

**SOLVING GEOPHYSICAL INVERSE PROBLEMS WITH SCIENTIFIC
MACHINE LEARNING**

A Dissertation
Presented to
The Academic Faculty

By

Ziyi Yin

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Computational Science and Engineering
College of Computing

Georgia Institute of Technology

August 2024

© Ziyi Yin 2024

SOLVING GEOPHYSICAL INVERSE PROBLEMS WITH SCIENTIFIC MACHINE LEARNING

Thesis committee:

Dr. Felix J. Herrmann
School of Computational Science and En-
gineering
Georgia Institute of Technology

Dr. J. Carlos Santamarina
School of Civil and Environmental Engi-
neering
Georgia Institute of Technology

Dr. Nisha Chandramoorthy
School of Computational Science and En-
gineering
Georgia Institute of Technology

Dr. Lars Ruthotto
Department of Mathematics
Emory University

Dr. Peng Chen
School of Computational Science and En-
gineering
Georgia Institute of Technology

Dr. Olav Møyner
Department of Mathematics and Cyber-
netics
SINTEF Digital

Date approved: June 25, 2024

To my family

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Prof. Felix J. Herrmann, for the opportunity to pursue my PhD degree under his supervision in the Seismic Laboratory for Imaging and Modeling (SLIM). This thesis would not exist without his genuine support, insightful advice, and successful leadership. I am especially grateful for his support for my travels to academic conferences, generous offering of coffee espresso, prompt follow-ups on my research progress, and encouragement to pursue my research directions with freedom.

I would also like to thank the committee members at my dissertation defense, including Prof. Nisha Chandramoorthy, Prof. Peng Chen, Prof. Juan Carlos Santamarina, Prof. Lars Ruthotto, and Dr. Olav Møyner, for their time, effort, and valuable feedback on my thesis and future research directions. Additionally, I appreciate Prof. Edmond Chow and Prof. Zhigang Peng for serving as committee members during my proposal.

My gratitude extends to the past and current colleagues at SLIM for their wonderful collaborations and discussions over the years. This includes Dr. Henryk Modzelewski, Dr. Mathias Louboutin, Dr. Philipp A. Witte, Dr. Gabrio Rizzuti, Dr. Mengmeng Yang, Dr. Shashin Sharan, Dr. Ali Siahkoohi, Dr. Yijun Zhang, Rafael Orozco, Abhinav Prakash Gahlot, Huseyin Tuna Erdinc, Thomas J. Grady II, Rosen Yu, and Thales Souza. Specifically, I would like to thank Dr. Mathias Louboutin for his many wonderful suggestions on my research projects and generous guidance on software design and development. I am particularly grateful to Rafael Orozco for the insightful discussions that contributed significantly to several chapters of my thesis. I would also like to thank Dr. Mathias Louboutin and Will Powell for maintaining the cruyff cluster, which greatly accelerated my research progress.

I am grateful for the collaboration with Dr. Olav Møyner from SINTEF on reservoir simulations, geological carbon storage, and numerical solvers, which significantly bene-

fited several chapters of my thesis and other research projects. I also appreciate the collaboration with Dr. Rishi Khan from Extreme Scale Solutions on Fourier neural operators and their deployment in multiphase flow modeling. Thanks to Charles Jones from Osokey for his suggestions on several of my publications that appear in this thesis. I am also thankful to the sponsors of the ML4SEISMIC center for their valuable feedback on my research during annual partners' meetings.

I would like to thank my colleagues at Chevron during my internship in the summer of 2023, including Dr. John Washbourne, Dr. Luke Decker, Dr. Mark Fan, Dr. Kurt Nihei, Dr. Sam Kamplan, Lei Wei, Dr. Anusha Sekar, Dr. Zhishuai Zhang, Dr. Hao Guo, Dr. Kristoffer Walker, Dr. Hermes Malcotti, and Dr. Tamas Nemeth, for kindly sharing their insights on industrial applications of geophysical inversion.

I feel very fortunate to be a student in the School of Computational Science and Engineering (CSE). I genuinely thank the faculty and staff members at CSE, especially Nirvana Edwards, Iris Hamilton, and Bryant Wine, as well as the CSE Graduate Student Association, for their efforts in providing such an enjoyable and delightful environment for me and other students. I would also like to thank my friends at Georgia Tech for accompanying me on this journey, including but not limited to Dr. Alex Rodriguez, Ben Wilfong, Conlain Kelly, Jinwoo Go, Hansol Suh, Max Hawkins, Kaan Sancak, Sarah Bi, Peimeng Guan, Ben Cobb, Ben Hoover, Aranya Banerjee, Dmitry Shribak, Sebastián Gutiérrez Hernández, Irene Simó Muñoz, Kalsuda Lapborisuth, Haoyun Li, Richard Rex, Grant Bruer, and Ziqi Zhang.

I owe much gratitude to my wonderful family for their unwavering love, support, and belief in me. I am grateful for everything I learned from my parents, Li Yin and Xin Cheng, my grandmother, Dongxia Kou, and my younger sister, Zirun Yin. My heartfelt appreciation goes to my girlfriend, Shui Yu, for her love, understanding, and support throughout my academic journey. I am also fortunate to have our dogs, Oscar and Ollie, accompanying me on this journey.

My research studies were funded by the Georgia Research Alliance, partners of the ML4SEISMIC Center, the US National Science Foundation grant OAC 2203821, and the Department of Energy grant No. DE-SC0021515. I am grateful for their support throughout my studies.

During the preparation of this dissertation, I gratefully used ChatGPT to refine sentence structures and improve readability. After using this service, I reviewed and edited the content as needed and take full responsibility for the content of this dissertation.

TABLE OF CONTENTS

Acknowledgments	iv
List of Tables	xiv
List of Figures	xv
Nomenclature	xxii
Summary	xxii
Chapter 1: Introduction	1
1.1 Objective	2
1.2 Full-waveform inversion	3
1.2.1 The inverse problem	4
1.2.2 Computational and numerical challenges	4
1.3 Geological carbon storage and time-lapse monitoring	6
1.3.1 Multiphase flow in porous media	7
1.3.2 Rock physics	8
1.3.3 Wave physics	9
1.3.4 The inverse problem	11
1.3.5 Computational and numerical challenges	11

1.4	Contributions	12
1.5	References	15
Chapter 2: Learned multiphysics inversion with differentiable programming and machine learning		20
2.1	Summary	20
2.2	Motivation	20
2.3	Design principles	22
2.4	Learned time-lapse end-to-end permeability inversion	24
2.4.1	Wave-equation-based inversion	25
2.4.2	Deep priors and normalizing flows	28
2.4.3	Fluid-flow simulation and permeability inversion	31
2.4.4	Fourier neural operator surrogates	34
2.5	Putting it all together	35
2.6	Preliminary inversion results	37
2.7	Remaining challenges	41
2.7.1	Scale-up normalizing flows	41
2.7.2	Scale-up neural operators	42
2.8	Conclusions	43
2.9	References	44
Chapter 3: Time-lapse full-waveform permeability inversion: a feasibility study .		51
3.1	Summary	51
3.2	Introduction	51

3.3	Permeability inversion framework	54
3.4	Feasibility study on the Compass model	56
3.4.1	Sensitivity with respect to starting models	57
3.4.2	Sensitivity with respect to forward modeling errors	59
3.4.3	CO ₂ plume estimation and forecast	61
3.4.4	Multiparameter inversion	61
3.5	Limitations	65
3.5.1	Reservoir simulation	65
3.5.2	Rock physics	66
3.5.3	Wave physics	66
3.6	Discussion and conclusion	66
3.7	Data availability	67
3.8	References	68

Chapter 4: Solving multiphysics-based inverse problems with learned surrogates and constraints 73

4.1	Summary	73
4.2	Introduction	74
4.3	Fourier neural operators	76
4.4	Inversion with learned surrogates	77
4.5	Learned constraints with normalizing flows	78
4.5.1	Training normalizing flows	80
4.5.2	Trained normalizing flows as constraints	80
4.5.3	Inversion with progressively relaxed learned constraints	85

4.6	Numerical Experiments	87
4.6.1	Unconstrained/constrained permeability inversion from CO ₂ saturation data	90
4.6.2	Unconstrained/constrained permeability inversion from well observations	91
4.6.3	Multiphysics end-to-end inversion	93
4.6.4	Jointly inverting time-lapse seismic data and well measurements	98
4.6.5	CO ₂ plume estimation and forecast	100
4.6.6	Analysis of computational gains	102
4.7	Discussion and conclusions	104
4.8	Availability of data and materials	107
4.9	References	108

Chapter 5: Derisking geologic carbon storage from high-resolution time-lapse seismic to explainable leakage detection 117

5.1	Summary	117
5.2	Introduction	118
5.3	Seismic monitoring with time-lapse imaging	119
5.3.1	Monitoring with the joint recovery model	120
5.3.2	Monitoring with curvelet-domain structure promotion	122
5.4	Numerical case study: Blunt sandstone in the Southern North Sea	123
5.4.1	Proxy seismic and fluid-flow models	123
5.4.2	Fluid-flow simulations	126
5.4.3	Rock-physics conversion	126
5.4.4	Time-lapse seismic simulations	127

5.4.5	Imaging with joint recovery model versus reverse-time migration . .	128
5.5	Deep neural network classifier for CO ₂ leakage detection	128
5.6	Class activation mapping based saliency map	131
5.7	Discussion and conclusion	133
5.8	References	135

Chapter 6: WISE: full-Waveform variational Inference via Subsurface Extensions 139

6.1	Summary	139
6.2	Introduction	139
6.3	Methodology	140
6.3.1	Full-waveform inversion	140
6.3.2	Full-waveform inference	140
6.3.3	Amortized variational inference with conditional normalizing flows	141
6.3.4	Physics-informed summary statistics	141
6.3.5	Common-image gathers as summary statistics	142
6.4	Synthetic case studies	143
6.4.1	Dataset generation and network training	143
6.4.2	Results	145
6.4.3	Quality control	145
6.4.4	Uncertainty quantification and downstream imaging	145
6.5	Discussion	146
6.6	Conclusions	149
6.7	References	151

Chapter 7: WISER: multimodal variational inference for full-waveform inversion without dimensionality reduction	155
7.1 Summary	155
7.2 Introduction	155
7.3 Amortized VI with WISE (lines 1—20)	157
7.3.1 Dataset generation (lines 3—9)	157
7.3.2 Network training (lines 11—12)	159
7.3.3 Online inference (lines 14—20)	159
7.4 Physics-based refinements (lines 22—32)	159
7.5 Case studies	161
7.5.1 Case 1: in distribution	162
7.5.2 Case 2: out of distribution	164
7.6 Discussion and conclusions	167
7.7 References	169
Chapter 8: Conclusion	173
8.1 Scalable, interoperable, and differentiable programming framework	173
8.2 Surrogate-assisted inversion with learned modeling operators	175
8.3 Semi-amortized variational inference with physics-informed summary statistics	176
8.4 Current limitations and future directions	178
8.4.1 Chapter 3	178
8.4.2 Chapter 4	178
8.4.3 Chapter 5	180

8.4.4	Chapter 6	181
8.4.5	Chapter 7	183
8.5	References	185
	Appendices	187
	Appendix A: More details of FWI: modeling, gradient, and extended gradient . .	188
	Appendix B: More results for Chapter 6	191
	Appendix C: Permissions to use copyrighted material	197

LIST OF TABLES

4.1	S/N (in dB) and SSIM values of permeability recovery.	100
-----	---	-----

LIST OF FIGURES

1.1	A schematic representation of a seismic survey, adapted from [16].	3
1.2	(a) Example of a 2D velocity model. (b) Synthetic seismic data generated by the acoustic wave modeling operator at three different source locations. The simulation is carried out using JUDI.jl [17].	5
1.3	Multiphase fluid-flow modeling. (a) Spatial distribution of subsurface permeability. (b) CO ₂ saturation snapshots at year 5, 15, 25, 45, 65. The simulation is carried out using JutulDarcyRules.jl [43], which utilizes the high-performant simulators implemented in JutulDarcy.jl [44].	8
1.4	(a) Brine-filled (pre-injection) wavespeed model. (b) Decrease in wavespeed induced by the CO ₂ saturation at 25th year shown as third plot in Figure 1.3b.	10
2.1	The multiphysics forward model. The permeability, \mathbf{K} , is generated from Gaussian noise with a pretrained NF, \mathcal{G} , followed by two-phase flow simulations through \mathcal{S} , rock physics denoted by \mathcal{R} , and time-lapse seismic data simulations via wave physics, \mathcal{F}	24
2.2	Experimental setup. The black \times symbol in the middle of the model indicates the CO ₂ injection location. The seismic sources are on the left-hand side of the model (shown as yellow \times symbols) and receivers are on the right-hand side of the model (shown as red dots). In grey color, we overlay the compressional wavespeed with simulated CO ₂ saturation modeled for 18 years.	25
2.3	Demonstration of Gaussinazation of Compass slices during training of normalizing flow. The data used for this didactic example is openly available and this figure fully in the InvertibleNetworks.jl repository.	29
2.4	Examples of Compass 2D slices used to train a normalizing flow prior. . . .	29

2.5	Generative samples from our trained prior. Their similarity to the training samples in Figure 2.4 suggests that our normalizing flow has learned a useful prior.	30
2.6	Results from using our normalizing flow learned prior in FWI. (a) Ground truth. (b) Traditional FWI without prior resulting in 12.98 dB PSNR. (c) Our FWI result with learned prior resulting in 14.77 dB PSNR.	32
2.7	Fifteen-year time-lapse seismic end-to-end permeability inversion with physics-based and surrogate fluid-flow simulations. (a) Ground truth permeability. (b) Initial permeability with homogeneous values in the reservoir. (c) Inverted permeability from physics-based inversion. (d) Inverted permeability with neural surrogate approximation.	39
2.8	CO ₂ plume estimation and prediction. The first two columns are the CO ₂ concentration snapshots at year 10 and year 15 of the first 15 years of simulation monitored seismically. The last three columns are forecasted snapshots at years 16, 17, 18, where no seismic data is available. First row corresponds to the ground truth CO ₂ plume simulated by the unseen ground truth permeability model. Second row contains plume simulations in the starting model, with a 10.99 dB SNR on the first 15 years of CO ₂ snapshots and a 8.51 dB on the last 3 years. Rows three and four contain estimated and predicted CO ₂ plumes for the physics-based and surrogate-based permeability inversion. The SNR values of the first 15 years of the estimated CO ₂ plume are 17.72 dB and 16.17 dB for the physics-based inversion and the surrogate-based inversion, respectively. The SNR values for the CO ₂ plume forecasts for the last 3 years are 15.69 dB and 14.05 dB for the physics-based inversion and the surrogate-based inversion.	40
3.1	Experimental configuration. (a) Setup of seismic acquisition and well control. Dark blue * denotes the CO ₂ injection well. While × and black · represent source and receiver locations, respectively. The gray curve delineates the shape of the CO ₂ plume at the 25th year. (b) Unseen ground truth spatial distribution of horizontal permeability. (c) Histogram of the common logarithm of the permeability model with Kernel Density Estimation (KDE).	53

3.2	<p>Multiphysics forward model. The reservoir simulator, \mathcal{S}, produces n_k snapshots of time-varying CO₂ saturation, compiled in \mathbf{c}, from the permeability model, \mathbf{K}. The rock physics model, \mathcal{R}, based on the porosity and the brine-filled velocity model, converts each CO₂ saturation snapshot, \mathbf{c}_k, to the altered velocity model, \mathbf{v}_k. The shaded area highlights the CO₂-induced changes in velocity. Finally, the wave modeling, \mathcal{F}, generates a time-lapse seismic dataset, \mathbf{d}_k, for each velocity model, \mathbf{v}_k. These datasets are collected in the vector, \mathbf{d}.</p>	55
3.3	<p>Permeability inversion results for cases 1 and 2. (a)(c)(e) display the initial permeability model in case 1, the logarithmic ratio of the ground truth permeability (Figure 3.1 (b)) to the initial one, and the logarithmic ratio of the inverted permeability to the initial one. (b)(d)(f) display the same but for case 2 with a distorted initial permeability model. Gray curve indicates the boundary of the CO₂ plume at the 25th year. “× initial” on the caption of the colorbar represents the factor by which the initial permeability is updated. 60</p>	60
3.4	<p>Permeability inversion results for case 3. (a) Inverted brine-filled baseline velocity used in permeability inversion. (b) The logarithmic ratio of the inverted permeability to the initial one. Gray curve indicates the boundary of the CO₂ plume at the 25th year.</p>	60
3.5	<p>Predicted CO₂ saturation for 5th, 15th, and 25th years, shown in first, second, and third columns, respectively. The first row shows the (unseen) ground truth CO₂ saturation. The second and fourth rows show the saturation predicted with initial permeability models in Figure 3.3 (a) and Figure 3.3 (b), respectively. The third, fifth, and sixth rows show the updated saturation after updating the initial permeability models by Figure 3.3 (e), Figure 3.3 (f), and Figure 3.4 (b), respectively. The boundaries of the (unseen) ground truth CO₂ saturation are shown in white curves.</p>	62
3.6	<p>CO₂ plume forecasts for 45th and 65th years, shown in first and second columns, respectively. The ordering of the rows remains the same as Figure 3.5. Purple regions display the CO₂ plume permanently stored via the residual trapping mechanism.</p>	63
3.7	<p>Porosity inversion in case 4. (a) Initial porosity. (b) Inverted porosity. (c) Unseen ground truth porosity. The gray curve delineates the shape of the CO₂ plume at the 25th year.</p>	65
4.1	<p>Permeability models. First row shows the realistic permeability samples for FNO and NF training. Second row shows the generative samples from the trained NF.</p>	81

4.2	Sample permeability models in the physical and latent space. (a) An in-distribution permeability model. (b) An out-of-distribution permeability model. (c) An in-distribution permeability model in the latent space. (d) An out-of-distribution permeability model in the latent space.	83
4.3	Projections onto increasing ℓ_2 -norm balls for the in- and out-of-distribution examples of Figure 4.2. Top row: projections of in-distribution sample. Bottom row: projections of out-of-distribution sample. Each column corresponds to setting $\alpha = 0, 0.1, 0.2, 0.4, 1$ in Equation 4.6.	84
4.4	Latent space projection experiments.(a) Relative ℓ_2 reconstruction error and FNO prediction error for in-distribution sample. (b) The same but for out-of-distribution sample. The blue curve shows the relative ℓ_2 misfit between the permeability models before and after latent space shrinkage. The orange curve shows the FNO prediction error on the permeability model after shrinking the ℓ_2 -norm ball. The red dashed line denotes the amplitude of standard Gaussian noise.	84
4.5	FNO prediction errors for the latent space shrinkage experiment in Equation 4.6 for 50 random realizations of standard Gaussian noise.	84
4.6	Permeability models. (a) unknown “ground-truth” permeability model from unseen test set, where the symbols \blacktriangleright and \blacktriangleleft denote the CO ₂ injection and brine production location, respectively; (b) initial permeability model, K_0	88
4.7	Five CO ₂ saturation snapshots after 400, 500, 600, 700, and 800 days. First row shows the CO ₂ saturation simulated by the PDE. Second row shows the CO ₂ saturation predicted by the trained FNO. Third row shows the $5\times$ difference between the first row and the second row.	89
4.8	Permeability inversion from fully observed time-lapse CO ₂ saturations. (a) Inversion result with PDE solvers. (b) The same but via the approximate FNO surrogate. (c) Same as (a) but with NF constraint. (d) Same as (b) but with NF constraint. (e) The FNO approximation errors as a function of the number of iterations for the result plotted in (b) and (d).	92
4.9	Permeability inversions from CO ₂ saturations sampled at three well locations at 6 early snapshots. The well locations are denoted by the red vertical lines. (a) Unconstrained inversion result based on PDE solves. (b) Same as (a) but now with FNO surrogate approximation. (c) Constrained inversion result based on PDE solves. (d) Same as (c) but now with FNO surrogate approximation. (e)-(h) The error of the permeability inversion results in (a)-(d) compared to the unseen ground truth shown in Figure 4.6a. (i) The FNO prediction errors as a function of the number of iterations for (b) and (d).	94

4.10	Seismic acquisition. The white \times represents the acoustic sources, and the red lines represent the dense receivers. The CO ₂ saturation snapshot at day 600 is plotted in the middle, which is the last snapshot that is monitored seismically.	96
4.11	Permeability inversions from time-lapse seismic data. (a) Inversion result using PDE solvers. (b) The same as (a) but for the FNO surrogate. (c) The same as (a) but with the NF-based constraint. (d) The same as (a) but now for the FNO surrogate with the NF-based constraint. (e)-(h) The error of the permeability inversion results in (a)-(d) compared to the unseen ground truth shown in Figure 4.6a. (i) The FNO prediction errors as a function of the number of iterations for (b) and (d).	97
4.12	Joint permeability inversions from both time-lapse seismic data and time-lapse well measurements. (a) Inversion result using PDE solvers. (b) The same as (a) but for the FNO surrogate. (c) The same as (a) but with the NF-based constraint. (d) The same as (a) but now for the FNO surrogate with the NF-based constraint. (e)-(h) The error of the permeability inversions in (a)-(d), compared to the unseen ground truth shown in Figure 4.6a. (i) The FNO prediction errors as a function of the number of iterations for (b) and (d).	99
4.13	CO ₂ plume estimation and forecast using FNO surrogates and NF constraints to invert different modalities of observed data. The first three columns represent past CO ₂ saturations at day 400, 500, and 600 of the first 600 days of CO ₂ saturation monitored either through the well measurements or time-lapse data. The last two columns include forecasts for the saturations at future days 700 and 800, where no observed data is available. The first row shows the past and future CO ₂ estimates yielded by inverting well measurements only. The second row is the same but now inverting time-lapse seismic data. The third row is the same but now jointly inverting well measurements and time-lapse seismic data. The fourth, fifth, and sixth rows show $5\times$ difference between the ground truth CO ₂ plume (first row of Figure 4.7) and the first, second, third row, respectively. The S/Ns for the first, the second, and the third rows are 15.26 dB, 20.14 dB, 20.46 dB, respectively.	103

5.1	Simulation-based monitoring design framework. Starting with a proxy model for the wavespeed and density (a), the workflow proceeds by converting these seismic properties into permeability and porosity (b). These fluid flow properties are used to simulated CO ₂ plumes that behave regularly or exhibit leakage outside the storage complex (c). Induced changes by the CO ₂ plume for the wavespeed and density are depicted in (d) and serve as input to simulations of time-lapse seismic data (SNR 8.0 dB) and shot-domain time-lapse differences (SNR -31.4 dB). Imaging results for regular and irregular plume developments are plotted in (f) and serve as input to the deep neural classifier (g), which determines whether the flow behaves regularly or leaks. Activation mappings in (h) show regions on which the network is basing its classification. As expected, the activation mapping is diffusive in case of regular CO ₂ plume development and focused on the leakage location when CO ₂ plume behaves irregularly.	124
5.2	Permeability and porosity derived from a 2D slice of Compass model. . . .	126
5.3	Reverse-time migration (RTM) versus inversion joint recovery model (JRM). (a) RTM image of the baseline; (b) JRM image of the baseline; (c) time-lapse difference and CO ₂ plume for independent RTM images with leakage; (d) time-lapse obtained by inverting the time-lapse data jointly with leakage; (e) time-lapse difference and CO ₂ plume for independent RTM images without leakage; (f) time-lapse obtained by inverting the time-lapse data jointly without leakage. Notice improvement in the time-lapse image quality. This improvement is reflected in the NRMS values that decrease from 8.48 % for RTM to 3.20 % for JRM.	129
5.4	NRMS values for 1000 time-lapse experiments.	130
5.5	Confusion matrix for classifier trained on recovery images from JRM. . . .	132
5.6	CAM for time-lapse difference images with a leaking plume and with a regular plume.	133
6.1	(a) an unseen ground-truth velocity model; (b) 1D initial FWI-velocity model; (c) conditional mean estimate for RTM as summary statistics (SSIM = 0.48); (d) conditional mean estimate from WISE (SSIM = 0.56); (e) CIGs calculated by the initial FWI-velocity model given by (b); (f) CIGs calculated by (d).	144

6.2	Variability in velocity models and imaged reflectors. (a) Posterior velocity samples from WISE visualized similar to CIGs by plotting the conditional mean (Figure 6.1(d)) in the central image. Above it shows the posterior sample traces at $Z = 2.4$ km. On the right shows the traces at $X = 3.4$ km. (b) Point-wise standard deviation of the posterior velocity samples. (c) Samples of imaged reflectors, where the central image displays imaged reflectors using the conditional mean estimate. The layout of the traces remains the same as (a). (d) Point-wise standard deviation of the imaged reflectors. (e) Point-wise maximum depth shift.	147
7.1	Comparison between WISE and WISER for an in-distribution case. (a) Unseen ground-truth velocity model. (b) Estimated velocity models from WISE. The conditional mean estimate (CM) is shown in the center. For posterior samples, horizontal traces at $Z = 2.7$ km and vertical traces at $X = 3.6$ km are displayed on the top and on the right, respectively. (d) Imaged reflectivity samples from WISE. (f) Zoom-in views of (d) overlaying on the CM of WISE. (c)(e)(g) are the counterparts from WISER, showcasing significant improvements.	163
7.2	OOD case study. (a) Curves for velocity-value perturbations; (b) histograms of values at the depth of 0.5 km and 2.8 km in the original velocity model (Figure 7.1a), perturbed velocity model (Figure 7.2c), posterior samples of WISE, and WISER, shown in red, blue, yellow and green color, respectively. (c)—(i) Comparison between WISE and WISER. The ordering remains the same as in Figure 7.1. (j) FWI result starting with a posterior sample from WISE. (k) A posterior sample from WISER.	166
B.1	(a) FWI result using Figure 1(b) in the main text as a starting model; (b) ground truth velocity model.	192
B.2	Vertical profiles.	193
B.3	CIGs for three posterior samples from WISE.	194
B.4	Applying WISE for two unseen test samples in Open FWI CurveFault-A dataset. \bar{y} denotes the type of summary statistics.	196

NOMENCLATURE

General usage of acronyms

AD	Automatic Differentiation
CCS	Carbon Capture and Storage
CIG	Common-Image Gather
CNF	Conditional Normalizing Flow
CNN	Convolutional Neural Network
FD	Finite Differences
FNO	Fourier Neural Operator
FWI	Full Waveform Inversion
GCS	Geological Carbon Storage
JRM	Joint Recovery Model
NF	Normalizing Flow
PDE	Partial Differential Equation
RTM	Reverse-Time Migration
SGD	Stochastic Gradient Descent
S/N	Signal-to-Noise Ratio
UQ	Uncertainty Quantification
VI	Variational Inference
WISE	full-Waveform variational Inference via Subsurface Extensions
WISER	full-Waveform variational Inference via Subsurface Extensions with Refinements

SUMMARY

Solving inverse problems involves estimating unknown parameters of interest from indirect measurements. Specifically, geophysical inverse problems seek to determine various Earth properties critical for geophysical exploration, carbon control, monitoring, and earthquake detection. These problems pose unique challenges: the parameters of interest are often high-dimensional, and the mapping from parameters to observables is computationally demanding. Moreover, these problems are typically non-convex and ill-posed, meaning that multiple sets of model parameters can adequately fit the observations, and inversion algorithms depend on accurate initial model parameters.

This thesis introduces several innovative methods to tackle these challenges using scientific machine learning techniques. It discusses algorithms and software frameworks that utilize surrogate and generative models to achieve scalable and reliable inversion. It also examines the integration of conditional generative models with physics for Bayesian variational inference and uncertainty quantification. These methods have been applied to two critical inverse problems in geophysical applications: monitoring geological carbon storage and full-waveform inversion, both of which are plagued by the aforementioned computational challenges.

The thesis consists of six papers. The first two papers present a scalable, interoperable, and differentiable programming framework for learned multiphysics inversion, showcased through realistic synthetic case studies in geological carbon storage monitoring. The third paper introduces a computationally efficient and reliable algorithm that employs surrogate models, particularly Fourier neural operators, to accelerate inversion. The reliability of this algorithm is ensured by using normalizing flows as learned constraints to safeguard the accuracy of the surrogate models throughout the inversion process. The subsequent paper explores a joint inversion approach and an explainable deep neural classifier for time-lapse seismic imaging and carbon dioxide leakage detection during geological carbon

storage. The final two papers introduce amortized and semi-amortized variational inference approaches that employ information-preserving physics-informed summary statistics and refinements to provide computationally feasible and reliable uncertainty quantification in high-dimensional full-waveform inversion problems. They also assess the impact of the inherent uncertainty in these ill-posed inversion problems on subsequent imaging tasks.

CHAPTER 1

INTRODUCTION

This thesis aims to address the challenges associated with solving large-scale inverse problems with computationally expensive forward modeling operators, particularly in geophysical applications. Solving an inverse problem involves estimating unknown parameters of interest from a set of indirect measurements, which may be corrupted by noise [1]. Geophysical inversion, specifically, seeks to estimate certain properties of the Earth's subsurface from measurements related to these properties [2]. The parameter-to-observable maps, also known as forward modeling operators, often represent physical phenomena. Geophysical inverse problems are crucial in earth science applications as they infer properties of the Earth that cannot be directly observed. Estimating these Earth properties is essential for geophysical exploration [3], geologic carbon storage monitoring [4], and earthquake detection [5], to name a few.

Solving geophysical inverse problems presents multiple challenges, with the high computational cost being the most significant. The Earth properties of interest are often parameters of partial differential equations (PDEs), discretized using finite difference [6], finite element [7], or finite volume [8] schemes. Achieving high-resolution and high-fidelity imaging of the subsurface requires fine discretization of the unknown Earth properties, leading to computationally expensive PDE solves during the evaluation of the forward modeling operators and the calculation of sensitivity with respect to model parameters. This problem is exacerbated in inverse problems, which typically rely on optimization algorithms requiring forward modeling and sensitivity calculations at each iteration. Additionally, PDE solvers in some domains do not naturally offer the capability to calculate sensitivities with respect to model parameters. In such cases, inverting for model parameters using derivative-free optimization methods can be even more computationally demanding [9]. Beyond the com-

computational cost of the modeling operators, the high dimensionality of the parameter space also poses significant challenges for Bayesian inference algorithms. Sampling-based methods, such as Markov-chain Monte Carlo (MCMC, [10]), often suffer from the curse of dimensionality and struggle with high-dimensional parameter spaces in geophysical inverse problems.

In addition to computational concerns, many geophysical inverse problems are non-convex and ill-posed. Specifically, their forward modeling operators have a non-trivial null space, leading to multiple model parameters that fit the observed data. Furthermore, the non-convexity results in a high reliance on initial model parameters in geophysical inversion algorithms. These challenges necessitate effective regularization techniques to encourage desired Earth properties [11, 12, 13] or uncertainty quantification (UQ, [14]) schemes to identify a range of model parameters that are both physical and compatible with the observed data.

1.1 Objective

The computational and numerical challenges of solving large-scale geophysical inverse problems necessitate the development of scientific machine learning (SciML) techniques. The aim of this thesis is to develop several SciML methods for solving two major geophysical inverse problems:

- full-waveform inversion (FWI);
- and geological carbon storage (GCS) monitoring.

These SciML methods are designed to enhance the quality and interpretability of the solutions, reduce the intensive computational cost, and accelerate the UQ workflows. The rest of this chapter is organized as follows. First, I introduce the GCS monitoring and FWI problems in detail. I discuss the forward problems associated with them, which represent the underlying physical phenomena in the Earth's subsurface. Then, I illustrate the

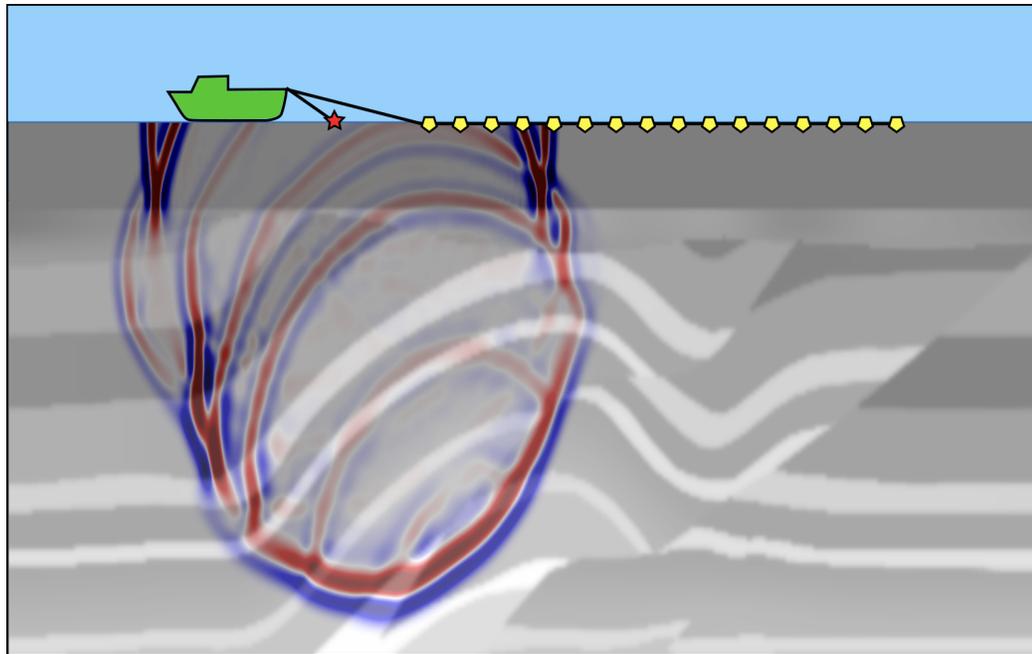


Figure 1.1: A schematic representation of a seismic survey, adapted from [16].

challenges of solving these two inverse problems. To meet these challenges, I outline my contributions in this thesis to conclude this chapter.

1.2 Full-waveform inversion

Seismic imaging and inversion play a pivotal role in geophysical exploration, primarily focusing on estimating the wave properties in the Earth's subsurface through observed seismic data collected at sensors located at the surface [15]. Figure 1.1 shows a schematic representation of a seismic survey. The seismic vessel, denoted in green, excites acoustic waves by firing seismic sources (the red star). The wave propagates according to the wave properties in the Earth's subsurface. The reflected and scattered waves are recorded by the receivers (yellow pentagons) as a function of time. A seismic survey involves many seismic sources firing at different locations.

1.2.1 The inverse problem

FWI aims to solve for the spatial distribution of the acoustic velocity in the subsurface from measured data collected at the receiver locations. This is achieved by minimizing the misfit between the observed seismic data of a seismic survey and its numerically modeled counterpart. The relationship between the model parameters and the observed data can be represented by:

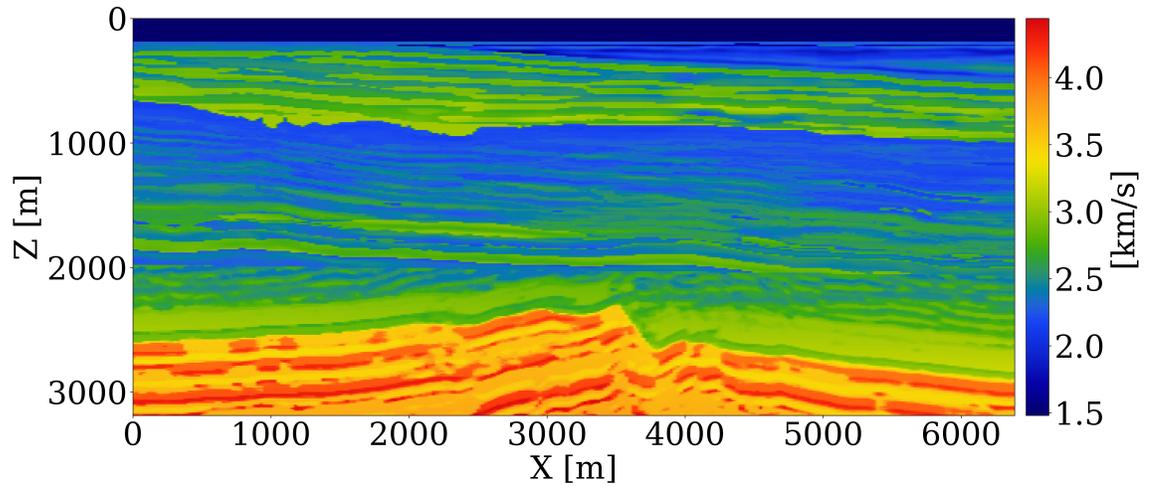
$$\mathbf{y} = \mathcal{F}(\mathbf{x}) + \boldsymbol{\epsilon}, \quad (1.1)$$

where the velocity model, \mathbf{x} , and the observed data, \mathbf{y} (corrupted by measurement noise, $\boldsymbol{\epsilon}$), are linked via the nonlinear modeling operator, \mathcal{F} . The operator, \mathcal{F} , encodes the source signature and receiver positions. More details about the wave physics can be found in Chapter A. An example of the model parameter, \mathbf{x} , is shown in Figure 1.2a. The corresponding noise-free seismic data generated by the operator, \mathcal{F} , is shown in Figure 1.2b, for three distinct source locations. Each column represents a recorded time trace for the corresponding receiver location.

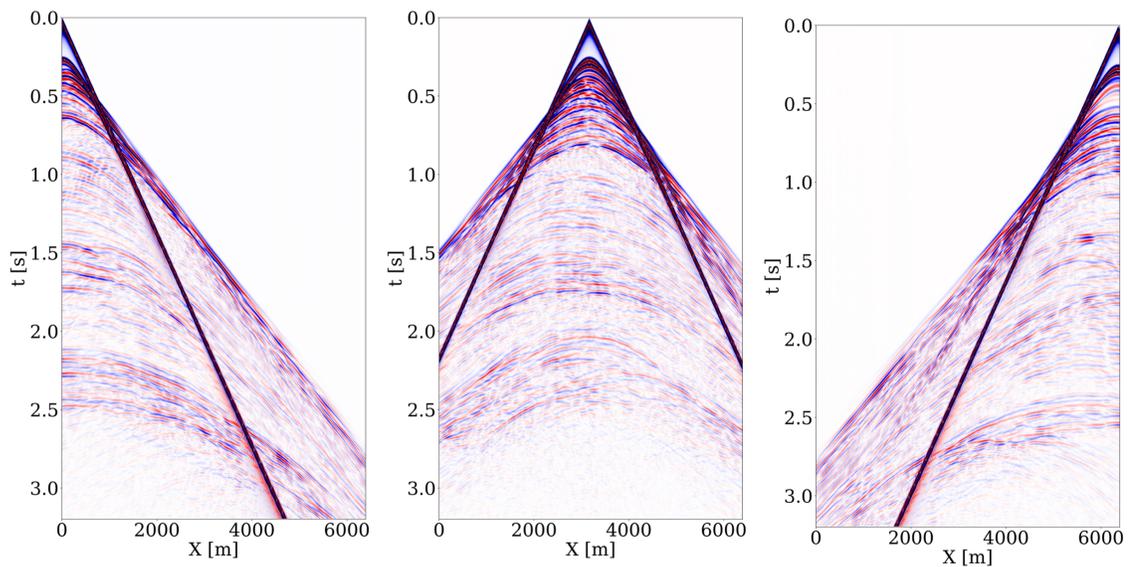
1.2.2 Computational and numerical challenges

The forward modeling operator in FWI is computationally expensive because every evaluation of the forward modeling operator, \mathcal{F} , requires solving the wave equation for all sources. For high-resolution imaging of the Earth's subsurface, the unknown parameter of interest, \mathbf{x} , needs to be finely discretized, leading to a high degree of freedom for this inverse problem. Consequently, the evaluation of the forward modeling operator, which involves solving the wave equation multiple times, becomes time-consuming. Additionally, the high dimensionality of the parameter spaces poses challenges for Bayesian statistical inference [18].

Apart from the generic challenges associated with high-dimensional Bayesian infer-



(a)



(b)

Figure 1.2: (a) Example of a 2D velocity model. (b) Synthetic seismic data generated by the acoustic wave modeling operator at three different source locations. The simulation is carried out using JUDI.jl [17].

ence, FWI is further complicated by its sensitivity to the initial models. When the initial velocity model is poor, FWI is likely to fall into local minima. This is known as the “cycle skipping” problem [19]. Moreover, a nontrivial null space of the forward modeling operator, \mathcal{F} , may result in non-unique solutions where different velocity models fit the observed data adequately well [20].

1.3 Geological carbon storage and time-lapse monitoring

The aforementioned FWI problem aims to estimate the subsurface properties through a single seismic survey. Time-lapse seismic monitoring extends this goal to estimate subsurface properties over time, delineating temporal changes in the Earth’s subsurface. In this thesis, I primarily focus on using time-lapse seismic monitoring for an application to combat climate change.

Carbon capture and storage (CCS) is among the very few, if not the only, scalable technologies capable of rapidly reducing emissions in industrial sectors that have limited options for decarbonization [21, 22]. The concept of CCS was first introduced in 1977 [23]. It involves capturing CO_2 emissions from industrial processes, such as steel and cement production, or burning fossil fuels, and then injecting the captured CO_2 into suitable geological formations (including depleted oil reservoirs [24] and saline aquifers [25]) in the Earth’s subsurface. This process, known as GCS [26, 27], has been pioneered by projects like the Sleipner CO_2 injection project in Norway [4].

While GCS technology has the potential to scale, its success depends on our ability to mitigate risks, particularly CO_2 leakage through faults, fractures, and abandoned wells [28]. An essential aspect of risk mitigation involves ensuring that the injected CO_2 remains within the storage complex. To minimize potential risks associated with GCS, it is crucial for practitioners to monitor the CO_2 dynamics over time. In GCS projects, various types of time-lapse data can be collected to monitor CO_2 plumes. These data modalities include measurements in wells [29, 30], and the collection of gravity [31, 32], electromagnetic

[33, 34], and seismic data [35, 36]. Among these modalities, time-lapse seismic data is the most commonly used and stands out for its ability to attain high-resolution and high-fidelity images of the Earth’s subsurface. Additionally, CO₂ is typically in a supercritical state at the injection depth, providing a sonic velocity contrast to the initially brine-filled reservoir, making time-lapse seismic data ideal for containment monitoring [37]. Next, I introduce several types of physics involved in GCS and time-lapse monitoring and define the inverse problem associated with GCS monitoring.

1.3.1 Multiphase flow in porous media

During GCS projects, supercritical CO₂ injected into the Earth’s subsurface replaces brine in the porous rocks [28]. This process, called multiphase flow in porous media, involves the two phases of CO₂ and brine. This is a slow-time dynamic flow process that occurs over a long time range (years), which can be mathematically represented as follows:

$$\mathbf{c} = \mathcal{S}(\mathbf{K}) \quad \text{where} \quad \mathbf{c} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{n_k}]. \quad (1.2)$$

In this expression, the nonlinear operator \mathcal{S} represents the multiphase flow modeling operator, which takes the gridded spatially varying permeability in the reservoir, \mathbf{K} , as input and produces n_k time-varying CO₂ saturation snapshots, encapsulated in \mathbf{c} . The forward modeling of \mathcal{S} involves solving time-dependent partial differential equations (PDEs) using implicit timestepping scheme [38, 39, 40], where permeability \mathbf{K} is the PDE parameter. The governing equations for the multiphase flow involve Darcy’s equation and the mass conservation law in the leading order. Detailed information on the governing equations, initial and boundary conditions, and numerical solution schemes can be found in [39, 41] and the references therein. Figure 1.3 demonstrates a numerical simulation of CO₂ injection into a saline aquifer. Figure 1.3a shows the spatial distribution of the horizontal permeability model, \mathbf{K} , in millidarcies (md). An injection well is set at around 1.7 km depth and 2.3 km lateral position, which keeps injecting supercritical CO₂ in a constant rate

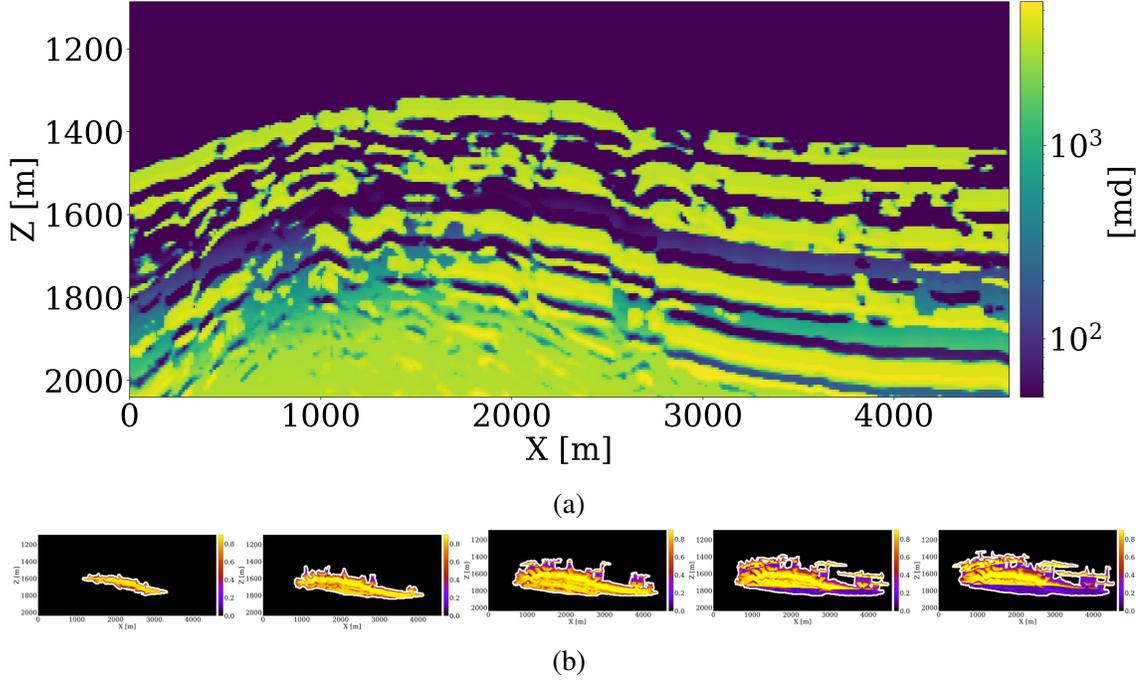


Figure 1.3: Multiphase fluid-flow modeling. (a) Spatial distribution of subsurface permeability. (b) CO₂ saturation snapshots at year 5, 15, 25, 45, 65. The simulation is carried out using JutulDarcyRules.jl [43], which utilizes the high-performant simulators implemented in JutulDarcy.jl [44].

in the first 25 years and shuts down after the 25th year. Figure 1.3b shows the CO₂ saturation snapshots, c at year 5, 15, 25, 45, 65, where the white curves enclose the boundaries of the CO₂ plumes. During the first 25 years, CO₂ plume grows up due to the injection well and follows the high permeability channels in the model. After the 25th year, the mass of the CO₂ remains constant in the storage complex. Due to the buoyancy effect, the CO₂ plume starts to migrate upwards, while a portion (approximately 10%) remains trapped in the pore spaces, indicated in purple. This phenomenon, known as residual trapping [42], is a critical factor in assessing the long-term storage capabilities of GCS projects.

1.3.2 Rock physics

The CO₂ saturation in the porous rocks and pressure change during CO₂ injection affects the compressibility of the rocks [45]. While other choices can be made, we follow the Patchy saturation model [46] for simplicity to mathematically describe the conversion from

CO₂ saturation to the decrease in the compressional wavespeed of the rock. This model follows the Gassmann's fluid substitution law and assumes that the medium as a mixture of CO₂ and brine is isotropic [47], which implies that the effective compressional modulus can be computed as a harmonic average of the CO₂-filled modulus and brine-filled modulus. This rock physics modeling does not involve any PDEs, and practically describes an elementwise relationship between CO₂ saturation and compressional wavespeed in the following form:

$$\mathbf{v}_i = \mathcal{R}(\mathbf{c}_p, \mathbf{c}_i) \quad \text{for } i = 1, 2, \dots, n_k. \quad (1.3)$$

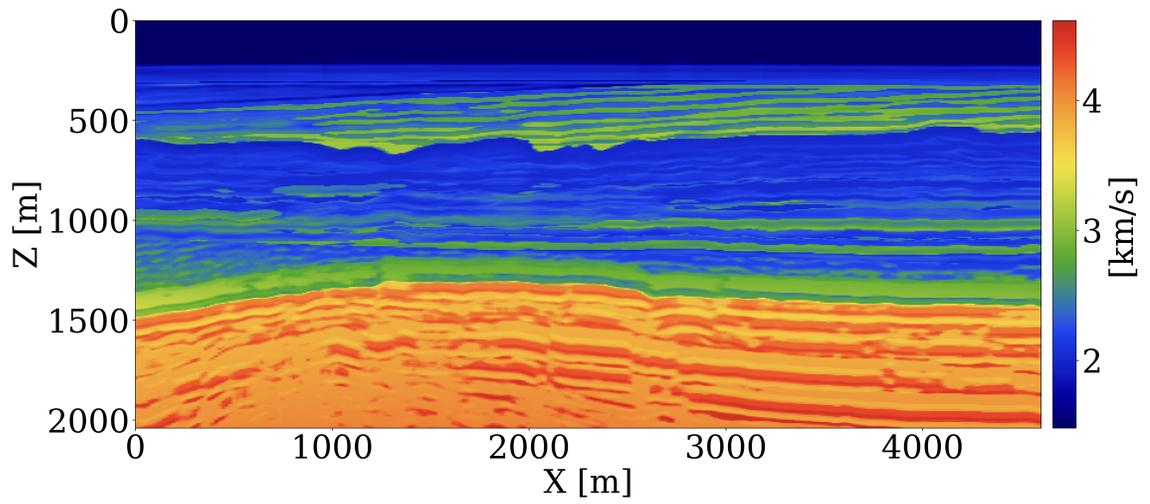
Here, the nonlinear modeling operator \mathcal{R} converts the brine-filled (pre-injection) compressional wavespeed, \mathbf{c}_p , and the CO₂ saturation snapshots, \mathbf{c}_i , to the altered wavespeed \mathbf{v}_i at time step i . Practically, this results in up to a few hundreds of meters per second decrease in compressional wavespeed of the rock due to CO₂ saturation. Figure 1.4b shows the decrease in compressional wavespeed according to the CO₂ saturation snapshot shown as the third plot in Figure 1.3b, where Figure 1.4a shows the brine-filled wavespeed model before CO₂ injection.

1.3.3 Wave physics

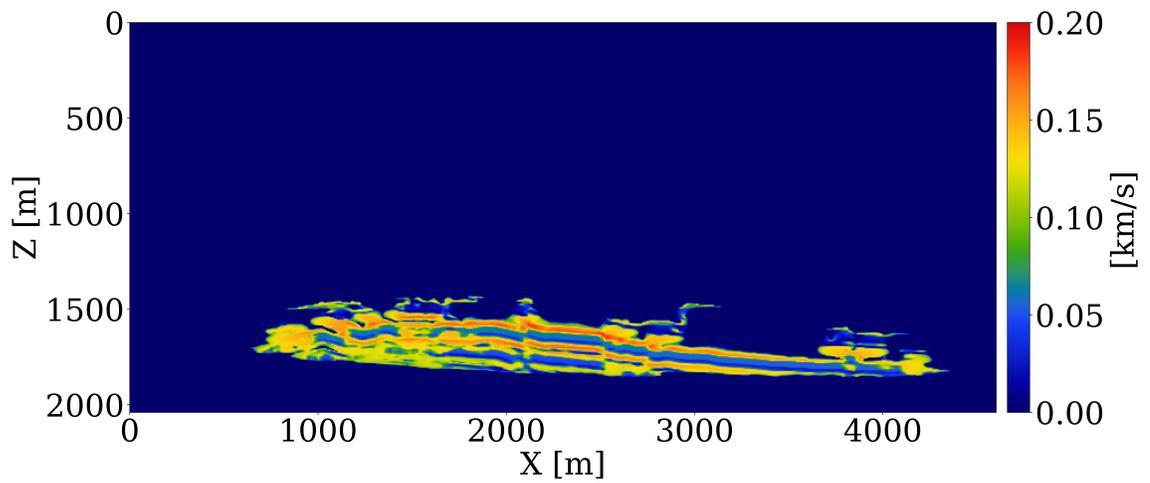
During GCS monitoring, multiple seismic surveys are collected to estimate the velocity model of the Earth subsurface and monitor the time-lapse changes. The CO₂-induced changes in compressional wavespeed of the rocks can be detected by time-lapse seismic data according to the seismic wave modeling in the following expression:

$$\mathbf{d}_i = \mathcal{F}_i(\mathbf{v}_i) \quad \text{for } i = 1, 2, \dots, n_k. \quad (1.4)$$

where \mathcal{F}_i represents the nonlinear seismic modeling operator (mentioned in Section 1.1) that simulates the seismic data at the i -th survey using wavespeed model \mathbf{v}_i .



(a)



(b)

Figure 1.4: (a) Brine-filled (pre-injection) wavespeed model. (b) Decrease in wavespeed induced by the CO₂ saturation at 25th year shown as third plot in Figure 1.3b.

1.3.4 The inverse problem

The primary goal of time-lapse monitoring is to estimate the past, current, and future behavior of CO₂ plume from available time-lapse seismic data. Mathematically, this involves solving an inverse problem. We observe the time-lapse seismic data, collected in $\{\mathbf{d}_i\}_{i=1}^{n_k}$, with some measurement noise $\{\epsilon_i\}_{i=1}^{n_k}$. We aim to recover time-varying changes of the subsurface properties, either wavespeed models at each survey, $\{\mathbf{v}_i\}_{i=1}^{n_k}$, or CO₂ saturation snapshots, $\{c_i\}_{i=1}^{n_k}$, or the intrinsic permeability model of the Earth, \mathbf{K} , such that the measurements modeled by these subsurface properties fit the observed ones. By comparing these properties at different time steps, we understand the dynamics of CO₂ during GCS projects.

1.3.5 Computational and numerical challenges

GCS monitoring shares many computational challenges with the FWI problem mentioned in Section 1.2.2: the parameter of interest is high-dimensional, and the parameter-to-observable map is computationally intensive. The solution space of this ill-posed inverse problem can be multimodal. In addition, it requires integration of different physics modeling operators. For example, estimation of the CO₂ saturation snapshots needs coupling of rock physics and wave physics. Estimation of the intrinsic permeability model of the Earth needs coupling between fluid-flow physics, rock physics, and wave physics. These couplings require not only propagating the parameter of interests to the observed time-lapse seismic data, but also calculating sensitivities of the observation with respect to the parameter of interests. In the field of geophysics imaging, however, software packages for modeling and inversion are typically based on monolithic low-level (C/Fortran) implementations. While these lead to performant codes for specific problems, maintaining the codebases and especially transporting the codebase to a different computing facility (i.e., cluster to cloud and vice versa), poses significant challenges. Consequently, coupling of different physics modeling units, which requires seamless integration of different software stacks including

seismic modeling and reservoir simulators, is considered challenging and costly. These intricacies are further compounded by the fact that few of open-source reservoir simulators support differentiation of the simulator’s output (time-varying CO_2 saturation) with respect to its input (permeability), with Jutul.jl [48] as one of the rare exceptions. These challenges call for a scalable, interoperable, and differentiable framework based on high-level software abstractions.

1.4 Contributions

This thesis consists of six papers, each of which partially addresses the challenges associated with solving large-scale geophysical inverse problems. Each chapter follows the general structure of a technical journal article and begins with a summary, followed by an introduction into the respective topic, and then describes the main contribution and provides numerical examples. Below I outline these chapters by summarizing the problem and my contribution.

- Chapter 2 (published in [49]) proposes a math-inspired differentiable programming framework based on software abstraction. At its core, this framework utilizes customized automatic differentiation (AD) rules to couple numerical simulators and deep neural networks for solving geophysical inversion problems at scale. My contribution includes design and implementation of this programming framework for multiphysics inversion. While the customized AD rules are already implemented for wave modeling operators in JUDI4Flux.jl [50] (integrated in JUDI.jl later), I extend this framework by implementing differentiation rules for reservoir simulators in JutulDarcy.jl [44], and develop the JutulDarcyRules.jl package [43]. Using this extension as an essential module, I implement an end-to-end permeability inversion workflow by integrating two independent state-of-the-art numerical modeling software libraries. In addition, I propose to use FNOs as a surrogate modeling operator for reservoir simulator, and to replace the numerical simulator by the trained FNOs. I

demonstrate the efficacy of this differentiable programming framework by designing and conducting a synthetic case study for geological carbon storage monitoring.

- Chapter 3 (accepted for publication in [51]) presents a feasibility study for time-lapse full-waveform permeability inversion using the differentiable programming framework in Chapter 2. My contributions include illustrating the framework of using time-lapse seismic data to invert for reservoir permeability in a GCS setting, and fully examining the sensitivity of the end-to-end inversion framework with respect to initial models, modeling errors, and potential crosstalk during multiparameter inversion. The case study also investigates the potential of the framework for forecasting long-term CO₂ storage under residual trapping mechanism, which is crucial for GCS monitoring and verification.
- Chapter 4 (published in [52]) proposes a data-driven algorithm to solve multiphysics-based inverse problems. My contributions include a constrained optimization method that combines computationally cheap learned surrogates with learned constraints. This method not only allows for computationally efficient inversion thanks to the orders of magnitude speed-up from learned surrogates (FNOs), the accuracy of the surrogates is also safeguarded via inclusion of a trained normalizing flow (NF) that forces the model iterates to remain in-distribution. I support this method with a synthetic case study in which the accuracy curves of the FNO surrogates under different inversion methods are compared, illustrating the importance of the NF-based constraints during inversion.
- Chapter 5 (published in [53]) introduces an integrated workflow for GCS monitoring, which is generally designed for scenarios where CO₂ dynamics is not assumed to adhere exactly to the multiphase flow equations. My contributions include a joint inversion algorithm for time-lapse seismic imaging and integration of a deep neural classifier for CO₂ leakage detection. I exploit the fact that the Earth models at dif-

ferent surveys, as the unknown parameters of interest, share a common component. Therefore, I reparameterize the unknown Earth models accordingly and derive an algorithm based on linearized Bregman methods to jointly invert for time-lapse seismic images. Through hundreds of inversion experiments performed on time-lapse seismic datasets, I demonstrate that this approach effectively produces images in higher quality without costly replication of seismic surveys. A deep neural classifier is further trained to automatically detect CO₂ leakage.

- Chapter 6 (published in [54]) proposes an amortized variational inference framework (called WISE) for solving the FWI problem. My contributions include a systematic approach to utilize the extended gradient (common-image gathers) as a suitable physics-informed summary statistics for FWI thanks to their capability to encapsulate the observed data. In this approach, I integrate generative artificial intelligence techniques with the physics-informed common-image gathers to quantify uncertainty in the velocity models and its impact on imaging. Considered case studies juxtapose the inference results when the standard gradient and the extended gradient are used as summary statistics, and demonstrate the necessity of using extended gradient for variational inference when the initial model is poor. I further translate the uncertainty in the velocity models to the uncertainty in imaged reflectors in the downstream tasks.
- Chapter 7 (preprint at [55]) builds on top of Chapter 6 to introduce a semi-amortized variational inference framework (called WISER) for solving the FWI problem. My contributions include a refinement approach with frugal use of the wave physics to mitigate the amortization gap existing in the WISE framework in Chapter 6. I verify the efficacy of this framework by juxtaposing the inference results from WISE and WISER, and further demonstrate the robustness of WISER with respect to potential out-of-distribution scenarios at inference.

Chapter 8 presents the conclusion of this thesis and discuss future research directions.

1.5 References

- [1] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter estimation and inverse problems*. Elsevier, 2018 (page 1).
- [2] R. Snieder and J. Trampert, *Inverse problems in geophysics*. Springer, 1999 (page 1).
- [3] D. Vigh, X. Cheng, J. Xu, and K. Jiao, “Forge ahead in acquisition and model building via full-waveform inversion: Gulf of Mexico case study,” *Geophysics*, vol. 88, no. 5, B285–B295, 2023 (page 1).
- [4] A.-K. Furre, O. Eiken, H. Alnes, J. N. Vevatne, and A. F. Kiær, “20 years of monitoring CO₂-injection at Sleipner,” *Energy Procedia*, vol. 114, pp. 3916–3926, 2017, 13th International Conference on Greenhouse Gas Control Technologies, GHGT-13, 14-18 November 2016, Lausanne, Switzerland (pages 1, 6).
- [5] R. Scarpa, R. I. Tilling, and S. McNutt, “Seismic monitoring and eruption forecasting of volcanoes: A review of the state-of-the-art and case histories,” *Monitoring and mitigation of volcano hazards*, pp. 99–146, 1996 (page 1).
- [6] J. C. Strikwerda, *Finite difference schemes and partial differential equations*. SIAM, 2004 (page 1).
- [7] B. Szabó and I. Babuška, “Finite element analysis: Method, verification and validation,” 2021 (page 1).
- [8] R. Eymard, T. Gallouët, and R. Herbin, “Finite volume methods,” *Handbook of numerical analysis*, vol. 7, pp. 713–1018, 2000 (page 1).
- [9] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer, 1999 (page 1).
- [10] M. K. Cowles and B. P. Carlin, “Markov chain Monte Carlo convergence diagnostics: A comparative review,” *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 883–904, 1996 (page 2).
- [11] G. H. Golub, P. C. Hansen, and D. P. O’Leary, “Tikhonov regularization and total least squares,” *SIAM journal on matrix analysis and applications*, vol. 21, no. 1, pp. 185–194, 1999 (page 2).
- [12] M. S. Zhdanov, *Geophysical inverse theory and regularization problems*. Elsevier, 2002, vol. 36 (page 2).

- [13] M. Duff, N. D. Campbell, and M. J. Ehrhardt, “Regularising inverse problems with generative machine learning models,” *Journal of Mathematical Imaging and Vision*, vol. 66, no. 1, pp. 37–56, 2024 (page 2).
- [14] Y. Gal, P. Koumoutsakos, F. Lanusse, G. Louppe, and C. Papadimitriou, “Bayesian uncertainty quantification for machine-learned models in physics,” *Nature Reviews Physics*, vol. 4, no. 9, pp. 573–577, 2022 (page 2).
- [15] J. Virieux and S. Operto, “An overview of full-waveform inversion in exploration geophysics,” *Geophysics*, vol. 74, no. 6, WCC1–WCC26, 2009 (page 3).
- [16] P. A. Witte, “Software and algorithms for large-scale seismic inverse problems,” 2020 (page 3).
- [17] P. A. Witte *et al.*, “A large-scale framework for symbolic implementations of seismic inversion algorithms in julia,” *GEOPHYSICS*, vol. 84, no. 3, F57–F71, 2019. eprint: <https://doi.org/10.1190/geo2018-0174.1> (page 5).
- [18] K. Cranmer, J. Brehmer, and G. Louppe, “The frontier of simulation-based inference,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 055–30 062, 2020 (page 4).
- [19] O. Gauthier, J. Virieux, and A. Tarantola, “Two-dimensional nonlinear inversion of seismic waveforms: Numerical results,” *geophysics*, vol. 51, no. 7, pp. 1387–1403, 1986 (page 6).
- [20] A. Tarantola, “Inversion of seismic reflection data in the acoustic approximation,” *Geophysics*, vol. 49, no. 8, pp. 1259–1266, 1984 (page 6).
- [21] E. J. Moniz, “Innovating a green real deal,” *Science*, vol. 364, no. 6445, pp. 1013–1013, 2019 (page 6).
- [22] M. E. Boot-Handford *et al.*, “Carbon capture and storage update,” *Energy & Environmental Science*, vol. 7, no. 1, pp. 130–189, 2014 (page 6).
- [23] C. Marchetti, “On geoengineering and the co2 problem,” *Climatic change*, vol. 1, no. 1, pp. 59–68, 1977 (page 6).
- [24] S. Whittaker *et al.*, “A decade of co2 injection into depleting oil fields: Monitoring and research activities of the iea ghg weyburn-midale co2 monitoring and storage project,” *Energy Procedia*, vol. 4, pp. 6069–6076, 2011 (page 6).
- [25] K. Michael *et al.*, “Geological storage of co2 in saline aquifers—a review of the experience from existing storage operations,” *International journal of greenhouse gas control*, vol. 4, no. 4, pp. 659–667, 2010 (page 6).

- [26] M. J. Bickle, “Geological carbon storage,” *Nature Geoscience*, vol. 2, no. 12, pp. 815–818, 2009 (page 6).
- [27] P. Newell and A. G. Ilgen, “Overview of geological carbon storage (GCS),” in *Science of Carbon Storage in Deep Saline Formations*, Elsevier, 2019, pp. 1–13 (page 6).
- [28] J. M. Nordbotten and M. A. Celia, “Geological storage of CO₂: Modeling approaches for large-scale simulation,” in *Geological Storage of CO₂: Modeling Approaches for Large-Scale Simulation*, John Wiley and Sons, 2011 (pages 6, 7).
- [29] B. M. Freifeld, T. M. Daley, S. D. Hovorka, J. Henniges, J. Underschultz, and S. Sharma, “Recent advances in well-based monitoring of co₂ sequestration,” *Energy Procedia*, vol. 1, no. 1, pp. 2277–2284, 2009 (page 6).
- [30] J. P. Noguees, J. M. Nordbotten, and M. A. Celia, “Detecting leakage of brine or co₂ through abandoned wells in a geological sequestration operation using pressure monitoring wells,” *Energy Procedia*, vol. 4, pp. 3620–3627, 2011 (page 6).
- [31] S. L. Nooner, O. Eiken, C. Hermanrud, G. S. Sasagawa, T. Stenvold, and M. A. Zumberge, “Constraints on the in situ density of co₂ within the utsira formation from time-lapse seafloor gravity measurements,” *international journal of greenhouse gas control*, vol. 1, no. 2, pp. 198–214, 2007 (page 6).
- [32] H. Alnes, O. Eiken, S. Nooner, G. Sasagawa, T. Stenvold, and M. Zumberge, “Results from sleipner gravity monitoring: Updated density and temperature distribution of the co₂ plume,” *Energy Procedia*, vol. 4, pp. 5504–5511, 2011 (page 6).
- [33] J. M. Carcione, D. Gei, S. Picotti, and A. Michelini, “Cross-hole electromagnetic and seismic modeling for co₂ detection and monitoring in a saline aquifer,” *Journal of petroleum science and engineering*, vol. 100, pp. 162–172, 2012 (page 7).
- [34] M. S. Zhdanov *et al.*, “Electromagnetic monitoring of co₂ sequestration in deep reservoirs,” *first break*, vol. 31, no. 2, 2013 (page 7).
- [35] R. Arts, O. Eiken, A. Chadwick, P. Zweigel, L. van der Meer, and B. Zinszner, “Monitoring of co₂ injected at sleipner using time-lapse seismic data,” *Energy*, vol. 29, no. 9-10, pp. 1383–1392, Jul. 2004 (page 7).
- [36] D. Lumley, “4d seismic monitoring of co₂ sequestration,” *The Leading Edge*, vol. 29, no. 2, pp. 150–155, 2010 (page 7).
- [37] L. Huang, “Geophysical monitoring for geologic carbon storage,” 2022 (page 7).

- [38] H. Gross and A. Mazuyer, “Geosx: A multiphysics, multilevel simulator designed for exascale computing,” in *SPE Reservoir Simulation Conference*, OnePetro, 2021 (page 7).
- [39] A. F. Rasmussen *et al.*, “The open porous media flow reservoir simulator,” *Computers & Mathematics with Applications*, vol. 81, pp. 159–185, 2021 (page 7).
- [40] K.-A. Lie, *An introduction to reservoir simulation using MATLAB/GNU Octave: User guide for the MATLAB Reservoir Simulation Toolbox (MRST)*. Cambridge University Press, 2019 (page 7).
- [41] O. Møyner, “Jutuldarcy.jl - a fully differentiable high-performance reservoir simulator based on automatic differentiation,” in *European Conference on the Mathematics of Geological Reservoirs*, Sep. 2024 (page 7).
- [42] T. Rahman, M. Lebedev, A. Barifcani, and S. Iglauer, “Residual trapping of supercritical co2 in oil-wet sandstone,” *Journal of Colloid and Interface Science*, vol. 469, pp. 63–68, May 2016 (page 8).
- [43] Z. Yin, G. Bruer, and M. Louboutin, *Slimgroup/jutuldarcyrules.jl: V0.2.5*, version v0.2.5, Apr. 2023 (pages 8, 12).
- [44] O. Møyner, G. Bruer, and Z. Yin, *Sintefmath/jutuldarcy.jl: V0.2.3*, version v0.2.3, Apr. 2023 (pages 8, 12).
- [45] G. Côté, H. Amini, and C. MacBeth, “Bayesian inversion of 4d seismic data to pressure and saturation changes: Application to a west of shetlands field,” *Geophysical Prospecting*, vol. 71, no. 2, pp. 292–321, 2023 (page 8).
- [46] P. Avseth, T. Mukerji, and G. Mavko, *Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk*. Cambridge university press, 2010 (page 8).
- [47] R. Hill, “Elastic properties of reinforced solids: Some theoretical principles,” *Journal of the Mechanics and Physics of Solids*, vol. 11, no. 5, pp. 357–372, 1963 (page 9).
- [48] O. Møyner, M. Johnsrud, H. M. Nilsen, X. Raynaud, K. O. Lye, and Z. Yin, *Sintefmath/jutul.jl: V0.2.6*, version v0.2.6, Apr. 2023 (page 12).
- [49] M. Louboutin *et al.*, “Learned multiphysics inversion with differentiable programming and machine learning,” *The Leading Edge*, vol. 42, no. 7, pp. 474–486, 2023 (page 12).

- [50] P. A. Witte, M. Louboutin, and F. J. Herrmann, *Judi4flux: Seismic modeling for deep learning*, version v0.1.1, Dec. 2020 (page 12).
- [51] Z. Yin, M. Louboutin, O. Møyner, and F. J. Herrmann, “Time-lapse full-waveform permeability inversion: A feasibility study,” *The Leading Edge*, vol. 43, 8 Aug. 2024, (The Leading Edge) (page 13).
- [52] Z. Yin, R. Orozco, M. Louboutin, and F. J. Herrmann, “Solving multiphysics-based inverse problems with learned surrogates and constraints,” *Advanced Modeling and Simulation in Engineering Sciences*, vol. 10, no. 1, p. 14, 2023 (page 13).
- [53] Z. Yin, H. T. Erdinc, A. P. Gahlot, M. Louboutin, and F. J. Herrmann, “Derisking geologic carbon storage from high-resolution time-lapse seismic to explainable leakage detection,” *The Leading Edge*, vol. 42, no. 1, pp. 69–76, 2023 (page 13).
- [54] Z. Yin, R. Orozco, M. Louboutin, and F. J. Herrmann, “Wise: Full-waveform variational inference via subsurface extensions,” *Geophysics*, vol. 89, no. 4, pp. 1–31, 2024 (page 14).
- [55] Z. Yin, R. Orozco, and F. J. Herrmann, “Wiser: Multimodal variational inference for full-waveform inversion without dimensionality reduction,” *arXiv preprint arXiv:2405.10327*, 2024 (page 14).

CHAPTER 2

LEARNED MULTIPHYSICS INVERSION WITH DIFFERENTIABLE PROGRAMMING AND MACHINE LEARNING

2.1 Summary

This chapter presents a differentiable programming software framework for computational geophysics and, more generally, inverse problems involving the wave-equation (e.g., seismic and medical ultrasound), regularization with learned priors, and learned neural surrogates for multiphase flow simulations. By integrating multiple layers of abstraction, our software is designed to be both readable and scalable. This allows researchers to easily formulate their problems in an abstract fashion while exploiting the latest developments in high-performance computing. We illustrate and demonstrate our design principles and their benefits by means of building a scalable prototype for permeability inversion from time-lapse crosswell seismic data, which aside from coupling of wave physics and multiphase flow, involves machine learning.

2.2 Motivation

Thanks to major advancements in high-performance computing (HPC) techniques, computational (exploration) geophysics has made giant leaps over the past decades. These developments have, for instance, led to the adoption of wave-equation-based inversion technologies such as full-waveform inversion (FWI) and reverse-time migration (RTM) that, thanks to their adherence to wave physics, have resulted in superior imaging in complex geologies. While these techniques certainly rank amongst the most sophisticated imaging technologies, their implementation relies with few exceptions—most notably iWave++ [1], Julia Devito Inversion framework (JUDI.jl, [2]), COFII framework [3]—on monolithic

low-level (C/Fortran) implementations. As a consequence, due to their lack of abstraction and modern programming constructs, these low-level implementations are difficult and very costly to maintain, especially when performance considerations prevail over best software practices. A noteworthy attempt at modernizing wave-equation inversion frameworks is deepwave [4], which implements FWI using pytorch [5]. Despite state-of-the-art examples and applications for 2D inversion, this work is limited by the aforementioned pitfalls as it relies on handwritten low level C/Cuda code reducing the flexibility and extensibility to new physics and three dimensional problems. It also does not integrate machine learning with full-waveform inversion as advocated in this chapter. While these implementation design choices lead to performant code for specific problems, such as FWI, they often hinder the implementation of new algorithms, e.g., based on different objective functions or constraints, as well as coupling existing code bases with external software libraries. For instance, combining wave-equation-based inversion with machine learning frameworks or coupling wave-physics with multiphase fluid-flow solvers are considered challenging and costly. These hurdles pose challenges on the geophysical researchers to innovate new algorithms and workflows. In this chapter, we present a flexible and agile software framework that aims to resolve these challenges and is designed to be scalable, differentiable, and interoperable. We first introduce the design principles of our software framework, followed by a concrete usage scenario for time-lapse seismic monitoring of geological carbon storage. This illustrative and didactic example involves the integration of multiple software modules for different types of physics with machine-learning techniques such as learned deep priors and neural surrogates. For each module, we explain the choices we made and how these modules are connected through software abstractions and overarching high-level programming language constructs. The advocacy of our proposed framework is demonstrated on a preliminary 2D case study involving the realistic Compass model. We conclude by discussing remaining challenges and future work directions.

2.3 Design principles

To address these important shortcomings of current software implementations that impede progress, we adopt contemporary software design practices that include high-level abstractions, software design principles, and the utilization of modern programming languages, such as Python [6] and Julia [7]. We also make extensive use of abstractions provided by domain-specific languages (DSLs), such as the Rice Vector Library [RVL, 8] and the Unified Form Language [UFL, 9, 10], and adopt reproducible research practices introduced by the trailblazing open-source initiative Madagascar [11], which successfully made use of version control and an abstraction based on the software construction tool `SCons`.

In an effort to meet the challenges of modern software design in a performance-critical environment, we adhere to three key principles—in addition to the fundamental principle of separation of concerns. First, we adopt mathematical language to inform our abstractions. Mathematics is concise, unambiguous, well understood, and leads to natural abstractions for the

- **wave physics**, through partial differential equations as put to practice by Devito, which relies on Symbolic Python (SymPy) [12] to define partial differential equations. Given the symbolic expressions, Devito automatically generates highly-optimized, possibly domain-decomposed, parallel C code that targets the available hardware with near-optimal performance for 3D acoustic, tilted-transverse-isotropic, or elastic wave-equations;
- **linear algebra**, through matrix-free linear operators, as in JUDI.jl [2]—a high-level linear algebra DSL for wave-equation-based modeling and inversion. These ideas date back to SPOT [13] with more recent implementations JOLI.jl [14] in Julia and PyLops in Python [15];
- **optimization**, through definition of objective functions, also known as loss functions, that need to be minimized—via SlimOptim.jl [16]—subject to mathematical

constraints, which can be imposed through `SetIntersectionProjection.jl` [17, 18].

Second, we exploit hierarchy within wave-equation-based inversion problems that naturally leads to a separation of concerns. At the highest level, we deal with linear operators, specifically matrix-free Jacobians of wave-based inversion, with `JUDI.jl` and parallel file input/output with `SegyIO.jl` [19] on premise, or in the Cloud (Azure) via `JUDI4Cloud.jl` [20] and `CloudSegyIO.jl` [21]. At the intermediate and lower level, we make extensive use of Devito [22, 23]—a just-in-time compiler for stencil-based time-domain finite-difference calculations.

Third, we build on the principles of differentiable programming as advocated by [24] and intrusive automatic differentiation introduced by [25] to integrate wave-physics with machine learning frameworks and multiphase flow. Specifically, we employ automatic differentiation (AD) through the use of the chain rule, including abstractions that allow the user to add derivative rules, as in `ChainRules.jl` [26, 27].

Aside from developing software for wave-equation-based inversion, we have more recently also been involved in the development of scalable machine learning solutions, including the Julia package `InvertibleNetworks.jl` [28], which implements memory-efficient invertible deep neural networks such as (conditional) normalizing flows [NFs, 29], and scalable distributed Fourier neural operators [FNOs, 30] in the `dfno` package [31, 32]. All of these will be described in more detail below.

To illustrate how these design principles can lead to solutions of complex learned coupled inversions, we consider in the ensuing sections end-to-end inversion of time-lapse seismic data for the spatial permeability distribution [25]. As can be seen from Figure 2.1, this inversion problem is rather complex and whose solution arguably benefits from our three design principles listed above. In this formulation, the latent representation for the permeability is taken via a series of nonlinear operations all the way to the time-lapse seismic data. In the remainder of this exposition, we will detail how the different components in this learned inversion problem are implemented so that the coupled inversion can be car-

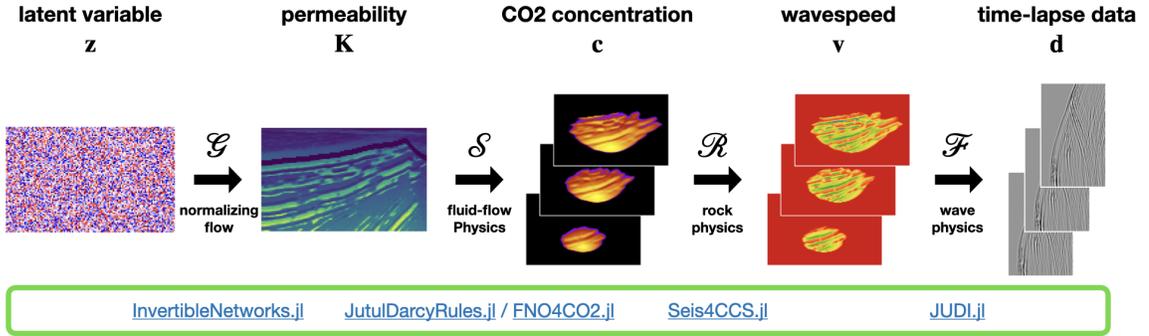


Figure 2.1: The multiphysics forward model. The permeability, K , is generated from Gaussian noise with a pretrained NF, \mathcal{G} , followed by two-phase flow simulations through \mathcal{S} , rock physics denoted by \mathcal{R} , and time-lapse seismic data simulations via wave physics, \mathcal{F} .

ried out. The results presented are preliminary representing a snapshot on how research is conducted according to the design principles.

2.4 Learned time-lapse end-to-end permeability inversion

Combating climate change and dealing with the energy transition call for solutions to problems of increasing complexity. Building seismic monitoring systems for geological CO₂ and/or H₂ storage falls in this category. To demonstrate how math-inspired abstractions can help, we consider inversion of permeability from crosswell time-lapse data (see Figure 2.2 for experimental setup) involving (i) coupling of wave physics with two-phase (brine/CO₂) flow using Jutul.jl [33], state-of-the-art reservoir modeling software in Julia; (ii) learned regularization with NFs with InvertibleNetworks.jl; (iii) learned surrogates for the fluid-flow simulations with FNOs. This type of inversion problem is especially challenging because it involves different types of physics to estimate the past, current, and future saturation and pressure distributions of CO₂ plumes from crosswell data in saline aquifers. In the subsequent sections, we demonstrate how we invert time-lapse data using the separate software packages listed in Figure 2.1.

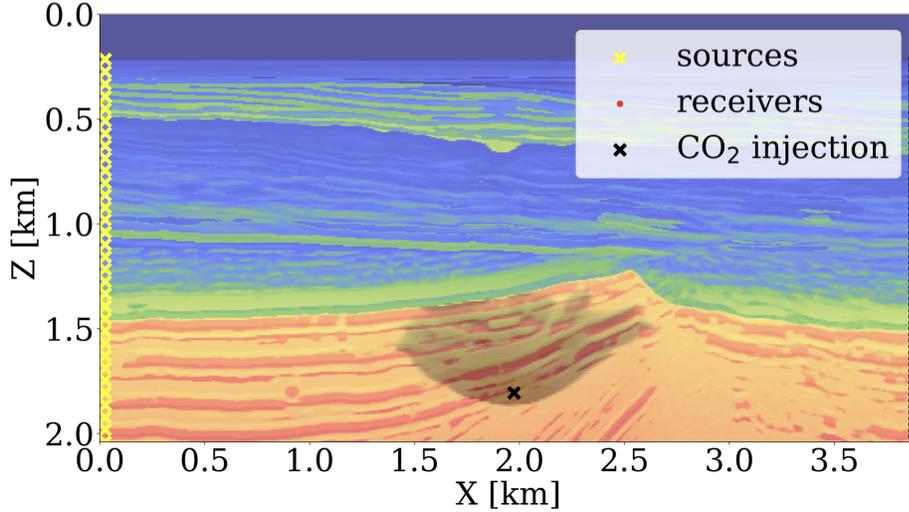


Figure 2.2: Experimental setup. The black \times symbol in the middle of the model indicates the CO_2 injection location. The seismic sources are on the left-hand side of the model (shown as yellow \times symbols) and receivers are on the right-hand side of the model (shown as red dots). In grey color, we overlay the compressional wavespeed with simulated CO_2 saturation modeled for 18 years.

2.4.1 Wave-equation-based inversion

Thanks to its unmatched ability to resolve CO_2 plumes, active-source time-lapse seismic is arguably the preferred imaging modality when monitoring geological storage [34]. In its simplest form for a single time-lapse vintage, FWI involves minimizing the ℓ_2 -norm misfit/loss function between observed and synthetic data—i.e., we have

$$\underset{\mathbf{m}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{F}(\mathbf{m})\mathbf{q} - \mathbf{d}\|_2^2 \quad \text{where} \quad \mathbf{F}(\mathbf{m}) = \mathbf{P}_r \mathbf{A}(\mathbf{m})^{-1} \mathbf{P}_s^\top. \quad (2.1)$$

In this formulation, the symbol $\mathbf{F}(\mathbf{m})$ represents the forward modeling operator (wave physics), parameterized by the squared slowness, \mathbf{m} . This forward operator acting on the sources consists of the composition of source injection operator, \mathbf{P}_s^\top with $^\top$ denoting the transpose operator, solution of the discretized wave equation via $\mathbf{A}(\mathbf{m})^{-1}$, and restriction to the receivers via the linear operator \mathbf{P}_r . The vector \mathbf{q} represents the seismic sources and the vector \mathbf{d} contains single-vintage seismic data, collected at the receiver locations. Thanks to our adherence to the math, the corresponding Julia code to invert for the unknown squared

slowness m with JUDI.jl reads

```
# Forward modeling to generate seismic data.
Pr = judiProjection(recGeometry) # setup receiver
Ps = judiProjection(srcGeometry) # setup sources
Ainv = judiModeling(model)      # setup wave-equation solver
F = Pr * Ainv * Ps'             # forward modeling operator
d = F(m_true) * q               # generate observed data
# Gradient descent to invert for the unknown squared slowness.
for it = 1:maxiter
    d0 = F(m) * q                # generate synthetic data
    J = judiJacobian(F(m), q)    # setup the Jacobian operator of F
    g = J' * (d0 - d)           # gradient w.r.t. squared slowness
    m = m - t * g               # gradient descent with steplength t
end
```

To obtain this concise and abstract formulation for FWI, we utilized hierarchical abstractions for propagators in Devito and linear algebra tools in JUDI.jl, including matrix-free implementations for F and its Jacobian J . While the above stand-alone implementation allows for (sparsity-promoting) seismic [35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46] and medical [47, 48, 49, 50] inversions, it relies on hand-derived implementations for the adjoint of the Jacobian J' and for the derivative of the loss function. Although this approach is viable, relying solely on hand-derived derivatives can become cumbersome when we want to utilize machine learning models or when we need to couple the wave equation to the multiphase flow equation.

To allow for this situation, we make use of Julia’s differentiable programming ecosystem that includes tools to use AD and to add differentiation rules via ChainRules.jl. Using this tool, the AD system can be taught how to differentiate JUDI.jl via the following differentiation rule for the forward propagator

```
# Custom AD rule for wave modeling operator.
```

```

function rrule(::typeof(*), F::judiModeling, q)
    y = F * q                # forward modeling
    # The pullback function for gradient calculations.
    pullback(dy) = NoTangent(), judiJacobian(F, q)' * dy, F' * dy
    return y, pullback
end

```

In this rule, the `pullback` function takes as input the data residual, dy , and outputs the gradient of $F * q$ with respect to the operator $*$ (no gradient), the model parameters, and the source distribution. With this differentiation rule, the above gradient descent algorithm can be implemented as follows:

```

# Define the loss function.
loss(m) = .5f0 * norm(F(m) * q - d)^2f0
# Gradient descent to invert for the squared slowness.
for it = 1:maxiter
    g = gradient(loss, m)[1]    # gradient computation via AD
    m = m - t * g              # gradient descent with steplength t
end

```

Compared to the original implementation, this code only needs $F(m)$ and the function `loss(m)`. With the help of the above `rrule`, Julia's AD system¹ is capable of computing the gradients (line 5). Aside from remaining performant, i.e., we still make use of the adjoint-state method to compute the gradients, the advantage of this approach is that it allows for much more flexibility, e.g., in situations where the squared slowness is parameterized in terms of a pretrained neural network or in terms of the output of multiphase flow simulations. In the next section, we show how trained NFs can serve as priors to improve the quality of FWI.

¹In this case, we used reverse AD provided by `Zygote.jl`, the AD system provided by Julia machine learning package `Flux.jl`. Because `ChainRules.jl` is AD system agnostic, another choice could have been made.

2.4.2 Deep priors and normalizing flows

Normalizing Flows are generative models that take advantage of invertible deep neural network architectures to learn complex distributions from training examples [51]. The term, “flow” refers to the transformation of data from a complex distribution to a simple one. The term “normalizing” refers to the standard Gaussian (Normal) target distribution the network learns to map images to. For example, in seismic inversion applications, we are interested in approximating the distribution of Earth models to use as priors in downstream tasks. NFs learn to map samples from the target distribution (i.e., Earth models) to zero-mean unit standard deviation Gaussian noise using a sequence of trainable nonlinear invertible layers. Once trained, one can resample new Gaussian noise and pass it through the inverse sequence of layers to obtain new generative samples from the target distribution. NFs are an attractive choice for generative models in seismic applications [52, 53, 54, 55, 56, 57, 58] because they provide fast sampling and allow for memory-efficient training due to their intrinsic invertibility, which eliminates the need to store intermediate activations during backpropagation. Memory efficiency is particularly important for seismic applications due to the 3D volumetric nature of the seismic models. Thus, our methods need to scale well in this regime.

To illustrate the practical use of NFs as priors in seismic inverse problems, we trained an NF on slices from the Compass model [59]. The training of a normalizing flow is laid out in Figure 2.3, where for illustrative purposes, we demonstrate a training run on small (64×64) slices of the Compass model. Each row shows the normalization (image \mathbf{m} transformed to \mathbf{Z}_m intended to be white zero-mean standard deviation one Gaussian noise) during training and its generative inverse (white noise $\mathbf{z} \sim \mathcal{N}(0, 1)$ to image $\tilde{\mathbf{m}}$) during each epoch. From Figure 2.5, we clearly observe the intended behavior. As the training proceeds, the NFs transform the true model better towards white noise while its inverse progressively generates more realistic looking generative velocity models. To perform a comparison with traditional FWI, we train an NF on full model size slices (512×256 grid

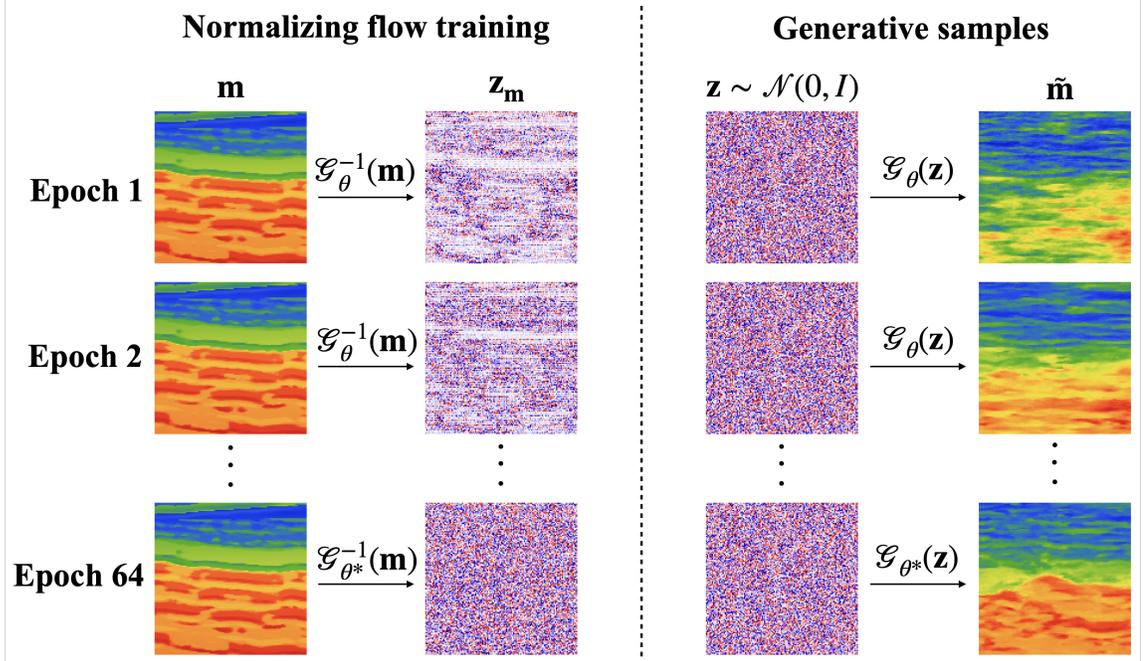


Figure 2.3: Demonstration of Gaussinazation of Compass slices during training of normalizing flow. The data used for this didactic example is openly available and this figure fully in the InvertibleNetworks.jl repository.

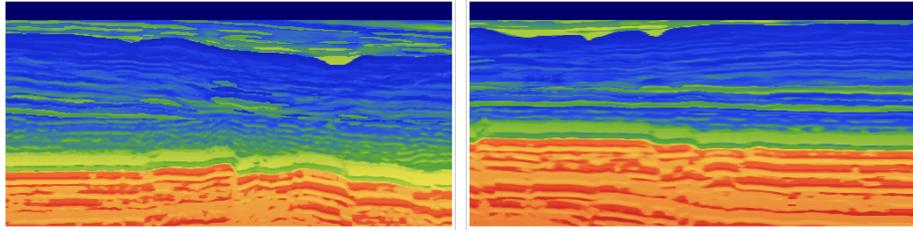


Figure 2.4: Examples of Compass 2D slices used to train a normalizing flow prior.

points). In Figure 2.5, we compare generative samples from the NF with the slices used to train the model shown in Figure 2.4. Although there are still irregularities, the model has learned important qualitative aspects of the model that will be useful in inverse problems. To demonstrate this usefulness, we test our prior on an FWI inverse problem. Because our NF prior is trained independently, it is flexible and can easily be plugged into different inverse problems.

Our FWI experiment includes: ocean bottom nodes, Ricker wavelet with no energy below 4Hz, additive colored Gaussian noise that has the same bandwidth as the noise-free

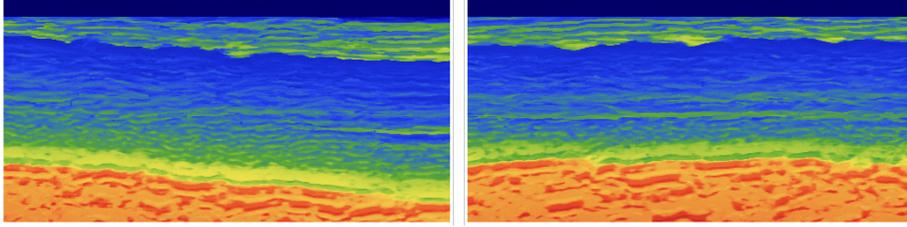


Figure 2.5: Generative samples from our trained prior. Their similarity to the training samples in Figure 2.4 suggests that our normalizing flow has learned a useful prior..

data. For FWI with our learned prior, we minimize

$$\underset{\mathbf{z}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{F}(\mathcal{G}_{\theta^*}(\mathbf{z}))\mathbf{q} - \mathbf{d}\|_2^2 + \frac{\lambda}{2} \|\mathbf{z}\|_2^2, \quad (2.2)$$

where \mathcal{G}_{θ^*} is a pretrained NF with weights θ^* . After training, the inverse of the NF maps realistic Compass-like Earth samples to white noise—i.e., $\mathcal{G}_{\theta^*}^{-1}(\mathbf{m}) = \mathbf{z} \sim \mathcal{N}(0, I)$. Since the NFs are designed to be invertible, the action of the pretrained NF, \mathcal{G}_{θ^*} , on Gaussian noise \mathbf{z} produces realistic samples of Earth models (see Figure 2.5). We use this capability in the above equation where the unknown model parameters in \mathbf{m} are reparameterized by $\mathcal{G}_{\theta^*}(\mathbf{z})$. The regularization term, $\frac{\lambda}{2} \|\mathbf{z}\|_2^2$, penalizes the latent variable \mathbf{z} with large ℓ_2 norm, where λ balances the misfit and regularization terms. Consequently, this learned regularizer encourages FWI results that are more likely to be realistic Earth models [60]. However, notice that the optimization routine now requires differentiation through both the physical operator (wave physics, \mathbf{F}) and the pretrained NF (\mathcal{G}_{θ^*}), and only a true invertible implementation like ours, with minimal memory imprint for both training and inference, can provide scalability.

Thanks to the JUDI.jl’s `rrule` for `F` and `InvertibleNetworks.jl`’s `rrule` for `G`, integration of machine learning with FWI becomes straightforward involving replacement of `m` by `G(z)` on line 6. Minimizing the objective function in 2.2 now translates to

```
# Load the pretrained NF and weights.
G = NetworkGlow(nc, nc_hidden, depth, nscales)
set_params!(G, theta)
```

```

# Set up the ADAM optimizer.
opt = ADAM()
# Define the reparameterized loss function including penalty term.
loss(z) = .5f0 * norm(F(G(z)) * q - d)^2f0 + .5f0 * lambda * norm(z)^2f0
# ADAM iterations.
for it = 1:maxiter
    g = gradient(loss, z)[1]      # gradient computation with AD
    update!(opt, z, g)           # update z with ADAM
end
# Convert latent variable to squared slowness.
m = G(z)

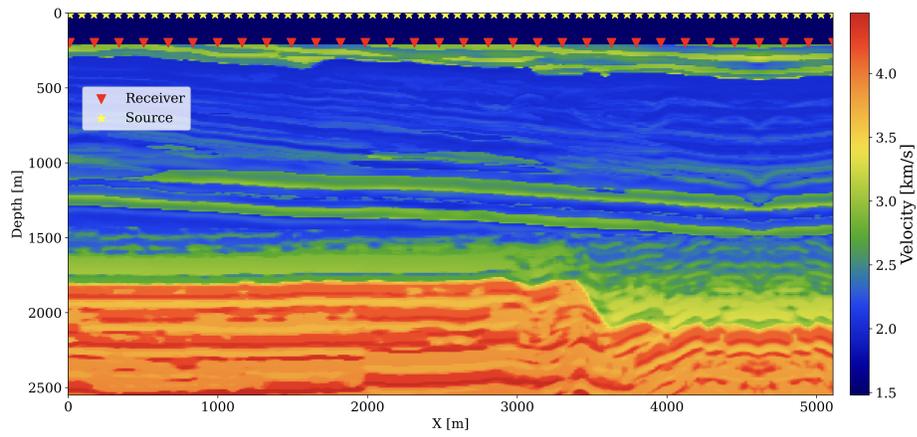
```

In Figure 2.6, we compare the results of FWI with our learned prior against unregularized FWI. Since our prior regularizes the solution towards realistic models, we obtain a velocity estimate that is closer to the ground truth. To measure the performance of our method, we use Peak Signal to Noise Ratio (PSNR) and see an increase from 12.98 dB with traditional FWI to 14.77 dB with the learned prior.

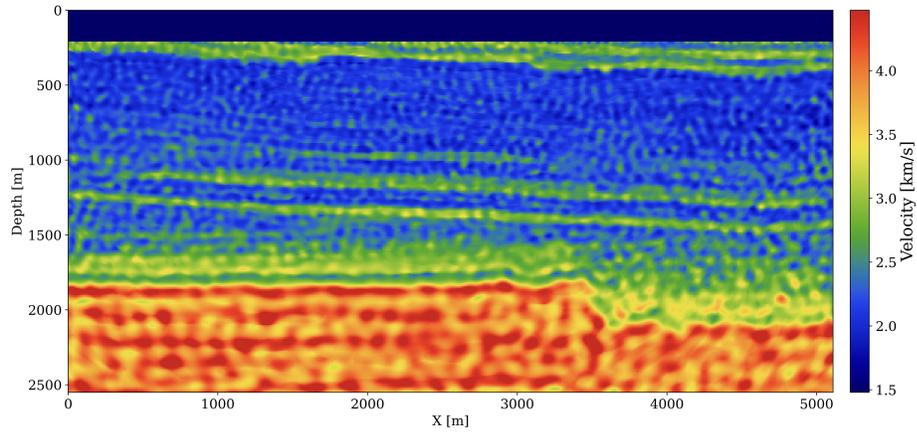
Through this simple example, we demonstrated the ability to easily integrate our state-of-the-art wave-equation propagators with the Julia’s differentiable programming system. By applying these design principles to other components of the end-to-end inversion, we design a seismic monitoring framework for real-world applications in subsurface reservoirs.

2.4.3 Fluid-flow simulation and permeability inversion

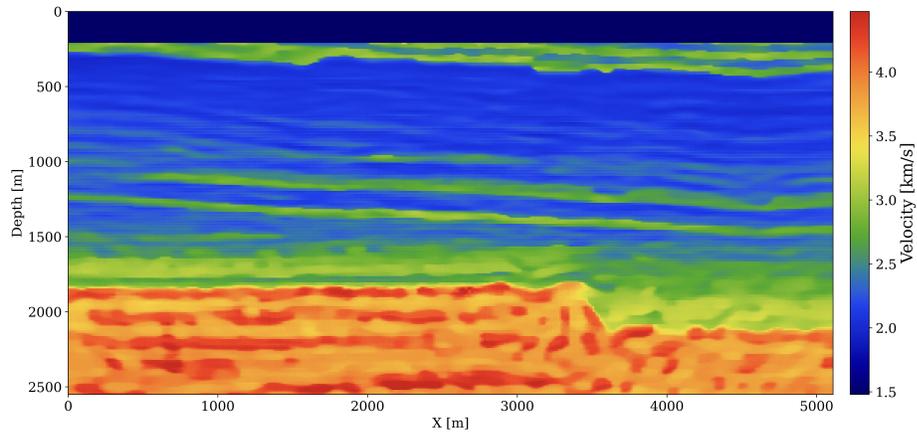
As stated earlier, our goal is to estimate the permeability from time-lapse crosswell monitoring data collected at a CO₂ injection site (cf. Figure 2.2). Compared to conventional seismic imaging, time-lapse monitoring of geological storage differs because it aims to image time-lapse changes in the CO₂ plume while obtaining estimates for the reservoir’s fluid-flow properties. This involves coupling wave modeling operators to fluid-flow physics to track the CO₂ plumes underground. The fluid-flow physics models the slow process of CO₂ partly replacing brine in the pore space of the reservoir, which involves solving the



(a)



(b)



(c)

Figure 2.6: Results from using our normalizing flow learned prior in FWI. (a) Ground truth. (b) Traditional FWI without prior resulting in 12.98 dB PSNR. (c) Our FWI result with learned prior resulting in 14.77 dB PSNR.

multiphase flow equations. For this purpose, we need access to reservoir simulation software capable of modeling two-phase (brine/CO₂) flow. While a number of proprietary and open-source reservoir simulators exist, including MRST [61], GEOSX [62] and Open Porous Media (OPM) [63], few support differentiation of the simulator’s output (CO₂ saturation) with respect to its input (the spatial permeability distribution \mathbf{K} in Figure 2.1). We use the recently developed external Julia package `JutulDarcy.jl` that supports Darcy flow and serves as a front-end to `Jutul.jl` [33], which provides accurate Jacobians with respect to \mathbf{K} . `Jutul.jl` is an implicit solver for finite-volume discretizations that internally uses AD to calculate the Jacobian. It has a performance and feature set comparable to commercial multiphase flow simulators and accounts for realistic effects (e.g., dissolution, inter-phase mass exchange, compressibility, capillary effects) and residual trapping mechanisms. It also provides accurate sensitivities through an adjoint formulation of the subsurface multiphase flow equations. To integrate the Jacobian of this software package into Julia’s differentiable programming system, we wrote the light “wrapper package” `JutulDarcyRules.jl` [64] that adds an `rrule` for the nonlinear operator $\mathcal{S}(\mathbf{K})$, which maps the permeability distribution, \mathbf{K} , to the spatially-varying CO₂ concentration snapshots, $\mathbf{c} = \{\mathbf{c}^i\}_{i=1}^{n_v}$, over n_v monitoring time-steps (cf. Figure 2.1). Addition of this `rrule` allows these packages to interoperate with other packages in Julia’s AD ecosystem. Below, we show a basic example where ADAM algorithm is used to invert for subsurface permeability given the full history of CO₂ concentration snapshots:

```
# Generate CO2 concentration.
c = S(K_true)

# Set up ADAM optimizer.
opt = ADAM()

# Define the loss function.
loss(K) = .5f0 * norm(S(K) - c)^2f0

# ADAM iterations.
for it = 1:maxiter
```

```

g = gradient(loss, K)[1]      # gradient computed with AD
update!(opt, K, g)           # update K with ADAM
end

```

During each iteration of the loop above, Julia’s machine learning package Flux.jl [65, 66] uses the custom gradient defined by the aforementioned `rrule`, calling the high-performance adjoint code from JutulDarcy.jl. Our adaptable software framework also facilitates effortless substitution of deep learning models in lieu of the numerical fluid-flow simulator. In the next section, we introduce distributed Fourier neural operators (dfno) and discuss how this neural surrogate contributes to our inversion framework.

2.4.4 Fourier neural operator surrogates

While the integration of multiphase flow modeling into Julia differentiable programming ecosystem opens the way to carry out end-to-end inversions (as explained below), fluid-flow simulations are computationally expensive—a notion compounded by the fact that these simulations have to be done many times during inversion. For this reason, we switch to a data-driven approach where a neural operator is first trained on simulation examples, pairs $\{\mathbf{K}, \mathcal{S}(\mathbf{K})\}$, to learn the mapping from permeability models, \mathbf{K} , to the corresponding CO₂ snapshots, $\mathbf{c} := \mathcal{S}(\mathbf{K})$. After incurring initial offline training costs, this neural surrogate provides a fast alternative to numerical solvers with acceptable accuracy. Fourier neural operators [FNOs, 30], a recently introduced neural network architecture based on spectral convolutions that capture the long range correlations rather than localized spatial convolutions, has recently been introduced as a surrogate for elliptic partial differential equations such as the Darcy or Burgers equation. This spectral architecture has been applied successfully to simulate two-phase flow during geological CO₂ storage projects [67]. Independently, [68] used a trained FNO to replace the fluid-flow simulations as part of end-to-end inversion and showed that AD of Julia’s machine learning package can be used to compute gradients with respect to the permeability using Flux.jl’s reverse-mode AD sys-

tem Zygote.jl [69]. After training, the above permeability inversion from concentration snapshots, \mathbf{c} , is carried out by simply replacing \mathcal{S} by $\mathcal{S}_{\mathbf{w}^*}$ with \mathbf{w}^* being the weights of the pretrained FNO. Thanks to the AD system, the gradient with respect to \mathbf{K} is computed automatically. Thus, after loading the trained FNO and redefining the operator S , the above code remains exactly the same. For implementation details on the FNO and its training, we refer to [68] and [31].

2.5 Putting it all together

As a final step in our end-to-end permeability inversion, we introduce a nonlinear rock physics model, denoted by \mathcal{R} . Based on the patchy saturation model [70], this model nonlinearly maps the time-lapse CO_2 saturations to decreases in the seismic properties (compressional wavespeeds, $\mathbf{v} = \{\mathbf{v}^i\}_{i=1}^{n_v}$) within the reservoir with the Julia code

```
# Patchy saturation function.
# Input: CO2 saturation, velocity, density, porosity.
# Optional: bulk modulus of mineral, brine, CO2; density of CO2, brine.
# Output: velocity, density.
function Patchy(sw, vp, rho, phi;
    bulk_min=36.6f9, bulk_fl1=2.735f9, bulk_fl2=0.125f9,
    rhow=7f2, rho0=1f3) where T
    # Relate vp to vs, set modulus properties.
    vs = vp ./ sqrt(3f0)
    bulk_sat1 = rho .* (vp.^2f0 .- 4f0/3f0 .* vs.^2f0)
    shear_sat1 = rho .* (vs.^2f0)
    # Calculate bulk modulus if filled with 100% CO2.
    patch_temp = bulk_sat1 ./ (bulk_min .- bulk_sat1)
        .- bulk_fl1 ./ phi ./ (bulk_min .- bulk_fl1)
        .+ bulk_fl2 ./ phi ./ (bulk_min .- bulk_fl2)
    bulk_sat2 = bulk_min ./ (1f0 ./ patch_temp .+ 1f0)
    # Calculate new bulk modulus as weighted harmonic average.
    bulk_new = 1f0 / ((1f0 .- sw) ./ (bulk_sat1 .+ 4f0/3f0 * shear_sat1))
end
```

```

+ sw ./ (bulk_sat2 + 4f0/3f0 * shear_sat1) - 4f0/3f0 * shear_sat1
# Calculate new density and velocity.
rho_new = rho + phi .* sw * (rhow - rhoo)
vp_new = sqrt.((bulk_new ./ 4f0/3f0 * shear_sat1) ./ rho_new)
return vp_new, rho_new
end

```

We map the changes in the wavespeeds to time-lapse seismic data, $\mathbf{d} = \{\mathbf{d}^i\}_{i=1}^{n_v}$, via the blockdiagonal seismic modeling² operator $\mathcal{F}(\mathbf{v}) = \text{diag}(\{\mathbf{F}^i(\mathbf{v}^i)\mathbf{q}^i\}_{i=1}^{n_v})$. In this formulation, the single vintage forward operators \mathbf{F}^i and corresponding sources, \mathbf{q}^i , are allowed to vary between vintages.

With the fluid-flow (surrogate) solver, \mathcal{S} , the rock physics module, \mathcal{R} , and wave physics module, \mathcal{F} , in place, along with regularization via reparametrization using \mathcal{G}_{θ^*} , we are now in a position to formulate the desired end-to-end inversion problem as

$$\underset{\mathbf{z}}{\text{minimize}} \quad \frac{1}{2} \|\mathcal{F} \circ \mathcal{R} \circ \mathcal{S}(\mathcal{G}_{\theta^*}(\mathbf{z})) - \mathbf{d}\|_2^2 + \frac{\lambda}{2} \|\mathbf{z}\|_2^2, \quad (2.3)$$

where the inverted permeability can be calculated by $\mathbf{K}^* = \mathcal{G}_{\theta^*}(\mathbf{z}^*)$ with \mathbf{z}^* the latent space minimizer of 2.3. As illustrated in Figure 2.1, we obtain the nonlinear end-to-end map by composing the fluid-flow, rock, and wave physics, according to $\mathcal{F} \circ \mathcal{R} \circ \mathcal{S}$. The corresponding Julia code reads

```

# Set up ADAM optimizer.
opt = ADAM()
# Define the reparameterized loss function including penalty term.
loss(z) = .5f0 * norm(F ∘ R ∘ S(G(z)) - d)^2f0 + .5f0 * lambda * norm(z)^2f0
# ADAM iterations.
for it = 1:maxiter
    g = gradient(loss, z)[1] # gradient computed by AD
    update!(opt, z, g) # update z by ADAM
end

```

²Note, we parameterized this forward modeling in terms of the compressional wavespeed.

```

end
# Convert latent variable to permeability.
K = G(z)

```

This end-to-end inversion procedure, which utilizes a learned deep prior and a pre-trained FNO surrogate, was successfully employed by [68] on a simple stylistic blocky high-low permeability model. The procedure involves using AD, with `rrule` for the wave and fluid physics, in combination with innate AD capabilities to compute the gradient of the objective in 2.3, which incorporates fluid-flow, rock, and wave physics. Below, we share early results from applying the proposed end-to-end inversion in a more realistic setting derived from real data (cf. Figure 2.2).

2.6 Preliminary inversion results

While initial results by [68] were encouraging and showed strong benefits from the learned prior, the permeability model and fluid flow simulations considered in their study were too simplistic. To evaluate the proposed end-to-end inversion methodology in a more realistic setting, we consider the permeability model plotted in Figure 2.7a, which we derived from a slice of the Compass model [59] shown in Figure 2.2. To generate realistic CO₂ plumes in this model, we generate immiscible and compressible two-phase flow simulations with `JutulDarcy.jl` over a period of 18 years with 5 snapshots plotted at years 10, 15, 16, 17, and 18. These CO₂ snapshots are shown in the first row of Figure 2.8. Next, given the fluid-flow simulation, we use the patchy saturation model [70] to convert each CO₂ concentration snapshot, \mathbf{c}^i , $i = 1 \dots n_v$ to corresponding wavespeed model, \mathbf{v}^i , $i = 1 \dots n_v$ with $\mathbf{v} = \mathcal{R}(\mathbf{c})$. We then use `JUDI.jl` to generate synthetic time-lapse data, \mathbf{d}^i , $i = 1 \dots n_v$, for each vintage.

During the inversion, the first 15 years of time-lapse data, \mathbf{d}^i , $i = 1 \dots 15$, from the above synthetic experiment are inverted with permeabilities within the reservoir initialized by a single reasonable value as shown in Figure 2.7b. Inversion results obtained after

25 passes through the data for the physics-driven two-phase flow solver and its learned neural surrogate approximation are included in Figure 2.7c and Figure 2.7d, respectively. Both results were obtained with 200 iterations of the code block shown above. Each time-lapse vintage consist of 960 receivers and 32 shots. To limit the number of wave-equation solves, gradients were calculated for only four randomly selected shots with replacement per iteration. While these results obtained without learned regularization are somewhat preliminary, they lead to the following observations. First, both inversion results for the permeability follow the inverted cone shape of the CO₂. This is to be expected because permeability can only be inverted where CO₂ has flown over the first 15 years. Second, the inverted permeability follows trends of this strongly heterogeneous model. Third, as expected details and continuity of the results obtained with the two-phase flow solver are better. In part, this can be explained by the fact that there are no guarantees that the model iterations remain with the statistical distribution on which the FNO was trained. Fourth, the implementation of this workflow greatly benefited from the software design principles listed above. For instance, the use of abstractions made it trivial to replace physics-driven two-phase flow solvers with their learned counterparts.

Despite being preliminary, the inversion results show that this framework is conducive to produce current CO₂ plume estimates and near-future forecasts. As described by [68], these capabilities can be achieved through use of the physics simulator or the trained FNO surrogate. The 18 year CO₂ simulations in both inverted permeability models are reasonable when comparing the true plume development plotted in the top row of Figure 2.8 with plumes simulated from the inverted models plotted in rows three and four of Figure 2.8. While certain details are missing in the estimates for the past, current, and predicted CO₂ concentrations, the inversion constitutes a considerable improvement compared to plumes generated in the starting model for the permeability plotted in the second row of 2.8.

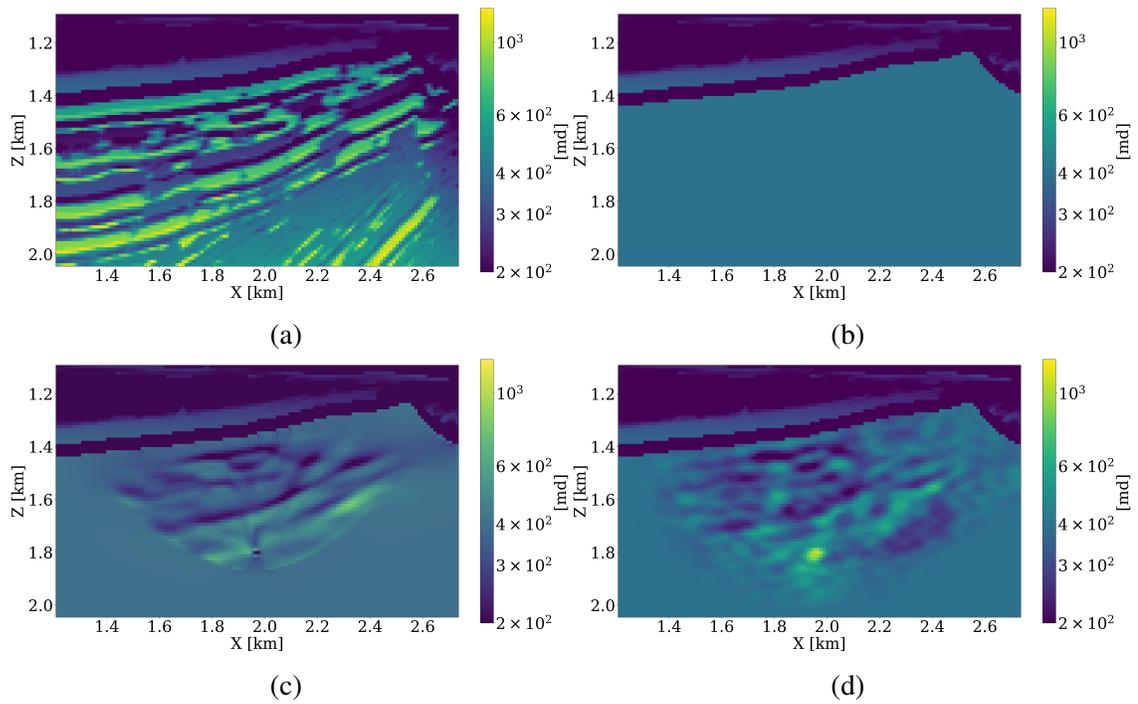


Figure 2.7: Fifteen-year time-lapse seismic end-to-end permeability inversion with physics-based and surrogate fluid-flow simulations. (a) Ground truth permeability. (b) Initial permeability with homogeneous values in the reservoir. (c) Inverted permeability from physics-based inversion. (d) Inverted permeability with neural surrogate approximation.

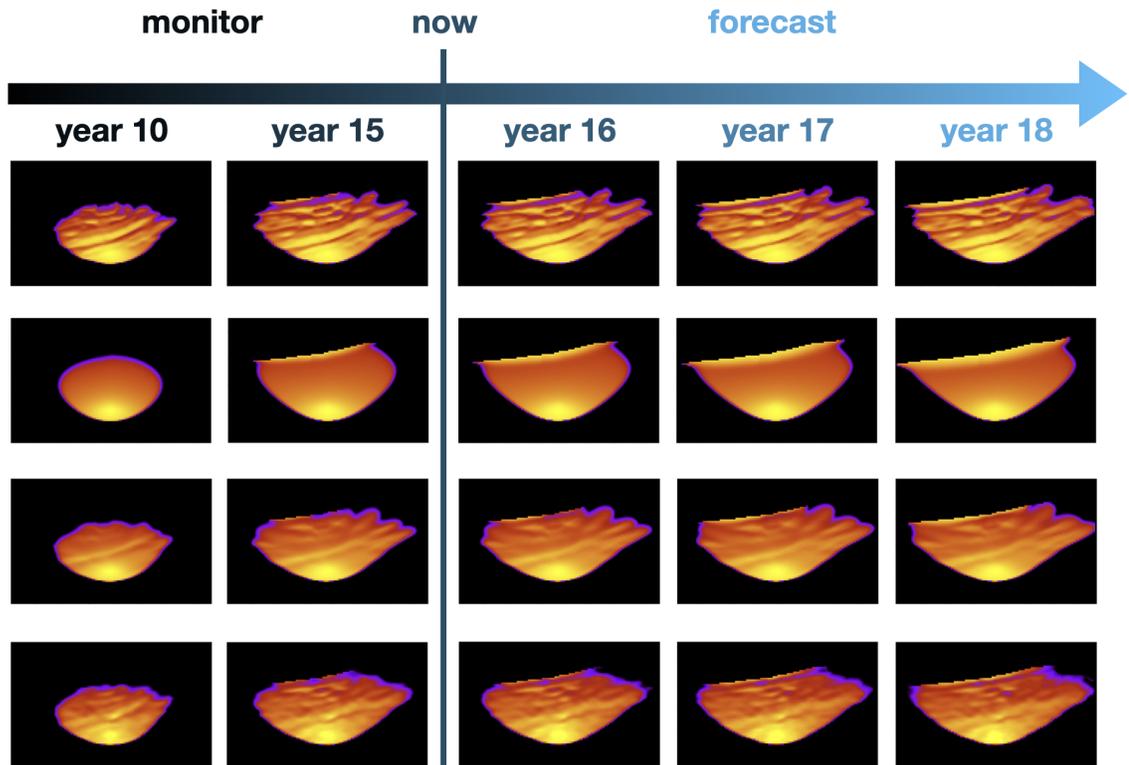


Figure 2.8: CO₂ plume estimation and prediction. The first two columns are the CO₂ concentration snapshots at year 10 and year 15 of the first 15 years of simulation monitored seismically. The last three columns are forecasted snapshots at years 16, 17, 18, where no seismic data is available. First row corresponds to the ground truth CO₂ plume simulated by the unseen ground truth permeability model. Second row contains plume simulations in the starting model, with a 10.99 dB SNR on the first 15 years of CO₂ snapshots and a 8.51 dB on the last 3 years. Rows three and four contain estimated and predicted CO₂ plumes for the physics-based and surrogate-based permeability inversion. The SNR values of the first 15 years of the estimated CO₂ plume are 17.72 dB and 16.17 dB for the physics-based inversion and the surrogate-based inversion, respectively. The SNR values for the CO₂ plume forecasts for the last 3 years are 15.69 dB and 14.05 dB for the physics-based inversion and the surrogate-based inversion.

2.7 Remaining challenges

The numerical case studies shown above illustrate that working with abstractions certainly has its benefits. Thanks to the math-inspired abstractions, which naturally lead to modularity and separation of concerns, we were able to accelerate the research and development cycle for the end-to-end inversion. As a result, we created a development environment that allowed us to include state-of-the-art industry-strength simulators and machine learning techniques. What we unfortunately not yet have been able to do is to demonstrate our ability to scale this end-to-end inversion to 3D, while both the Devito-based propagators and Jutul.jl’s fluid-flow simulations both have been demonstrated on industry-scale problems. Unfortunately, lack of access to large-scale computational resources makes it challenging in an academic environment to validate the proposed methodology on 4D synthetic and field data even though the computational toolchain presented in this chapter is fully differentiable and in principle capable of scale-up. Most components have been separately tested and verified on realistic 3D examples [31, 71, 72, 73] and efforts are underway to remove fundamental memory and other bottlenecks.

2.7.1 Scale-up normalizing flows

Generative models, and NFs included, call for relatively large training sets and large computational resources for training. While efforts have been made to create training sets for the more traditional machine learning tasks, no public-domain training set exists that contains realistic 3D examples. The good news is that normalizing flows [29] have a small memory footprint compared to diffusion models [74], so training this type of network will be feasible when training sets and compute become available. Concurrent studies have already shown the success to train and evaluate NFs on 3D models in size of $128 \times 128 \times 128$ [75]. In some cases where geophysicists might not have enough samples for velocity/permeability models, one could use in-house legacy models to train the NFs as a preparation

step for inverting the seismic data. We leave the potential investigation to future studies.

2.7.2 Scale-up neural operators

Since the seminal paper by [30], there has been a flurry of publications on the use of FNOs as neural surrogates for expensive multiphase fluid-flow solvers used to simulate CO₂ injection as part of geological storage projects [67, 76]. While there is good reason for this excitement, challenges remain when scaling this technique to realistic 3D problems. In that case, additional measures have to be taken. For instance, by nesting FNOs [76] were able to divide 3D domains into smaller hierarchical subdomains centered around the wells, an approach that is only viable when certain assumptions are met. Because of this nested decomposition, these authors avoid the large memory footprint of 3D FNOs and report many orders of magnitude speedup. Given the potential impact of irregular CO₂ flow, e.g., leakage, we as much as possible try to avoid making assumptions on the flow behavior and propose an accurate distributed Fourier neural operator (dfno) structure based on a domain decomposition of the network’s input and network weights [31]. By using DistDL [77], a software package that supports “model parallelism” in machine learning, our dfno partitions the input data and network weights across multiple GPUs such that each partition is able to fit in the memory of a single GPU. As reported by [31], our work demonstrated validity of dfno on realistic problem and reasonable training set (permeability/CO₂ concentration pairs) sizes, for permeability models derived from the Sleipner benchmark model [78]. On 16 timesteps and models of size $64 \times 118 \times 263$, we reported from our perspective a more realistic speedup of over $1300\times$ compared to the simulation time on Open Porous Media [63], one of the leading open-source reservoir simulators. These results confirm a similar independent approach advocated by [79].

2.8 Conclusions

In this chapter, we introduced a software framework for geophysical inverse problems and machine learning that provides a scalable, portable, and interoperable environment for research and development at scale. We showed that through carefully chosen design principles, software with math-inspired abstractions can be created that naturally leads to desired modularity and separation of concerns without sacrificing performance. We achieve this by combining Devito's automatic code generation for wave propagators with Julia's modern highly performant and scalable programming capabilities, including differentiable programming. Thanks to these features, we were able to quickly implement a prototype, in principle scalable to 3D, for permeability inversion from time-lapse crosswell seismic data. Aside from the use of proper abstractions, our approach to solving this relatively complex multiphysics problem extensively relied on Julia's innate algorithmic differentiation capabilities, supplemented by auxiliary performant derivatives for the wave/fluid-flow physics, and for components of the machine learning. On account of these design choices, we developed an agile and relatively easy to maintain compact software stack where low-level code is hidden through a combination of math-inspired abstractions, modern programming practices, and automatic code generation.

2.9 References

- [1] D. Sun and W. W. Symes, “Iwave implementation of adjoint state method,” Tech. Rep. 10-06, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA, Tech. Rep., 2010 (page 20).
- [2] P. A. Witte *et al.*, “A large-scale framework for symbolic implementations of seismic inversion algorithms in julia,” *GEOPHYSICS*, vol. 84, no. 3, F57–F71, 2019. eprint: <https://doi.org/10.1190/geo2018-0174.1> (pages 20, 22).
- [3] J. Washbourne *et al.*, “Chevron optimization framework for imaging and inversion (cofi) – an open source and cloud friendly julia language framework for seismic modeling and inversion,” in *First International Meeting for Applied Geoscience & Energy Expanded Abstracts*, 2021, pp. 792–796. eprint: <https://library.seg.org/doi/pdf/10.1190/segam2021-3594362.1> (page 20).
- [4] A. Richardson, *Seismic full-waveform inversion using deep learning tools and techniques*, 2018. arXiv: 1801.07232 [physics.geo-ph] (page 21).
- [5] A. Paszke *et al.*, “Pytorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., 2019, pp. 8024–8035 (page 21).
- [6] G. van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009, ISBN: 1441412697 (page 22).
- [7] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, “Julia: A fresh approach to numerical computing,” *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017 (page 22).
- [8] A. D. Padula, S. D. Scott, and W. W. Symes, “A software framework for abstract expression of coordinate-free linear algebra and optimization algorithms,” *ACM Trans. Math. Softw.*, vol. 36, no. 2, Apr. 2009 (page 22).
- [9] F. Rathgeber *et al.*, “Firedrake: Automating the finite element method by composing abstractions,” *ACM Trans. Math. Softw.*, vol. 43, no. 3, Dec. 2016 (page 22).
- [10] M. S. Alnaes *et al.*, “The FEniCS project version 1.5,” *Archive of Numerical Software*, vol. 3, 2015 (page 22).
- [11] S. Fomel, P. Sava, I. Vlad, Y. Liu, and V. Bashkardin, “Madagascar: Open-source software project for multidimensional data analysis and reproducible computational experiments,” *Journal of Open Research Software*, vol. 1, no. 1, 2013 (page 22).

- [12] A. Meurer *et al.*, “SymPy: Symbolic computing in python,” *PeerJ Computer Science*, vol. 3, e103, Jan. 2017 (page 22).
- [13] E. van den Berg and M. P. Friedlander, “Spot: A linear-operator toolbox for matlab,” in *SCAIM*, SCAIM Seminar, University of British Columbia: SCAIM Seminar, 2009 (page 22).
- [14] H. Modzelewski, M. Louboutin, Z. Yin, D. Karrasch, and R. Orozco, *Slimgroup/joli.jl: V0.8.5*, version v0.8.5, Mar. 2023 (page 22).
- [15] M. Ravasi and I. Vasconcelos, “Pylops—a linear-operator python library for scalable algebra and optimization,” *SoftwareX*, vol. 11, p. 100 361, 2020 (page 22).
- [16] M. Louboutin, Z. Yin, and F. J. Herrmann, *Slimgroup/slimoptim.jl: V0.2.0*, version v0.2.0, Aug. 2022 (page 22).
- [17] B. Peters and F. J. Herrmann, “Algorithms and software for projections onto intersections of convex and non-convex sets with applications to inverse problems,” *arXiv preprint arXiv:1902.09699*, 2019 (page 23).
- [18] B. Peters, M. Louboutin, and H. Modzelewski, *Slimgroup/setintersectionprojection.jl: V0.2.4*, version v0.2.4, Oct. 2022 (page 23).
- [19] K. Lensink, H. Modzelewski, M. Louboutin, yzhang3198, and Z. Yin, *Slimgroup/s-egyio.jl: V0.8.3*, version v0.8.3, Jan. 2023 (page 23).
- [20] M. Louboutin, Z. Yin, and F. J. Herrmann, *Slimgroup/judi4cloud.jl: First public release*, version v0.2.1, Mar. 2022 (page 23).
- [21] H. Modzelewski and M. Louboutin, *Slimgroup/cloudsegyio.jl: V1.0.1*, version v1.0.1, Dec. 2022 (page 23).
- [22] M. Louboutin *et al.*, “Devito (v3.1.0): An embedded domain-specific language for finite differences and geophysical exploration,” *Geoscientific Model Development*, vol. 12, no. 3, pp. 1165–1187, 2019 (page 23).
- [23] F. Luporini *et al.*, “Architecture and performance of devito, a system for automated stencil computation,” *ACM Trans. Math. Softw.*, vol. 46, no. 1, Apr. 2020 (page 23).
- [24] M. Innes *et al.*, “A differentiable programming system to bridge machine learning and scientific computing,” *arXiv preprint arXiv:1907.07587*, 2019 (page 23).
- [25] D. Li, K. Xu, J. M. Harris, and E. Darve, “Coupled time-lapse full-waveform inversion for subsurface flow problems using intrusive automatic differentiation,” *Water Resources Research*, vol. 56, no. 8, e2019WR027032, 2020 (page 23).

- [26] F. C. White *et al.*, *Juliadiff/chainrulescore.jl: V1.15.6*, version v1.15.6, Sep. 2022 (page 23).
- [27] F. C. White *et al.*, *Juliadiff/chainrules.jl: V1.47.0*, version v1.47.0, Feb. 2023 (page 23).
- [28] P. Witte *et al.*, *Slimgroup/invertiblenetworks.jl: V2.2.5*, version v2.2.5, Apr. 2023 (page 23).
- [29] D. Rezende and S. Mohamed, “Variational inference with normalizing flows,” in *International conference on machine learning*, PMLR, 2015, pp. 1530–1538 (pages 23, 41).
- [30] Z. Li *et al.*, *Fourier neural operator for parametric partial differential equations*, 2020. arXiv: 2010.08895 [cs.LG] (pages 23, 34, 42).
- [31] T. J. Grady *et al.*, “Model-parallel fourier neural operators as learned surrogates for large-scale parametric pdes,” *Computers & Geosciences*, vol. 178, p. 105 402, 2023 (pages 23, 35, 41, 42).
- [32] T. J. Grady, Infinoid, and M. Louboutin, *Slimgroup/dfno: Optimal comm*, version 0.3, Aug. 2022 (page 23).
- [33] O. Møyner, M. Johnsrud, H. M. Nilsen, X. Raynaud, K. O. Lye, and Z. Yin, *Sintef-math/jutul.jl: V0.2.6*, version v0.2.6, Apr. 2023 (pages 24, 33).
- [34] P. Ringrose, *How to store CO₂ underground: Insights from early-mover CCS Projects*. Springer, 2020 (page 25).
- [35] P. A. Witte, M. Louboutin, F. Luporini, G. J. Gorman, and F. J. Herrmann, “Compressive least-squares migration with on-the-fly fourier transforms,” *Geophysics*, vol. 84, no. 5, R655–R672, 2019 (page 26).
- [36] M. Louboutin, P. Witte, and F. J. Herrmann, “Effects of wrong adjoints for rtm in tti media,” in *SEG Technical Program Expanded Abstracts 2018*. 2018, pp. 331–335. eprint: <https://library.seg.org/doi/pdf/10.1190/segam2018-2996274.1> (page 26).
- [37] M. Louboutin and F. Herrmann, “Extending the search space of time-domain adjoint-state fwi with randomized implicit time shifts,” vol. 2017, no. 1, pp. 1–5, 2017 (page 26).
- [38] F. J. Herrmann, A. Siahkoohi, and G. Rizzuti, “Learned imaging with constraints and uncertainty quantification,” in *Neural Information Processing Systems (NeurIPS) 2019 Deep Inverse Workshop*, Dec. 2019 (page 26).

- [39] M. Yang, Z. Fang, P. A. Witte, and F. J. Herrmann, “Time-domain sparsity promoting least-squares reverse time migration with source estimation,” *Geophysical Prospecting*, vol. 68, no. 9, pp. 2697–2711, Aug. 2020, (Geophysical Prospecting) (page 26).
- [40] G. Rizzuti, A. Siahkoohi, P. A. Witte, and F. J. Herrmann, “Parameterizing uncertainty by deep invertible networks, an application to reservoir characterization,” in *90th Annual International Meeting*, Society of Exploration Geophysicists, Expanded Abstracts, Sep. 2020, pp. 1541–1545 (page 26).
- [41] G. Rizzuti, M. Louboutin, R. Wang, and F. J. Herrmann, “A dual formulation of wavefield reconstruction inversion for large-scale seismic inversion,” *Geophysics*, vol. 86, no. 6, 1ND–Z3, Oct. 2021, (Geophysics) (page 26).
- [42] A. Siahkoohi, G. Rizzuti, and F. J. Herrmann, “A deep-learning based bayesian approach to seismic imaging and uncertainty quantification,” in *EAGE Annual Conference Proceedings*, Accepted in EAGE, Jan. 2020 (page 26).
- [43] A. Siahkoohi, G. Rizzuti, and F. J. Herrmann, “Uncertainty quantification in imaging and automatic horizon tracking—a bayesian deep-prior based approach,” in *SEG Technical Program Expanded Abstracts*, (SEG, virtual), Sep. 2020, pp. 1636–1640 (page 26).
- [44] A. Siahkoohi, G. Rizzuti, and F. J. Herrmann, “Weak deep priors for seismic imaging,” in *SEG Technical Program Expanded Abstracts 2020*, Society of Exploration Geophysicists, 2020, pp. 2998–3002 (page 26).
- [45] Z. Yin, M. Louboutin, and F. J. Herrmann, “Compressive time-lapse seismic monitoring of carbon storage and sequestration with the joint recovery model,” in *SEG International Exposition and Annual Meeting*, SEG, 2021, D011S145R001 (page 26).
- [46] Z. Yin, H. T. Erdinc, A. P. Gahlot, M. Louboutin, and F. J. Herrmann, “Derisking geologic carbon storage from high-resolution time-lapse seismic to explainable leakage detection,” *The Leading Edge*, vol. 42, no. 1, pp. 69–76, 2023 (page 26).
- [47] Z. Yin, R. Orozco, P. Witte, M. Louboutin, G. Rizzuti, and F. J. Herrmann, “Extended source imaging—a unifying framework for seismic and medical imaging,” in *SEG Technical Program Expanded Abstracts 2020*, Society of Exploration Geophysicists, 2020, pp. 3502–3506 (page 26).
- [48] R. Orozco, A. Siahkoohi, G. Rizzuti, T. van Leeuwen, and F. J. Herrmann, “Photoacoustic imaging with conditional priors from normalizing flows,” in *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems*, Dec. 2021 (page 26).

- [49] R. Orozco, A. Siahkoohi, G. Rizzuti, T. van Leeuwen, and F. Herrmann, “Adjoint operators enable fast and amortized machine learning based bayesian uncertainty quantification,” *Medical Imaging 2023: Image Processing*, I. Išgum and O. Colliot, Eds., Apr. 3, 2023 (page 26).
- [50] R. Orozco, M. Louboutin, A. Siahkoohi, G. Rizzuti, T. van Leeuwen, and F. J. Herrmann, “Amortized normalizing flows for transcranial ultrasound with uncertainty quantification,” Mar. 2023 (page 26).
- [51] L. Dinh, J. Sohl-Dickstein, and S. Bengio, “Density estimation using real nvp,” *arXiv preprint arXiv:1605.08803*, 2016 (page 28).
- [52] X. Zhang and A. Curtis, “Seismic tomography using variational inference methods,” *Journal of Geophysical Research: Solid Earth*, vol. 125, no. 4, e2019JB018589, 2020 (page 28).
- [53] X. Zhao, A. Curtis, and X. Zhang, “Bayesian seismic tomography using normalizing flows,” *Geophysical Journal International*, vol. 228, no. 1, pp. 213–239, Jul. 2021. eprint: <https://academic.oup.com/gji/article-pdf/228/1/213/40348424/ggab298.pdf> (page 28).
- [54] X. Zhang and A. Curtis, “Bayesian geophysical inversion using invertible neural networks,” *Journal of Geophysical Research: Solid Earth*, vol. 126, no. 7, e2021JB022320, 2021 (page 28).
- [55] A. Siahkoohi and F. J. Herrmann, “Learning by example: Fast reliability-aware seismic imaging with normalizing flows,” in *First International Meeting for Applied Geoscience & Energy*, Society of Exploration Geophysicists, Expanded Abstracts, Sep. 2021, pp. 1580–1585 (page 28).
- [56] A. Siahkoohi, G. Rizzuti, M. Louboutin, P. Witte, and F. J. Herrmann, “Preconditioned training of normalizing flows for variational inference in inverse problems,” in *3rd Symposium on Advances in Approximate Bayesian Inference*, Jan. 2021 (page 28).
- [57] A. Siahkoohi, G. Rizzuti, and F. J. Herrmann, “Deep Bayesian inference for seismic imaging with tasks,” *Geophysics*, vol. 87, no. 5, S281–S302, Sep. 2022 (page 28).
- [58] A. Siahkoohi, G. Rizzuti, R. Orozco, and F. J. Herrmann, “Reliable amortized variational inference with physics-based latent distribution correction,” *Geophysics*, vol. 88, no. 3, R297–R322, 2023 (page 28).
- [59] C. Jones, J. Edgar, J. Selvage, and H. Crook, “Building complex synthetic models to evaluate acquisition geometries and velocity inversion technologies,” in *74th EAGE*

Conference and Exhibition incorporating EUROPEC 2012, European Association of Geoscientists & Engineers, 2012, cp–293 (pages 28, 37).

- [60] M. Asim, M. Daniels, O. Leong, A. Ahmed, and P. Hand, “Invertible generative models for inverse problems: Mitigating representation error and dataset bias,” in *International conference on machine learning*, PMLR, 2020, pp. 399–409 (page 30).
- [61] K.-A. Lie and O. Møyner, *Advanced modelling with the MATLAB reservoir simulation toolbox*. Cambridge University Press, 2021 (page 33).
- [62] H. Gross and A. Mazuyer, “Geosx: A multiphysics, multilevel simulator designed for exascale computing,” in *SPE Reservoir Simulation Conference*, OnePetro, 2021 (page 33).
- [63] A. F. Rasmussen *et al.*, “The open porous media flow reservoir simulator,” *Computers & Mathematics with Applications*, vol. 81, pp. 159–185, 2021 (pages 33, 42).
- [64] Z. Yin, G. Bruer, and M. Louboutin, *Slimgroup/jutuldarcyrules.jl: V0.2.5*, version v0.2.5, Apr. 2023 (page 33).
- [65] M. Innes *et al.*, “Fashionable modelling with flux,” *CoRR*, vol. abs/1811.01457, 2018. arXiv: 1811.01457 (page 34).
- [66] M. Innes, “Flux: Elegant machine learning with julia,” *Journal of Open Source Software*, vol. 3, no. 25, p. 602, 2018 (page 34).
- [67] G. Wen, Z. Li, K. Azizzadenesheli, A. Anandkumar, and S. M. Benson, “U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow,” *Advances in Water Resources*, vol. 163, p. 104 180, 2022 (pages 34, 42).
- [68] Z. Yin, A. Siahkoohi, M. Louboutin, and F. J. Herrmann, “Learned coupled inversion for carbon sequestration monitoring and forecasting with fourier neural operators,” in *International Meeting for Applied Geoscience and Energy Expanded Abstracts*, (IMAGE, Houston), May 2022 (pages 34, 35, 37, 38).
- [69] M. Innes, “Don’t unroll adjoint: Differentiating ssa-form programs,” *CoRR*, vol. abs/1810.07951, 2018. arXiv: 1810.07951 (page 35).
- [70] P. Avseth, T. Mukerji, and G. Mavko, *Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk*. Cambridge university press, 2010 (pages 35, 37).
- [71] O. Møyner, G. Bruer, and Z. Yin, *Sintefmath/jutuldarcy.jl: V0.2.3*, version v0.2.3, Apr. 2023 (page 41).

- [72] M. Louboutin and F. J. Herrmann, “Enabling wave-based inversion on gpus with randomized trace estimation,” in *83rd EAGE Annual Conference & Exhibition*, EAGE Publications BV, vol. 2022, 2022, pp. 1–5 (page 41).
- [73] M. Louboutin and F. J. Herrmann, “Wave-based inversion at scale on graphical processing units with randomized trace estimation,” *Geophysical Prospecting*, vol. 72, no. 2, pp. 353–366, 2024 (page 41).
- [74] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, “Score-based generative modeling through stochastic differential equations,” *arXiv preprint arXiv:2011.13456*, 2020 (page 41).
- [75] R. Orozco, A. Siahkoohi, M. Louboutin, and F. J. Herrmann, “Aspire: Iterative amortized posterior inference for bayesian inverse problems,” *arXiv preprint arXiv:2405.05398*, 2024 (page 41).
- [76] G. Wen, Z. Li, Q. Long, K. Azizzadenesheli, A. Anandkumar, and S. Benson, “Real-time high-resolution co2 geological storage prediction using nested fourier neural operators,” *Energy & Environmental Science*, 2023 (page 42).
- [77] R. J. Hewett and T. J. Grady II, “A linear algebraic approach to model parallelism in deep learning,” *arXiv preprint arXiv:2006.03108*, 2020 (page 42).
- [78] A.-K. Furre, O. Eiken, H. Alnes, J. N. Vevatne, and A. F. Kiær, “20 years of monitoring CO₂-injection at sleipner,” *Energy Procedia*, vol. 114, pp. 3916–3926, 2017, 13th International Conference on Greenhouse Gas Control Technologies, GHGT-13, 14-18 November 2016, Lausanne, Switzerland (page 42).
- [79] P. A. Witte, R. J. Hewett, K. Saurabh, A. Sojoodi, and R. Chandra, “Sciai4industry—solving pdes for industry-scale problems with deep learning,” *arXiv preprint arXiv:2211.12709*, 2022 (page 42).

CHAPTER 3

TIME-LAPSE FULL-WAVEFORM PERMEABILITY INVERSION: A FEASIBILITY STUDY

3.1 Summary

Time-lapse seismic monitoring necessitates integrated workflows that combine seismic and reservoir modeling to enhance reservoir property estimation. We present a feasibility study of an end-to-end inversion framework that directly inverts for permeability from multiple prestack time-lapse seismic datasets. To assess the method’s robustness, we design experiments focusing on its sensitivity to initial models and potential errors in modeling. Our study leverages the publicly available Compass model to simulate CO₂ storage in saline aquifers. This model is derived from well and seismic data from the North Sea in an area that is currently considered for geological carbon storage.

3.2 Introduction

Despite significant advancements in reservoir monitoring over recent decades, time-lapse seismic technology continues to face challenges related to cost and efficiency [1, 2, 3, 4]. Employing 4D seismic workflows, including time-lapse full-waveform inversion (TL-FWI) [5, 6], has become a common practice for estimating changes in the Earth’s elastic properties, facilitating the quantitative interpretation of these changes as indicators of reservoir attributes like fluid content and pressure [7, 8]. Recent methodologies aim to leverage time-lapse seismic data for the joint estimation of both elastic and reservoir properties, with a focus on parameters such as saturation and porosity [9, 10]. However, the integration of seismic imaging workflows with reservoir simulation tools remains limited, constraining the direct application of time-lapse seismic data for permeability estimation directly

from multiple time-lapse seismic surveys. A few exceptions exist. For example, [11] uses ensemble Kalman filtering to refine permeability and porosity estimates. [12] and [13] have explored using time-lapse seismic data for linearized inversion to update permeability. Despite these initial attempts, a more systematic and integrated approach for reservoir characterization and monitoring deserves further investigation.

This chapter introduces a novel 4D processing framework for estimating permeability directly from prestack time-lapse seismic data, offering a streamlined, geophysics based inversion process. Unlike traditional methods, this framework, tested on various synthetic case studies [14, 15, 16, 17], updates permeability models by exclusively matching against multiple observed time-lapse seismic surveys. Despite the potential for rapid model updates, initial results have not yet demonstrated significant alterations in fluid saturation predictions, and the resulting permeability models often lack the heterogeneity necessary for detailed analysis. To address these limitations and to assess the framework’s real-world applicability for 4D monitoring, we undertake a feasibility study using a 2D slice of the Compass model shown in Figure 3.1 [18]. The geological structures of this model were derived from well logs and imaged seismic from the South-West North Sea area — a region under consideration for CO₂ storage [19, 20]. This region comprises a storage unit composed of Bunter sandstone 300 – 500m thick, depicted in orange in Figure 3.1 (a) and characterized by high permeability values shown in Figure 3.1 (b). On top of the storage unit there is a primary seal about 50 m thick, made of the low-permeable Rot Halite Member, and a secondary seal, over 300 m, made of low-permeability mudstone in the Haisborough Group, illustrated in blue and green in Figure 3.1 (a).

This study evaluates the coupled fluid-flow, rock physics, seismic inversion framework’s sensitivity to different starting models, forward modeling errors, and crosstalk during multiparameter inversion, omitting regularization techniques to focus on the impact of time-lapse seismic data on permeability updates. We explore the framework’s ability to recover relatively fine-scale permeability structures, predict CO₂ dynamics within the

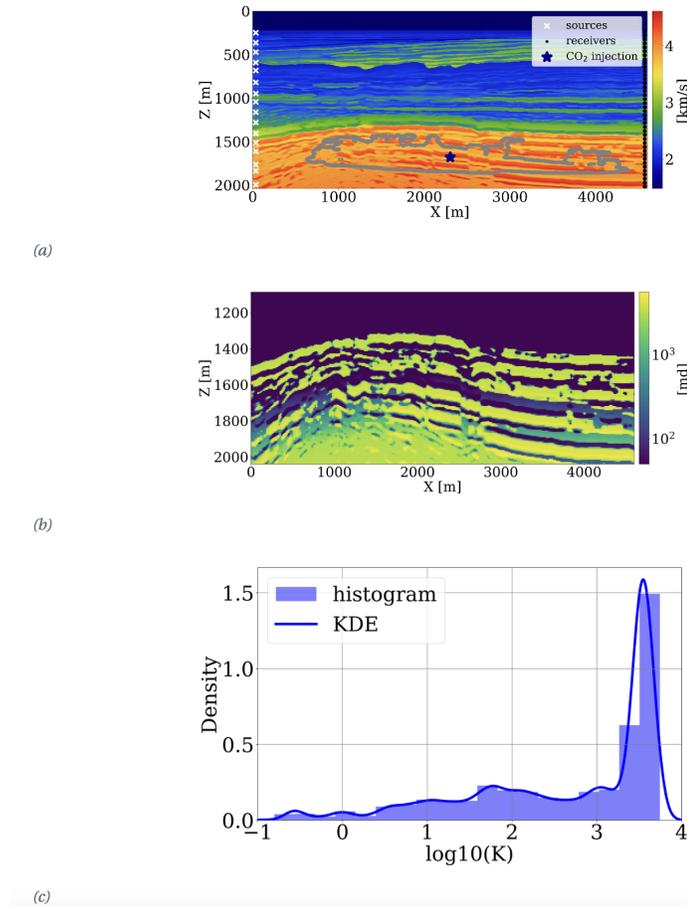


Figure 3.1: Experimental configuration. (a) Setup of seismic acquisition and well control. Dark blue \star denotes the CO₂ injection well. While \times and black \cdot represent source and receiver locations, respectively. The gray curve delineates the shape of the CO₂ plume at the 25th year. (b) Unseen ground truth spatial distribution of horizontal permeability. (c) Histogram of the common logarithm of the permeability model with Kernel Density Estimation (KDE).

seismic monitoring period, and forecast CO₂ dynamics in near future without any seismic observation. Recognizing the limitations of our simplifying assumptions, we conclude with suggestions for future research to advance this promising approach.

3.3 Permeability inversion framework

Our feasibility study examines the time-lapse seismic monitoring of geological carbon storage (GCS), focusing on the integration of three fundamental physics disciplines: fluid-flow physics, rock physics, and wave physics, as illustrated in Figure 3.2. The dynamics of the CO₂ plume during injection are modeled using multiphase flow equations [21], processed through a reservoir simulator [22, 23, 24, 25]. While these simulations require detailed inputs, including well operation parameters and the spatial distribution of porosity and permeability, in this exposition we focus on the permeability, \mathbf{K} , particularly, as the parameter of interest. The output from the reservoir simulator, \mathcal{S} , primarily the time-varying CO₂ saturation snapshots, compiled in \mathbf{c} , serves as the input to the rock physics model, \mathcal{R} . Based on the porosity and the brine-filled baseline velocity before CO₂ injection, this model translates each CO₂ saturation snapshot into altered seismic velocity models, compiled in the vector, \mathbf{v} , using the patchy saturation model proposed by [26]. Lastly, based on the velocity models for each snapshot, the wave modeling operator [27], \mathcal{F} , is used to generate the time-lapse seismic dataset, \mathbf{d} , which collects the seismic data from each vintage.

In practice, the prestack time-lapse seismic dataset, \mathbf{d} , is observed from the field, with the objective to estimate the past, current, and future dynamics of the CO₂ plume in GCS projects. Our methodology diverges from traditional workflows that typically proceed from seismic inversion to quantitative interpretation and subsequent reservoir parameter updates based on the derived wave properties. Instead, we propose an integrated, end-to-end approach that directly inverts the time-lapse surveys collected in \mathbf{d} for permeability, \mathbf{K} , by reducing the time-lapse seismic data misfit objective through an automatic optimization procedure. The core of our method is the composition of three physics-based modeling

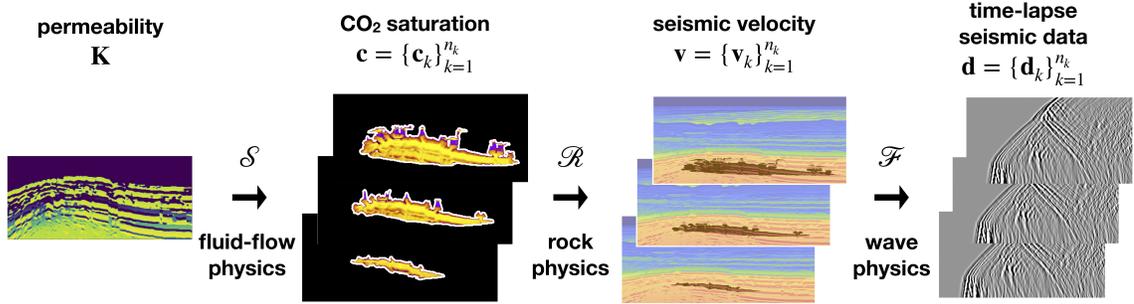


Figure 3.2: Multiphysics forward model. The reservoir simulator, \mathcal{S} , produces n_k snapshots of time-varying CO_2 saturation, compiled in \mathbf{c} , from the permeability model, \mathbf{K} . The rock physics model, \mathcal{R} , based on the porosity and the brine-filled velocity model, converts each CO_2 saturation snapshot, \mathbf{c}_k , to the altered velocity model, \mathbf{v}_k . The shaded area highlights the CO_2 -induced changes in velocity. Finally, the wave modeling, \mathcal{F} , generates a time-lapse seismic dataset, \mathbf{d}_k , for each velocity model, \mathbf{v}_k . These datasets are collected in the vector, \mathbf{d} .

operators, formulated to minimize the following objective function:

$$\underset{\mathbf{K}}{\text{minimize}} \quad \|\mathcal{F} \circ \mathcal{R} \circ \mathcal{S}(\mathbf{K}) - \mathbf{d}\|_2^2. \quad (3.1)$$

This optimization problem is reached with the assumption that the permeability model, \mathbf{K} , is the only unknown parameter during inversion. Especially, the porosity and the brine-filled velocity model before CO_2 injection (as inputs to the patchy saturation model, \mathcal{R}) are assumed known and fixed during the inversion. This objective is minimized via an iterative procedure that includes:

- Generating synthetic time-lapse seismic data using an initial guess for the permeability model;
- Calculating the gradient of the permeability by backpropagating the residuals of the time-lapse seismic datasets;
- Updating the permeability model to reduce the misfit between the synthetic and observed time-lapse seismic datasets.

The advantage of this end-to-end inversion framework lies in its ability to break down silos through multiphysics integration. Specifically, it eliminates the need for intermediate

processing steps to update the saturation and velocity models. As we demonstrate in the subsequent feasibility study, the inverted permeability can produce accurate fluid saturation and velocity models, even for the near future without any additional seismic observations.

3.4 Feasibility study on the Compass model

We evaluate the performance of this inversion framework through a synthetic case study on the Compass model [18]. This model has a grid spacing of 6m in both the horizontal and vertical directions. Compared to conventional reservoir models that often have nearly homogeneous layers and a coarse discretization in the horizontal direction (e.g., 100 m), the high-resolution and spatially heterogeneous Compass model can help reveal the potential of inverting fine-scale geological structures in the permeability model.

Using five vintages of prestack time-lapse seismic surveys, we aim to invert for the spatial distribution of permeability. To this end, we utilize the aforementioned 2D slice of the velocity model, included in Figure 3.1 (a), where the orange region signifies the storage unit. Since the Compass model only includes velocities and densities and no permeability or porosity values, we build a fully heterogeneous ground truth permeability model, displayed in Figure 3.1 (b), by assuming the elementwise relationship in Equation 3.2, between the entries of the brine-filled velocity model, \mathbf{c}_p , in km/s, and the horizontal permeability, \mathbf{K} , in millidarcies (md)¹.

$$\mathbf{K} = \begin{cases} 3000 \exp(\mathbf{c}_p - 4) & \text{if } \mathbf{c}_p \geq 4 \\ 0.01 \exp(25.22(\mathbf{c}_p - 3.5)) & \text{if } \mathbf{c}_p \geq 3.5 \\ 0.01 \exp(\mathbf{c}_p - 3.5) & \text{else} \end{cases} \quad (3.2)$$

Because the baseline seismic Compass model is derived from well and imaged seismic, it contains subtle changes in the seismic properties related to sub-wavelength interference.

¹This relationship is only used to create the ground truth permeability model. When solving the optimization problem in Equation 4.1, no assumptions are made on relationships between the permeability, porosity, and velocity.

The above elementwise relationship between the seismic velocity and permeability ensures that the derived permeability model shares the same spatial heterogeneity as exhibited by the seismic baseline. By construction, it also features significant permeability contrasts within different layers within the storage unit. The low-permeability layers range from approximately 10^{-3} to 1 md, while the high-permeability layers vary between 600 to 6000 md. The histogram of the common logarithm of the permeability model is shown in Figure 3.1 (3), which demonstrates that the permeability values do not follow log-normal distribution. A CO₂ injection well, marked with a dark blue \star , is placed centrally to inject supercritical CO₂ for 25 years at a constant rate of two million metric tons per year. We assume the porosity and the kv/kh ratio to be constant and given by 25% and 10%, respectively. The simulation of compressible and immiscible two-phase flow, where CO₂ displaces brine in porous rocks, is performed using a fully implicit method implemented in JutulDarcy.jl [28, 29]. The boundary of the CO₂ plume at the 25th year is depicted in grey in Figure 3.1 (a). After converting the CO₂ saturation into seismic velocity models, v , via the patchy saturation model, acoustic time-lapse seismic data is generated with constant density for five vintages at years 5, 10, 15, 20, and 25 using Devito [30, 31] and JUDI.jl [32, 33], employing a Ricker wavelet with a central frequency of 20 Hz. The well-bore source and receiver geometries are shown in Figure 3.1 (a).

3.4.1 Sensitivity with respect to starting models

To evaluate the efficacy of our end-to-end inversion framework, particularly its sensitivity to initial permeability models, we examine two distinct initial permeability models. In case 1, the initial permeability model, shown in Figure 3.3 (a), features homogeneous permeability values (100md) across the entire reservoir. This model allows us to explore the extent of permeability updates achievable from a non-informative permeability model. In case 2, we apply a spatial distortion [34] to the unseen ground truth permeability in Figure 3.1 (b) to obtain the initial permeability model, shown in Figure 3.3 (b). The values of

different permeability layers are near accurate, but the positions are misplaced.

In both cases, we employ a methodological shortcut often referred to as committing an “inversion crime”, where the data generation and inversion processes share the same computational kernel. This ideal setup is used here to show what is ideally achievable by this inversion framework. To add a layer of realism, we incorporate 8 dB of incoherent band-limited Gaussian noise into the observed time-lapse datasets, which severely contaminates the seismic signal in the time-lapse difference.

To invert for the permeability model, we run 100 iterations of stochastic gradient descent (SGD), starting with Figure 3.3 (a) and Figure 3.3 (b). During each iteration, we randomly draw four sources out of the total of 32 sources to calculate the misfit and the gradient with respect to permeability model [35]. This amounts to 12.5 datapasses through the entire time-lapse seismic dataset. We display permeability updates in logarithmic scale for both cases, in Figure 3.3 (e) and Figure 3.3 (f), respectively. Additionally, Figure 3.3 (c) and Figure 3.3 (d) offer a visualization of “ideal” updates by showing the logarithmic differences between the ground truth and the initial permeability models.

The following observations can be made: First, the permeability updates are primarily confined to areas directly influenced by the CO₂ plume’s flow, as delineated by the gray curves. This outcome is expected since the time-lapse variations in wave properties are attributed to changes in fluid saturation exclusively. Consequently, without additional information, this inversion method does not alter permeability values outside the CO₂ plume’s extent where the flow of CO₂ has not occurred. Second, the inverted permeability within the CO₂ plume largely reflects the trend of the ground truth permeability model. In case 1, the framework successfully identifies major permeability layers — both high and low (depicted in red and blue, respectively) at approximately 1600 m depth — accurately capturing their depth and lateral distribution in alignment with the actual layers. In case 2, the inversion process introduces high-resolution details to the layers affected by the plume, aligning well with the ideal updates shown in Figure 3.3 (d). Despite these successes, the

full magnitude of permeability contrasts is not entirely captured, pointing to the inherently ill-posed nature of permeability inversion [36], necessitating workflows that include uncertainty quantification for future investigations [37].

3.4.2 Sensitivity with respect to forward modeling errors

To extend our investigation beyond overly idealized scenarios, we examine the framework’s robustness next, during scenarios that avoid “committing the inversion crime”. A critical area of focus in time-lapse seismic is the error in the brine-filled baseline velocity model before CO₂ injection. Errors in this baseline model, which feeds into the rock physics model, can produce inaccurate velocity models of the CO₂-filled reservoir, leading to inaccuracies in the simulated time-lapse seismic datasets.

To construct an inaccurate but realistic baseline velocity model, we use a Ricker wavelet with central frequency of 20 Hz to generate cross-well and surface seismic data before CO₂ injection, and employ the SGD method to run 10 datapass of FWI with a kinematically correct but smooth initial velocity model to obtain the inverted velocity model depicted in Figure 3.4 (a). This imperfect brine-filled velocity model is subsequently fed into the rock physics model, \mathcal{R} , for permeability inversion (case 3).

We employ the initial permeability model from Figure 3.3 (a) to assess the impact of modeling errors on the inversion results. The update to the permeability, shown in logarithmic scale in Figure 3.4 (b), reveals some artifacts outside the CO₂ plume area due to the modeling inaccuracies. Additionally, a high permeability zone within the plume is slightly misplaced when compared to the updates in Figure 3.3 (e), yet the overall trend of permeability changes is correctly captured.

Following this permeability update, we proceed with reservoir simulations using the updated permeability models to assess the corrections made to the CO₂ plume predictions. This step is crucial for validating the practical utility of the inversion framework for real-world seismic monitoring scenarios.

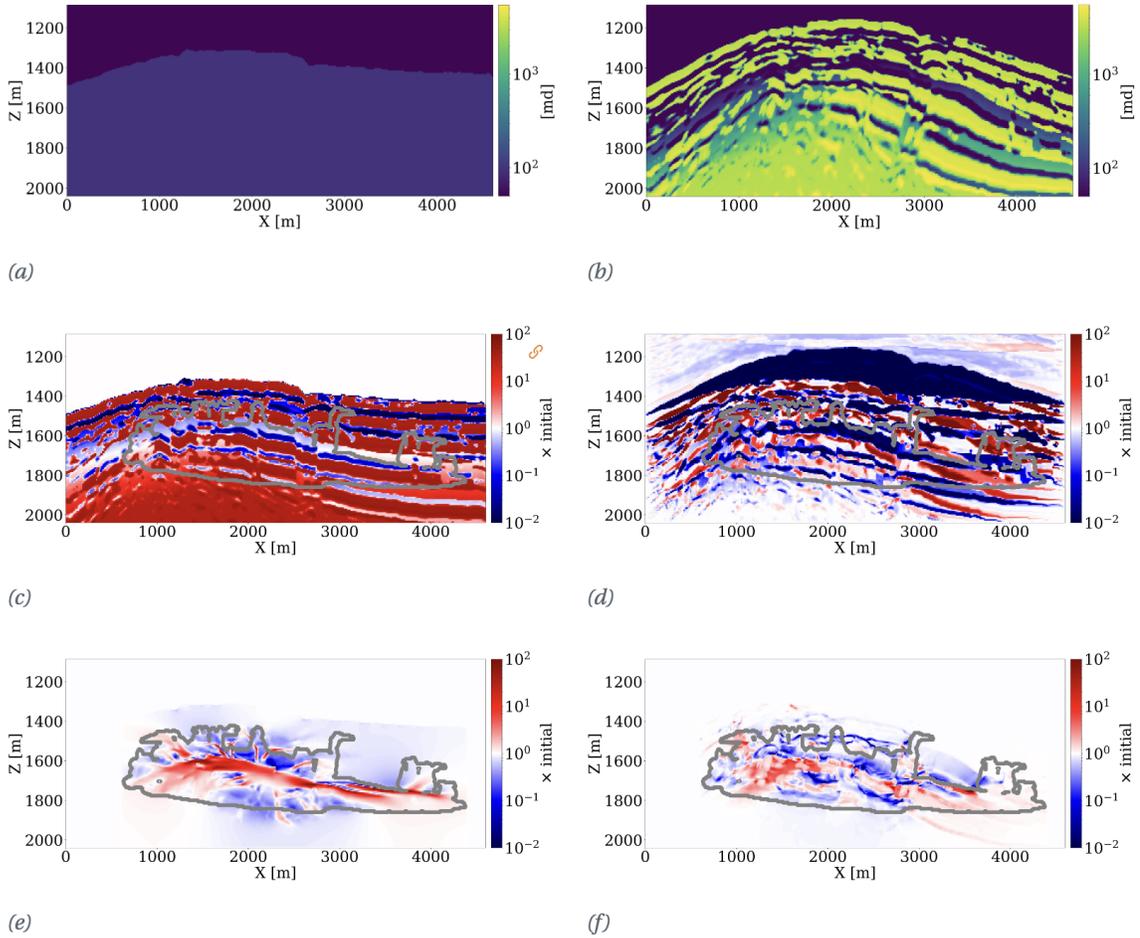


Figure 3.3: Permeability inversion results for cases 1 and 2. (a)(c)(e) display the initial permeability model in case 1, the logarithmic ratio of the ground truth permeability (Figure 3.1 (b)) to the initial one, and the logarithmic ratio of the inverted permeability to the initial one. (b)(d)(f) display the same but for case 2 with a distorted initial permeability model. Gray curve indicates the boundary of the CO₂ plume at the 25th year. “ \times initial” on the caption of the colorbar represents the factor by which the initial permeability is updated.

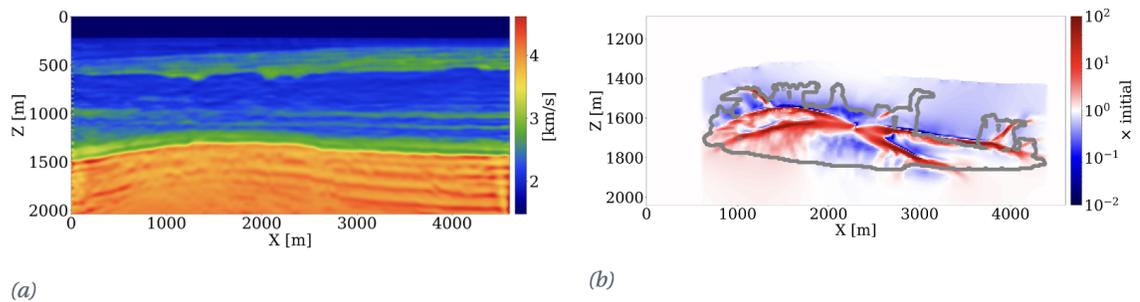


Figure 3.4: Permeability inversion results for case 3. (a) Inverted brine-filled baseline velocity used in permeability inversion. (b) The logarithmic ratio of the inverted permeability to the initial one. Gray curve indicates the boundary of the CO₂ plume at the 25th year.

3.4.3 CO₂ plume estimation and forecast

The primary objective of our end-to-end inversion framework is to accurately estimate reservoir permeability, a crucial step towards the ultimate goal of predicting CO₂ saturation both historically and in the near future. Based on initial, inverted, and ground truth permeability models, we conduct a quality control involving CO₂ saturation simulations, as depicted in Figure 3.5. Across all simulations, we note substantial improvements in predictions of the CO₂ plume shape, closely aligning with the boundaries of the ground truth CO₂ plume. Notably, the initial simulations significantly misjudged the lateral spread of the CO₂ plume. The corrections applied through the updated permeability models, however, yield accurate representations of the plume’s lateral extent.

Expanding our analysis to future forecasting, Figure 3.6 illustrates the predicted movement of the CO₂ plume over a 40-year period, following a 25-year injection phase, without further CO₂ injection or seismic observations. During this forecasted period, the CO₂ plume primarily ascends due to buoyancy, while a portion (approximately 10%) remains trapped in the pore spaces, indicated in purple. This phenomenon, known as residual trapping [38], is a critical factor in assessing the long-term storage capabilities of GCS projects. Initial forecasts tend to underestimate the extent of CO₂ sequestration through residual trapping. In contrast, simulations driven by the updated permeability models not only provide a more accurate estimation of the permanently stored CO₂ volume but also closely match the ground truth CO₂ plume’s boundaries, even without collecting further monitoring data.

3.4.4 Multiparameter inversion

While case 1-3 demonstrate the performance of the inversion framework for permeability estimation, further scrutiny is in order to investigate its performance when multiple parameters are unknown and need to be jointly estimated. To this end, we design case 4 where spatial distributions of both porosity and permeability are unknown. While permeability only appears in the two-phase flow equations, porosity appears as an input not

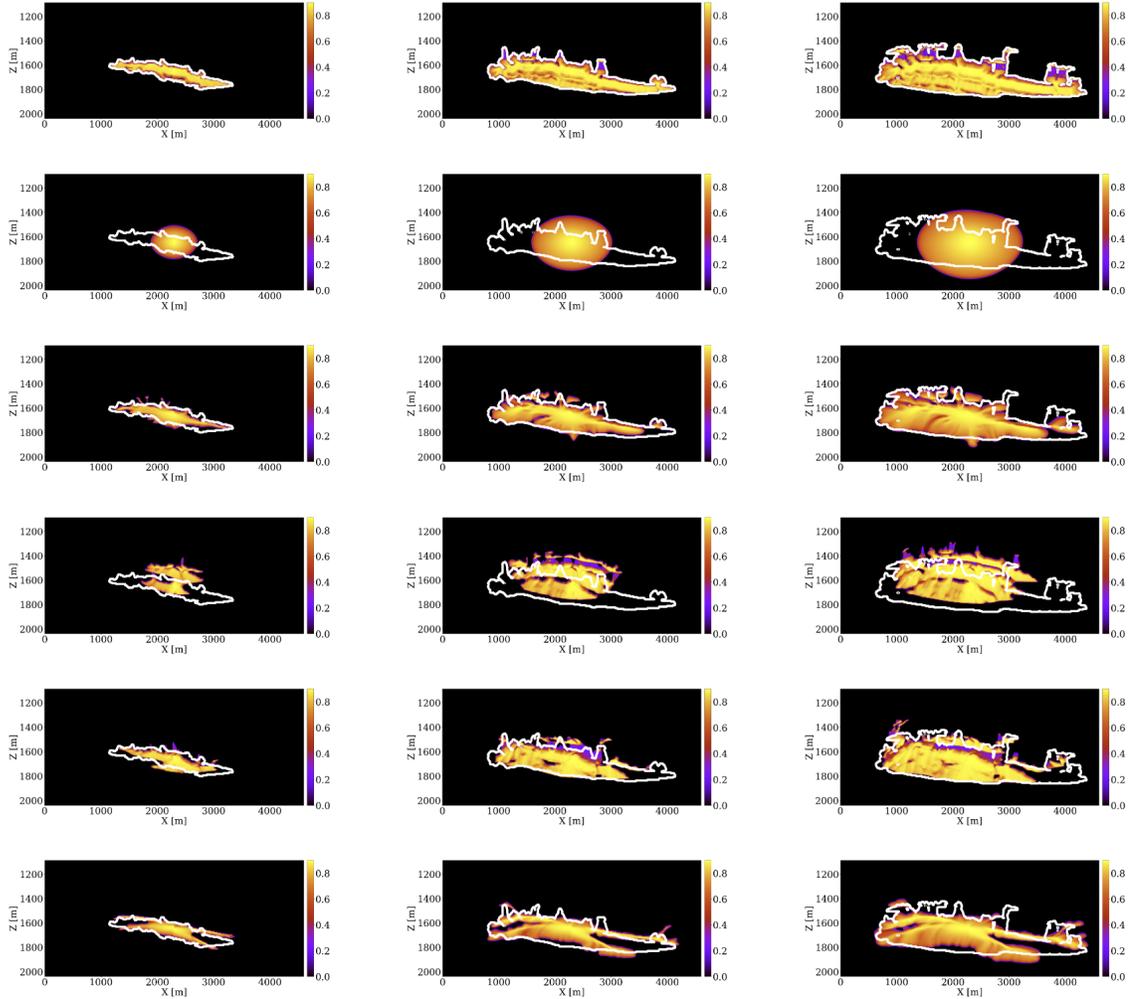


Figure 3.5: Predicted CO_2 saturation for 5th, 15th, and 25th years, shown in first, second, and third columns, respectively. The first row shows the (unseen) ground truth CO_2 saturation. The second and fourth rows show the saturation predicted with initial permeability models in Figure 3.3 (a) and Figure 3.3 (b), respectively. The third, fifth, and sixth rows show the updated saturation after updating the initial permeability models by Figure 3.3 (e), Figure 3.3 (f), and Figure 3.4 (b), respectively. The boundaries of the (unseen) ground truth CO_2 saturation are shown in white curves.

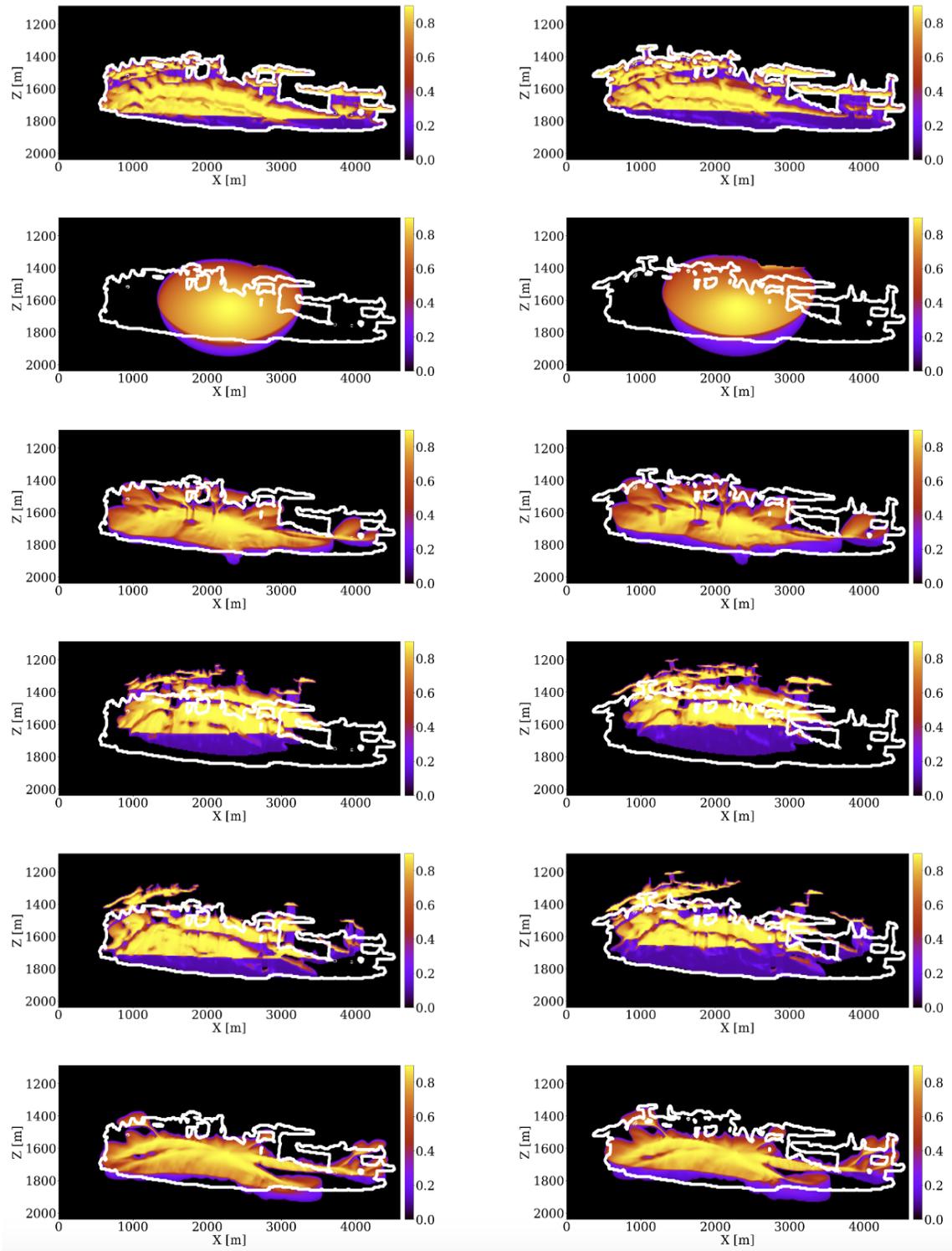


Figure 3.6: CO₂ plume forecasts for 45th and 65th years, shown in first and second columns, respectively. The ordering of the rows remains the same as Figure 3.5. Purple regions display the CO₂ plume permanently stored via the residual trapping mechanism.

only in the two-phase flow reservoir simulator, \mathcal{S} , but also in the patchy saturation model, \mathcal{R} . This results in a more challenging inverse problem because porosity affects more than one physics-based forward models in Figure 3.2, and because there can be crosstalk in the gradient calculations during the inversion.

As a proof of concept, we simplify this multiparameter inversion problem by assuming that the porosity, ϕ , and the horizontal permeability, \mathbf{K} , are related by the following elementwise Kozeny-Carman relationship [39]:

$$\mathbf{K} = \mathcal{T}(\phi) = 3.65 \times 10^4 \frac{\phi^3}{(1 - \phi)^2}. \quad (3.3)$$

Following this relationship, we artificially create a ground truth porosity model, shown in Figure 3.7 (c), according to the permeability values in Figure 3.1 (b), and then use the ground truth permeability and porosity models to simulate CO₂ saturation, velocity models, and the five time-lapse seismic datasets. During inversion, we parameterize permeability by porosity according to Equation 3.3, and minimize the following objective function to invert for the porosity:

$$\underset{\phi}{\text{minimize}} \quad \|\mathcal{F}(\mathcal{R}(\phi, \mathcal{S}(\phi, \mathcal{T}(\phi)))) - \mathbf{d}\|_2^2. \quad (3.4)$$

The scalable and differentiable programming framework, proposed by [16], allows for effortless and accurate gradient calculation with respect to porosity, ϕ , which otherwise requires labor-intensive and error-prone derivation of cross-gradient terms by hand. We initialize the reservoir with homogeneous porosity values of 12%, shown in Figure 3.7 (a). After 100 iterations of SGD, the inverted porosity is shown in Figure 3.7 (b). While some layers in the inverted porosity are slightly misplaced in this preliminary study, the overall trend of porosity is adequately estimated in the center of the model.

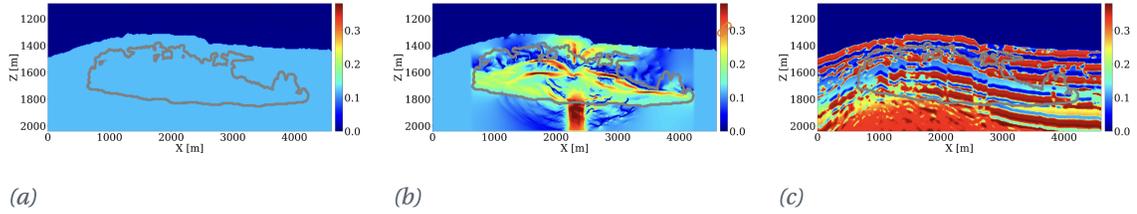


Figure 3.7: Porosity inversion in case 4. (a) Initial porosity. (b) Inverted porosity. (c) Unseen ground truth porosity. The gray curve delineates the shape of the CO₂ plume at the 25th year.

3.5 Limitations

While our case studies offer promising insights, it is crucial to acknowledge the assumptions underpinning our approach and recognize the inherent limitations that merit further investigation. Additionally, we explore the potential for integrating this 4D processing workflow with other reservoir characterization and management strategies.

3.5.1 Reservoir simulation

Our study assumes known values for all multiphase flow model parameters, including relative permeability functions, residual water saturation, temperature and capillary pressure. These parameters were kept constant in the simulations to isolate the impact of permeability on seismic data, but there can be significant rock-dependent variations in practice. A multiparameter inversion, indicated by the preliminary case study in case 4, is worthwhile for future investigation to extend this inversion framework through joint estimation of these parameters. Further exploration is also required to understand the crosstalk between these parameters. In addition, our assumption that supercritical CO₂ miscibility in the resident brine is low could be removed by considering a compositional flow model that introduces additional uncertain parameters. The feasibility of such approaches hinges on the availability of a differentiable reservoir simulator, like `JutulDarcy.jl`, or the use of deep neural networks to approximate the physics of multiphase flow [40, 16, 17] and serve as a surrogate during inversion. Moreover, multiphase flow equations may not hold in scenarios

involving CO₂ leakage, necessitating robust leakage detection methodologies [41, 42].

3.5.2 Rock physics

The case studies currently ignore the pressure effect on the wave properties [43, 44]. While this can be justified for some GCS projects where the pressure change is relatively subtle, the inversion framework can be extended to honor the relationship between pressure and wave properties, and include geomechanical effects [45, 46]. The patchy saturation model may also not fully capture the complexities of real-world reservoirs [47], indicating a need for calibration of the rock physics model against actual reservoir and seismic data.

3.5.3 Wave physics

The omission of updates to the brine-filled baseline velocity model represents a simplification that warrants further exploration. Future research could extend the framework to jointly update this baseline alongside permeability, incorporating additional parameters like shear velocity and density, which are currently ignored in the modeling and inversion. Quantifying uncertainties in velocity [48, 49, 50] and permeability models remains a critical challenge for enhancing the reliability of inversion results.

3.6 Discussion and conclusion

Our feasibility studies demonstrate the performance of this inversion framework in estimating permeability models directly from multiple prestack time-lapse seismic datasets in a cross-well setting. The recovered horizontal details in the permeability, especially shown in Figure 3.3 (f) using a 6 m grid spacing, highlight the capability of this end-to-end inversion framework to provide high-resolution spatial information of the permeability model. This framework also has the potential to significantly reduce cycle time in 4D processing workflows by avoiding labor-intensive, step-by-step inversion schemes where seismic velocity and CO₂ saturation are subsequently inverted from right to left in Figure 3.2. Further-

more, the proposed framework differs from conventional workflows [51, 52] by coupling fluid-flow, rock, and wave physics, leveraging the sensitivities of the reservoir simulator through a differential programming software framework [16]. By utilizing the rock physics model that links changes in CO₂ saturation to changes in seismic properties, we gain access to these sensitivities from the time-lapse seismic data, allowing us to invert for the permeability directly.

Opportunities for future research remain. This inversion framework can be further enhanced to incorporate multimodal observations, such as a combination of well measurements and seismic data. This enhancement can be readily achieved by integrating additional misfit terms into the objective function, as detailed by [17]. Moreover, the resolution of permeability inversion deserves more exploration. Currently, the inversion is limited to the resolution achievable by seismic methods. Adopting a multimodal data assimilation approach necessitates further studies into the upscaling of the permeability model [53], which we leave for future work. Additionally, the permeability inversion framework is well-suited for integration with the digital twin framework, as reported by [54] and in other ongoing projects at our research group. When time-lapse seismic and well measurements are collected from the field, performing permeability inversion facilitates more accurate estimations of reservoir properties. These estimations can then be used to forecast the CO₂ plume's behavior and optimize well injectivity in GCS projects, thereby maximizing injection volumes while minimizing fracturing risks [55].

3.7 Data availability

Software with this research can be accessed at <http://github.com/slimgroup/TL-FWPI.jl/>, with the DOI link <https://doi.org/10.5281/zenodo.10910283>. The full 3D Compass model is open access with CC BY license, available at <ftp://slim.gatech.edu/data/synth/Compass/>.

3.8 References

- [1] D. E. Lumley, “Time-lapse seismic reservoir monitoring,” *Geophysics*, vol. 66, no. 1, pp. 50–53, 2001 (page 51).
- [2] R. Chadwick, D. Noy, R. Arts, and O. Eiken, “Latest time-lapse seismic data from sleipner yield new insights into co2 plume development,” *Energy Procedia*, vol. 1, no. 1, pp. 2103–2110, Feb. 2009 (page 51).
- [3] A. Chadwick *et al.*, “Quantitative analysis of time-lapse seismic monitoring data at the sleipner CO₂ storage operation,” *The Leading Edge*, vol. 29, no. 2, pp. 170–177, 2010 (page 51).
- [4] A.-K. Furre, O. Eiken, H. Alnes, J. N. Vevatne, and A. F. Kiær, “20 years of monitoring CO₂-injection at sleipner,” *Energy Procedia*, vol. 114, pp. 3916–3926, 2017, 13th International Conference on Greenhouse Gas Control Technologies, GHGT-13, 14-18 November 2016, Lausanne, Switzerland (page 51).
- [5] D. Lumley, “4d seismic monitoring of co 2 sequestration,” *The Leading Edge*, vol. 29, no. 2, pp. 150–155, 2010 (page 51).
- [6] E. Hicks, H. Hoerber, M. Houbiers, S. Lescoffit, A. Ratcliffe, and V. Vinje, “Time-lapse full-waveform inversion as a reservoir-monitoring tool — a north sea case study,” *The Leading Edge*, vol. 35, no. 10, pp. 850–858, Oct. 2016 (page 51).
- [7] M. Bosch, T. Mukerji, and E. F. Gonzalez, “Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review,” *GEOPHYSICS*, vol. 75, no. 5, 75A165–75A176, Sep. 2010 (page 51).
- [8] L. Wei, P. Roy, T. Dygert, D. Grimes, and M. Edwards, “Estimation of reservoir pressure and saturation changes from 4d inverted elastic properties,” *SEG Technical Program Expanded Abstracts 2017*, Aug. 17, 2017 (page 51).
- [9] M. Bosch, L. Cara, J. Rodrigues, A. Navarro, and M. Díaz, “A monte carlo approach to the joint estimation of reservoir and elastic parameters from seismic amplitudes,” *GEOPHYSICS*, vol. 72, no. 6, O29–O39, Nov. 2007 (page 51).
- [10] Q. Hu, D. Grana, and K. A. Innanen, “Feasibility of seismic time-lapse monitoring of co2 with rock physics parametrized full waveform inversion,” *Geophysical Journal International*, vol. 233, no. 1, pp. 402–419, Nov. 29, 2022 (page 51).

- [11] K. S. Eikrem, G. Nævdal, M. Jakobsen, and Y. Chen, “Bayesian estimation of reservoir properties—effects of uncertainty quantification of 4d seismic data,” *Computational Geosciences*, vol. 20, no. 6, pp. 1211–1229, Aug. 16, 2016 (page 52).
- [12] D. W. Vasco, A. Datta-Gupta, R. Behrens, P. Condon, and J. Rickett, “Seismic imaging of reservoir flow properties: Time-lapse amplitude changes,” *GEOPHYSICS*, vol. 69, no. 6, pp. 1425–1442, Nov. 2004 (page 52).
- [13] D. W. Vasco, H. Keers, J. Khazanehdari, and A. Cooke, “Seismic imaging of reservoir flow properties: Resolving water influx and reservoir permeability,” *GEOPHYSICS*, vol. 73, no. 1, O1–O13, Jan. 2008 (page 52).
- [14] D. Li, K. Xu, J. M. Harris, and E. Darve, “Coupled time-lapse full-waveform inversion for subsurface flow problems using intrusive automatic differentiation,” *Water Resources Research*, vol. 56, no. 8, e2019WR027032, 2020 (page 52).
- [15] Z. Yin, A. Siahkoohi, M. Louboutin, and F. J. Herrmann, “Learned coupled inversion for carbon sequestration monitoring and forecasting with fourier neural operators,” in *International Meeting for Applied Geoscience and Energy Expanded Abstracts*, (IMAGE, Houston), May 2022 (page 52).
- [16] M. Louboutin, *slimgroup/ImageGather.jl: v0.2.6*. Zenodo, Jun. 2, 2023 (pages 52, 64, 65, 67).
- [17] Z. Yin, R. Orozco, M. Louboutin, and F. J. Herrmann, “Solving multiphysics-based inverse problems with learned surrogates and constraints,” *Advanced Modeling and Simulation in Engineering Sciences*, vol. 10, no. 1, p. 14, 2023 (pages 52, 65, 67).
- [18] C. Jones, J. Edgar, J. Selvage, and H. Crook, “Building complex synthetic models to evaluate acquisition geometries and velocity inversion technologies,” in *74th EAGE Conference and Exhibition incorporating EUROPEC 2012*, European Association of Geoscientists & Engineers, 2012, cp–293 (pages 52, 56).
- [19] C. Kolster, S. Agada, N. Mac Dowell, and S. Krevor, “The impact of time-varying CO₂ injection rate on large scale storage in the UK bunter sandstone,” *International Journal of Greenhouse Gas Control*, vol. 68, pp. 77–85, 2018 (page 52).
- [20] R. Meneguolo, N. Thompson, C. Acuna, A.-K. Furre, and E. Milovanova, “Sub-surface maturation in a saline aquifer ccs project development. experience from the northern lights project, offshore norway,” *Geoenergy*, geoenergy2024–013, 2024 (page 52).
- [21] K. Pruess and J. Nordbotten, “Numerical simulation studies of the long-term evolution of a CO₂ plume in a saline aquifer with a sloping caprock,” *Transport in porous media*, vol. 90, no. 1, pp. 135–151, 2011 (page 54).

- [22] S. Krogstad, K.-.-A. Lie, O. Møyner, H. M. Nilsen, X. Raynaud, and B. Skaflestad, “Mrst-ad—an open-source framework for rapid prototyping and evaluation of reservoir simulation problems,” in *SPE Reservoir Simulation Conference*, SPE, 2015, D022S002R004 (page 54).
- [23] H. Gross and A. Mazuyer, “Geosx: A multiphysics, multilevel simulator designed for exascale computing,” in *SPE Reservoir Simulation Conference*, OnePetro, 2021 (page 54).
- [24] A. F. Rasmussen *et al.*, “The open porous media flow reservoir simulator,” *Computers & Mathematics with Applications*, vol. 81, pp. 159–185, 2021 (page 54).
- [25] R. W. Stacey and M. J. Williams, “Validation of eclipse reservoir simulator for geothermal problems,” *GRC Transactions*, vol. 41, pp. 2095–2109, 2017 (page 54).
- [26] P. Avseth, T. Mukerji, and G. Mavko, *Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk*. Cambridge university press, 2010 (page 54).
- [27] A. Tarantola, “Inversion of seismic reflection data in the acoustic approximation,” *Geophysics*, vol. 49, no. 8, pp. 1259–1266, 1984 (page 54).
- [28] O. Møyner and G. Bruer, *sintefmath/JutulDarcy.jl: v0.2.1*. Zenodo, Feb. 21, 2023 (page 57).
- [29] Z. Yin, G. Bruer, and M. Louboutin, *Slimgroup/jutuldarcyrules.jl: V0.2.5*, version v0.2.5, Apr. 2023 (page 57).
- [30] M. Louboutin *et al.*, “Devito (v3.1.0): An embedded domain-specific language for finite differences and geophysical exploration,” *Geoscientific Model Development*, vol. 12, no. 3, pp. 1165–1187, 2019 (page 57).
- [31] F. Luporini *et al.*, “Architecture and performance of devito, a system for automated stencil computation,” *ACM Trans. Math. Softw.*, vol. 46, no. 1, Apr. 2020 (page 57).
- [32] P. A. Witte *et al.*, “A large-scale framework for symbolic implementations of seismic inversion algorithms in julia,” *GEOPHYSICS*, vol. 84, no. 3, F57–F71, 2019. eprint: <https://doi.org/10.1190/geo2018-0174.1> (page 57).
- [33] M. Louboutin *et al.*, *slimgroup/JUDI.jl: v3.3.8*. Zenodo, Sep. 18, 2023 (page 57).
- [34] M. D. Bloice, C. Stocker, and A. Holzinger, “Augmentor: An image augmentation library for machine learning,” 2017 (page 57).

- [35] F. J. Herrmann *et al.*, “Frugal full-waveform inversion: From theory to a practical algorithm,” *The Leading Edge*, vol. 32, no. 9, pp. 1082–1092, 2013 (page 58).
- [36] Z. Zhang, B. Jafarpour, and L. Li, “Inference of permeability heterogeneity from joint inversion of transient flow and temperature data,” *Water Resources Research*, vol. 50, no. 6, pp. 4710–4725, Jun. 2014 (page 59).
- [37] A. P. Gahlot, H. T. Erdinc, R. Orozco, Z. Yin, and F. J. Herrmann, “Inference of co2 flow patterns—a feasibility study,” *arXiv preprint arXiv:2311.00290*, 2023 (page 59).
- [38] T. Rahman, M. Lebedev, A. Barifcani, and S. Iglauer, “Residual trapping of supercritical co2 in oil-wet sandstone,” *Journal of Colloid and Interface Science*, vol. 469, pp. 63–68, May 2016 (page 61).
- [39] A. Costa, “Permeability-porosity relationship: A reexamination of the kozeny-carman equation based on a fractal pore-space geometry assumption,” *Geophysical research letters*, vol. 33, no. 2, 2006 (page 64).
- [40] T. J. Grady *et al.*, “Model-parallel fourier neural operators as learned surrogates for large-scale parametric pdes,” *Computers & Geosciences*, vol. 178, p. 105 402, 2023 (page 65).
- [41] H. T. Erdinc, A. P. Gahlot, Z. Yin, M. Louboutin, and F. J. Herrmann, “De-risking carbon capture and sequestration with explainable co2 leakage detection in time-lapse seismic monitoring images,” 2022 (page 66).
- [42] Z. Yin, H. T. Erdinc, A. P. Gahlot, M. Louboutin, and F. J. Herrmann, “Derisking geologic carbon storage from high-resolution time-lapse seismic to explainable leakage detection,” *The Leading Edge*, vol. 42, no. 1, pp. 69–76, 2023 (page 66).
- [43] C. MacBeth, “A classification for the pressure-sensitivity properties of a sandstone rock frame,” *GEOPHYSICS*, vol. 69, no. 2, pp. 497–510, Mar. 2004 (page 66).
- [44] C. MacBeth, M. Floricich, and J. Soldo, “Going quantitative with 4d seismic analysis,” *Geophysical Prospecting*, vol. 54, no. 3, pp. 303–317, Apr. 12, 2006 (page 66).
- [45] R. Chen *et al.*, “Coupled fluid flow and geomechanical modeling of seismicity in the azle area (north texas),” *SPE Reservoir Evaluation & Engineering*, vol. 23, no. 03, pp. 1006–1018, 2020 (page 66).
- [46] M. Nagao, C. Yao, T. Onishi, H. Chen, and A. Datta-Gupta, “An efficient deep learning-based workflow for co2 plume imaging with distributed pressure and temperature measurements,” *SPE Journal*, pp. 1–15, 2023 (page 66).

- [47] F. Allo and L. Vernik, “Overcoming gassmann’s equation limitations in reservoir rocks,” *The Leading Edge*, vol. 43, no. 5, pp. 278–284, 2024 (page 66).
- [48] Z. Yin, R. Orozco, M. Louboutin, and F. J. Herrmann, “Wise: Full-waveform variational inference via subsurface extensions,” *Geophysics*, vol. 89, no. 4, pp. 1–31, 2024 (page 66).
- [49] Z. Yin, R. Orozco, and F. J. Herrmann, “Wiser: Multimodal variational inference for full-waveform inversion without dimensionality reduction,” *arXiv preprint arXiv:2405.10327*, 2024 (page 66).
- [50] R. Orozco, A. Siahkoohi, M. Louboutin, and F. J. Herrmann, “Aspire: Iterative amortized posterior inference for bayesian inverse problems,” *arXiv preprint arXiv:2405.05398*, 2024 (page 66).
- [51] M. Hatab and C. MacBeth, “Assessing data error for 4d seismic history matching: Uncertainties from processing workflow,” in *EAGE Annual Conference & Exhibition*, European Association of Geoscientists & Engineers, vol. 2021, 2021, pp. 1–5 (page 67).
- [52] M. Hatab and C. MacBeth, “Assessment of data error for 4d quantitative interpretation,” in *First International Meeting for Applied Geoscience & Energy*, Society of Exploration Geophysicists, 2021, pp. 3439–3443 (page 67).
- [53] X.-H. Wu, Y. Efendiev, and T. Y. Hou, “Analysis of upscaling absolute permeability,” *Discrete and Continuous Dynamical Systems Series B*, vol. 2, no. 2, pp. 185–204, 2002 (page 67).
- [54] F. J. Herrmann, “President’s page: Digital twins in the era of generative ai,” *The Leading Edge*, vol. 42, no. 11, pp. 730–732, 2023 (page 67).
- [55] A. P. Gahlot, H. Li, Z. Yin, R. Orozco, and F. J. Herrmann, “A digital twin for geological carbon storage with controlled injectivity,” *arXiv preprint arXiv:2403.19819*, 2024 (page 67).

CHAPTER 4

SOLVING MULTIPHYSICS-BASED INVERSE PROBLEMS WITH LEARNED SURROGATES AND CONSTRAINTS

4.1 Summary

Solving multiphysics-based inverse problems for geological carbon storage monitoring can be challenging when multimodal time-lapse data are expensive to collect and costly to simulate numerically. We overcome these challenges by combining computationally cheap learned surrogates with learned constraints. Not only does this combination lead to vastly improved inversions for the important fluid-flow property, permeability, it also provides a natural platform for inverting multimodal data including well measurements and active-source time-lapse seismic data. By adding a learned constraint, we arrive at a computationally feasible inversion approach that remains accurate. This is accomplished by including a trained deep neural network, known as a normalizing flow, which forces the model iterates to remain in-distribution, thereby safeguarding the accuracy of trained Fourier neural operators that act as surrogates for the computationally expensive multiphase flow simulations involving partial differential equation solves. By means of carefully selected experiments, centered around the problem of geological carbon storage, we demonstrate the efficacy of the proposed constrained optimization method on two different data modalities, namely time-lapse well and time-lapse seismic data. While permeability inversions from both these two modalities have their pluses and minuses, their joint inversion benefits from either, yielding valuable superior permeability inversions and CO₂ plume predictions near, and far away, from the monitoring wells.

4.2 Introduction

In this chapter, we introduce a novel learned inversion algorithm designed to address inverse problems based on partial differential equations (PDEs). These problems can be represented using the following general form:

$$\mathbf{d} = \mathcal{H} \circ \mathcal{S}(\mathbf{K}) + \epsilon. \quad (4.1)$$

In this expression, the nonlinear operator \mathcal{S} represents the solution operator of a nonlinear parametric PDE mapping the coefficients \mathbf{K} to the solution. Given numerical solutions of the PDE, partially observed data, collected in the vector \mathbf{d} , are modeled by compounding the solution operator with the measurement operator, \mathcal{H} , followed by adding the noise term ϵ with noise level of σ —i.e., $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. This problem is quite general and pertinent to various physical applications, including geophysical exploration [1, 2], medical imaging [3], and experimental design [4].

Without loss of generality, we focus on time-lapse seismic monitoring of geological carbon storage (GCS), which involves underground storage of supercritical CO_2 captured from the atmosphere or from industrial smoke stacks [5]. We consider GCS in saline aquifers, which involves multiphase flow physics where CO_2 replaces brine in the porous rocks [6]. In this context, the PDE solution operator, \mathcal{S} , serves as the multiphase flow simulator, which takes the gridded spatially varying permeability in the reservoir, \mathbf{K} , as input and produces n_t time-varying CO_2 saturation snapshots, $\mathbf{c} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{n_t}]$, as output. The governing equations for the multiphase flow involve Darcy’s and the mass conservation law. Detailed information on the governing equations, initial and boundary conditions, and numerical solution schemes can be found in [7] and the references therein. To ensure safety, confinement, and containment of GCS projects, various kinds of time-lapse data are collected to monitor CO_2 plumes. These different data modalities include measurements in wells [8, 9], and the collection of gravity [10, 11], electromagnetic [12, 13], and seismic time-lapse

data [14, 15, 16] that can be used to follow the plume and invert for reservoir properties such as the permeability, \mathbf{K} . The latter is the property of interest in this exposition.

Overall, solving for the reservoir model parameter, \mathbf{K} , poses significant challenges for two primary reasons:

- the forward modeling operator, $\mathcal{H} \circ \mathcal{S}$, can be ill-posed, resulting in multiple model parameters that fit the observed data equally well. This necessitates the use of regularizers [17, 2] in the form of penalties or constraints [18].
- The PDE modeling operator \mathcal{S} , and the sensitivity calculations with respect to the model parameters can be computationally expensive for large-scale problems, limiting the efficacy of iterative methods such as gradient-based [19] or Markov chain Monte Carlo [20] methods.

To overcome the second challenge, numerous attempts have been made to replace computationally expensive PDE solves with more affordable approximate alternatives [21, 22], which include the use of radial basis functions to learn the complex models from few sample points [23] or reduced-order modeling where the dimension of the model space is reduced [24, 25]. More recently, various deep learning techniques have emerged as cheap alternatives to numerical PDE solves [26, 27, 28, 29, 30, 31, 32]. After incurring initial training costs, these neural operators lead to vastly improved computation of PDE solves. Data-driven methods have also been used successfully to learn coarse-to-fine grid mappings of PDEs solves. Because of their advertised performance on approximating solution operators of the multiphase flow in porous media [33, 34, 35, 36, 37], we will consider Fourier neural operators [38, 30] in this work even though alternative choices can be made. Once trained, FNOs produce approximate PDE solutions orders of magnitude faster than traditional solvers [38, 39, 35, 40]. In addition, [41, 42, 43] demonstrated that trained FNOs can replace PDE solution operators during inversion. This latest development is especially beneficial to applications such as GCS where trained FNOs can be used in lieu of

numerically costly flow simulators [44, 7, 45]. However, despite their promising results, unconstrained inversion formulations offer little to no guarantees that the model iterates remain within the statistical distribution on which the FNO was trained initially during inversion. As a consequence, FNOs may no longer produce accurate fluid-flow simulations throughout the iterations, which can lead to erroneous inversion results when the errors become too large, possibly rendering surrogate modeling by FNOs ineffective. To avoid this situation, we propose a constrained formulation where a trained normalizing flow (NF, [46]) is included as a learned constraint. This added constraint guarantees that the model iterates remain within the desired statistical distribution. Because our approach safeguards the FNO’s accuracy, it allows FNOs to act as reliable low-cost neural surrogates replacing costly fluid-flow simulations and gradient calculations that rely on numerically expensive PDE solves during inversion.

The organization of this chapter is as follows: First, we introduce FNOs and explore the possibility of replacing the forward modeling operator with a trained FNO surrogate. Next, NFs are introduced. By means of a motivational example, we demonstrate how these learned generative networks control the prediction error of FNOs by ensuring that the model iterates remain in distribution. Based on this motivational example, we propose our novel method for using trained NFs as a learned constraint to guarantee performance of FNO surrogates during inversion. Through four synthetic experiments related to GCS monitoring, the efficacy of our method will be demonstrated.

4.3 Fourier neural operators

There is an extensive literature on training deep neural networks to serve as affordable alternatives to computationally expensive numerical simulators [26, 29, 30, 47, 48]. Without loss of generality, we limit ourselves in this exposition to the training of a special class of neural operators known as Fourier neural operators (FNOs). These FNOs are designed to approximate numerical solution operators of the PDE solution operator, \mathcal{S} , by minimizing

the following objective:

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \quad \frac{1}{N} \sum_{j=1}^N \|\mathcal{S}_{\boldsymbol{\theta}}(\mathbf{K}^{(j)}) - \mathbf{c}^{(j)}\|_2^2 \quad \text{where} \quad \mathbf{c}^{(j)} = \mathcal{S}(\mathbf{K}^{(j)}). \quad (4.2)$$

Here, $\mathcal{S}_{\boldsymbol{\theta}}$ denotes the FNO with network weights $\boldsymbol{\theta}$. The optimization aims to minimize the ℓ_2 misfit between numerically simulated PDE solutions, $\mathbf{c}^{(j)}$, and solutions approximated by the FNO, across N training samples (permeability models), $\{\mathbf{K}^{(j)}\}_{j=1}^N$ compiled by domain experts. Once trained, FNOs can generate approximate PDE solutions for unseen model parameters orders of magnitude faster than numerical simulations [35, 40]. For model parameters that fall within the distribution used to train, approximation by FNOs are reasonably accurate—i.e., $\mathcal{S}_{\boldsymbol{\theta}^*}(\mathbf{K}) \approx \mathcal{S}(\mathbf{K})$, with $\boldsymbol{\theta}^*$ being the minimizer of Equation 4.2. We refer to the numerical examples section for details calculating these weights. Before studying the impact of applying these surrogates on samples for the permeability that are out of distribution, let us first consider an example where data is inverted using surrogate modeling.

4.4 Inversion with learned surrogates

Replacing PDE solutions by approximate solutions yielded by trained FNO surrogates has two main advantages when solving inverse problems. First, as mentioned earlier, FNOs are orders of magnitude faster than numerical PDE solves, which allows for many simulations at negligible costs [49, 50]. Second, existing software for multiphase flow simulations may not always support computationally efficient calculations of sensitivity, e.g. via adjoint-state calculations [51, 52, 53] of the simulations with respect to their input. In such cases, FNO surrogates are favorable because automatic differentiation on the trained network [54, 42, 41, 55, 43] readily provides access to gradients with respect to model parameters. As a result, the PDE solver, \mathcal{S} , in Equation 4.1 can be replaced by trained surrogate, $\mathcal{S}_{\boldsymbol{\theta}^*}$ —i.e., we have

$$\underset{\mathbf{K}}{\text{minimize}} \quad \|\mathbf{d} - \mathcal{H} \circ \mathcal{S}_{\theta^*}(\mathbf{K})\|_2^2 \quad (4.3)$$

where θ^* represent the optimized weights minimizing Equation 4.2. While the above formulation in terms of trained surrogates has been applied successfully during permeability inversion from time-lapse seismic data [56, 41, 43], this type of inversion is only valid as long as the (intermediate) permeabilities remain within distribution during the inversion. Practically, this means two things. First, the data need to be in the range of permeability models that are in distribution. This means that there can not be too much noise neither can the observed data be the result of an out-of-distribution permeability. Second, there are no guarantees that the permeability model iterates remain in distribution during inversion even though some bias of the gradients of the surrogate towards in-distribution permeabilities may be expected. To overcome this challenge, we propose to add a learned constraint to Equation 4.3 that offers guarantees that the model iterates remain in distribution.

4.5 Learned constraints with normalizing flows

As demonstrated by [57, 58, 59], regularization of non-linear inverse problems, such as full-waveform inversion, with constraints, e.g., total-variation [58] or transform-domain sparsity with ℓ_1 -norms [60], offers distinct advantages over regularizations based adding these norms as penalties. Even though constraint and penalty formulations are equivalent for linear inverse problems for the appropriate choice of the Lagrange multiplier, minimizing the constraint formulation leads to completely different solution paths compared to adding a penalty term to the data misfit objective [61]. In the constrained formulation, the model iterates remain at all times within the constraint set while model iterates yielded by the penalty formulation does not offer these guarantees. [57] demonstrated this importance difference for the non-convex problem of full-waveform inversion. For this problem, it proved essential to work with a homotopy where the intersection of multiple handcrafted

constraints (intersection of box and size of total-variation-norm ball constraints) are relaxed slowly during the inversion, so the model iterates remain physically feasible and local minima are avoided.

Motivated by these results, we propose a similar approach but now for “data-driven” learned constraints based on normalizing flows (NFs, [46]). NFs are powerful deep generative neural networks capable of learning to generate samples from complex distributions [62, 63, 64, 65, 43]. Designed to be invertible, these NFs require the latent and model spaces to share identical dimensions, which confers several advantages:

- unlike variational autoencoders [66] or generative adversarial networks (GANs, [67]), which both have a lower-dimensional latent space, NFs do not impose any intrinsic dimensionality constraints. This flexibility lets NFs capture model space characteristics across high dimensions [68]. Relevantly, concurrent literature has delved into the intrinsic dimensionality of NFs, indicating the potential to using NFs to generate models with inherently lower dimensions [69].
- NFs’ inherent invertibility negates the need to store state variables during gradient calculations, enabling memory-efficient training and inversion in large-scale 3D applications, such as in geophysics [70, 71, 72, 73, 74] and ultrasound imaging [75, 64, 76, 65, 77, 78].
- because of their invertibility NFs guarantee unique latent codes for all model space samples, including out-of-distribution ones. Therefore, they can still be used to invert for out-of-distribution model parameters, while other methods like GANs may introduce bias [79].

Aside from being invertible, NFs are trained to map samples from a target distribution in the physical space to samples from the standard zero-mean Gaussian distribution noise in the latent space. After training is completed, samples from the target distribution are generated by running the NF in reverse on samples in the latent space from the standard

Gaussian distribution. Below, we will demonstrate how NFs can be used to guarantee that the permeability remains in distribution during the inversion.

4.5.1 Training normalizing flows

Given samples from the permeability distribution, $\{\mathbf{K}^{(j)}\}_{j=1}^N$, training NFs entails minimizing the Kullback-Leibler divergence between the base and target distributions [80]. This involves solving the following variational problem:

$$\underset{\mathbf{w}}{\text{minimize}} \quad \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{2} \|\mathcal{G}_{\mathbf{w}}^{-1}(\mathbf{K}_i)\|_2^2 - \log |\det J_{\mathcal{G}_{\mathbf{w}}^{-1}}(\mathbf{K}_i)| \right). \quad (4.4)$$

In this optimization problem, $\mathcal{G}_{\mathbf{w}}^{-1}$ represents the NF, which is parameterized by its network weights \mathbf{w} , while $J_{\mathcal{G}_{\mathbf{w}}^{-1}}$ denotes its Jacobian. By minimizing the ℓ_2 -norm, the objective imposes a Gaussian distribution on the network’s output and the second log det term prevents trivial solutions, i.e., cases where $\mathcal{G}_{\mathbf{w}}^{-1}$ produces zeros. To ensure alignment between the permeability distributions, Equation 4.2 and Equation 4.4 are trained on the same dataset consisting of 2000 permeability models examples of which are included in Figure 4.1. Each 64×64 permeability model consists of a randomly generated highly permeable channels (120 mD) in a low-permeable background of 20 mD, where mD denotes millidarcy. Generative examples produced by the trained NF are included in the second row of Figure 4.1, which confirm the NF’s ability to learn distributions from examples. Aside from generating samples from the learned distribution, trained NFs are also capable of carrying out density calculations, an ability we will exploit below.

4.5.2 Trained normalizing flows as constraints

As we mentioned before, adding constraints to the solution of non-convex optimization problems offers guarantees that model iterates remain within constrained sets. When solving inverse problems with learned surrogates, it is important that model iterates remain “in

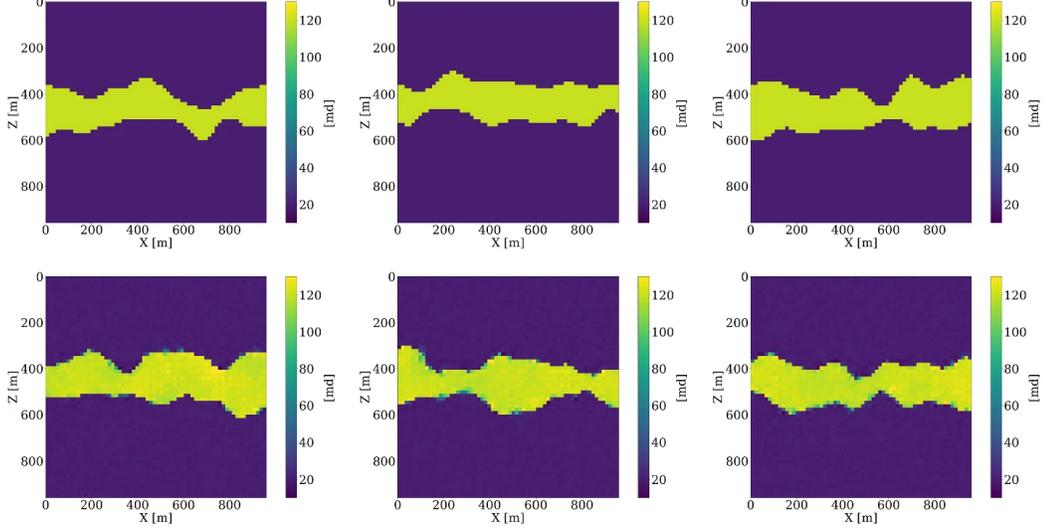


Figure 4.1: Permeability models. First row shows the realistic permeability samples for FNO and NF training. Second row shows the generative samples from the trained NF.

distribution”, which can be achieved by recasting the optimization problem in Equation 4.3 into the following constrained form:

$$\underset{\mathbf{z}}{\text{minimize}} \quad \|\mathbf{d} - \mathcal{H} \circ \mathcal{S}_{\theta^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2^2 \quad \text{subject to} \quad \|\mathbf{z}\|_2 \leq \tau. \quad (4.5)$$

To arrive at this constrained optimization problem, two important changes were made. First, the permeability \mathbf{K} is replaced by the output of a trained NF with trained weights \mathbf{w}^* obtained by minimizing Equation 4.4. This reparameterization in terms of the latent variable, \mathbf{z} , produces permeabilities that are in distribution as long as \mathbf{z} remains distributed according to the standard normal distribution. Second, we added a constraint on this latent space variable in Equation 4.5, which ensures that the latent variable \mathbf{z} remains within an ℓ_2 -norm ball of size τ .

To better understand the behavior of a trained normalizing flow in conjunction with the ℓ_2 -norm constraint for in- and out-of-distribution examples, we include Figure 4.2 and Figure 4.3. In the latter Figure, nonlinear projections (via latent space shrinkage),

$$\tilde{\mathbf{K}} = \mathcal{G}_{\mathbf{w}^*}(\alpha \mathbf{z}) \quad \text{where} \quad \mathbf{z} = \mathcal{G}_{\mathbf{w}^*}^{-1}(\mathbf{K}) \quad (4.6)$$

are plotted as a function of increasing α . We also plot in Figure 4.4 the NF’s relative nonlinear approximation error, $\|\tilde{\mathbf{K}} - \mathbf{K}\|_2/\|\mathbf{K}\|_2$, and the corresponding relative FNO prediction error, $\|\mathcal{S}_{\theta^*}(\tilde{\mathbf{K}}) - \mathcal{S}(\tilde{\mathbf{K}})\|_2/\|\mathcal{S}(\tilde{\mathbf{K}})\|_2$ as a function of increasing $0 \leq \alpha \leq 1$. From these plots, we can make the following observations. First, the latent representations (Figure 4.2c and Figure 4.2d) of the in- and out-of-distribution samples (Figure 4.2a and Figure 4.2b) clearly show that NF applied to out-of-distribution samples produces a latent variable far from the standard normal distribution, while the latent variable corresponding to the in-distribution example is close to being white Gaussian noise. Quantitatively, the ℓ_2 norm of the latent variables are $0.99\|\mathcal{N}(0, \mathbf{I})\|_2$ and $3.11\|\mathcal{N}(0, \mathbf{I})\|_2$, respectively, where $\|\mathcal{N}(0, \mathbf{I})\|_2$ corresponds to the ℓ_2 -norm of the standard normal distribution. Second, we observe from Figure 4.3 that for small ℓ_2 -norm balls ($\alpha \ll 1$) the projected solutions tend to be close to the most probable sample, which is a flat permeability channel in the middle. This is true for both the in- and out-of-distribution example. As α increases, the in-distribution example is reconstructed accurately when the ℓ_2 norm of the scaled latent variable, $\|\alpha\mathbf{z}\|_2$, is close to the $\|\mathcal{N}(0, \mathbf{I})\|_2$. Clearly, this is not the case for the out-of-distribution example. When $\|\alpha\mathbf{z}\|_2 \approx \|\mathcal{N}(0, \mathbf{I})\|_2$, the reconstruction still looks like an in-distribution permeability sample and is not close to the out-of-distribution sample. However, if $\alpha = 1$, which makes $\|\alpha\mathbf{z}\|_2$ well beyond the norm of the standard normal distribution, the out-of-distribution example is recovered accurately by virtue of the invertibility of NFs, irrespective on their input and what they have been trained on. Third, the relative FNO prediction error for the in-distribution example (Figure 4.4a) remains flat while the error of the FNO surrogate increases as soon as $\alpha \approx 0.25$. At that value for α , the projection, $\tilde{\mathbf{K}}$, is gradually transitioning from being in-distribution to out-of-distribution, which occurs at a non-linear approximation error of about 45%. As expected the plots in Figure 4.4 also show a monotonous decay of the nonlinear approximation error as a function of increasing α . To further analyze the effects of the nonlinear projections in Equation 4.6, we draw 50 random realizations from the standard normal distribution, scale each of them by $0 \leq \alpha \leq 2$,

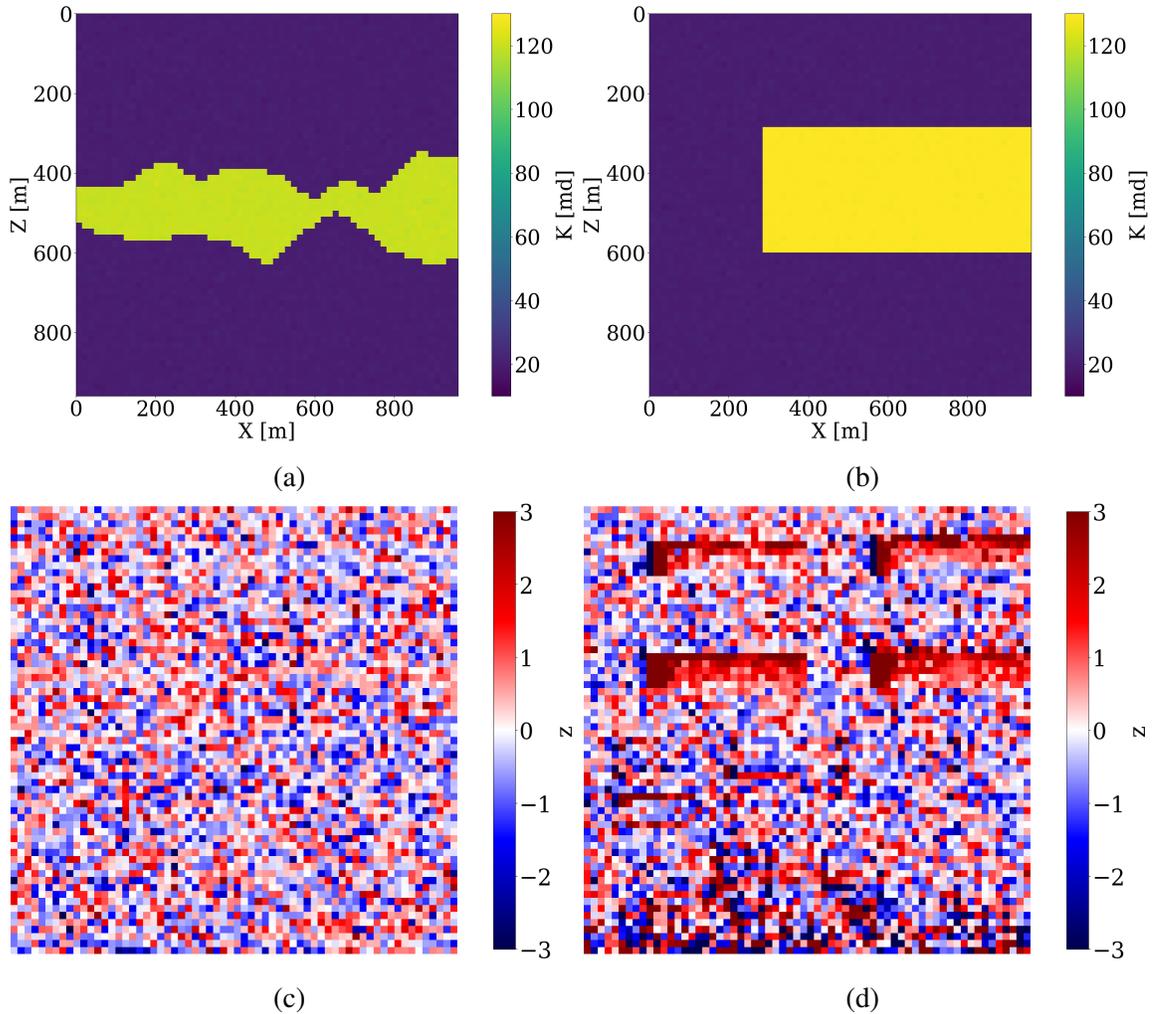


Figure 4.2: Sample permeability models in the physical and latent space. (a) An in-distribution permeability model. (b) An out-of-distribution permeability model. (c) An in-distribution permeability model in the latent space. (d) An out-of-distribution permeability model in the latent space.

and calculate the FNO prediction errors on these samples. Figure 4.5 includes the results of this exercise where each column represents the FNO prediction error calculated for $0 \leq \alpha \leq 2$. From these experiments, we make the following two observations. First, when $\alpha < 0.8$, the FNO consistently makes accurate predictions for all projected samples. Second, as expected, the FNO starts to make less accurate predictions for $\alpha > 1$ with errors that increase as the size of the ℓ_2 -norm ball of the latent space expands, demarcating the transition from being in distribution to being out-of-distribution.

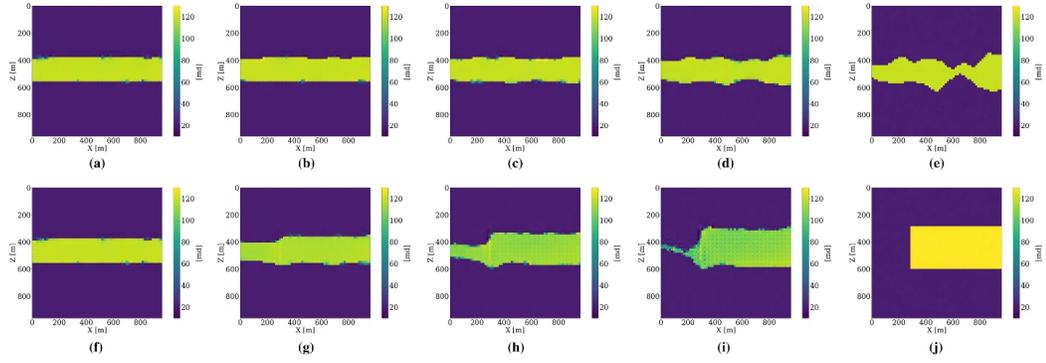


Figure 4.3: Projections onto increasing ℓ_2 -norm balls for the in- and out-of-distribution examples of Figure 4.2. Top row: projections of in-distribution sample. Bottom row: projections of out-of-distribution sample. Each column corresponds to setting $\alpha = 0, 0.1, 0.2, 0.4, 1$ in Equation 4.6.

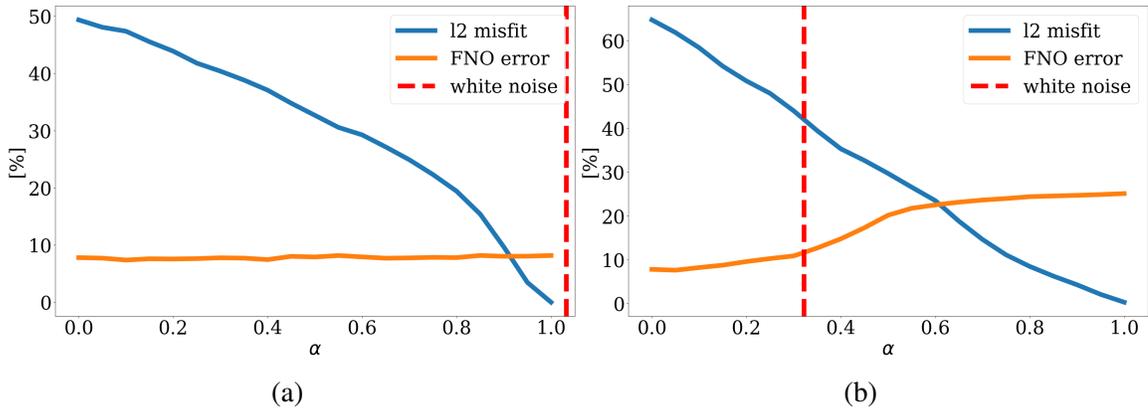


Figure 4.4: Latent space projection experiments. (a) Relative ℓ_2 reconstruction error and FNO prediction error for in-distribution sample. (b) The same but for out-of-distribution sample. The blue curve shows the relative ℓ_2 misfit between the permeability models before and after latent space shrinkage. The orange curve shows the FNO prediction error on the permeability model after shrinking the ℓ_2 -norm ball. The red dashed line denotes the amplitude of standard Gaussian noise.

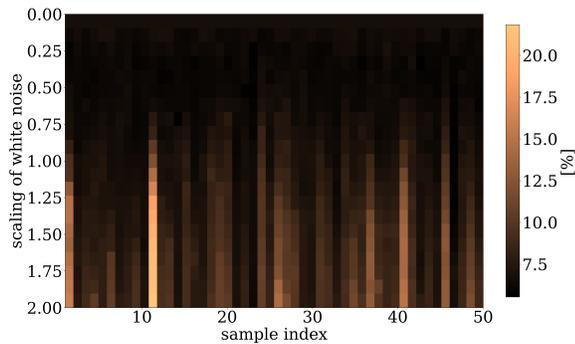


Figure 4.5: FNO prediction errors for the latent space shrinkage experiment in Equation 4.6 for 50 random realizations of standard Gaussian noise.

In summary, the experiments of Figure 4.2 to Figure 4.4 indicate that FNO errors remain small and relatively constant for the in-distribution example. Irrespective of the value of α , the generated samples remain in distribution while moving from the most likely—i.e., a flat high-permeability channel in the middle, to the in-distribution sample as α increases. Conversely, the projection of the out-of-distribution example morphs from being in distribution to being out-of-distribution for $\alpha \geq 0.25$. The FNO prediction errors also increase during this transition from an in-distribution sample to an out-of-distribution sample. Therefore, shrinkage in the latent space by multiplying with a small α can serve as an effective projection that ensures relatively low FNO prediction errors. We will use this unique ability to control the distribution during inversion.

4.5.3 Inversion with progressively relaxed learned constraints

Our main objective is to perform inversions where the multiphase flow equations are replaced with pretrained FNO surrogates. To make sure the learned surrogates remain accurate, we propose working with a continuation scheme where the learned constraint in Equation 4.5 is steadily relaxed by increasing the size of the ℓ_2 -norm ball constraint. Compared to the more common penalty formulation, where regularization entails adding a Lagrange-multiplier weighted ℓ_2 -norm squared, constrained formulations offer guarantees that the model iterates for the latent variable, \mathbf{z} , remain within the constraint set—i.e., within the ℓ_2 -norm ball of size τ . Using the argument of the previous section, this implies that permeability distributions generated by the trained NF remain in distribution as long as the size of the initial ℓ_2 -norm ball, τ_{init} , is small enough (e.g., smaller than $0.6\|\mathcal{N}(0, \mathbf{I})\|_2$, following the observations from Figure 4.5). Taking advantage of this trained NF in a homotopy, we propose Algorithm 1.

Given observed data, \mathbf{d} , trained networks, \mathcal{S}_{θ^*} and $\mathcal{G}_{\mathbf{w}^*}$, the initial guess for the permeability distribution, \mathbf{K}_0 , the initial size of the ℓ_2 -norm ball, τ_{init} , and the final size of the ℓ_2 -norm ball, τ_{final} , Algorithm 1 proceeds by solving a series of constrained optimization

Algorithm 1 Inversion with relaxed learned constraints

- 1: **Input:** initial model parameter $\mathbf{K}_0 \in \mathbb{R}^N$, observed data \mathbf{d} , noise level σ
 - 2: **Input:** trained FNO \mathcal{S}_{θ^*} , trained NF $\mathcal{G}_{\mathbf{w}^*}$
 - 3: **Input:** number of inner-loop iterations $maxiter$
 - 4: **Input:** initial ℓ_2 ball size τ_{init} , multiplier $\beta > 1$, final ℓ_2 ball size τ_{final}
 - 5: $\mathbf{z} = \mathcal{G}_{\mathbf{w}^*}^{-1}(\mathbf{K}_0)$
 - 6: $\tau = \tau_{init}$
 - 7: **while** $\|\mathbf{d} - \mathcal{H} \circ \mathcal{S}_{\theta^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2 > \sigma \|\mathcal{N}(0, \mathbf{I})\|_2$ and $\tau \leq \tau_{final}$ **do**
 - 8: **for** $iter = 1 : maxiter$ **do**
 - 9: $\mathbf{g} = \nabla_{\mathbf{z}} \|\mathbf{d} - \mathcal{H} \circ \mathcal{S}_{\theta^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2^2$
 - 10: $\mathbf{z} = \mathcal{P}_{\tau}(\mathbf{z} - \gamma \mathbf{g})$
 - 11: **end for**
 - 12: $\tau = \beta \tau$
 - 13: **end while**
 - 14: **Output:** inverted model parameter $\mathbf{K} = \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})$
-

problems where the size of the constraint set is increased by a factor of β after each iteration (cf. line 12 in Algorithm 1). The constrained optimization subproblems themselves (cf. line 8 to 11 of Algorithm 1) are solved with projected gradient descent [81]. Each iteration of the projected gradient descent method first calculates the gradient (cf. line 9 of Algorithm 1), followed by the much cheaper projection of the updated latent variable back onto the ℓ_2 -norm ball of size τ via the projection operator \mathcal{P}_{τ} (cf. line 10 in Algorithm 1). This projection is a trivial scaling operation if the updated latent variable ℓ_2 -norm exceeds the constraint — i.e.,

$$\mathcal{P}_{\tau}(\mathbf{z}) = \begin{cases} \mathbf{z} & \text{if } \|\mathbf{z}\|_2 \leq \tau \\ \tau \mathbf{z} / \|\mathbf{z}\|_2 & \text{if } \|\mathbf{z}\|_2 > \tau \end{cases} \quad (4.7)$$

A line search determines the steplength γ [82] for each iteration shown in line 8 to 11. As is common in continuation methods, the relaxed gradient-descent iterations are warm-started with the optimization result from the previous iteration, which at the first iteration is initialized by the latent representation of the initial permeability model, \mathbf{K}_0 (cf. line 5 in Algorithm 1). Practically, each subproblem does not need to be fully solved, but only need a few iterations instead. The number of iterations to solve each subproblem is denoted by

maxiter in line 8 of Algorithm 1. This continuation strategy serves two purposes. First, for small τ 's it makes sure the model iterates remain in distribution, so accuracy of the learned surrogate is preserved. Second, by relaxing the constraint slowly, the data residual is gradually allowed to decrease, bringing in more and more features derived from the data. By slowly relaxing the constraint, we find a careful balance between these two purposes as long as progress is made towards the solution when solving the subproblem (cf. line 8 to 11 in Algorithm 1). One notable distinction of the surrogate-assisted inversion, compared to the conventional inversion with relaxed constraints [58], is that the size of the ℓ_2 -norm projection ball cannot increase far beyond the ℓ_2 -norm of the standard Gaussian white noise on which the NFs are trained. Otherwise, there is no guarantee the learned surrogate is accurate because the NF may generate samples that are out-of-distribution (cf. Figure 4.5). This is explicitly incorporated into the stopping criteria, $\tau \leq \tau_{\text{final}}$, in line 7 of Algorithm 1.

4.6 Numerical Experiments

To showcase the advocacy of the proposed optimization method with relaxed learned constraints, a series of carefully chosen experiments of increasing complexity are conducted. These experiments are designed to be relevant to GCS, which in its ultimate form involves coupling of multiphase flow with the wave equation to perform end-to-end inversion for the permeability given multimodal data. To convince ourselves of the validity of our approach, at all times comparisons will be made between inversion results involving numerical solves of the multiphase equations and inversions yielded by approximations with our learned surrogate.

For all numerical experiments, the “ground-truth” permeability model will be selected from the unseen test set and is shown in Figure 4.6a. The inversions will be initiated with the smooth permeability model depicted in Figure 4.6b. This initial model, \mathbf{K}_0 , represents the arithmetic mean of all permeability samples in the training dataset. To ensure that the model iterates remain in distribution, we set the starting ℓ_2 -norm ball size to

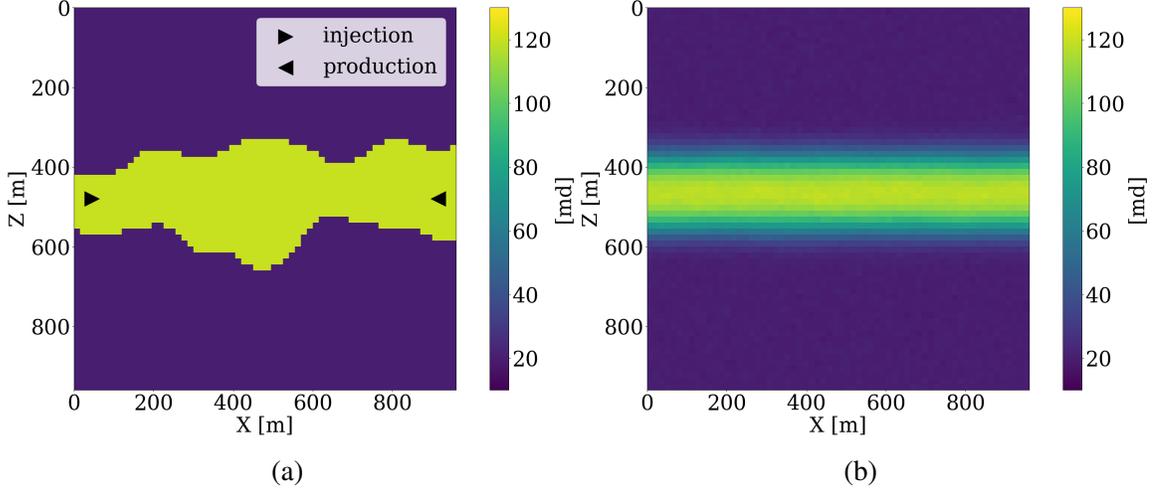


Figure 4.6: Permeability models. (a) unknown “ground-truth” permeability model from unseen test set, where the symbols ► and ◄ denote the CO₂ injection and brine production location, respectively; (b) initial permeability model, \mathbf{K}_0 .

$\tau_{\text{init}} = 0.6 \|\mathcal{N}(0, \mathbf{I})\|_2$ —i.e., $0.6 \times$ the ℓ_2 -norm of standard white Gauss noise realizations for the discrete permeability model of 64 by 64 gridpoints. To gradually relax the learned constraint, the multiplier of the projection ball size is taken to be $\beta = 1.2$, and we set the ultimate projection ball size τ_{final} in Algorithm 1 to be 1.2 times the norm of standard white noise. To limit computational costs of solving the subproblems, we allow each constrained subproblem (cf. line 8 to 11 in Algorithm 1) to perform 8 iterations of projected gradient descent to solve for the latent variable. From practical experience, we found that the proposed inversions are not very sensitive to the choice of these hyperparameters.

To simulate the evolution of injected CO₂ plumes, we make use of the open-source software package Jutul.jl [83, 84, 85], which for each permeability model, $\mathbf{K}^{(j)}$, solves the immiscible and compressible two-phase flow equations for the CO₂ and brine saturation. As shown in Figure 4.6a, an injection well is set up on the left-hand side of the model, which injects supercritical CO₂ with density 700 kg/m³ at a constant rate of 0.005 m³/s. To relieve pressure, a production well is included on the right-hand side of the model, which produces brine with density 1000 kg/m³ with a constant rate of also 0.005 m³/s. This finally results in approximately a 6% storage capacity after 800 days of CO₂

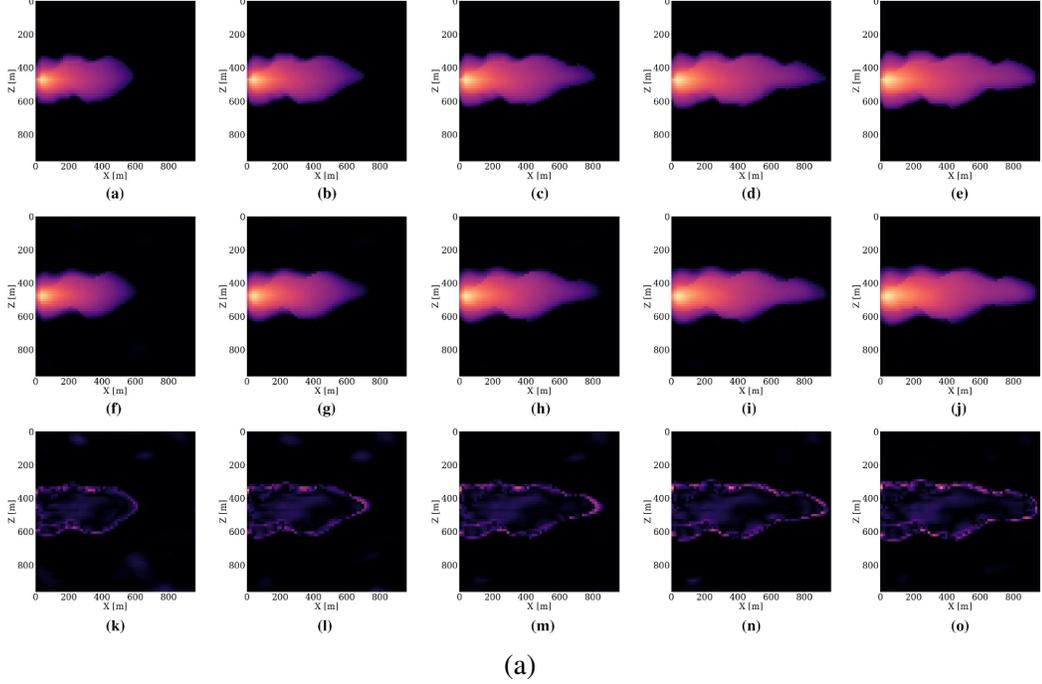


Figure 4.7: Five CO_2 saturation snapshots after 400, 500, 600, 700, and 800 days. First row shows the CO_2 saturation simulated by the PDE. Second row shows the CO_2 saturation predicted by the trained FNO. Third row shows the $5\times$ difference between the first row and the second row.

injection. From these simulations, we collect eight snapshots for the CO_2 concentration, $\mathbf{c} = [c_1, c_2, \dots, c_{n_t}]$ with $n_t = 8$ the number of snapshots that cover a total time period of 800 days. The last five snapshots of these simulations are included in the top row of Figure 4.6a. Due to buoyancy effects and well control, the CO_2 plume gradually moves from the left to the right and upwards.

Given these simulated CO_2 concentrations, the optimized weights, \mathbf{w}^* , for the FNO surrogate are calculated by minimizing Equation 4.2 for $N = 1900$ training pairs, $\{\mathbf{K}^{(j)}, \mathbf{c}^{(j)}\}_{j=1}^N$. Another 100 training pairs are used for validation. After training with 350 epochs, an average of 7% prediction error is achieved for permeability samples from the unseen test set. As observed from Figure 4.7, the approximation errors of the FNO are mostly concentrated at the leading edge of the CO_2 plumes. The same permeability models are used to train the NF by minimizing Equation 4.4 for 245 epochs using the open-source software package InvertibleNetworks.jl [86]. We use the HINT network structure [87] for the NF. Three gen-

erative samples are shown in the second row of Figure 4.1. From these examples, we can see that the trained NF is capable of generating random permeability models that resemble the ones in the training samples closely, despite minor noisy artifacts.

4.6.1 Unconstrained/constrained permeability inversion from CO₂ saturation data

To demonstrate that permeability inversion with surrogates is indeed feasible, we first consider the idealized, impossible in practice, situation where we assume to have access to the time-lapse CO₂ concentration, $\mathbf{c} = [c_1, c_2, \dots, c_{n_t}]$, everywhere, and for all $n_t = 8$ timesteps. In that case, the measurement operator, \mathcal{H} in Equation 4.1, corresponds to the identity matrix. Given CO₂ concentrations simulated from the “ground-truth” permeability distribution plotted in Figure 4.6a, we invert for the permeability by minimizing the unconstrained formulation (cf. Equation 4.3) for the correct, yielded by the PDE, and approximate fluid-flow physics, yielded by the trained FNO. The results of these inversions after 100 iterations of gradient descent with back-tracking linesearch [82] are plotted in Figure 4.8a and Figure 4.8b. From these plots, we observe that the inversion results using PDE solvers delineates most of the upper boundary of the channel accurately. Because there is a null space in the fluid-flow modeling—i.e., this null space mostly corresponds to regions of the permeability model that are barely touched by the CO₂ plume (e.g. bottom and right-hand side of the channel) — artifacts are present in the high-permeability channel itself. As expected, the reconstruction of the permeability is also not perfect at the bottom and at the far right of the model. The inversion result with the FNO surrogate is similar but introduces unrealistic artifacts in the high-permeability channel and also outside the channel. These more severe artifacts can be explained by the behavior of the FNO approximation error plotted as the orange curve in Figure 4.8e. The error value increases rapidly to 13%, and finally saturates at 10%. This behavior of the error is a manifestation of out-of-distribution model iterates that explain the erroneous behavior of the surrogate and its gradient with respect to the permeability.

Inversions yielded by the relaxed constrained formulation with the trained NF (see Algorithm 1), on the other hand, show virtually artifact free inversion results (see Figure 4.8c and Figure 4.8d) that compare favorably with the “ground-truth” permeability plotted in Figure 4.6. While adding the NF as a constraint obviously adds information, explaining the improved inversion for the accurate physics (Figure 4.8c), it also renders the approximate surrogates more accurate, as can be observed from the blue curve in Figure 4.8e, where the FNO approximation error is controlled thanks to adding the constraint to the inversion. This behavior underlines the importance of ensuring model iterates to remain within distribution. It also demonstrates the benefits of a solution strategy where we start with a small τ , followed by relaxing the constraint slowly by increasing the size of the constraint set gradually.

4.6.2 Unconstrained/constrained permeability inversion from well observations

While the example of the previous section established feasibility of constrained permeability inversion, it relied on having access to the CO₂ saturation everywhere, which is unrealistic in practice. To address this issue, we first consider permeability inversion from CO₂ saturations, collected at three equally spaced monitoring well locations, for only the first 6 timesteps over the period of 600 days [88]. In this more realistic setting, the measurement operator, \mathcal{H} in Equation 4.1, corresponds to a restriction operator that extracts simulated CO₂ saturations at each well location in first six snapshots. The objective function reads

$$\underset{\mathbf{z}}{\text{minimize}} \quad \|\mathbf{d}_w - \mathbf{M} \circ \mathcal{S}_{\theta^*} \circ \mathcal{G}_{w^*}(\mathbf{z})\|_2^2 \quad \text{subject to} \quad \|\mathbf{z}\|_2 \leq \tau, \quad (4.8)$$

where \mathbf{d}_w represents the well measurements collected at three well locations through the linear restriction operator \mathbf{M} . The goal is to invert for the permeability by minimizing the misfit of the well measurements of the CO₂ saturation without and with constraints on the ℓ_2 -norm ball in the latent space. The results of these numerical experiments are included in the first row of Figure 4.9, where the differences with respect to the ground truth permeabil-

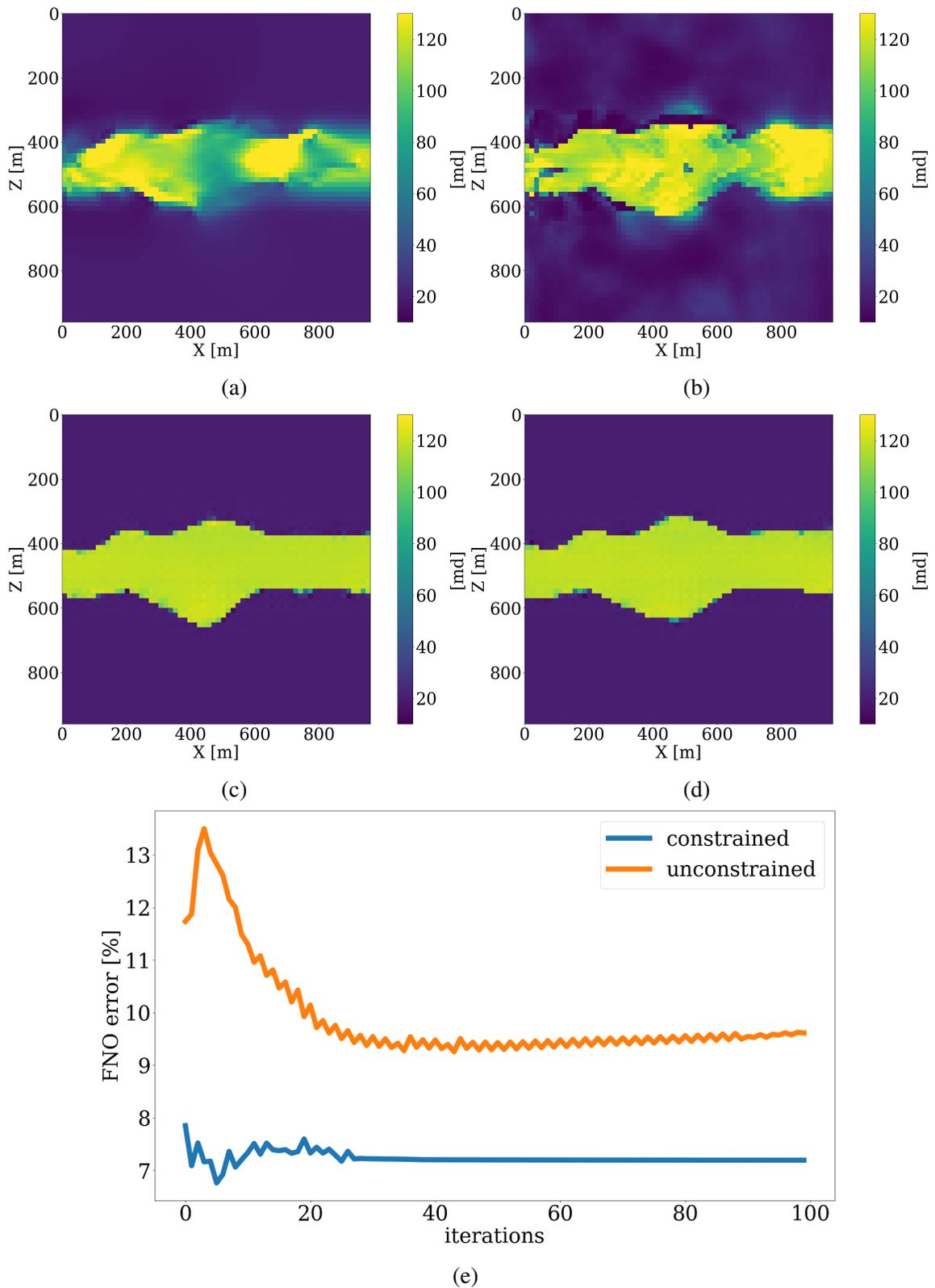


Figure 4.8: Permeability inversion from fully observed time-lapse CO_2 saturations. (a) Inversion result with PDE solvers. (b) The same but via the approximate FNO surrogate. (c) Same as (a) but with NF constraint. (d) Same as (b) but with NF constraint. (e) The FNO approximation errors as a function of the number of iterations for the result plotted in (b) and (d).

ity shown in Figure 4.6a are plotted in the second row. Because the part of the permeability that is not touched by the CO₂ plume lives in the null space, we highlight the CO₂ plume in the difference plots by dark color and focus on analyzing errors within the plume region. As expected, the unconstrained inversions based on PDE solves (Figure 4.9a) and surrogate approximations (Figure 4.9b) are both poorly resolved because of the limited spatial information on the saturation. Contrasting these unconstrained inversions with results for the constrained inversions for the PDE (Figure 4.9c) and surrogate (Figure 4.9d) again shows the importance of adding constraints to the inversion. Figure 4.9i clearly demonstrates that the FNO prediction errors remain relatively constant during constrained inversion while the error continues to grow during the unconstrained iterations eventually exceeding 14%. Both constrained results improve significantly, even though they converge to different solutions in the end. This is because history matching is typically an ill-posed problem with many distinctive solutions [89]. This observation further motivates us to consider the experiment below, where time-lapse seismic data are jointly inverted for the subsurface permeability.

4.6.3 Multiphysics end-to-end inversion

Next, we consider the alternative setting for seismic monitoring of geological carbon storage, where the dynamics of the CO₂ plumes are indirectly observed from time-lapse seismic data. In this case, the measurement operator, \mathcal{H} , involves the composition of the rock physics modeling operator, \mathcal{R} , which converts CO₂ saturations to decreases in the compressional wavespeeds for rocks within the reservoir [90], and the seismic modeling operator, \mathcal{F} , which generates time-lapse seismic data recorded at the receiver locations and based on acoustic wave equation modeling [91]. The multiphysics end-to-end inversion process estimates permeability from time-lapse seismic data via inversion of these nested physics operators for the flow, rock physics, and waves [56]. Following earlier work by [41] and [43], the fluid-flow PDE modeling is replaced by the trained FNO (cf. Equation 4.5), resulting in the following optimization problem:

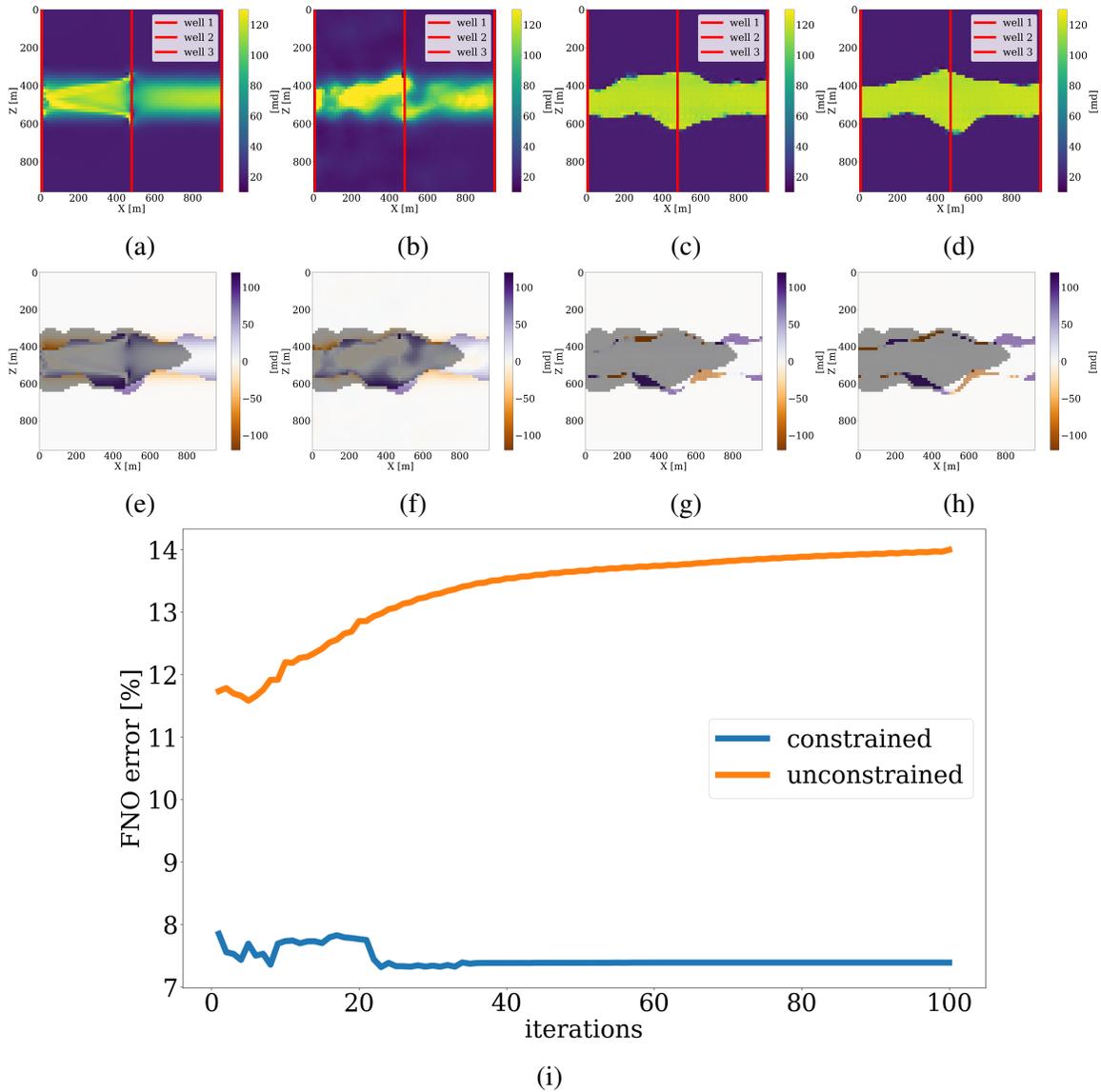


Figure 4.9: Permeability inversions from CO_2 saturations sampled at three well locations at 6 early snapshots. The well locations are denoted by the red vertical lines. (a) Unconstrained inversion result based on PDE solves. (b) Same as (a) but now with FNO surrogate approximation. (c) Constrained inversion result based on PDE solves. (d) Same as (c) but now with FNO surrogate approximation. (e)-(h) The error of the permeability inversion results in (a)-(d) compared to the unseen ground truth shown in Figure 4.6a. (i) The FNO prediction errors as a function of the number of iterations for (b) and (d).

$$\underset{\mathbf{z}}{\text{minimize}} \quad \|\mathbf{d}_s - \mathcal{F} \circ \mathcal{R} \circ \mathcal{S}_{\theta^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2^2 \quad \text{subject to} \quad \|\mathbf{z}\|_2 \leq \tau, \quad (4.9)$$

where \mathbf{d}_s represents the observed time-lapse seismic data. While this end-to-end inversion problem benefits from having remote access to changes in the compressional wavespeed, it may now suffer from null spaces associated with the flow, \mathcal{S}_{θ^*} , and the wave/rock physics, $\mathcal{F} \circ \mathcal{R}$. For instance, the latter suffers from bandwidth limitation of the source function and from limited aperture. Because important components are missing in the observed data, inversion based on the data objective alone in Equation 4.9 are likely to suffer from artifacts that can easily drive the intermediate permeability model iterates out-of-distribution.

To demonstrate capabilities of the proposed relaxed inversion procedure with surrogates for the fluid flow, we assume the baseline to be known—i.e, we assume the brine-filled reservoir with 25% porosity to be acoustically homogeneous prior to CO₂ injection with a compressional wavespeed of 3500m/s. We use the patchy saturation model [90] to convert the time-dependent CO₂ saturation resulting in $< 300\text{m/s}$ decreases in the wavespeed within the CO₂ plumes. We collect six seismic surveys at the first six snapshots for the CO₂ saturation from day 100 to day 600, which are the same snapshots as the ones used in the previous experiment. For each time-lapse seismic survey, 16 active-seismic sources are located within a well on the left-hand side of the model. We also position 16 sources on the top of the model. Each active source uses a Ricker wavelet with a central frequency of 50Hz. The transmitted and reflected wavefields are collected by 480 receivers on the top and 480 receivers on the right-hand side of the model. The seismic acquisition is shown in Figure 4.10, where the plume at the last seismic vintage (at day 600) is plotted in the middle.

To avoid numerical dispersion, the velocity model is upsampled by a factor of two in both the horizontal and vertical directions, which results in a 7.5m grid spacing. For the simulations, use is made of the open-source software package JUDI.jl [92, 43] to generate the time-lapse seismic data at the first six snapshots. The fact that this software is based

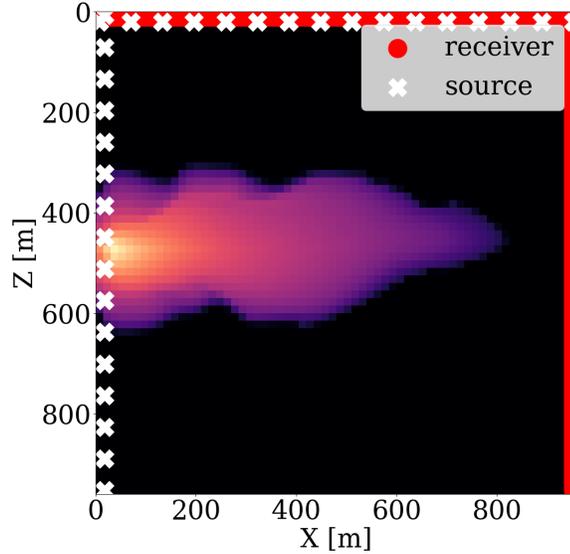


Figure 4.10: Seismic acquisition. The white \times represents the acoustic sources, and the red lines represent the dense receivers. The CO_2 saturation snapshot at day 600 is plotted in the middle, which is the last snapshot that is monitored seismically.

on Devito’s wave propagators [93, 94] allows us to do this quickly. For realism, we add 10 dB Gaussian noise to the time-lapse seismic data. Given these six time-lapse vintages, our goal is to invert for the permeability in the reservoir by minimizing the time-lapse seismic data misfit through the nested physics operators shown in Equation 4.9.

Inversion results obtained by solving the PDEs for the fluid flow during the inversion are shown in Figure 4.11a and Figure 4.11c. As before, the inversions benefit majorly from adding the trained NF as a constraint. Remarkably, the end-to-end inversion results shown in Figure 4.11a, Figure 4.11c, and Figure 4.11d are close to the results plotted in Figure 4.8a, Figure 4.8c, and Figure 4.8d, which was obtained with access to the CO_2 saturation everywhere. This reaffirms the notion that time-lapse seismic can indeed provide useful spatial information away from the monitoring wells to estimate the reservoir permeability, which aligns with earlier observations by [56, 41, 43]. Juxtaposing the results for the FNO surrogate without (Figure 4.11b) and with the constraint (Figure 4.11d) again underlines the importance of adding constraints especially in situations where the forward (wave) operator has a non-trivial nullspace. The presence of this nullspace has a detrimen-

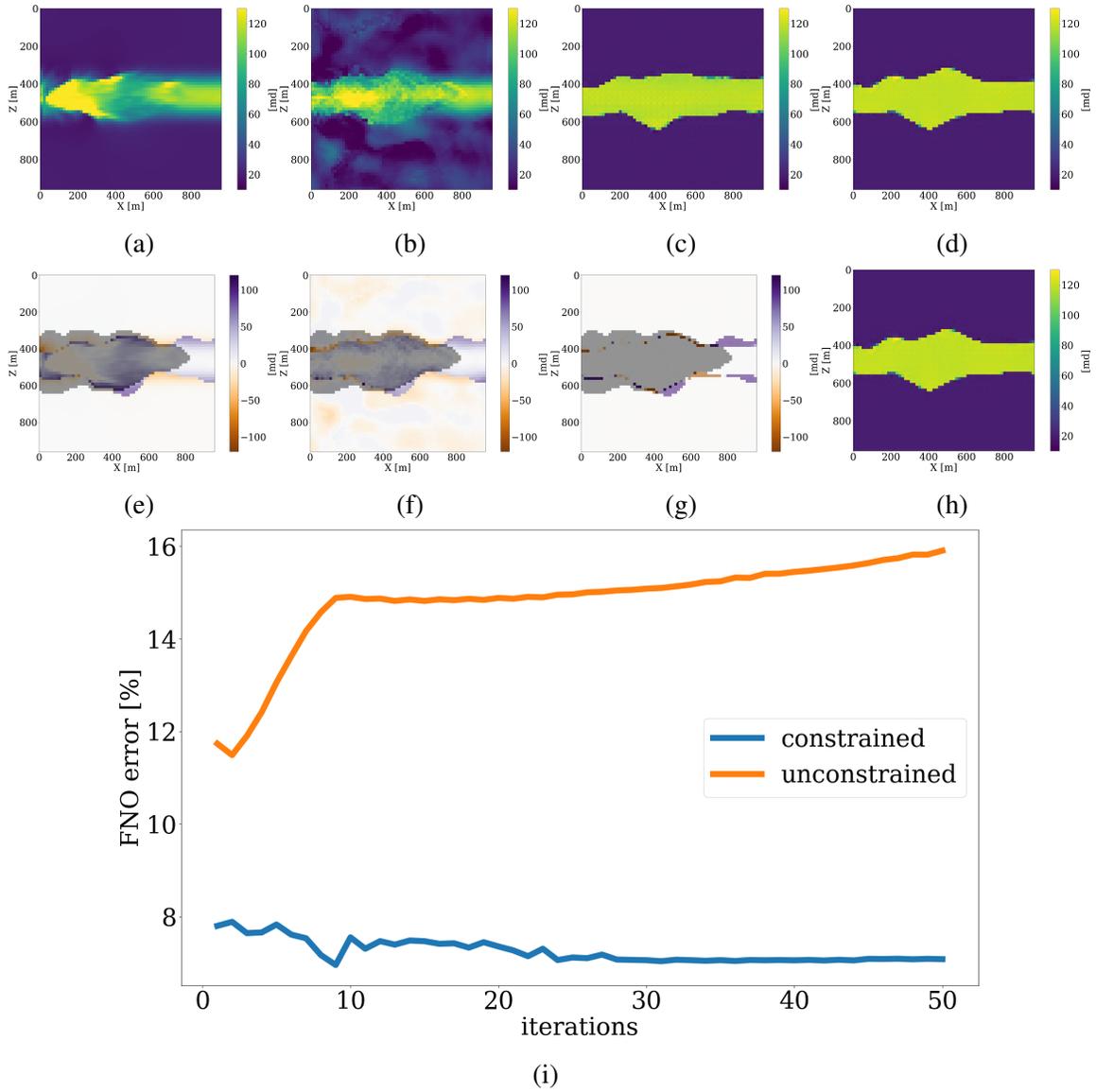


Figure 4.11: Permeability inversions from time-lapse seismic data. (a) Inversion result using PDE solvers. (b) The same as (a) but for the FNO surrogate. (c) The same as (a) but with the NF-based constraint. (d) The same as (a) but now for the FNO surrogate with the NF-based constraint. (e)-(h) The error of the permeability inversion results in (a)-(d) compared to the unseen ground truth shown in Figure 4.6a. (i) The FNO prediction errors as a function of the number of iterations for (b) and (d).

tal affect on the unconstrained result obtained by the FNO. Contrary to solutions yielded by the PDE, trained FNOs offer little to no control on the feasibility of the solution, which explains the strong artifacts in Figure 4.11b. As we can see from Figure 4.11i, these artifacts are mainly due to the FNO-approximation errors that dominate and grow after a few iterations. Conversely, the errors for the constrained case remain more or less flatlined between 7% and 8%. In contrast, using the trained NF as a learned constraint yields better recovery where the errors are minor within the plume region and mostly live on the edges, shown in the second row of Figure 4.11.

4.6.4 Jointly inverting time-lapse seismic data and well measurements

Finally, we consider the most preferred scenario for GCS monitoring, where multiple modalities of data are jointly inverted for the reservoir permeability [95, 96]. In our experiment, we consider to jointly invert time-lapse seismic data and well measurements by minimizing the following objective function:

$$\underset{\mathbf{z}}{\text{minimize}} \quad \|\mathbf{d}_s - \mathcal{F} \circ \mathcal{R} \circ \mathcal{S}_{\theta^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2^2 + \lambda \|\mathbf{d}_w - \mathbf{M} \circ \mathcal{S}_{\theta^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z})\|_2^2 \quad \text{subject to} \quad \|\mathbf{z}\|_2 \leq \tau. \quad (4.10)$$

This objective function includes both the time-lapse seismic data misfit from Equation 4.9 and the time-lapse well measurement misfit from Equation 4.8 with a balancing term λ . While better choices can be made, we select this λ in our numerical experiment to be 10, so that the magnitudes of the two terms are relatively the same. The inversion results and differences from the unseen ground truth permeability are shown in Figure 4.12, where we again observe large artifacts for the recovery when FNO surrogate is inverted without NF constraints. This behavior is confirmed by the plot for the FNO error curve as a function of the number of iterations. This error finally reaches a value over 15%.

We report quantitative measures for the permeability inversions for all optimization methods and different types of observed data in Table 4.1 for the signal-to-noise ratios

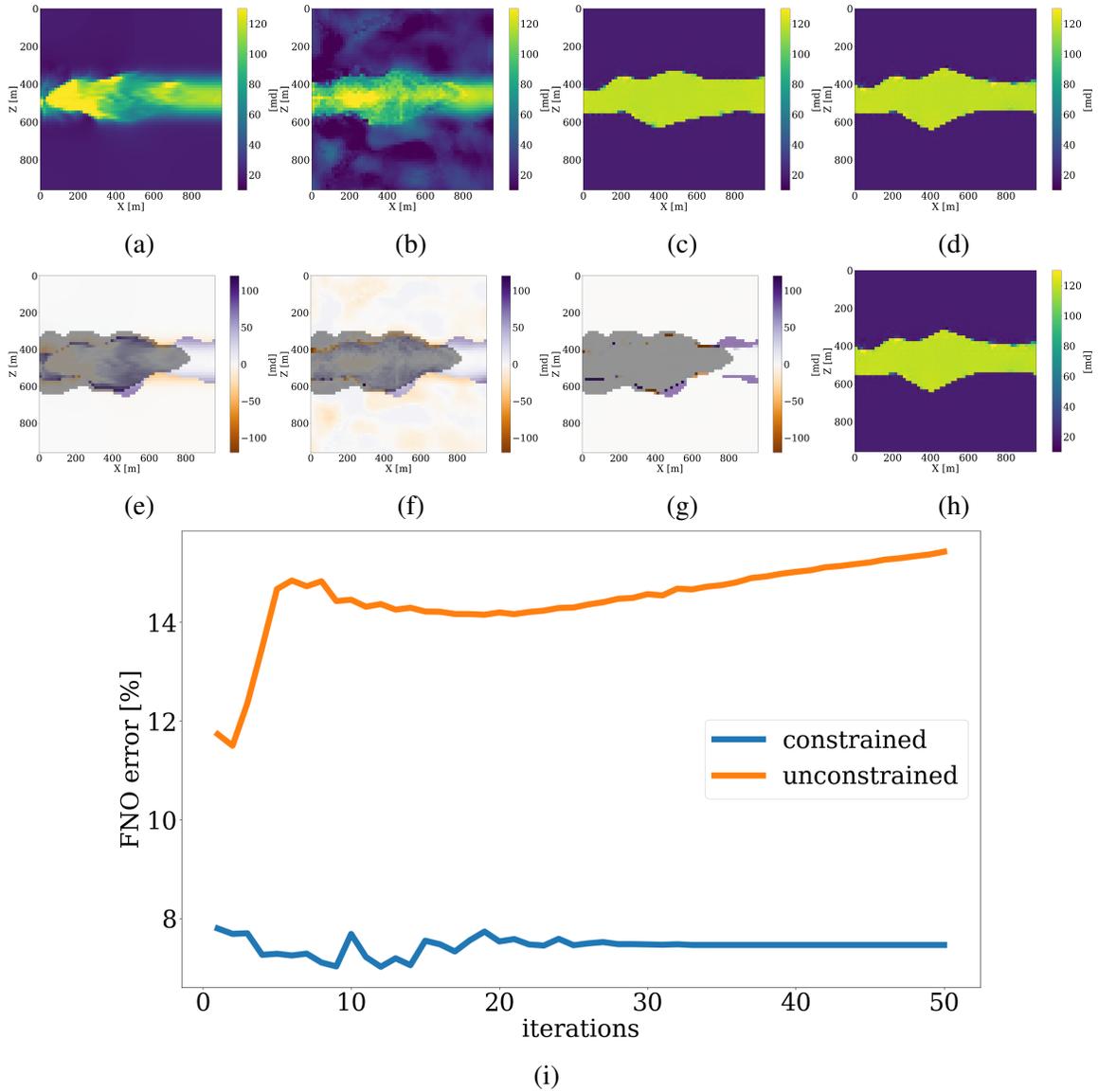


Figure 4.12: Joint permeability inversions from both time-lapse seismic data and time-lapse well measurements. (a) Inversion result using PDE solvers. (b) The same as (a) but for the FNO surrogate. (c) The same as (a) but with the NF-based constraint. (d) The same as (a) but now for the FNO surrogate with the NF-based constraint. (e)-(h) The error of the permeability inversions in (a)-(d), compared to the unseen ground truth shown in Figure 4.6a. (i) The FNO prediction errors as a function of the number of iterations for (b) and (d).

(S/Ns) and the structural similarity index measure (SSIM, [97]). To avoid undue influence of the null space for the permeability, we only calculate the S/N and SSIM values based on the parts of the models that are touched by CO₂ plume. From these values, following observations can be made. First, the NF-constrained permeability inversion are superior in both S/Ns and SSIMs, which demonstrates the efficacy of the learned constraint. Second, by virtue of this NF constraint, the results yielded by either the PDE solver or by the FNO surrogate produce very similar S/Ns and SSIMs. This behavior reaffirms that the trained FNO behavior is similar to the behavior yielded by PDE solver when its inputs remain in-distribution, which is controlled by the NF constraints.

Table 4.1: S/N (in dB) and SSIM values of permeability recovery.

Inversion method	Well	Time-lapse	
	measurement	seismic	Both
Unconstrained inversion with PDE solvers	9.34 dB / 0.67	10.50 dB / 0.73	10.70 dB / 0.73
Unconstrained inversion with FNO surrogates	9.64 dB / 0.68	11.94 dB / 0.72	11.98 dB / 0.72
Constrained inversion with PDE solvers	12.2 dB / 0.77	14.18 dB / 0.80	15.20 dB / 0.85
Constrained inversion with FNO surrogates	11.06 dB / 0.74	14.16 dB / 0.81	14.92 dB / 0.83

4.6.5 CO₂ plume estimation and forecast

While end-to-end permeability inversion from time-lapse data provides novel access to this important fluid-flow property, the real interest in monitoring GCS lies in determining where CO₂ plumes are and will be in the foreseeable future, say of 100 and 200 days ahead. To demonstrate the value of the proposed surrogates and of the use of time-lapse seismic

data, as opposed to time-lapse saturation data measured at the wells only, we in Figure 4.7 juxtapose CO₂ predictions obtained from fluid-flow simulations based on the inverted permeabilities in situations where either well data is available (first row), or where time-lapse seismic data is available (second row), or where both data modalities are available (third row). These results are achieved by first inverting for permeabilities using FNO surrogates and NF constraints, followed by running the fluid-flow simulations for additional time steps given the inverted permeabilities yielded by well-only (Figure 4.9d), time-lapse data (Figure 4.11d), and both (Figure 4.12d). From these plots, we draw the following two conclusion. First, the predicted CO₂ plumes estimated from seismic data are significantly more accurate than those obtained by inverting time-lapse saturations measured at the wells only. As expected, there are large errors in the regions away from the wells for the CO₂ plumes estimated from wells shown in the fourth row of Figure 4.13. Second, thanks to the NF-constraint, the CO₂ predictions obtained with the computationally beneficial surrogate approximation remain close to the ground truth CO₂ plume plotted in the first row of Figure 4.7, with only minor artifacts at the edges. Third, using both seismic data and well measurements produces CO₂ plume predictions with the smallest errors, while the uplift of well measurements on top of seismic observations is modest (comparing the second and the third rows of Figure 4.13). Finally, while the CO₂ plume estimates for the past (monitored) vintages (i.e. first three columns of the third row of Figure 4.13) are accurate, the near-future forecasts without time-lapse well or seismic data (i.e. last two columns of the third row of Figure 4.13) could be less accurate. This is because the right-hand side and the bottom of the permeability model are not touched yet by the CO₂ plume during the first 600 days. As a result, the error on the permeability recovery on the right-hand side leads to the slightly larger errors on the CO₂ plume forecast. Overall, these CO₂ forecasts for the future 100 and 200 days match the general trend of the CO₂ plume without any observed data despite minor errors. A continuous monitoring system, where multiple modalities of data are being acquired and inverted throughout the GCS project, could allow for updating

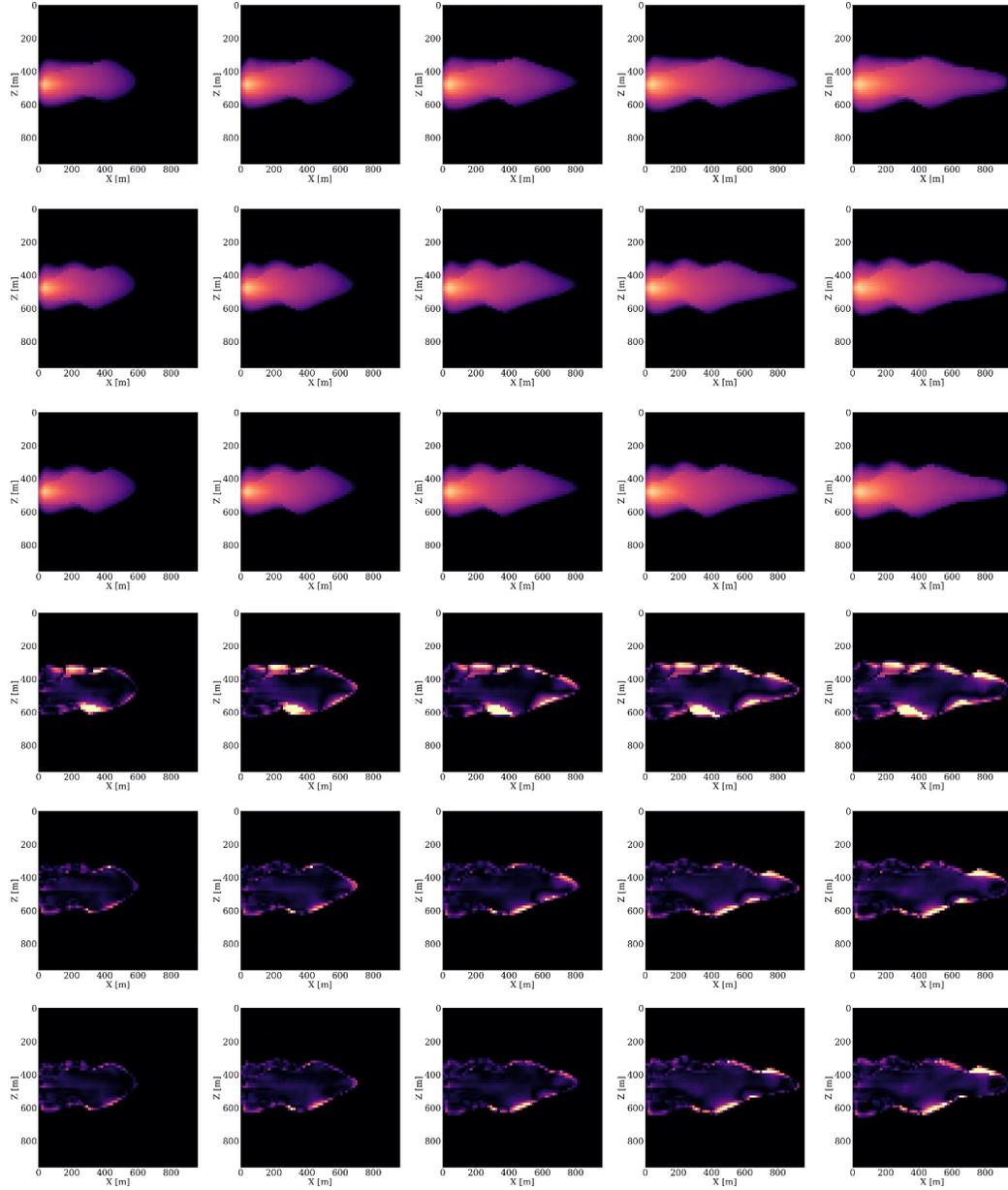
the reservoir permeability and forecasting the CO₂ plume consistently.

4.6.6 Analysis of computational gains

FNOs, and deep neural surrogates in general, have the potential to be orders of magnitude faster than conventional PDE solvers [38], and this speed-up is generally problem-dependent. In our numerical experiments, the PDE solver from Jutul.jl [83, 84, 85] currently only supports CPUs and we find an average runtime for both the forward and gradient on the 64×64 model to be 10.6 seconds on average on an 8-core Intel(R) Xeon(R) W-2245 CPU. The trained FNO, implemented using modules from Flux.jl [98], takes 16.4 milliseconds on average for both the forward and gradient. This means that the trained FNO in our case provides $646\times$ speed up compared to conventional PDE solvers. The training of FNO takes about 4 hours on an NVIDIA T1000 8GB GPU. Given these numbers, we can calculate the break-even point — i.e., the point where using FNO surrogate becomes cheaper in terms of the overall runtime, by the following formula:

$$\text{breakeven} = \frac{\text{generating training set time} + \text{training time}}{\text{PDE solver runtime} - \text{FNO runtime}} \approx 3364. \quad (4.11)$$

This means that after 3364 calls to the forward simulator and its gradients, the computational savings gained from using the FNO surrogate evaluations during the inversion process balances out the initial upfront costs. These upfront costs include the generation of the training dataset and the training of the FNO. Therefore, after this break-even point of 3364 calls, the use of the FNO surrogate becomes more cost-effective compared to the conventional PDE solver. Because the trained FNO has the potential to generalize to different kinds of inversion problems, and potentially also different GCS sites, 3364 calls is justifiable in practice. However, we acknowledge that a more detailed analysis on a more realistic 4D scale problem will be necessary to understand the potential computational gains and tradeoffs of the proposed methodology. For details on a high-performance computing parallel implementation of FNOs, we refer to [35] who also conducted a real-



(a)

Figure 4.13: CO₂ plume estimation and forecast using FNO surrogates and NF constraints to invert different modalities of observed data. The first three columns represent past CO₂ saturations at day 400, 500, and 600 of the first 600 days of CO₂ saturation monitored either through the well measurements or time-lapse data. The last two columns include forecasts for the saturations at future days 700 and 800, where no observed data is available. The first row shows the past and future CO₂ estimates yielded by inverting well measurements only. The second row is the same but now inverting time-lapse seismic data. The third row is the same but now jointly inverting well measurements and time-lapse seismic data. The fourth, fifth, and sixth rows show $5\times$ difference between the ground truth CO₂ plume (first row of Figure 4.7) and the first, second, third row, respectively. The S/Ns for the first, the second, and the third rows are 15.26 dB, 20.14 dB, 20.46 dB, respectively.

istic performance on large-scale 4D multiphase fluid-flow problems. Even in cases where the computational advances are perhaps challenging to justify, the use of FNOs has the additional benefit by providing access to the gradient with respect to model parameters (i.e. permeability) through automatic differentiation. This feature is important since it is an enabler for inversion problems that involve complex PDE solvers for which gradients are often not readily available, e.g. [45]. By training FNOs on input-output pairs, “gradient-free” gradient-based inversion is made possible in situations where the simulator does not support gradients.

4.7 Discussion and conclusions

Monitoring of geological carbon storage is challenging because of excessive computational needs and demands on data collection by drilling monitor wells or by collecting time-lapse seismic data. To offset the high computational costs of solving multiphase flow equations and to improve permeability inversions from possibly multimodal time-lapse data, we introduce the usage of trained Fourier neural operators (FNOs) that act as surrogates for the fluid-flow simulations. We propose to do this in combination with trained normalizing flows (NFs), which serve as regularizers to keep the inversion and the accuracy of the FNOs in check. Since the computational expense of FNO’s online evaluation is negligible compared to numerically expensive partial differential equation solves, FNOs only incur upfront offline training costs. While this obviously presents a major advantage, the approximation accuracy of FNOs is, unfortunately, only guaranteed when its argument, the permeability, is in distribution—i.e., is drawn from the same distribution as the FNO was trained on. This creates a problem because there is, thanks to the non-trivial null space of permeability inversion, no guarantee the model iterates remain in-distribution. Quite the opposite, our numerical examples show that these iterates typically move out-of-distribution during the (early) iterations. This results in large errors in the FNO and in rather poor inversion results for the permeability.

To overcome this out-of-distribution dilemma for the model iterates during permeability inversion with FNOs, we propose adding learned constraints, which ensure that model iterates remain in-distribution during the inversion. We accomplish this by training a NF on the same training set for the permeability used to train the FNO. After training, the NF is capable of generating in-distribution samples for the permeability from random realizations of standard Gaussian noise in the latent space. We employ this learned ability by parameterizing the unknown permeability in the latent space, which offers additional control on whether the model iterates remain in-distribution during the inversion. After establishing that out-of-distribution permeability models can be mapped to in-distribution models by restricting the ℓ_2 -norm of their latent representation, we introduce permeability inversion as a constrained optimization problem where the data misfit is minimized subject to a constraint on the ℓ_2 -norm of the latent space. Compared to adding pretrained NFs as priors via additive penalty terms, use of constraints ensures that model iterates remain at all times in-distribution. We show that this holds as long as the size of constraint set does not exceed the size of the ℓ_2 -norm ball of the standard normal distribution. As a result, we arrive at a computationally efficient continuation scheme, known as a homotopy, during which the ℓ_2 -norm constraint is relaxed slowly, so the data misfit objectives can be minimized while the model iterates remain in distribution.

By means of a series of carefully designed numerical experiments, we were able to establish the advocacy of combining learned surrogates and constraints, yielding solutions to permeability inversion problems that are close to solutions yielded by costly PDE-based methods. The examples also clearly show the advantages of working with gradually relaxed constraints where model iterates remain at all times in distribution with the additional joint benefit of slowly building up the model while bringing down the data misfit, an approach known to mitigate the effects of local minima [57, 58, 59]. Consequently, the quality of all time-lapse inversions improved significantly without requiring information that goes beyond having access to the training set of permeability models.

While we applied the proposed method to gradient-based iterative inversion, a similar approach can be used for other types of inversion methods, including inference with Markov chain Monte Carlo methods for uncertainty quantification [50]. We also envisage extensions of the proposed method to other physics-based inverse problems [99, 55] and simulation-based inference problems [100], where numerical simulations often form the computational bottleneck.

Despite the encouraging results from the numerical experiments, the presented approach leaves room for improvements, which we will leave for future work. For instance, the gradient with respect to the model parameters (permeability) derived from the neural surrogate is not guaranteed to be accurate—e.g. close to the gradient yielded by the adjoint-state method. As recent work by [101] has shown, this potential source of error can be addressed by training neural surrogates on the simulator’s gradient with respect to the model parameters, provided it is available. Unfortunately, deriving gradients of complex HPC implementations of numerical PDE solvers is often extremely challenging, explaining why this information is often not available. Because our method solely relies on gradients of the surrogate, which are readily available through algorithmic differentiation, we only need access to numerical PDE solvers available in legacy codes. While this approach may go at the expense of some accuracy, this feature offers a distinct practical advantage. However, as with many other machine learning approaches, our learned methods may also suffer from time-lapse observations that are out-of-distribution—i.e., produced by a permeability model that is out-of-distribution. While this is a common problem in data-driven methods, recent developments [63] may remedy this problem by applying latent space corrections, a solution that is amenable to our approach. On the other hand, expanding the latent space’s ℓ_2 norm ball during inversion would allow NFs to generate any out-of-distribution model parameter. However, in that case the accuracy of the learned surrogate is not guaranteed. For such cases, transitioning from the learned surrogate to the numerical solver during later iterations may be advantageous and merits further study. The choice for the size of the

ℓ_2 -norm ball at the beginning and at the end can also be further investigated [102].

While this chapter primarily presents a proof of concept through a relatively small 2D experiment, our inversion strategy is designed to scale to large-scale 3D problems. NFs, with their inherent memory efficiency due to invertibility, are already primed for extension to 3D problems. For the learned surrogates, [35] showcases model-parallel FNOs, demonstrating success in simulating 4D multiphase flow physics of over 2.6 billion variables. By combining these strengths, we are optimistic scaling this inversion strategy to 3D.

To end on a positive note and forward looking note, we argue that the presented approach makes a strong case for the inversion of multimodal data, consisting of time-lapse well and seismic data. While inversions from time-lapse saturation data collected from wells are feasible and fall within the realm of reservoir engineering, their performance, as expected, degrades away from the well. We argue that adding active-source seismic provides essential fill-in away from the wells. As such, it did not come to our surprise that joint inversion of multimodal data resulted in the best permeability estimates. From our perspective, our successful combination of these often disjoint data modalities holds future promise when addressing challenges that come with monitoring and control of geological carbon storage and enhanced geothermal systems.

4.8 Availability of data and materials

The scripts to reproduce the experiments are available on the SLIM GitHub page <https://github.com/slimgroup/FNO-NF.jl>.

4.9 References

- [1] A. Tarantola, “Inversion of seismic reflection data in the acoustic approximation,” *Geophysics*, vol. 49, no. 8, pp. 1259–1266, 1984 (page 74).
- [2] A. Tarantola, *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005 (pages 74, 75).
- [3] S. R. Arridge, “Optical tomography in medical imaging,” *Inverse problems*, vol. 15, no. 2, R41, 1999 (page 74).
- [4] A. Alexanderian, “Optimal experimental design for infinite-dimensional bayesian inverse problems governed by pdes: A review,” *Inverse Problems*, vol. 37, no. 4, p. 043 001, 2021 (page 74).
- [5] A.-K. Furre, O. Eiken, H. Alnes, J. N. Vevatne, and A. F. Kiær, “20 years of monitoring CO₂-injection at sleipner,” *Energy Procedia*, vol. 114, pp. 3916–3926, 2017, 13th International Conference on Greenhouse Gas Control Technologies, GHGT-13, 14-18 November 2016, Lausanne, Switzerland (page 74).
- [6] J. M. Nordbotten and M. A. Celia, “Geological storage of CO₂: Modeling approaches for large-scale simulation,” in *Geological Storage of CO₂: Modeling Approaches for Large-Scale Simulation*, John Wiley and Sons, 2011 (page 74).
- [7] A. F. Rasmussen *et al.*, “The open porous media flow reservoir simulator,” *Computers & Mathematics with Applications*, vol. 81, pp. 159–185, 2021 (pages 74, 76).
- [8] B. M. Freifeld, T. M. Daley, S. D. Hovorka, J. Henniges, J. Underschultz, and S. Sharma, “Recent advances in well-based monitoring of co₂ sequestration,” *Energy Procedia*, vol. 1, no. 1, pp. 2277–2284, 2009 (page 74).
- [9] J. P. Noguees, J. M. Nordbotten, and M. A. Celia, “Detecting leakage of brine or co₂ through abandoned wells in a geological sequestration operation using pressure monitoring wells,” *Energy Procedia*, vol. 4, pp. 3620–3627, 2011 (page 74).
- [10] S. L. Nooner, O. Eiken, C. Hermanrud, G. S. Sasagawa, T. Stenvold, and M. A. Zumberge, “Constraints on the in situ density of co₂ within the utsira formation from time-lapse seafloor gravity measurements,” *international journal of greenhouse gas control*, vol. 1, no. 2, pp. 198–214, 2007 (page 74).

- [11] H. Alnes, O. Eiken, S. Nooner, G. Sasagawa, T. Stenvold, and M. Zumberge, “Results from sleipner gravity monitoring: Updated density and temperature distribution of the co2 plume,” *Energy Procedia*, vol. 4, pp. 5504–5511, 2011 (page 74).
- [12] J. M. Carcione, D. Gei, S. Picotti, and A. Michelini, “Cross-hole electromagnetic and seismic modeling for co2 detection and monitoring in a saline aquifer,” *Journal of petroleum science and engineering*, vol. 100, pp. 162–172, 2012 (page 74).
- [13] M. S. Zhdanov *et al.*, “Electromagnetic monitoring of co2 sequestration in deep reservoirs,” *first break*, vol. 31, no. 2, 2013 (page 74).
- [14] R. Arts, O. Eiken, A. Chadwick, P. Zweigel, L. Van der Meer, and B. Zinszner, “Monitoring of co2 injected at sleipner using time-lapse seismic data,” *Energy*, vol. 29, no. 9-10, pp. 1383–1392, 2004 (page 75).
- [15] D. Lumley, “4d seismic monitoring of co 2 sequestration,” *The Leading Edge*, vol. 29, no. 2, pp. 150–155, 2010 (page 75).
- [16] Z. Yin, H. T. Erdinc, A. P. Gahlot, M. Louboutin, and F. J. Herrmann, “Derisking geologic carbon storage from high-resolution time-lapse seismic to explainable leakage detection,” *The Leading Edge*, vol. 42, no. 1, pp. 69–76, 2023 (page 75).
- [17] G. H. Golub, P. C. Hansen, and D. P. O’Leary, “Tikhonov regularization and total least squares,” *SIAM journal on matrix analysis and applications*, vol. 21, no. 1, pp. 185–194, 1999 (page 75).
- [18] B. Peters and F. J. Herrmann, “Algorithms and software for projections onto intersections of convex and non-convex sets with applications to inverse problems,” *arXiv preprint arXiv:1902.09699*, 2019 (page 75).
- [19] D. C. Liu and J. Nocedal, “On the limited memory bfgs method for large scale optimization,” *Mathematical programming*, vol. 45, no. 1-3, pp. 503–528, 1989 (page 75).
- [20] M. K. Cowles and B. P. Carlin, “Markov chain monte carlo convergence diagnostics: A comparative review,” *Journal of the American statistical Association*, vol. 91, no. 434, pp. 883–904, 1996 (page 75).
- [21] S. Razavi, B. A. Tolson, and D. H. Burn, “Review of surrogate modeling in water resources,” *Water Resources Research*, vol. 48, no. 7, 2012 (page 75).
- [22] M. J. Asher, B. F. Croke, A. J. Jakeman, and L. J. Peeters, “A review of surrogate models and their application to groundwater modeling,” *Water Resources Research*, vol. 51, no. 8, pp. 5957–5973, 2015 (page 75).

- [23] M. J. Powell, “Radial basis functions for multivariable interpolation: A review.” *Algorithms for the Approximation of Functions and Data.*, 1985 (page 75).
- [24] W. H. Schilders, H. A. Van der Vorst, and J. Rommes, *Model order reduction: theory, research aspects and applications*. Springer, 2008, vol. 13 (page 75).
- [25] K. Lu *et al.*, “Review for order reduction based on proper orthogonal decomposition and outlooks of applications in mechanical systems,” *Mechanical Systems and Signal Processing*, vol. 123, pp. 264–297, 2019 (page 75).
- [26] L. Lu, P. Jin, and G. E. Karniadakis, “Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators,” *arXiv preprint arXiv:1910.03193*, 2019 (pages 75, 76).
- [27] R. Pestourie, Y. Mroueh, T. V. Nguyen, P. Das, and S. G. Johnson, “Active learning of deep surrogates for pdes: Application to metasurface design,” *npj Computational Materials*, vol. 6, no. 1, p. 164, 2020 (page 75).
- [28] E. Qian, B. Kramer, B. Peherstorfer, and K. Willcox, “Lift & learn: Physics-informed machine learning for large-scale nonlinear dynamical systems,” *Physica D: Non-linear Phenomena*, vol. 406, p. 132 401, 2020 (page 75).
- [29] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, “Physics-informed machine learning,” *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021 (pages 75, 76).
- [30] N. Kovachki *et al.*, “Neural operator: Learning maps between function spaces with applications to pdes,” *Journal of Machine Learning Research*, vol. 24, no. 89, pp. 1–97, 2023 (pages 75, 76).
- [31] M. A. Rahman, Z. E. Ross, and K. Azizzadenesheli, “U-no: U-shaped neural operators,” *arXiv preprint arXiv:2204.11127*, 2022 (page 75).
- [32] S. Hijazi, M. Freitag, and N. Landwehr, “Pod-galerkin reduced order models and physics-informed neural networks for solving inverse problems for the navier–stokes equations,” *Advanced Modeling and Simulation in Engineering Sciences*, vol. 10, no. 1, pp. 1–38, 2023 (page 75).
- [33] G. Wen, Z. Li, K. Azizzadenesheli, A. Anandkumar, and S. M. Benson, “U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow,” *Advances in Water Resources*, vol. 163, p. 104 180, 2022 (page 75).
- [34] G. Wen, Z. Li, Q. Long, K. Azizzadenesheli, A. Anandkumar, and S. Benson, “Real-time high-resolution co2 geological storage prediction using nested fourier neural operators,” *Energy & Environmental Science*, 2023 (page 75).

- [35] T. J. Grady *et al.*, “Model-parallel fourier neural operators as learned surrogates for large-scale parametric pdes,” *Computers & Geosciences*, vol. 178, p. 105 402, 2023 (pages 75, 77, 102, 107).
- [36] P. A. Witte, W. Redmond, R. J. Hewett, and R. Chandra, “Industry-scale co2 flow simulations with model-parallel fourier neural operators,” in *NeurIPS 2022 Workshop Tackling Climate Change with Machine Learning*, 2022 (page 75).
- [37] P. A. Witte, R. J. Hewett, K. Saurabh, A. Sojoodi, and R. Chandra, “Sciai4industry—solving pdes for industry-scale problems with deep learning,” *arXiv preprint arXiv:2211.12709*, 2022 (page 75).
- [38] Z. Li *et al.*, *Fourier neural operator for parametric partial differential equations*, 2020. arXiv: 2010.08895 [cs.LG] (pages 75, 102).
- [39] Z. Li *et al.*, “Physics-informed neural operator for learning partial differential equations,” *arXiv preprint arXiv:2111.03794*, 2021 (page 75).
- [40] M. De Hoop, D. Z. Huang, E. Qian, and A. M. Stuart, “The cost-accuracy trade-off in operator learning with neural networks,” *arXiv preprint arXiv:2203.13181*, 2022 (pages 75, 77).
- [41] Z. Yin, A. Siahkoochi, M. Louboutin, and F. J. Herrmann, “Learned coupled inversion for carbon sequestration monitoring and forecasting with fourier neural operators,” in *International Meeting for Applied Geoscience and Energy Expanded Abstracts*, (IMAGE, Houston), May 2022 (pages 75, 77, 78, 93, 96).
- [42] M. Louboutin *et al.*, “Accelerating innovation with software abstractions for scalable computational geophysics,” in *International Meeting for Applied Geoscience & Energy Expanded Abstracts*, (IMAGE, Houston), May 2022 (pages 75, 77).
- [43] M. Louboutin *et al.*, “Learned multiphysics inversion with differentiable programming and machine learning,” *The Leading Edge*, vol. 42, no. 7, pp. 474–486, 2023 (pages 75, 77–79, 93, 95, 96).
- [44] K.-A. Lie, *An introduction to reservoir simulation using MATLAB/GNU Octave: User guide for the MATLAB Reservoir Simulation Toolbox (MRST)*. Cambridge University Press, 2019 (page 76).
- [45] H. Gross and A. Mazuyer, “Geosx: A multiphysics, multilevel simulator designed for exascale computing,” in *SPE Reservoir Simulation Conference*, OnePetro, 2021 (pages 76, 104).

- [46] D. Rezende and S. Mohamed, “Variational inference with normalizing flows,” in *International conference on machine learning*, PMLR, 2015, pp. 1530–1538 (pages 76, 79).
- [47] K. Kontolati, S. Goswami, G. E. Karniadakis, and M. D. Shields, “Learning in latent spaces improves the predictive accuracy of deep neural operators,” *arXiv preprint arXiv:2304.07599*, 2023 (page 76).
- [48] J. A. L. Benitez, T. Furuya, F. Faucher, X. Tricoche, and M. V. de Hoop, “Fine-tuning neural-operator architectures for training and generalization,” *arXiv preprint arXiv:2301.11509*, 2023 (page 76).
- [49] R. Chandra, D. Azam, A. Kapoor, and R. D. Müller, “Surrogate-assisted bayesian inversion for landscape and basin evolution models,” *Geoscientific Model Development*, vol. 13, no. 7, pp. 2959–2979, 2020 (page 77).
- [50] S. Lan, S. Li, and B. Shahbaba, “Scaling up bayesian uncertainty quantification for inverse problems using deep neural networks,” *SIAM/ASA Journal on Uncertainty Quantification*, vol. 10, no. 4, pp. 1684–1713, 2022 (pages 77, 106).
- [51] Y. Cao, S. Li, L. Petzold, and R. Serban, “Adjoint sensitivity analysis for differential-algebraic equations: The adjoint dae system and its numerical solution,” *SIAM journal on scientific computing*, vol. 24, no. 3, pp. 1076–1089, 2003 (page 77).
- [52] R.-E. Plessix, “A review of the adjoint-state method for computing the gradient of a functional with geophysical applications,” *Geophysical Journal International*, vol. 167, no. 2, pp. 495–503, 2006 (page 77).
- [53] J. D. Jansen, “Adjoint-based optimization of multi-phase flow through porous media—a review,” *Computers & Fluids*, vol. 46, no. 1, pp. 40–51, 2011 (page 77).
- [54] A. Griewank *et al.*, “On automatic differentiation,” *Mathematical Programming: recent developments and applications*, vol. 6, no. 6, pp. 83–107, 1989 (page 77).
- [55] Y. Yang, A. F. Gao, K. Azizzadenesheli, R. W. Clayton, and Z. E. Ross, “Rapid seismic waveform modeling and inversion with neural operators,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023 (pages 77, 106).
- [56] D. Li, K. Xu, J. M. Harris, and E. Darve, “Coupled time-lapse full-waveform inversion for subsurface flow problems using intrusive automatic differentiation,” *Water Resources Research*, vol. 56, no. 8, e2019WR027032, 2020 (pages 78, 93, 96).
- [57] B. Peters and F. J. Herrmann, “Constraints versus penalties for edge-preserving full-waveform inversion,” *The Leading Edge*, vol. 36, no. 1, pp. 94–100, 2017 (pages 78, 105).

- [58] E. Esser, L. Guasch, T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, “Total variation regularization strategies in full-waveform inversion,” *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 376–406, 2018 (pages 78, 87, 105).
- [59] B. Peters, B. R. Smithyman, and F. J. Herrmann, “Projection methods and applications for seismic nonlinear inverse problems with multiple constraints,” *Geophysics*, vol. 84, no. 2, R251–R269, 2019 (pages 78, 105).
- [60] X. Li, A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, “Fast randomized full-waveform inversion with compressive sensing,” *Geophysics*, vol. 77, no. 3, A13–A17, 2012 (page 78).
- [61] G. Hennenfent, E. v. d. Berg, M. P. Friedlander, and F. J. Herrmann, “New insights into one-norm solvers from the pareto curve,” *Geophysics*, vol. 73, no. 4, A23–A26, 2008 (page 78).
- [62] L. Dinh, J. Sohl-Dickstein, and S. Bengio, “Density estimation using real nvp,” *arXiv preprint arXiv:1605.08803*, 2016 (page 79).
- [63] A. Siahkoohi, G. Rizzuti, R. Orozco, and F. J. Herrmann, “Reliable amortized variational inference with physics-based latent distribution correction,” *Geophysics*, vol. 88, no. 3, R297–R322, 2023 (pages 79, 106).
- [64] R. Orozco, M. Louboutin, A. Siahkoohi, G. Rizzuti, T. van Leeuwen, and F. J. Herrmann, “Amortized normalizing flows for transcranial ultrasound with uncertainty quantification,” Mar. 2023 (page 79).
- [65] R. Orozco, A. Siahkoohi, M. Louboutin, and F. J. Herrmann, “Refining amortized posterior approximations using gradient-based summary statistics,” *arXiv preprint arXiv:2305.08733*, 2023 (page 79).
- [66] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013 (page 79).
- [67] I. Goodfellow *et al.*, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014 (page 79).
- [68] I. Kobyzev, S. J. Prince, and M. A. Brubaker, “Normalizing flows: An introduction and review of current methods,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 11, pp. 3964–3979, 2020 (page 79).
- [69] C. Horvat and J.-P. Pfister, “Intrinsic dimensionality estimation using normalizing flows,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 12 225–12 236, 2022 (page 79).

- [70] X. Zhang and A. Curtis, “Seismic tomography using variational inference methods,” *Journal of Geophysical Research: Solid Earth*, vol. 125, no. 4, e2019JB018589, 2020 (page 79).
- [71] X. Zhang and A. Curtis, “Bayesian geophysical inversion using invertible neural networks,” *Journal of Geophysical Research: Solid Earth*, vol. 126, no. 7, e2021JB022320, 2021 (page 79).
- [72] A. Siahkoohi and F. J. Herrmann, “Learning by example: Fast reliability-aware seismic imaging with normalizing flows,” in *First International Meeting for Applied Geoscience & Energy*, Society of Exploration Geophysicists, Expanded Abstracts, Sep. 2021, pp. 1580–1585 (page 79).
- [73] X. Zhao, A. Curtis, and X. Zhang, “Bayesian seismic tomography using normalizing flows,” *Geophysical Journal International*, vol. 228, no. 1, pp. 213–239, Jul. 2021. eprint: <https://academic.oup.com/gji/article-pdf/228/1/213/40348424/ggab298.pdf> (page 79).
- [74] K. Lensink, B. Peters, and E. Haber, “Fully hyperbolic convolutional neural networks,” *Research in the Mathematical Sciences*, vol. 9, no. 4, p. 60, 2022 (page 79).
- [75] R. Orozco, A. Siahkoohi, G. Rizzuti, T. van Leeuwen, and F. J. Herrmann, “Photoacoustic imaging with conditional priors from normalizing flows,” in *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems*, Dec. 2021 (page 79).
- [76] R. Orozco, M. Louboutin, and F. J. Herrmann, “Memory efficient invertible neural networks for 3d photoacoustic imaging,” *arXiv preprint arXiv:2204.11850*, 2022 (page 79).
- [77] R. Orozco, A. Siahkoohi, G. Rizzuti, T. van Leeuwen, and F. J. Herrmann, “Photoacoustic imaging with conditional priors from normalizing flows,” in *Neural Information Processing Systems (NeurIPS)*, (NIPS, virtual), Dec. 2021 (page 79).
- [78] J. Wen, R. Ahmad, and P. Schniter, “A conditional normalizing flow for accelerated multi-coil mr imaging,” *Proceedings of machine learning research*, vol. 202, p. 36 926, 2023 (page 79).
- [79] M. Asim, M. Daniels, O. Leong, A. Ahmed, and P. Hand, “Invertible generative models for inverse problems: Mitigating representation error and dataset bias,” in *International conference on machine learning*, PMLR, 2020, pp. 399–409 (page 79).
- [80] L. Ardizzone *et al.*, “Analyzing inverse problems with invertible neural networks,” *arXiv preprint arXiv:1808.04730*, 2018 (page 80).

- [81] A. Beck, *Introduction to nonlinear optimization: Theory, algorithms, and applications with MATLAB*. SIAM, 2014 (page 86).
- [82] P. S. Stanimirović and M. B. Miladinović, “Accelerated gradient descent methods with line search,” *Numerical Algorithms*, vol. 54, pp. 503–520, 2010 (pages 86, 90).
- [83] O. Møyner, M. Johnsrud, H. M. Nilsen, X. Raynaud, K. O. Lye, and Z. Yin, *Sintefmath/jutul.jl: V0.2.6*, version v0.2.6, Apr. 2023 (pages 88, 102).
- [84] O. Møyner, G. Bruer, and Z. Yin, *Sintefmath/jutuldarcy.jl: V0.2.3*, version v0.2.3, Apr. 2023 (pages 88, 102).
- [85] Z. Yin, G. Bruer, and M. Louboutin, *Slimgroup/jutuldarcyrules.jl: V0.2.5*, version v0.2.5, Apr. 2023 (pages 88, 102).
- [86] P. Witte *et al.*, *Slimgroup/invertiblenetworks.jl: V2.2.5*, version v2.2.5, Apr. 2023 (page 89).
- [87] J. Kruse, G. Detommaso, U. Köthe, and R. Scheichl, “Hint: Hierarchical invertible neural transport for density estimation and bayesian inference,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021, pp. 8191–8199 (page 89).
- [88] L. Mosser, O. Dubrule, and M. J. Blunt, “Deepflow: History matching in the space of deep generative models,” *arXiv preprint arXiv:1905.05749*, 2019 (page 91).
- [89] S. W. Canchumuni, A. A. Emerick, and M. A. C. Pacheco, “History matching geological facies models based on ensemble smoother and deep generative models,” *Journal of Petroleum Science and Engineering*, vol. 177, pp. 941–958, 2019 (page 93).
- [90] P. Avseth, T. Mukerji, and G. Mavko, *Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk*. Cambridge university press, 2010 (pages 93, 95).
- [91] R. E. Sheriff and L. P. Geldart, *Exploration seismology*. Cambridge university press, 1995 (page 93).
- [92] P. A. Witte *et al.*, “A large-scale framework for symbolic implementations of seismic inversion algorithms in julia,” *GEOPHYSICS*, vol. 84, no. 3, F57–F71, 2019. eprint: <https://doi.org/10.1190/geo2018-0174.1> (page 95).

- [93] M. Louboutin *et al.*, “Devito (v3.1.0): An embedded domain-specific language for finite differences and geophysical exploration,” *Geoscientific Model Development*, vol. 12, no. 3, pp. 1165–1187, 2019 (page 96).
- [94] F. Luporini *et al.*, “Architecture and performance of devito, a system for automated stencil computation,” *ACM Trans. Math. Softw.*, vol. 46, no. 1, Apr. 2020 (page 96).
- [95] L. Huang, “Geophysical monitoring for geologic carbon storage,” 2022 (page 98).
- [96] M. Liu, D. Vashisth, D. Grana, and T. Mukerji, “Joint inversion of geophysical data for geologic carbon sequestration monitoring: A differentiable physics-informed neural network model,” *Journal of Geophysical Research: Solid Earth*, vol. 128, no. 3, e2022JB025372, 2023 (page 98).
- [97] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004 (page 100).
- [98] M. Innes, “Flux: Elegant machine learning with julia,” *Journal of Open Source Software*, vol. 3, no. 25, p. 602, 2018 (page 102).
- [99] S. Heidenreich, H. Gross, M. Henn, C. Elster, and M. Bär, “A surrogate model enables a bayesian approach to the inverse problem of scatterometry,” in *Journal of Physics: Conference Series*, IOP Publishing, vol. 490, 2014, p. 012 007 (page 106).
- [100] K. Cranmer, J. Brehmer, and G. Louppe, “The frontier of simulation-based inference,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 055–30 062, 2020 (page 106).
- [101] T. O’Leary-Roseberry, P. Chen, U. Villa, and O. Ghattas, “Derivative-informed neural operator: An efficient framework for high-dimensional parametric derivative learning,” *Journal of Computational Physics*, vol. 496, p. 112 555, 2024 (page 106).
- [102] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter estimation and inverse problems*. Elsevier, 2018 (page 107).

CHAPTER 5

DERISKING GEOLOGIC CARBON STORAGE FROM HIGH-RESOLUTION TIME-LAPSE SEISMIC TO EXPLAINABLE LEAKAGE DETECTION

5.1 Summary

Geological carbon storage represents one of the few truly scalable technologies capable of reducing the CO₂ concentration in the atmosphere. While this technology has the potential to scale, its success hinges on our ability to mitigate its risks. An important aspect of risk mitigation concerns assurances that the injected CO₂ remains within the storage complex. Amongst the different monitoring modalities, seismic imaging stands out with its ability to attain high resolution and high fidelity images. However, these superior features come, unfortunately, at prohibitive costs and time-intensive efforts potentially rendering extensive seismic monitoring undesirable. To overcome this shortcoming, we present a methodology where time-lapse images are created by inverting non-replicated time-lapse monitoring data jointly. By no longer insisting on replication of the surveys to obtain high fidelity time-lapse images and differences, extreme costs and time-consuming labor are averted. To demonstrate our approach, hundreds of realistic synthetic noisy time-lapse seismic datasets are simulated that contain imprints of regular CO₂ plumes and irregular plumes that leak. These time-lapse datasets are subsequently inverted to produce time-lapse difference images used to train a deep neural classifier. The testing results show that the classifier is capable of detecting CO₂ leakage automatically on unseen data and with a reasonable accuracy. We consider the use of this classifier as a first step in the development of an automatic workflow designed to handle the large number of continuously monitored CO₂ injection sites needed to help combat climate change.

5.2 Introduction

For various reasons, seismic monitoring of geological carbon storage (GCS) comes with its own set of unique challenges. Amongst these challenges, the need for low-cost highly repeatable, high resolution, and high fidelity images ranks chiefly. While densely sampled and replicated time-lapse surveys—which rely on permanent reservoir monitoring systems or on replicated streamer or node surveys—may be able to provide images conducive to interpretation and reservoir management, these approaches are often too costly and may require too much handholding to be of practical use for GCS at many injection sites.

To overcome these challenges, we replace the current paradigm of costly replicated acquisition, cumbersome time-lapse processing, and interpretation, by a joint inversion framework mapping time-lapse data to high fidelity and high resolution images from sparse non-replicated time-lapse surveys. We demonstrate that we arrive at an imaging framework that is suitable for automatic detection of pressure-induced CO₂ leakage, which represents one of the possible leakage scenarios. Rather than relying on meticulous 4D workflows where baseline and monitoring surveys are processed separately to yield accurate and artifact-free time-lapse differences, our approach exposes information that is shared amongst the different vintages by formulating the imaging problem in terms of an unknown fictitious common component, and innovations of the baseline and monitor surveys with respect to this common component. Because the common component is informed by all time-lapse surveys, its image quality improves when the surveys bring complementary information, which is the case when the surveys are not replicated. In turn, the enhanced common component results in improved images for the baseline, monitor, and their time-lapse difference(s). Our joint wave-equation based imaging formulation is versatile and capable of accounting for real data time-lapse issues such as changes in the background velocity model, calibration errors in shot and receiver locations [1], and noise [2, 3]. The same applies to corrections for the source signature using on-the-fly source estimations [4,

5]. However, we also acknowledge that the robustness of our method to such real data issues needs to be validated and is a topic for future study.

To showcase the achievable imaging gains and how these can be used in a GCS setting where CO₂ leakage is of major consideration, we create hundreds of time-lapse imaging experiments involving CO₂ plumes whose behavior is determined by the two-phase flow equations. To mimic irregular flow due to pressure-induced opening of fractures, we increase the permeability in the seal at random locations and pressure thresholds. The resulting flow simulations are used to generate time-lapse datasets that serve as input to our joint imaging scheme. The produced time-lapse difference images are subsequently used to train and test a neural network that as an explainable classifier determines whether the CO₂ plume behaves regularly or shows signs of leakage.

Our contributions are organized as follows. First, we discuss the time-lapse seismic imaging problem and its practical difficulties. Next, we introduce the joint recovery model that takes explicit advantage of information shared by multiple surveys. By means of a carefully designed synthetic case study involving saline aquifers made of Blunt sandstone in the Southern North Sea, we demonstrate the uplift of the joint recovery model and how its images can be used to train a deep neural network classifier to detect erroneous growth of the CO₂ plume automatically. Aside from determining whether the CO₂ plume behaves regularly or not, our network also provides class activation mappings that visualize areas in the image on which the network is basing its classification.

5.3 Seismic monitoring with time-lapse imaging

To keep track of CO₂ plume development during geological carbon storage (GCS) projects, multiple time-lapse surveys are collected. Baseline surveys are acquired before the supercritical CO₂ is injected into the reservoir. These baseline surveys, denoted by the index $j = 1$, are followed by one or more monitor surveys, collected at later times and indexed by $j = 2, \dots, n_v$ with n_v the total number of surveys.

Seismic monitoring of GCS brings its own unique set of challenges that stem from the fact that its main concern is (early) detection of possible leakage of CO₂ from the storage complex. To be successful with this task, monitoring GCS calls for a time-lapse imaging modality that is capable of

- detecting weak time-lapse signals associated with small rock-physics changes induced by CO₂ leakage
- attaining high lateral resolution from active-source surface seismic data to detect vertically moving leakage
- handling an increasing number of not perfectly calibrated seismic surveys collected over long periods of time ($\sim 50 - 100$ years)
- reducing costs drastically by no longer insisting on replication of time-lapse surveys to attain high degrees of repeatability
- lowering the cumulative environmental imprint of active source acquisition

5.3.1 Monitoring with the joint recovery model

To meet these challenges, we choose a linear imaging framework where observed linearized data for each vintage is related to perturbations in the acoustic impedance via

$$\mathbf{b}_j = \mathbf{A}_j \mathbf{x}_j + \mathbf{e}_j \quad \text{for } j = 1, 2, \dots, n_v. \quad (5.1)$$

In this expression, the matrix \mathbf{A}_j stands for the linearized Born scattering operator for the j th vintage. Observed linearized data, collected for all shots in the vector \mathbf{b}_j , is generated by applying the \mathbf{A}_j 's to the (unknown) impedance perturbations denoted by \mathbf{x}_j for $j = 1, 2, \dots, n_v$. The task of time-lapse imaging is to create high resolution, high fidelity and true amplitude estimates for the time-lapse images, $\{\widehat{\mathbf{x}}_j\}_{j=1}^{n_v}$, from non-replicated sparsely sampled noisy time-lapse data.

We argue that our choice for linearized imaging is justified for four reasons. First,

CO₂ injection sites undergo baseline studies, involving vintage data and possible follow-up surveys, which means that accurate information on the background velocity model is generally available. Second, changes in the acoustic parameters induced by CO₂ injection are typically small, so it suffices to work with one and the same background model for the baseline and monitor surveys. Third, when the background model is sufficiently close to the true model, linearized inversion, which corresponds to a single Gauss-Newton iteration of full-waveform inversion, converges quadratically. Fourth, because the forward model is linear, it is conducive to the use of the joint recovery model where inversions are carried out with respect to the common component, which is shared between all vintages, and innovations with respect to the common component. Because the common component represents an average, we expect this joint imaging method to be relatively robust with respect to kinematic changes induced by time-lapse effects or by lack of calibration of the acquisition [1].

By parameterizing time-lapse images, $\{\mathbf{x}_j\}_{j=1}^{n_v}$, in terms of the common component, \mathbf{z}_0 , and innovations with respect to the common component, $\{\mathbf{z}_j\}_{j=1}^{n_v}$, we arrive at the joint recovery model where representations for the images are given by

$$\mathbf{x}_j = \frac{1}{\gamma} \mathbf{z}_0 + \mathbf{z}_j \quad \text{for } j = 1, 2, \dots, n_v. \quad (5.2)$$

Here, the parameter, γ , controls the balance between the common component, \mathbf{z}_0 , and innovation components, $\{\mathbf{z}_j\}_{j=1}^{n_v}$ [6]. Compared to traditional time-lapse approaches, where data are imaged separately or where time-lapse surveys are subtracted, inversions for time-lapse images based on the above parameterization are carried out jointly and involve inverting the following matrix:

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\gamma} \mathbf{A}_1 & \mathbf{A}_1 & & \\ \vdots & & \ddots & \\ \frac{1}{\gamma} \mathbf{A}_{n_v} & & & \mathbf{A}_{n_v} \end{bmatrix}. \quad (5.3)$$

While traditional time-lapse imaging approaches strive towards maximal replication between the surveys to suppress acquisition related artifacts, imaging with the joint recovery model—which entails inverting the underdetermined system in Equation 5.3 using structure-promotion techniques (e.g. via ℓ_1 -norm minimization)—improves the image quality of the vintages themselves in situations where the surveys are not replicated. This occurs in cases where $\mathbf{A}_i \neq \mathbf{A}_j$ for $\forall i \neq j$, or in situations where there is significant noise. This remarkable result was shown to hold for sparsity-promoting denoising of time-lapse field data [2, 3], for various wavefield reconstructions of randomized simultaneous-source dynamic (towed-array) and static (OBC/OBN) marine acquisitions [1, 7, 8], and for wave-based inversion, including least-squares reverse-time migration and full-waveform inversion [9, 10]. The observed quality gains in these applications can be explained by improvements in the common component resulting from complementary information residing in non-replicated time-lapse surveys. This enhanced recovery of the common component in turn improves the recovery of the innovations and therefore the vintages themselves. The time-lapse differences themselves also improve, or at the very least, remain relatively unaffected when the surveys are not replicated. Relaxing replication of surveys obviously leads to reduction in cost and environmental impact. Below, we show how GCS monitoring also benefits from this approach.

5.3.2 Monitoring with curvelet-domain structure promotion

To obtain high resolution and high fidelity time-lapse images, we invert the system in Equation 5.3 [4, 11, 12] with

$$\begin{aligned} \underset{\mathbf{z}}{\text{minimize}} \quad & \lambda \|\mathbf{Cz}\|_1 + \frac{1}{2} \|\mathbf{Cz}\|_2^2 \\ \text{subject to} \quad & \|\mathbf{b} - \mathbf{Az}\|_2^2 \leq \sigma, \end{aligned} \tag{5.4}$$

where \mathbf{C} is the forward curvelet transform, λ the threshold parameter, and σ the magnitude of the noise. At iteration k and for $\sigma = 0$, solving Equation 5.4 corresponds to computing

the following iterations:

$$\begin{aligned}\mathbf{u}_{k+1} &= \mathbf{u}_k - t_k \mathbf{A}_k^\top (\mathbf{A}_k \mathbf{z}_k - \mathbf{b}_k) \\ \mathbf{z}_{k+1} &= \mathbf{C}^\top S_\lambda(\mathbf{C} \mathbf{u}_{k+1}),\end{aligned}\tag{5.5}$$

where \mathbf{A}_k , with a slight abuse of notation, represents the matrix in Equation 5.3 for a subset of shots randomly selected from sources in each survey. The vector \mathbf{b}_k contains the extracted shot records from \mathbf{b} and the symbol $^\top$ refers to the adjoint. The dynamic steplength t_k is given by $t_k = \|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|_2^2 / \|\mathbf{A}_k^\top (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)\|_2^2$ [13]. Sparsity is promoted via curvelet-domain soft thresholding, $S_\lambda(\cdot) = \max(|\cdot| - \lambda, 0) \text{sign}(\cdot)$, where, λ , is the threshold. The vectors \mathbf{u}_k and \mathbf{z}_k contain the baseline and innovation components.

5.4 Numerical case study: Blunt sandstone in the Southern North Sea

Before discussing the impact of high resolution and high fidelity time-lapse imaging with the joint recovery model on the down-stream task of automatic leakage detection with a neural network classifier, we first detail the setup of our numerical experiments using techniques from simulation-based acquisition design as described by [12]. In order to generate realistic time-lapse data and training sets for the automatic leakage classifier, we follow the workflow summarized in Figure 5.1. In this approach, use is made of proxy models for seismic properties derived from real 3D imaged seismic and well data [14]. With rock physics, these seismic models are converted to fluid-flow models that serve as input to two-phase flow simulations. The resulting datasets, which include pressure-induced leakage, will be used to create time-lapse data used to train our classifier. For more detail, refer to the caption of Figure 5.1.

5.4.1 Proxy seismic and fluid-flow models

Amongst the various CO₂ injection projects, GCS in offshore saline aquifers has been most successful in reaching scale and in meeting injection targets [15]. For that reason, we con-

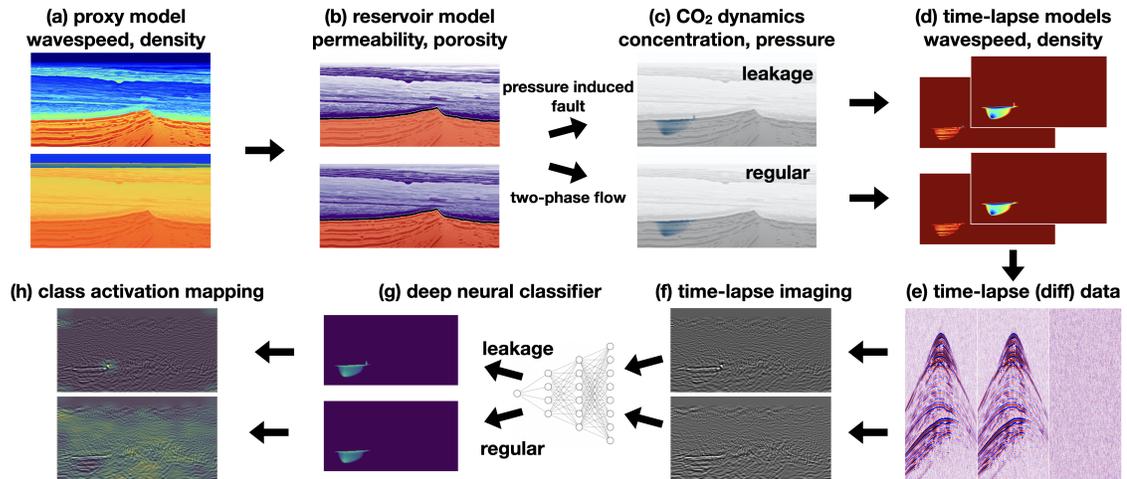


Figure 5.1: Simulation-based monitoring design framework. Starting with a proxy model for the wavespeed and density (a), the workflow proceeds by converting these seismic properties into permeability and porosity (b). These fluid flow properties are used to simulated CO₂ plumes that behave regularly or exhibit leakage outside the storage complex (c). Induced changes by the CO₂ plume for the wavespeed and density are depicted in (d) and serve as input to simulations of time-lapse seismic data (SNR 8.0 dB) and shot-domain time-lapse differences (SNR -31.4 dB). Imaging results for regular and irregular plume developments are plotted in (f) and serve as input to the deep neural classifier (g), which determines whether the flow behaves regularly or leaks. Activation mappings in (h) show regions on which the network is basing its classification. As expected, the activation mapping is diffusive in case of regular CO₂ plume development and focused on the leakage location when CO₂ plume behaves irregularly.

sider a proxy model derived from real 3D imaged seismic and well data [14] and representative for CO₂ injection in the South of the North Sea involving a saline aquifer made of the highly permeable Blunt sandstone. This area, which is actively being considered for GCS [16], consists of the following three geologic sections (see Figure 5.2 for the permeability and porosity distribution):

- (i) the highly porous (average 33%) and permeable (> 170 mD) Blunt sandstone reservoir of about 300 – 500 m thick. This section, denoted by red colors in Figure 5.2, corresponds to the saline aquifer and serves as the reservoir for CO₂ injection;
- (ii) the primary seal (permeability $10^{-4} - 10^{-2}$ mD) made of the Rot Halite Member, which is 50 m thick and continuous (black layer in Figure 5.2);
- (iii) the secondary seal made of the Haisborough group, which is > 300 m thick and consists of low-permeable (permeability 15 – 18 mD) mudstone (purple section in Figure 5.2).

To arrive at the fluid-flow models, we consider 2D subsets of the 3D Compass model [14] and convert these seismic models to fluid-flow properties (see Figure 5.1 (b)) by assuming a linear relationship between compressional wavespeed and permeability in each stratigraphic section. For further details on the conversion of compressional wavespeed and density to permeability and porosity, we refer to empirical relationships reported in [17]. During conversion, an increase of 1km/s in compressional wavespeed is assumed to correspond to an increase of 1.63 mD in permeability. From this, porosity is calculated with the Kozeny-Carman equation [18] $K = \phi^3 \left(\frac{1.527}{0.0314*(1-\phi)} \right)^2$, where K and ϕ denote permeability (mD) and porosity (%) with constants taken from the Strategic UK CCS Storage Appraisal Project report.

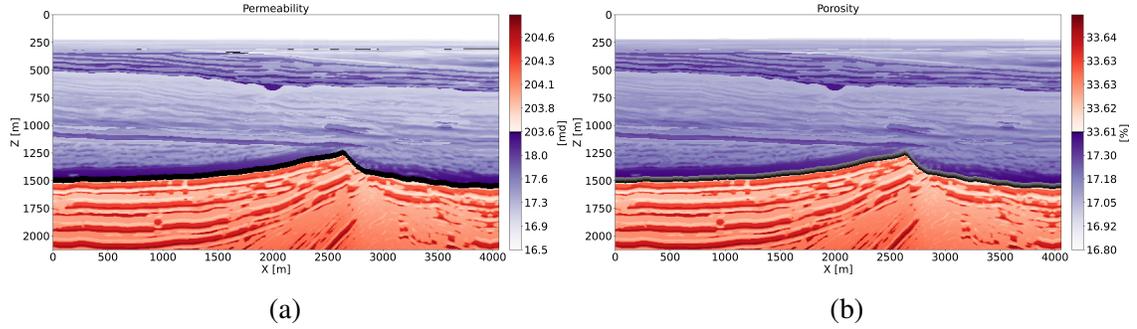


Figure 5.2: Permeability and porosity derived from a 2D slice of Compass model.

5.4.2 Fluid-flow simulations

To model CO_2 plumes that behave regularly and irregularly, the latter due to leakage, we solve the two-phase flow equations numerically¹ for both pressure and concentration [19, 20]. To mimic possible pressure-induced CO_2 leakage, we increase the permeability at random distances away from the injection well within the seal from 10^{-4} mD to 500 mD when the pressure exceeds ~ 15 MPa. At that depth, the pressure is below the fracture gradient [21]. Since pressure-induced fractures come in different sizes, we also randomly vary the width of the pressure-induced fracture openings from 12.5 m to 62.5 m. Examples of fluid-flow simulations without and with leakage are shown in Figure 5.1 (c).

5.4.3 Rock-physics conversion

To monitor temporal variations in the plume's CO_2 concentration seismically, we use the patchy saturation model [22] to convert the CO_2 concentration to decrease in compressional wavespeed and density. These changes are shown in Figure 5.1 (d). The fact that these induced changes in the time-lapse differences in seismic properties are relatively small in spatial extent (~ 800 m for the plume and < 62.5 m for the leakage) and amplitude (1.68% time-lapse change in the acoustic impedance) calls for a time-lapse imaging modality with small normalized root-mean-square (NRMS) values [23].

¹We used the open-source software FwiFlow.jl [19, 20] to solve the two-phase flow equations for both the pressure and concentration.

5.4.4 Time-lapse seismic simulations

To train and validate automatic detection of CO₂ leakage from the storage complex requires the creation of realistic synthetic time-lapse datasets that contain the seismic imprint of regular as well as irregular (leakage) plume development. To this end, baseline surveys are simulated prior to CO₂ injection for different subsets of the Compass model. Monitor surveys are simulated 200 days after leakage occurs to verify that potential leakage can be detected automatically early on. For regular plume development, we shoot monitor surveys for each subset at random times after CO₂ injection. To strike a balance between acquisition productivity and time-lapse image quality, use is made of dense semi-permanent acoustic monitoring at the seafloor with 25 m receiver spacing. Contrary to expensive permanent reservoir monitoring systems with multi-component geophones, our system works with hydrophones connected to underwater buoys located 2 m above the ocean bottom. Aside from being relatively low-cost, this system also avoids complications arising from elastic wave interactions at the seabed. Time-lapse acquisition costs are further reduced by non-replicated coarse shooting with the source towed at 10 m depth below the ocean surface. Subsampling artifacts are reduced by using a randomized technique from compressive sensing where 32 sources are located at non-replicated jittered [24] source positions, yielding an average source sampling of 125 m. Given this acquisition geometry, linear data is generated² with Equation 5.1 for a 25 Hz Ricker wavelet and with the band-limited noise term set so that the data's signal-to-noise ratio (SNR) is 8.0 dB. This noise level leads to an extremely poor SNR of -31.4 dB for time-lapse differences in the shot domain. See Figure 5.1 (e).

²We used the open-source software JUDI.jl [25, 26] to model the wave propagation. This Julia package implements highly optimized propagators using Devito [27, 28, 29].

5.4.5 Imaging with joint recovery model versus reverse-time migration

Given the simulated time-lapse datasets with and without leakage, time-lapse difference images are created according to two different imaging scenarios, namely via independent reverse-time migration (RTM), conducted on the baseline and monitor surveys separately, and via inversion of the joint recovery model (cf. Equations 5.3 and 5.4). To limit the computational cost of the Bregman iterations (Equation 5.5), four shot records are selected per iteration at random from each survey for imaging [30, 11, 4, 12], limiting the cost of the joint inversion to the equivalent of three RTMs. The recovered baseline images are shown in Figures 5.3a for RTM and 5.3b for JRM. For the leakage scenario, the time-lapse differences are plotted in Figures 5.3c and 5.3d, for RTM and JRM respectively. For the regular plume, the time-lapse differences are plotted in Figures 5.3e and 5.3f, for RTM and JRM respectively. From these images, it is clear that joint inversion leads to relatively artifact-free recovery of the vintages and time-lapse differences. This observation is reflected in the NRMS values, which improve considerably as shown by the histograms in Figure 5.4 for 1000 imaging experiments. Not only do the NRMS values shift towards the left, their values are also more concentrated when inverting time-lapse data with the joint recovery model. Both features are beneficial to automatic leakage detection.

5.5 Deep neural network classifier for CO₂ leakage detection

The injection of supercritical CO₂ into the storage complex perturbs the physical, chemical and thermal environment of the reservoir [31]. Because CO₂ injection increases the pressure, this process may trigger CO₂ leakage across the seal when the pressure increase induces opening of pre-existing faults or fractures zones [32, 21]. To ensure safe operations of CO₂ storage, we develop a quantitative leakage detection tool based on a deep neural classifier. This classifier is trained on time-lapse images that contain the imprint of CO₂ plumes that behave regularly and irregularly. In case of irregular flow, CO₂ escapes the

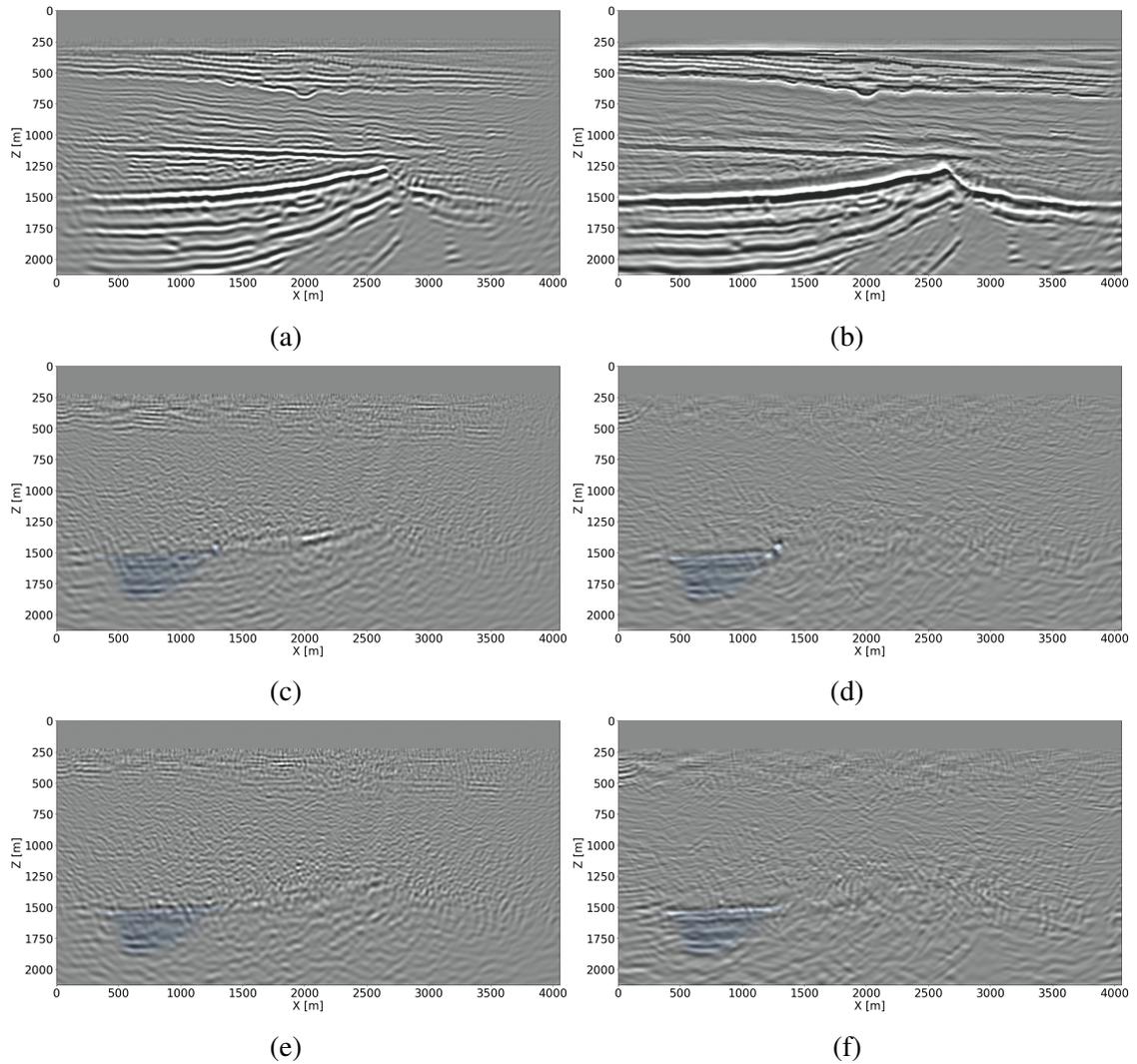


Figure 5.3: Reverse-time migration (RTM) versus inversion joint recovery model (JRM). (a) RTM image of the baseline; (b) JRM image of the baseline; (c) time-lapse difference and CO₂ plume for independent RTM images with leakage; (d) time-lapse obtained by inverting the time-lapse data jointly with leakage; (e) time-lapse difference and CO₂ plume for independent RTM images without leakage; (f) time-lapse obtained by inverting the time-lapse data jointly without leakage. Notice improvement in the time-lapse image quality. This improvement is reflected in the NRMS values that decrease from 8.48 % for RTM to 3.20 % for JRM.

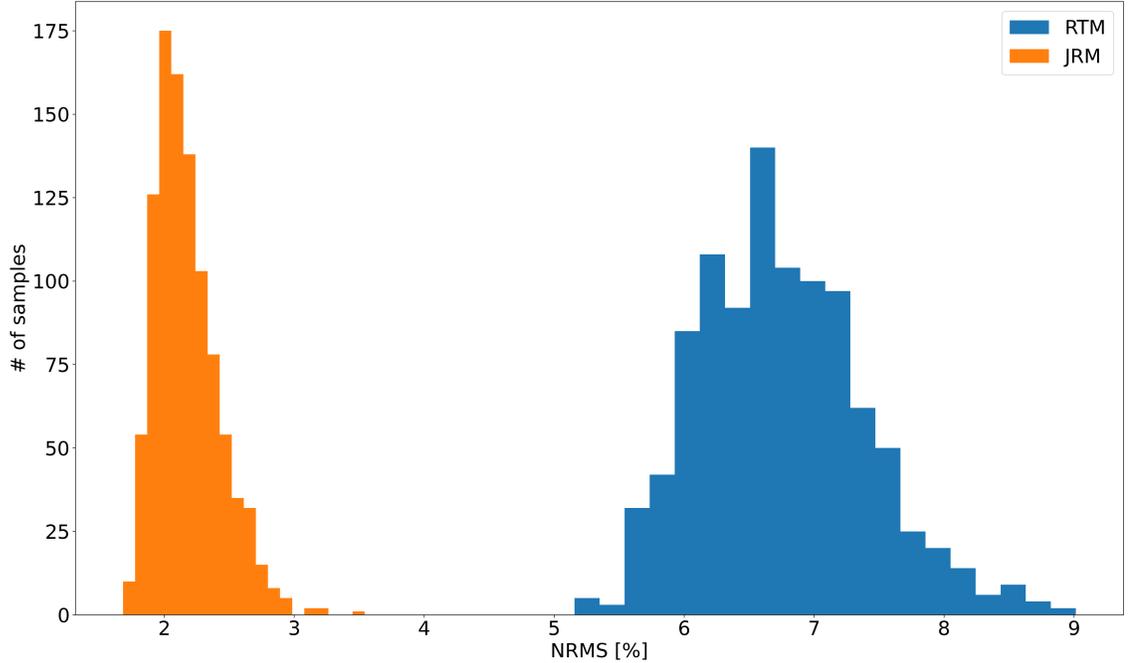


Figure 5.4: NRMS values for 1000 time-lapse experiments.

storage complex through a pressure induced opening in the seal, which causes a localized increase in permeability (shown in Figure 5.3d).

Because time-lapse differences are small in amplitude, and strongly localized laterally when leakage occurs, highly sensitive learned classifiers are needed. For this purpose, we follow [33] and adopt the Vision Transformer (ViT) [34]. This state-of-the-art classifier originated from the field of natural language processing (NLP) [35]. Thanks to their attention mechanism, ViTs have been shown to achieve superior performance on image classification tasks where image patches are considered as word tokens by the transformer network. As a result, ViTs have much less image-specific inductive bias compared to convolutional neural networks [34].

To arrive at a practical and performant ViT classifier, we start from a ViT that is pre-trained on image tasks with 16×16 patches and apply transfer learning [36] to fine-tune this network on 1576 labeled time-lapse images. Catastrophic forgetting is avoided by freezing the initial layers, which are responsible for feature extraction, during the initial training. After the initial training of the last dense layers, all network weights are updated

for several epochs while keeping the learning rate small. The labeled (regular vs. irregular flow) training set itself consists of 1576 time-lapse datasets divided equally between regular and irregular flow.

After the training is completed, baseline and monitor surveys are simulated for 394 unseen Earth models with regular and irregular plumes. These simulated time-lapse datasets are imaged with JRM by inverting the matrix in Equation 5.3 via Bregman iterations in Equation 5.5. The resulting time-lapse difference images (see Figures 5.3d and 5.3f for two examples) serve as input to the ViT classifier. Refer to Figure 5.5 for performance, which corresponds to a two by two confusion matrix. The first row denotes the classification results for samples with regular plume (negative samples), where 193 (true negative) out of 206 samples are classified correctly. The second row denotes the classification results for samples with CO₂ leakage over the seal (positive samples), where 147 (true positive) out of 188 samples are classified correctly. Due to the fact that JRM recovers relatively artifact-free time-lapse differences, the classifier does not pick up too many artifacts related to finite acquisition as CO₂ leakage. This leads to much fewer false alarms for CO₂ leakage.

5.6 Class activation mapping based saliency map

While our ViT classifier is capable of achieving good performance (see Figure 5.5), making intervention decisions during GCS projects calls for interpretability and trustworthiness of our classifier [37, 38, 39]. To enhance these features, we take advantage of class activation mappings (CAM) [40]. These saliency maps help us to identify the discriminative spatial regions in each image that support a particular class decision. In our application, these regions correspond to areas where the classifier deems the CO₂ plume to behave irregularly (if the classification result is leakage). By overlaying time-lapse difference images with these maps, interpretation is facilitated, assisting practitioners to make decisions on how to proceed with GCS projects and take associated actions. Figure 5.6 illustrates how

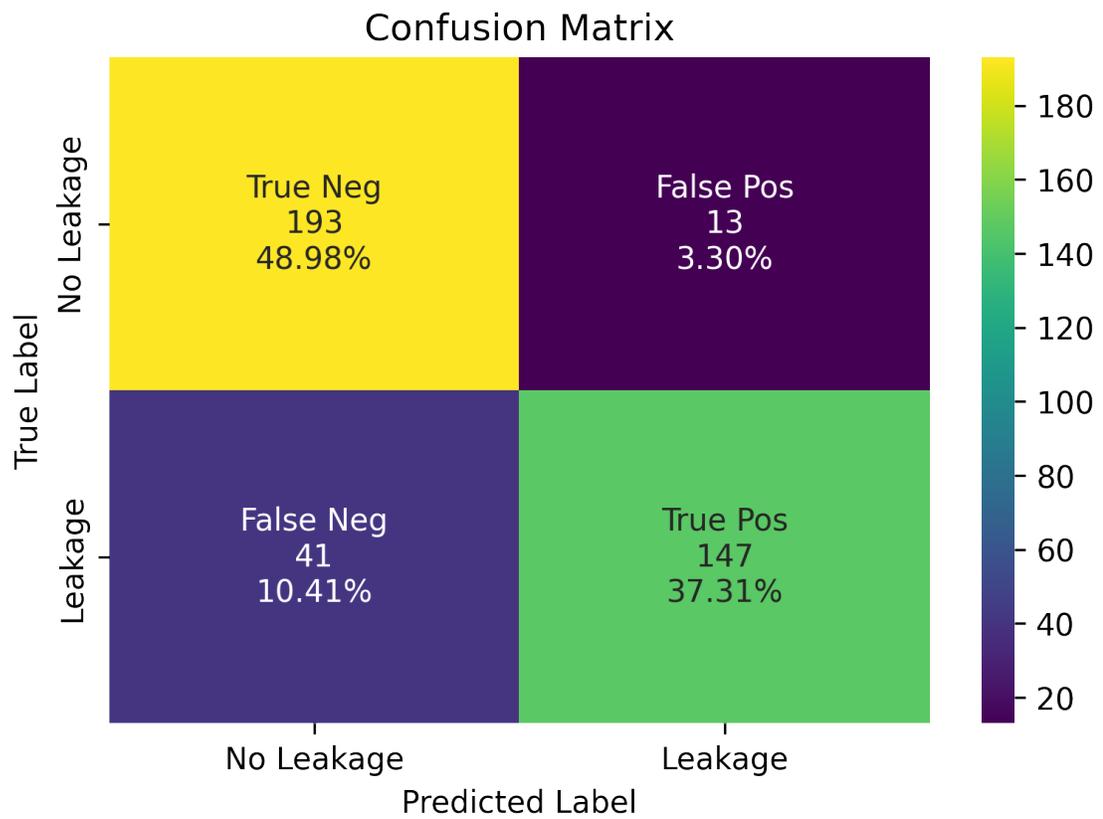


Figure 5.5: Confusion matrix for classifier trained on recovery images from JRM.

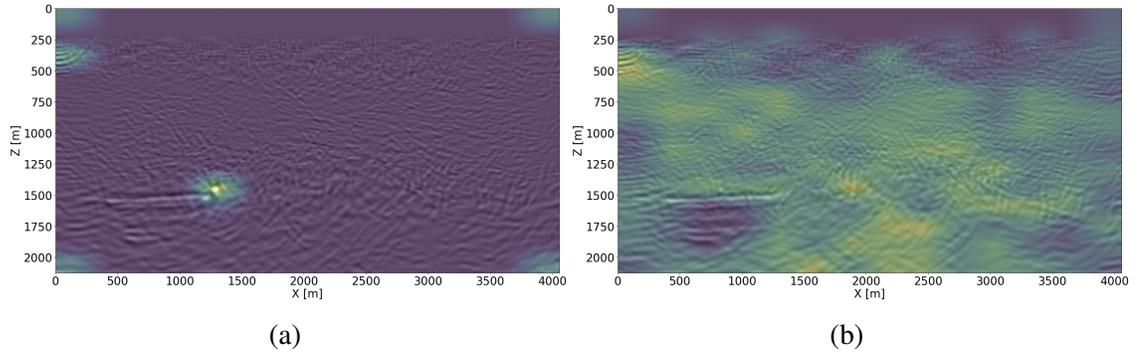


Figure 5.6: CAM for time-lapse difference images with a leaking plume and with a regular plume.

the Score CAM approach [41] serves this purpose³. Figure 5.6a shows the CAM result for a time-lapse difference image classified as a CO₂ leakage (in Figure 5.3d). Despite few artifacts around the image, the CAM clearly focuses on the CO₂ leakage over the seal, which could potentially alert the practitioners of GCS. When the plume is detected as growing regularly, the CAM result is diffusive (shown in Figure 5.6b). This shows that the classification decision is based on the entire image and not only at the plume area. The scripts to reproduce the experiments are available on the SLIM GitHub page <https://github.com/slimgroup/GCS-CAM>.

5.7 Discussion and conclusion

As a first step in the development of scalable automatic workflows for seismic monitoring of geologic carbon storage, we propose a methodology for low-cost time-lapse imaging that exploits commonality between baseline and monitor surveys through the joint recovery model. By means of carefully designed realistic synthetic time-lapse seismic experiments, we have shown that highly repeatable, high resolution and high fidelity images are achievable without insisting on replication of the baseline and monitor surveys. Because our method relies on a joint inversion methodology, it also averts labor-intensive 4D processing to compensate for less-than-ideal acquisitions. Aside from establishing our claim

³We used the open-source software PyTorch library for CAM methods [42] to calculate the CAM images.

of relaxing the need for replication empirically, through hundreds of synthetic time-lapse experiments yielding significant improvements in time-lapse image quality and NMRS values, we also showed that a deep neural classifier can be trained to detect CO₂ leakage automatically. While the classification results are encouraging, false positives and negatives remain. We argue that these may be acceptable since decisions to intervene, e.g. to stop injection of CO₂, typically involve other complementary sources of information such as pressure drops at the wellhead. In future work, we plan to extend our methodology to different leakage scenarios and quantification of uncertainty. We also intend to further investigate robustness of the proposed joint imaging methodology with respect to calibration errors and variations in the source signature within and across different surveys. Finally, interpretability of the neural classifier's output and different leakage scenarios and their impact on the shape of the CO₂ plume will also be further investigated.

5.8 References

- [1] F. Oghenekohwo and F. J. Herrmann, “Highly repeatable time-lapse seismic with distributed compressive sensing—mitigating effects of calibration errors,” *The Leading Edge*, vol. 36, no. 8, pp. 688–694, 2017 (pages 118, 121, 122).
- [2] L. Wei, Y. Tian, C. Li, S. Oppert, and G. Hennenfent, “Improve 4d seismic interpretability with joint sparsity recovery,” in *SEG Technical Program Expanded Abstracts 2018*, Society of Exploration Geophysicists, 2018, pp. 5338–5342 (pages 118, 122).
- [3] Y. Tian, L. Wei, C. Li, S. Oppert, and G. Hennenfent, “Joint sparsity recovery for noise attenuation,” in *SEG Technical Program Expanded Abstracts 2018*, Society of Exploration Geophysicists, 2018, pp. 4186–4190 (pages 118, 122).
- [4] M. Yang, Z. Fang, P. A. Witte, and F. J. Herrmann, “Time-domain sparsity promoting least-squares reverse time migration with source estimation,” *Geophysical Prospecting*, vol. 68, no. 9, pp. 2697–2711, Aug. 2020, (Geophysical Prospecting) (pages 118, 122, 128).
- [5] Z. Yin, R. Orozco, P. Witte, M. Louboutin, G. Rizzuti, and F. J. Herrmann, “Extended source imaging—a unifying framework for seismic and medical imaging,” in *SEG Technical Program Expanded Abstracts 2020*, Society of Exploration Geophysicists, 2020, pp. 3502–3506 (page 119).
- [6] X. Li, “A weighted ℓ_1 -minimization for distributed compressive sensing,” Ph.D. dissertation, University of British Columbia, 2015 (page 121).
- [7] M. Kotsi, “Time-lapse seismic imaging and uncertainty quantification,” Ph.D. dissertation, Memorial University of Newfoundland, 2020 (page 122).
- [8] W. Zhou and D. Lumley, “Non-repeatability effects on time-lapse 4d seismic full waveform inversion for ocean-bottom node data,” *Geophysics*, vol. 86, no. 4, pp. 1–60, 2021 (page 122).
- [9] F. Oghenekohwo, “Economic time-lapse seismic acquisition and imaging—Reaping the benefits of randomized sampling with distributed compressive sensing,” (PhD), Ph.D. dissertation, The University of British Columbia, Vancouver, Aug. 2017 (page 122).
- [10] F. Oghenekohwo and F. J. Herrmann, “Improved time-lapse data repeatability with randomized sampling and distributed compressive sensing,” in *EAGE Annual Conference Proceedings*, (EAGE, Paris), Jun. 2017 (page 122).

- [11] P. A. Witte, M. Louboutin, F. Luporini, G. J. Gorman, and F. J. Herrmann, “Compressive least-squares migration with on-the-fly fourier transforms,” *Geophysics*, vol. 84, no. 5, R655–R672, 2019 (pages 122, 128).
- [12] Z. Yin, M. Louboutin, and F. J. Herrmann, “Compressive time-lapse seismic monitoring of carbon storage and sequestration with the joint recovery model,” in *SEG International Exposition and Annual Meeting*, SEG, 2021, D011S145R001 (pages 122, 123, 128).
- [13] D. A. Lorenz, F. Schopfer, and S. Wenger, “The linearized bregman method via split feasibility problems: Analysis and generalizations,” *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 1237–1262, 2014 (page 123).
- [14] C. Jones, J. Edgar, J. Selvage, and H. Crook, “Building complex synthetic models to evaluate acquisition geometries and velocity inversion technologies,” in *74th EAGE Conference and Exhibition incorporating EUROPEC 2012*, European Association of Geoscientists & Engineers, 2012, cp–293 (pages 123, 125).
- [15] K. Michael *et al.*, “Geological storage of co2 in saline aquifers—a review of the experience from existing storage operations,” *International journal of greenhouse gas control*, vol. 4, no. 4, pp. 659–667, 2010 (page 123).
- [16] C. Kolster, S. Agada, N. Mac Dowell, and S. Krevor, “The impact of time-varying CO₂ injection rate on large scale storage in the UK bunter sandstone,” *International Journal of Greenhouse Gas Control*, vol. 68, pp. 77–85, 2018 (page 125).
- [17] T. Klimentos, “The effects of porosity-permeability-clay content on the velocity of compressional waves,” *Geophysics*, vol. 56, no. 12, pp. 1930–1939, 1991 (page 125).
- [18] A. Costa, “Permeability-porosity relationship: A reexamination of the kozeny-carman equation based on a fractal pore-space geometry assumption,” *Geophysical research letters*, vol. 33, no. 2, 2006 (page 125).
- [19] D. Li and K. Xu, *Lidongzh/fwiflow.jl: V0.3.1*, version v0.3.1, Sep. 2021 (page 126).
- [20] D. Li, K. Xu, J. M. Harris, and E. Darve, “Coupled time-lapse full-waveform inversion for subsurface flow problems using intrusive automatic differentiation,” *Water Resources Research*, vol. 56, no. 8, e2019WR027032, 2020 (page 126).
- [21] P. Ringrose, *How to store CO₂ underground: Insights from early-mover CCS Projects*. Springer, 2020 (pages 126, 128).
- [22] P. Avseth, T. Mukerji, and G. Mavko, *Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk*. Cambridge university press, 2010 (page 126).

- [23] E. Kragh and P. Christie, “Seismic repeatability, normalized rms, and predictability,” *The Leading Edge*, vol. 21, no. 7, pp. 640–647, 2002 (page 126).
- [24] F. J. Herrmann and G. Hennenfent, “Non-parametric seismic data recovery with curvelet frames,” *Geophysical Journal International*, vol. 173, pp. 233–248, Apr. 2008 (page 127).
- [25] P. A. Witte *et al.*, “A large-scale framework for symbolic implementations of seismic inversion algorithms in julia,” *GEOPHYSICS*, vol. 84, no. 3, F57–F71, 2019. eprint: <https://doi.org/10.1190/geo2018-0174.1> (page 127).
- [26] M. Louboutin *et al.*, *Slimgroup/judi.jl: V3.1.9*, version v3.1.9, Sep. 2022 (page 127).
- [27] M. Louboutin *et al.*, “Devito (v3.1.0): An embedded domain-specific language for finite differences and geophysical exploration,” *Geoscientific Model Development*, vol. 12, no. 3, pp. 1165–1187, 2019 (page 127).
- [28] F. Luporini *et al.*, “Architecture and performance of devito, a system for automated stencil computation,” *ACM Trans. Math. Softw.*, vol. 46, no. 1, Apr. 2020 (page 127).
- [29] F. Luporini *et al.*, *Devitocodes/devito: V4.7.1*, version v4.7.1, Aug. 2022 (page 127).
- [30] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, “Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing,” *SIAM Journal on Imaging sciences*, vol. 1, no. 1, pp. 143–168, 2008 (page 128).
- [31] P. Newell and A. G. Ilgen, “Overview of geological carbon storage (GCS),” in *Science of Carbon Storage in Deep Saline Formations*, Elsevier, 2019, pp. 1–13 (page 128).
- [32] K. Pruess, “On CO₂ behavior in the subsurface, following leakage from a geologic storage reservoir,” Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), Tech. Rep., 2006 (page 128).
- [33] H. T. Erdinc, A. P. Gahlot, Z. Yin, M. Louboutin, and F. J. Herrmann, “De-risking carbon capture and sequestration with explainable CO₂ leakage detection in time-lapse seismic monitoring images,” in *AAAI 2022 Fall Symposium: The Role of AI in Responding to Climate Challenges*, (AAAI 2022 Fall Symposium: The Role of AI in Responding to Climate Challenges, Arlington), Aug. 2022 (page 130).
- [34] A. Dosovitskiy *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” in *International Conference on Learning Representations*, 2021 (page 130).

- [35] A. Vaswani *et al.*, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017 (page 130).
- [36] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?” *Advances in neural information processing systems*, vol. 27, 2014 (page 130).
- [37] S. Hooker, D. Erhan, P.-J. Kindermans, and B. Kim, “A benchmark for interpretability methods in deep neural networks,” *Advances in neural information processing systems*, vol. 32, 2019 (page 131).
- [38] Y. Zhang, P. Tiño, A. Leonardis, and K. Tang, “A survey on neural network interpretability,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2021 (page 131).
- [39] R. Mackowiak, L. Ardizzone, U. Kothe, and C. Rother, “Generative classifiers as a basis for trustworthy image classification,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2971–2981 (page 131).
- [40] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning deep features for discriminative localization,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921–2929 (page 131).
- [41] H. Wang *et al.*, “Score-CAM: Score-weighted visual explanations for convolutional neural networks,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 24–25 (page 133).
- [42] J. Gildenblat and contributors, *Pytorch library for CAM methods*, <https://github.com/jacobgil/pytorch-grad-cam>, 2021 (page 133).

CHAPTER 6

WISE: FULL-WAVEFORM VARIATIONAL INFERENCE VIA SUBSURFACE EXTENSIONS

6.1 Summary

We introduce a probabilistic technique for full-waveform inversion, employing variational inference and conditional normalizing flows to quantify uncertainty in migration-velocity models and its impact on imaging. Our approach integrates generative artificial intelligence with physics-informed common-image gathers, reducing reliance on accurate initial velocity models. Considered case studies demonstrate its efficacy producing realizations of migration-velocity models conditioned by the data. These models are used to quantify amplitude and positioning effects during subsequent imaging.

6.2 Introduction

Full-waveform inversion (FWI) plays a pivotal role in exploration, primarily focusing on estimating Earth’s subsurface properties from observed seismic data. The inherent complexity of FWI stems from its nonlinearity, further complicated by ill-posedness and computational intensiveness of the wave modeling. To address these challenges, we introduce a computationally cost-effective probabilistic framework that generates multiple migration-velocity models conditioned on observed seismic data. By combining deep learning with physics, our approach harnesses advancements in variational inference (VI) [1] and generative artificial intelligence (AI) [2, 3, 4]. We achieve this by forming common-image gathers (CIGs), followed by training conditional normalizing flows (CNFs) that quantify uncertainties in migration-velocity models.

This chapter is organized as follows. First, we delineate the FWI problem and its in-

herent challenges. Subsequently, we explore VI to quantify FWI’s uncertainty. To reduce VI’s computational costs, we introduce *physics-informed summary statistics* and justify the use of CIGs as these statistics. Our framework’s capabilities are validated through two case studies, which include studying the effects of uncertainty in the generated migration-velocity models on migration.

6.3 Methodology

We present a Bayesian inference approach to FWI by briefly introducing FWI and VI used as a framework for uncertainty quantification (UQ).

6.3.1 Full-waveform inversion

Estimation of unknown migration-velocity models, \mathbf{x} , from noisy seismic data, \mathbf{y} involves inverting nonlinear forward operator, \mathcal{F} , which links \mathbf{x} to \mathbf{y} via $\mathbf{y} = \mathcal{F}(\mathbf{x}) + \epsilon$ with ϵ measurement noise. Source/receiver signatures are assumed known and absorbed into \mathcal{F} . Solving this nonlinear inverse problem is challenging because of the noise, the non-convexity of the objective function, and the non-trivial null-space of the modeling [5]. As a result, multiple migration-velocity models fit the data, necessitating a Bayesian framework for UQ.

6.3.2 Full-waveform inference

Rather than seeking a single migration-velocity model, our goal is to invert for a range of models compatible with the data, termed “full-waveform inference”. From a Bayesian perspective, this involves determining the posterior distribution of migration-velocity models given the data, $p(\mathbf{x}|\mathbf{y})$. We focus on amortized VI, which exchanges the computational cost of posterior sampling for neural network training [6, 7, 8, 9, 10, 11]. Specifically, we employ amortized VI, which incurs offline computational training cost but enables cheap online posterior inference on many datasets \mathbf{y} [12]. Next, we discuss how to use CNFs for

amortized VI.

6.3.3 Amortized variational inference with conditional normalizing flows

During VI, the posterior distribution $p(\mathbf{x}|\mathbf{y})$ is approximated by the surrogate, $p_{\theta}(\mathbf{x}|\mathbf{y})$, with learnable parameters, θ . Given the sample pairs $\{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^N$, CNFs are suitable to act as surrogates for the posterior because of their low-cost training and rapid sampling [13, 14]. Their training involves minimization of the Kullback-Leibler divergence between the true and surrogate posterior distribution. In practice, this requires access to N training pairs of migration-velocity model and observed data to minimize the following objective:

$$\underset{\theta}{\text{minimize}} \quad \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{2} \|f_{\theta}(\mathbf{x}^{(i)}; \mathbf{y}^{(i)})\|_2^2 - \log |\det \mathbf{J}_{f_{\theta}}| \right). \quad (6.1)$$

Here, f_{θ} is the CNF with network parameters, θ , and Jacobian, $\mathbf{J}_{f_{\theta}}$. It transforms each velocity model, $\mathbf{x}^{(i)}$, into white noise (as indicated by the ℓ_2 -norm), conditioned on the observation, $\mathbf{y}^{(i)}$. After training, the inverse of CNF turns random realizations of the standard Gaussian distribution into posterior samples (migration-velocity models) conditioned on any seismic observation that is in the same statistical distribution as the training data.

6.3.4 Physics-informed summary statistics

While CNFs are capable of approximating the posterior distribution, training the CNFs on pairs (\mathbf{x}, \mathbf{y}) presents challenges when changes in the acquisition occur or when physical principles simplifying the mapping between model and data are lacking, both of which lead to increasing training costs. To tackle these challenges, [15] introduced fixed reduced-size *summary statistics* that encapsulate observed data and inform the posterior distribution. Building on this concept, [16] uses the gradient as the set of *physics-informed summary statistics*, partially reversing the forward map and therefore accelerating CNF training. For linear inverse problems with Gaussian noise, these statistics are unbiased — maintaining

the same posterior distribution, whether conditioned on original shot data or on the gradient. Based on this principle, [17] used reverse-time migration (RTM), given by the action of the adjoint of the linearized Born modeling, to summarize data and quantify imaging uncertainties for a fixed accurate migration-velocity model.

We aim to extend this approach to the nonlinear FWI problems. While RTM transfers information from the data to the image domain, its performance diminishes for incorrect migration velocities. [18] showed that least-squares migration can perfectly fit the data for correct migration-velocity models, but this fit fails for inaccurate velocity models. This highlights a fundamental limitation in cases where the velocity model is inaccurate and RTM does not correctly summarize the original shot data, which leads to a biased posterior. For an inaccurate initial FWI-velocity model \mathbf{x}_0 , $p(\mathbf{x}|\mathbf{y}) \neq p(\mathbf{x}|\nabla\mathcal{F}(\mathbf{x}_0)^\top\mathbf{y})$ with $\nabla\mathcal{F}$ Born modeling and $^\top$ the adjoint. To avoid this problem, more robust *physics-informed summary statistics* are needed to preserve information.

6.3.5 Common-image gathers as summary statistics

Migration-velocity analysis has a rich history in the literature [19]. Following [18], we employ relatively artifact-free subsurface-offset extended Born modeling to calculate summary statistics. More information can be found in Section A.3. Thanks to being closer to an isometry—i.e., the adjoint of extended Born modeling is closer to its inverse [20, 21] and therefore preserves information — its adjoint can nullify residuals even when the FWI-velocity model is incorrect as shown by [18]. [22] further demonstrated that neural networks can be used to map CIGs to velocity models. Both these findings shed important light on the role of CIGs during VI because CIGs preserve more information than the gradient, which leads to less biased physics-informed summary statistics when given an inaccurate initial FWI-velocity model. Formally, this means $p(\mathbf{x}|\mathbf{y}) \approx p(\mathbf{x}|\overline{\nabla\mathcal{F}}(\mathbf{x}_0)^\top\mathbf{y})$, where $\overline{\nabla\mathcal{F}}$ is extended Born modeling. Leveraging this mathematical observation, we propose WISE, short for full-**W**aveform variational **I**nference via **S**ubsurface **E**xtensions.

The core of this technique is to train CNFs with pairs of velocity models, \mathbf{x} , and CIGs, $\overline{\nabla \mathcal{F}}(\mathbf{x}_0)^\top \mathbf{y}$, guided by the objective of Equation 6.1. Our case studies will demonstrate that even with inaccurate initial FWI-velocity models, CIGs encapsulate more information, enabling the trained CNFs to generate accurate migration-velocity models consistent with the observed shot data.

6.4 Synthetic case studies

Our study evaluates the performance of WISE through synthetic case studies on 2D slices of the Compass dataset [23], known for its “velocity kickback” challenge for FWI algorithms. For a poor initial FWI-velocity model, we aim to compare the quality of posterior samples informed by RTM alone versus those informed by CIGs to verify the superior information content of CIGs. We also illustrate how uncertainty in migration-velocity models can be converted into uncertainties in amplitude and positioning of imaged reflectors.

6.4.1 Dataset generation and network training

We take 800 2D slices of the Compass model of 6.4 km by 3.2 km, with 512 equally spaced sources towed at 12.5m depth and 64 ocean-bottom nodes (OBNs) located at jittered sampled horizontal positions [24, 25]. This sampling scheme utilizes compressive sensing techniques to improve acquisition productivity in various situations [26, 27, 28, 29]. The surface is assumed absorbing. Using a 15Hz central frequency Ricker wavelet with energy below 3Hz removed for realism, acoustic data is simulated with Devito [30, 31] and JUDI.jl [32]. Uncorrelated band-limited Gaussian noise is added (S/N 12dB). The arithmetic mean over all velocity models is used as the 1D initial FWI-velocity model (shown in Figure 6.1(b)). 51 horizontal subsurface offsets ranging from -500m to $+500\text{m}$ are used to compute CIGs (shown in Figure 6.1(e)). Each offset is input to the network as a separate channel. We use the conditional glow network structure [33] for the CNFs thanks to its capability to generate superior natural [34] and seismic [14] images.

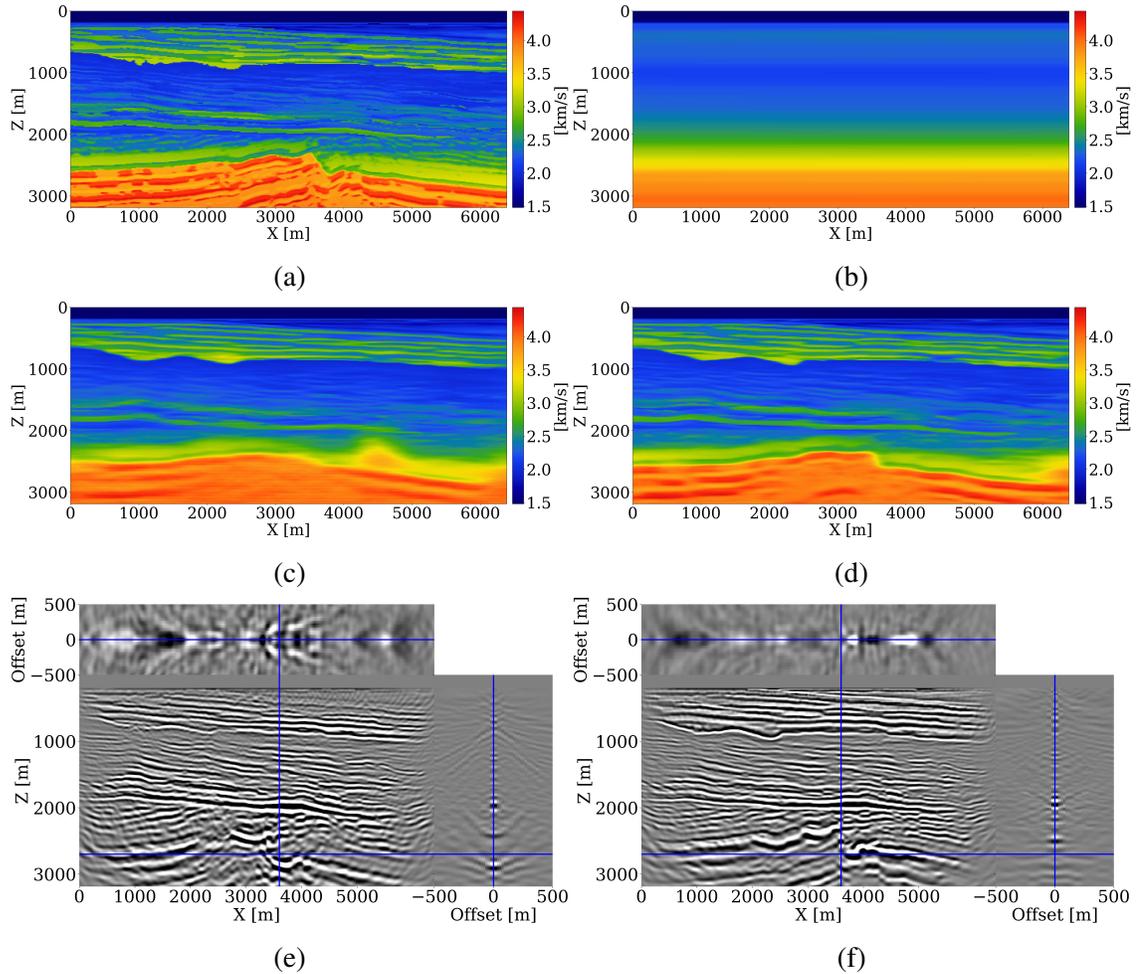


Figure 6.1: (a) an unseen ground-truth velocity model; (b) 1D initial FWI-velocity model; (c) conditional mean estimate for RTM as summary statistics (SSIM = 0.48); (d) conditional mean estimate from WISE (SSIM = 0.56); (e) CIGs calculated by the initial FWI-velocity model given by (b); (f) CIGs calculated by (d).

6.4.2 Results

After CNF training, our method’s performance is evaluated on an unseen 2D Compass slice shown in Figure 6.1(a). When RTM is used to summarize the data, the conditional mean estimate (Figure 6.1(c)) does not capture the shape of the unconformity. Thanks to the CIGs, WISE captures more information and as a result produces a more accurate conditional mean (Figure 6.1(d)). For the 50 test samples, the structural similarity index measure (SSIM) with CIGs yields a mean of 0.63, outperforming RTM-based statistics with a mean SSIM of 0.52.

6.4.3 Quality control

To verify the inferred migration-velocity model as the conditional mean of the posterior, CIGs calculated for the initial FWI-velocity model (Figure 6.1(b)), plotted in Figure 6.1(e), are juxtaposed against CIGs calculated for the inferred migration-velocity model (Figure 6.1(d)), plotted in Figure 6.1(f). Significant improvement in near-offset focused energy is observed in the CIGs for the inferred migration-velocity model. A similar focusing behavior is noted for the posterior samples themselves, as shown in Section B.

6.4.4 Uncertainty quantification and downstream imaging

While access to the posterior represents an important step towards grasping uncertainty, understanding its impact on imaging with (30Hz) RTMs is more relevant because it concerns uncertainty in the final product. For this purpose, we display the posterior velocity samples in Figure 6.2(a) and the point-wise standard deviation in Figure 6.2(b). These deviations increase with depth and correlate with complex geology where the RTM-based inference struggled. To understand how this uncertainty propagates to imaged reflectors, forward uncertainty is assessed by carrying out RTMs for different posterior samples with results shown in Figure 6.2(c) and the standard deviations plotted in Figure 6.2(d). These amplitude deviations are different because mapping migration-velocities to RTMs is highly

nonlinear, leading to large areas of intense amplitude variation and dimming at the edges caused by the Born modeling’s null-space. While these amplitude sensitivities are useful, deviations in the migration velocities also leads to differences in reflector positioning. Vertical shifts between the envelope of the reference image (central image in Figure 6.2(c)) and the envelopes of RTMs for different posterior samples are calculated with a local cross-correlation technique and included in Figure 6.2(d) where blue/red areas correspond to up/down shifts. As expected, these shifts are most notable in the deeper regions and at the edges where velocity variations are the largest.

6.5 Discussion

Once the offline costs of computing 800 CIGs and network training are covered, WISE enables generation of velocity models for unseen seismic data at the low computational cost of a single set of CIGs for a poor initial FWI-velocity model. The Open FWI case study in Section B demonstrates WISE’s capability of producing realistic posterior samples and conditional means for a broad range of unseen velocity models. In the case of the Compass model, the initial FWI-velocity model was poor. Still, CIGs obtained from a single 1D initial model capture relevant information from the non-zero offsets. From this information, the network learns to produce migration-velocity models at inference that focus CIGs. WISE also produced two types of uncertainty, namely (i) inverse uncertainty in migration-velocity model estimation, which arises from both the non-trivial null-space of FWI and the measurement noise, and (ii) forward uncertainty where uncertainty in migration-velocity models is propagated to uncertainty in amplitude and positioning of imaged reflectors.

Opportunities for future research remain. One area concerns dealing with the “amortization gap” where CNFs tend to maximize performance across multiple datasets rather than excelling at a single observation [35]. While we discovered that training CNFs on a diverse set of samples enhances generalization, applying AI techniques to unseen, out-of-distribution samples remains a challenge. However, our WISE framework is compatible

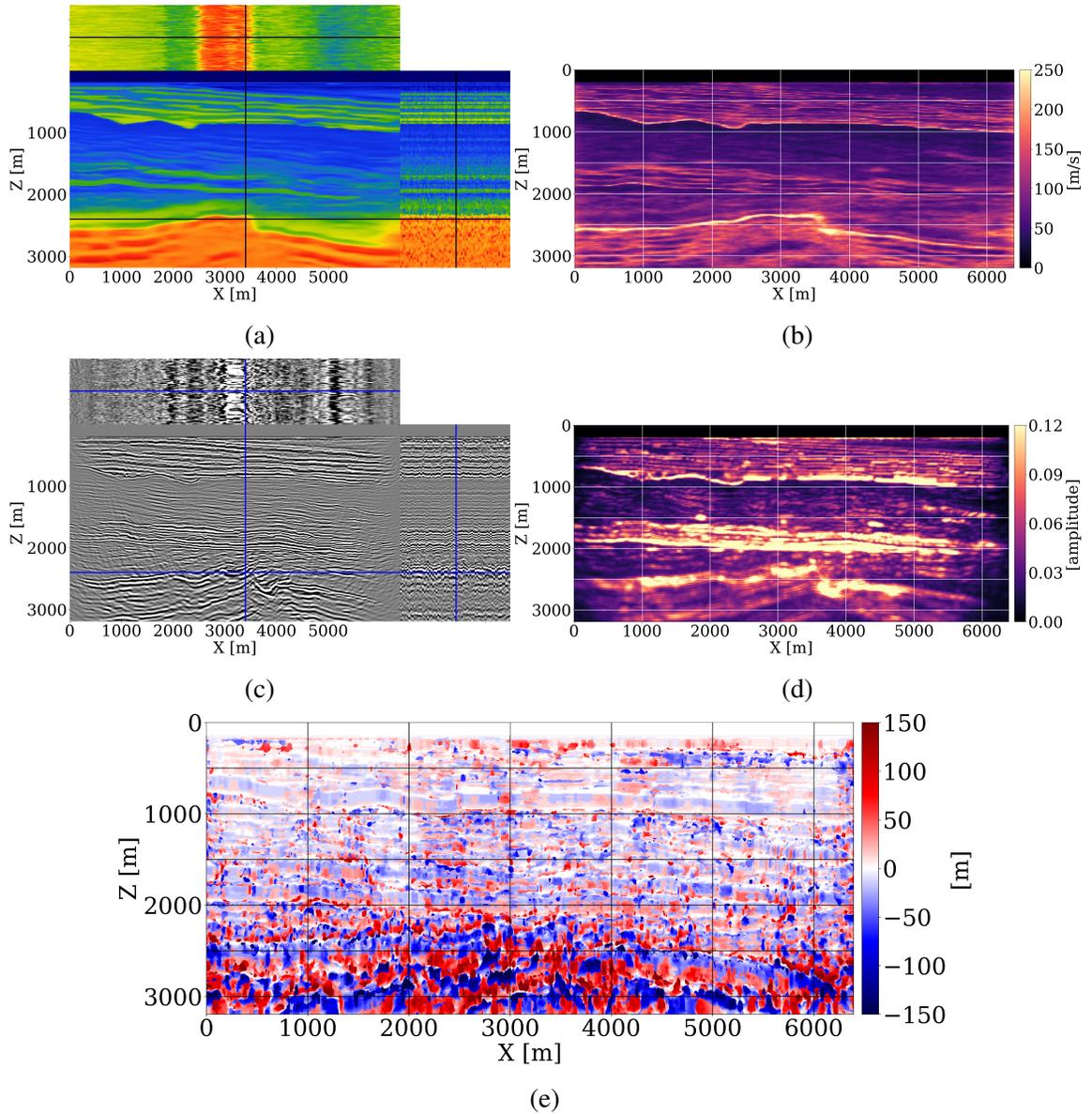


Figure 6.2: Variability in velocity models and imaged reflectors. (a) Posterior velocity samples from WISE visualized similar to CIGs by plotting the conditional mean (Figure 6.1(d)) in the central image. Above it shows the posterior sample traces at $Z = 2.4$ km. On the right shows the traces at $X = 3.4$ km. (b) Point-wise standard deviation of the posterior velocity samples. (c) Samples of imaged reflectors, where the central image displays imaged reflectors using the conditional mean estimate. The layout of the traces remains the same as (a). (d) Point-wise standard deviation of the imaged reflectors. (e) Point-wise maximum depth shift.

with several fine-tuning approaches. To improve single-observation performance, particularly for out-of-distribution samples, computationally more expensive latent space corrections [17] can be employed that incorporate the physics. Recent studies also have indicated that trained CNFs can act as preconditioners or regularizers for physics-based, non-amortized inference [6, 36]. These correction methods can enhance the fit of posterior samples to observed data, as shown in [17], or enable the generation of more focused CIGs through migration velocity analysis. Moreover, velocity continuation methods [37] could be used including recent advances in neural operators [38]. These could offset the cost of running RTMs for each posterior sample, thus accelerating forward uncertainty propagation. While we observed that providing more offsets can enhance the quality of the inference, we recognize the resulting increase in CIG computation costs and CNF memory consumption. This necessitates cost-effective frameworks for determining optimal offset numbers or sampling strategies for CIGs. In this context, recent work on using CNFs for Bayesian optimal experimental design [39] seamlessly integrates as an advancement to the WISE framework. Considering low-rank approximations of CIGs [20] may reduce computational demands. Additionally, exploring other conditional generative models like diffusion models [40] may be worthwhile. Our case studies have yet to account for inverse uncertainty due to modeling errors, such as attenuation effects, multiples, or residual shear wave energy, which could be addressed through Bayesian model misspecification techniques [41]. Recent advances suggest that transfer learning could correct these modeling inaccuracies [42, 43], a solution that is amenable to our approach.

Incurred computational cost on an NVIDIA A100 GPU can be broken down as follows: generating training pairs requires generation of 64 OBN datasets and corresponding CIGs for 800 models, totaling approximately 80 hours of runtime. After generating the training set, training the CNF takes around 16 hours. With these initial runtime investments, the cost for a single inference involves only a single CIG computation, which takes about 6 minutes. For context, running a single FWI starting from the velocity model included in

Figure 6.1(b) requires 12.5 data passes taking roughly 50 minutes to complete (the final result is shown in Section B). Traditional UQ methods require the compute equivalent to hundreds of FWI runs [10], but here we estimate at least 50 FWI runs. Based on these numbers, the computational savings from employing CNF surrogates offset the upfront costs after inference on approximately 3 datasets. We emphasize that as long as the statistics of the underlying geology remains similar, our amortized network can be applied to different observed datasets in the complete basin without retraining. Furthermore, the parallel execution of training pair generation on clusters can significantly reduce initial computational time. Although our study primarily demonstrates a proof of concept on a realistic 2D experiment, the WISE software tool chain is designed for large-scale 3D problems. CNFs, favored for their memory efficiency through invertibility [33], are well-suited for 3D problems. In addition, memory consumption of CIG computation can be reduced significantly with random trace estimation techniques [44]. Since our work requires training samples of Earth models, we envision these samples coming from legacy proxy models and future work will explore automatic workflows for generating these from field observations.

6.6 Conclusions

We present WISE, full-Waveform variational Inference via Subsurface Extensions, for computationally efficient uncertainty quantification of FWI. This framework underscores the potential of generative AI in addressing FWI challenges, paving the way for a new seismic inversion and imaging paradigm that is uncertainty-aware. By having common-image gathers act as information-preserving summary statistics, a principled approach to UQ is achieved where generative AI is successfully combined with wave physics. Because WISE automatically produces distributions for migration-velocity models conditioned by the data, it moves well beyond traditional velocity model building. It was shown that this distributional information can be employed to quantify uncertainties in the migration-velocity models that can be used to better understand amplitude and positioning uncertainty

in migration.

6.7 References

- [1] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, “An introduction to variational methods for graphical models,” *Machine learning*, vol. 37, pp. 183–233, 1999 (page 139).
- [2] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013 (page 139).
- [3] I. Goodfellow *et al.*, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014 (page 139).
- [4] D. J. Rezende, S. Mohamed, and D. Wierstra, “Stochastic backpropagation and approximate inference in deep generative models,” in *International conference on machine learning*, PMLR, 2014, pp. 1278–1286 (page 139).
- [5] A. Tarantola, “Inversion of seismic reflection data in the acoustic approximation,” *GEOPHYSICS*, vol. 49, no. 8, pp. 1259–1266, Aug. 1984 (page 140).
- [6] G. Rizzuti, A. Siahkoohi, P. A. Witte, and F. J. Herrmann, “Parameterizing uncertainty by deep invertible networks: An application to reservoir characterization,” *SEG Technical Program Expanded Abstracts 2020*, Sep. 30, 2020 (pages 140, 148).
- [7] Y. Ren, P. A. Witte, A. Siahkoohi, M. Louboutin, Z. Yin, and F. J. Herrmann, “Seismic velocity inversion and uncertainty quantification using conditional normalizing flows,” in *AGU Fall Meeting 2021*, AGU, 2021 (page 140).
- [8] A. Siahkoohi and F. J. Herrmann, “Learning by example: Fast reliability-aware seismic imaging with normalizing flows,” *First International Meeting for Applied Geoscience & Energy Expanded Abstracts*, Sep. 1, 2021 (page 140).
- [9] X. Zhang, M. A. Nawaz, X. Zhao, and A. Curtis, “An introduction to variational inference in geophysical inverse problems,” in *Advances in geophysics*, vol. 62, Elsevier, 2021, pp. 73–140 (page 140).
- [10] X. Zhang, A. Lomas, M. Zhou, Y. Zheng, and A. Curtis, “3-d bayesian variational full waveform inversion,” *Geophysical Journal International*, vol. 234, no. 1, pp. 546–561, 2023 (pages 140, 149).
- [11] A. P. Gahlot, H. T. Erdinc, R. Orozco, Z. Yin, and F. J. Herrmann, “Inference of co2 flow patterns—a feasibility study,” *arXiv preprint arXiv:2311.00290*, 2023 (page 140).

- [12] J. Kruse, G. Detommaso, U. Köthe, and R. Scheichl, “Hint: Hierarchical invertible neural transport for density estimation and bayesian inference,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021, pp. 8191–8199 (page 140).
- [13] D. Rezende and S. Mohamed, “Variational inference with normalizing flows,” in *International conference on machine learning*, PMLR, 2015, pp. 1530–1538 (page 141).
- [14] M. Louboutin *et al.*, “Learned multiphysics inversion with differentiable programming and machine learning,” *The Leading Edge*, vol. 42, no. 7, pp. 474–486, 2023 (pages 141, 143).
- [15] S. T. Radev, U. K. Mertens, A. Voss, L. Ardizzone, and U. Köthe, “Bayesflow: Learning complex stochastic models with invertible neural networks,” *IEEE transactions on neural networks and learning systems*, vol. 33, no. 4, pp. 1452–1466, 2020 (page 141).
- [16] R. Orozco, A. Siahkoohi, G. Rizzuti, T. van Leeuwen, and F. Herrmann, “Adjoint operators enable fast and amortized machine learning based bayesian uncertainty quantification,” *Medical Imaging 2023: Image Processing*, I. Išgum and O. Colliot, Eds., Apr. 3, 2023 (page 141).
- [17] A. Siahkoohi, G. Rizzuti, R. Orozco, and F. J. Herrmann, “Reliable amortized variational inference with physics-based latent distribution correction,” *Geophysics*, vol. 88, no. 3, R297–R322, 2023 (pages 142, 148).
- [18] J. Hou and W. W. Symes, “Accelerating extended least-squares migration with weighted conjugate gradient iteration,” *GEOPHYSICS*, vol. 81, no. 4, S165–S179, Jul. 2016 (page 142).
- [19] W. W. Symes, “Migration velocity analysis and waveform inversion,” *Geophysical prospecting*, vol. 56, no. 6, pp. 765–790, 2008 (page 142).
- [20] M. Yang, M. Graff, R. Kumar, and F. J. Herrmann, “Low-rank representation of omnidirectional subsurface extended image volumes,” *GEOPHYSICS*, vol. 86, no. 3, S165–S183, Mar. 11, 2021 (pages 142, 148).
- [21] F. t. Kroode, “An omnidirectional seismic image extension,” *Inverse Problems*, vol. 39, no. 3, p. 035 003, Jan. 31, 2023 (page 142).
- [22] Z. Geng, Z. Zhao, Y. Shi, X. Wu, S. Fomel, and M. Sen, “Deep learning for velocity model building with common-image gather volumes,” *Geophysical Journal International*, vol. 228, no. 2, pp. 1054–1070, 2022 (page 142).

- [23] C. Jones, J. Edgar, J. Selvage, and H. Crook, “Building complex synthetic models to evaluate acquisition geometries and velocity inversion technologies,” in *74th EAGE Conference and Exhibition incorporating EUROPEC 2012*, European Association of Geoscientists & Engineers, 2012, cp–293 (page 143).
- [24] G. Hennenfent and F. J. Herrmann, “Simply denoise: Wavefield reconstruction via jittered undersampling,” *Geophysics*, vol. 73, no. 3, pp. V19–V28, 2008 (page 143).
- [25] F. J. Herrmann, “Randomized sampling and sparsity: Getting more information from fewer samples,” *Geophysics*, vol. 75, no. 6, WB173–WB187, 2010 (page 143).
- [26] H. Wason and F. J. Herrmann, “Time-jittered ocean bottom seismic acquisition,” in *SEG International Exposition and Annual Meeting*, SEG, 2013, SEG–2013 (page 143).
- [27] F. Oghenekohwo, H. Wason, E. Esser, and F. J. Herrmann, “Low-cost time-lapse seismic with distributed compressive sensing—part 1: Exploiting common information among the vintages,” *Geophysics*, vol. 82, no. 3, P1–P13, 2017 (page 143).
- [28] H. Wason, F. Oghenekohwo, and F. J. Herrmann, “Low-cost time-lapse seismic with distributed compressive sensing—part 2: Impact on repeatability,” *Geophysics*, vol. 82, no. 3, P15–P30, 2017 (page 143).
- [29] Z. Yin, H. T. Erdinc, A. P. Gahlot, M. Louboutin, and F. J. Herrmann, “Derisking geologic carbon storage from high-resolution time-lapse seismic to explainable leakage detection,” *The Leading Edge*, vol. 42, no. 1, pp. 69–76, 2023 (page 143).
- [30] M. Louboutin *et al.*, “Devito (v3.1.0): An embedded domain-specific language for finite differences and geophysical exploration,” *Geoscientific Model Development*, vol. 12, no. 3, pp. 1165–1187, 2019 (page 143).
- [31] F. Luporini *et al.*, “Architecture and performance of devito, a system for automated stencil computation,” *ACM Trans. Math. Softw.*, vol. 46, no. 1, Apr. 2020 (page 143).
- [32] P. A. Witte *et al.*, “A large-scale framework for symbolic implementations of seismic inversion algorithms in julia,” *GEOPHYSICS*, vol. 84, no. 3, F57–F71, 2019. eprint: <https://doi.org/10.1190/geo2018-0174.1> (page 143).
- [33] R. Orozco *et al.*, “Invertiblenetworks.jl: A julia package for scalable normalizing flows,” *arXiv preprint arXiv:2312.13480*, 2023 (pages 143, 149).
- [34] D. P. Kingma and P. Dhariwal, “Glow: Generative flow with invertible 1x1 convolutions,” *Advances in neural information processing systems*, vol. 31, 2018 (page 143).

- [35] J. Marino, Y. Yue, and S. Mandt, “Iterative amortized inference,” in *International Conference on Machine Learning*, PMLR, 2018, pp. 3403–3412 (page 146).
- [36] A. Siahkoohi, G. Rizzuti, M. Louboutin, P. A. Witte, and F. J. Herrmann, “Preconditioned training of normalizing flows for variational inference in inverse problems,” *arXiv preprint arXiv:2101.03709*, 2021 (page 148).
- [37] S. Fomel, “Time-migration velocity analysis by velocity continuation,” *Geophysics*, vol. 68, no. 5, pp. 1662–1672, 2003 (page 148).
- [38] A. Siahkoohi, M. Louboutin, and F. J. Herrmann, “Velocity continuation with fourier neural operators for accelerated uncertainty quantification,” in *SEG International Exposition and Annual Meeting*, SEG, 2022, D011S092R004 (page 148).
- [39] R. Orozco, F. J. Herrmann, and P. Chen, “Probabilistic bayesian optimal experimental design using conditional normalizing flows,” *arXiv preprint arXiv:2402.18337*, 2024 (page 148).
- [40] L. Baldassari, A. Siahkoohi, J. Garnier, K. Solna, and M. V. de Hoop, “Conditional score-based diffusion models for bayesian inference in infinite dimensions,” *Advances in Neural Information Processing Systems*, vol. 36, 2024 (page 148).
- [41] M. Schmitt, P.-C. Bürkner, U. Köthe, and S. T. Radev, “Detecting model misspecification in amortized bayesian inference with neural networks,” *arXiv preprint arXiv:2112.08866*, 2021 (page 148).
- [42] A. Siahkoohi, M. Louboutin, and F. J. Herrmann, “The importance of transfer learning in seismic modeling and imaging,” *Geophysics*, vol. 84, no. 6, A47–A52, 2019 (page 148).
- [43] J. Yao, L. Guasch, and M. Warner, “Neural networks as a tool for domain translation of geophysical data,” *Geophysics*, vol. 88, no. 3, pp. V267–V275, 2023 (page 148).
- [44] M. Louboutin and F. J. Herrmann, “Wave-based inversion at scale on graphical processing units with randomized trace estimation,” *Geophysical Prospecting*, vol. 72, no. 2, pp. 353–366, 2024 (page 149).

CHAPTER 7

WISER: MULTIMODAL VARIATIONAL INFERENCE FOR FULL-WAVEFORM INVERSION WITHOUT DIMENSIONALITY REDUCTION

7.1 Summary

We present a semi-amortized variational inference framework designed for computationally feasible uncertainty quantification in 2D full-waveform inversion to explore the multimodal posterior distribution without dimensionality reduction. The framework is called **WISER**, short for full-Waveform variational Inference via Subsurface Extensions with Refinements. WISER leverages the power of generative artificial intelligence to perform approximate amortized inference that is low-cost albeit showing an amortization gap. This gap is closed through non-amortized refinements that make frugal use of acoustic wave physics. Case studies illustrate that WISER is capable of full-resolution, computationally feasible, and reliable uncertainty estimates of velocity models and imaged reflectivities.

7.2 Introduction

Full-waveform inversion (FWI) aims to estimate unknown multi-dimensional ($D \geq 2$) velocity models, denoted as \mathbf{x} , from noisy seismic data, \mathbf{y} , by inverting the nonlinear forward operator, \mathcal{F} , which relates \mathbf{x} and \mathbf{y} via $\mathbf{y} = \mathcal{F}(\mathbf{x}) + \epsilon$ with ϵ measurement noise [1]. FWI poses significant challenges, as it requires solving a high-dimensional, non-convex, and ill-posed inverse problem, with a computationally demanding forward operator in multiple dimensions. In addition, the inherent nonuniqueness of FWI results leads to multiple possible Earth models compatible with the observed data, underscoring the need for uncertainty quantification (UQ) to handle this multimodality.

The trade-off between accuracy and computational cost is a critical consideration in any

high-dimensional inference routine with expensive forward operators [2], including FWI. To circumvent the costs associated with global optimization, several approaches have attempted localized UQ [3, 4, 5, 6] based on the Laplace approximation. However, these approaches may not capture the full complexities of multimodal parameter spaces. In contrast, a Bayesian inference approach offers a costly but comprehensive resolution of the posterior distribution, $p(\mathbf{x} \mid \mathbf{y})$.

Bayesian inference algorithms are broadly categorized into two groups. The first, sampling-based methods, like Markov-chain Monte Carlo [MCMC, 7], struggle with high-dimensional parameter spaces. To meet this challenge, they often rely on too restrictive low-dimensional parameterizations to reduce the number of sampling iterations [8, 9, 10, 11, 12, 13], which could bias the inference results, rendering them impractical for multi-D UQ studies especially when solutions are nonunique.

The second category, optimization-based methods, like variational inference [VI, 14], seek to approximate the posterior distribution using classes of known parameterized distributions. VI can be subdivided into amortized and non-amortized methods. Amortized VI involves a computationally intensive offline training phase, leveraging advances in generative artificial intelligence (genAI), particularly with models like conditional diffusion [15] and conditional normalizing flows [CNFs, 16]. After training, amortized VI provides rapid sampling during inference [17, 18], exemplified by the WISE framework [19] for multi-D FWI problems. However, these methods may suffer from an *amortization gap* — implying that the amortized networks may only deliver suboptimal inference for a single observation at inference time, particularly when trained with limited examples or when there exists a discrepancy between training and inference [20]. Conversely, non-amortized VI dedicates entire computational resources to the online inference [21, 22, 23]. They result in more accurate inference, but the costly optimization has to be carried out repeatedly for new observations, and integrating realistic priors can be challenging since the prior needs to be embedded involving density evaluations [24].

This chapter introduces WISER as a semi-amortized VI framework to facilitate computationally feasible and reliable UQ for multi-D FWI without dimensionality reduction. Building on WISE, we train CNFs for efficient, suboptimal amortized inference, but then follow up with a crucial refinement step that only needs frugal use of the forward operator and its gradient. The refinement step aligns the posterior samples with the observation during inference, effectively bridging the amortization gap and enhancing inference accuracy.

Our contributions are organized as follows. We begin by outlining WISER in Algorithm 2. We explore the algorithm by explaining amortized VI with WISE, followed by computationally feasible multi-D physics-based refinement. The performance of WISER is demonstrated through realistic synthetic 2D case studies using the Compass model [25], showcasing improvements over WISE for both in- and out-of-distribution scenarios.

7.3 Amortized VI with WISE (lines 1—20)

WISER starts with an offline training phase that leverages conditional generative models to approximate the posterior distribution. This is achieved by WISE [19], which involves generating a training dataset (lines 3—9 of Algorithm 2) and training the CNFs (line 11—12).

7.3.1 Dataset generation (lines 3—9)

We begin by drawing N velocity models from the prior distribution, denoted by $p(\mathbf{x})$ (line 5). For each sample, $\mathbf{x}^{(i)}$, we simulate the observed data, $\mathbf{y}^{(i)}$, by performing the wave modeling and adding a random noise term (lines 6—7). Next, we compute common-image gathers [CIGs, 26] for each observed data with an initial smooth 1D migration-velocity model, \mathbf{x}_0 , which can be rather inaccurate. These CIGs, represented by $\bar{\mathbf{y}}^{(i)}$, are produced by applying the adjoint of the extended migration operator, $\overline{\nabla \mathcal{F}}(\mathbf{x}_0)^\top$, to the observed data. Using CIGs as the set of physics-informed summary statistics not only preserves information from the observed seismic data [27] but also enhances the training of CNFs in

Algorithm 2 WISER: full-Waveform variational Inference via Subsurface Extensions with Refinements

```

1: Offline training phase
2:
3: Dataset generation
4: for  $i = 1 : N$  do
5:    $\mathbf{x}^{(i)} \sim p(\mathbf{x})$ 
6:    $\boldsymbol{\epsilon}^{(i)} \sim p(\boldsymbol{\epsilon})$ 
7:    $\mathbf{y}^{(i)} = \mathcal{F}(\mathbf{x}^{(i)}) + \boldsymbol{\epsilon}^{(i)}$ 
8:    $\bar{\mathbf{y}}^{(i)} = \nabla \bar{\mathcal{F}}(\mathbf{x}_0)^\top \mathbf{y}^{(i)}$ 
9: end for
10:
11: Network training
12:  $\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \quad \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{2} \|f_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}; \bar{\mathbf{y}}^{(i)})\|_2^2 - \log |\det \mathbf{J}_{f_{\boldsymbol{\theta}}}| \right)$ 
13:
14: Online inference phase
15:
16:  $\bar{\mathbf{y}}_{\text{obs}} = \nabla \bar{\mathcal{F}}(\mathbf{x}_0)^\top \mathbf{y}_{\text{obs}}$ 
17: for  $i = 1 : M$  do
18:    $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ 
19:    $\mathbf{x}_i = f_{\boldsymbol{\theta}^*}^{-1}(h_{\phi}(\mathbf{z}_i); \bar{\mathbf{y}}_{\text{obs}})$ 
20: end for
21:
22: Physics-based refinements
23: for  $ii = 1 : \text{maxiter}_1$  do
24:   for  $i = 1 : M$  do
25:      $\mathbf{g}_i = \nabla_{\mathbf{x}_i} \left[ \frac{1}{2\sigma^2} \|\mathcal{F}(\mathbf{x}_i) - \mathbf{y}_{\text{obs}}\|_2^2 + \frac{1}{2\gamma^2} \|\mathbf{x}_i - f_{\boldsymbol{\theta}^*}^{-1}(h_{\phi}(\mathbf{z}_i); \bar{\mathbf{y}}_{\text{obs}})\|_2^2 \right]$ 
26:      $\mathbf{x}_i = \mathbf{x}_i - \tau \mathbf{g}_i$ 
27:   end for
28:   for  $iii = 1 : \text{maxiter}_2$  do
29:      $\mathcal{L}(\phi) = \sum_{i=1}^M \left[ \frac{1}{2\gamma^2} \|\mathbf{x}_i - f_{\boldsymbol{\theta}^*}^{-1}(h_{\phi}(\mathbf{z}_i); \bar{\mathbf{y}}_{\text{obs}})\|_2^2 + \frac{1}{2} \|h_{\phi}(\mathbf{z}_i)\|_2^2 - \log |\det \mathbf{J}_{h_{\phi}}| \right]$ 
30:      $\phi \leftarrow \text{ADAM}(\mathcal{L}(\phi))$ 
31:   end for
32: end for
33:
34: Output:  $\{f_{\boldsymbol{\theta}^*}^{-1}(h_{\phi}(\mathbf{z}_i); \bar{\mathbf{y}}_{\text{obs}})\}_{i=1}^M$  as samples of  $p(\mathbf{x}|\mathbf{y}_{\text{obs}})$ 

```

the next stage [28, 29], as they help to decode the wave physics, translating prestack data to the image (subsurface-offset) domain.

7.3.2 Network training (lines 11—12)

CNFs are trained with pairs of velocity models and CIGs via minimization of the objective in line 12. The symbol f_{θ} denotes the CNFs, characterized by their network weights, θ , and the Jacobian, $\mathbf{J}_{f_{\theta}}$. The term “normalizing” within CNFs implies their functionality to transform realizations of velocity models, $\mathbf{x}^{(i)}$, into Gaussian noise from a standard multivariate normal distribution (as defined by the ℓ_2 norm), conditioned on the summary statistics (CIGs).

7.3.3 Online inference (lines 14—20)

The aforementioned data generation and CNF training procedures conclude the offline training phase. During online inference, amortized VI is enabled by leveraging the inherent invertibility of CNFs. For a given observation, \mathbf{y}_{obs} , the online cost is merely generation of a single set of CIGs (line 16). Subsequently, the posterior samples are generated by applying the inverse of the CNFs to Gaussian noise realizations, conditioned on these CIGs (lines 18—19)¹.

7.4 Physics-based refinements (lines 22—32)

Consider a single observation, \mathbf{y}_{obs} , and its corresponding posterior samples, $\mathbf{x}_i \sim p(\mathbf{x} \mid \bar{\mathbf{y}}_{\text{obs}})$. The latent representations generated by the trained CNFs, $\hat{\mathbf{z}}_i = f_{\theta^*}(\mathbf{x}_i; \bar{\mathbf{y}}_{\text{obs}})$, may not conform exactly to the standard Gaussian distribution during inference. To address this issue, we follow [17] to apply latent space corrections to fine-tune the trained CNFs. This involves integrating a shallower, yet invertible, network², specifically trained to map realizations of true Gaussian noise to the corresponding latent codes, $\hat{\mathbf{z}}_i$. Adhering to a transfer learning approach, we maintain the weights of the trained CNFs while solely updating the

¹We slightly abuse the notation to assume h_{ϕ} as an identity operator here.

²For linear inverse problems in seismic imaging, [17] show that an elementwise scaling and shift mechanism is adequate to bridge the gap. However, given the complex, non-convex nature of FWI, we employ h_{ϕ} as a generic invertible network.

weights of the shallower network by minimizing the following objective:

$$\underset{\phi}{\text{minimize}} \quad \mathbb{E}_{\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \left[\frac{1}{2\sigma^2} \|\mathcal{F} \circ f_{\theta^*}^{-1}(h_{\phi}(\mathbf{z}); \bar{\mathbf{y}}_{\text{obs}}) - \mathbf{y}_{\text{obs}}\|_2^2 + \frac{1}{2} \|h_{\phi}(\mathbf{z})\|_2^2 - \log |\det \mathbf{J}_{h_{\phi}}| \right]. \quad (7.1)$$

Here, the refinement network, h_{ϕ} , mitigates the amortization gap by adjusting the latent variable \mathbf{z} before feeding it to the inverse of the trained CNFs, $f_{\theta^*}^{-1}$. Intuitively, minimizing the first terms ties the posterior samples closer to the observed data. The second and third terms prevent the corrected latent space from being far from the Gaussian distribution, which implicitly takes advantage of the prior information existing in the amortized training phase.

Equation 7.1 offers a fine-tuning approach that leverages the full multi-D wave physics to refine the amortized VI framework for a single observation at inference phase. However, it introduces notable computational demands because it necessitates the coupling of the modeling operator and the networks. Specifically, every update to the network weights, ϕ , requires costly wave modeling operations. Given that network training typically involves numerous iterations, these computational demands can render it impractical for realistic FWI applications.

To relieve this computational burden, we adopt a strategy from [30] to reformulate Equation 7.1 into a weak form by allowing the network output to be only weakly enforced (in an ℓ_2 sense) to be the corrected velocity models. The objective function for this weak formulation reads:

$$\underset{\mathbf{x}_{1:M}, \phi}{\text{minimize}} \quad \left[\frac{1}{M} \sum_{i=1}^M \frac{1}{2\sigma^2} \|\mathcal{F}(\mathbf{x}_i) - \mathbf{y}_{\text{obs}}\|_2^2 + \frac{1}{2\gamma^2} \|\mathbf{x}_i - f_{\theta^*}^{-1}(h_{\phi}(\mathbf{z}_i); \bar{\mathbf{y}}_{\text{obs}})\|_2^2 + \frac{1}{2} \|h_{\phi}(\mathbf{z}_i)\|_2^2 - \log |\det \mathbf{J}_{h_{\phi}}| \right]. \quad (7.2)$$

We strategically decouple the computationally expensive forward operator, \mathcal{F} , from the more cheap-to-evaluate networks, f_{θ^*} and h_{ϕ} . This is achieved in a penalty form with the assumption that the misfit between the network outputs and the posterior samples adheres to a Gaussian distribution, $\mathcal{N}(\mathbf{0}, \gamma^2 \mathbf{I})$, where γ is a hyperparameter dictating the trade-off between data misfit and regularization. Setting γ to 0 equates this weak formulation to the constrained formulation in Equation 7.1. This weak formulation also supports optimization strategies for updating the velocity models with physical constraints [31, 32] and multiscale optimization techniques [33].

WISER takes full computational advantage of this weak formulation by employing a nested loop structure. The outer loop is dedicated to updating M velocity models, \mathbf{x}_i , through costly gradient descent steps (lines 24—27 of Algorithm 2), while the inner loop (lines 28—31) focuses on more updates (with the ADAM optimizer) to network weights, ϕ , without computationally expensive physics modeling. To achieve a balance, we conduct $\text{maxiter}_2 = 128$ iterations in the inner loop. After refinements, WISER first evaluates the refined network on the latent variables to obtain refined latent codes. Subsequently, the amortized network uses the refined codes conditioned on the CIGs to compute the corrected posterior samples (line 34).

7.5 Case studies

Evaluation of WISER is conducted through synthetic case studies utilizing 2D slices of the Compass model and 2D acoustic wave physics. The parameter of interest is discretized into 512×256 gridpoints with a spatial resolution of 12.5 m, resulting in over 10^5 degrees of freedom. The forward operator, \mathcal{F} , simulates acoustic data with absorbing boundaries. A Ricker wavelet with a central frequency of 15 Hz and an energy cut below 3 Hz is employed. We use 512 sources towed at 12.5 m depth and 64 ocean-bottom nodes (OBNs) located at jittered sampled horizontal positions [34]. We employ source-receiver reciprocity during the modeling and sensitivity calculations. The observed data, \mathbf{y}^{obs} , is perturbed with band-

limited Gaussian noise to achieve a signal-to-noise ratio (S/N) of 12 dB. The training of the CNFs uses $N = 800$ pairs of velocity models and CIGs. To demonstrate WISER’s efficacy in mitigating the amortization gap, we compare results from WISE and WISER under two scenarios during inference:

- (i) observed shot data is generated using an in-distribution velocity model with the same forward operator;
- (ii) observed shot data is produced by an out-of-distribution (OOD) velocity model and also a slightly altered forward operator.

7.5.1 Case 1: in distribution

The ground-truth velocity model is an unseen 2D slice from the Compass model, shown in Figure 7.1a. Following Algorithm 2, we initiate WISER by drawing $M = 16$ Gaussian noise realizations to create the initial set of 16 velocity models, depicted in Figure 7.1b. To minimize computational demands, stochastic gradients [35] are calculated in line 25 of Algorithm 2. Each particle’s gradient is estimated using only 1 randomly selected OBN gather from the observed data. We also add box constraints to the velocity models to restrict their range to 1.48 to 4.44 km/s. Following $\text{maxiter}_1 = 80$ outer iterations—equivalent to 20 data passes or 2560 PDE solves³—we obtain the posterior samples from WISER in Figure 7.1c.

Observations

The conditional mean estimate (CM) from WISE lacks finer details, particularly beneath the unconformity at depths below 2.4 km (in red). This is attributed to the excessive variability in structural details of the posterior samples, visible on the right panel of Figure 7.1b.

In contrast, WISER generates more consistent and accurate posterior samples. In Figure 7.1c, the right panel shows that the uncertainty from WISER is reduced below the un-

³1 PDE solve means solving the wave equation for a single source. A gradient requires 2 PDE solves (forward and adjoint).

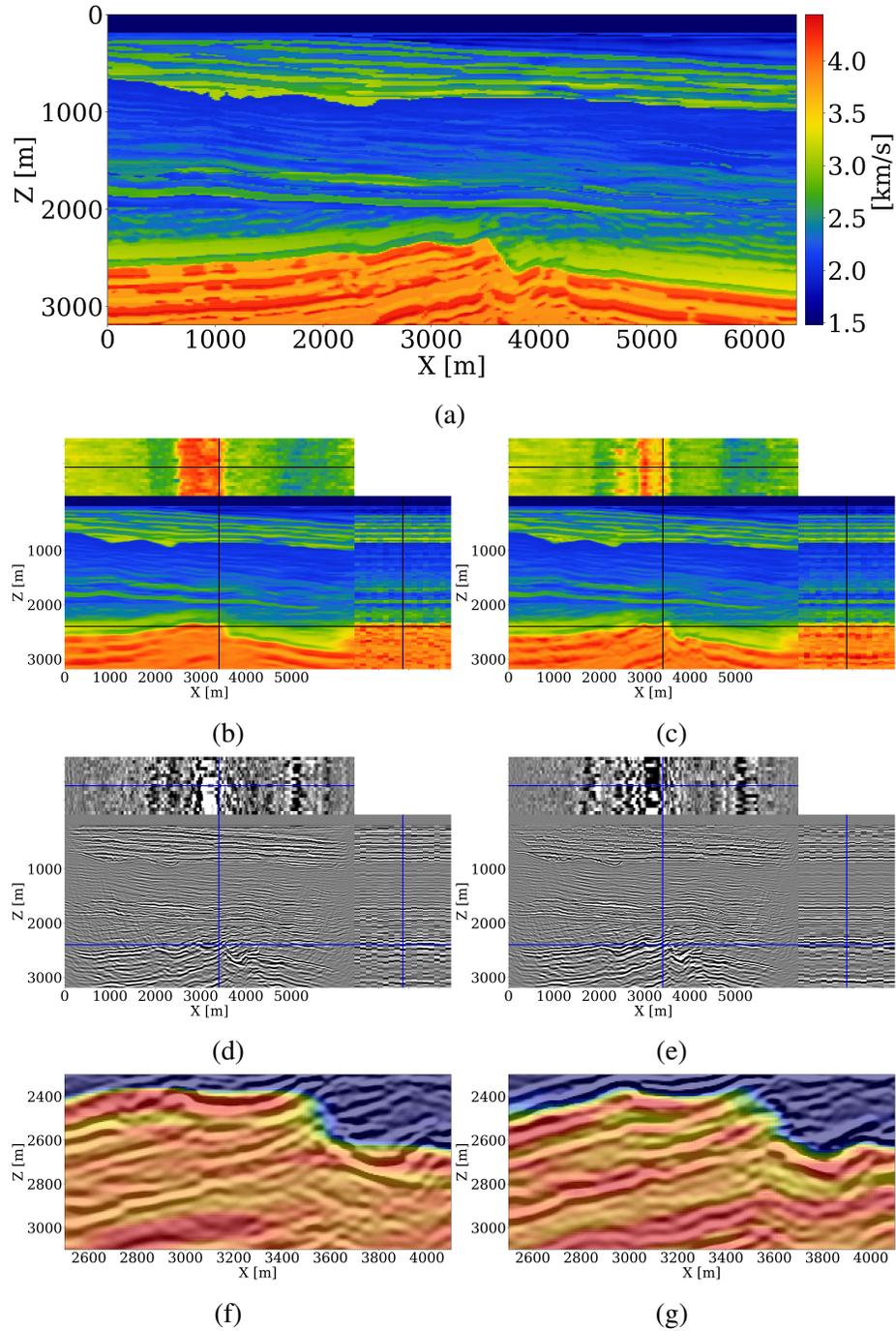


Figure 7.1: Comparison between WISE and WISER for an in-distribution case. (a) Unseen ground-truth velocity model. (b) Estimated velocity models from WISE. The conditional mean estimate (CM) is shown in the center. For posterior samples, horizontal traces at $Z = 2.7$ km and vertical traces at $X = 3.6$ km are displayed on the top and on the right, respectively. (d) Imaged reflectivity samples from WISE. (f) Zoom-in views of (d) overlaying on the CM of WISE. (c)(e)(g) are the counterparts from WISER, showcasing significant improvements.

conformity. The upper panel illustrates that the uncertainty is more focused on the dipping events at the unconformity, highlighting areas of poor illumination.

Impact on imaging

To assess the impact of uncertainty in velocity models on downstream tasks, we conduct high-frequency imaging using a Ricker wavelet with central frequency of 30 Hz and compare the imaged reflectivities derived from the posterior samples of both WISE and WISER, shown in Figure 7.1d and Figure 7.1e, respectively.

The imaged reflectivities produced by CM from WISER exhibit superior continuity and a better correlation with the CM migration-velocity model, particularly noticeable in Figure 7.1g under the unconformity. Also, reflectivity samples produced by WISER demonstrate improved alignment among themselves compared to those produced by WISE. In addition, notable vertical shifts observed in the imaged reflectivities from WISE to WISER indicate significant adjustments in the positioning of subsurface reflectors, underlining the necessity of the refinement procedure for precisely estimating migration-velocity models that locate subsurface reflectors more accurately.

7.5.2 Case 2: out of distribution

To test the robustness and adaptability of WISER when faced with unexpected variations at inference, we also evaluate WISER’s performance under OOD scenarios. We introduce alterations to the velocity model depicted in Figure 7.1a through an elementwise perturbation shown in Figure 7.2a. This manipulation modifies the velocity values across different depth levels, resulting in a significant shift in their statistical distribution, illustrated in Figure 7.2b. We use the perturbed velocity as the unseen ground-truth velocity model in this case study, shown in Figure 7.2c. To further expand the amortization gap, we modify the encoding of the forward operator by introducing a higher amplitude of band-limited Gaussian noise (S/N 0 dB).

These complexities present substantial challenges for WISE, leading to biased inference results as depicted in Figure 7.2d. The yellow histograms in Figure 7.2b show that the velocity values of the posterior samples from WISE closely resemble those of the original velocity model, despite the different distribution of the ground-truth velocity model. This indicates that WISE tends to incorporate an inductive bias from the training samples. In WISER, we conduct $\text{maxiter}_1 = 160$ outer iterations, using $M = 16$ particles and 1 OBN per gradient. We also employ the frequency continuation method [33] to compute the gradient in line 25 of Algorithm 2, transitioning gradually from low-frequency to high-frequency data. This results in 40 datapasses or 5120 PDE solves in total.

Observations

WISER produces more accurate posterior samples shown in Figure 7.2e. Furthermore, the statistical distribution of the velocity values in the WISER posterior samples (green histogram in Figure 7.2b) aligns better with the distribution of the unseen ground-truth velocity values (blue histogram in Figure 7.2b), demonstrating WISER’s robustness against potential distribution shifts during inference.

To further showcase WISER’s robustness under severe measurement noise, we compare a posterior sample from WISER (Figure 7.2k) to a velocity model estimated by FWI (Figure 7.2j), derived by minimizing only the data likelihood (the first term in line 25 of Algorithm 2), while starting from the same initial model as WISER. The FWI result is significantly impacted by noise, while the posterior samples from WISER remain relatively noise-free and capture all pertinent geological structures.

Impact on imaging

The imaging results from WISE (Figure 7.2f) and WISER (Figure 7.2g) reveal noticeable discrepancies in quality. The CM migration-velocity model from WISE leads to discontinuities in the imaged reflectivities, particularly at the horizontal layer around 1.8 km depth

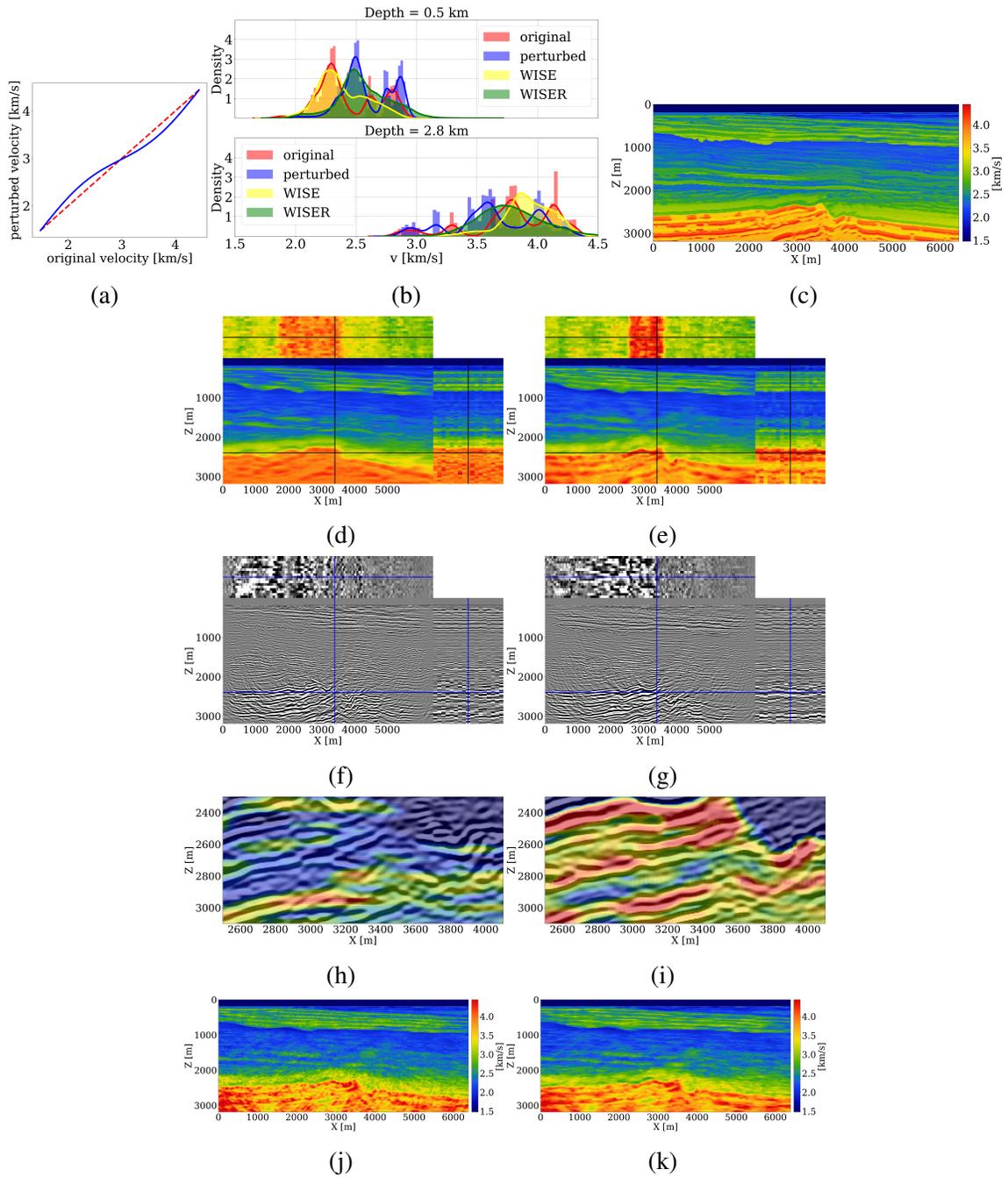


Figure 7.2: OOD case study. (a) Curves for velocity-value perturbations; (b) histograms of values at the depth of 0.5 km and 2.8 km in the original velocity model (Figure 7.1a), perturbed velocity model (Figure 7.2c), posterior samples of WISE, and WISER, shown in red, blue, yellow and green color, respectively. (c)—(i) Comparison between WISE and WISER. The ordering remains the same as in Figure 7.1. (j) FWI result starting with a posterior sample from WISE. (k) A posterior sample from WISER.

and more so below the unconformity. In contrast, the CM from WISER significantly improves the continuity of the imaged reflectivities across the entire seismic section. The imaged reflectivity samples from WISER not only show better consistency among themselves but also align more accurately with the estimated CM migration-velocity model, particularly visible in Figure 7.2i.

7.6 Discussion and conclusions

The primary contribution of WISER is to leverage both genAI and physics to achieve a semi-amortized VI framework for scalable ($D \geq 2$) and reliable UQ for FWI even in situations where local approximations are unsuitable. At its core, WISER harnesses the strengths of both amortized and non-amortized VI: the amortized posterior obtained through offline training provides a low-fidelity but fast mapping, and the physics-based refinements offer reliable and accurate inference. Both approaches benefit from information preservation exhibited by CIGs, rendering our inference successful where conventional FWI fails due to cycle skipping.

Compared to MCMC methods that rely on low-dimensional parameterizations, WISER does not impose intrinsic dimensionality reductions or simplifications of the forward model. Therefore, WISER is capable of delivering full-resolution UQ for realistic multi-D FWI problems. CNFs are primed for large-scale 3D inversion thanks to their invertibility, which allows for memory-efficient training and inference [36].

Compared to non-amortized VI methods, WISER also requires significantly less computational resources during inference. This is because WISE, as a precursor of WISER, already provides near-accurate posterior samples, making the refinement procedure computationally feasible. [22] show that non-amortized VI, without access to realistic prior information, requires $O(10^6)$ to $O(10^8)$ PDE solves, while WISER only needs $O(10^3)$ PDE solves. Apart from the computational cost reduction, WISER ensures that the posterior samples realistically resemble Earth models, thanks to the integration of the conditional

prior information from WISE. Contrary to non-amortized VI, which needs density evaluations to embed the prior (i.e., $p(\mathbf{x})$ in line 5 of Algorithm 2) to produce realistic Earth models, WISER only needs access to samples of the prior distribution (i.e., $\mathbf{x}^{(i)}$ in line 5).

Opportunities for future research remain. Although case 2 demonstrates the robustness of WISER concerning OOD issues, these issues could be fundamentally addressed by diversifying the training set of WISE through a foundation model [37]. Also, our OOD case study has not yet explored scenarios where the likelihood term is more pathologically misspecified, such as the presence of unremoved shear wave energy outside the range of the forward operator, which calls for further investigations. Our approach will also benefit from calibration of the estimated posterior, including application of WISE(R) in 3D [38].

In conclusion, this chapter sets the stage for deploying genAI models to facilitate high-dimensional Bayesian inference with computationally intensive multi-D forward operators. Deep learning and AI have been criticized for their reliance on realistic training samples, but WISER alleviates this reliance and still offers computationally feasible and reliable inference through a blend of offline training and online frugal physics-based refinements, preparing our approach for large 3D deployment.

7.7 References

- [1] J. Virieux and S. Operto, “An overview of full-waveform inversion in exploration geophysics,” *GEOPHYSICS*, vol. 74, no. 6, WCC1–WCC26, Nov. 2009 (page 155).
- [2] K. Cranmer, J. Brehmer, and G. Louppe, “The frontier of simulation-based inference,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 055–30 062, 2020 (page 156).
- [3] Z. Fang, C. Da Silva, R. Kuske, and F. J. Herrmann, “Uncertainty quantification for inverse problems with weak partial-differential-equation constraints,” *Geophysics*, vol. 83, no. 6, R629–R647, 2018 (page 156).
- [4] S. D. Keating and K. A. Innanen, “Null-space shuttles for targeted uncertainty analysis in full-waveform inversion,” *Geophysics*, vol. 86, no. 1, R63–R76, 2021 (page 156).
- [5] M. Izzatullah, M. Ravasi, and T. Alkhalifah, “Physics reliable frugal uncertainty analysis for full waveform inversion,” *arXiv preprint arXiv:2305.07921*, 2023 (page 156).
- [6] A. Hoffmann, R. Brossier, L. Métivier, and A. Tarayoun, “Local uncertainty quantification for 3-d time-domain full-waveform inversion with ensemble kalman filters: Application to a north sea obc data set,” *Geophysical Journal International*, vol. 237, no. 3, pp. 1353–1383, 2024 (page 156).
- [7] M. K. Cowles and B. P. Carlin, “Markov chain monte carlo convergence diagnostics: A comparative review,” *Journal of the American statistical Association*, vol. 91, no. 434, pp. 883–904, 1996 (page 156).
- [8] Z. Fang, H. Fang, and L. Demanet, “Deep generator priors for bayesian seismic inversion,” in *Advances in geophysics*, vol. 61, Elsevier, 2020, pp. 179–216 (page 156).
- [9] Z. Liang, F. Wellmann, and O. Ghattas, “Uncertainty quantification of geologic model parameters in 3d gravity inversion by hessian-informed markov chain monte carlo,” *Geophysics*, vol. 88, no. 1, G1–G18, 2023 (page 156).
- [10] X. Wei, J. Sun, and M. K. Sen, “Quantifying uncertainty of salt body shapes recovered from gravity data using trans-dimensional markov chain monte carlo sampling,” *Geophysical Journal International*, vol. 232, no. 3, pp. 1957–1978, 2023 (page 156).
- [11] X. Wei, J. Sun, and M. Sen, “3d monte carlo geometry inversion using gravity data,” *Geophysics*, vol. 89, no. 3, pp. 1–62, 2024 (page 156).

- [12] X. Wei, J. Sun, and M. K. Sen, “Reconstruction of multiple target bodies using trans-dimensional bayesian inversion with different constraints,” *IEEE Transactions on Geoscience and Remote Sensing*, 2024 (page 156).
- [13] N. Dhabaria and S. C. Singh, “Hamiltonian monte carlo based elastic full-waveform inversion of wide-angle seismic data,” *Geophysical Journal International*, vol. 237, no. 3, pp. 1384–1399, 2024 (page 156).
- [14] X. Zhang, M. A. Nawaz, X. Zhao, and A. Curtis, “An introduction to variational inference in geophysical inverse problems,” in *Advances in geophysics*, vol. 62, Elsevier, 2021, pp. 73–140 (page 156).
- [15] L. Baldassari, A. Siahkoohi, J. Garnier, K. Solna, and M. V. de Hoop, “Conditional score-based diffusion models for bayesian inference in infinite dimensions,” *Advances in Neural Information Processing Systems*, vol. 36, 2024 (page 156).
- [16] C. Winkler, D. Worrall, E. Hoogeboom, and M. Welling, “Learning likelihoods with conditional normalizing flows,” *arXiv preprint arXiv:1912.00042*, 2019 (page 156).
- [17] A. Siahkoohi, G. Rizzuti, R. Orozco, and F. J. Herrmann, “Reliable amortized variational inference with physics-based latent distribution correction,” *Geophysics*, vol. 88, no. 3, R297–R322, 2023 (pages 156, 159).
- [18] R. Orozco, M. Louboutin, A. Siahkoohi, G. Rizzuti, T. van Leeuwen, and F. J. Herrmann, “Amortized normalizing flows for transcranial ultrasound with uncertainty quantification,” in *Medical Imaging with Deep Learning*, 2023 (page 156).
- [19] Z. Yin, R. Orozco, M. Louboutin, and F. J. Herrmann, “Wise: Full-waveform variational inference via subsurface extensions,” *Geophysics*, vol. 89, no. 4, pp. 1–31, 2024 (pages 156, 157).
- [20] J. Marino, Y. Yue, and S. Mandt, “Iterative amortized inference,” in *International Conference on Machine Learning*, PMLR, 2018, pp. 3403–3412 (page 156).
- [21] X. Zhao, A. Curtis, and X. Zhang, “Bayesian seismic tomography using normalizing flows,” *Geophysical Journal International*, vol. 228, no. 1, pp. 213–239, 2022 (page 156).
- [22] X. Zhang, A. Lomas, M. Zhou, Y. Zheng, and A. Curtis, “3-d bayesian variational full waveform inversion,” *Geophysical Journal International*, vol. 234, no. 1, pp. 546–561, 2023 (pages 156, 167).
- [23] X. Zhang and A. Curtis, “Bayesian variational time-lapse full waveform inversion,” *Geophysical Journal International*, vol. 237, no. 3, pp. 1624–1638, 2024 (page 156).

- [24] J. Kruse, G. Detommaso, U. Köthe, and R. Scheichl, “Hint: Hierarchical invertible neural transport for density estimation and bayesian inference,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 2021, pp. 8191–8199 (page 156).
- [25] C. Jones, J. Edgar, J. Selvage, and H. Crook, “Building complex synthetic models to evaluate acquisition geometries and velocity inversion technologies,” in *74th EAGE Conference and Exhibition incorporating EUROPEC 2012*, European Association of Geoscientists & Engineers, 2012, cp–293 (page 157).
- [26] J. Hou and W. W. Symes, “Accelerating extended least-squares migration with weighted conjugate gradient iteration,” *GEOPHYSICS*, vol. 81, no. 4, S165–S179, Jul. 2016 (page 157).
- [27] F. ten Kroode, “An omnidirectional seismic image extension,” *Inverse Problems*, vol. 39, no. 3, p. 035 003, 2023 (page 157).
- [28] S. T. Radev, U. K. Mertens, A. Voss, L. Ardizzone, and U. Köthe, “Bayesflow: Learning complex stochastic models with invertible neural networks,” *IEEE transactions on neural networks and learning systems*, vol. 33, no. 4, pp. 1452–1466, 2020 (page 158).
- [29] R. Orozco, A. Siahkoochi, G. Rizzuti, T. van Leeuwen, and F. Herrmann, “Adjoint operators enable fast and amortized machine learning based bayesian uncertainty quantification,” *Medical Imaging 2023: Image Processing*, I. Išgum and O. Colliot, Eds., Apr. 3, 2023 (page 158).
- [30] A. Siahkoochi, G. Rizzuti, and F. J. Herrmann, “Weak deep priors for seismic imaging,” in *SEG Technical Program Expanded Abstracts 2020*, Society of Exploration Geophysicists, 2020, pp. 2998–3002 (page 160).
- [31] E. Esser, L. Guasch, T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, “Total variation regularization strategies in full-waveform inversion,” *SIAM Journal on Imaging Sciences*, vol. 11, no. 1, pp. 376–406, 2018 (page 161).
- [32] B. Peters, B. R. Smithyman, and F. J. Herrmann, “Projection methods and applications for seismic nonlinear inverse problems with multiple constraints,” *Geophysics*, vol. 84, no. 2, R251–R269, 2019 (page 161).
- [33] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, “Multiscale seismic waveform inversion,” *Geophysics*, vol. 60, no. 5, pp. 1457–1473, 1995 (pages 161, 165).
- [34] G. Hennenfent and F. J. Herrmann, “Simply denoise: Wavefield reconstruction via jittered undersampling,” *Geophysics*, vol. 73, no. 3, pp. V19–V28, 2008 (page 161).

- [35] F. J. Herrmann *et al.*, “Frugal full-waveform inversion: From theory to a practical algorithm,” *The Leading Edge*, vol. 32, no. 9, pp. 1082–1092, 2013 (page 162).
- [36] R. Orozco *et al.*, “Invertiblenetworks.jl: A julia package for scalable normalizing flows,” *arXiv preprint arXiv:2312.13480*, 2023 (page 167).
- [37] H. Sheng, X. Wu, X. Si, J. Li, S. Zhang, and X. Duan, “Seismic foundation model (sfm): A new generation deep learning model in geophysics,” *arXiv preprint arXiv:2309.02791*, 2023 (page 168).
- [38] R. Orozco, A. Siahkoohi, M. Louboutin, and F. J. Herrmann, “Aspire: Iterative amortized posterior inference for bayesian inverse problems,” *arXiv preprint arXiv:2405.05398*, 2024 (page 168).

CHAPTER 8

CONCLUSION

In conclusion, this thesis presents several innovative methods based on scientific machine learning for solving geophysical inverse problems at scale. Two main problems addressed in this thesis are GCS monitoring and FWI. They both draw their complexity from several computational challenges: the intensive cost, non-convexity and ill-posedness of the modeling operator, and high-dimensional multimodal solution spaces. To meet these challenges, this thesis makes contributions by

- designing interoperable and differentiable programming framework that supports learned multiphysics inversion at scale;
- exploring the usage of deep neural networks as surrogate models for cost-effective Bayesian inference in large-scale inverse problems, which includes learning either the forward map or the (nonunique) inverse map;
- employing the proposed scientific algorithms to solve FWI and GCS monitoring problems.

8.1 Scalable, interoperable, and differentiable programming framework

Time-lapse monitoring of GCS requires resolving multiple physical systems, including fluid-flow, rock, and wave physics. Specifically, multiphysics inversion requires different physics modeling operators to interoperate both in the forward simulation and the gradient calculations. Conventional software modules for different physics modeling operators typically rely on monolithic low-level implementations, which are difficult and costly to maintain and innovate. While these practices lead to performant software for a specific

problem, they pose significant challenges in innovation in geophysical research and development, especially in the case for GCS monitoring where different physics modeling operators (and their associated software modules) need to be coupled.

To meet these challenges, I proposed a scalable, interoperable, and differentiable programming framework in Chapter 2. Based on high-level math-inspired software abstraction techniques, I presented numerical examples to leverage customized AD rules for different software modules in the end-to-end permeability inversion framework. I also illustrated that this adaptable and flexible software framework supports extremely fast innovations in geophysical research, with a salient example where the numerical simulators in the fluid-flow physics module are replaced by trained deep neural surrogates during inversion. This effortless substitution demonstrates that the proposed framework, thanks to the modern software practices, facilitates effortless substitution of deep learning models in lieu of numerical simulators. The feasibility study presented in Chapter 3 also verifies the computational feasibility of this software framework on the Compass model. The geological structures of this Compass model were derived from well logs and imaged seismic from the South-West North Sea area — a region under consideration for GCS [1]. Because the software framework leverages differentiable programming principles to calculate gradients with customized AD rules, I was able to conduct a multiparameter inversion study in Chapter 3, where the unknown parameter of the interest, the porosity model, appears as an input in more than one physics modeling operators, and has a deterministic relationship with the permeability model. Calculating the gradient of the objective function with respect to the porosity can be rather complex, error-prone, and labor-intensive as it requires deriving cross gradients by hand if different physics modeling operators are separated. In contrast, integration of different software stacks in the proposed differentiable programming framework allows for effortless and accurate gradient calculation.

8.2 Surrogate-assisted inversion with learned modeling operators

After presenting the scalable and differentiable programming framework for surrogate-assisted inversion, Chapter 4 further analyzes the accuracy of deep neural surrogates during inversion. I proposed a proof of concept case study to examine the accuracy of surrogates during gradient-based optimization, and observe that the accuracy of the surrogates decreases along iterations. This is because there is no guarantee that the model parameter iterates remain in the statistical distribution on which the surrogate models were initially trained. As a consequence, the surrogates may not produce accurate simulations during the inversion, which can lead to erroneous inversion results eventually.

To safeguard the accuracy of surrogate models during inversion, I proposed an algorithm that combines a trained generative model with the learned surrogates to achieve reliable and cost-effective inversion. Through a motivational example, I illustrated that NFs, trained on the same statistical distribution as FNO surrogates, can be used to effectively project an out-of-distribution sample to an in-distribution sample. This was achieved by shrinkage in the latent space of the NFs. I also demonstrated that applying latent space shrinkage to an out-of-distribution sample can successfully reduce the FNO prediction error on the projected (in-distribution) sample.

To take advantage of the interplay between NFs and FNOs, I proposed a surrogate-assisted inversion algorithm with progressively relaxed learned constraints. This results in a continuation scheme where the constraints on the latent space are gradually relaxed throughout the iterations so that the model iterates always remain in-distribution and the objective function can be minimized eventually. I conducted a series of case studies to examine this algorithm with different kinds of observed data in the GCS monitoring problem, and verified that the accuracy of the FNO surrogates remains relatively constant in all these studies, while the unconstrained inversion leads to gradually decreasing accuracy of the FNOs throughout the inversion.

To further investigate the GCS monitoring problem, specifically in the scenario where the CO₂ plumes leaks and therefore fails to follow the multiphase flow equations, I proposed to use joint recovery model in Chapter 5 to jointly image the CO₂ saturation history from multiple time-lapse seismic surveys. This linearized joint inversion framework relies on a relatively mild assumption that the Earth properties in different surveys share a fictitious common component. By explicitly solving for this common component and the innovations of each survey with respect to this common component, I demonstrated a streamlined workflow for time-lapse imaging where the time-lapse changes in wave properties are robustly estimated from multiple seismic surveys with non-replicated acquisition geometries, which shows great potential to drastically reduce the operational cost in time-lapse acquisition and processing. I finalized this automated workflow by incorporating a deep neural classifier to detect potential CO₂ leakage in the estimated time-lapse seismic images.

8.3 Semi-amortized variational inference with physics-informed summary statistics

In contrast to previous chapters where the modeling operators are learned, Chapter 6 and Chapter 7 take a learned variational inference approach to tackle the FWI problem. Due to the nonuniqueness and nonconvexity of the FWI objective, there can be multiple model parameters that fit the observation adequately. To this end, Bayesian inference and uncertainty quantification techniques are utilized to solve for the posterior distribution of model parameters (velocity models) conditioned on the observation (seismic data). Due to the high-dimensional multimodal posterior distribution and the computationally expensive forward modeling operator, VI is chosen for its scalability to large-scale problems.

Chapter 6 proposed an amortized VI approach for solving FWI. The notable challenge of amortized VI for FWI is that the mapping between model parameters and observations is extremely complicated, and therefore, very difficult to learn. To meet this challenge, physics-informed summary statistics is necessary to encapsulate the information in the ob-

served data [2]. While prior art in [3] showed that adjoint operators in linear inverse problems preserves information, FWI is inherently a nonlinear and non-convex inverse problem, implying that the adjoint of the Jacobian operator cannot preserve all the information in the data if the initial model is far from the ground truth model. To this end, I proposed to use the common-image gathers as the physics-informed summary statistics and argued that it preserves most of the information in the observed data [4] even if an inaccurate initial model is used to calculate them. I designed representative case studies that compare the inference results when the summary statistics is only the gradient (obtained by applying the adjoint of the standard Jacobian operator on the data) or the CIGs (obtained by applying the adjoint of the extended Jacobian), and demonstrated that CIGs successfully encapsulate more information in the data to inform the posterior.

Chapter 7 further improves the inference framework in Chapter 6 through a physics-based refinement process. This refinement process is crucial when an amortization gap exists in an amortized VI approach. To close this gap, I proposed a semi-amortized variational inference framework that finetunes the amortized networks via frugal usage of the wave physics. This is achieved by composing a refinement network to the amortized network, and optimizing the network weights in the refinement network to tie the posterior samples better to the observed data through likelihood evaluations. I verified the efficacy of this physics-based refinement approach in both in-distribution and out-of-distribution scenarios and illustrated that the refined posterior samples are less biased by the training samples compared to the amortized posterior samples. Chapter 6 and Chapter 7 together presents a computationally feasible and accurate VI approach for large-scale inverse problems with computationally expensive forward modeling operators. I also demonstrated that the uncertainty in the FWI solution is crucial for downstream imaging tasks where the positioning and amplitude variations in the imaged reflectivities can also be accessed.

8.4 Current limitations and future directions

8.4.1 Chapter 3

A multiparameter inversion study, including relative permeability curves and capillary pressure functions to only name a few, would allow further investigation on potential crosstalks between these parameters, paving the way for a more practical usage of the time-lapse seismic data for inverting reservoir parameters. Also, an uncertainty quantification study is worthwhile for further investigation to find different permeability models that produce the same CO₂ saturation history. Moreover, the impact of pore pressure and geomechanical effects can be considered to enhance this framework. An upscaling study is necessary to examine the resolution of this permeability inversion framework when multimodal observations (well measurements, time-lapse seismic data) are jointly inverted.

8.4.2 Chapter 4

Parameterization

The Algorithm 1 proposed in Chapter 4 presents a proof of concept to use trained NFs as a learned constraint during surrogate-assisted inversion to improve the accuracy of the surrogate throughout the iterations. Despite the success in the preliminary experiments, there are alternative ways to carry out the optimization and add the constraint. Future research can examine the inversion trajectory by juxtaposing inversion results obtained by two different parameterizations, shown in Equation 8.1 and Equation 8.2, respectively.

$$\underset{\mathbf{z}}{\text{minimize}} \quad \|\mathcal{H} \circ \mathcal{S}_{\theta^*} \circ \mathcal{G}_{\mathbf{w}^*}(\mathbf{z}) - \mathbf{d}\|_2^2 \quad \text{subject to} \quad P_Z(\mathbf{z}) \geq \Psi_z, \quad (8.1)$$

$$\underset{\mathbf{K}}{\text{minimize}} \quad \|\mathcal{H} \circ \mathcal{S}_{\theta^*}(\mathbf{K}) - \mathbf{d}\|_2 \quad \text{subject to} \quad P_K(\mathbf{K}) \geq \Psi_K. \quad (8.2)$$

In these formulations, \mathcal{S}_{θ^*} represents the trained FNO with optimized network weights,

θ^* . $\mathcal{G}_{\mathbf{w}^*}$ represents the trained NF, with optimized network weights, \mathbf{w}^* . \mathcal{H} is the measurement operator on the solution of the PDEs, and \mathbf{d} is the observed data. Ψ_z and Ψ_K represents the threshold level for the probability density in the latent space, $P_Z(\mathbf{z})$, and the density in the image space, $P_K(\mathbf{K})$, respectively. These two probability density functions can be evaluated via the following:

$$\begin{aligned}
-\log P_Z(\mathbf{z}) &= \frac{1}{2}\|\mathbf{z}\|_2^2 + \text{const} \\
-\log p_K(\mathbf{K}) &= -\log p_Z(\mathcal{G}_{\mathbf{w}^*}^{-1}(\mathbf{K})) + \log \det |\mathbf{J}_{\mathcal{G}_{\mathbf{w}^*}}(\mathbf{K})| \\
&= -\frac{1}{2}\|\mathcal{G}_{\mathbf{w}^*}^{-1}(\mathbf{K})\|_2^2 + \log \det |\mathbf{J}_{\mathcal{G}_{\mathbf{w}^*}}(\mathbf{K})| + \text{const}.
\end{aligned} \tag{8.3}$$

In Equation 8.3, const denotes constant terms not related to either \mathbf{z} or \mathbf{K} . Based on these different parameterization routines, a constrained formulation can be reached with either Equation 8.1 or Equation 8.2 if the threshold levels, Ψ_z and Ψ_K , are gradually relaxed (decreasing) throughout the constrained iterations. Both of these formulations can begin with an unknown, either \mathbf{z} or \mathbf{K} , with high likelihood, and then gradually relax the constraint to allow for data fitting. In fact, Equation 4.5 is equivalent to Equation 8.1. Optimizing over the latent variable, \mathbf{z} , brings the advantage that the threshold level is relatively easier to choose, because \mathbf{z} follows multivariate Gaussian distribution. Choosing a threshold level for Equation 8.2 is comparatively more difficult because there is not easily a value for reference. A rough estimation of the likelihood for the unseen ground truth model parameter can be explored and used to examine the optimization routines in Equation 8.1 and Equation 8.2 to investigate the choice of parameterizations. These parameterization routines both exist in the literature [5, 6]. It is an enticing avenue of future research to examine these formulations specifically for surrogate-assisted inversion, where the generative network is utilized to improve the accuracy of the surrogate model.

Derivative-informed surrogate-assisted inversion

Although Algorithm 1 demonstrates efficacy to enhance the accuracy of the surrogate models during forward evaluation in the inversion process, the gradient of the surrogate models with respect to the model parameters is not guaranteed. This is because the training objective in Equation 4.2 only includes a misfit between the forward evaluations of the simulator and the surrogate model. If the simulator is already differentiable, a hybrid objective function can be considered to include both misfit in the forward evaluation and the misfit in gradient calculations [7]. Future work can investigate the benefit and necessity for accurate gradient calculations by comparing the results using surrogate models trained by two different objective functions. Also, it is interesting to investigate the optimal fiducial point to compare the gradients during training.

Overall, future investigations would benefit from a more detailed mathematical analysis on a simpler problem to draw theoretical conclusions regarding how the projection improves the accuracy of the surrogate models. The algorithm would also benefit from a case study in a larger scale with more complex geologies in the model parameter (permeability) to validate its practicality.

8.4.3 Chapter 5

The joint inversion scheme presented in Chapter 5 only assumes the existence of a common component shared by all time-lapse surveys, but has not yet taken advantage of the fact that the time-varying changes of the Earth properties are induced by CO₂ plumes. To further enhance the joint inversion scheme, an alternative way to parameterize the joint recovery model in GCS monitoring is to assume the common component is the pre-injection brine-filled wavespeed at the baseline (therefore not related to the time dynamics), and the innovation components are the CO₂ saturation snapshots that can be solely described by the time-varying fluid-flow dynamics. In this context, we could consider to extend the feasibility study in Chapter 3 to simultaneously invert for the permeability and the brine-filled

wavespeed model at the baseline.

8.4.4 Chapter 6

Theoretical analysis on uncertainty quantification

The WISE framework presented in Chapter 6 demonstrates the superior performance of CIGs (extended gradients) in summarizing the observed data compared to RTMs (gradients). It is worthwhile for future research to include a case study on a toy problem to further verify that these CIGs preserve all information in the observed data — i.e., whether the posterior conditioned on CIGs is the same as the posterior conditioned on the observed data. If there is an approximation error, future research can also investigate whether the introduced error is related to the choice of the fiducial point (initial model) used to calculate the CIGs.

Choice of initial models

Even if a theoretical proof in the idealized scenarios, as suggested by Section 8.4.4, could indicate that the CIGs lead to an unbiased posterior, it is still worthwhile to investigate the practical choice of the initial models (fiducial points) where the CIGs (extended gradients) are calculated. This is because there is always a possibility that an amortization gap exists due to the limited amount of training data and computational resources during the training phase. To help choose the initial model, there are a few ways to further interpret the role of the initial model during the CIG calculation and the amortized VI process:

- The initial model can be thought as an experimental configuration — i.e., different initial models produce different CIGs, which might lead to different posterior distributions approximated by the CNFs. In this context, we can optimize for the initial model in order for it to maximally inform the posterior distribution. Thanks to our principled Bayesian inference approach, this results in a Bayesian experimental design problem where the quality of the the experimental configurations (i.e. initial

models) can be measured by the expected information gain [8]. Suggested by [9], minimizing the training objective of CNFs over both the network weights and the initial model can maximize the expected information gain. This reads:

$$\underset{\boldsymbol{\theta}, \mathbf{x}_0}{\text{minimize}} \quad \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{2} \|f_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}; \overline{\nabla \mathcal{F}}(\mathbf{x}_0)^\top \mathbf{y}^{(i)})\|_2^2 - \log |\det \mathbf{J}_{f_{\boldsymbol{\theta}}}| \right). \quad (8.4)$$

The definitions of the symbols in this objective remain the same as in Equation 6.1 in Chapter 6. This objective is minimized over both network parameters, $\boldsymbol{\theta}$, and the initial model, \mathbf{x}_0 . This can be achieved by gradient-based optimization algorithms. There remains a computational challenge that every gradient-based iteration requires evaluating the action of $\overline{\nabla \mathcal{F}}(\mathbf{x}_0)^\top$ and differentiating the objective with respect to the initial model, \mathbf{x}_0 , both of which involves the computationally intensive wave modeling. To alleviate this problem, recently developed frameworks can use FNOs as a surrogate to map a set of CIGs associated with one background model to CIGs associated with another initial model [10]. To take advantage of this neural surrogate, future work can investigate the potential to replace the operator, $\overline{\nabla \mathcal{F}}(\mathbf{x}_0)^\top$, with the neural surrogates to achieve accelerated evaluation and gradient calculation. This surrogate-assisted optimization framework could also benefit by incorporating the algorithm presented in Chapter 4 to safeguard the accuracy of the surrogates during optimization.

- From a totally different perspective, the initial model can also be considered as a nuisance parameter in the framework of simulation-based inference (SBI) [11]. In this context, a simulator, \mathcal{G} , takes the input of velocity model, \mathbf{x} from the prior distribution, initial model, \mathbf{x}_0 , and random realizations of Gaussian noise, $\boldsymbol{\epsilon}$, and eventually outputs the CIGs, $\bar{\mathbf{y}}$, as following

$$\bar{\mathbf{y}} = \mathcal{G}(\mathbf{x}, \mathbf{x}_0, \boldsymbol{\epsilon}) = \overline{\nabla \mathcal{F}}(\mathbf{x}_0)^\top (\mathcal{F}(\mathbf{x}) + \boldsymbol{\epsilon}). \quad (8.5)$$

Equation 8.5 presents a stylish SBI framework as the unknown parameter of interest, \mathbf{x} , and observables, \mathbf{y} , are related through a simulator with two nuisance parameters, \mathbf{x}_0 , and ϵ . In this SBI context [2], a training dataset can be created by not only randomly drawing samples from the prior and noise from the Gaussian distribution, but also randomly drawing initial models, \mathbf{x}_0 , from a distribution that can either be independent or dependent on the velocity model, \mathbf{x} . After generating the training pairs, a conditional density estimator (e.g. CNF) can be trained to draw samples of \mathbf{x} conditioned on $\bar{\mathbf{y}}$.

The benefit of following the SBI framework is that the posterior estimation is robust with respect to the nuisance parameter, \mathbf{x}_0 . This practically means that any initial model, \mathbf{x}_0 , drawn from the same statistical distribution as the one used in training, can be used during inference, while the original WISE framework presented in Chapter 6 is designed to only use a single initial model during both training and inference. It is worthwhile to investigate whether stochastic sampling can improve the accuracy of inference as indicated by [12]. However, this interpretation deviates from the mainstream SBI literature because the nuisance parameter, \mathbf{x}_0 , can be observed (chosen) at the inference phase, while the conventional SBI only assumes to know the distribution of the nuisance parameter. This can be an enticing avenue for future research.

8.4.5 Chapter 7

Although the WISER framework demonstrates a proof of concept that distribution shifts can be mitigated via physics-based refinements, future investigations are necessary to validate the efficacy of WISER on more challenging OOD scenarios, especially using a velocity model with a salt body to generate the observed data as the difficulty of FWI for salt is well recognized in the literature [13]. Also, further OOD studies can be performed where the observed data is simulated with a kernel slightly different from the kernel used in inversion

to represent more realistic inversion scenarios. It is also worthwhile to investigate whether the part of the observed data that is out of the range of the modeling operator leads to higher uncertainty in the velocity model estimation accordingly, or the networks are insensitive to those events.

8.5 References

- [1] C. Jones, J. Edgar, J. Selvage, and H. Crook, “Building complex synthetic models to evaluate acquisition geometries and velocity inversion technologies,” in *74th EAGE Conference and Exhibition incorporating EUROPEC 2012*, European Association of Geoscientists & Engineers, 2012, cp–293 (page 174).
- [2] S. T. Radev, U. K. Mertens, A. Voss, L. Ardizzone, and U. Köthe, “Bayesflow: Learning complex stochastic models with invertible neural networks,” *IEEE transactions on neural networks and learning systems*, vol. 33, no. 4, pp. 1452–1466, 2020 (pages 177, 183).
- [3] R. Orozco, A. Siahkoohi, G. Rizzuti, T. van Leeuwen, and F. Herrmann, “Adjoint operators enable fast and amortized machine learning based bayesian uncertainty quantification,” *Medical Imaging 2023: Image Processing*, I. Išgum and O. Colliot, Eds., Apr. 3, 2023 (page 177).
- [4] J. Hou and W. W. Symes, “Accelerating extended least-squares migration with weighted conjugate gradient iteration,” *GEOPHYSICS*, vol. 81, no. 4, S165–S179, Jul. 2016 (page 177).
- [5] M. Asim, M. Daniels, O. Leong, A. Ahmed, and P. Hand, “Invertible generative models for inverse problems: Mitigating representation error and dataset bias,” in *International conference on machine learning*, PMLR, 2020, pp. 399–409 (page 179).
- [6] D. Li, H. Denli, C. MacDonald, K. Basler-Reeder, A. Baumstein, and J. Daves, “Multiparameter geophysical reservoir characterization augmented by generative networks,” in *First International Meeting for Applied Geoscience & Energy*, Society of Exploration Geophysicists, 2021, pp. 1364–1368 (page 179).
- [7] T. O’Leary-Roseberry, P. Chen, U. Villa, and O. Ghattas, “Derivative-informed neural operator: An efficient framework for high-dimensional parametric derivative learning,” *Journal of Computational Physics*, vol. 496, p. 112 555, 2024 (page 180).
- [8] K. Chaloner and I. Verdinelli, “Bayesian experimental design: A review,” *Statistical science*, pp. 273–304, 1995 (page 182).
- [9] R. Orozco, F. J. Herrmann, and P. Chen, “Probabilistic bayesian optimal experimental design using conditional normalizing flows,” *arXiv preprint arXiv:2402.18337*, 2024 (page 182).

- [10] R. Rex, Z. Yin, and F. J. Herrmann, “Velocity continuation for common image gathers with fourier neural operators,” in *International Meeting for Applied Geoscience & Energy*, May 2024 (page 182).
- [11] K. Cranmer, J. Brehmer, and G. Louppe, “The frontier of simulation-based inference,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 055–30 062, 2020 (page 182).
- [12] Y. Zeng, R. Orozco, Z. Yin, and F. J. Herrmann, “Enhancing full-waveform variational inference through stochastic resampling and data augmentation,” in *International Meeting for Applied Geoscience & Energy*, Mar. 2024 (page 183).
- [13] P. Wang, Z. Zhang, J. Mei, F. Lin, and R. Huang, “Full-waveform inversion for salt: A coming of age,” *The Leading Edge*, vol. 38, no. 3, pp. 204–213, 2019 (page 183).

Appendices

APPENDIX A

MORE DETAILS OF FWI: MODELING, GRADIENT, AND EXTENDED GRADIENT

A.1 Wave modeling

The forward modeling operator in FWI, denoted by \mathcal{F} , maps the velocity model, \mathbf{x} , to the seismic data collected at the receiver locations, collected in \mathbf{y} . In its simplest form, \mathcal{F} can be evaluated by solving the wave equation and restricting the solution wavefield to the receiver positions and recording time windows [1], as follows:

$$\mathcal{F}(\mathbf{x}) = \begin{bmatrix} \mathbf{P}\mathbf{A}(\mathbf{x})^{-1}\mathbf{q}_1 \\ \mathbf{P}\mathbf{A}(\mathbf{x})^{-1}\mathbf{q}_2 \\ \dots \\ \mathbf{P}\mathbf{A}(\mathbf{x})^{-1}\mathbf{q}_{n_s} \end{bmatrix}. \quad (\text{A.1})$$

In this equation, we slightly abuse the notation to let \mathbf{x} be the squared slowness given by the reciprocal of the squared velocity. $\{\mathbf{q}_i\}_{i=1}^{n_s}$ represents the sources, with n_s being the number of sources. The matrix, $\mathbf{A}(\mathbf{x})$, represents the wave equation with the coefficient, \mathbf{x} . The wave equation can be written as:

$$\mathbf{A}(\mathbf{x})\mathbf{u} = \left(\text{diag}(\mathbf{x})\frac{\partial^2}{\partial t^2} - \nabla^2 \right) \mathbf{u} = \mathbf{q}_i. \quad (\text{A.2})$$

In this equation, $\text{diag}(\mathbf{x})$ represents a diagonal matrix with \mathbf{x} on its diagonal. $\frac{\partial^2}{\partial t^2}$ represents second-order time derivative operator, and ∇^2 represents Laplacian operator. The wave operator relates its solution (also called wavefield), \mathbf{u} , to each source, \mathbf{q}_i . The inverse of $\mathbf{A}(\mathbf{x})$ solves the wave equation for each source, and the matrix, \mathbf{P} , is a restriction operator that measures the wavefield at the receiver locations. In practice, this wave oper-

ator, $\mathbf{A}(\mathbf{x})$, is never explicitly formed, nor its inverse mathematically derived. It is instead evaluated using finite-difference time-stepping propagation [2, 3].

A.2 Gradient

Algorithms for solving FWI typically use the gradient of the objective function with respect to the model parameter as an update direction. Here, we choose the ℓ_2 misfit as the objective function for simplicity:

$$\phi(\mathbf{x}) = \frac{1}{2} \|\mathcal{F}(\mathbf{x}) - \mathbf{y}\|_2^2. \quad (\text{A.3})$$

Then, the gradient is given by:

$$\mathbf{g} = \frac{\partial \phi}{\partial \mathbf{x}} = \nabla \mathcal{F}(\mathbf{x})^\top (\mathcal{F}(\mathbf{x}) - \mathbf{y}), \quad (\text{A.4})$$

where $\nabla \mathcal{F}(\mathbf{x})$ represents the Jacobian matrix, and \top denotes the adjoint. This is also referred as the reverse-time migration (RTM, [4]) in the geophysical literature.

A.3 Extended gradient

While there exist many approaches to extend the wave modeling and its gradient calculations [5, 6], my thesis focuses on an extension to the wave equation in Equation A.2 where the instantaneous action in the wave equation is relaxed, allowing for the action of the wavefield over a “distance”, also known as the subsurface offset [7, 8, 9]. In this context, the wave equation is extended to honor this action through a non-diagonal matrix before the second-order time-derivative term in Equation A.2. The extended wave equation reads:

$$\overline{\mathbf{A}}(\mathbf{X})\mathbf{u} = \left(\mathbf{X} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) \mathbf{u} = \mathbf{q}_i, \quad (\text{A.5})$$

where $\mathbf{X} = \text{diag}(\mathbf{x})$ reduces it to the standard scalar wave equation in Equation A.2.

Following Equation A.5 to parameterize the wave equation and solve the FWI problem, the objective function in Equation A.3 is modified to:

$$\bar{\phi}(\mathbf{X}) = \frac{1}{2} \|\bar{\mathcal{F}}(\mathbf{X}) - \mathbf{y}\|_2^2 \quad \text{where} \quad \bar{\mathcal{F}}(\mathbf{X}) = \mathbf{P}\bar{\mathbf{A}}(\mathbf{X})^{-1}\mathbf{q}_i. \quad (\text{A.6})$$

The extended gradient, also known as the common-image gathers (CIGs, [10]) in the geophysical literature, is the gradient of the objective function in Equation A.6 with respect to the extended model parameter when the objective is linearized at $\text{diag}(\mathbf{x})$:

$$\bar{\mathbf{g}} = \left. \frac{\partial \bar{\phi}}{\partial \mathbf{X}} \right|_{\mathbf{X}=\text{diag}(\mathbf{x})} = \overline{\nabla \mathcal{F}}(\text{diag}(\mathbf{x}))^\top (\bar{\mathcal{F}}(\text{diag}(\mathbf{x})) - \mathbf{y}) = \overline{\nabla \mathcal{F}}(\text{diag}(\mathbf{x}))^\top (\mathcal{F}(\mathbf{x}) - \mathbf{y}). \quad (\text{A.7})$$

For simplicity, $\overline{\nabla \mathcal{F}}(\text{diag}(\mathbf{x}))$ is written as $\overline{\nabla \mathcal{F}}(\mathbf{x})$ in Chapter 6.

APPENDIX B

MORE RESULTS FOR CHAPTER 6

B.1 Summary

This section is structured as follows: First, we provide a more detailed analysis of the case studies on the Compass model. Next, we introduce an additional case study on the CurveFault-A dataset from Open FWI dataset [11].

B.2 Compass

In the Compass case study, WISE showcases its ability to generate velocity models consistent with the observed shot data. Here, we delve into a detailed analysis of these results in this section to affirm its superior performance. To benchmark WISE against traditional FWI methods, we initiate by conducting FWI on the 1D initial model depicted in Figure 1(b) in the main text. Subsequently, we examine several posterior samples from WISE, focusing on the analysis of CIG focusing.

B.2.1 Full-waveform inversion from the 1D initial model

We conducted an FWI on the 1D initial model depicted in Figure 1(b) in the main text, using 200 iterations of gradient descent. In each iteration, we selected 4 random OBNs with replacement to determine the update direction. The entire inversion amounts to 12.5 datapasses. To address cycle-skipping, we employed a frequency continuation strategy, progressively inverting data from low to high frequencies [12]. The FWI result, presented in Figure B.1(a), reveals mispositioning of several major reflectors and poor recovery of layers beneath the unconformity. This issue becomes more pronounced in the vertical profiles shown in Figure B.2. The cycle-skipping problem led to the FWI's failure in accurately es-

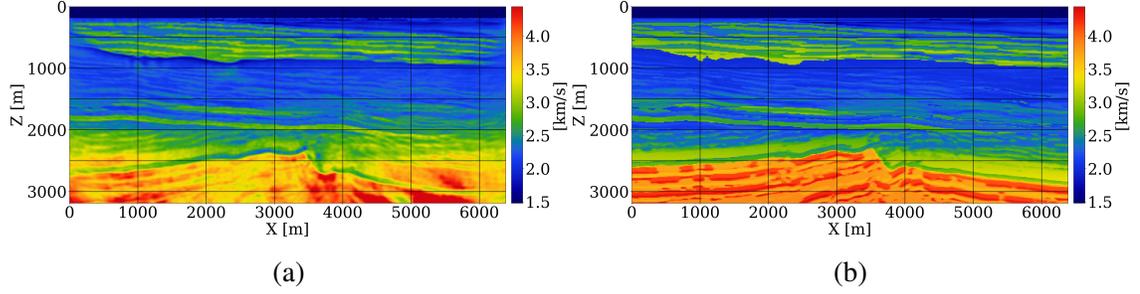


Figure B.1: (a) FWI result using Figure 1(b) in the main text as a starting model; (b) ground truth velocity model.

timating the bottom of the velocity “kick-back” layer at approximately 800 m depth, resulting in mispositioned layers, including the unconformity at about 2200 m depth. Conversely, the conditional mean estimate from WISE closely matches the ground truth velocity trend, and the 95% confidence interval successfully encompasses the ground truth velocity model at nearly all locations.

B.2.2 Common-image gathers of posterior samples

To evaluate the focusing of CIGs for all posterior samples, we calculated the percentage of energy within near offsets (specifically, between -60 , m and 60 , m) as

$$\frac{\|\text{energy in near offsets}\|_2}{\|\text{energy in all offsets}\|_2}.$$

The CIGs derived from the initial 1D velocity model (depicted in Figure (1)b of the main text) contained only 73.6% of their energy in near offsets, whereas the conditional mean estimate from WISE (illustrated in Figure (1)d) accounted for 81.6% of the energy. This quantitatively confirms the enhanced focusing of the conditional mean estimate. Moreover, we computed the focused energy percentage in near offsets for CIGs produced by all posterior samples, resulting in an average of 74.3% — an improvement over the initial 1D velocity model. The standard deviation among these measurements was 0.005%. Three example CIGs are presented in Figure B.3, which, while not as sharply focused as the conditional mean estimate CIGs (shown in Figure 1(f) in the main text), exhibit visibly

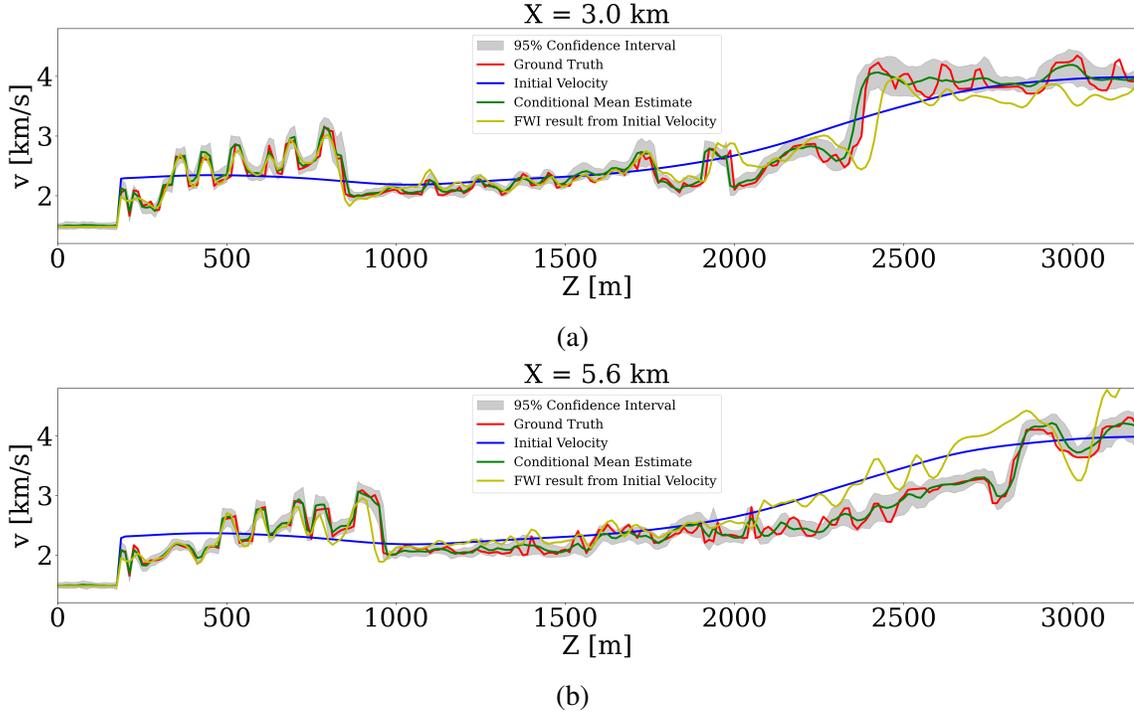
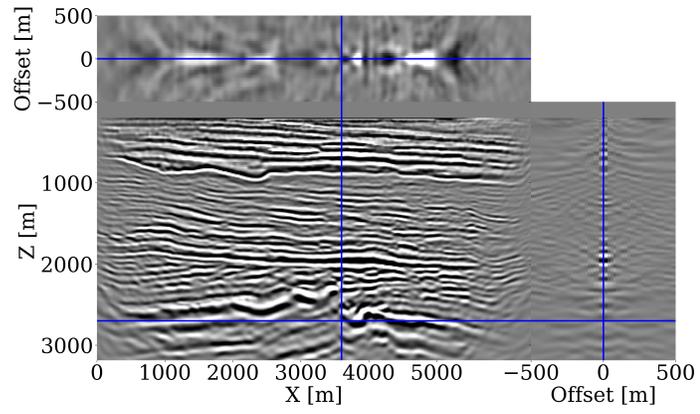


Figure B.2: Vertical profiles.

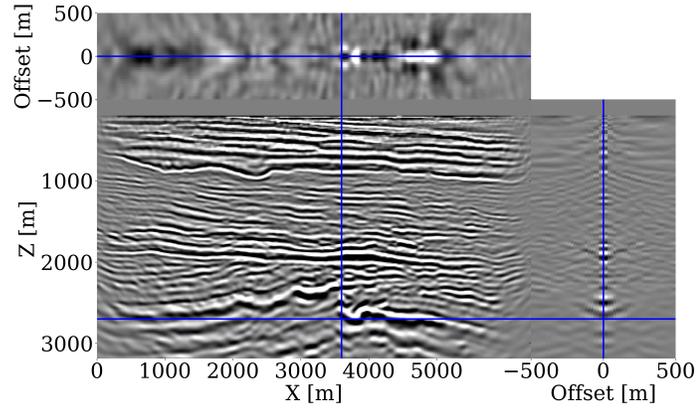
better focusing than those generated by the 1D velocity model (shown in Figure 1(e) in the main text). For further analysis, practitioners might choose posterior samples based on this metric or other criteria, such as the curvature of gathers determined by migration velocity analysis. We propose to explore these alternative selection methodologies in future research.

B.3 Open FWI

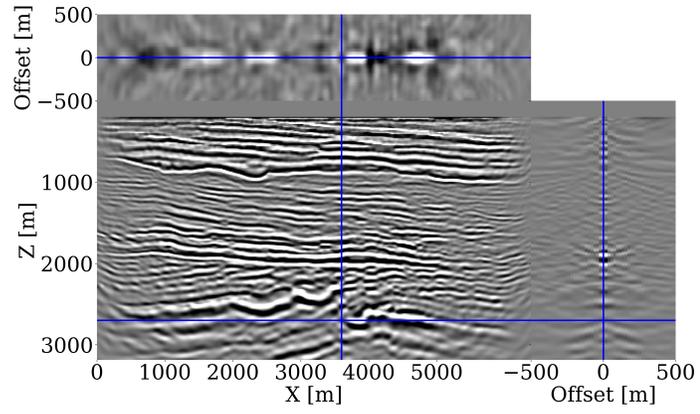
We present a case study using the CurveFault-A models in the Open FWI dataset, which is a public deep learning benchmark dataset for FWI. The CurveFault-A dataset comprises velocity models with significant variability across samples, which poses challenges for deep learning methods. This is further compounded by faults and dipping events while observations contain only reflected energy. Testing on this dataset allows us to test WISE’s velocity-model generation capabilities.



(a)



(b)



(c)

Figure B.3: CIGs for three posterior samples from WISE.

B.3.1 Dataset generation and network training

We select 2800 velocity models of 640 m by 640 m, each with 64 equally spaced receivers at 10m tow depth and 16 randomly placed sources. The surface is assumed absorbing. Using a 15Hz central frequency Ricker wavelet with energy below 3Hz removed for realism, acoustic data is simulated with Devito [2, 3] and JUDI.jl [13]. Uncorrelated band-limited Gaussian noise is added (S/N 12dB) before migrating each dataset with a 1D initial FWI-velocity model calculated by averaging the corresponding true model horizontally. CIGs are computed for 101 horizontal subsurface offsets ranging from -250m to $+250\text{m}$. Two CNFs are trained: one with velocity-RTM pairs and another with velocity-CIGs pairs.

B.3.2 Results

Results on two tested samples by our CNFs are included in Figure B.4 and reveal notable variation in the posterior samples for sharp boundaries and smooth transitions in the velocity. While the conditional mean estimate does not fully replicate the true velocity, the standard deviations meaningfully correlate with the errors, indicating that the uncertainty represented by the standard deviation is informative. Across 50 test samples, the mean SSIM score for CIGs-based statistics is 0.87, surpassing the 0.85 mean for RTM-based statistics. We also observe that CIGs better inform the posterior compared to RTM in both test samples.

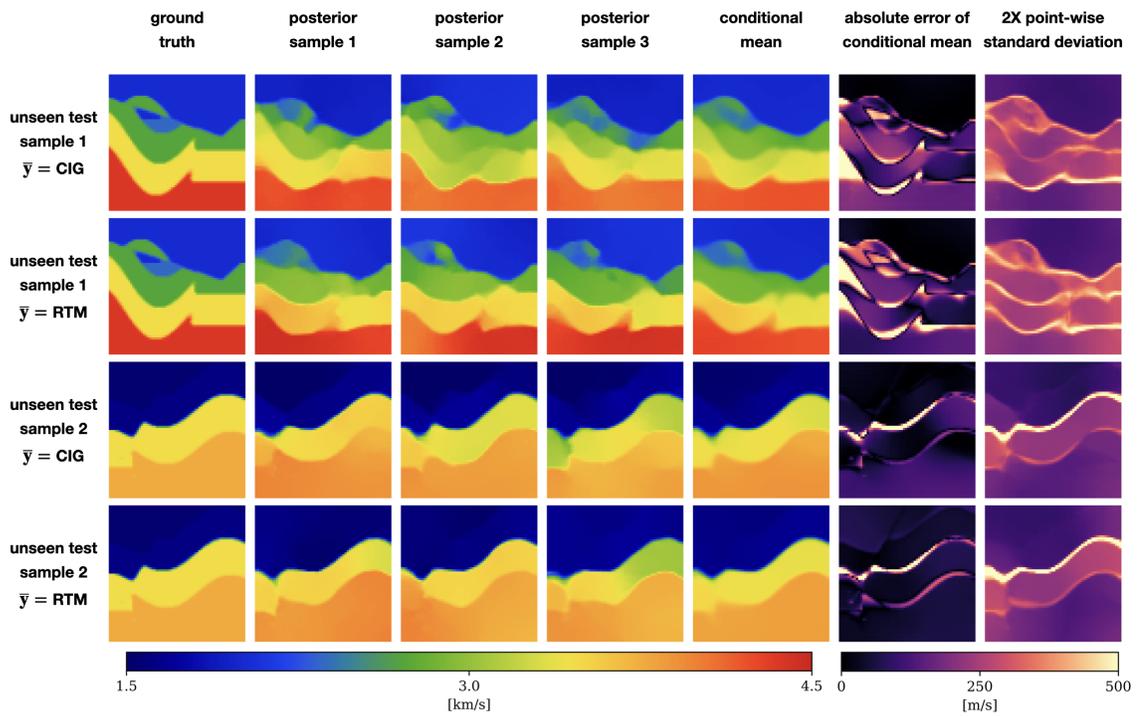


Figure B.4: Applying WISE for two unseen test samples in Open FWI CurveFault-A dataset. \bar{y} denotes the type of summary statistics.

APPENDIX C
PERMISSIONS TO USE COPYRIGHTED MATERIAL

C.1 Chapter 2

The content of Chapter 2 was published as a technical article in *The Leading Edge* under the title “Learned multiphysics inversion with differentiable programming and machine learning” [14]:

- Louboutin, Mathias, Ziyi Yin, Rafael Orozco, Thomas J. Grady, Ali Siahkoochi, Gabrio Rizzuti, Philipp A. Witte, Olav Møyner, Gerard J. Gorman, and Felix J. Herrmann. “Learned multiphysics inversion with differentiable programming and machine learning.” *The Leading Edge* 42, no. 7 (2023): 474-486.
- DOI: 10.1190/tle42070474.1.
- The article is published open-access under Creative Commons Attribution license (CC BY).
- The authors retain copyright of the work, as stated in SEG Policy on Open-Access Publishing at <https://library.seg.org/page/policies/open-access>.

C.2 Chapter 3

The content of Chapter 3 was accepted for publication as a technical article in *The Leading Edge* in May 2024 under the title “Time-lapse full-waveform permeability inversion: a feasibility study” [15]:

- Yin, Ziyi, Mathias Louboutin, Olav Møyner, and Felix J. Herrmann. “Time-lapse full-waveform permeability inversion: a feasibility study.” Accepted for publication in *The Leading Edge* (May 2024).

- A preprint of this article is available as a preprint on ArXiv under Creative Commons Attribution license (CC BY) with DOI: 10.48550/arXiv.2403.04083.
- Copyright © 2024 Society of Exploration Geophysicists.
- The author retains the right to reuse all of part of the work in a thesis or dissertation, as stated in the Copyright Agreement.

The Leading Edge Transfer of Copyright Agreement

Agreement must be signed by lead or corresponding author and returned to the SEG Business Office before manuscript receives final acceptance.

Article title: Time-lapse full-waveform permeability inversion: a feasibility study

Names of all authors: Ziyi Yin, Mathias Louboutin, Olav Møyner, Felix J. Herrmann

For and in consideration of the potential publication of the article listed above (the “Work”) by the Society of Exploration Geophysicists (“SEG”) in *The Leading Edge (TLE)*, I, the undersigned, as lead and/or corresponding author and/or rightsholders and acting on behalf of all authors and/or owners of copyright and other rights in the Work (“Authors”), hereby transfer, assign, and convey all right, title, interest, and worldwide copyright in the Work to SEG, effective if and when the Work is accepted for publication, subject only to limitations expressed in this Transfer of Copyright (the “Agreement”). I warrant that I am authorized and empowered to represent all Authors with respect to the Agreement, which is executed jointly and severally by Authors if copyright transfer is from multiple individuals or entities. The Work includes the article and all material to be published within and with the article in any and all media, including but not limited to tables, figures, graphs, source code, movies, and other multimedia. Responses to discussions of the Work, errata, and other similar material directly related to the Work that may arise subsequent to publication also are considered part of the Work. In the event the Work incorporates copyrighted material of others, Authors warrant that all required permissions or releases have been secured. Authors agree to indemnify, defend, and hold SEG, its directors, officers, employees, and agents harmless against any claims to the contrary.

Authors shall retain the following royalty-free rights:

1. All proprietary rights in the Work that are not transferred to SEG in the Agreement, including the right to any patentable subject matter that may be contained in the Work
2. The right to reproduce and distribute part or all of the Work, including figures, drawings, tables, and abstracts of the Work, with proper attribution and copyright acknowledgment, in connection with Authors’ teaching and technical collaborations
3. The right to make oral presentation of the same or similar information as that contained in the Work provided acknowledgment is made of SEG copyright ownership and publication status
4. The right to post a final accepted version of the manuscript or the final SEG-formatted version on Authors’ personal Web sites (not including social-network sites used for article sharing such as ResearchGate), employers’ Web sites, or in institutional repositories operated and controlled exclusively by Authors’ employers provided that (a) the TLE-accepted or TLE-published version is presented without modification unless modification is noted and fully described; (b) SEG copyright notice and a full citation appear with the paper; (c) a live link to the version of record in the SEG Library using a Digital Object Identifier (DOI) permalink is provided; (d) the posting is noncommercial in nature, and the paper is made available to users without charge; (e) notice is provided that use is subject to SEG terms of use and conditions; and (f) a posting of the Work in an institutional repository neither carries nor is implied as carrying a license in conflict with SEG copyright and terms of use
5. The right to prepare and hold copyright in derivative publications based on the Work provided that the derivative work is published subsequent to the official date of the Work’s publication by SEG and the Work is cited
6. The right to reuse all or part of the Work in a thesis or dissertation that the Author writes and is required to submit to satisfy criteria of degree-granting institutions, with full citation of the Work, its copyright status, and a DOI permalink to the version of record
7. The nonexclusive right, after publication by SEG, to republish and distribute print versions of the Work or excerpts therefrom without obtaining permission from SEG, provided that (a) the paper is not republished in a journal, book, or collection of conference abstracts or proceedings; and (b) no fee is charged for the printed versions. Permission must be obtained from SEG for other republication of the Work.

SEG may republish the Work or portions thereof in any future SEG publication or compilation in any form and in any language, and SEG retains exclusive right to license third parties to do the same.

This Agreement entitles Authors (or in the case of a Work Made for Hire, the employer) to retain all rights not transferred, assigned, or conveyed to SEG in the Agreement. Authors confirm that the Work has not been published previously elsewhere, nor is it under consideration by any other publisher, that the Work does not infringe any copyright or invade any right of privacy or publicity, and that Authors have the full power to enter into this Agreement and to make the grants contained herein. Authors warrant that they have complied and will comply with Ethical Guidelines for SEG Publications and that they agree to the provisions of these guidelines and all other SEG publications policies applicable to authors. This Agreement shall be binding on Authors’ heirs, executors, administrators, and assigns, and shall be construed in accordance with the laws of the State of Oklahoma, United States of America.

IN WITNESS WHEREOF, I have executed this Transfer of Copyright on this 29 day of May, 2024.

Ziyi Yin

Name of author (print or type)

 
Digitally signed by Ziyi Yin
Date: 2024.05.29 11:09:21 -07'00'

Signature

Company for which work was performed (if applicable)

Authorized by/Title

FOR AUTHORS EMPLOYED BY U.S. GOVERNMENT

I certify that the article named above was prepared by a U.S. government employee(s) as part of his/her (their) official duties and therefore legally cannot be copyrighted. Authors agree to all other terms of this agreement.

Name (print or type)

Date

Signature

C.3 Chapter 4

The content of Chapter 4 was published as a technical article in *Advanced Modeling and Simulation in Engineering Sciences* under the title “Solving multiphysics-based inverse problems with learned surrogates and constraints” [16]:

- Yin, Ziyi, Rafael Orozco, Mathias Louboutin, and Felix J. Herrmann. “Solving multiphysics-based inverse problems with learned surrogates and constraints.” *Advanced Modeling and Simulation in Engineering Sciences* 10, no. 1 (2023): 14.
- DOI: 10.1186/s40323-023-00252-0.
- The article is published open-access under Creative Commons Attribution license (CC BY).
- The authors retain copyright of the work, as stated in the Copyright Agreement.

Licensee:	Computational Structural Mechanics Association	(the 'Licensee')
Journal Name:	Advanced Modeling and Simulation in Engineering Sciences	(the 'Journal')
Manuscript Number:	AMOS-D-23-00027R2	
Proposed Title of Article:	Solving multiphysics-based inverse problems with learned surrogates and constraints	(the 'Article')
Author(s) [Please list all named Authors]:	Ziyi Yin, Rafael Orozco, Mathias Louboutin, Felix J. Herrmann	(the 'Author')
Corresponding Author Name:	Ziyi Yin	

Licence Applicable to the Article:

Creative Commons licence CC BY: This licence allows readers to copy, distribute and transmit the Article as long as it is attributed back to the author. Readers are permitted to alter, transform or build upon the Article, and to use the Article for commercial purposes. Please read the full licence for further details at - <http://creativecommons.org/licenses/by/4.0/>

Subject to editorial acceptance of the Article, it will be published under the Creative Commons licence shown above.

1 Grant of Rights

- a) For good and valuable consideration, the Author hereby grants to the Licensee the perpetual, non-exclusive, irrevocable, world-wide, assignable, sublicensable and unlimited right to: publish, reproduce, copy, distribute, communicate, display publicly, sell, rent and/ or otherwise make available the article identified above, including any supplementary information and graphic elements therein (e.g. illustrations, charts, moving images) (the "Article") in any language, in any versions or editions in any and all forms and/or media of expression (including without limitation in connection with any and all end-user devices), whether now known or developed in the future. Without limitation, the above grant includes: (i) the right to edit, alter, adapt, adjust and prepare derivative works; (ii) all commercial use, advertising, and marketing rights, including without limitation graphic elements on the cover of the journal and in relation to social media; (iii) rights for any training, educational and/or instructional purposes; (iv) the right to add and/or remove links or combinations with other media/works; and (v) the right to create, use and/or license and/or sublicense content data or metadata of any kind in relation to the Article (including abstracts and summaries) without restriction. The above rights are granted in relation to the Article as a whole or any part and with or in relation to any other works.
- b) Without limiting the rights granted above, Licensee is granted the rights to use the Article for the purposes of analysis, testing, and development of publishing- and research-related workflows, systems, products, projects, and services; to confidentially share the Article with select third parties to do the same; and to retain and store the Article and any associated correspondence/files/forms to maintain the historical record, and to facilitate research integrity investigations. The grant of rights set forth in this clause (b) is irrevocable.
- c) The Licensee will have the right, but not the obligation, to exercise any or all of the rights granted herein. If the Licensee elects not to publish the Article for any reason, all publishing rights under this Agreement as set forth in clause 1.a) above will revert to the Author.

2 Copyright

Ownership of copyright in the Article will be vested in the name of the Author. When reproducing the Article or extracts from it, the Author will acknowledge and reference first publication in the Journal.

3 Use of Article Versions

- a) For purposes of this Agreement: (i) references to the "Article" include all versions of the Article; (ii) "Submitted Manuscript" means the version of the Article as first submitted by the Author; (iii) "Accepted Manuscript" means the version of the Article accepted for publication, but prior to copy-editing and typesetting; and (iv) "Version of Record" means the version of the Article published by the Licensee, after copy-editing and typesetting. Rights to all versions of the Manuscript are granted on a non-exclusive basis.

C.4 Chapter 5

The content of Chapter 5 was published as a technical article in *The Leading Edge* under the title “Derisking geologic carbon storage from high-resolution time-lapse seismic to explainable leakage detection” [17]:

- Yin, Ziyi, Huseyin Tuna Erdinc, Abhinav Prakash Gahlot, Mathias Louboutin, and Felix J. Herrmann. “Derisking geologic carbon storage from high-resolution time-lapse seismic to explainable leakage detection.” *The Leading Edge* 42, no. 1 (2023): 69-76.
- DOI: 10.1190/tle42010069.1.
- Copyright © 2023 Society of Exploration Geophysicists.
- The author retains the right to reuse all of part of the work in a thesis or dissertation, as stated in the Copyright Agreement.

The Leading Edge Transfer of Copyright Agreement

Agreement must be signed by lead or corresponding author and returned to the SEG Business Office before manuscript receives final acceptance.

Article title: De-risking geological carbon storage from high resolution time-lapse seismic to ex

Names of all authors: Ziyi Yin, Husevin Tuna Erdinc, Abhinav Prakash Gahlot, Mathias Louboutin,

For and in consideration of the potential publication of the article listed above (the “Work”) by the Society of Exploration Geophysicists (“SEG”) in *The Leading Edge (TLE)*, I, the undersigned, as lead and/or corresponding author and/or rightsholders and acting on behalf of all authors and/or owners of copyright and other rights in the Work (“Authors”), hereby transfer, assign, and convey all right, title, interest, and worldwide copyright in the Work to SEG, effective if and when the Work is accepted for publication, subject only to limitations expressed in this Transfer of Copyright (the “Agreement”). I warrant that I am authorized and empowered to represent all Authors with respect to the Agreement, which is executed jointly and severally by Authors if copyright transfer is from multiple individuals or entities. The Work includes the article and all material to be published within and with the article in any and all media, including but not limited to tables, figures, graphs, source code, movies, and other multimedia. Responses to discussions of the Work, errata, and other similar material directly related to the Work that may arise subsequent to publication also are considered part of the Work. In the event the Work incorporates copyrighted material of others, Authors warrant that all required permissions or releases have been secured. Authors agree to indemnify, defend, and hold SEG, its directors, officers, employees, and agents harmless against any claims to the contrary.

Authors shall retain the following royalty-free rights:

1. All proprietary rights in the Work that are not transferred to SEG in the Agreement, including the right to any patentable subject matter that may be contained in the Work
2. The right to reproduce and distribute part or all of the Work, including figures, drawings, tables, and abstracts of the Work, with proper attribution and copyright acknowledgment, in connection with Authors’ teaching and technical collaborations
3. The right to make oral presentation of the same or similar information as that contained in the Work provided acknowledgment is made of SEG copyright ownership and publication status
4. The right to post a final accepted version of the manuscript or the final SEG-formatted version on Authors’ personal Web sites (not including social-network sites used for article sharing such as ResearchGate), employers’ Web sites, or in institutional repositories operated and controlled exclusively by Authors’ employers provided that (a) the *TLE*-accepted or *TLE*-published version is presented without modification unless modification is noted and fully described; (b) SEG copyright notice and a full citation appear with the paper; (c) a live link to the version of record in the SEG Library using a Digital Object Identifier (DOI) permalink is provided; (d) the posting is noncommercial in nature, and the paper is made available to users without charge; (e) notice is provided that use is subject to SEG terms of use and conditions; and (f) a posting of the Work in an institutional repository neither carries nor is implied as carrying a license in conflict with SEG copyright and terms of use
5. The right to prepare and hold copyright in derivative publications based on the Work provided that the derivative work is published subsequent to the official date of the Work’s publication by SEG and the Work is cited
6. The right to reuse all or part of the Work in a thesis or dissertation that the Author writes and is required to submit to satisfy criteria of degree-granting institutions, with full citation of the Work, its copyright status, and a DOI permalink to the version of record
7. The nonexclusive right, after publication by SEG, to republish and distribute print versions of the Work or excerpts therefrom without obtaining permission from SEG, provided that (a) the paper is not republished in a journal, book, or collection of conference abstracts or proceedings; and (b) no fee is charged for the printed versions. Permission must be obtained from SEG for other republication of the Work.

SEG may republish the Work or portions thereof in any future SEG publication or compilation in any form and in any language, and SEG retains exclusive right to license third parties to do the same.

This Agreement entitles Authors (or in the case of a Work Made for Hire, the employer) to retain all rights not transferred, assigned, or conveyed to SEG in the Agreement. Authors confirm that the Work has not been published previously elsewhere, nor is it under consideration by any other publisher, that the Work does not infringe any copyright or invade any right of privacy or publicity, and that Authors have the full power to enter into this Agreement and to make the grants contained herein. Authors warrant that they have complied and will comply with Ethical Guidelines for SEG Publications and that they agree to the provisions of these guidelines and all other SEG publications policies applicable to authors. This Agreement shall be binding on Authors’ heirs, executors, administrators, and assigns, and shall be construed in accordance with the laws of the State of Oklahoma, United States of America.

IN WITNESS WHEREOF, I have executed this Transfer of Copyright on this 25 day of October, 2022.

Ziyi Yin

Name of author (print or type)

尹梓屹 Ziyi Yin

Signature

Company for which work was performed (if applicable)

Authorized by/Title

**SIGN HERE IF U.S. GOVERNMENT EMPLOYED ALL
AUTHORS WHEN WORK WAS PREPARED.**

I certify that the article named above was prepared solely by a U.S. government employee(s) as part of his/her (their) official duties and therefore legally cannot be copyrighted. Authors agree to all other terms of this agreement.

Name (print or type)

Date

Signature

C.5 Chapter 6

The content of Chapter 6 was published as a technical article in *Geophysics* under the title “WISE: full-Waveform variational Inference via Subsurface Extensions” [18]:

- Yin, Ziyi, Rafael Orozco, Mathias Louboutin, and Felix J. Herrmann. “WISE: full-Waveform variational Inference via Subsurface Extensions.” *Geophysics* (2024).
- DOI: 10.1190/geo2023-0744.1.
- Copyright © 2024 Society of Exploration Geophysicists.
- The author retains the right to reuse all of part of the work in a thesis or dissertation, as stated in the Copyright Agreement (signed in April 2024).

GEOPHYSICS Transfer of Copyright Agreement

Agreement must be signed by lead or corresponding author and uploaded to ScholarOne Manuscripts before manuscript receives final acceptance.

Article title: **WISE: full-Waveform variational Inference via Subsurface Extensions**

Names of all authors: Ziyi Yin, Rafael Orozco, Mathias Louboutin, Felix J. Herrmann

For and in consideration of the potential publication of the article listed above (the “Work”) by the Society of Exploration Geophysicists (“SEG”), I, the undersigned, as lead and/or corresponding author and/or rightsholders and acting on behalf of all authors and/or owners of copyright and other rights in the Work (“Authors”), hereby transfer, assign, and convey all right, title, interest, and worldwide copyright in the Work to SEG, effective if and when the Work is accepted for publication, subject only to limitations expressed in this Transfer of Copyright (the “Agreement”). I warrant that I am authorized and empowered to represent all Authors with respect to the Agreement, which is executed jointly and severally by Authors if copyright transfer is from multiple individuals or entities. The Work includes the article and all material to be published within and with the article in any and all media, including but not limited to tables, figures, graphs, source code, movies, and other multimedia. Responses to discussions of the Work, errata, and other similar material directly related to the Work that may arise subsequent to publication also are considered part of the Work. In the event the Work incorporates copyrighted material of others, Authors warrant that all required permissions or releases have been secured. Authors agree to indemnify, defend, and hold SEG, its directors, officers, employees, and agents harmless against any claims to the contrary.

Authors shall retain the following royalty-free rights:

1. All proprietary rights in the Work that are not transferred to SEG in the Agreement, including the right to any patentable subject matter that may be contained in the Work
2. The right to reproduce and distribute part or all of the Work, including figures, drawings, tables, and abstracts of the Work, with proper attribution and copyright acknowledgment, in connection with Authors’ teaching and technical collaborations
3. The right to make oral presentation of the same or similar information as that contained in the Work provided acknowledgment is made of SEG copyright ownership and publication status
4. The right to post a final accepted version of the manuscript or the final SEG-formatted version on Authors’ personal Web sites (not including social-network sites used for article sharing such as ResearchGate), employers’ Web sites, or in institutional repositories operated and controlled exclusively by Authors’ employers provided that (a) the GEOPHYSICS-accepted or GEOPHYSICS-published version is presented without modification unless modification is noted and fully described; (b) SEG copyright notice and a full citation appear with the paper; (c) a live link to the version of record in the SEG Library using a Digital Object Identifier (DOI) permalink is provided; (d) the posting is noncommercial in nature, and the paper is made available to users without charge; (e) notice is provided that use is subject to SEG terms of use and conditions; and (f) a posting of the Work in an institutional repository neither carries nor is implied as carrying a license in conflict with SEG copyright and terms of use
5. The right to prepare and hold copyright in derivative publications based on the Work provided that the derivative work is published subsequent to the official date of the Work’s publication by SEG and the Work is cited
6. The right to reuse all or part of the Work in a thesis or dissertation that the Author writes and is required to submit to satisfy criteria of degree-granting institutions, with full citation of the Work, its copyright status, and a DOI permalink to the version of record
7. The nonexclusive right, after publication by GEOPHYSICS, to republish and distribute print versions of the Work or excerpts therefrom without obtaining permission from SEG, provided that (a) the paper is not republished in a journal, book, or collection of conference abstracts or proceedings; and (b) no fee is charged for the printed versions. Permission must be obtained from SEG for other republication of the Work.

SEG may republish the Work or portions thereof in any future SEG publication or compilation in any form and in any language, and SEG retains exclusive right to license third parties to do the same.

This Agreement entitles Authors (or in the case of a Work Made for Hire, the employer) to retain all rights not transferred, assigned, or conveyed to SEG in the Agreement. Authors confirm that the Work has not been published previously elsewhere, nor is it under consideration by any other publisher, that the Work does not infringe any copyright or invade any right of privacy or publicity, and that Authors have the full power to enter into this Agreement and to make the grants contained herein. Authors warrant that they have complied and will comply with Ethical Guidelines for SEG Publications and that they agree to the provisions of these guidelines and all other SEG publications policies applicable to authors. This Agreement shall be binding on Authors’ heirs, executors, administrators, and assigns, and shall be construed in accordance with the laws of the State of Oklahoma, United States of America.

IN WITNESS WHEREOF, I have executed this Transfer of Copyright on this 2ⁿ day of April, 202.

Ziyi Yin

Name of author (print or type)

Ziyi Yin

Digitally signed by Ziyi Yin
Date: 2024.04.02 21:43:41 -04'00'

Signature

Georgia Institute of Technology

Company for which work was performed (if applicable)

Felix J. Herrmann

Authorized by/Title

FOR AUTHORS EMPLOYED BY U.S. GOVERNMENT

I certify that the article named above was prepared by a U.S. government employee(s) as part of his/her (their) official duties and therefore legally cannot be copyrighted. Authors agree to all other terms of this agreement.

Name (print or type)

Date

Signature

C.6 Chapter 7

The content of Chapter 7 is available as a technical article on *ArXiv* in May 2024 under the title “WISER: multimodal variational inference for full-waveform inversion without dimensionality reduction” [19]:

- Yin, Ziyi, Rafael Orozco, and Felix J. Herrmann. “WISER: multimodal variational inference for full-waveform inversion without dimensionality reduction.” Preprint on *ArXiv* (May 2024).
- Currently, the paper is open-access under Creative Commons Attribution license (CC BY) with DOI: 10.48550/arXiv.2405.10327.

C.7 References

- [1] M. Louboutin *et al.*, “Full-waveform inversion - part 1: Forward modeling,” *The Leading Edge*, vol. 36, no. 12, pp. 1033–1036, Dec. 2017, (The Leading Edge) (page 188).
- [2] M. Louboutin *et al.*, “Devito (v3.1.0): An embedded domain-specific language for finite differences and geophysical exploration,” *Geoscientific Model Development*, vol. 12, no. 3, pp. 1165–1187, 2019 (pages 189, 195).
- [3] F. Luporini *et al.*, “Architecture and performance of devito, a system for automated stencil computation,” *ACM Trans. Math. Softw.*, vol. 46, no. 1, Apr. 2020 (pages 189, 195).
- [4] E. Baysal, D. D. Kosloff, and J. W. Sherwood, “Reverse time migration,” *Geophysics*, vol. 48, no. 11, pp. 1514–1524, 1983 (page 189).
- [5] P. Sava and S. Fomel, “Time-shift imaging condition in seismic migration,” *Geophysics*, vol. 71, no. 6, S209–S217, 2006 (page 189).
- [6] I. Jones, “Tutorial: Migration imaging conditions,” *First break*, vol. 32, no. 12, 2014 (page 189).
- [7] J. Hou and W. W. Symes, “An approximate inverse to the extended born modeling operator,” *Geophysics*, vol. 80, no. 6, R331–R349, 2015 (page 189).
- [8] J. Hou and W. W. Symes, “Accelerating extended least-squares migration with weighted conjugate gradient iteration,” *GEOPHYSICS*, vol. 81, no. 4, S165–S179, Jul. 2016 (page 189).
- [9] J. Hou and W. W. Symes, “Inversion velocity analysis in the subsurface-offset domain,” *Geophysics*, vol. 83, no. 2, R189–R200, 2018 (page 189).
- [10] W. W. Symes, “Migration velocity analysis and waveform inversion,” *Geophysical prospecting*, vol. 56, no. 6, pp. 765–790, 2008 (page 190).
- [11] C. Deng *et al.*, “Openfwi: Large-scale multi-structural benchmark datasets for full waveform inversion,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 6007–6020, 2022 (page 191).
- [12] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, “Multiscale seismic waveform inversion,” *Geophysics*, vol. 60, no. 5, pp. 1457–1473, 1995 (page 191).

- [13] P. A. Witte *et al.*, “A large-scale framework for symbolic implementations of seismic inversion algorithms in julia,” *GEOPHYSICS*, vol. 84, no. 3, F57–F71, 2019. eprint: <https://doi.org/10.1190/geo2018-0174.1> (page 195).
- [14] M. Louboutin *et al.*, “Learned multiphysics inversion with differentiable programming and machine learning,” *The Leading Edge*, vol. 42, no. 7, pp. 474–486, 2023 (page 197).
- [15] Z. Yin, M. Louboutin, O. Møyner, and F. J. Herrmann, “Time-lapse full-waveform permeability inversion: A feasibility study,” *The Leading Edge*, vol. 43, 8 Aug. 2024, (The Leading Edge) (page 197).
- [16] Z. Yin, R. Orozco, M. Louboutin, and F. J. Herrmann, “Solving multiphysics-based inverse problems with learned surrogates and constraints,” *Advanced Modeling and Simulation in Engineering Sciences*, vol. 10, no. 1, p. 14, 2023 (page 200).
- [17] Z. Yin, H. T. Erdinc, A. P. Gahlot, M. Louboutin, and F. J. Herrmann, “Derisking geologic carbon storage from high-resolution time-lapse seismic to explainable leakage detection,” *The Leading Edge*, vol. 42, no. 1, pp. 69–76, 2023 (page 202).
- [18] Z. Yin, R. Orozco, M. Louboutin, and F. J. Herrmann, “Wise: Full-waveform variational inference via subsurface extensions,” *Geophysics*, vol. 89, no. 4, pp. 1–31, 2024 (page 204).
- [19] Z. Yin, R. Orozco, and F. J. Herrmann, “Wiser: Multimodal variational inference for full-waveform inversion without dimensionality reduction,” *arXiv preprint arXiv:2405.10327*, 2024 (page 206).