WISE: full-Waveform variational Inference via Subsurface Extensions

Ziyi Yin[1,*], Rafael Orozco[1,*], Mathias Louboutin[2], Felix J. Herrmann[1]
[1]Georgia Institute of Technology, [2]Devito Codes Ltd

Abstract

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.

1 Introduction

Full-waveform inversion (FWI) plays a pivotal role in exploration, primarily focusing on estimating Earth's subsurface properties from observed seismic data [1]. 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 [2] with physics, our approach harnesses advancements in variational inference [VI, 3] and generative artificial intelligence [AI, 4, 5]. We achieve this by forming common-image gathers [CIGs, 6], followed by training conditional normalizing flows [CNFs, 7] that quantify uncertainties in migration-velocity models.

Outline: First, we delineate the FWI problem and its inherent 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.

2 Methodology

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

2.1 Full-waveform inversion

Estimation of unknown migration-velocity models, $\mathbf{x}$, from noisy seismic data, $\mathbf{y}$ involves inverting nonlinear forward modeling, $\mathcal{F}$, which links $\mathbf{x}$ to $\mathbf{y}$ via $\mathbf{y} = \mathcal{F}(\mathbf{x}) + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon}$ measurement noise. Source/receiver signatures are assumed known and absorbed in $\mathcal{F}$. Solving this nonlinear inverse problem is challenging because of the noise and the non-trivial null-space of the modeling [8]. As a result, multiple migration-velocity models fit the data, necessitating a Bayesian framework for UQ.

* First two authors contributed equally. Correspondence: ziyi.yin@gatech.edu

2.2 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, $p(\mathbf{x}|\mathbf{y})$. Algorithms to calculate statistics from the posterior can be divided into: (i) sampling based algorithms, such as MCMC [9–12]; or (ii) optimization based algorithms, such as VI [3].

Due to the high-dimensionality and expensive wave-based modeling, MCMC becomes impractical for FWI [13]. For this reason, we focus on low-cost VI, which exchanges the computational cost of posterior sampling for neural network training [14–29]. Specifically, we employ amortized VI, which incurs offline computational training cost but enables cheap online posterior inference on many datasets $\mathbf{y}$ [30]. Next, we discuss how to use CNFs for amortized VI.

2.3 Amortized variational inference with conditional normalizing flows

During VI, the posterior distribution $p(\mathbf{x}|\mathbf{y})$ is approximated by the surrogate, $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{y})$, with learnable parameters, $\boldsymbol{\theta}$. Because of their low-cost training and rapid sampling [31, 32], CNFs are suitable to act as surrogates for the posterior with training that 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:

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

Here, $f_{\boldsymbol{\theta}}$ is the CNF with network parameters, $\boldsymbol{\theta}$, and Jacobian, $\mathbf{J}_{f_{\boldsymbol{\theta}}}$. 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 turn 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.

2.4 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, Radev et al. [33] introduced fixed reduced-size *summary statistics* that encapsulate observed data and inform the posterior distribution. Building on this concept, Orozco et al. [21] uses the gradient as the *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 data or on the gradient. Based on this principle, Siahkoohi et al. [20] and Siahkoohi et al. [22] used reverse-time migration [RTM, 34], given by the action of the adjoint of the linearized Born modeling [35–37], 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. Hou and Symes [38] showed that least-squares migration [39] can perfectly fit the data for correct migration-velocity models, but this fit fails for wrong velocity models. This highlights a fundamental limitation in cases where the velocity model is wrong and RTM does not capture the information, which leads to a biased posterior. For a wrong initial FWI-velocity model $\mathbf{x}_0$, $p\left(\mathbf{x}|\mathbf{y}\right) \neq p\left(\mathbf{x}\middle|\nabla\mathcal{F}\left(\mathbf{x}_0\right)^\top \mathbf{y}\right)$ 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.

2.5 Common-image gathers as summary statistics

Migration-velocity analysis has a rich history in the literature [6, 40]. Following Liu et al. [41];Hou and Symes [42];Hou and Symes [43], we employ relatively artifact-free subsurface-offset extended

Born modeling to calculate summary statistics. Thanks to being closer to an isometry—i.e, the adjoint of extended Born modeling is closer to its inverse [44, 45] and therefore preserves information, its adjoint can nullify residuals even when the FWI-velocity model is incorrect as shown by Hou and Symes [38]. Geng et al. [46] 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, which leads to less biased physics-informed summary statistics for a wrong initial FWI-velocity model. Formally, this means $p\left(\mathbf{x}|\mathbf{y}\right) \approx p\left(\mathbf{x}\middle|\overline{\nabla\mathcal{F}}\left(\mathbf{x}_0\right)^\top \mathbf{y}\right)$, 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, $\overline{\nabla\mathcal{F}}\left(\mathbf{x}_0\right)^\top \mathbf{y}$, guided by the objective of Equation 1. Our case studies will demonstrate that even with wrong initial FWI-velocity models, CIGs encapsulate more information, enabling the trained CNFs to generate accurate migration-velocity models conditioned by the data.

3 Synthetic case studies

Our study evaluates the performance of WISE through synthetic case studies on two datasets: the CurveFault-A dataset of Open FWI [47] and 2D slices of the Compass dataset [48]. We aim to compare the quality of posterior samples informed by RTM alone versus those informed by CIGs.

3.1 Open FWI

The CurveFault-A dataset comprises velocity models with significant variability across samples, which poses challenges for deep learning methods [49, 50]. 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.

Dataset generation and network training. We select 3000 velocity models of 640 m by 640 m, each with 64 equally spaced receivers at $10\mathrm{m}$ tow depth and 16 randomly placed sources [51]. The surface is assumed absorbing. Using a $15\mathrm{Hz}$ central frequency Ricker wavelet with energy below $3\mathrm{Hz}$ removed for realism, acoustic data is simulated with Devito [52, 53] and JUDI.jl [54, 55]. Uncorrelated band-limited Gaussian noise is added (S/N $12\mathrm{dB}$) before migrating each dataset with a 1D initial FWI-velocity model calculated by averaging the corresponding true model horizontally. CIGs are computed for 101 subsurface offsets ranging from $-250\mathrm{m}$ to $+250\mathrm{m}$ [56, 57]. The dataset is split into 2800 training, 150 validation, and 50 test samples. Two CNFs are trained: one with velocity-RTM pairs and another with velocity-CIGs pairs.

Results. Results on two untested samples by our CNFs are included in Figure 1 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. Motivated by these results, we will examine a more realistic example with intricate geological structures next.

3.2 Compass model

To validate WISE in a more realistic setting and examine uncertainty in imaging, we consider the Compass dataset, known for its "velocity kickback" challenge for FWI algorithms. By comparing, for a poor initial FWI-velocity model, the conditional mean of samples of the migration-velocity model informed by CIGs and RTM, we 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 [58].

Dataset generation and network training. We take 1040 2D slices of the Compass model of 6.4 km by 3.2 km, with 512 equally spaced sources towed at $12.5\mathrm{m}$ depth and 64 ocean-bottom nodes at random locations. Source and noise setup remains the same. The arithmetic mean over all velocity models is used as the 1D initial FWI-velocity model (shown in Figure 2(b)). 51 subsurface offsets

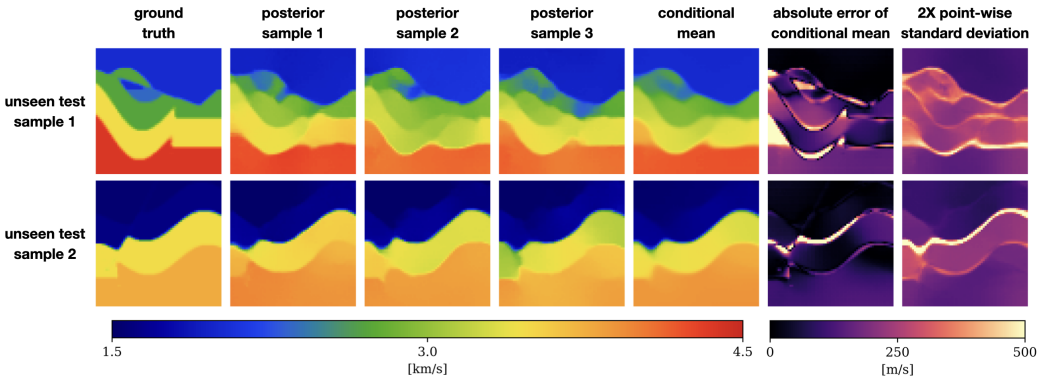

Figure 1: Applying WISE for two unseen test samples in Open FWI CurveFault-A dataset.

ranging from -500m to $+500\text{m}$ are used to compute CIGs (shown in Figure 3(a)). The dataset is divided into 800 training, 190 validation, and 50 test samples, with CNFs trained over 200 epochs.

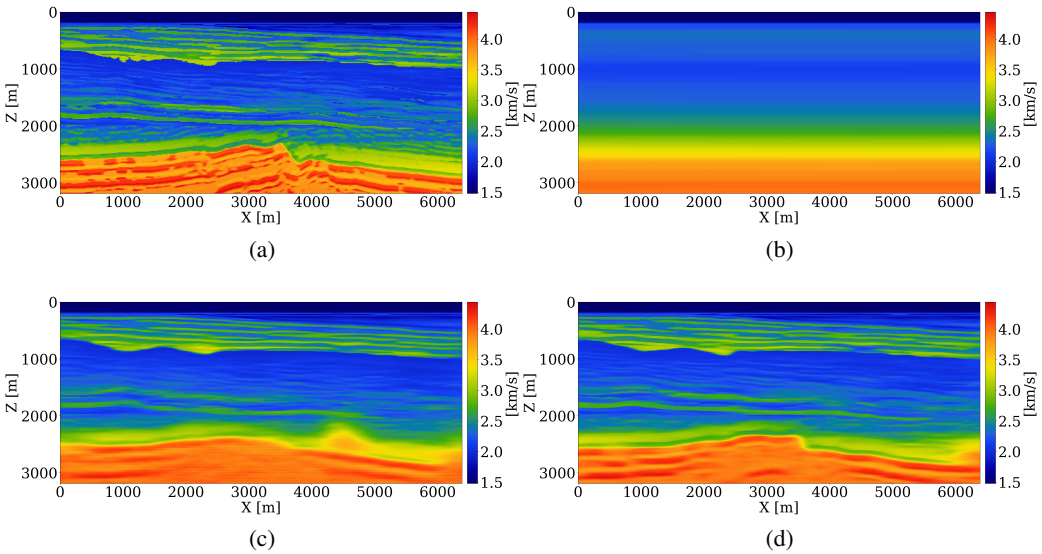

Figure 2: (a) an unseen ground-truth velocity model; (b) 1D initial FWI-velocity model; (c) conditional mean estimate for RTM as summary statistics (SSIM $= 0.48$); (d) conditional mean estimate from WISE (SSIM $= 0.56$).

Results. Our method's performance is evaluated on an unseen 2D Compass slice shown in Figure 2(a). When RTM is used to summarize the data, the conditional mean estimate (Figure 2(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 2(d)). For the 50 test samples, the SSIM scores with CIGs yield a mean of 0.63, outperforming RTM-based statistics with a mean SSIM of 0.52.

Quality control. To verify the inferred migration-velocity model, CIGs calculated for the initial FWI-velocity model (Figure 2(b)), plotted in Figure 3(a), are juxtaposed against CIGs calculated for the inferred migration-velocity model (Figure 2(d)), plotted in Figure 3(b). 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.

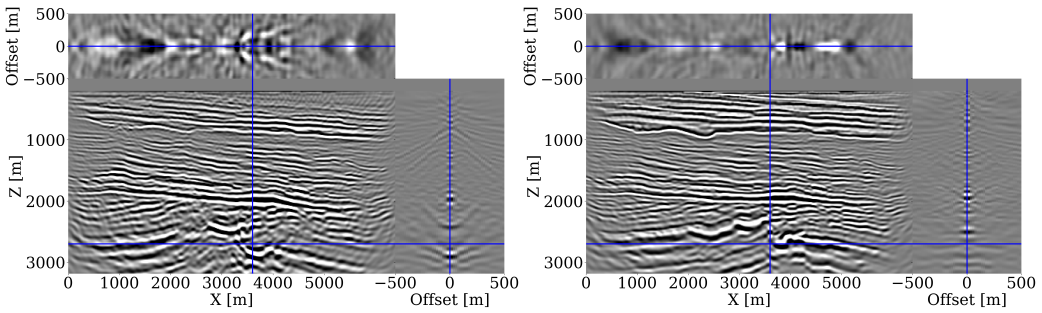

Figure 3: CIGs calculated by the initial FWI-velocity model given by (a) the 1D background model (shown in Figure 2(b)) or by (b) the conditional mean estimate (shown in Figure 2(d)).

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, comparisons are made between the RTM computed for the conditional mean (Figure 4(a)) and RTMs computed for individual posterior samples that vary significantly judged by the point-wise standard deviations in Figure 4(b). These deviations increase with depth and correlate with complex geology where the RTM-based inference struggled. To understand how this uncertainty propagates to RTM images, forward uncertainty is assessed by carrying out RTMs for different posterior samples with results for the standard deviations plotted in Figure 4(c). 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 reference image (Figure 4(a)) and RTMs for different posterior samples are calculated with a local cross-correlation technique [59] and included in Figure 4(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.

4 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 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 that focus at inference. WISE also produced two types of uncertainty, namely (i) inverse uncertainty in migration-velocity model estimation from noisy data, and (ii) forward uncertainty where uncertainty in migration-velocity models is propagated to uncertainty in amplitude and positioning of imaged reflectors. The latter can also be assessed for tasks like horizon tracking [12, 60] and seismic interpretation [61].

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 [62]. To improve single-observation performance, particularly for out-of-distribution samples, computationally more expensive latent space corrections [22] can be employed that incorporate the physics. Moreover, velocity continuation methods [44, 63] could be used including recent advances in neural operators [64]. These could offset the cost of running RTMs for each posterior sample.

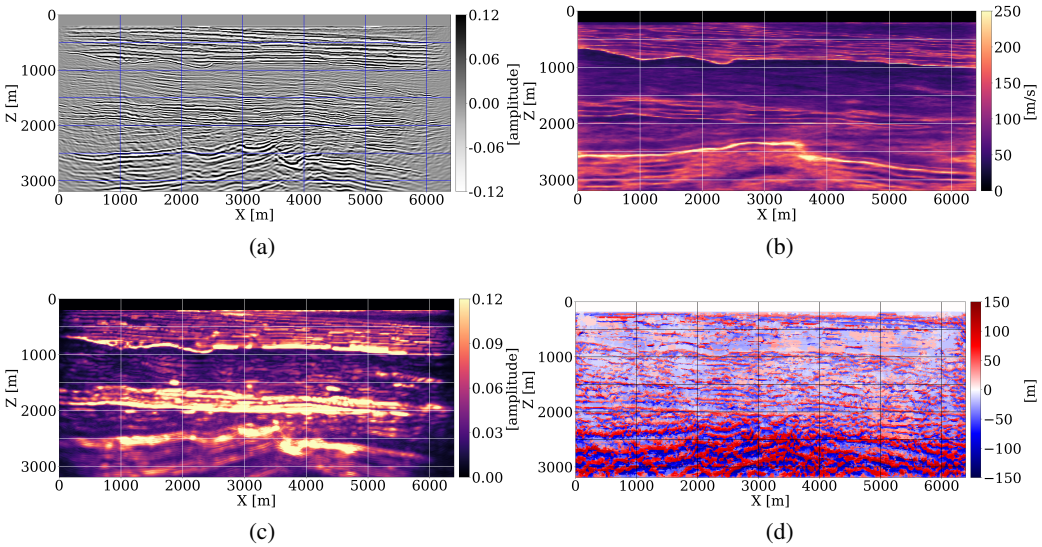

Figure 4: Using WISE in the downstream imaging task: (a) imaged reflectors using the conditional mean estimate (shown in Figure 2(d)) as the migration velocity; (b) point-wise standard deviation of the posterior velocity samples; (c) point-wise deviation of the imaged reflectivities; (d) point-wise maximum depth shift. Note: (a) and (c) are normalized by the same constant. The images are all grided for visualization purpose.

5 Conclusions

We present WISE, full-**W**aveform variational **I**nference via **S**ubsurface **E**xtensions, 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 Acknowledgement

This research was carried out with the support of Georgia Research Alliance and partners of the ML4Seismic Center. The authors would like to thank Charles Jones (Osokey) for the constructive discussion. The authors gratefully acknowledge the contribution of OpenAI's ChatGPT for refining sentence structure and enhancing the overall readability of this manuscript.

References

[1] Jean Virieux and Stéphane Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, 2009.

[2] Siwei Yu and Jianwei Ma. Deep learning for geophysics: Current and future trends. *Reviews of Geophysics*, 59(3):e2021RG000742, 2021.

[3] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine learning*, 37:183–233, 1999.

[4] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022.

[5] Felix J Herrmann. President's page: Digital twins in the era of generative AI. *The Leading Edge*, 42(11):730–732, 2023.

[6] William W Symes. Migration velocity analysis and waveform inversion. *Geophysical prospecting*, 56(6):765–790, 2008.

[7] Christina Winkler, Daniel Worrall, Emiel Hoogeboom, and Max Welling. Learning likelihoods with conditional normalizing flows. *arXiv preprint arXiv:1912.00042*, 2019.

[8] Albert Tarantola. Inversion of seismic reflection data in the acoustic approximation. *GEOPHYSICS*, 49(8):1259–1266, 08 1984. doi: 10.1190/1.1441754. URL http://dx.doi.org/10.1190/1.1441754.

[9] Max Welling and Yee W Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 681–688, 2011.

[10] Gregory Ely, Alison Malcolm, and Oleg V Poliannikov. Assessing uncertainties in velocity models and images with a fast nonlinear uncertainty quantification method. *Geophysics*, 83(2): R63–R75, 2018.

[11] M Kotsi, A Malcolm, and G Ely. Uncertainty quantification in time-lapse seismic imaging: A full-waveform approach. *Geophysical Journal International*, 222(2):1245–1263, 2020.

[12] Ali Siahkoohi, Gabrio Rizzuti, and Felix J. Herrmann. Deep bayesian inference for seismic imaging with tasks. *GEOPHYSICS*, 87(5):S281–S302, 08 2022. doi: 10.1190/geo2021-0666.1. URL http://dx.doi.org/10.1190/geo2021-0666.1.

[13] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 07 2003. doi: 10.1201/9780429258480. URL http://dx.doi.org/10.1201/9780429258480.

[14] Gabrio Rizzuti, Ali Siahkoohi, Philipp A. Witte, and Felix J. Herrmann. Parameterizing uncertainty by deep invertible networks: An application to reservoir characterization. *SEG Technical Program Expanded Abstracts 2020*, 09 2020. doi: 10.1190/segam2020-3428150.1. URL http://dx.doi.org/10.1190/segam2020-3428150.1.

[15] Yuxiao Ren, Philipp A Witte, Ali Siahkoohi, Mathias Louboutin, Ziyi Yin, and Felix J Herrmann. Seismic velocity inversion and uncertainty quantification using conditional normalizing flows. In *AGU Fall Meeting 2021*. AGU, 2021.

[16] Rajiv Kumar, Maria Kotsi, Ali Siahkoohi, and Alison Malcolm. Enabling uncertainty quantification for seismic data preprocessing using normalizing flows (nf) – an interpolation example. *First International Meeting for Applied Geoscience & Energy Expanded Abstracts*, 09 2021. doi: 10.1190/segam2021-3583705.1. URL http://dx.doi.org/10.1190/segam2021-3583705.1.

[17] Ali Siahkoohi and Felix J. Herrmann. Learning by example: Fast reliability-aware seismic imaging with normalizing flows. *First International Meeting for Applied Geoscience & Energy Expanded Abstracts*, 09 2021. doi: 10.1190/segam2021-3581836.1. URL http://dx.doi.org/10.1190/segam2021-3581836.1.

[18] Ali Siahkoohi, Gabrio Rizzuti, Mathias Louboutin, Philipp A Witte, and Felix J Herrmann. Preconditioned training of normalizing flows for variational inference in inverse problems. *arXiv preprint arXiv:2101.03709*, 2021.

[19] Xin Zhang, Muhammad Atif Nawaz, Xuebin Zhao, and Andrew Curtis. An introduction to variational inference in geophysical inverse problems. In *Advances in Geophysics*, volume 62, pages 73–140. Elsevier, 2021.

[20] Ali Siahkoohi, Rafael Orozco, Gabrio Rizzuti, and Felix J Herrmann. Wave-equation-based inversion with amortized variational bayesian inference. *arXiv preprint arXiv:2203.15881*, 2022.

[21] Rafael Orozco, Ali Siahkoohi, Gabrio Rizzuti, Tristan van Leeuwen, and Felix J Herrmann. Adjoint operators enable fast and amortized machine learning based bayesian uncertainty quantification. In *Medical Imaging 2023: Image Processing*, volume 12464, pages 357–367. SPIE, 2023.

[22] Ali Siahkoohi, Gabrio Rizzuti, Rafael Orozco, and Felix J. Herrmann. Reliable amortized variational inference with physics-based latent distribution correction. *GEOPHYSICS*, 88 (3):R297–R322, 04 2023. doi: 10.1190/geo2022-0472.1. URL http://dx.doi.org/10.1190/geo2022-0472.1.

[23] Jeffrey Wen, Rizwan Ahmad, and Philip Schniter. A conditional normalizing flow for accelerated multi-coil mr imaging. *arXiv preprint arXiv:2306.01630*, 2023.

[24] Rafael Orozco, Mathias Louboutin, Ali Siahkoohi, Gabrio Rizzuti, Tristan van Leeuwen, and Felix Herrmann. Amortized normalizing flows for transcranial ultrasound with uncertainty quantification. *arXiv preprint arXiv:2303.03478*, 2023.

[25] Xin Zhang, Angus Lomas, Muhong Zhou, York Zheng, and Andrew Curtis. 3-d bayesian variational full waveform inversion. *Geophysical Journal International*, 234(1):546–561, 2023.

[26] Rafael Orozco, Ali Siahkoohi, Mathias Louboutin, and Felix J Herrmann. Refining amortized posterior approximations using gradient-based summary statistics. *arXiv preprint arXiv:2305.08733*, 2023.

[27] Xin Zhang and Andrew Curtis. Bayesian variational time-lapse full-waveform inversion. *arXiv preprint arXiv:2308.08805*, 2023.

[28] Dominik Strutz and Andrew Curtis. Variational bayesian experimental design for geophysical applications. *arXiv preprint arXiv:2307.01039*, 2023.

[29] Abhinav Prakash Gahlot, Huseyin Tuna Erdinc, Rafael Orozco, Ziyi Yin, and Felix J Herrmann. Inference of CO2 flow patterns–a feasibility study. *arXiv preprint arXiv:2311.00290*, 2023.

[30] Jakob Kruse, Gianluca Detommaso, Ullrich Köthe, and Robert Scheichl. Hint: Hierarchical invertible neural transport for density estimation and bayesian inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 8191–8199, 2021.

[31] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.

[32] Mathias Louboutin, Ziyi Yin, Rafael Orozco, Thomas J Grady, Ali Siahkoohi, 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(7): 474–486, 2023.

[33] Stefan T. Radev, Ulf K. Mertens, Andreas Voss, Lynton Ardizzone, and Ullrich Kothe. Bayesflow: Learning complex stochastic models with invertible neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(4):1452–1466, 04 2022. doi: 10.1109/tnnls.2020.3042395. URL http://dx.doi.org/10.1109/TNNLS.2020.3042395.

[34] Edip Baysal, Dan D Kosloff, and John WC Sherwood. Reverse time migration. *Geophysics*, 48 (11):1514–1524, 1983.

[35] Gerard T. Schuster. Least-squares cross-well migration. *SEG Technical Program Expanded Abstracts 1993*, 01 1993. doi: 10.1190/1.1822308. URL http://dx.doi.org/10.1190/1.1822308.

[36] Tamas Nemeth, Chengjun Wu, and Gerard T. Schuster. Least-squares migration of incomplete reflection data. *GEOPHYSICS*, 64(1):208–221, 01 1999. doi: 10.1190/1.1444517. URL http://dx.doi.org/10.1190/1.1444517.

[37] Justin Alsing and Benjamin Wandelt. Generalized massive optimal data compression. *Monthly Notices of the Royal Astronomical Society: Letters*, 476(1):L60–L64, 2018.

[38] Jie Hou and William W. Symes. Accelerating extended least-squares migration with weighted conjugate gradient iteration. *GEOPHYSICS*, 81(4):S165–S179, 07 2016. doi: 10.1190/geo2015-0499.1. URL http://dx.doi.org/10.1190/geo2015-0499.1.

[39] Chong Zeng, Shuqian Dong, and Bin Wang. Least-squares reverse time migration: Inversion-based imaging toward true reflectivity. *The Leading Edge*, 33(9):962–968, 09 2014. doi: 10.1190/tle33090962.1. URL http://dx.doi.org/10.1190/tle33090962.1.

[40] Paul Sava and Biondo Biondi. Wave-equation migration velocity analysis. i. theory. *Geophysical Prospecting*, 52(6):593–606, 2004.

[41] Yujin Liu, William W Symes, and Zhenchun Li. Multisource least-squares extended reverse-time migration with preconditioning guided gradient method. In *SEG International Exposition and Annual Meeting*, pages SEG–2013. SEG, 2013.

[42] Jie Hou and William W Symes. An approximate inverse to the extended born modeling operator. *Geophysics*, 80(6):R331–R349, 2015.

[43] Jie Hou and William W Symes. Inversion velocity analysis in the subsurface-offset domain. *Geophysics*, 83(2):R189–R200, 2018.

[44] Mengmeng Yang, Marie Graff, Rajiv Kumar, and Felix J. Herrmann. Low-rank representation of omnidirectional subsurface extended image volumes. *GEOPHYSICS*, 86(3):S165–S183, 03 2021. doi: 10.1190/geo2020-0152.1. URL http://dx.doi.org/10.1190/geo2020-0152.1.

[45] Fons ten Kroode. An omnidirectional seismic image extension. *Inverse Problems*, 39(3):035003, 01 2023. doi: 10.1088/1361-6420/acb285. URL http://dx.doi.org/10.1088/1361-6420/acb285.

[46] Zhicheng Geng, Zeyu Zhao, Yunzhi Shi, Xinming Wu, Sergey Fomel, and Mrinal Sen. Deep learning for velocity model building with common-image gather volumes. *Geophysical Journal International*, 228(2):1054–1070, 2022.

[47] Chengyuan Deng, Shihang Feng, Hanchen Wang, Xitong Zhang, Peng Jin, Yinan Feng, Qili Zeng, Yinpeng Chen, and Youzuo Lin. Openfwi: Large-scale multi-structural benchmark datasets for full waveform inversion. *Advances in Neural Information Processing Systems*, 35: 6007–6020, 2022.

[48] C. E. Jones, J. A. Edgar, J. I. Selvage, and H. Crook. Building complex synthetic models to evaluate acquisition geometries and velocity inversion technologies. *Proceedings*, 06 2012. doi: 10.3997/2214-4609.20148575. URL http://dx.doi.org/10.3997/2214-4609.20148575.

[49] Peng Jin, Xitong Zhang, Yinpeng Chen, Sharon Xiaolei Huang, Zicheng Liu, and Youzuo Lin. Unsupervised learning of full-waveform inversion: Connecting cnn and partial differential equation in a loop. *arXiv preprint arXiv:2110.07584*, 2021.

[50] Peng Jin, Yinan Feng, Shihang Feng, Hanchen Wang, Yinpeng Chen, Benjamin Consolvo, Zicheng Liu, and Youzuo Lin. Does full waveform inversion benefit from big data? *arXiv preprint arXiv:2307.15388*, 2023.

[51] Felix J Herrmann. Randomized sampling and sparsity: Getting more information from fewer samples. *Geophysics*, 75(6):WB173–WB187, 2010.

[52] M. Louboutin, M. Lange, F. Luporini, N. Kukreja, P. A. Witte, F. J. Herrmann, P. Velesko, and G. J. Gorman. Devito (v3.1.0): an embedded domain-specific language for finite differences and geophysical exploration. *Geoscientific Model Development*, 12(3):1165–1187, 2019. doi: 10.5194/gmd-12-1165-2019. URL https://www.geosci-model-dev.net/12/1165/2019/.

[53] Fabio Luporini, Mathias Louboutin, Michael Lange, Navjot Kukreja, Philipp Witte, Jan Hückelheim, Charles Yount, Paul H. J. Kelly, Felix J. Herrmann, and Gerard J. Gorman. Architecture and performance of devito, a system for automated stencil computation. *ACM Trans. Math. Softw.*, 46(1), apr 2020. ISSN 0098-3500. doi: 10.1145/3374916. URL https://doi.org/10.1145/3374916.

[54] Philipp A. Witte, Mathias Louboutin, Navjot Kukreja, Fabio Luporini, Michael Lange, Gerard J. Gorman, and Felix J. Herrmann. A large-scale framework for symbolic implementations of seismic inversion algorithms in julia. *GEOPHYSICS*, 84(3):F57–F71, 2019. doi: 10.1190/geo2018-0174.1. URL https://doi.org/10.1190/geo2018-0174.1.

[55] Mathias Louboutin, Philipp Witte, Ziyi Yin, Henryk Modzelewski, Kerim, Carlos da Costa, and Peterson Nogueira. slimgroup/judi.jl: v3.2.3, March 2023. URL https://doi.org/10.5281/zenodo.7785440.

[56] Mathias Louboutin. *slimgroup/ImageGather.jl: v0.2.6*. Zenodo, 06 2023. doi: 10.5281/ZENODO.7997993. URL https://zenodo.org/record/7997993.

[57] Mathias Louboutin and Felix J. Herrmann. Wave-based inversion at scale on graphical processing units with randomized trace estimation. *Geophysical Prospecting*, 07 2023. doi: 10.1111/1365-2478.13405. URL http://dx.doi.org/10.1111/1365-2478.13405.

[58] Oleg V Poliannikov and Alison E Malcolm. The effect of velocity uncertainty on migrated reflectors: Improvements from relative-depth imaging. *Geophysics*, 81(1):S21–S29, 2016.

[59] Dave Hale. An efficient method for computing local cross-correlations of multi-dimensional signals. *CWP Report*, 656:282, 2006.

[60] Xinming Wu and Sergey Fomel. Least-squares horizons with local slopes and multigrid correlations. *Geophysics*, 83(4):IM29–IM40, 2018.

[61] Harpreet Kaur, Qie Zhang, Philipp Witte, Luming Liang, Long Wu, and Sergey Fomel. Deep-learning-based 3d fault detection for carbon capture and storage. *Geophysics*, 88(4):IM101–IM112, 2023.

[62] Joe Marino, Yisong Yue, and Stephan Mandt. Iterative amortized inference. In *International Conference on Machine Learning*, pages 3403–3412. PMLR, 2018.

[63] Sergey Fomel. Time-migration velocity analysis by velocity continuation. *Geophysics*, 68(5):1662–1672, 2003.

[64] Ali Siahkoohi, Mathias Louboutin, and Felix J Herrmann. Velocity continuation with fourier neural operators for accelerated uncertainty quantification. In *SEG International Exposition and Annual Meeting*, page D011S092R004. SEG, 2022.