.19.

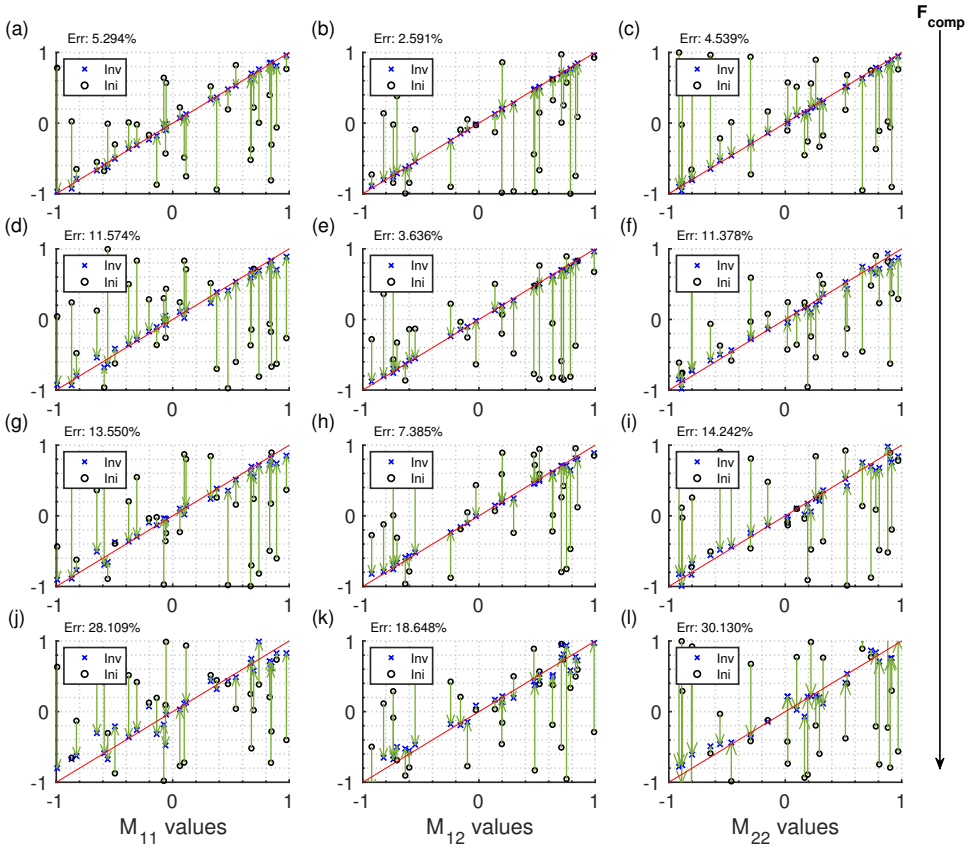

Figure 3.21: Inversion of moment tensors with different frequency datasets. The x-axis in each subplot represents the initial values, while the y-axis in each subplot represents the true values. The black arrow denotes spectra with less comprehensive low frequencies, and the green arrows are from the initial data points (black circles) to the final data points (blue circles). The red line in cross plots represents the line of perfect agreement. (a)-(c) correspond to F_1 in Figure 3.19. (d)-(f) correspond to F_2 in Figure 3.19. (g)-(i) correspond to F_3 in Figure 3.19. (j)-(l) correspond to F_4 in Figure 3.19.

3.5 Discussion

The synthetic examples show that a simultaneous inversion for recovering the subsurface models and unknown source terms is feasible in a surface + SWD acquisition setting. The

simulations allow the impact of the additional ray paths to be understood and analyzed by comparing the FWI results with surface-only and surface-well acquisitions. In this SWD-FWI algorithm, I mitigate the nonlinearity by assuming known isotropic source radiation in the conventional acquisition, thus reducing the number of inversion variables, which helps the inversion of V_P and ρ. More accurate V_P and ρ models can be obtained with explosive surface sources with generally radiating SWD sources. The drilling path can also be depicted with the estimation of discrete sources via this inversion approach, which suggests the potential in managing the drilling programs.

As illustrated in Figure 3.4, Figure 3.8, and Figure 3.12, the upper-middle portion of the models displays well-resolved features of low-velocity and low-density lenses across most examples. However, the lower layer continuity with higher values exhibits more distinguishable variations, particularly in the angle test section. To investigate this phenomenon, a ray-tracing test was conducted on the model. Results in Figure 3.22 (a) indicate that when the angle is fixed at 45°, the ray paths from the subsurface sources to the ground cannot effectively convey information about that layer, suggesting that a variation of source numbers or drill path extension may not be effective. However, Figure 3.22 (b) demonstrates that high-velocity and high-density layers can cause some ray paths to bend upwards and become detectable, which helps to reduce the discontinuity of those layers by providing the inversion with information about the high-velocity and density layers between $3 - 4$ km. The ray-tracing outcomes imply that underground models may play a significant role in implementing SWD-FWI. It is worth noting that while ray-tracing provides only first-order information about the wavefield and applies a high-frequency approximation to demonstrate the paths the energy follows, the FWI uses full waveform information that includes both first and higher-order reflections, also the lower-frequency components. This explains why the model portions not covered by the ray paths are also well recovered.

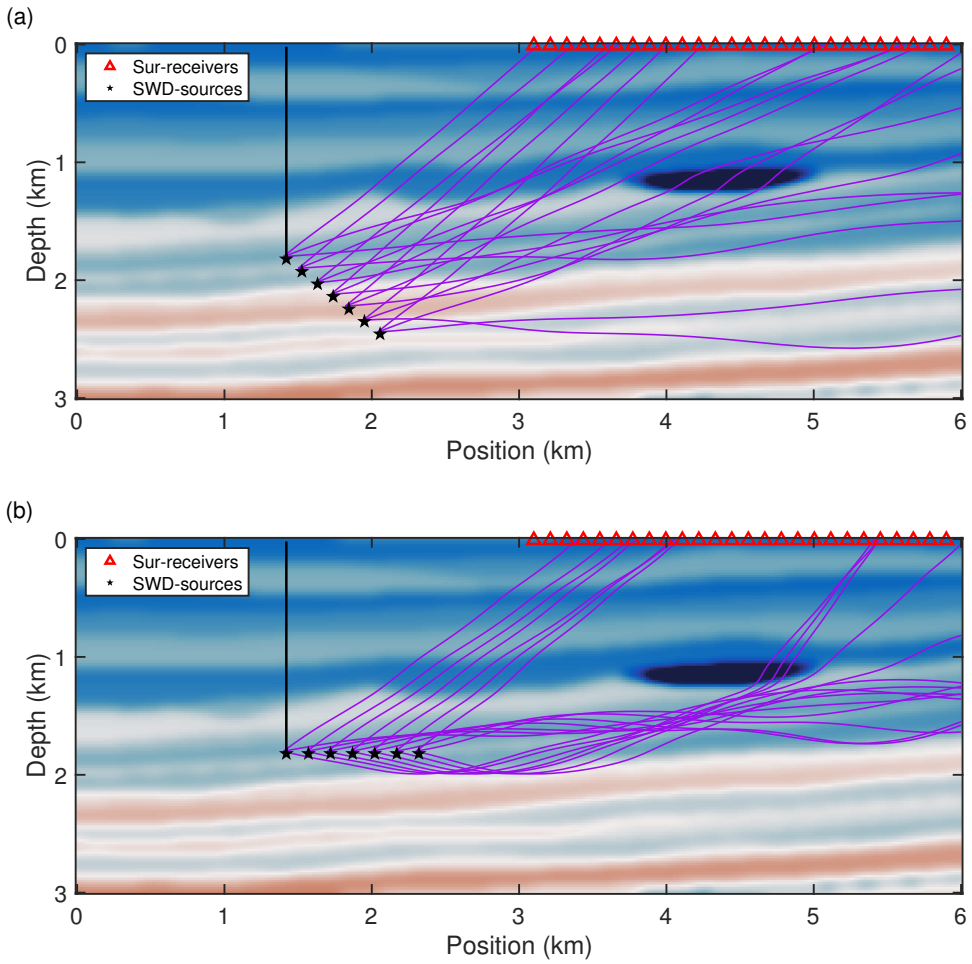

Figure 3.22: Ray-tracing with different SWD settings. (a) $\theta = 45°$. (b) $\theta = 90°$.

While it is possible to formulate a time-domain FWI scheme, I made a thoughtful decision to prioritize the frequency-domain approach. This choice stems from its notable compatibility with multi-parameter inversion and its demonstrated effectiveness in handling vectorized parameters, as evidenced in previous studies (e.g., Métivier et al., 2013). Moreover, for our specific application involving SWD datasets, the frequency-domain approach brings inherent convenience by seamlessly integrating data with diverse frequency components into the inversion process, offering a significant advantage for our research.

I believe in this formulation that the critical elements needed to explain the SWD signature and begin to use it are in place. This study also suggests some important directions

for further research. First, although I have not allowed it to be a free unknown yet, this framework can estimate source-related unknowns, thus enabling SWD-FWI technology to help refine drill position estimates. Moment tensor decomposition can also be applied to analyze the mechanisms of the bit-rock interaction if force couples can describe the discrete SWD sources. Such inversion has the potential to help the ahead-of-the-bit estimation in SWD, though computational speed and expense would need to be reduced for this to be practically realized. Another challenge is the source frequency spectrum in this problem. The dominant frequencies of SWD sources with the bit penetration could be higher than the typical frequency band in FWI, and there are some related issues, as previously discussed, to be considered in this simultaneous inversion approach since it is currently conducted with lower frequencies. However, it is possible to design a more realistic representation considering both the drill-bit-rock mechanism with moment tensors and frequency dependence. Promisingly, this methodology can be potentially used for jointly inverting active surface seismic and passive microseismic data, beyond just SWD. For instance, previous studies (Wagner et al., 2007; Vesnaver et al., 2010; Goertz-Allmann et al., 2017) have introduced joint inversion methods of multiple data types to enhance subsurface imaging. Hence, this adapted FWI scheme can be a valuable tool for improving the imaging of complex geologic structures and identifying hydrocarbon reservoirs in the subsurface.

3.6 Conclusions

In this chapter, I explore the potential of taking advantage of the SWD data to compensate for the incomplete surface acquisition in simultaneous FWI. Numerical examples demonstrate that the additional ray paths provided in an SWD dataset help to provide a better FWI result. The inclusion of SWD data improves the inversion of elastic properties, and this leads to conclude that SWD data offers the potential to enhance inversion results. Besides, the radiations and positions of the underground sources can also be precisely depicted,

which shows a promising possibility that the drilling application can be monitored while implementing its datasets to the FWI. Further research is still required to provide more comprehensive conclusions, especially concerning a more precise moment tensor representation of the drill-bit-rock interaction mechanisms and an advanced inversion strategy that will fit a more practical case. Additionally, the source signature of various types of drill bits should be quantified to detail the P and S-wave components.

Chapter 4

Uncertainty quantification in time-lapse full waveform inversion of vertical seismic profiling data: contrasting Hamiltonian Monte Carlo and Stein Variational Gradient Descent methods

4.1 Summary

Time-lapse FWI is increasingly important in energy-transition applications, particularly for monitoring CO_2 geological storage. Due to the nonlinearity of the FWI problem, the presence of diverse sources of uncertainty between baseline and monitor surveys, and the typically sparse or variable data coverage in practical deployments, incorporating robust uncertainty quantification strategies is essential. Two promising approaches are HMC and SVGD, each offering distinct advantages. HMC leverages Hamiltonian dynamics to explore the model space more efficiently than traditional Monte Carlo Markov Chain (MCMC) methods, while SVGD approximates the posterior distribution by iteratively updating a finite set of particles to minimize the KL divergence. In this chapter, I conduct a feasibility analysis of 2D TL-FWI using VSP acquisition, comparing HMC and SVGD under equivalent computational budgets. The results indicate that, while both approaches provide meaningful uncertainty quantification, they exhibit distinct trade-offs. HMC produces more global posterior estimates and potentially more comprehensive model updates, but the risk of accepting erroneous model parameters in poorly illuminated regions increases. This challenge can further be exacerbated in TL-FWI scenarios. In contrast, SVGD, although sensitive to survey non-repeatability due to its deterministic nature, achieves faster convergence and better scalability, making it

well-suited for rapid or large-scale uncertainty assessments. The findings underscore the importance of aligning algorithmic choice with inversion goals and computational constraints, especially in real-world TL-FWI monitoring applications.

4.2 Introduction

The effectiveness of FWI is hindered by several inherent complexities. As explored in previous chapters, uncertainties in FWI arise from various sources, including sparse and noisy observational data, errors in acquisition, and limitations in forward modeling theories (Parker, 1977). These factors contribute to a fundamental ambiguity: multiple sets of model parameters can produce similar seismic responses, leading to non-uniqueness in the inversion results. This issue reflects the ill-posed nature of FWI, where subtle variations in subsurface properties may not be distinguishable given the available data, complicating the interpretation of the estimated models through FWI.

The complexity of FWI further escalates in time-lapse studies, where consecutive inversions are conducted on baseline and monitor datasets to track changes in subsurface properties over time. These surveys introduce additional challenges due to inconsistencies in data acquisition, such as non-repeatability caused by environmental factors or equipment variations, and differences in the geometry of surveys conducted at different times (Mosegaard and Tarantola, 1995; Kotsi et al., 2020). Time-lapse FWI also inherits the uncertainties inherent in individual inversions, compounding the difficulties of building reliable models of temporal change. In such scenarios, it becomes crucial to move beyond a single "best" model by assessing a spectrum of plausible solutions that fit the data—known as the ensemble of solutions. Properly characterizing this ensemble requires strategies for quantifying the uncertainties in FWI models and identifying the range of possible subsurface configurations that align with the data (Backus and Gilbert, 1967, 1968).

Given the inherent uncertainties in FWI, probabilistic frameworks are important for cap-

turing the full range of potential solutions. MC-based methods represent one of the most widely used approaches for uncertainty quantification in FWI. However, the computational cost of this realm of methods becomes prohibitive for large-scale problems, particularly in time-lapse FWI, where two full sampling processes—one for each dataset—must be conducted (Bellman, 2003). The scalability challenge limits the direct application of traditional MC methods in practical 3D or time-lapse studies, necessitating more efficient alternatives. A promising alternative is the HMC method (Duane et al., 1987), which combines gradient-based optimization with sampling techniques. HMC leverages Hamiltonian dynamics (Hamilton, 1834) to explore the model space efficiently, and reduces the chances of getting trapped in local minima, facilitating a more targeted exploration of the parameter space (Neal, 1993; Brooks et al., 2011). This makes it well-suited for FWI applications, where complex, high-dimensional parameter spaces are common. More discussions of HMC in FWI can be found in Fichtner et al. (2018); Gebraad et al. (2020); Kotsi et al. (2020); de Lima et al. (2023, 2024).

In addition to MC-based methods, variational inference techniques offer another path for uncertainty quantification. These methods reformulate the Bayesian problem as an optimization task, where the goal is to approximate the posterior distribution of the model parameters with a simpler, tractable distribution (Jordan et al., 1998). This transformation reduces computational demands, making variational inference particularly attractive for large-scale FWI problems where traditional sampling methods are impractical. Variational inference methods aim to minimize the difference between the true posterior and the approximating distribution, typically measured using divergence metrics such as the KL divergence (Blei et al., 2017). Among various methods for minimizing the KL divergence, the SVGD method has gained prominence in geophysical applications. Unlike other methods that require explicit analytical transformations between distributions, SVGD uses kernel-based transformations to iteratively adjust model parameters toward regions of high

posterior probability (Liu and Wang, 2019). This property makes SVGD flexible and well-suited for FWI, where the analytical forms of posterior distributions are often unavailable. Applications of SVGD in geophysical studies include works by Zhang et al. (2018); Nawaz and Curtis (2018, 2019); Zhang and Curtis (2020); Zhang et al. (2023), demonstrating its potential in handling complex inversion problems with improved efficiency and scalability.

While there is a comparison between these two methods in elastic FWI with transmissive acquisition (e.g., (Zhang and Curtis, 2020; Gebraad et al., 2020)), it is generally difficult to make broad claims about the comparative strengths of methods like HMC and SVGD, case-specific synthetic experiments can reveal meaningful differences in their behavior. I assume that, given a sufficiently well-defined setup, lessons can be drawn that are relevant not only to the tested scenario but also to others of similar structure. The problem of time-lapse VSP-FWI for CO_2 plume monitoring provides such a context. de Lima et al. (2024) have demonstrated the use of HMC for uncertainty assessment in seismic monitoring with FWI, while Zhang and Curtis (2024) have explored stochastic SVGD for monitoring applications. Related Bayesian approaches include physics-structured variational inference that shows promise for 4D FWI (Zhao and Curtis, 2024), and sparse seismic survey design (Zhao and Curtis, 2025). In this chapter, I first briefly review the theoretical foundations of Bayesian FWI, HMC, and SVGD, emphasizing their role in probabilistic inversion. I then design synthetic experiments to evaluate HMC and SVGD within a near-parallel acoustic TL-FWI framework. The setup simulates CO_2 injection using a permanent VSP receiver array and varying surface source geometries between baseline and monitor surveys. Varying levels of noise are then added to both datasets to further amplify the non-repeatability in TL-FWI. The results include a comparative analysis of the two methods in terms of uncertainty structure, convergence behavior, and sensitivity to acquisition geometry.

4.3 Theory

4.3.1 Pseudo-code for HMC- and SVGD-based FWI

The core steps involved in both HMC and SVGD have been outlined in Chapter 2. Building on those descriptions, I summarize the workflows for these methods in the following algorithms. All relevant formulations and theoretical foundations are also discussed in detail in Chapter 2.

Algorithm 1 provides a pseudo-code representation of a single iteration of the HMC algorithm in FWI, with the input data denoted as $\mathbf{d}$. The current and proposed states are indicated by the subscripts *cur* and *new*, respectively. The gradient of the potential energy is calculated by the adjoint method, which is also presented in Chapter 2.

Algorithm 1: HMC iterations

for $M = 1$ to nsamples **do**
  Randomly generate $\mathbf{p}_{cur} \sim \mathcal{N}(0, \mathbf{M}_{mass})$;
  $\mathbf{p}_{new} = \mathbf{p}_{cur}$
  $\mathbf{m}_{new} = \mathbf{m}_{cur}$
  $U(\mathbf{m}_{cur}) = -logP_{m_{cur}}(\mathbf{m}_{cur}|\mathbf{d})$
  $K(\mathbf{p}_{cur}) = \frac{\mathbf{p}_{cur}^T \mathbf{M}_{mass}^{-1} \mathbf{p}_{cur}}{2}$
  $H(\mathbf{p}_{cur}, \mathbf{m}_{cur}) = U(\mathbf{m}_{cur}) + K(\mathbf{p}_{cur})$
  $\mathbf{p}_{new} = \mathbf{p}_{new} - \delta t \frac{\nabla U(\mathbf{m}_{new})}{2}$
  for $N = 1$ to Leapfrog iterations - 1 **do**
    $\mathbf{m}_{new} = \mathbf{m}_{new} + \delta t \nabla K(\mathbf{p}_{new})$
    $\mathbf{p}_{new} = \mathbf{p}_{new} - \delta t \nabla U(\mathbf{m}_{new})$
  end for
  $\mathbf{m}_{new} = \mathbf{m}_{new} + \delta t \nabla K(\mathbf{p}_{new})$
  $\mathbf{p}_{new} = \mathbf{p}_{new} - \delta t \frac{\nabla U(\mathbf{m}_{new})}{2}$
  $\mathbf{p}_{new} = -\mathbf{p}_{new}$
  $U(\mathbf{m}_{new}) = -logP_{m_{new}}(\mathbf{m}_{new}|\mathbf{d})$
  $K(\mathbf{p}_{new}) = \frac{\mathbf{p}_{new}^T \mathbf{M}_{mass}^{-1} \mathbf{p}_{new}}{2}$
  $H(\mathbf{p}_{new}, \mathbf{m}_{new}) = U(\mathbf{m}_{new}) + K(\mathbf{p}_{new})$
  if randnum $\leq exp(-(H_{new} - H_{cur}))$ **then**
    $\mathbf{m}_{cur} = \mathbf{m}_{new}$
  else
    $\mathbf{m}_{cur} = \mathbf{m}_{cur}$
  end if
end for

In practice, the distribution is approximated using a collection of particles. I summarize this method in Algorithm 2 below. The gradient of the likelihood function is also acquired by the adjoint method.

Algorithm 2: Variational Inference with Stein Variational Gradient Descent

1: Draw a set of N particles $\mathbf{m}^0 = \left[\mathbf{m}_j^0\right]_{j=1}^N \in q_0(\mathbf{m})$, where $q_0(\mathbf{m})$ is the initial distribution;
2: **for** $i = 1$ to iterations **do**
3: $\quad \Phi_i^*(\mathbf{m}) = \frac{1}{N}\sum_{j=1}^N \left[k\left(\mathbf{m}_j^i, \mathbf{m}\right)\nabla_{\mathbf{m}_j^i} log\rho\left(\mathbf{m}_j^i|\mathbf{d}\right) + \nabla_{\mathbf{m}_j^i} k\left(\mathbf{m}_j^i, \mathbf{m}\right)\right]$
4: $\quad \mathbf{m}_j^{i+1} = \mathbf{m}_j^i + \alpha^i V_i^*\left(\mathbf{m}_j^i\right)$
5: **end for**

Generally, parameters have natural constraints in many geophysical inverse problems. Those constraints complicate the SVGD optimization processes because the algorithms must respect these boundaries, which can be challenging. Thus, I encode parameters to an unconstrained space by equation 4.1, and transfer these back after the SVGD process using equation 4.2, where i denotes the parameter index, and a_i and b_i are the lower and upper boundaries on model parameters m_i. An example can be seen in Figure 4.1.

$$m_i^T = T(m_i) = log(m_i - a_i) - log(b_i - m_i), \tag{4.1}$$

$$m_i = T^{-1}\left(m_i^T\right) = a_i + \frac{(b_i - a_i)}{1 + exp\left(-m_i^T\right)}, \tag{4.2}$$

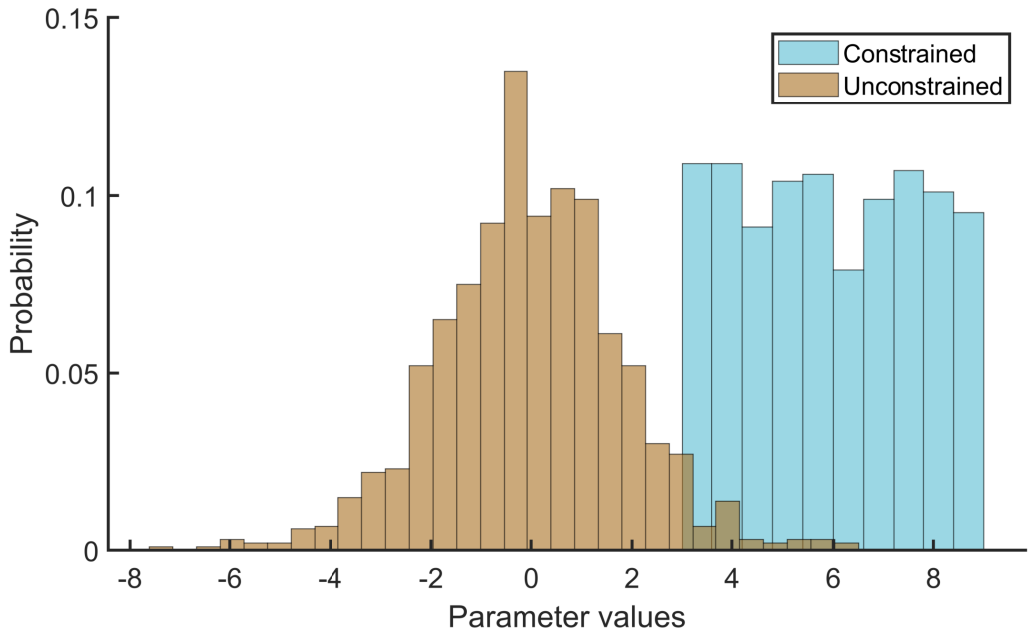

Figure 4.1: An example of the constrained and unconstrained parameter space.

4.4 Numerical Experiments

4.4.1 Configurations and workflow

In this section, I present several acoustic time-lapse FWI experiments conducted in the frequency domain using HMC and SVGD methods with comparable computational workload, which is set to 1 day on the same workstation. The experiments focus on estimating P-wave velocity variations in a synthetic model both before and after the CO_2 injection (refer to Figure 4.2). The horizontal and vertical grids contain 300 and 100 points, respectively, with a grid interval of 10 meters. Across all numerical tests, I deploy 40 permanent VSP receivers in the middle of the model, simulating the observation well for CO_2 sequestration. For the baseline inversion, there are 80 explosive sources separated by 4 grid intervals on the surface. The number of sources is halved in the monitor inversion. The synthetic datasets

are discretized into a set of 8 frequencies, ranging from 5 to 25 Hz. I adopt a parallel time-lapse inversion strategy in this research, thus, as illustrated in Figure 4.3, the initial models for all HMC-FWI and SVGD-FWI tests are selected from a uniform distribution configured from a smoothed version of the true baseline model. The upper and lower boundaries of the uniform distribution are set to be ± 600 m/s.

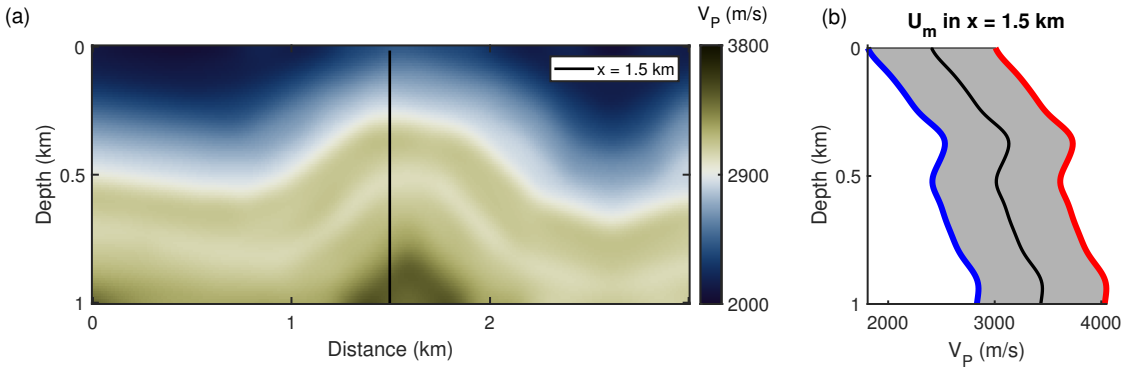

Figure 4.2: P-wave velocity models and schematic acquisition geometry. The source layout is represented by white stars, and the observation well is shown by the black line with receivers indicated by red triangles. (a) true baseline model. (b) true monitor model.

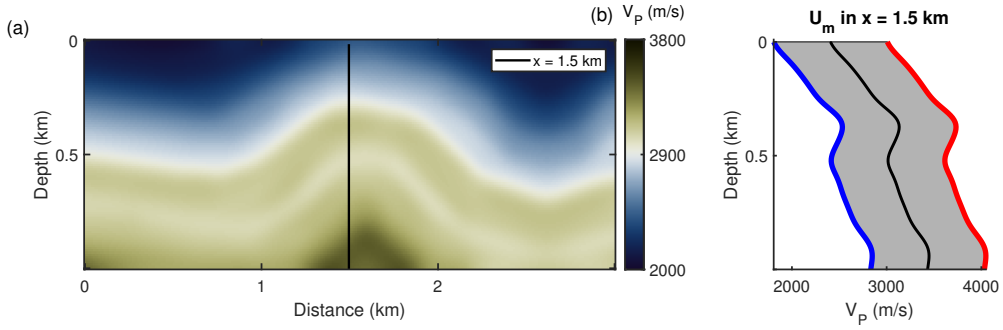

Figure 4.3: The initial model for baseline and monitor surveys. (a) Smoothed baseline model. (b) An example of the range of the uniform distribution.

The workflow consists of two main phases. In phase 1, the baseline inversion is performed using both HMC and SVGD under consistent parameter settings but with the surface source

coverage reduced by half to simulate sparser monitoring conditions. Phase 2 involves inversions using the same methods but different noise levels added in baseline and monitor datasets to enhance the non-repeatability. Noise with SNR of 8 is applied to the baseline data and an SNR of 6 to the monitor data. To address the scale dependency of standard deviation in uncertainty visualization and to enable a fair comparison between the baseline and monitor inversion results, I use the coefficient of variation (CV), which normalizes the standard deviation $\sigma(\mathbf{m})$ by the corresponding mean model $\overline{\mathbf{m}}$ as below:

$$CV(\mathbf{m}) = \frac{\sigma(\mathbf{m})}{\overline{\mathbf{m}}}. \tag{4.3}$$

CV values are presented specifically when comparing the baseline and monitor inversions. For the time-lapse results, however, I present the standard deviation directly, as it more effectively highlights the absolute uncertainty in the estimated time-lapse changes. This relative metric, used for both baseline and monitor results, is positively correlated with the level of uncertainty. The time-lapse differences are then analyzed using the model mean and standard deviation of the changes, expressed as $\overline{\mathbf{m}_m - \overline{\mathbf{m}}_b}$ and $\sigma(\mathbf{m}_m - \overline{\mathbf{m}}_b)$, respectively.

4.4.2 Time-lapse FWI with HMC

Consistent with the adaptive tuning strategy (Fichtner et al., 2018) in HMC, I employ dynamic integration lengths, randomly selected from the range of 9 to 15. I initialize the starting time step, denoted as δt, to be twice the maximum value within the model vector, scaled by dimension$^{-1/4}$ multiplied by the iteration number in the Leapfrog method. The adjustment of the time step is contingent on the acceptance rate within the sampling subset. If the acceptance rate falls below 45%, I decrease δt by a factor of 0.8. Conversely, if the acceptance rate within a subset exceeds 75%, I increase the time step by the reciprocal of 0.8. The mass matrix in HMC is set to be the identity matrix.

In Figure 4.4, I present the HMC-FWI results for both the baseline and monitor inversions under noise-free conditions. A total of $54,322$ samples are generated for the baseline, with an

acceptance rate of 67.04%, resulting in 36, 417 accepted samples. For the monitor inversion, which uses half the number of sources and thus is faster, 77, 596 samples are generated with a slightly lower acceptance rate of 65.21%, producing 50, 600 accepted samples. The posterior means of the baseline and monitor models are shown in panels (a) and (d), while the corresponding uncertainty estimates at three reference points are displayed in panels (b) and (e). These reference points include one near the VSP array and two located farther away (a shallow and a deep point; see Figure 4.2). The full posterior distributions at these locations are given in panels (c) and (f). The results highlight a dependence of posterior behavior on illumination. The first two points from both the baseline and monitor inversions exhibit near-Gaussian distributions with mild skewness. In contrast, multi-modal patterns are presented in the deeper point (purple star) where the illumination is limited. This indicates the low reliability of the uncertainty estimation of such poorly illuminated areas.

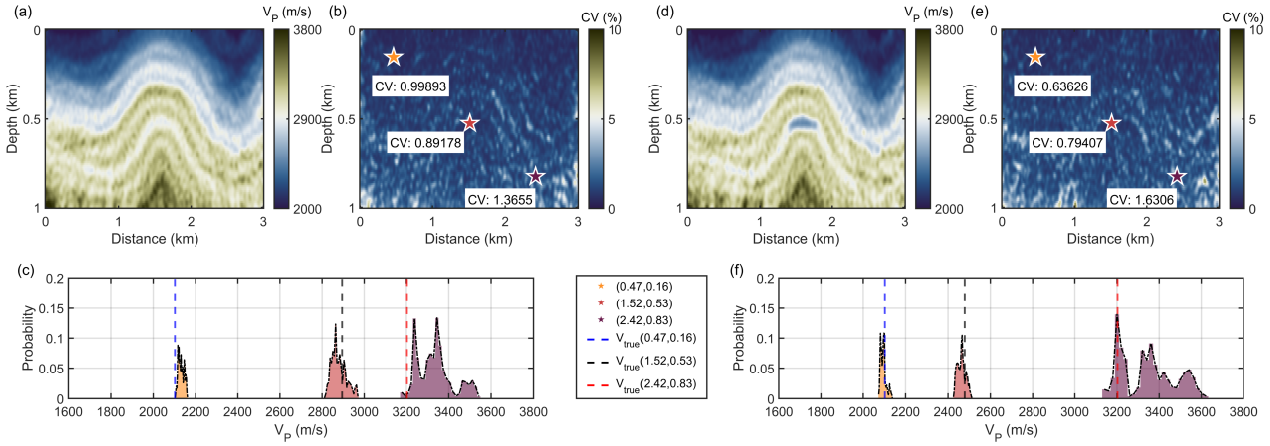

Figure 4.4: Baseline and monitor HMC-FWI results with noise-free data. (a) and (d) are baseline and monitor model means $\overline{\mathbf{m}}_b$ and $\overline{\mathbf{m}}_m$. (b) and (e) are coefficient of variation plots of the baseline and monitor models. (c) and (f) are histograms of the posterior distribution of the model parameters on three reference points shown by the three stars in (b) and (e). The three dashed lines denote the true values.

I present the time-lapse results in Figure 4.5. As shown in panel (a), the time-lapse mean

model successfully captures the P-wave velocity changes in the injection area, indicating good resolution of the velocity variation. Panel (b) displays the standard deviation of the changes in time lapse, calculated as

$$\sigma(\mathbf{m}_{TL}) = sqrt[variance(\mathbf{m}_m) + variance(\mathbf{m}_b)]. \tag{4.4}$$

In interpreting these results, I assume independence between the baseline and monitor ensembles, consistent with the parallel inversion strategy adopted here. Figure 4.5(c) displays the posterior distributions of the time-lapse velocity differences at the selected reference points. For the two points closer to the VSP array, the distributions deviate only slightly from Gaussianity, with mild skewness, and thus provide reasonable uncertainty estimates. In contrast, the distribution at the deeper purple point exhibits pronounced multi-modality, reflecting the compounded effects of limited illumination and differencing between baseline and monitor inversions. This behavior highlights that uncertainty quantification becomes less reliable in poorly illuminated regions, and that the degradation is amplified when assessing time-lapse changes.

I then present the time-lapse results in Figure 4.5. As shown in Figure 4.5 (a), the time-lapse mean model successfully captures the P-wave velocity changes in the injection area, indicating good resolution of the velocity variation. Figure 4.5 (b) displays the standard deviation of the changes in time lapse, calculated as $sqrt[variance(\mathbf{m}_m) + variance(\mathbf{m}_b)]$, assuming independence between the baseline and monitor samples due to the parallel inversion strategy, which results in $covariance(\mathbf{m}_b, \mathbf{m}_m) = 0$. Additionally, Figure 4.5 (c) shows the distributions of time-lapse differences at the selected locations. These distributions exhibit slight skewness from a standard Gaussian shape but overall reflect the underlying uncertainty estimates.

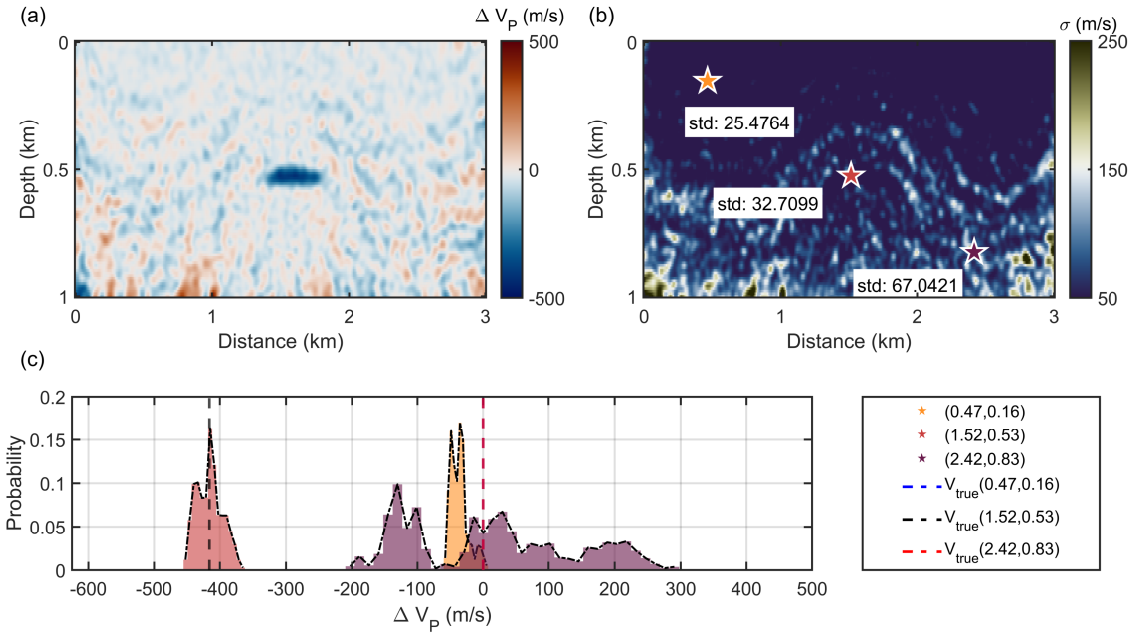

Figure 4.5: Time-lapse HMC-FWI results with noise-free data. (a) and (b) are mean value and standard deviation plots of the time-lapse variation, and (c) shows the posterior distributions of the reference points, with the true time-lapse changes denoted by the three dashed lines. Note that the blue and red lines are overlaying each other as they both represent 0.

Three vertical profiles across the reference points are shown in Figure 4.6. In general, higher uncertainties are observed in the deeper sections of the model, particularly in the first and third profiles, primarily due to insufficient illumination from the VSP acquisition in those regions. Despite this, the histograms along all three profiles align well with the true time-lapse changes, indicating a generally accurate estimation of the subsurface variations.

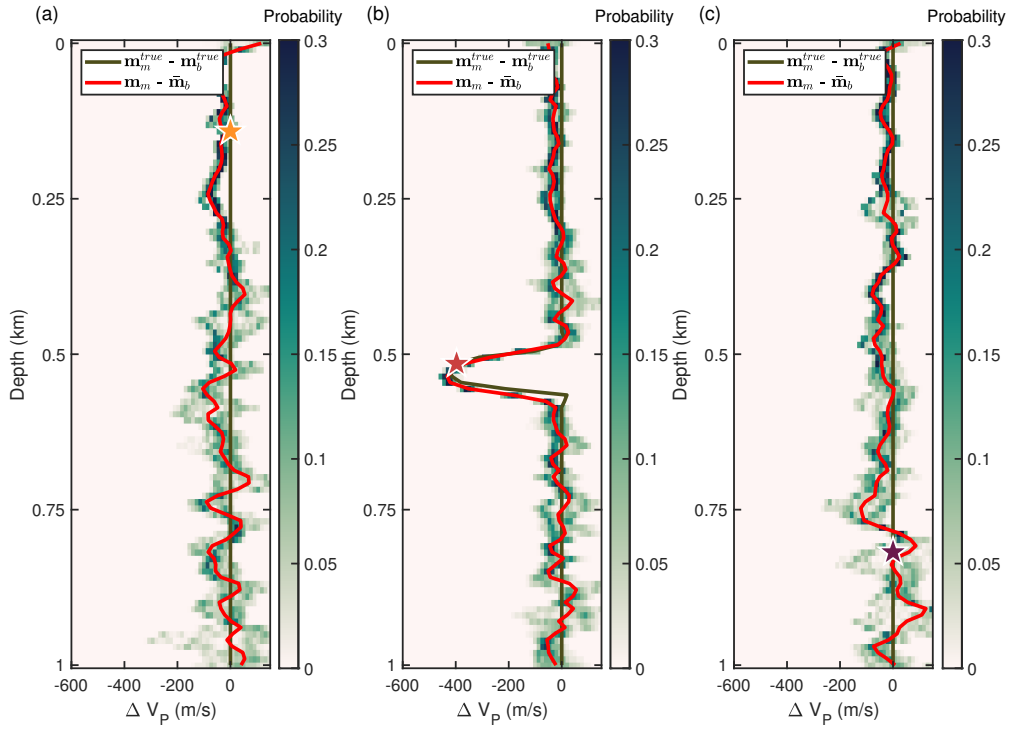

Figure 4.6: HMC time-lapse variation profiles with noise-free data. (a) to (c) are profiles across the three reference points, with their locations shown in Figure 4.5. The dark-green line is the true time-lapse change, the red line is the time-lapse model mean, and colors along the profile denote probability values.

Next, I evaluate the performance of HMC-FWI using noisy datasets. With the same computational budget as in the noise-free case, the total number of sample attempts remains comparable. However, due to the added noise, the acceptance rates decrease to 59.92% for the baseline and 58.44% for the monitor. Figure 4.7 presents the HMC-FWI results under the noisy condition. Compared to the noise-free case, the mean models display increased artifacts, and the associated uncertainties are higher. However, the estimated model means still show good agreement with the true models, demonstrating the robustness of the inversion framework. The posterior distribution at the bottom-right reference point becomes skewed and exhibits a multi-modal structure, especially in the monitor inversion, reflecting the compounded challenges of noise and acquisition non-repeatability.

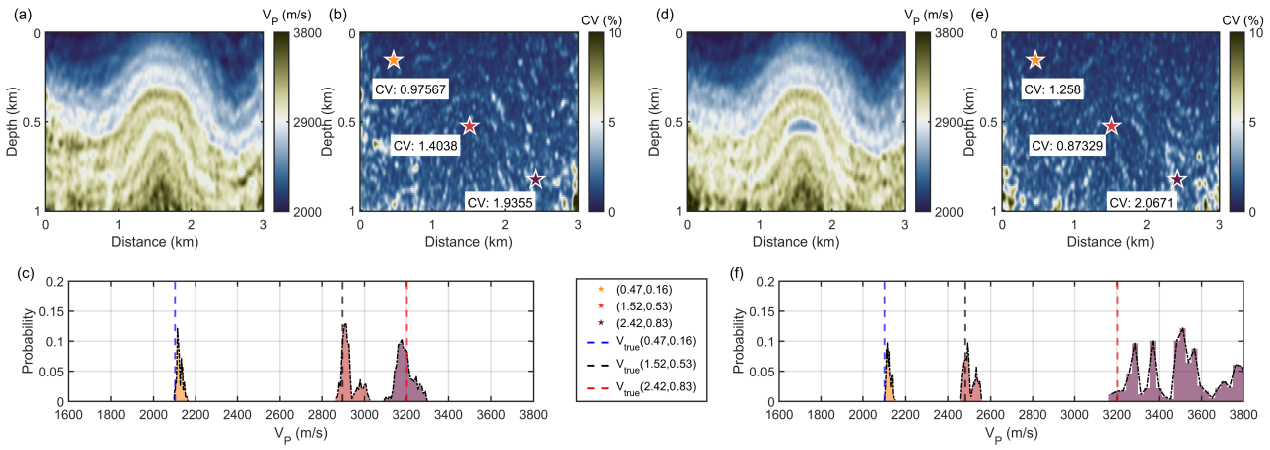

Figure 4.7: Baseline and monitor HMC-FWI results with noisy data. (a) and (d) are baseline and monitor model means $\overline{\mathbf{m}}_b$ and $\overline{\mathbf{m}}_m$. (b) and (e) are coefficient of variation plots of the baseline and monitor models. (c) and (f) are histograms of the posterior distribution of the model parameters on three reference points shown by the three stars in (b) and (e). The three dashed lines denote the true values.

Without a more detailed analysis, the influence of noise does not appear to be significantly amplified. In the model mean shown in Figure 4.8 (a), the primary velocity anomaly remains clearly identifiable, despite the presence of additional artifacts. Larger uncertainties and more irregular posterior distributions emerge in regions poorly constrained by the VSP acquisition, yet around the injection zone the uncertainty levels remain comparable, as illustrated in Figures 4.8 (b) and (c). However, the compounded effects of noise and acquisition non-repeatability become evident when examining the time-lapse variations through the vertical profiles in Figure 4.9. While Figure 4.9 (b) demonstrates good agreement between the true and estimated time-lapse changes, along with their associated uncertainties, stronger deterioration is observed in panels (a) and (c), where the estimated posterior distributions deviate considerably from the true variations.

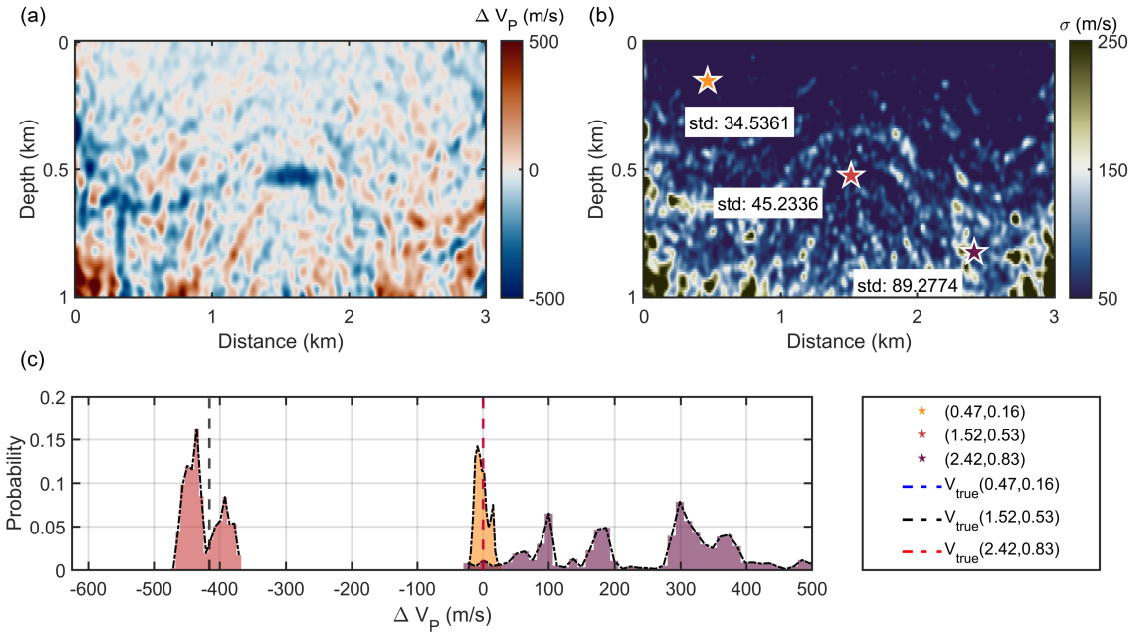

Figure 4.8: Time-lapse HMC-FWI results with noisy data. (a) and (b) are mean value and standard deviation plots of the time-lapse variation, and (c) shows the posterior distributions of the reference points, with the true time-lapse changes denoted by the three dashed lines. Note that the blue and red lines are overlaying each other as they both represent 0.

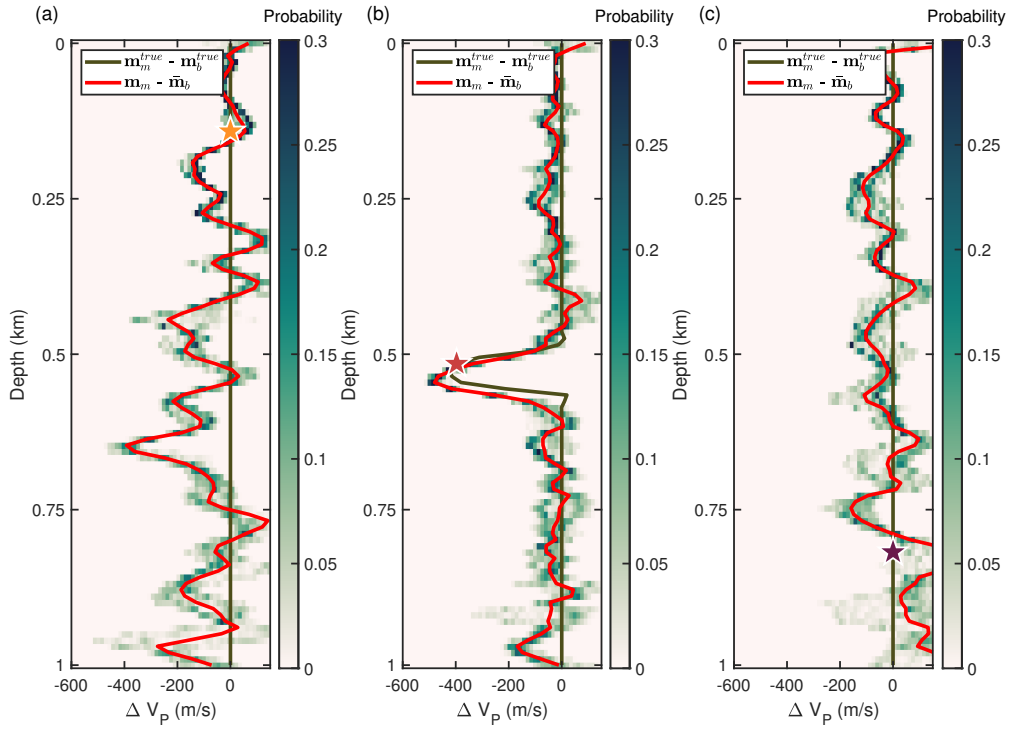

Figure 4.9: HMC time-lapse variation profiles with noisy data. (a) to (c) are profiles across the three reference points, with their locations shown in Figure 4.8. The dark-green line is the true time-lapse change, the red line is the time-lapse model mean, and colors along the profile denote probability values.

4.4.3 Time-lapse FWI with SVGD

In SVGD-FWI, 400 particles are set in the SVGD method. I apply the Adam optimization (Kingma and Ba, 2014), and the master stepsize is set to be 0.01. The computational budget allows approximately 450 iterations for the baseline SVGD-FWI and around 700 iterations for the monitor inversion. The noise-free results are shown in Figure 4.10, which highlight several differences compared to HMC-FWI. The baseline model mean appears generally less noisy, yet artifacts emerge in the monitor inversion due to the reduced number of sources. The CV values are consistently higher than those in the HMC-FWI results, reflecting greater estimated uncertainty. The uncertainty maps also reveal a dependence on

acquisition geometry, as uncertainties remain low near the VSP receiver array but increase with distance. In the monitor case, zones of higher uncertainty appear in the shallow region, likely caused by the reduced number of sources. The posterior distributions at the three reference points are approximately Gaussian, though often with long tails on both sides. This behavior can be attributed to the repulsive property of the SVGD kernel, which pushes particles apart when they cluster too closely.

Figure 4.11 shows the corresponding time-lapse results from SVGD-FWI, which exhibit different behavior from those obtained with HMC-FWI. The model mean in panel (a) successfully captures the main anomaly, although the recovered time-lapse changes appear smoother. While the central anomaly is well resolved, artifacts are present across the model, especially inside the VSP illumination zone. The uncertainty map in panel (b) indicates generally higher values than those from HMC, with the largest uncertainties concentrated in the top-left and bottom-right corners. Sensitivity to acquisition geometry is also observed, as uncertainty is lower near the VSP array and increases with distance on both sides of the model. As shown in panel (c), the posterior distributions at the selected reference points generally follow Gaussian-like shapes, but are broader and occasionally exhibit long tails. This behavior suggests that SVGD-FWI maintains a wider particle spread, which can provide more faithful uncertainty estimates when the true distribution is complicated (Liu and Wang, 2016). However, this same characteristic may limit its ability to resolve fine-scale features or deliver high-confidence estimates in regions strongly constrained by the data, such as the injection area around the VSP receiver array in this study. Consequently, the selection of kernel functions and hyperparameter tuning remains a case-dependent challenge and an important direction for further investigation.

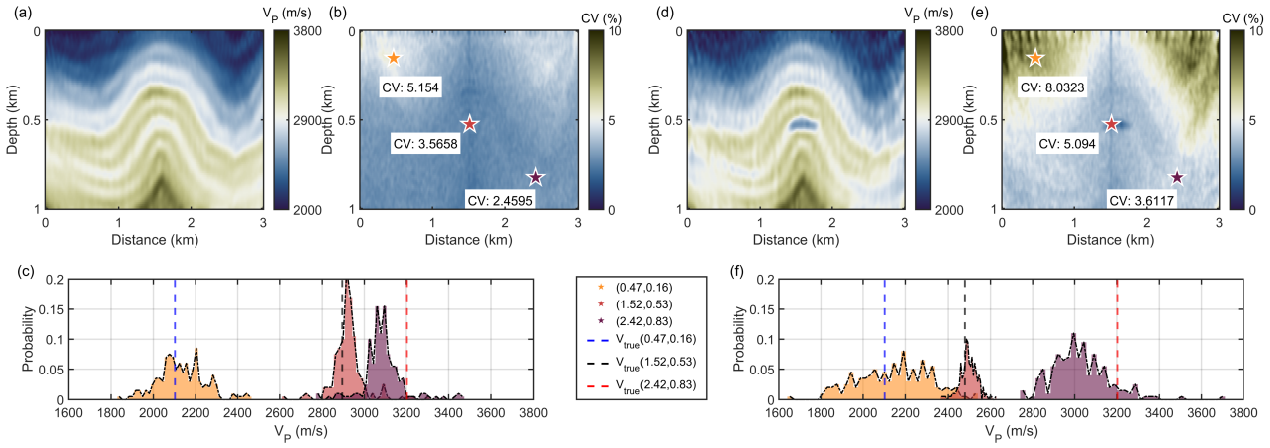

Figure 4.10: Baseline and monitor SVGD-FWI results with noise-free data. (a) and (d) are baseline and monitor model means $\overline{\mathbf{m}}_b$ and $\overline{\mathbf{m}}_m$. (b) and (e) are coefficient of variation plots of the baseline and monitor models. (c) and (f) are histograms of the posterior distribution of the model parameters on three reference points shown by the three stars in (b) and (e). The three dashed lines denote the true values.

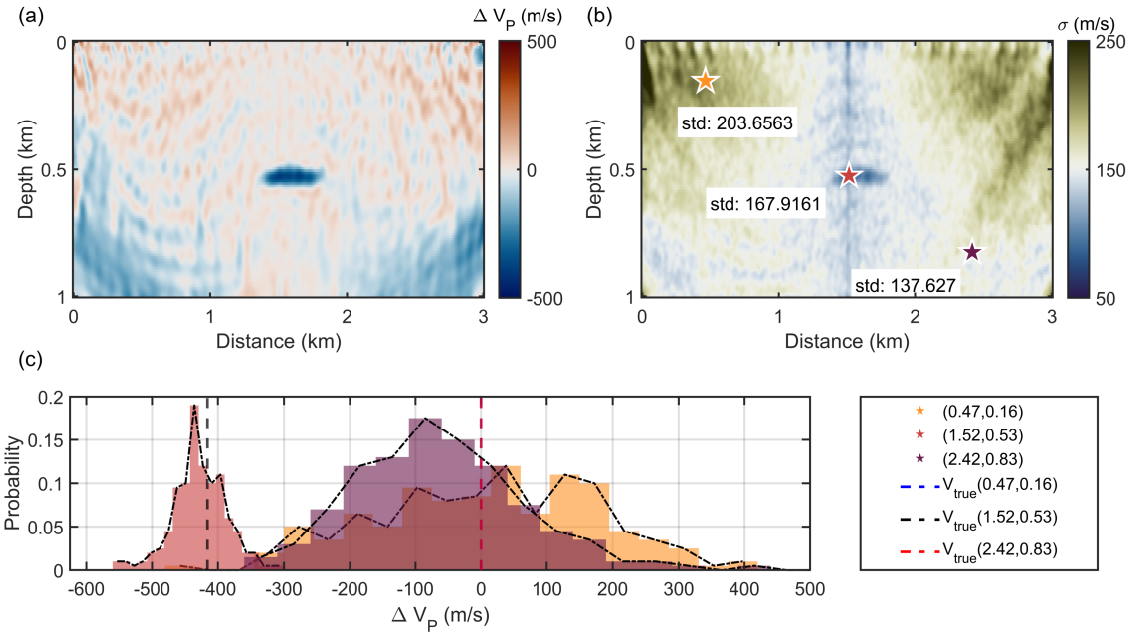

Figure 4.11: Time-lapse SVGD-FWI results with noise-free data. (a) and (b) are mean value and standard deviation plots of the time-lapse variation, and (c) shows the posterior distributions of the reference points, with the true time-lapse changes denoted by the three dashed lines. Note that the blue and red lines are overlaying each other as they both represent 0.

The time-lapse profiles in Figure 4.12 show that SVGD-FWI generally aligns well with the true time-lapse P-wave velocity variations, while also highlighting its sensitivity to acquisition geometry. In panels (a) and (c), which are farther from the VSP receiver array, the posterior distributions are broader and the model means exhibit larger oscillations. In contrast, panel (b), located closer to the VSP array, shows that the model mean aligns closely with the true time-lapse changes, accompanied by lower uncertainty. This confirms the higher confidence in regions near the injection area.

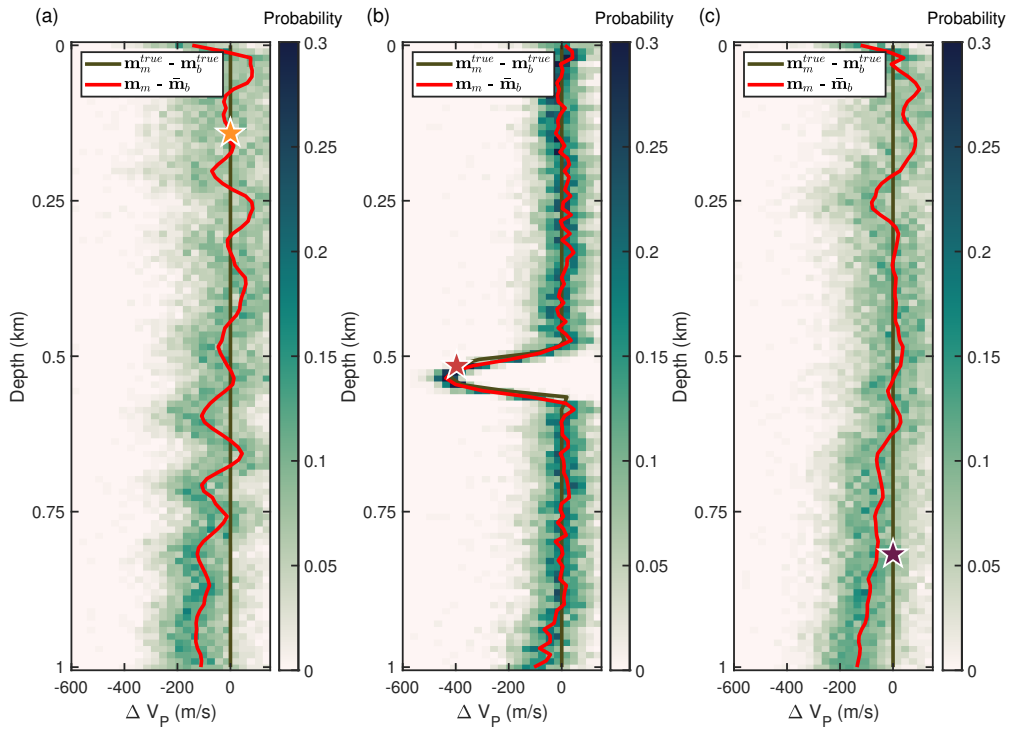

Figure 4.12: SVGD time-lapse variation profiles with noise-free data. (a) to (c) are profiles across the three reference points, with their locations shown in Figure 4.11. The dark-green line is the true time-lapse change, the red line is the time-lapse model mean, and colors along the profile denote probability values.

Next, I evaluate the performance of SVGD-FWI under noise conditions. The baseline and monitor inversion results are shown in Figure 4.13. Compared to the noise-free scenario, the mean models reveal increased structural artifacts. In addition, the uncertainty becomes more spatially extensive and prominent, indicating a deterioration due to noise and acquisition gaps. The posterior distributions at the selected reference points retain an approximately Gaussian shape but become generally diffused, reflecting reduced confidence in the model estimates.

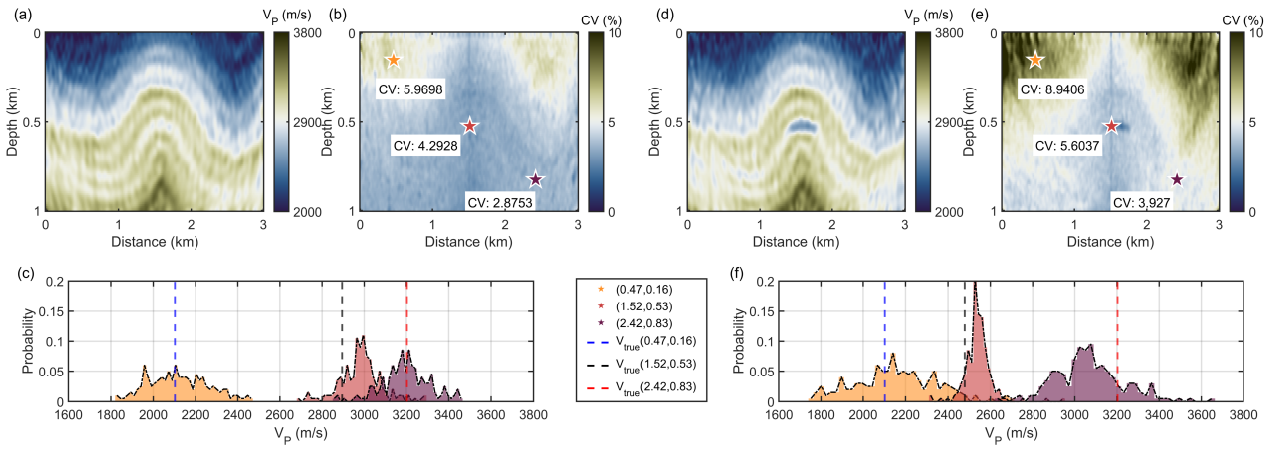

Figure 4.13: Baseline and monitor SVGD-FWI results with noisy data. (a) and (d) are baseline and monitor model means $\overline{\mathbf{m}}_b$ and $\overline{\mathbf{m}}_m$. (b) and (e) are coefficient of variation plots of the baseline and monitor models. (c) and (f) are histograms of the posterior distribution of the model parameters on three reference points shown by the three stars in (b) and (e). The three dashed lines denote the true values.

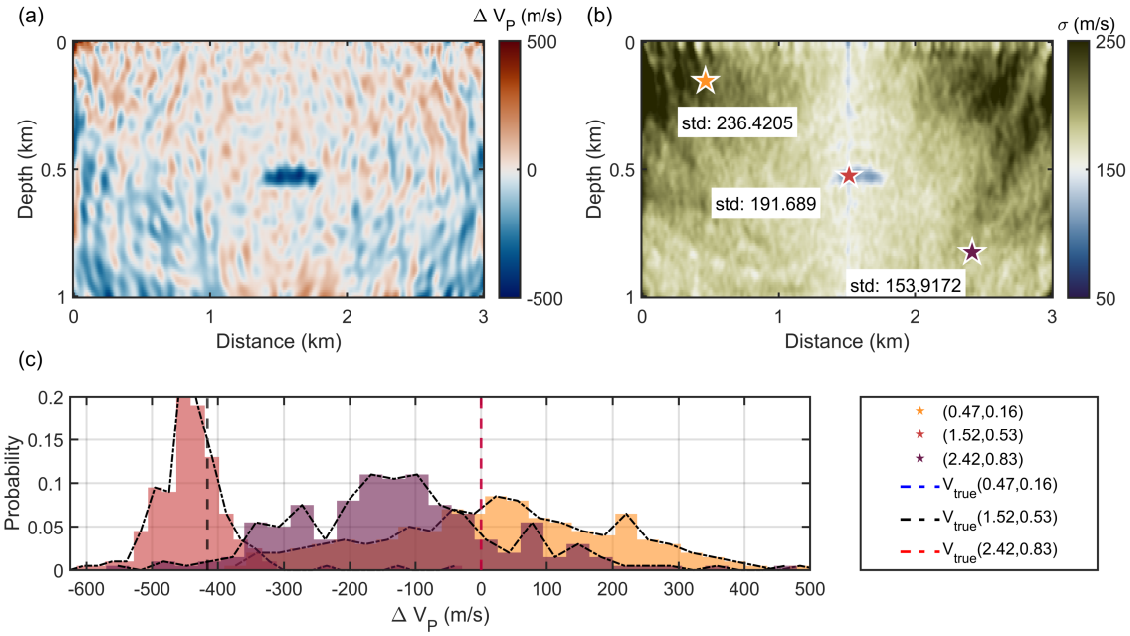

Figure 4.14: Time-lapse SVGD-FWI results with noisy data. (a) and (b) are mean value and standard deviation plots of the time-lapse variation, and (c) shows the posterior distributions of the reference points, with the true time-lapse changes denoted by the three dashed lines. Note that the blue and red lines are overlaying each other as they both represent 0.

Figure 4.14 presents the time-lapse inversion results with noisy data. Compared to the noise-free case, the model mean in panel (a) shows a noticeable increase in artifacts, and the uncertainty is elevated across the model domain. Despite this degradation, the region surrounding the VSP array, particularly the injection zone, remains relatively well resolved, as highlighted in panel (b), where the uncertainty is moderate and the anomaly is clearly delineated. In contrast, uncertainty increases substantially in areas farther from the VSP array, reflecting the limited constraint provided by the data. This spatial variation is further illustrated in panel (c), where the posterior distributions at the reference points that are located away from the VSP region exhibit greater variance and more dispersed profiles. The observations in Figure 4.14 are further supported by the time-lapse profiles shown in Figure 4.15.

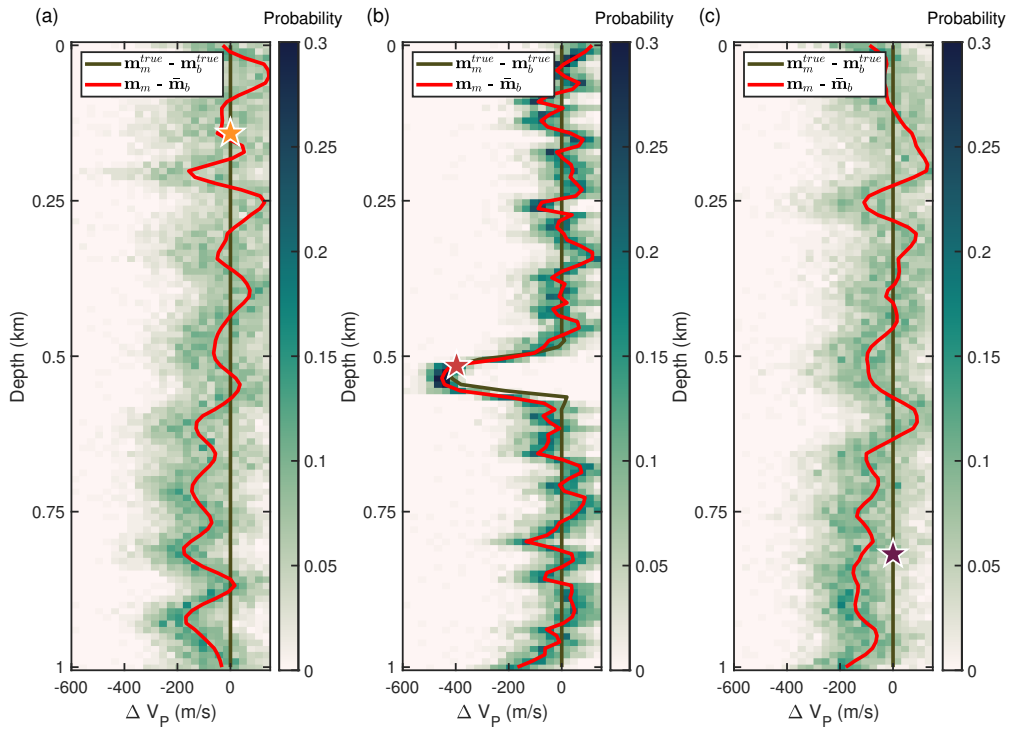

Figure 4.15: SVGD time-lapse variation profiles with noisy data. (a) to (c) are profiles across the three reference points, with their locations shown in Figure 4.14. The dark-green line is the true time-lapse change, the red line is the time-lapse model mean, and colors along the profile denote probability values.

4.5 Discussion

4.5.1 Convergence behaviors

As discussed earlier, HMC begins from a state of complete uncertainty and gradually constructs the posterior distribution through stochastic sampling, while SVGD starts from a uniform prior and iteratively refines it using deterministic particle-based dynamics. The evolution of the monitor models with noisy datasets, shown in Figures 4.16 and 4.17 for HMC-FWI and SVGD-FWI, respectively, reveals notably different convergence behaviors.

Although the model means in both methods remain relatively stable during the middle

and final stages of inversion, the evolution of uncertainty diverges clearly. HMC exhibits a more uniform, global estimation of uncertainty without strong spatial focus, reflecting the stochastic nature of its sampling process. In contrast, SVGD shows a more targeted evolution, where the uncertainty quickly becomes concentrated around the injection zone in which data sensitivity is highest. This targeted localization makes SVGD particularly suitable for time-lapse monitoring applications such as CO_2 sequestration.

Both methods, being gradient-based, are sensitive to subsurface illumination. However, their responses in poorly illuminated zones differ. In these areas, where data provide weak constraints, HMC tends to accept a wide range of model proposals, resulting in a potentially chaotic posterior distribution that may be difficult to interpret or unreliable. SVGD, on the other hand, updates such regions very slowly, or sometimes halts updating, producing wider posterior distributions while maintaining controlled estimates.

Even with a limited computational budget preventing full convergence, the evolution of uncertainty maps allows us to predict future behaviors. In HMC (Figure 4.16), the monitor mean models gradually accumulate a wider range of values in poorly illuminated regions, especially in the lower-left and lower-right corners of the model, indicating that continued sampling could yield posterior features that are challenging to interpret. More constrained HMC variants (e.g., Fichtner et al., 2018) may improve control over the sampling process, but they require careful tuning and are case-dependent. In contrast, SVGD (Figure 4.17) tends to gradually fix the uncertainty estimation within a stable range, although this range depends on the choice of kernel and hyperparameters. Without knowing the true uncertainty in poorly illuminated areas, it is difficult to determine which method is more accurate. However, SVGD clearly provides a more controlled and stable estimate of uncertainty.

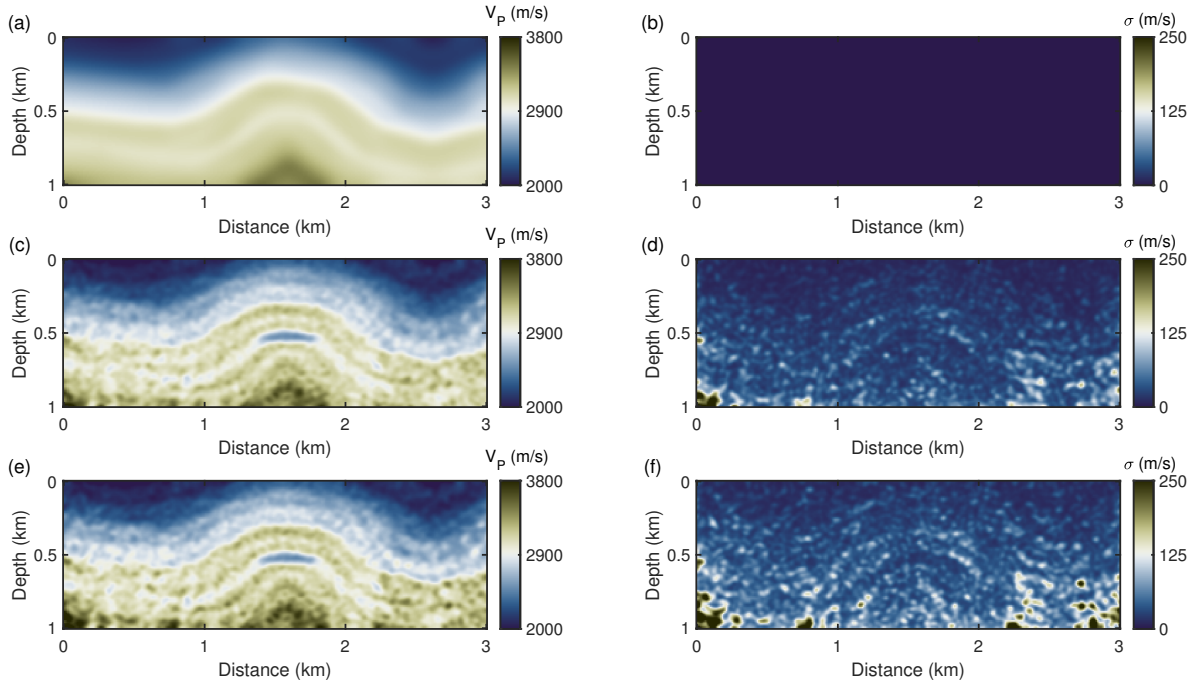

Figure 4.16: Model evolutions in HMC-FWI with noisy data. (a) and (b) are the mean model and the standard deviation from the beginning phase. (c) and (d) are the mean model and the standard deviation in the middle of the optimization. (e) and (f) are the mean model and the standard deviation after the optimization is done.

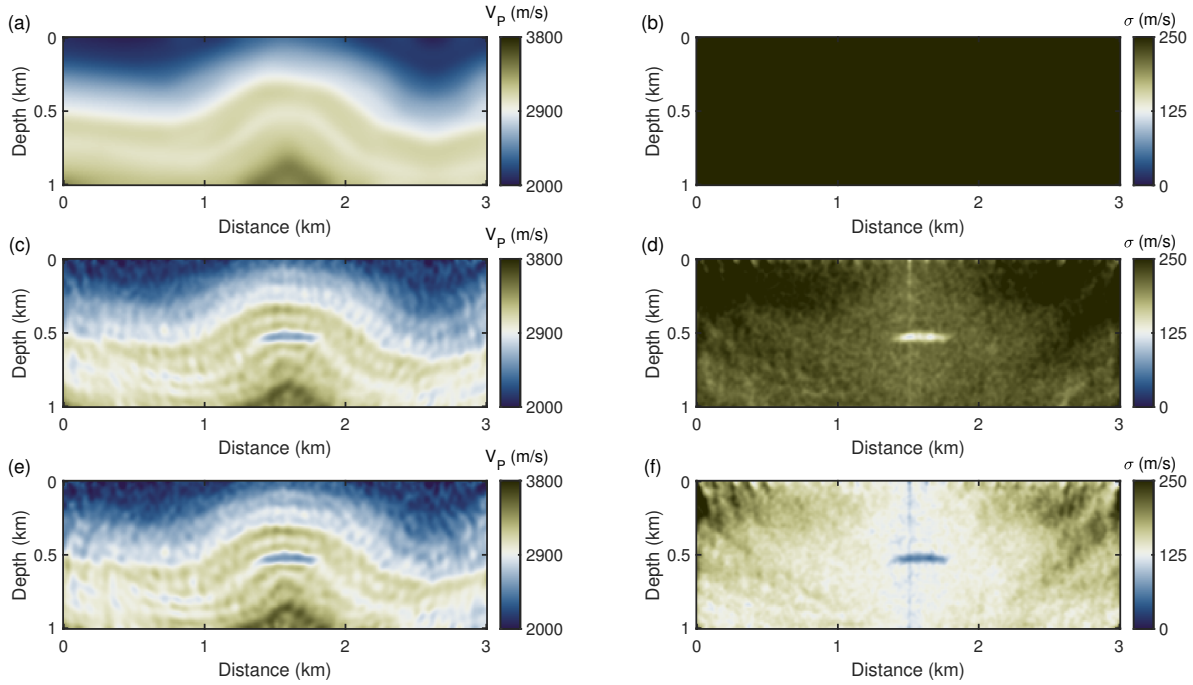

Figure 4.17: Model evolutions in SVGD-FWI with noisy data. (a) and (b) are the mean model and the standard deviation from the beginning phase. (c) and (d) are the mean model and the standard deviation in the middle of the optimization. (e) and (f) are the mean model and the standard deviation after the optimization is done.

The above discussions, similar to those in Zhang and Curtis (2020); Gebraad et al. (2020), suggest that, under limited computational budgets, SVGD offers a more efficient and targeted characterization of time-lapse changes than HMC, particularly around the injection zone where seismic data exhibit the most significant differences.

4.5.2 Robustness to survey nonrepeatability

Across the numerical tests, the nonrepeatability lies in choosing different acquisition and noise levels between the baseline and monitor FWI runs. The uncertainty estimations of time-lapse changes show accordingly different behaviors and robustness. HMC tends to amplify and propagate the errors, especially in areas that lack sufficient illumination. As suggested by Figures 4.4, 4.5, 4.7, and 4.8, HMC produces reasonable mean models and

uncertainty maps when nonrepeatability exists in both the survey design and datasets. However, the propagated errors escalates the interpretation challenges in evaluating time-lapse variations. As seen in the vertical profiles in Figures 4.6 and 4.9, some parts, especially the deeper sections of the HMC posterior tend to deviate from the true time-lapse variations, likely due to insufficient sampling and incomplete convergence. Although many model grids exhibit near-Gaussian distributions as a reasonable product from most of MC methods, these distributions should not be taken as evidence of reliable uncertainty estimates. Such deviations from the ground truth can be misleading in real-data applications. In contrast, SVGD profiles in Figures 4.12 and 4.15 demonstrate better alignment with the true model. Again and admittedly, this improved consistency in SVGD is partly attributed to the use of predefined value boundaries. However, when these bounds are well chosen to encompass the true parameter range, SVGD can more effectively capture the posterior and reduce interpretational ambiguity. Additionally, the use of momentum in the Adam optimization algorithm contributes to an improved convergence, which is an advantage not present in HMC in rapid assessment contexts.

4.5.3 Computational complexity

I compare HMC and SVGD under comparable computational budgets, specifically, one day of runtime on 16 CPU cores. Both methods provide reasonable uncertainty estimates, but they differ significantly in computational behavior, scalability, and optimization flexibility. HMC builds the posterior through stochastic sampling from arbitrary initial conditions, gradually exploring the model space via simulated Hamiltonian dynamics. This more thorough sampling, however, is computationally intensive. Achieving sufficient coverage typically requires over $100,000$ samples (Gebraad et al., 2020). While parallelization across multiple chains can accelerate sampling, it poses challenges in maintaining diversity. Without proper inter-chain interaction, chains may converge to a single posterior mode. In the frequency-domain FWI implementation, I parallelize across frequencies rather than chains, which is simpler but

limits the extent of model-space exploration. In contrast, SVGD employs a deterministic, particle-based approach that iteratively transforms an initial ensemble to approximate the posterior. Its update mechanism, driven by gradient-based interactions, naturally supports parallelization. In my setup, 4 cores are allocated to particle updates and 4 to frequency parallelization. It can be reasonably inferred that, as core counts increase, SVGD scales efficiently, making it a more flexible and resource-friendly alternative for large-scale uncertainty quantification. In 3D FWI applications, SVGD presents greater potential for practical implementation due to its inherent scalability and efficient parallelization.

4.5.4 Future work

While this study provides a controlled comparison between HMC and SVGD, there remains substantial scope for further investigation and optimization. Both methods involve several hyperparameters and features that can influence their performance and convergence behavior. For instance, in HMC, the choice of the mass matrix plays a crucial role in shaping the sampling trajectory, potentially improving sampling efficiency. An example can be seen in Gebraad et al. (2020). Similarly, in SVGD, the selection of the kernel function governs the repulsion and attraction dynamics among particles, impacting the diversity and fidelity of the approximated posterior ensemble. In this study, default or commonly used values are employed to maintain a fair and controlled condition for comparison. However, future work could involve systematic tuning or adaptive strategies to optimize these features, potentially leading to improved inversion performance and faster convergence for both methods in TL-FWI. Additionally, while I focus on a parallel inversion strategy where statistical features are easier to derive, future experiments could explore various time-lapse inversion frameworks, which may present additional challenges but new findings. Zhang et al. (2023) has presented a 3D SVGD-FWI, its time-lapse applications might also be worth investigating. Extending the current framework to elastic TL-FWI would also enable capturing the cross-talk effects, which are critical in certain monitoring environments. Finally, applying and validating both

HMC and SVGD on real field datasets will be essential to assess their robustness and practical utility.

4.6 Conclusions

In this study, I explore the application of HMC and SVGD to time-lapse VSP-FWI through synthetic experiments simulating a CO_2 storage monitoring scenario, with a focus on their capabilities for uncertainty quantification and their respective strengths and limitations. HMC demonstrates stronger capabilities in exploring more comprehensive model spaces owing to its rigorous sampling-based formulation, but its scalability and sampling cost pose practical limitations. SVGD, although potentially more sensitive to initial distributions and limited in model-space exploration, presents a more resource-efficient and scalable alternative. When computational resources are constrained or fast turnaround is required, SVGD offers greater potential for optimization and real-time deployment in TL-FWI problems. Future research should aim to systematically explore hyperparameter optimization in both methods, evaluate performance across varied time-lapse strategies and elastic wave physics, and advance toward applications involving real data and 3D time-lapse inversion scenarios.

Chapter 5

3D targeted nullspace shuttles in 4D FWI: synthetic time-lapse FWI experiments for CO_2 monitoring configured for the Snowflake dataset

5.1 Summary

FWI is a critical tool for monitoring subsurface CO_2 sequestration, an essential component of CCUS. However, interpreting time-lapse FWI results can be challenging due to artifacts arising from non-repeatability between baseline and monitor surveys. The targeted nullspace shuttle technique addresses this issue by enhancing structural consistency between baseline and monitor inversions without increasing the overall data misfit. Previous studies have demonstrated its effectiveness in 2D settings, showing improved reliability in the interpretation of time-lapse differences. In this chapter, I extend the targeted nullspace shuttle framework to 3D time-lapse FWI and assess it in 4D seismic monitoring scenarios through synthetic acoustic inversion experiments. The synthetic model is derived from the Snowflake VSP dataset, based on the CaMI FRS of CMC, jointly developed by the Consortium for Research in Elastic Wave Exploration Seismology (CREWES). The scenario simulates a realistic injection process in which 60-tonnes of CO_2 is injected. I evaluate the detectability of the injected plume using the nullspace shuttle method under varying acquisition sparsity and aperture conditions and quantify the relationship between CO_2 saturation levels and observable P-wave velocity changes. The results indicate that the targeted nullspace shuttle enables effective plume delineation in sparse acquisition scenarios. However, the limited aperture inherent to VSP-based FWI restricts the resolution of finer-scale plume features, leading to a suboptimal or failed time-lapse recovery. The findings highlight the potential

of the targeted nullspace shuttle as a focused and computationally efficient strategy for 4D FWI applications, particularly in improving the robustness of geo-monitoring.

5.2 Introduction

While FWI offers high-resolution imaging of the subsurface, its time-lapse application presents several fundamental challenges. Chief among these is the nonlinearity of the inverse problem, which results in non-uniqueness where multiple velocity models may produce similarly low data misfits yet differ significantly in structural and physical details. Even after convergence, the obtained FWI model may only represent one of many plausible solutions that fit the data. Such challenges are especially pronounced when monitoring subtle or spatially localized variations, such as CO_2 migration or small-scale pressure changes. The combined effects of solution non-uniqueness and acquisition-related inconsistencies give rise to significant uncertainty in interpreting time-lapse differences, fundamentally limiting the reliability of conventional FWI for geo-monitoring. Hence, robust strategies are required to suppress spurious variations and extract meaningful geophysical signals related to real reservoir dynamics.

To address the aforementioned limitations, the targeted nullspace shuttle technique, originally proposed by Keating and Innanen (2021), offers a novel inversion framework that systematically explores the inversion nullspace, which is defined as the set of models that yield similarly acceptable data fits near convergence (Deal and Nolet, 2007). Instead of relying solely on independent baseline and monitor inversions, this approach probes the inversion nullspace by minimizing a secondary objective function to identify model updates that preserve data consistency while generate physically plausible differences and suppress inversion artifacts in time-lapse models. This technique has demonstrated promising results in 2D time-lapse FWI studies, with validation under varying acquisition geometries, noise levels, and inversion strategies (Keating and Innanen, 2024; Pike et al., 2024a,b). However,

its practical deployment has thus far been limited to 2D applications. Given the inherent directional nature of subsurface changes and acquisition configurations in time-lapse FWI, extending the nullspace shuttle method to 3D FWI represents a crucial step toward making it viable for field monitoring tasks.

In this chapter, I present a 3D extension of the targeted nullspace shuttle for time-lapse acoustic FWI in the frequency domain. Building on the theoretical framework established by Keating and Innanen (2021), I validate this approach using synthetic models derived from the Snowflake VSP dataset and numerical simulations of a 60-tonne CO_2 injection at the CaMI FRS. The dataset offers a controlled CO_2 injection and monitoring environment with realistic geological changes and acquisition geometry. The numerical experiments demonstrate that the 3D targeted nullspace shuttle method effectively suppresses non-repeatable artifacts introduced by survey inconsistencies, while enhancing the detectability of true P-wave velocity changes associated with CO_2 plume migration. The results underscore the promise of the 3D nullspace shuttle approach as a practical and computationally efficient solution for 4D seismic monitoring in settings such as geological carbon storage. However, comparisons reveal that the limited aperture inherent to VSP geometry constrains the lateral resolution of subtle plume features. These limitations persist even when the targeted nullspace shuttle is applied.

5.3 Theory

In time-lapse applications where the optimal differences between baseline and monitor inversion outcomes are pursued, the shuttling problem can be framed as the following dual-

optimization process:

$$\begin{aligned}\Psi\left(\Delta\mathbf{m}_b^{'},\Delta\mathbf{m}_m^{'}\right)=\underset{\Delta\mathbf{m}_b^{'},\mathbf{m}_m^{'}}{argmin}\Psi\left(\mathbf{m}_b+\Delta\mathbf{m}_b^{'},\mathbf{m}_m+\Delta\mathbf{m}_m^{'}\right),\\ s.t.\Phi_b\left(\mathbf{m}_b+\Delta\mathbf{m}_b^{'}\right)\approx\Phi_b\left(\mathbf{m}_b\right),\\ \Phi_m\left(\mathbf{m}_m+\Delta\mathbf{m}_m^{'}\right)\approx\Phi_m\left(\mathbf{m}_m\right),\end{aligned}\tag{5.1}$$

where the subscriptions b and m refer to baseline and monitor. Following Keating and Innanen (2024); Pike et al. (2024a,b), I take the Huber norm (Guitton and Symes, 2003) for the shuttling objective function Ψ:

$$\Psi\left(\Delta\mathbf{m}_b^{'},\Delta\mathbf{m}_m^{'}\right)=\sum_{i=1}^{N_{grid}}E_\epsilon\left(e_i\right),\tag{5.2}$$

where N_{grid} is the overall grid points in the entire subsurface space, and

$$E_\epsilon\left(e_i\right)=\begin{cases}\frac{e_i^2}{2\varepsilon} & 0\leq|e_i|\leq\varepsilon,\\ |e_i|-\frac{\varepsilon}{2} & \varepsilon<|e_i|.\end{cases}\tag{5.3}$$

In equation 5.3, the threshold ε is a predefined scalar that defines the inversion nullspace of the baseline and monitor FWI, and

$$e_i=m_{m,i}-m_{b,i}.\tag{5.4}$$

Such an L_1 L_2 flexible norm promotes the sparsity in the time-lapse model when there are large differences and better handles the non-repeatable artifacts with the L_1 penalty, and the L_2 transition ensures a better convexity near zeros.

The framework described above is naturally extensible to 3D. The 3D inversion relies on accurate wavefield modeling, which is achieved by solving the acoustic Helmholtz equation in the frequency domain using a second-order finite-difference scheme on staggered grids in Cartesian coordinates. The implementation closely follows the 3D frequency-domain modeling approach detailed in Chapter 2.

5.4 Numerical experiments

In this section, I present a series of numerical experiments. All examples assume 3D constant-density acoustic media. For conventional FWI runs, we sequentially invert one frequency at a time across successive frequency bands ranging from 8 to 72 Hz. The L-BFGS optimization approach is used in the regular FWI and the nullspace shuttle implementation. In regular FWI, we use 2 outer iterations to update the models, and 10 inner iterations to calculate the updating directions. The nullspace shuttle implementation follows the frequency sensitivity insights of Li et al. (2024), using 3 shuttle iterations within a single band consisting of 4 frequencies between 45 and 70 Hz. A parallel time-lapse inversion strategy is employed throughout all tests.

5.4.1 P-wave velocity model from FRS

The FRS features a robust VSP monitoring system designed to evaluate time-lapse geophysical responses to injected CO_2. To date, more than 85 tonnes of CO_2 have been injected into the Basal Belly River Sandstone (BBRS) formation, a relatively flat-lying unit situated between 295 and 305 meters depth (Kolkman-Quinn et al., 2024). This mid-level stratigraphic unit offers favorable conditions for controlled plume migration and high-resolution monitoring. Following recent studies (e.g., Isaac and Lawton, 2024; Macquet et al., 2019; Leaney et al., 2019; Hu and Innanen, 2021; Leaney et al., 2022; Kolkman-Quinn et al., 2024), I simulate a scenario involving the injection of 60 tonnes of CO_2. As shown in Figure 5.1, the resulting saturation model exhibits a spatial gradient, with CO_2 saturations decreasing from approximately 30% near the injection well to about 5% at greater distances. I construct the baseline, monitor, and initial velocity models as illustrated in Figure 5.2. All models are discretized on a uniform grid with 5-meter spacing.

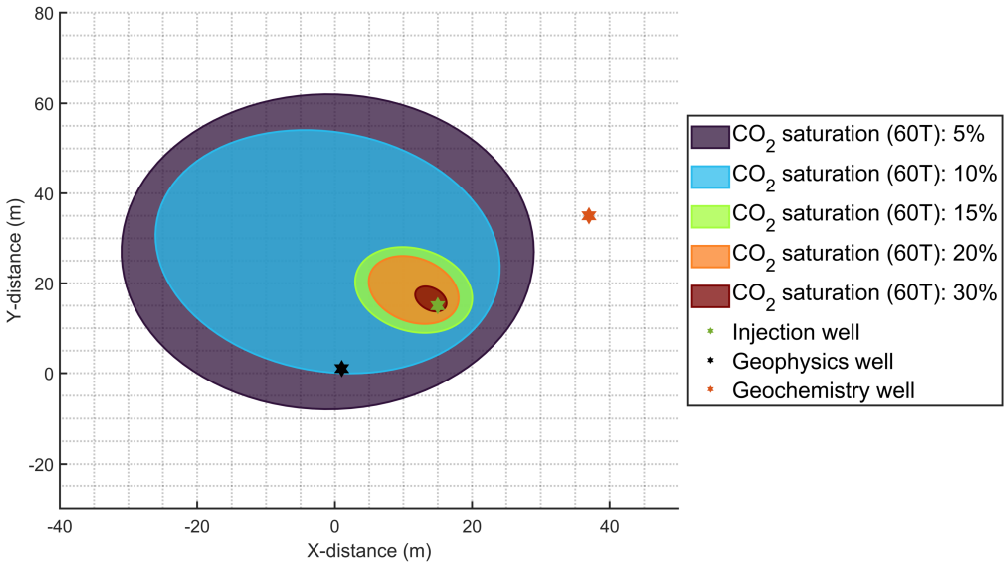

Figure 5.1: CO_2 saturation around the injection well. The injection, geophysics, and geochemistry wells are indicated with stars.

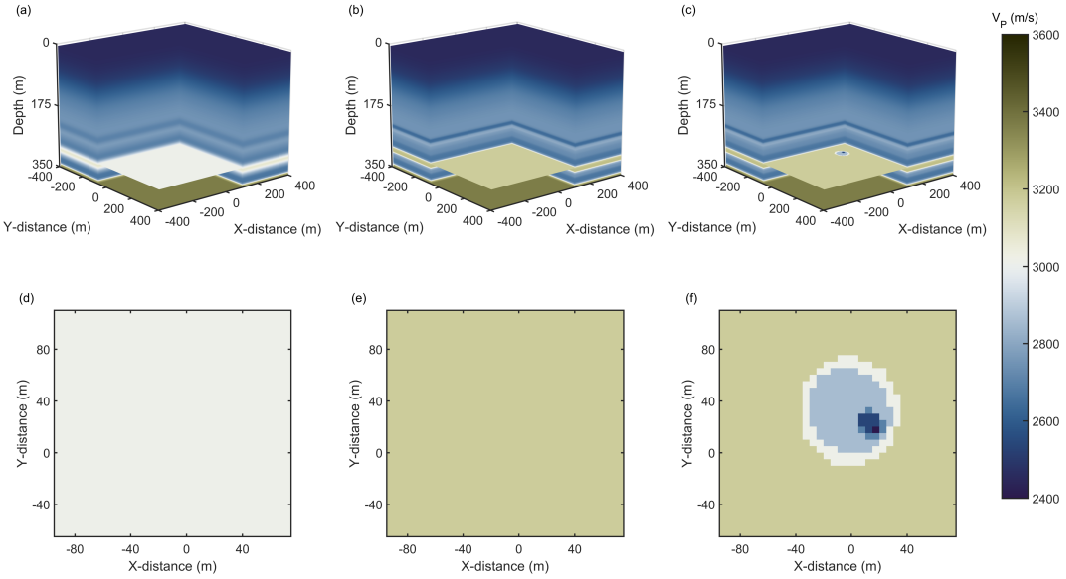

Figure 5.2: Initial, baseline, and monitor models. (a)-(c) Initial, baseline, and monitor models. (d)-(e) Depth slice at 295 meters from (a)-(c).

5.4.2 Regular time-lapse FWI

Following the design considerations outlined in Hall et al. (2019), the acquisition layout is simplified by removing practical complexities that could interfere with controlled analysis. In this configuration, a total of 415 surface sources are deployed, as illustrated in Figure 5.3, with acquisition lines and source positions indicated.

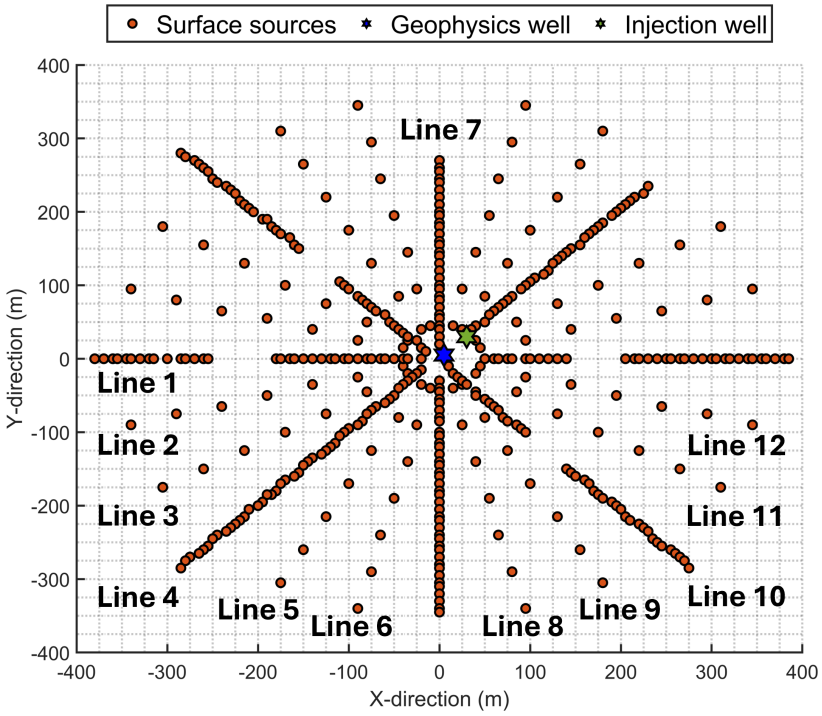

Figure 5.3: Source layout adapted from real CaMI dataset.

To simulate a realistic monitoring scenario, I begin by characterizing noise levels using SNRs estimated from pre-processed baseline and monitor field datasets reported in Cai et al. (2024). As shown in Figure 5.4, the baseline data exhibit notably lower SNRs, especially in the mid-to-high-frequency range. Based on the observations, I apply artificial Gaussian noise to the synthetic datasets, scaling the noise levels to match the frequency-dependent SNRs observed in the field data.

Figure 5.5 shows the baseline and monitor inversion results obtained using full and consistent acquisition geometry. Owing to the lower SNR in the baseline data, the baseline FWI

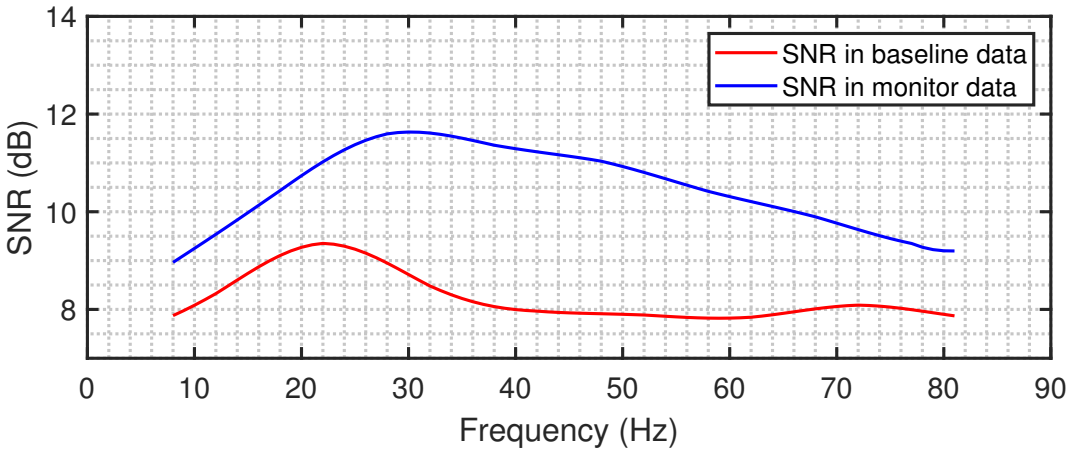

Figure 5.4: Signal-to-noise ratios from baseline (red) and monitor (blue) datasets.

model displays more pronounced artifacts, particularly around the VSP receiver locations, where noise-induced degradation is most evident. In contrast, the monitor model is less affected by noise, yielding a cleaner result that successfully highlights the low-velocity anomaly near the injection well, corresponding to the CO_2 plume. Both models reveal limitations in subsurface illumination, especially in deeper and lateral regions, suggesting that the VSP acquisition geometry imposes intrinsic aperture constraints on the inversion quality.

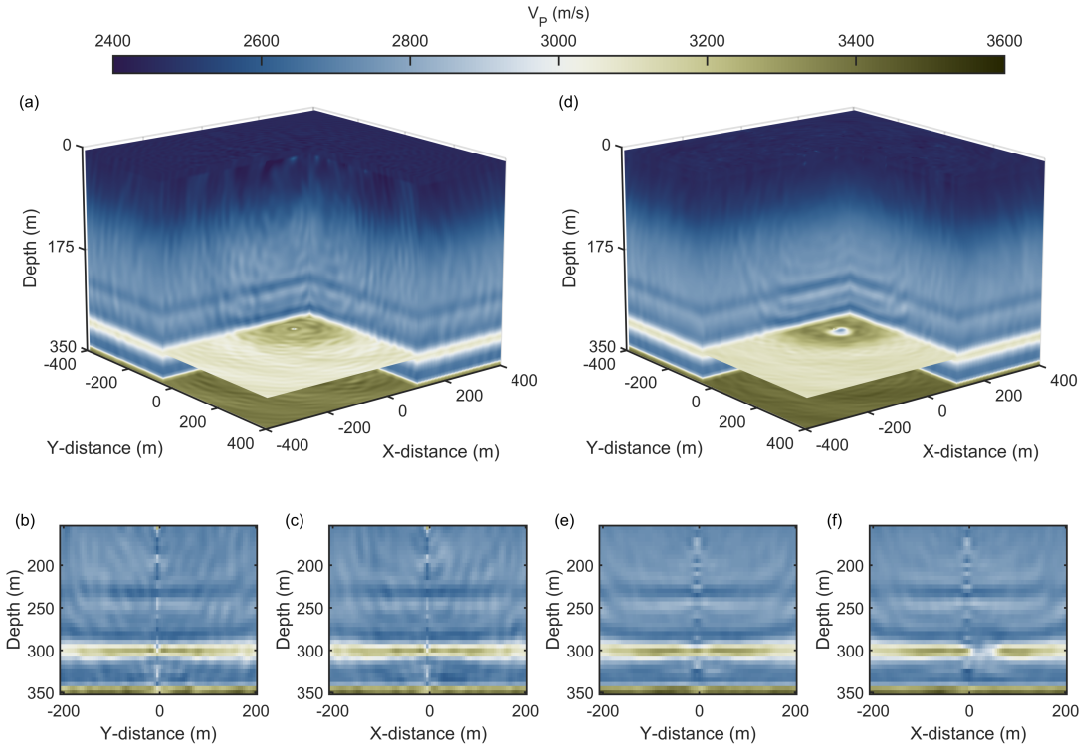

Figure 5.5: Baseline and monitor inversion results with full acquisition geometry. (a)/(d) Overview of the estimated baseline/monitor models. (b)/(e) Cross-sections between −200 to 200 meters in Line 7 in baseline/monitor models. (c)/(f) Cross-sections between -200 to 200 meters in baseline/monitor models. The depth range in the cross-sections is 150 to 350 meters.

5.4.3 Feasibility of sparse acquisition for plume monitoring

To evaluate the feasibility of using a sparser source layout for time-lapse FWI, I perform monitor inversions using the same inversion parameters and azimuthal coverage, while systematically reducing the source density to 100%, 75%, 50%, 25%, and 12.5% of the original CaMI acquisition layout. The corresponding acquisition layouts are shown in Figure 5.6. In the 75% case, the number of sources is halved along the northeast-southwest and northwest-southeast lines, while in the cases 50%, 25%, and 12.5%, every second, third, and fourth

source, respectively, is selected.

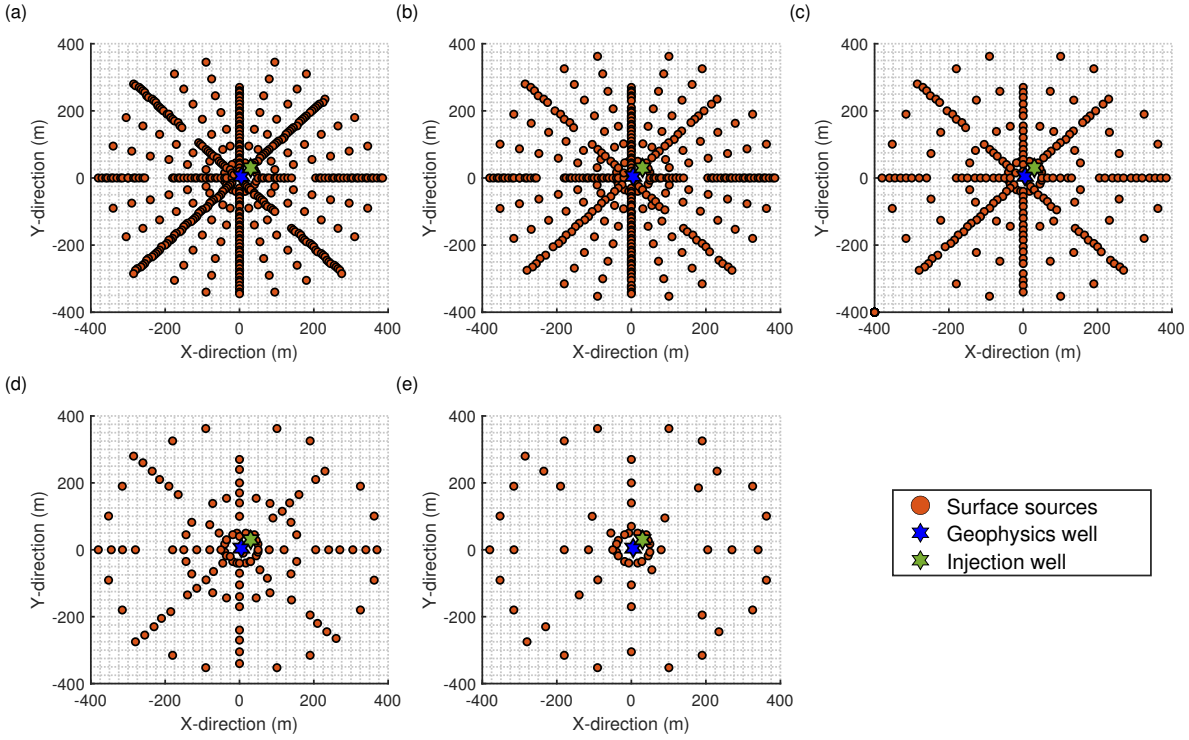

Figure 5.6: Surface source layouts with varying densities. (a) to (e) represent results using 100%, 75%, 50%, 25%, and 12.5% source density.

In Figure 5.7, the recovered time-lapse velocity changes for each case are illustrated. Across all scenarios, the central low-velocity anomaly associated with the CO_2 plume remains clearly identifiable, indicating that plume detection is achievable even with reduced acquisition effort. However, both the resolution and spatial accuracy of the plume delineation degrade as the source density decreases, with the extent of degradation varying between cases. In the 100%, 75%, and even 50% cases, the recovered anomalies closely match the expected plume geometry, with a well-defined outer boundary corresponding to approximately 5% CO_2 saturation. The differences among these three depth slices are minimal. With 25% source density, the plume remains discernible, although the outer boundary appears more diffuse. This effect becomes more pronounced at 12.5% source density, where the boundary

alignment with the 5%–10% saturation contours weakens, suggesting reduced confidence in resolving low-saturation margins.

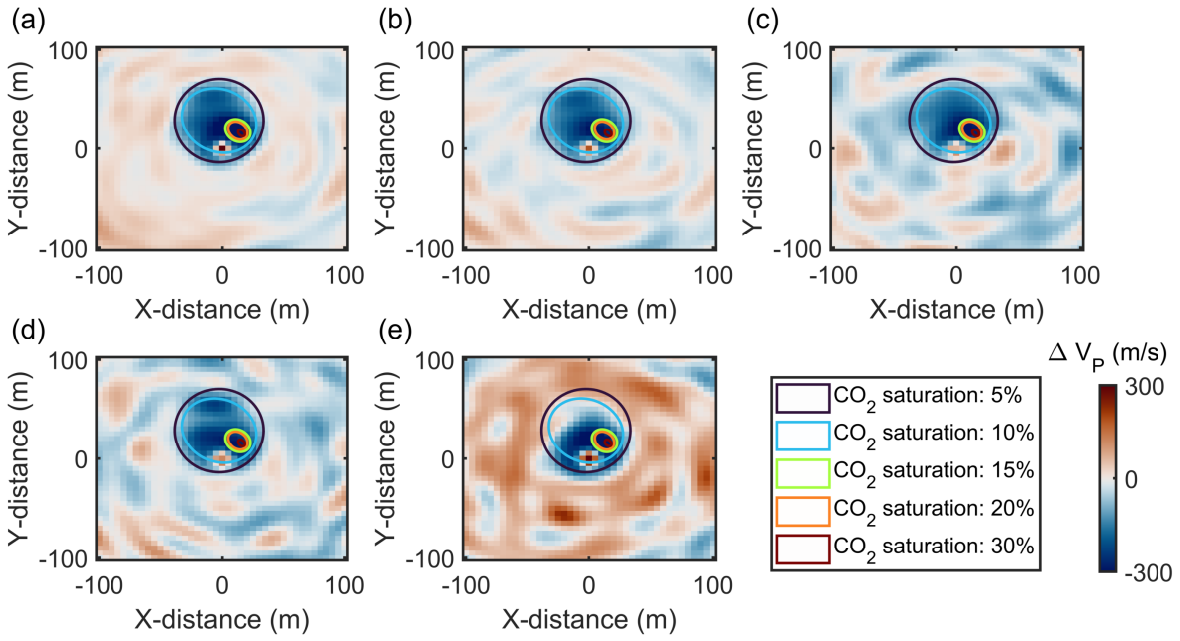

Figure 5.7: Depth slices reflecting time-lapse velocity changes in depth of 295 meters with different source sparsity. The axis is constrained to $(-100m, +100m)$ in X- and Y-directions to zoom at the plume. (a) to (e) represent results using 100%, 75%, 50%, 25%, and 12.5% source sparsity.

In all acquisition scenarios, the time-lapse inversion results are affected by noise and acquisition non-repeatability, with these effects becoming more pronounced as source density decreases. The reduced number of surface sources amplifies spurious artifacts, resulting in increasingly noisy estimates of P-wave velocity changes. To address these challenges, I apply the targeted nullspace shuttling technique as a post-inversion refinement to the 50%, 25%, and 12.5% source density cases where degradation is more evident. This step is not applied to the full and 75% source cases, as they already yield sufficiently clear time-lapse results. The resulting time-lapse differences after applying the nullspace shuttle are shown in Figures 5.8, 5.9, and 5.10 for the three acquisition scenarios, respectively. The

shuttling process leads to a substantial reduction in overburden artifacts and enhances the structural coherence of the recovered time-lapse anomaly. In all the cases, the plume becomes more distinctly identifiable, even in the presence of sparse acquisition, demonstrating the effectiveness of the shuttling procedure in recovering meaningful time-lapse signals. Despite these improvements, differences persist across source sparsity levels. As shown in Figure 5.8, the recovered plume structure with 50% of the sources maintains a shape and extent closely matching the expected distribution, while Figure 5.9 shows that the result with 25% of the sources produces a discontinuous plume that could hinder interpretation, particularly along the plume edges. In the last case (Figure 5.10), the visible plume size is significantly smaller than in all other scenarios. The above observations suggest that although higher sparsity (for example, 25%) can still produce coherent plume delineation after shuttling, the velocity anomaly corresponding to the 5% CO_2 saturation becomes less distinct or even undetectable. This deterioration is even more pronounced at 12.5% of the source density, indicating that aggressive reduction of sources can substantially reduce the resolvability of the plume, even with refinement.

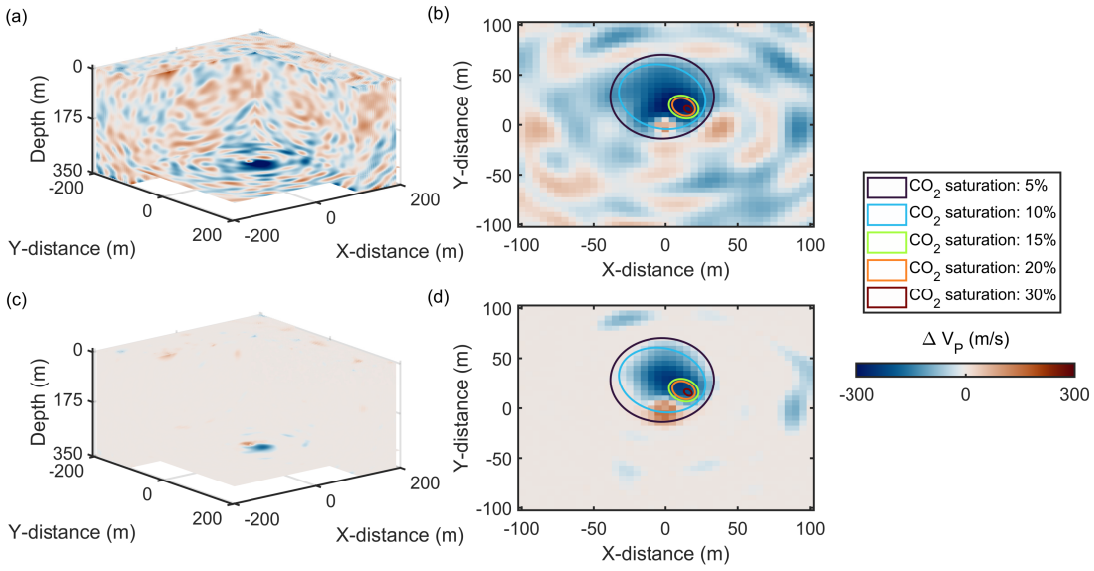

Figure 5.8: Time-lapse inversion and nullspace shuttle results with 50% sources. (a) and (b) display the models before shuttling, and (c) and (d) show the after-shuttling results. Panels (a) and (c) are 3D views, and (b) and (d) are depth slices at 295 meters.

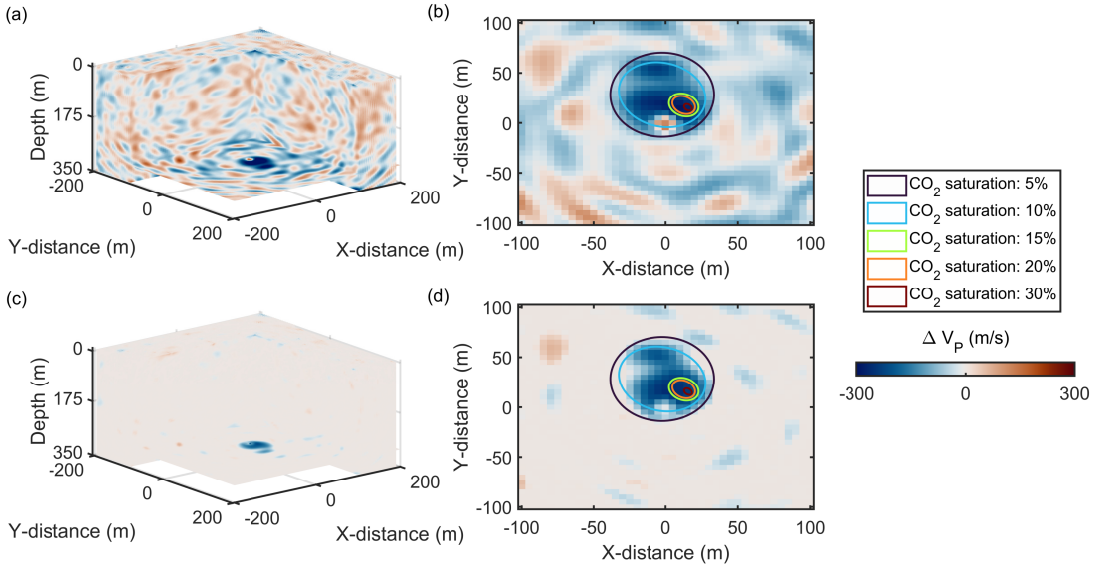

Figure 5.9: Time-lapse inversion and nullspace shuttle results with 25% sources. (a) and (b) display the models before shuttling, and (c) and (d) show the after-shuttling results. Panels (a) and (c) are 3D views, and (b) and (d) are depth slices at 295 meters.

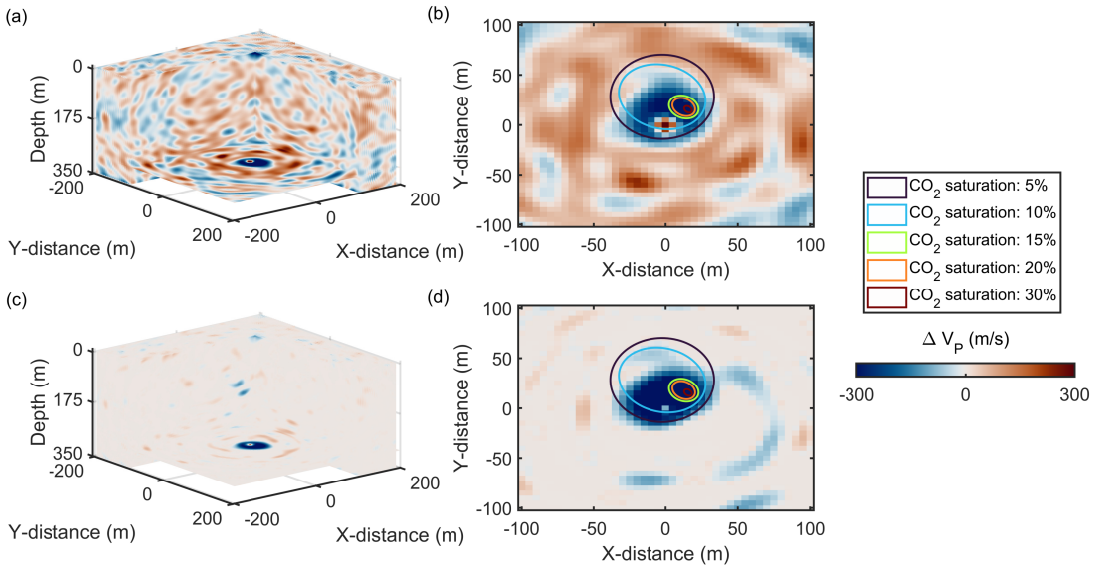

Figure 5.10: Time-lapse inversion and nullspace shuttle results with 12.5% sources. (a) and (b) display the models before shuttling, and (c) and (d) show the after-shuttling results. Panels (a) and (c) are 3D views, and (b) and (d) are depth slices at 295 meters.

5.4.4 Impact of offsets

As suggested by the results in Figure 5.5, aperture and offset might play a critical role in the effectiveness of plume delineation in VSP-FWI. To assess the impact of acquisition aperture, I conduct a series of tests using three different offsets, including 400, 300, and 200 meters. The widths are measured symmetrically from the observation well and are with 100%, 75%, and 50% of the original offsets. To ensure a fair comparison and isolate the offset effect, the number of surface sources is fixed at 25% of the original layout for all tests. This configuration maintains consistent fold coverage and is supported by previous results, which demonstrated that the 25% source density provides a reasonable balance between acquisition effort and plume detectability. The three layouts can be referred to by Figure 5.11.

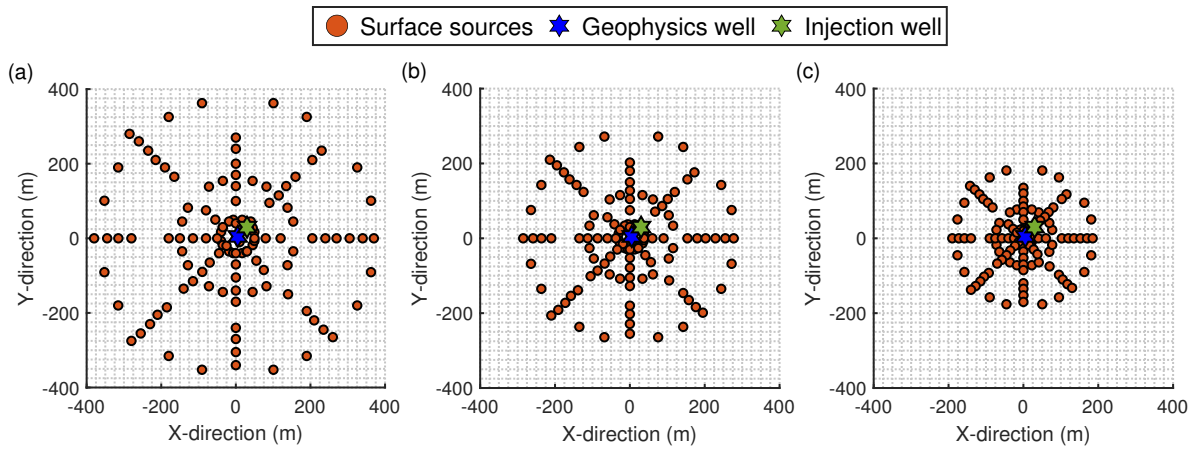

Figure 5.11: Surface source layouts with varying offsets. (a) to (e) represent results using 100%, 75%, and 50% offsets.

The depth slices in Figure 5.12 demonstrate that the ability to resolve fine-scale time-lapse variations is highly sensitive to acquisition aperture. With a 75% offset, the time-lapse result shows only minor differences compared to the full-aperture case, as seen in panels (a) and (b). However, in panel (c), the recovered anomaly is highly ambiguous. Key features of the plume are lost, and the surrounding area is so severely degraded that interpretation becomes discouraging. As the nullspace-shuttled results for the full-aperture scenario are already shown in Figure 5.9, I present the corresponding refinements for the cases with 75% and 50% offsets in Figures 5.13 and 5.14. Although shuttling moderately improves the 75%-aperture result, yielding partial alignment with the true saturation contours, finer features remain poorly defined, as indicated by the small and faint outer plume boundary in panel (d). In the 50%-aperture case, even after nullspace refinement, the plume is barely visible, and the recovered velocity anomaly fails to correlate with the expected saturation distribution. These findings suggest that a critical aperture threshold exists for effective time-lapse monitoring in VSP-FWI for targets of a given spatial extent, below which even advanced post-inversion techniques provide only limited or no benefit.

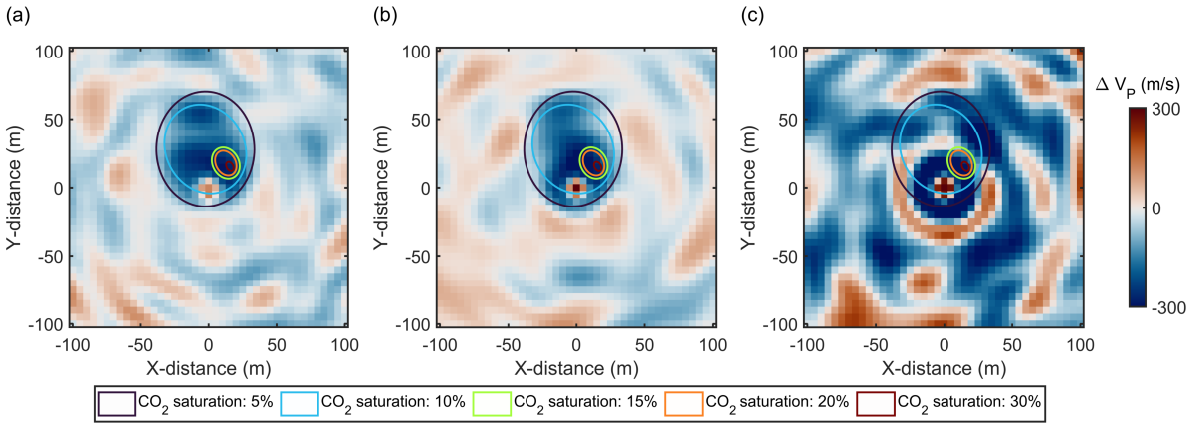

Figure 5.12: Time-lapse velocity changes in depth of 295 meters with different offsets. (a) to (c) represent results using 400, 300, and 200 meters as one-sided offset from the observation well.

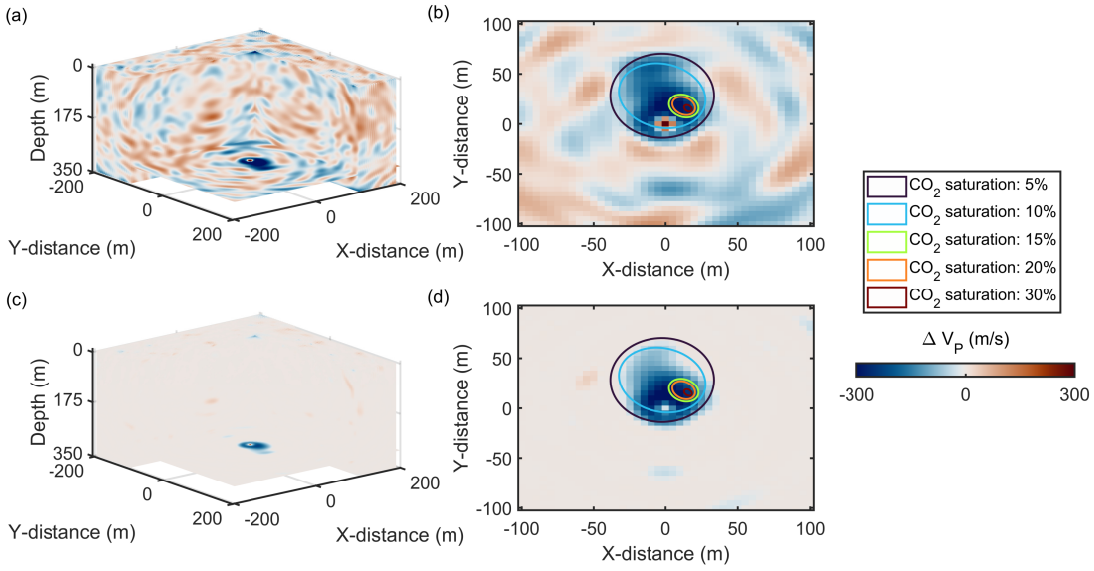

Figure 5.13: Time-lapse inversion and nullspace shuttle results with 75% offset. (a) and (b) display the models before shuttling, and (c) and (d) show the after-shuttling results. Panels (a) and (c) are 3D views, and (b) and (d) are depth slices at 295 meters.

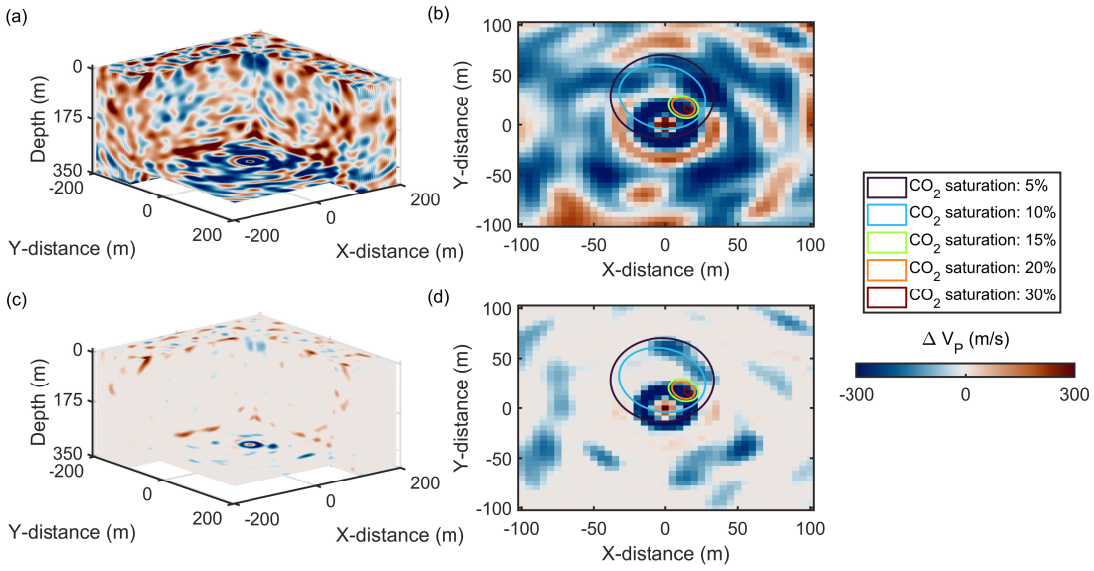

Figure 5.14: Time-lapse inversion and nullspace shuttle results with 50% offset. (a) and (b) display the models before shuttling, and (c) and (d) show the after-shuttling results. Panels (a) and (c) are 3D views, and (b) and (d) are depth slices at 295 meters.

To further assess the relationship between aperture size and spatial illumination range, I examine cross-sections along Line 1 and Line 7, as shown in Figure 5.15. The results indicate that the effective horizontal illumination range at the BBRS formation is positively correlated with the one-sided offset, as FWI becomes less capable of resolving extended lateral variations in P-wave velocity as the offset decreases. This suggests that while reducing offsets can lower acquisition costs, it does so at the expense of resolution, particularly in detecting subtle horizontal velocity changes. As a result, aperture selection must be carefully optimized to balance cost efficiency with the level of detail required for effective CO_2 plume monitoring using FWI.

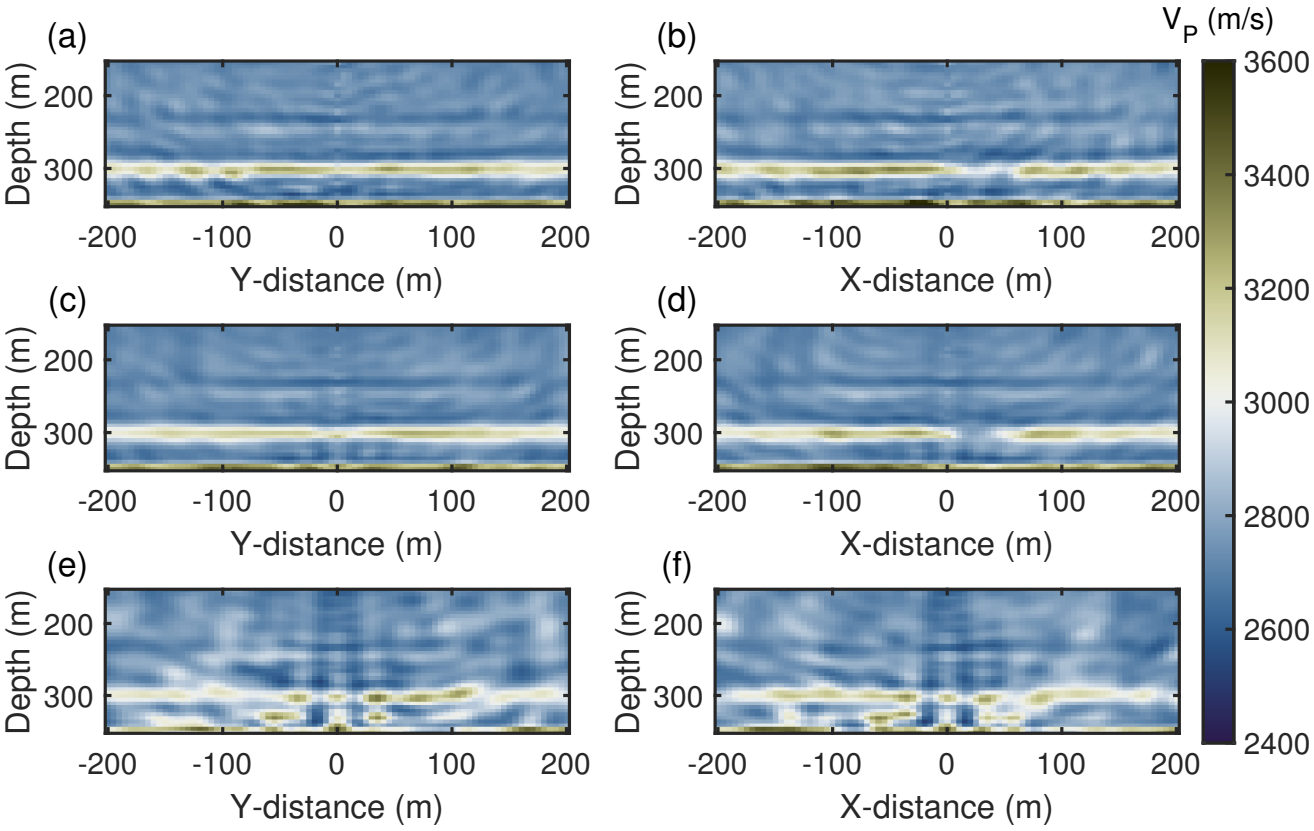

Figure 5.15: Monitor inversion sections with different apertures. (a)/(b) Cross-sections between -200 to 200 meters in Line 7/Line 1 with 400-meter offset. (c)/(d) Cross-sections between -200 to 200 meters in Line 7/Line 1 with 300-meter offset. (e)/(f) Cross-sections between -200 to 200 meters in Line 7/Line 1 with 200-meter offset. The depth range in the cross-sections is 150 to 350 meters.

5.5 Discussion

5.5.1 Towards the sparse monitoring

As shown in Figure 5.7, accurate plume detection remains possible even when the number of surface sources is reduced to about 25% of the original layout, provided that both azimuthal and aperture coverage are preserved. Figure 5.12 further demonstrates that with a moderate

aperture reduction (e.g., 75% of the original offset range), a reasonable plume delineation can still be yielded, whereas reducing the offset range to only half its original value causes time-lapse FWI to fail in recovering the plume. These results demonstrate both the robustness and the limitations of sparse monitoring under realistic acquisition and noise conditions, pointing toward a practical route for cost-effective yet reliable monitoring of CO_2 plume evolution in 4D seismic applications.

5.5.2 Role of the post-inversion refinement with targeted nullspace shuttle

Across most of the numerical experiments, the 3D targeted nullspace shuttle approach proves to be an effective post-inversion refinement strategy for addressing persistent challenges in 4D seismic FWI monitoring. As shown in Figures 5.8, 5.9, 5.10, and 5.13, it consistently enhances the clarity and spatial coherence of time-lapse velocity anomalies by mitigating non-repeatable artifacts that remain after conventional inversion. However, its robustness is not unconditional. When source density is reduced aggressively down to 12.5% of the original layout, the resulting loss of subsurface illumination leads to a pronounced deterioration in time-lapse FWI resolution. This effect is most severe at the outer plume boundaries, where lower CO_2 saturation levels (about 5%–10% in some cases) become extremely difficult to recover. A clear failure case is shown in Figure 5.14, where halving the offset produces an unsuccessful monitor FWI result. Under such circumstances, the targeted nullspace shuttle can provide little or no improvement because it will focus on refining the entire model rather than prioritizing the actual change that produces the most evident data variations. These findings indicate that while short-offset designs, such as those discussed in Cai et al. (2024), may suffice for rapid or preliminary assessments, a moderate offset that is capable of resolving enough lateral changes offers a more reliable basis for confident conformance interpretation.

5.5.3 Future work

Future research can focus on the refinement of rock physics models to enable more feasibility assessments. Incorporating more geologically complex models will allow the evaluation of sparse monitoring under conditions that better reflect real-world scenarios. Alternative time-lapse inversion strategies should also be investigated, as the present study adopts a parallel inversion approach that requires minimal tailoring between baseline and monitor FWI runs. Finally, applying the proposed methodologies to field datasets would provide critical validation of their practical applicability and performance in operational monitoring campaigns.

Looking forward, this 3D nullspace shuttle framework also opens several pathways for future research. One is its extension to time-lapse multiparameter inversion (e.g., Vp-Vs, or Vp-density), where parameter crosstalk and varying sensitivity may pose additional challenges. The pivotal discussion in 2D visco-elastic FWI can be found in Keating and Innanen (2021). Another direction involves integrating formal uncertainty quantification metrics, such as posterior variance estimation or ensemble spread, into the shuttling framework to provide quantitative confidence in the refined models. Additionally, this method could be combined with model similarity constraints or regularizations to further enhance stability and interpretability. Finally, as aforementioned, choices of the shuttling objective functions can be explored.

5.5.4 Conclusions

In this chapter, 4D numerical FWI tests are conducted to evaluate sparse monitoring of CO_2 plume migration using the Snowflake dataset from CaMI FRS. The results show that with moderate reductions in source density, the plume remains observable in most cases. Aperture appears to be a significant factor influencing plume detectability, and a critical aperture that is related to the estimated target size should be carefully considered. These

observations highlight the feasibility of sparse geo-monitoring with FWI. In addition, the targeted nullspace shuttling framework is extended to 3D FWI and evaluated in synthetic 4D seismic monitoring experiments. Most results demonstrate that the targeted nullspace shuttle can effectively delineate the plume by removing non-repeatable variations in the models. However, aperture limitations in VSP-based FWI are still factors that influence the success of this post-inversion refinement.

Chapter 6

Conclusions

6.1 Summary

In the context of the ongoing energy transition, FWI plays a crucial role in advancing subsurface imaging for various applications. While FWI theories continue to evolve, significant challenges remain in integrating FWI with energy transition projects. This thesis aims to address some of these key challenges, as discussed in Chapters 3 to 5. Specifically, I have proposed strategies for improving FWI accuracy while tracking the mechanisms and positions of SWD sources, developing uncertainty quantification methods for time-lapse FWI, and utilizing 4D FWI to capture spatial subsurface changes induced by CO_2 injection. These contributions provide insights that can inform future data acquisition strategies for carbon sequestration projects. In this section, I will summarize the key findings from my numerical investigations on these topics.

In geothermal drilling projects, accurately tracking drill bit locations and drill-bit-rock interactions is essential. While FWI provides high-resolution subsurface imaging, its effectiveness is often constrained by incomplete acquisition and limited illumination. Integrating the SWD data, which introduces independent ray paths, has the potential to significantly enhance FWI accuracy. However, incorporating SWD into FWI also introduces new uncertainties that must be carefully addressed. In Chapter 3, I develop a multi-parameter FWI algorithm that treats source radiation patterns and positions as additional unknowns alongside velocity and density. Through numerical experiments, I demonstrate that the inclusion of SWD data improves FWI results by providing additional ray paths that enhance model accuracy. The integration of SWD data not only refines the inversion of elastic properties but also allows for precise tracking of subsurface source locations and radiation patterns. These

findings suggest that SWD data has the dual benefit of improving inversion quality while enabling real-time monitoring of drilling operations, highlighting its potential for enhanced subsurface imaging in geothermal applications.

Due to the nonlinearity of the FWI problem, the presence of various uncertainties between baseline and monitor surveys in time-lapse FWI, and the often sparse or variable data coverage in these applications, incorporating robust uncertainty quantification methods is crucial. HMC and SVGD provide effective approaches for assessing uncertainty in FWI, each with distinct advantages. In Chapter 4, I investigate the feasibility of HMC and SVGD in time-lapse VSP-FWI through synthetic experiments simulating a CO_2 storage monitoring scenario. The focus is on their ability to quantify uncertainty, as well as their respective strengths and limitations. HMC excels at exploring complex, nonlinear model spaces due to its rigorous sampling-based formulation, but its high computational cost and limited scalability pose practical challenges. In contrast, SVGD offers a more resource-efficient and scalable approach, although it can be sensitive to initial distributions and inherent limitations in FWI as being deterministic. The numerical tests with a comparable computational budget indicate that SVGD is particularly attractive when computational resources are limited or rapid assessment is required, offering potential for optimization and near real-time deployment in larger-scale or 3D time-lapse FWI applications.

In Chapter 5, I evaluate the feasibility of 4D seismic monitoring through synthetic acoustic frequency-domain FWI experiments. Using the Snowflake VSP framework with walk-away and walk-around geometries, I refine the CaMI well log and construct synthetic models incorporating plume migration from a 60-tonne CO_2 injection. Synthetic baseline and monitor FWI tests on the Snowflake dataset demonstrate the potential for sparse monitoring. Furthermore, I extend the targeted nullspace shuttle technique to 3D time-lapse FWI as a post-inversion refinement, highlighting authentic structural changes while suppressing non-repeatable artifacts. The results indicate that moderate reductions in source density still

allow reasonable plume detection, but a critical aperture, estimated from the spatial layout of the target, must be maintained.

6.2 Future work

Beyond the future work discussed separately in Chapters 3, 4, and 5, there is considerable potential in combining the approaches presented in this thesis. First, the uncertainty quantification algorithms could be extended to elastic multi-parameter FWI, enabling a robust assessment of uncertainties within the model–source elastic FWI framework discussed in Chapter 3, and providing quantitative insights for SWD applications to support improved decision-making. Second, 3D frequency-domain acoustic FWI can be generalized to a multi-parameter framework by incorporating frequency-dependent parameters, such as attenuation. Although this increases the number of inversion variables and computational demands, further development of efficient 3D solvers could mitigate these challenges. Another promising direction is applying the uncertainty measurement approaches and targeted nullspace shuttle technique to multi-parameter time-lapse scenarios, where investigating temporal interactions among multiple, inherently coupled parameters in elastic FWI will be challenging but rewarding. Finally, all methods developed in this thesis have the potential to be applied to real data, paving the way for practical implementation and field validation.

Bibliography

Aki, K., and Richards, P., 2002, Quantitative Seismology, Geology Seismology: University Science Books.

Amestoy, P., Duff, I. S., Koster, J., and L'Excellent, J.-Y., 2001, A fully asynchronous multifrontal solver using distributed dynamic scheduling: SIAM Journal on Matrix Analysis and Applications, **23**, No. 1, 15–41.

Archer, R., 2020, 20 - geothermal energy, *in* Letcher, T. M., Ed., Future Energy (Third Edition), Elsevier, third edition edn., 431–445.

Armijo, L., 1966, Minimization of functions having Lipschitz continuous first partial derivatives.: Pacific Journal of Mathematics, **16**, No. 1, 1 – 3.

Arts, R. J., Eiken, O., Chadwick, A., Zweigel, P., van der Meer, L., and Zinszner, B., 2004, Monitoring of CO_2 injected at Sleipner using time-lapse seismic data: Energy, **29**, 1383–1392.

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P., and Virieux, J., 2015, Time-lapse seismic imaging using regularized full-waveform inversion with a prior model: which strategy?: Geophysical Prospecting, **63**, No. 1, 78–98.

Auriol, J., Kazemi, N., and Niculescu, S.-I., 2021, Sensing and computational frameworks for improving drill-string dynamics estimation: Mechanical Systems and Signal Processing, **160**, 107,836.

Auriol, J., Kazemi, N., Shor, R. J., Innanen, K. A., and Gates, I. D., 2019, A sensing and computational framework for estimating the seismic velocities of rocks interacting with the drill bit: IEEE Transactions on geoscience and remote sensing, **58**, No. 5, 3178–3189.

Backus, G. E., and Gilbert, J. F., 1967, Numerical applications of a formalism for geophysical inverse problems: Geophysical Journal International, **13**, No. 1-3, 247–276.

Backus, G. E., and Gilbert, J. F., 1968, The Resolving Power of Gross Earth Data: Geophysical Journal International, **16**, No. 2, 169–205.

Barbier, E., 2002, Geothermal energy technology and current status: an overview: Renewable and Sustainable Energy Reviews, **6**, No. 1, 3–65.

Bellman, R., 2003, Dynamic Programming, Dover Books on Computer Science Series: Dover Publications.

Bertelli, L., and di Cesare, F., 1999, Improving the subsurface geological model while drilling: First Break, **17**, No. 6.

Betancourt, M., 2017, A conceptual introduction to hamiltonian monte carlo: arXiv: Methodology.

Bishop, C. M., and Nasrabadi, N. M., 2006, Pattern recognition and machine learning, vol. 4: Springer.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D., 2017, Variational inference: A review for statisticians: Journal of the American Statistical Association, **112**, No. 518, 859–877.

Bregman, N. D., Chapman, C. H., and Bailey, R. C., 1989, Travel time and amplitude analysis in seismic tomography: Journal of Geophysical Research: Solid Earth, **94**, No. B6, 7577–7587.

Brittan, J., and Jones, I., 2019, FWI evolution — From a monolith to a toolkit: The Leading Edge, **38**, No. 3, 179–184.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., 2011, Handbook of Markov Chain Monte Carlo, CRC press.

Brossier, R., Etienne, V., Operto, S., and Virieux, J., 2010, Frequency-domain numerical modelling of visco-acoustic waves with finite-difference and finite-element discontinuous Galerkin methods, *in* Dissanayake, D., Ed., Acoustic Waves, Mechanical Engineering, SCIYO editions, 434 p.

Brossier, R., Operto, S., and Virieux, J., 2009, Seismic imaging of complex structures by 2D elastic frequency-domain full-waveform inversion: Geophysics, **74**, WCC105–WCC118.

Brossier, R., Virieux, J., and Operto, S., 2008, Parsimonious finite-volume frequency-domain method for 2-D P–SV-wave modelling: Geophysical Journal International, **175**, No. 2, 541–559.

Broyden, C. G., 1970, The convergence of a class of double-rank minimization algorithms 1. General considerations: IMA Journal of Applied Mathematics, **6**, No. 1, 76–90.

Bunks, G., Saleck, F. M., Zaleski, S., and Chavent, G., 1995, Multiscale seismic waveform inversion: Geophysics, **60**, No. 5, 1457–1473.

Butzer, S., 2015, 3D elastic time-frequency full-waveform inversion: Ph.D. thesis.

Butzer, S., Kurzmann, A., and Bohlen, T., 2013, 3D elastic full-waveform inversion of small-scale heterogeneities in transmission geometry: Geophysical Prospecting, **61**, No. 6, 1238–1251.

Cai, X., Innanen, K. A., Hu, Q., Eaid, M., Keating, S., Hall, K. W., and Lawton, D. C., 2024, VSP imaging of CO2 injection with rapid-repeat time lapse full waveform inversion, CREWES Research Report, 34, 5, 15.

Castellanos, C., Etienne, V., Hu, G., Operto, S., Brossier, R., and Virieux, J., 2011, Algorithmic and methodological developments towards full waveform inversion in 3D elastic media, 2793–2798.

Chadwick, A., Williams, G., Delepine, N., Clochard, V., Labat, K., Sturton, S., Buddensiek, M.-L., Dillen, M., Nickel, M., Lima, A. L., Arts, R., Neele, F., and Rossi, G., 2010, Quantitative analysis of time-lapse seismic monitoring data at the Sleipner CO_2 storage operation: The Leading Edge, **29**, No. 2, 170–177.

Cheadle, S. P., Brown, R. J., and Lawton, D. C., 1991, Orthorhombic anisotropy: A physical seismic modeling study: Geophysics, **56**, No. 10, 1603–1613.

Chi, B., Dong, L., and Liu, Y., 2015, Correlation-based reflection full-waveform inversion: Geophysics, **80**, No. 4, R189–R202.

Chwialkowski, K., Strathmann, H., and Gretton, A., 2016, A kernel test of goodness of fit, *in* Balcan, M. F., and Weinberger, K. Q., Eds., Proceedings of the 33rd International Conference on Machine Learning, vol. 48 of *Proceedings of Machine Learning Research*, PMLR, New York, NY, USA, 2606–2615.

Danecek, P., and Seriani, G., 2008, An efficient parallel chebyshev pseudo-spectral method for large scale 3d seismic forward modelling, *in* 70th EAGE Conference and Exhibition – Workshops and Technical Conference, European Association of Geoscientists and Engineers.

Davey, K., 2009, What Is Gibbs's canonical distribution?: Philosophy of Science, **76**.

de Lima, P. D. S., Corso, G., Ferreira, M. S., and de Araújo, J. M., 2023, Acoustic full waveform inversion with Hamiltonian Monte Carlo method: Physica A: Statistical Mechanics and its Applications, **617**, No. C.

de Lima, P. D. S., Ferreira, M. S., Corso, G., and de Araújo, J. M., 2024, Bayesian time-lapse full waveform inversion using hamiltonian monte carlo: Geophysical Prospecting, **72**, No. 9, 3381–3398.

Deal, M., and Nolet, G., 2007, Nullspace shuttles: Geophysical Journal International, **124**, 372 – 380.

Dosso, S. E., Dettmer, J., Steininger, G., and Holland, C. W., 2014, Efficient trans-dimensional bayesian inversion for geoacoustic profile estimation: Inverse Problems, **30**, No. 11, 114,018.

Duane, S., Kennedy, A., Pendleton, B. J., and Roweth, D., 1987, Hybrid Monte Carlo: Physics Letters B, **195**, No. 2, 216–222.

Dubbeldam, D., Calero, S., Ellis, D. E., and Snurr, R. Q., 2016, Raspa: molecular simulation software for adsorption and diffusion in flexible nanoporous materials: Molecular Simulation, **42**, No. 2, 81–101.

Duong, C., Bower, C., Hume, K., Rock, L., and Tessarolo, S., 2019, Quest carbon capture and storage offset project: Findings and learnings from 1st reporting period: International Journal of Greenhouse Gas Control, **89**, 65–75.

Dziejarski, B., Krzyżyńska, R., and Andersson, K., 2023, Current status of carbon capture, utilization, and storage technologies in the global economy: A survey of technical assessment: Fuel, **342**, 127,776.

Elhatisari, S., Lee, D., Rupak, G., Epelbaum, E., Krebs, H., Lähde, T. A., Luu, T., and Meißner, U.-G., 2015, Ab initio alpha–alpha scattering: Nature, **528**, No. 7580, 111–114.

Engquist, B., and Froese, B., 2014, Application of the Wasserstein metric to seismic signals: Communications in Mathematical Sciences, **12**, No. 5, 979–988.

Esser, E., Guasch, L., van Leeuwen, T., Aravkin, A. Y., and Herrmann, F. J., 2018, Total variation regularization strategies in full-waveform inversion: SIAM Journal on Imaging Sciences, **11**, No. 1, 376–406.

Fichtner, A., 2010, Full Seismic Waveform Modelling and Inversion, Advances in Geophysical and Environmental Mechanics and Mathematics: Springer, Berlin, Heidelberg.

Fichtner, A., and Zunino, A., 2019, Hamiltonian nullspace shuttles: Geophysical Research Letters, **46**, No. 2, 644–651.

Fichtner, A., Zunino, A., and Gebraad, L., 2018, Hamiltonian Monte Carlo solution of tomographic inverse problems: Geophysical Journal International, **216**, No. 2, 1344–1363.

Fichtner, A., Zunino, A., Gebraad, L., and Boehm, C., 2021, Autotuning Hamiltonian Monte Carlo for efficient generalized nullspace exploration: Geophysical Journal International, **227**, No. 2, 941–968.

Fletcher, R., 1970, A new approach to variable metric algorithms: The Computer Journal, **13**, No. 3, 317–322.

Fu, X., and Innanen, K. A., 2022, A time-domain multisource Bayesian/Markov chain Monte Carlo formulation of time-lapse seismic waveform inversion: Geophysics, **87**, No. 4, R349–R361.

Furre, A.-K., Eiken, O., Alnes, H., Vevatne, J. N., and KiÃ¦r, A. F., 2017, 20 Years of Monitoring CO_2-injection at Sleipner: Energy Procedia, **114**, 3916–3926, 13th International Conference on Greenhouse Gas Control Technologies, GHGT-13, 14-18 November 2016, Lausanne, Switzerland.

Futterman, W. I., 1962, Dispersive body waves: Journal of Geophysical Research (1896-1977), **67**, No. 13, 5279–5291.

Gebraad, L., Boehm, C., and Fichtner, A., 2020, Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo: Journal of Geophysical Research: Solid Earth, **125**, No. 3, e2019JB018,428.

Germay, C., Denoël, V., and Detournay, E., 2009, Multiple mode analysis of the self-excited vibrations of rotary drilling systems: Journal of Sound and Vibration, **325**, No. 1, 362–381.

Goertz, A., Atkinson, B., Thiem, T., and Bergfjord, E. V., 2020, Reservoir imaging-while-drilling with PRM arrays: First Break, **38**, No. 11, 77–83.

Goertz-Allmann, B., Jordan, M., Bauer, R., Oye, V., and Greenberg, S., 2017, Integrating active with passive seismic data to best constrain CO_2 injection monitoring, *in* EAGE/SEG Research Workshop 2017, Aug 2017, European Association of Geoscientists & Engineers.

Goldfarb, D., 1970, A family of variable-metric methods derived by variational means: Mathematics of Computation, **24**, 23–26.

Groos, L., Schäfer, M., Forbriger, T., and Bohlen, T., 2017, Application of a complete workflow for 2D elastic full-waveform inversion to recorded shallow-seismic Rayleigh waves: Geophysics, **82**, No. 2, R109–R117.

Guasch, L., Warner, M., Nangoo, T., Morgan, J., Umpleby, A., Stekl, I., and Shah, N., 2012, Elastic 3d full-waveform inversion, *in* SEG Technical Program Expanded Abstracts 2012, Society of Exploration Geophysicists, 1–5.

Guennebaud, G., Jacob, B. et al., 2010, Eigen v3, http://eigen.tuxfamily.org.

Guitton, A., and Symes, W. W., 2003, Robust inversion of seismic data using the huber norm: Geophysics, **68**, No. 4, 1310–1319.

Gutmann, M., and Hyvärinen, A., 2012, Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics.: Journal of Machine Learning Research, **13**, 307–361.

Hall, K. W., Bertram, K. L., Bertram, M., Innanen, K., and Lawton, D. C., 2019, Simultaneous accelerometer and optical fibre multi-azimuth walk-away VSP experiment: Newell County, Alberta, Canada, 5340–5344.

Hamilton, W. R., 1834, On a general method in dynamics; by which the study of the motions of all free systems of attracting or repelling points is reduced to the search and differentiation of one central relation, or characteristic function: Philosophical Transactions of the Royal Society of London, **124**, 247–308.

Honkela, A., Peltonen, J., Topa, H., Charapitsa, I., Matarese, F., Grote, K., Stunnenberg, H. G., Reid, G., Lawrence, N. D., and Rattray, M., 2015, Genome-wide modeling of transcription kinetics reveals patterns of rna production delays: Proceedings of the National Academy of Sciences, **112**, No. 42, 13,115–13,120.

Hu, Q., Grana, D., and Innanen, K. A., 2022, Feasibility of seismic time-lapse monitoring of CO_2 with rock physics parametrized full waveform inversion: Geophysical Journal International, **233**, No. 1, 402–419.

Hu, Q., and Innanen, K. A. H., 2021, Rock physics analysis of well-log data, CREWES Research Report, 33, 16, 20.

Hustedt, B., Operto, S., and Virieux, J., 2004, Mixed-grid and staggered-grid finite-difference methods for frequency-domain acoustic wave modelling: Geophysical Journal International, **157**, No. 3, 1269–1296.

Innanen, K. A., 2014, Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion: Geophysical Journal International, **199**, No. 2, 717–734.

Isaac, J. H., and Lawton, D. C., 2024, Brooks revisited, CREWES Research Report, 28, 33, 10.

Iserles, A., 1986, Generalized Leapfrog Methods: IMA Journal of Numerical Analysis, **6**, No. 4, 381–392.

Jaakkola, T. S., and Jordan, M. I., 1997, A variational approach to Bayesian logistic regression models and their extensions, *in* Madigan, D., and Smyth, P., Eds., Proceedings of

the Sixth International Workshop on Artificial Intelligence and Statistics, vol. R1 of *Proceedings of Machine Learning Research*, PMLR, 283–294, reissued by PMLR on 30 March 2021.

James W. Rector, I., and Hardage, B. A., 1992, Radiation pattern and seismic waves generated by a working roller-cone drill bit: Geophysics, **57**, No. 10, 1319–1333.

Jannane, M., Beydoun, W., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G., Singh, S., Snieder, R., Tarantola, A., Trezeguet, D., and Xie, M., 1989, Wavelengths of earth structures that can be resolved from seismic reflection data: Geophysics, **54**, No. 7, 906–910.

Jo, C., Shin, C., and Suh, J., 1996, An optimal 9-point, finite-difference, frequency-space, 2-D Scalar wave extrapolator: Geophysics, **61**, 529–537.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K., 1998, An Introduction to Variational Methods for Graphical Models, Springer Netherlands, Dordrecht, 105–161.

Kamath, N., Tsvankin, I., and Naeini, E. Z., 2017, Facies-constrained FWI: Toward application to reservoir characterization: The Leading Edge, **36**, No. 11, 924–930.

Katz, L. J., 1984, Method and System for Seismic Continuous Bit Positioning. U.S. Patent No. 4,460,059.

Kazemi, N., Auriol, J., Innanen, K. A. H., Shor, R., and Gates, I., 2021, Successive full-waveform inversion of surface seismic and seismic-while-drilling datasets without low frequencies, **2021**, No. 1, 1–5.

Kazemi, N., and Sacchi, M. D., 2014, Sparse multichannel blind deconvolution: Geophysics, **79**, No. 5, V143–V152.

Kazemi, N., Shor, R., and Innanen, K. A. H., 2018, Illumination compensation with seismic-while-drilling plus surface seismic imaging, *in* 78th EAGE Conference and Exhibition, vol. 2018, European Association of Geoscientists and Engineers, 1–5.

Kazemi, N., Wong, J., Zhang, H., Bertram, K., Innanen, K., and Shor, R., 2020, Seismic illumination analysis through physical modeling, *in* GeoConvention 2020, GeoConvention, Calgary, Canada.

Keating, S., and Innanen, K. A., 2020, Parameter crosstalk and leakage between spatially separated unknowns in viscoelastic full-waveform inversion: Geophysics, **85**, No. 4, R397–R408.

Keating, S. D., and Innanen, K. A., 2021, Null-space shuttles for targeted uncertainty analysis in full-waveform inversion: Geophysics, **86**, No. 1, R63–R76.

Keating, S. D., and Innanen, K. A., 2024, Targeted nullspace shuttling in time-lapse full-waveform inversion: Geophysics, **89**, No. 6, R541–R549.

Kerrison, H., Fallon, P., Kaszycka, E., Cichy, K., Ratcliffe, A., and Masmoudi, N., 2021, Impact of streamer acquisition geometry on FWI Imaging, **2021**, No. 1, 1–5.

Khaled, O. S., Al-Ateeqi, A. M., James, A. R., and Meehan, R. J., 1996, Seismic-while-drilling in Kuwait results and applications: GeoArabia, **1**, No. 4, 531–550.

Kim, Y., Liu, Q., and Tromp, J., 2011, Adjoint centroid-moment tensor inversions: Geophysical Journal International, **186**, No. 1, 264–278.

Kingma, D. P., and Ba, J., 2014, Adam: A method for stochastic optimization: CoRR, **abs/1412.6980**.

Klaveness, A., 1980, Seismic well logging system and method. U.S. Patent No.5,012,453.

Kolkman-Quinn, B., Lawton, D., Bertram, M., and Macquet, M., 2024, Sparse seismic monitoring of geologic carbon storage, *in* Proceedings of the 17th Greenhouse Gas Control Technologies Conference (GHGT-17), GHGT, Oslo, Norway.

Kolsky, H., 1956, Lxxi. the propagation of stress pulses in viscoelastic solids: The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics, **1**, No. 8, 693–710.

Kotsi, M., Malcolm, A., and Ely, G., 2020, Time-lapse full-waveform inversion using hamiltonian monte carlo: A proof of concept, 845–849.

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., and Blei, D. M., 2016, Automatic differentiation variational inference: J. Mach. Learn. Res., **18**, 14:1–14:45.

Kullback, S., and Leibler, R. A., 1951, On Information and Sufficiency: The Annals of Mathematical Statistics, **22**, No. 1, 79 – 86.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations, *in* Bednar, J. B., Robinson, E., and Weglein, A., Eds., Conference on Inverse Scattering—Theory and Application, SIAM, Philadelphia, 206–220.

Lawton, D. C., Dongas, J., Osadetz, K., Saeedfar, A., and Macquet, M., 2019, Development and Analysis of a Geostatic Model for Shallow CO_2 Injection at the Field Research Station, Southern Alberta, Canada, Cambridge University Press, 280–296.

Leaney, W. S., Arteaga, E., Lawton, D., and Innanen, K. A., 2019, Sonic anisotropy and monoclinic modelling at the cami site, *in* GeoConvention 2019.

Leaney, W. S., Chapman, C., and Bratton, T., 2022, Anisotropy in das data and ava for CO_2 monitoring, *in* GeoConvention 2022.

Li, J., Keating, S., Kazemi, N., Shor, R., and Innanen, K. A., 2022, Simultaneous waveform inversion of seismic-while-drilling data for P-wave velocity, density, and source parameters, *in* Second International Meeting for Applied Geoscience & Energy, Society of Exploration Geophysicists, 942–946.

Li, J., Pike, K., Innanen, K. A. H., and Hall, K. W., 2024, 4D synthetic time-lapse FWI experiments for CO_2 monitoring configured for the Snowflake dataset, CREWES Research Report, 36, 21, 23.

Liu, Q., Lee, J., and Jordan, M., 2016, A Kernelized Stein discrepancy for goodness-of-fit tests, *in* Balcan, M. F., and Weinberger, K. Q., Eds., Proceedings of The 33rd International Conference on Machine Learning, vol. 48 of *Proceedings of Machine Learning Research*, PMLR, New York, New York, USA, 276–284.

Liu, Q., and Peter, D., 2020, Square-root variable metric-based nullspace shuttle: A characterization of the nonuniqueness in elastic full-waveform inversion: Journal of Geophysical Research: Solid Earth, **125**, No. 2, e2019JB018,687.

Liu, Q., and Wang, D., 2016, Stein variational gradient descent: a general purpose bayesian inference algorithm, *in* Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16, Curran Associates Inc., Red Hook, NY, USA, 2378–2386.

Liu, Q., and Wang, D., 2019, Stein Variational Gradient Descent: A General purpose Bayesian inference algorithm.

Lumley, D. E., 2001, Time-lapse seismic reservoir monitoring: Geophysics, **66**, No. 1, 50–53.

Ma, T., Chen, P., and Zhao, J., 2016, Overview on vertical and directional drilling technologies for the exploration and exploitation of deep petroleum resources: Geomechanics and Geophysics for Geo-Energy and Geo-Resources, **2**, 365–395.

Macquet, M., and Lawton, D. C., 2019, Exploring continuous seismic data for monitoring CO_2 injection at the CaMI Field Research Station, Alberta, Canada, 4913–4917.

Macquet, M., Lawton, D. C., Saeedfar, A., and Osadetz, K. G., 2019, A feasibility study for detection thresholds of CO_2 at shallow depths at the cami field research station, newell county, alberta, canada: Petroleum Geoscience, **25**, No. 4, 509–518.

Marfurt, K. J., 1984, Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations: Geophysics, **49**, No. 5, 533–549.

Martinez, A., Useche, M., and Guerra, R., 2020, Use of fiber optic acoustics to improve drilling efficiency and well placement, *in* OTC Offshore Technology Conference, vol. Day 1 Mon, May 04, 2020.

Métivier, L., Brossier, R., Operto, S., and Virieux, J., 2015, Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: preconditioned truncated Newton approach, 1198–1203.

Métivier, L., Brossier, R., Virieux, J., and Operto, S., 2013, Full waveform inversion and the Truncated Newton method: SIAM Journal on Scientific Computing, **35**, No. 2, B401–B437.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., 1953, Equation of state calculations by fast computing machines: The Journal of Chemical Physics, **21**, No. 6, 1087–1092.

Min, D.-J., Shin, C., Pratt, R. G., and Yoo, H. S., 2003, Weighted-averaging finite-element method for 2D elastic wave equations in the frequency domain: Bulletin of the Seismological Society of America, **93**, No. 2, 904–921.

Miranda, F., Aleotti, L., Abramo, F., Poletto, F., Craglietto, A., Persoglia, S., and Rocca,

F., 1996, Impact of the seismic while drilling technique on exploration wells: First Break, **14**, No. 2.

Mispel, J., Furre, A., Sollid, A., and Maaø, F. A., 2019, High frequency 3d fwi at sleipner: A closer look at the CO_2 plume, *in* 81st EAGE Conference and Exhibition 2019, European Association of Geoscientists and Engineers, 1–5.

Morin, D., 2008, Introduction to Classical Mechanics: With Problems and Solutions: Cambridge University Press, Cambridge.

Mosegaard, K., and Tarantola, A., 1995, Monte carlo sampling of solutions to inverse problems: Journal of Geophysical Research: Solid Earth, **100**, No. B7, 12,431–12,447.

Mothi, S., Schwarz, K., and Zhu, H., 2014, Impact of full-azimuth and long-offset acquisition on Full Waveform Inversion in deep water Gulf of Mexico, 317–321.

Nakata, R., Nakata, N., Girard, A., Ichikawa, M., Kato, A., Lumley, D., and Xue, Z., 2022, Time-lapse crosswell seismic monitoring of CO_2 injection at the Nagaoka CCS site using elastic full-waveform inversion, 802–806.

Naville, C., Serbutoviez, S., Throo, A., Vincké, O., and Cecconi, F., 2004, Seismic while drilling (SWD) techniques with downhole measurements, introduced by Ifp and its Partners in 1990-2000: Oil & Gas Science and Technology-revue De L Institut Francais Du Petrole, **59**, 371–403.

Nawaz, M. A., and Curtis, A., 2018, Variational Bayesian inversion (VBI) of quasi-localized seismic attributes for the spatial distribution of geological facies: Geophysical Journal International, **214**, No. 2, 845–875.

Nawaz, M. A., and Curtis, A., 2019, Rapid discriminative Variational Bayesian inversion of geophysical data for the spatial distribution of geological properties: Journal of Geophysical Research: Solid Earth, **124**, No. 6, 5867–5887.

Neal, R. M., 1993, Probabilistic inference using markov chain monte carlo methods, Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto.

Neal, R. M., 1996, Bayesian Learning for Neural Networks: Springer-Verlag, Berlin, Heidelberg.

Nemeth, T., Wu, C., and Schuster, G. T., 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, No. 1, 208–221.

Nocedal, J., and Wright, S., 2006, Numerical Optimization, Springer Series in Operations Research and Financial Engineering: Springer New York.

Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Metivier, L., and Virieux, J., 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: The Leading Edge, **32**, No. 9, 1040–1054.

Operto, S., Virieux, J., Amestoy, P., L'Excellent, J.-Y., Giraud, L., and Ali, H. B. H., 2007, 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study: Geophysics, **72**, No. 5, SM195–SM211.

Ostrander, W. J., 1984, Plane-wave reflection coefficients for gas sands at nonnormal angles of incidence: Geophysics, **49**, No. 10, 1637–1648.

Pan, W., Innanen, K., and Geng, Y., 2018, Elastic full-waveform inversion and parametrization analysis applied to walk-away vertical seismic profile data for unconventional (heavy oil) reservoir characterization: Geophysical Journal International, **213**, 1934–1968.

Pan, W. Y., Innanen, K. A., Margrave, G. F., Fehler, M. C., Fang, X. D., and Li, J. X., 2016, Estimation of elastic constants for HTI media using Gauss-Newton and full-Newton multiparameter full-waveform inversion: Geophysics, **81**, No. 5, R275–R291.

Park, S.-E., and Oh, J.-W., 2023, 3D frequency-domain acoustic full-waveform inversion using diffraction-angle filtering: Journal of Applied Geophysics, **218**, 105,200.

Parker, P., Northrop, S., Valencia, J., Foglesong, R., and Duncan, W., 2011, CO_2 management at ExxonMobil's LaBarge field, Wyoming, USA: Energy Procedia, **4**, 5455–5470.

Parker, R. L., 1977, Understanding Inverse Theory: Annual Review of Earth and Planetary Sciences, **5**, 35.

Pike, K., Innanen, K. A. H., and Keating, S., 2024a, Time-lapse nullspace shuttles: VSP vs surface acquisition, shuttling to zero, and sparse monitoring prospects, CREWES Research Report, 36, 29, 18.

Pike, K. A., Innanen, K. A., and Keating, S. D., 2024b, Targeted nullspace shuttles as enablers of time-lapse co_2 monitoring with sparse acquisition, *in* Fourth International Meeting for Applied Geoscience and Energy (IMAGE) 2024, Society of Exploration Geophysicists and American Association of Petroleum Geologists (SEG/AAPG), 2623–2627.

Plessix, R., Stopin, A., Kuehl, H., Goh, V., and Overgaag, K., 2016, Visco-acoustic full waveform inversion, **2016**, No. 1, 1–5.

Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, No. 2, 495–503.

Plessix, R.-E., 2009, Three-dimensional frequency-domain full-waveform inversion with an iterative solver: Geophysics, **74**, No. 6, WCC149–WCC157.

Plessix, R.-E., and Perkins, C., 2010, Thematic set: Full waveform inversion of a deep water ocean bottom seismometer dataset: First Break, **28**, No. 4.

Podgornova, O., Leaney, S., and Liang, L., 2018, Resolution of VTI anisotropy with elastic full-waveform inversion: Theory and basic numerical examples: Geophysical Journal International, **214**, 200–218.

Poletto, F., Corubolo, P., Farina, B., Schleifer, A., Pollard, J., Peronio, M., and Böhm, G., 2012, Drill-bit SWD and seismic interferometry for imaging around geothermal wells, *in* SEG Technical Program Expanded Abstracts 2011, Society of Exploration Geophysicists, 4319–4324.

Poletto, F., Goertz, A., Bellezza, C., Bergfjord, E. V., Corubolo, P., Lindgård, J. E., and Moskvil, L. M., 2022, Seismic-while-drilling by drill-bit source and large-aperture ocean-bottom array: Geophysics, **87**, No. 2, D33–D45.

Poletto, F., and Miranda, F., 2004, Seismic While Drilling: Fundamentals of Drill-Bit Seismic for Exploration, Handbook of Geophysical Exploration: Seismic Exploration: Elsevier Science.

Poletto, F., and Miranda, F., 2022, Seismic While Drilling: Fundamentals of Drill-Bit Seismic for Exploration: Elsevier.

Poletto, F., Miranda, F., Farina, B., and Schleifer, A., 2020, Seismic-while-drilling drill-bit source by ground force: concept and application: Geophysics, **85**, No. 3, MR167–MR178.

Poletto, F., Petronio, L., Malusa, M., Schleifer, A., Corubolo, P., Bellezza, C., Miranda, F., Miandro, R., and Gressetvold, B., 2004, Prediction and 3D imaging while drilling by drill-bit 3D RVSP: World oil, **225**.

Pratt, R., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888–901.

Pratt, R., Shin, C., and Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341 – 362.

Pratt, R., and Shipp, R., 1999, Seismic waveform inversion in the frequency domain, part 2: Fault delineation in sediments using crosshole data: Geophysics, **64**, 902–914.

Pratt, R. G., 1990, Frequency-domain elastic wave modeling by finite differences: A tool for crosshole seismic imaging: Geophysics, **55**, No. 5, 626–632.

Pratt, R. G., Song, Z.-M., Williamson, P., and Warner, M., 1996, Two-dimensional velocity models from wide-angle seismic data by wavefield inversion: Geophysical Journal International, **124**, No. 2, 323–340.

Regier, J., Miller, A., McAuliffe, J., Adams, R., Hoffman, M., Lang, D., Schlegel, D., and Prabhat, 2015, Celeste: Variational inference for a generative model of astronomical images, arXiv:1506.03127, `1506.03127`.

Richard, T., Germay, C., and Detournay, E., 2004, Self-excited stick–slip oscillations of drill bits: Comptes Rendus Mécanique, **332**, No. 8, 619–626.

Rittgers, J., Revil, A., Mooney, M., Karaoulis, M., Wodajo, L., and Hickey, C., 2016, Time-lapse joint inversion of geophysical data with automatic joint constraints and dynamic attributes: Geophysical Journal International, **207**, No. 3, 1401–1419.

Riyanti, C. D., Erlangga, Y. A., Plessix, R.-E., Mulder, W. A., Vuik, C., and Oosterlee, C., 2006, A new iterative solver for the time-harmonic wave equation: Geophysics, **71**, No. 5, E57–E63.

Rocca, F., Vassallo, M., and Bernasconi, G., 2005, Three-dimensional seismic-while-drilling (SWD) migration in the angular frequency domain: Geophysics, **70**, No. 6, S111–S120.

Romdhane, A., Querendez, E., and Ravaut, C., 2014, CO_2 Thin-layer Detection at the Sleipner Field with Full Waveform Inversion: Application to Synthetic and Real Data: Energy Procedia, **51**, 281–288.

Saad, Y., 2003, Domain decomposition methods, *in* Iterative Methods for Sparse Linear Systems, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 451–493.

Schenk, O., and Gärtner, K., 2011, PARDISO, Springer US, Boston, MA, 1458–1464.

Schuster, G. T., Yu, J., Sheng, J., and Rickett, J., 2004, Interferometric/daylight seismic imaging: Geophysical Journal International, **157**, No. 2, 838–852.

Seah, S., Nimmrichter, S., and Scarani, V., 2018, Work production of quantum rotor engines: New Journal of Physics, **20**, No. 4, 043,045.

Shanno, D. F., 1970, Conditioning of Quasi-Newton Methods for Function Minimization: Mathematics of Computation, **24**, No. 111, 647–656.

Sharmin, T., Khan, N. R., Akram, M. S., and Ehsan, M. M., 2023, A state-of-the-art review on geothermal energy extraction, utilization, and improvement strategies: Conventional, hybridized, and enhanced geothermal systems: International Journal of Thermofluids, **18**, 100,323.

Sheriff, R. E., and Geldart, L. P., 1995, Exploration Seismology: Cambridge University Press, 2 edn.

Silvestrov, I., Bakulin, A., Aldawood, A., Hemyari, E., and Egorov, A., 2023, Improving shallow and deep seismic-while-drilling with a downhole pilot in a desert environment: Geophysics, **88**, No. 1, D1–D12.

Sirgue, L., Barkved, O. I., Dellinger, J., Etgen, J., Albertin, U., and Kommedal, J. H., 2010, Thematic set: Full waveform inversion: The next leap forward in imaging at valhall: First Break, **28**, No. 4.

Staron, P., Arens, G., and Gros, P., 1988, Method of instantaneous acoustic logging within a wellbore. U.S. Patent No. 4,718,048.

Stein, C. M., 1972, A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, *in* Le Cam, L. M., and Neyman, J., Eds., Proceedings

of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory, University of California Press, Berkeley, CA, 583–602.

Stekl, I., and Pratt, R. G., 1998, Accurate viscoelastic modeling by frequency-domain finite differences using rotated operators: Geophysics, **63**, No. 5, 1779–1794.

Stock, G., David, S., Cho, E., Kroode, F., Baldock, S., Stokes, S., and Zuberi, A., 2024, Imaging a CO_2 Plume Over the Sleipner CCS Facility Using FWI of Sparse OBN Data, *in* 85th EAGE Annual Conference & Exhibition, 1–5.

Stolt, R. H., 1978, Migration by fourier transform: Geophysics, **43**, No. 1, 23–48.

Tabouy, T., Barbillon, P., and Chiquet, J., 2020, Variational Inference for Stochastic Block Models From Sampled Data: Journal of the American Statistical Association, **115**, No. 529, 455–466.

Tape, W., and Tape, C., 2013, The classical model for moment tensors: Geophysical Journal International, **195**, 1701–1720.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, No. 8, 1259–1266.

Tarantola, A., 2004, Inverse Problem Theory and Methods for Model Parameter Estimation: Society for Industrial and Applied Mathematics, USA.

Toksöz, M., Johnston, D., and of Exploration Geophysicists, S., 1981, Seismic Wave Attenuation, Geophysics reprint series: Society of Exploration Geophysicists.

Vasconcelos, I., and Snieder, R., 2008, Interferometry by deconvolution: Part 2 — theory for elastic waves and application to drill-bit seismic imaging: Geophysics, **73**, No. 3, S129–S141.

Vavrycuk, V., 2005, Focal mechanisms in anisotropic media: Geophysical Journal International, **161**, 334 – 346.

Vesnaver, A., Lovisa, L., and Böhm, G., 2010, Joint 3d processing of active and passive seismic data: Geophysical Prospecting, **58**, No. 5, 831–844.

Vigh, D., Cheng, X., Jiao, K., Kang, W., and Brand, N., 2021, The impact of acquisition geometry on full-waveform inversion updates: The Leading Edge, **40**, No. 5, 335–341.

Virieux, J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, No. 4, 889–901.

Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, No. 6, WCC1–WCC26.

Wagner, D., Koulakov, I., Rabbel, W., Luehr, B.-G., Wittwer, A., Kopp, H., Bohm, M., Asch, G., and Scientists, M., 2007, Joint inversion of active and passive seismic data in Central Java: Geophysical Journal International, **170**, No. 2, 923–932.

Wang, Y., and Lawton, D., 2023, Time-lapse impedance variations using DAS VSP data during CO_2 injection at the CMC Newell County Facility, Alberta, Canada, 1750–1754.

Wang, Z., Wang, J., Sun, W., Huang, J., Li, Z., and Wang, Y., 2024, Waveform inversion with structural regularizing constraint based on gradient decomposition: Journal of Geophysics and Engineering, **21**, No. 4, 1119–1137.

Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., Vinje, V., Štekl, I., Guasch, L., Win, C., Conroy, G., and Bertrand, A., 2013, Anisotropic 3D full-waveform inversion: Geophysics, **78**, No. 2, R59–R80.

Wei, Q., Dobigeon, N., and Tourneret, J.-Y., 2015, Bayesian fusion of multi-band images: IEEE Journal of Selected Topics in Signal Processing, **9**, No. 6, 1117–1127.

Willacy, C., van Dedem, E., Minisini, S., Li, J., Blokland, J.-W., Das, I., and Droujinine, A., 2019, Full-waveform event location and moment tensor inversion for induced seismicity: Geophysics, **84**, No. 2, KS39–KS57.

Wolfe, P., 1969, Convergence conditions for ascent methods: SIAM Review, **11**, No. 2, 226–235.

Yang, Y., and Engquist, B., 2018, Analysis of optimal transport and related misfit functions in full-waveform inversion: Geophysics, **83**, No. 1, A7–A12.

Zhang, X., and Curtis, A., 2020, Variational full-waveform inversion: Geophysical Journal International, **222**, No. 1, 406–411.

Zhang, X., and Curtis, A., 2024, Bayesian variational time-lapse full waveform inversion: Geophysical Journal International, **237**, No. 3, 1624–1638.

Zhang, X., Curtis, A., Galetti, E., and de Ridder, S., 2018, 3-D Monte Carlo surface wave tomography: Geophysical Journal International, **215**, No. 3, 1644–1658.

Zhang, X., Lomas, A., Zhou, M., Zheng, Y., and Curtis, A., 2023, 3-D Bayesian variational full waveform inversion: Geophysical Journal International, **234**, No. 1, 546–561.

Zhang, Y.-G., and Ballmann, J., 1997, Two techniques for the absorption of elastic waves using an artificial transition layer: Wave Motion, **25**, No. 1, 15–33.

Zhang, Z., and Huang, L., 2013, Double-difference elastic-waveform inversion with prior information for time-lapse monitoring: Geophysics, **78**, No. 6, R259–R273.

Zhao, X., and Curtis, A., 2024, Physically structured variational inference for bayesian full waveform inversion: Journal of Geophysical Research: Solid Earth, **129**, No. 11, e2024JB029,557.

Zhao, X., and Curtis, A., 2025, On the design of ultra-sparse seismic surveys for monitoring subsurface CO_2 storage sites using full waveform inversion: International Journal of Greenhouse Gas Control, **146**, 104,433.

Zunino, A., Ghirotto, A., Armadillo, E., and Fichtner, A., 2022, Hamiltonian Monte Carlo Probabilistic Joint Inversion of 2D (2.75D) Gravity and Magnetic Data: Geophysical Research Letters, **49**, No. 20, e2022GL099,789.