

GEOPHYSICS®

WISER: multimodal variational inference for full-waveform inversion without dimensionality reduction

Journal:	Geophysics
Manuscript ID	GEO-2024-0483.R1
Manuscript Type:	Letters
Keywords:	full-waveform inversion, imaging, artificial intelligence, machine learning, inversion
Manuscript Focus Area:	Geophysics Letters

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

WISER: Multimodal variational inference for full-waveform inversion without dimensionality reduction

Ziyi Yin, Rafael Orozco, Felix J. Herrmann

Georgia Institute of Technology

(November 27, 2024)

GEO-2024-0483.R1

Running head: **cost-effective full-waveform inference**

ABSTRACT

We present a semi-amortized variational inference framework designed for computationally feasible uncertainty quantification in full-waveform inversion to explore the multimodal posterior distribution without dimensionality reduction. The framework is called **WISER**, short for full-**W**aveform variational **I**nference via **S**ubsurface **E**xtensions with **R**efinements. WISER builds on top of a supervised generative artificial intelligence method that performs 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 wave physics. Case studies illustrate that WISER is capable of full-resolution, computationally feasible, and reliable uncertainty estimates of velocity models and imaged reflectivities.

INTRODUCTION

Full-waveform inversion (FWI) aims to estimate unknown multi-dimensional velocity models, \mathbf{x} , from noisy seismic data, \mathbf{y} , by inverting the nonlinear forward operator, \mathcal{F} , which relates \mathbf{x} and \mathbf{y} via $\mathbf{y} = \mathcal{F}(\mathbf{x}) + \boldsymbol{\epsilon}$ with measurement noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. 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 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 always a critical consideration of UQ. To circumvent the costs associated with global optimization, several approaches have attempted localized UQ (Fang et al., 2018; Keating and Innanen, 2021; Izzatullah et al., 2024; Hoffmann et al., 2024). However, these approaches may not capture the full complexities of multimodal parameter spaces. In contrast, a Bayesian inference approach offers a costly but comprehensive characterization of the posterior distribution, $p(\mathbf{x} | \mathbf{y})$.

Bayesian inference algorithms are broadly categorized into two groups. The first, sampling-based methods, like Markov-chain Monte Carlo (MCMC), 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 (Fang et al., 2020; Liang et al., 2023; Wei et al., 2023, 2024a,b; Dhabaria and Singh, 2024), which could bias the inference results, rendering them impractical for multidimensional

UQ studies especially when solutions are nonunique.

The second category, optimization-based methods, like variational inference (VI, Zhang et al., 2021), 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 (Baldassari et al., 2024) and conditional normalizing flows (CNFs, Winkler et al., 2019). After training, amortized VI provides rapid sampling during inference (Siahkoobi et al., 2023; Orozco et al., 2023a; Sun et al., 2024), exemplified by the WISE framework (Yin et al., 2024b) for FWI problems. However, these methods may suffer from an *amortization gap* — implying that the amortized networks may only deliver suboptimal inference for a given observation during inference, particularly when trained with limited examples or when there exists a discrepancy between training and inference (Marino et al., 2018). Conversely, non-amortized VI dedicates all computational resources to the online inference (Zhao et al., 2022; Zhang et al., 2023; Zhang and Curtis, 2024). They result in more accurate inference, but the costly optimization has to be carried out repeatedly for new observations. Also, incorporating realistic priors requires embedding them through density evaluations, such as those enabled by NFs (Kruse et al., 2021; Louboutin et al., 2023; Yin et al., 2023).

This paper introduces the WISER algorithm (shown in Algorithm 1) as a semi-amortized VI framework to facilitate computationally feasible and reliable UQ for

multidimensional 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 makes 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. The performance of WISER is demonstrated through realistic synthetic 2D case studies using the Compass model (Jones et al., 2012).

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, which involves generating a training dataset (lines 3—9 of Algorithm 1) and training the CNFs (line 11—12).

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, Hou and Symes, 2016) for each observed data with a 1D initial velocity model, \mathbf{x}_0 . 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

Algorithm 1 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:    $\mathbf{y}^{(i)} = \mathcal{F}(\mathbf{x}^{(i)}) + \boldsymbol{\epsilon}^{(i)}$  # forward modeling
8:    $\bar{\mathbf{y}}^{(i)} = \nabla \mathcal{F}(\mathbf{x}_0)^\top \mathbf{y}^{(i)}$  # extended migration
9: end for
10:
11: Network training
12:  $\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \quad \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{2} \|\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{x}^{(i)}; \bar{\mathbf{y}}^{(i)})\|_2^2 - \log |\det \mathbf{J}_{f_{\boldsymbol{\theta}}}| \right)$ 
13:
14: Online inference phase
15:
16:  $\bar{\mathbf{y}}_{\text{obs}} = \nabla \mathcal{F}(\mathbf{x}_0)^\top \mathbf{y}_{\text{obs}}$  # extended migration
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 # physics-based gradient update
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  $p(\mathbf{x}|\mathbf{y}_{\text{obs}})$ 

```

CIGs as the set of physics-informed summary statistics not only preserves information from the observed seismic data (ten Kroode, 2023) but also enhances the training of CNFs in the next stage (Radev et al., 2020; Orozco et al., 2023b), as they help to decode the wave physics, translating prestack data to the image (subsurface-offset) domain.

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, $\mathbf{x}^{(i)}$, conditioned on the summary statistics (CIGs), into Gaussian noise from a standard multivariate normal distribution (as evident by use of the ℓ_2 norm).

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.

PHYSICS-BASED REFINEMENTS (LINES 22—32)

Consider a given observation, \mathbf{y}_{obs} , and its corresponding posterior samples, $\mathbf{x}_i \sim p(\mathbf{x} \mid \bar{\mathbf{y}}_{\text{obs}})$. The latent space representations generated by the trained CNFs, $\hat{\mathbf{z}}_i = f_{\boldsymbol{\theta}^*}(\mathbf{x}_i; \bar{\mathbf{y}}_{\text{obs}})$, may not conform exactly to the standard Gaussian distribution during inference. Thus, using $f_{\boldsymbol{\theta}^*}^{-1}(\mathbf{z} \mid \bar{\mathbf{y}}_{\text{obs}})$ for inference with $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ will not result in the desired posterior distribution. We refer to the distribution of $\hat{\mathbf{z}}_i$ as the shifted latent distribution, denoted by $p_{\boldsymbol{\theta}^*}(\mathbf{z} \mid \bar{\mathbf{y}}_{\text{obs}})$. Mathematically, the amortization gap is quantified as $\mathbb{KL}(\mathcal{N}(\mathbf{0}, \mathbf{I}) \parallel p_{\boldsymbol{\theta}^*}(\mathbf{z} \mid \bar{\mathbf{y}}_{\text{obs}})) > 0$, where \mathbb{KL} is the Kullback-Leibler (KL) divergence. To address this issue, we follow Siahkoochi et al. (2023) to apply latent space corrections to fine-tune the trained CNFs. This involves integrating a shallower, yet invertible, network, specifically trained to map samples from the standard Gaussian distribution to the (unknown) shifted latent distribution, thereby reducing this amortization gap.

In pursuit of this goal, we adopt a transfer learning approach by freezing the weights of the trained CNFs and updating only the weights of the shallower network.

¹Here, we use h_{ϕ} to mean the identity operator. It is a placeholder for later where h_{ϕ} will be repurposed to represent more complex operators based on physics-based corrections.

The objective function reads:

$$\begin{aligned} \underset{\phi}{\text{minimize}} \quad & \mathbb{E}_{\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \quad \mathbb{KL}(p(h_{\phi}(\mathbf{z})) \| p_{\theta^*}(\mathbf{z} | \bar{\mathbf{y}}_{\text{obs}})) \\ & = \frac{1}{2\sigma^2} \|\mathcal{F} \circ f_{\theta^*}^{-1}(h_{\phi}(\mathbf{z}); \bar{\mathbf{y}}_{\text{obs}}) - \mathbf{y}_{\text{obs}}\|_2^2 + \frac{1}{2} \|h_{\phi}(\mathbf{z})\|_2^2 - \log |\det \mathbf{J}_{h_{\phi}}|. \end{aligned} \quad (1)$$

Equation 1 minimizes the reverse KL divergence between the distribution of the corrected latent variables, $p(h_{\phi}(\mathbf{z}))$, and the shifted latent distribution, $p_{\theta^*}(\mathbf{z} | \bar{\mathbf{y}}_{\text{obs}})$, with h_{ϕ} being the refinement network. A detailed proof can be found in Siahkoohi et al. (2021); Siahkoohi et al. (2023), along with related references. For linear inverse problems in seismic imaging, Siahkoohi et al. (2023) demonstrated that an elementwise scaling and shift mechanism is sufficient to close the gap. However, given the nonconvex nature of FWI, we employ h_{ϕ} as a more general invertible network.

From an engineering perspective, 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 term ties the posterior samples closer to the observed data. The second term prevents 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. The third term prevents mode collapse (Rezende and Mohamed, 2015).

Equation 1 offers a fine-tuning approach that leverages the full multidimensional wave physics to refine WISE for a given observation during inference. 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 weak deep prior strategy (Siahkoohi et al., 2020) to reformulate equation 1 into a weak form by allowing the network output to be only “weakly” enforced to be the corrected velocity models. The objective function for this weak formulation reads:

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

Here, $\{\mathbf{z}_i\}_{i=1}^M$ are M realizations of standard Gaussian noise. $\{\mathbf{x}_i\}_{i=1}^M$ are their corresponding particles (i.e., velocity models) initialized in line 17-19 of Algorithm 1. 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, $N(\mathbf{0}, \gamma^2 \mathbf{I})$, 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 1. This weak formulation also supports optimization strategies for updating the velocity models with physical constraints and multiscale optimization techniques (Esser et al., 2018; Peters et al., 2019).

WISER takes full computational advantage of this weak formulation by employing

a nested loop structure. The outer loop is dedicated to updating \mathbf{M} velocity models, \mathbf{x}_i , through costly gradient descent steps (lines 24—27 of Algorithm 1), 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).

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. We follow Yin et al. (2024b) to train the amortized CNFs.

To test the robustness and adaptability of WISER when faced with unexpected conditioning data at inference, we evaluate WISER’s performance under out-of-distribution (OOD) scenarios. We introduce alterations to an unseen velocity model drawn from the same statistical distribution on which the CNFs were trained, depicted in Figure 1(c), through an elementwise perturbation shown in Figure 1(a). This manipulation modifies the velocity values across different depth levels, resulting in a significant shift in their statistical distribution, illustrated in Figure 1(b). We use the perturbed velocity as the unseen ground-truth velocity model in this case study, shown

in Figure 1(d). 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 1(e). The yellow histograms in Figure 1(b) 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 ocean-bottom node (OBN) per gradient. We also employ the frequency continuation method (Bunks et al., 1995) to compute the gradient in line 25 of the WISER algorithm, transitioning gradually from low-frequency to high-frequency data. This results in 40 datapasses or 5120 PDE solves² in total.

[Figure 1 about here.]

Observations

WISER produces more accurate posterior samples shown in Figure 1(f). Furthermore, the statistical distribution of the velocity values in the WISER posterior samples (green histogram in Figure 1(b)) aligns better with the distribution of the unseen ground-truth velocity values (blue histogram in Figure 1(b)), demonstrating WISER’s robustness against potential distribution shifts during inference.

²Datapass means a forward and an adjoint pass for the entire OBN dataset.

Impact on imaging

The imaging results from WISE (Figure 1(g)) and WISER (Figure 1(h)) 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 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 also align more accurately with the estimated CM migration-velocity model, particularly visible in Figure 1 in the ancillary material.

DISCUSSION

WISER leverages both genAI and physics to achieve a semi-amortized VI framework for scalable 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.

Related work and computational cost analysis

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 multidimensional FWI problems.

UQ problems inherently involve a trade-off between offline and online computational costs. Access to a larger number of velocity models and CIGs during the offline phase enhances the accuracy of amortized inference in WISE. Additionally, training samples that closely resemble actual Earth models could further improve this accuracy. Preliminary studies have demonstrated that genAI workflows can create Earth models from well measurements and imaged seismic data (Erdinc et al., 2024). Leveraging these automated workflows allows us to generate high-fidelity velocity models for the subsequent training of WISE, potentially reducing the online costs required in WISER. Striking the right balance between these offline and online costs remains a crucial consideration. In Table 1, we list the required number of PDE solves for several Bayesian inference methods applied to FWI.

[Table 1 about here.]

Compared to non-amortized VI methods, WISER requires significantly fewer computational resources during the online phase. This efficiency arises because WISE, the precursor to WISER, already provides near-accurate posterior samples, making the subsequent refinement procedure computationally feasible. Additionally, stochastic optimization techniques can be employed in line 25 of Algorithm 1 to control online

refinement costs. Therefore, we anticipate that $\text{datapass}_1 \ll \text{datapass}_2$ in Table 1 and expect that datapass_1 will not be prohibitively large. We believe this computational cost analysis would benefit from a benchmark case study where different methods are applied and fairly compared, which we leave for future work.

Furthermore, WISER employs WISE as a *conditional* prior. During the inference phase, the approximated posterior from WISE is conditioned on the observed data, and WISER uses this approximated posterior as its prior in the objective function, effectively performing sequential Bayesian updating (Rubio et al., 2023). The conditional prior offers a more accurate approximation to the posterior compared to an unconditional prior trained solely on prior samples (Orozco et al., 2021). Arguably, this approach makes WISER more cost-efficient than non-amortized inference methods that only have uninformative priors (Zhang et al., 2023) or that utilize a data-driven but unconditional prior (Kelkar et al., 2021). Beyond reducing computational costs, WISER also ensures that the posterior samples realistically resemble Earth models, thanks to the integration of conditional prior information from WISE.

In Wang et al. (2023) and Taufik et al. (2024), the authors trained a non-conditional diffusion model and conditioned it on various observations by performing gradient steps on the data likelihood, using the learned denoiser as a prior. While diffusion models have shown promise in generating realistic-looking samples, using them for density estimation is not straightforward (Song et al., 2021), whereas it is straightforward with NFs. The WISER algorithm was built using NFs, motivated by memory efficiency, but it could be extended to use diffusion models as the backbone generative model, as

demonstrated by Orozco et al. (2024a).

Correctness of posterior approximation

Since h_ϕ is trained to minimize the reverse KL divergence in equation 1, its composition with the amortized network, f_{θ^*} , is theoretically capable of representing the full Bayesian posterior (Winkler et al., 2019), which is likely multimodal in FWI problems. Especially, the amortized network minimizes the forward KL divergence, preventing mode collapse (Minka et al., 2005). Also, WISER combines forward and reverse KL divergence minimization, guiding the sampling towards high-probability regions (Vaitl et al., 2024). This theoretical capability has been empirically validated on toy examples (Orozco et al., 2024c) and post-stack seismic inversion examples (Rizzuti and Vasconcelos, 2024), where analytical posterior solutions exist.

Figure 1 in the ancillary material shows that WISER’s posterior samples respect both likelihood and prior terms, indicating successful refinement of WISE’s amortized posterior. However, verifying the correctness or multimodality of the approximated posterior remains challenging due to FWI’s high dimensionality and nonlinearity. A low-dimensional FWI case study could employ calibration tests (Guo et al., 2017) and compare WISER’s posterior samples with those from MCMC and other inference methods — an endeavor we reserve for future work.

Scalability

Neural network implementations often face significant memory challenges in 3D problems due to the need to store intermediate activations for backpropagation. However, NFs, composed of invertible layers, mitigate this issue by recomputing intermediate activations during backpropagation instead of storing them in memory. Orozco et al. (2024d) exploits this property for memory efficiency by manually implementing gradient calculations, ensuring that memory consumption during backpropagation remains constant regardless of the number of layers (as shown in Figure 2 of Orozco et al. (2024d)). Practically, this means a CNF with greatly many layers can be used for maximal expressiveness, provided that a constant factor of memory (related to the input and the condition) fits on the device. This inherent memory efficiency makes CNFs particularly suitable for 3D applications, as demonstrated by the 3D WISE case study in Orozco et al. (2024c).

While our computational toolchain supports 3D forward and gradient calculations, deploying WISER in large-scale 3D FWI remains a valuable direction for future research. In our case study, the dimensions of network parameters θ and ϕ are $O(10^6)$ and $O(10^5)$, respectively, with the velocity model size being $O(10^5)$. Further studies should explore how the required number of training samples and network parameters for accurate WISER inference scale with the dimensionality of the parameter of interest.

Simulation gap and out-of-distribution inference

The presented case study illustrates WISER’s capability when the ground-truth velocity model used during inference is geologically similar to those in the training samples. In practical terms, this suggests that WISER can be effectively applied when WISE is trained using velocity models that are drawn from the same region and share similar geological structures, e.g., from the same area in a basin. While these preliminary results are promising, further investigation is needed to determine the extent to which WISER can handle OOD scenarios during inference. Another way to fundamentally address the OOD issues is to diversify the training samples of WISE through a foundation model (Sheng et al., 2023).

We acknowledge that a simulation gap between synthetic and field data is inevitable. However, by using CIGs, we can map information into the extended image space even when the data originates from a misspecified forward model (e.g., elastic versus acoustic). Improving network training to better process this information remains an important avenue for future research. While the simulation gap may persist during WISE inference, WISER is specifically designed to mitigate this issue by employing a physics-based fine-tuning process that aligns the posterior samples with the observed data.

Extension to other geophysical inverse problems

Finally, the WISE(R) framework can extend to quantify uncertainty in other Bayesian inference problems, such as time-lapse monitoring (Gahlot et al., 2023, 2024b; Yin et al., 2024a). The estimated uncertainties can guide reservoir management decisions within a digital twin framework (Herrmann, 2023), including controlling injectivities (Gahlot et al., 2024a) and designing follow-up monitoring surveys and wells (Orozco et al., 2024b). Additionally, mappings learned from CIGs to the velocity domain can be applied to salt segmentation problems (Muller et al., 2023, 2024).

CONCLUSION

WISER sets the stage for deploying genAI models to facilitate high-dimensional Bayesian inference with computationally intensive forward operators. Deep learning and AI have been criticized for their reliance on realistic training samples, but WISER helps to reduce this dependence 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.

ACKNOWLEDGEMENTS

The authors would like to thank Mathias Louboutin (Devito Codes) for the constructive discussion. The authors would like to thank the associate editor, Anatoly Baumstein, the reviewer, Brent Wheelock, and two anonymous reviewers for the thorough reviews

and valuable suggestions to the original manuscript. This research was carried out with the support of Georgia Research Alliance and partners of the ML4Seismic Center.

**DECLARATION OF GENERATIVE AI AND AI-ASSISTED
TECHNOLOGIES IN THE WRITING PROCESS**

During the preparation of this work, the authors used ChatGPT to improve readability and language. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

- Baldassari, L., A. Siahkoohi, J. Garnier, K. Solna, and M. V. de Hoop, 2024, Conditional score-based diffusion models for bayesian inference in infinite dimensions: Advances in Neural Information Processing Systems, **36**.
- Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.
- Dhabaria, N., and S. C. Singh, 2024, Hamiltonian monte carlo based elastic full-waveform inversion of wide-angle seismic data: Geophysical Journal International, **237**, 1384–1399.
- Erdinc, H. T., R. Orozco, and F. J. Herrmann, 2024, Generative geostatistical modeling from incomplete well and imaged seismic observations with diffusion models: arXiv preprint arXiv:2406.05136.
- Esser, E., L. Guasch, T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, 2018, Total variation regularization strategies in full-waveform inversion: SIAM Journal on Imaging Sciences, **11**, 376–406.
- Fang, Z., C. Da Silva, R. Kuske, and F. J. Herrmann, 2018, Uncertainty quantification for inverse problems with weak partial-differential-equation constraints: Geophysics, **83**, R629–R647.
- Fang, Z., H. Fang, and L. Demanet, 2020, Deep generator priors for bayesian seismic inversion.
- Gahlot, A. P., H. T. Erdinc, R. Orozco, Z. Yin, and F. J. Herrmann, 2023, Inference of co2 flow patterns—a feasibility study: NeurIPS 2023 Workshop - Tackling Climate

Change with Machine Learning.

Gahlot, A. P., H. Li, Z. Yin, R. Orozco, and F. J. Herrmann, 2024a, A digital twin for geological carbon storage with controlled injectivity: arXiv preprint arXiv:2403.19819.

Gahlot, A. P., R. Orozco, Z. Yin, and F. J. Herrmann, 2024b, An uncertainty-aware digital shadow for underground multimodal co2 storage monitoring: arXiv preprint arXiv:2410.01218.

Guo, C., G. Pleiss, Y. Sun, and K. Q. Weinberger, 2017, On calibration of modern neural networks: International conference on machine learning, PMLR, 1321–1330.

Herrmann, F. J., 2023, President’s page: Digital twins in the era of generative ai: The Leading Edge, **42**, 730–732.

Hoffmann, A., R. Brossier, L. Métivier, and A. Tarayoun, 2024, 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, **237**, 1353–1383.

Hou, J., and W. W. Symes, 2016, Accelerating extended least-squares migration with weighted conjugate gradient iteration: Geophysics, **81**, S165–S179.

Izzatullah, M., A. Alali, M. Ravasi, and T. Alkhalifah, 2024, Physics-reliable frugal local uncertainty analysis for full waveform inversion: Geophysical Prospecting.

Jones, C., J. Edgar, J. Selvage, and H. Crook, 2012, Building complex synthetic models to evaluate acquisition geometries and velocity inversion technologies: 74th EAGE Conference and Exhibition incorporating EUROPEC 2012, European Association of Geoscientists & Engineers, cp–293.

Keating, S. D., and K. A. Innanen, 2021, Null-space shuttles for targeted uncertainty

- analysis in full-waveform inversion: *Geophysics*, **86**, R63–R76.
- Kelkar, V. A., S. Bhadra, and M. A. Anastasio, 2021, Compressible latent-space invertible networks for generative model-constrained image reconstruction: *IEEE transactions on computational imaging*, **7**, 209–223.
- Kruse, J., G. Detommaso, U. Köthe, and R. Scheichl, 2021, Hint: Hierarchical invertible neural transport for density estimation and bayesian inference: *Proceedings of the AAAI Conference on Artificial Intelligence*, 8191–8199.
- Liang, Z., F. Wellmann, and O. Ghattas, 2023, Uncertainty quantification of geologic model parameters in 3d gravity inversion by hessian-informed markov chain monte carlo: *Geophysics*, **88**, G1–G18.
- Louboutin, M., Z. Yin, R. Orozco, T. J. Grady, A. Siahkoohi, G. Rizzuti, P. A. Witte, O. Møyner, G. J. Gorman, and F. J. Herrmann, 2023, Learned multiphysics inversion with differentiable programming and machine learning: *The Leading Edge*, **42**, 474–486.
- Marino, J., Y. Yue, and S. Mandt, 2018, Iterative amortized inference: *International Conference on Machine Learning*, PMLR, 3403–3412.
- Minka, T., et al., 2005, Divergence measures and message passing: Technical report, Technical report, Microsoft Research.
- Muller, A. P., C. R. Bom, J. C. Costa, M. Klatt, E. L. Faria, B. d. S. Silva, M. P. de Albuquerque, and M. P. de Albuquerque, 2023, Deep-tomography: iterative velocity model building with deep learning: *Geophysical Journal International*, **232**, 975–989.
- Muller, A. P., B. Fraga, M. Klatt, J. C. Costa, C. R. Bom, E. L. Faria, M. P.

de Albuquerque, and M. P. de Albuquerque, 2024, Deep-salt: Complete three-dimensional salt segmentation from inaccurate migrated subsurface offset gathers using deep learning: Geophysical Prospecting.

Orozco, R., H. T. Erdinc, Y. Zeng, M. Louboutin, and F. J. Herrmann, 2024a, Machine learning-enabled velocity model building with uncertainty quantification: arXiv preprint arXiv:2411.06651.

Orozco, R., A. Gahlot, and F. J. Herrmann, 2024b, Beacon: Bayesian experimental design acceleration with conditional normalizing flows – a case study in optimal monitor well placement for co2 sequestration: arXiv preprint arXiv:2404.00075.

Orozco, R., M. Louboutin, A. Siahkoohi, G. Rizzuti, T. van Leeuwen, and F. J. Herrmann, 2023a, Amortized normalizing flows for transcranial ultrasound with uncertainty quantification: Presented at the Medical Imaging with Deep Learning.

Orozco, R., A. Siahkoohi, M. Louboutin, and F. J. Herrmann, 2024c, Aspire: Iterative amortized posterior inference for bayesian inverse problems: arXiv preprint arXiv:2405.05398.

Orozco, R., A. Siahkoohi, G. Rizzuti, T. van Leeuwen, and F. Herrmann, 2023b, Adjoint operators enable fast and amortized machine learning based bayesian uncertainty quantification: Medical Imaging 2023: Image Processing.

Orozco, R., A. Siahkoohi, G. Rizzuti, T. van Leeuwen, and F. J. Herrmann, 2021, Photoacoustic imaging with conditional priors from normalizing flows: Presented at the NeurIPS 2021 Workshop on Deep Learning and Inverse Problems.

Orozco, R., P. Witte, M. Louboutin, A. Siahkoohi, G. Rizzuti, B. Peters, and F. J. Herrmann, 2024d, Invertiblenetworks.jl: A julia package for scalable normalizing

- flows: *Journal of Open Source Software*, **9**, 6554.
- Peters, B., B. R. Smithyman, and F. J. Herrmann, 2019, Projection methods and applications for seismic nonlinear inverse problems with multiple constraints: *Geophysics*, **84**, R251–R269.
- Radev, S. T., U. K. Mertens, A. Voss, L. Ardizzone, and U. Köthe, 2020, Bayesflow: Learning complex stochastic models with invertible neural networks: *IEEE transactions on neural networks and learning systems*, **33**, 1452–1466.
- Rezende, D., and S. Mohamed, 2015, Variational inference with normalizing flows: *International conference on machine learning*, PMLR, 1530–1538.
- Rizzuti, G., and I. Vasconcelos, 2024, Multiscale uncertainty quantification for post-stack seismic inversion with wavelet flows: 85th EAGE Annual Conference & Exhibition (including the Workshop Programme), European Association of Geoscientists & Engineers, 1–5.
- Rubio, P.-B., Y. Marzouk, and M. Parno, 2023, A transport approach to sequential simulation-based inference: *arXiv preprint arXiv:2308.13940*.
- Sheng, H., X. Wu, X. Si, J. Li, S. Zhang, and X. Duan, 2023, Seismic foundation model (sfm): a new generation deep learning model in geophysics: *arXiv preprint arXiv:2309.02791*.
- Siahkoohi, A., G. Rizzuti, and F. J. Herrmann, 2020, Weak deep priors for seismic imaging, *in* SEG Technical Program Expanded Abstracts 2020: Society of Exploration Geophysicists, 2998–3002.
- Siahkoohi, A., G. Rizzuti, M. Louboutin, P. Witte, and F. Herrmann, 2021, Preconditioned training of normalizing flows for variational inference in inverse problems:

Presented at the Third Symposium on Advances in Approximate Bayesian Inference.

Siahkoohi, A., G. Rizzuti, R. Orozco, and F. J. Herrmann, 2023, Reliable amortized variational inference with physics-based latent distribution correction: *Geophysics*, **88**, R297–R322.

Song, Y., C. Durkan, I. Murray, and S. Ermon, 2021, Maximum likelihood training of score-based diffusion models: *Advances in neural information processing systems*, **34**, 1415–1428.

Sun, C., A. Malcolm, R. Kumar, and W. Mao, 2024, Enabling uncertainty quantification in a standard full waveform inversion method using normalizing flows: *Geophysics*, **89**, 1–69.

Taufik, M. H., F. Wang, and T. Alkhalifah, 2024, Learned regularizations for multi-parameter elastic full waveform inversion using diffusion models: *Journal of Geophysical Research: Machine Learning and Computation*, **1**, e2024JH000125.

ten Kroode, F., 2023, An omnidirectional seismic image extension: *Inverse Problems*, **39**, 035003.

Vaitl, L., L. Winkler, L. Richter, and P. Kessel, 2024, Fast and unified path gradient estimators for normalizing flows: Presented at the The Twelfth International Conference on Learning Representations.

Wang, F., X. Huang, and T. A. Alkhalifah, 2023, A prior regularized full waveform inversion using generative diffusion models: *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 1–11.

Wei, X., J. Sun, and M. Sen, 2024a, 3d monte carlo geometry inversion using gravity data: *Geophysics*, **89**, 1–62.

- Wei, X., J. Sun, and M. K. Sen, 2023, Quantifying uncertainty of salt body shapes recovered from gravity data using trans-dimensional markov chain monte carlo sampling: *Geophysical Journal International*, **232**, 1957–1978.
- , 2024b, Reconstruction of multiple target bodies using trans-dimensional bayesian inversion with different constraints: *IEEE Transactions on Geoscience and Remote Sensing*.
- Winkler, C., D. Worrall, E. Hoozeboom, and M. Welling, 2019, Learning likelihoods with conditional normalizing flows: *arXiv preprint arXiv:1912.00042*.
- Yin, Z., M. Louboutin, O. Møyner, and F. J. Herrmann, 2024a, Time-lapse full-waveform permeability inversion: A feasibility study: *The Leading Edge*, **43**, 544–553.
- Yin, Z., R. Orozco, M. Louboutin, and F. J. Herrmann, 2023, Solving multiphysics-based inverse problems with learned surrogates and constraints: *Advanced Modeling and Simulation in Engineering Sciences*, **10**, 14.
- , 2024b, Wise: full-waveform variational inference via subsurface extensions: *Geophysics*, **89**, 1–31.
- Zhang, X., and A. Curtis, 2024, Bayesian variational time-lapse full waveform inversion: *Geophysical Journal International*, **237**, 1624–1638.
- Zhang, X., A. Lomas, M. Zhou, Y. Zheng, and A. Curtis, 2023, 3-d bayesian variational full waveform inversion: *Geophysical Journal International*, **234**, 546–561.
- Zhang, X., M. A. Nawaz, X. Zhao, and A. Curtis, 2021, An introduction to variational inference in geophysical inverse problems.
- Zhao, X., A. Curtis, and X. Zhang, 2022, Bayesian seismic tomography using normal-

izing flows: Geophysical Journal International, **228**, 213–239.

LIST OF FIGURES

- 1 Out-of-distribution example. (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 1(c)), perturbed velocity model (Figure 1(d)), posterior samples of WISE, and WISER, shown in red, blue, yellow and green color, respectively. (c) An unseen in-distribution velocity model. (d) Unseen ground-truth velocity model. (e) 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. (g) Imaged reflectivity samples from WISE. (f)(h) are the same as (e)(g) but for WISER, showcasing significant improvements. 29

LIST OF TABLES

1 Number of PDE solves for several Bayesian inference methods for FWI.
2 N represents number of training pairs. n_{OBN} represents number of
3 OBNs in the seismic survey. datapass_1 and datapass_2 represent number
4 of datapasses through the entire OBN dataset using WISER and a
5 non-amortized inference method, such as the ones presented in Zhang
6 et al. (2023). 31

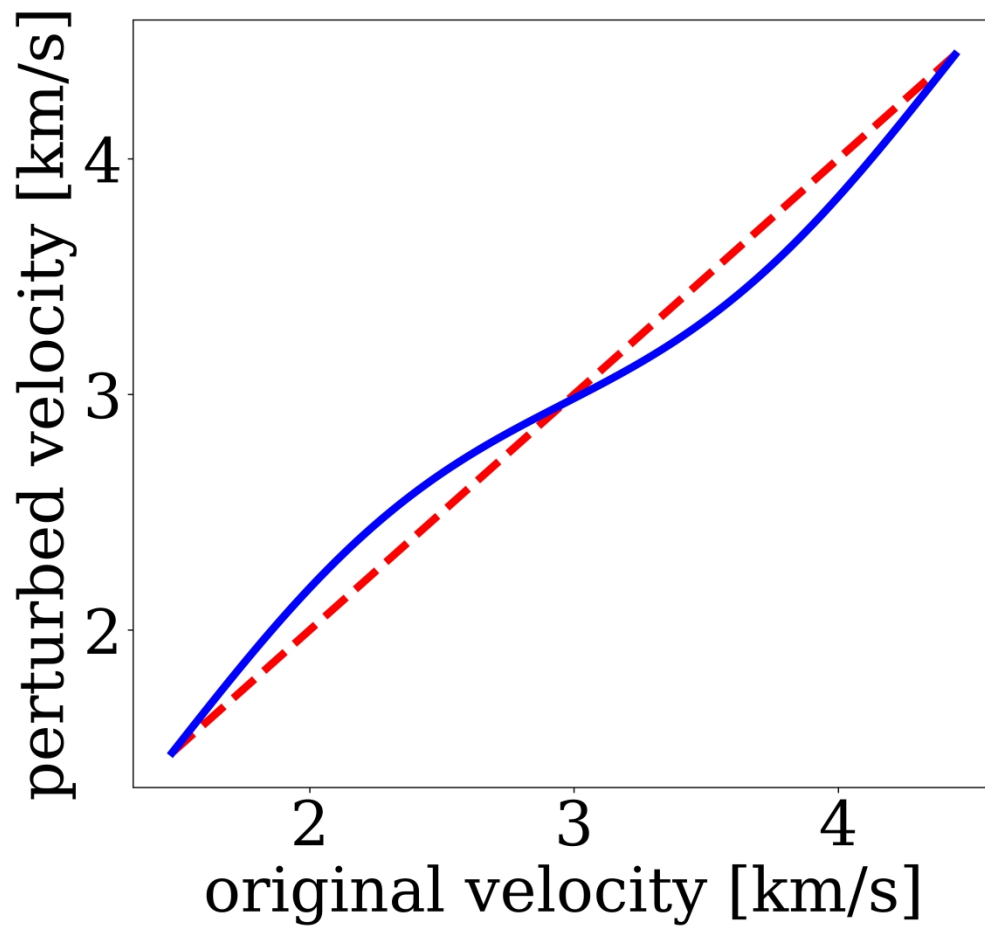


Figure 1a: Curves for velocity-value perturbations.

581x538mm (236 x 236 DPI)

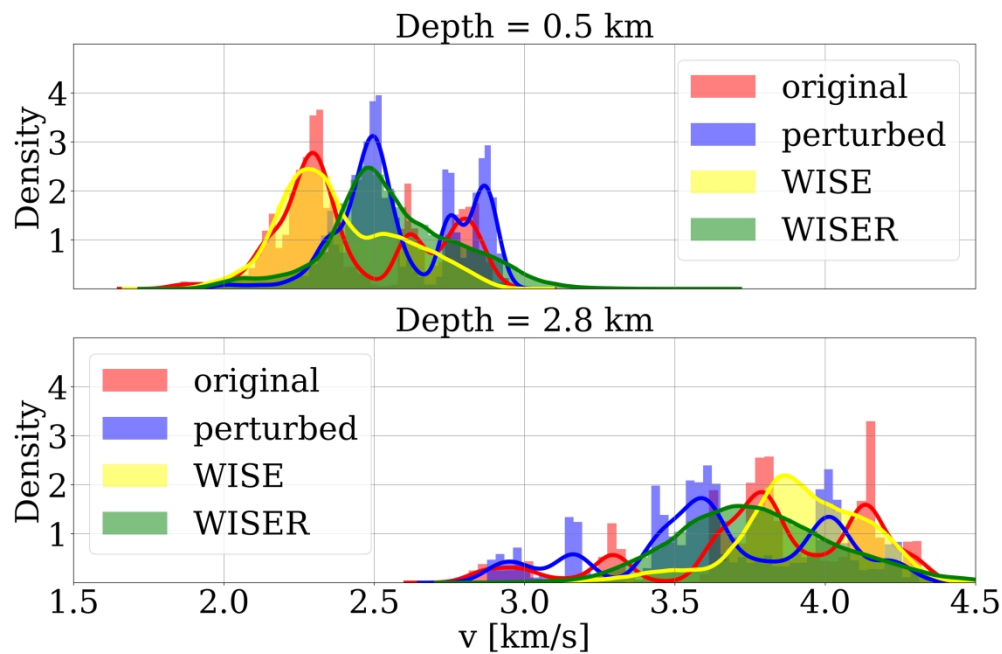


Figure 1b: Histograms of values at the depth of 0.5 km and 2.8 km in the original velocity model, perturbed velocity model, posterior samples of WISE, and WISER, shown in red, blue, yellow and green color, respectively.

148x96mm (600 x 600 DPI)

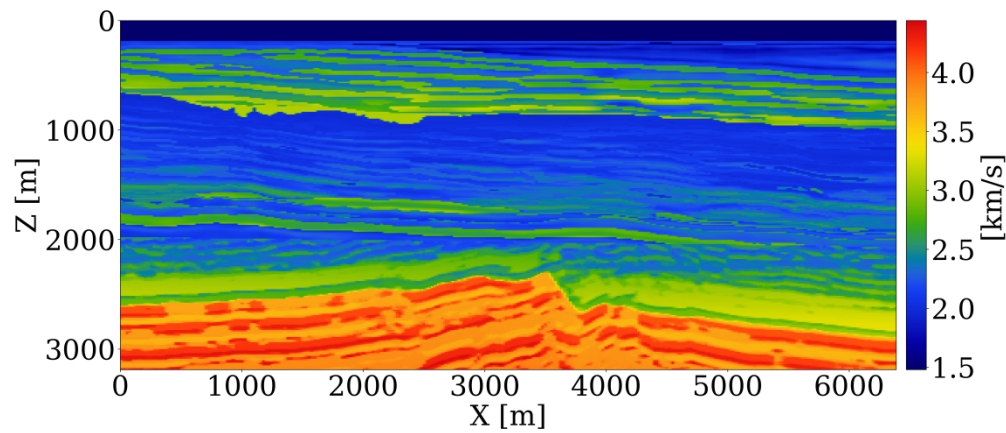


Figure 1c: An unseen in-distribution velocity model.

1290x556mm (118 x 118 DPI)

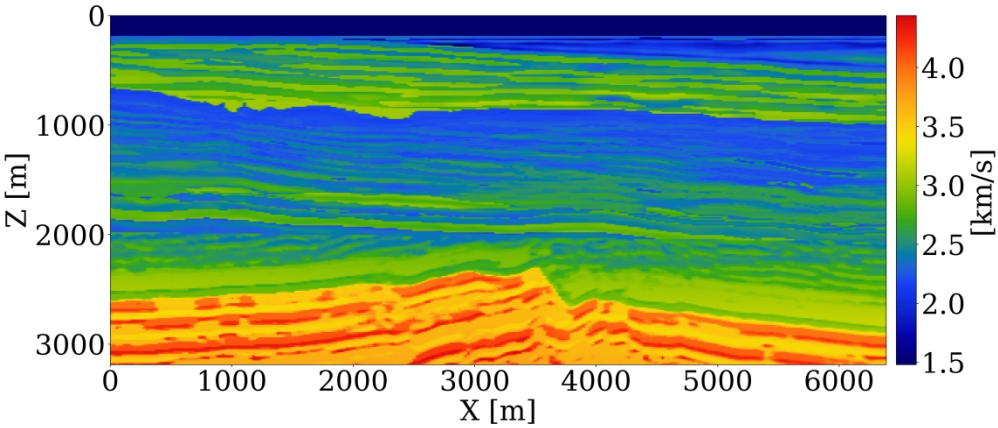


Figure 1d: Unseen ground-truth velocity model.

1290x556mm (118 x 118 DPI)

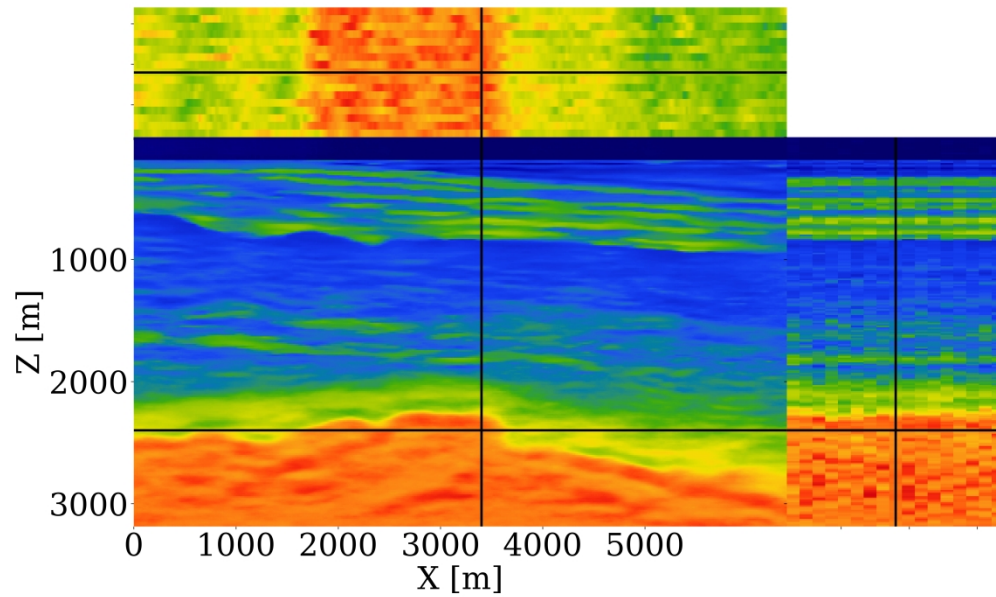


Figure 1e: Estimated velocity models from WISE. The conditional mean estimate is shown in the center. For posterior samples, horizontal traces at $Z = 2.7$ km and vertical traces at $X = 3.6$ km are displayed on the top and on the right, respectively.

1149x688mm (118 x 118 DPI)

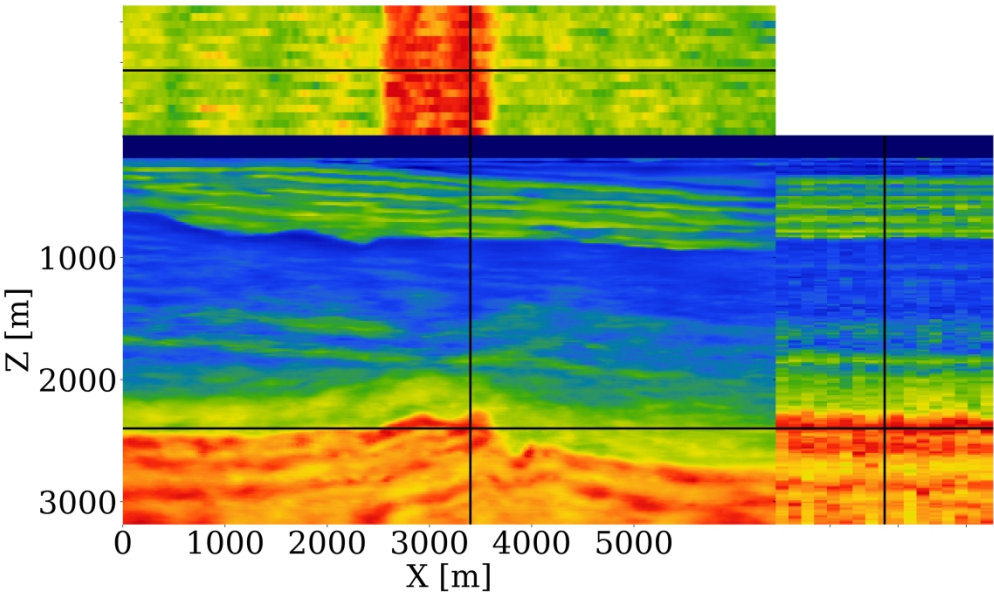


Figure 1f: The same as Figure 1e but for WISER.

148x88mm (600 x 600 DPI)

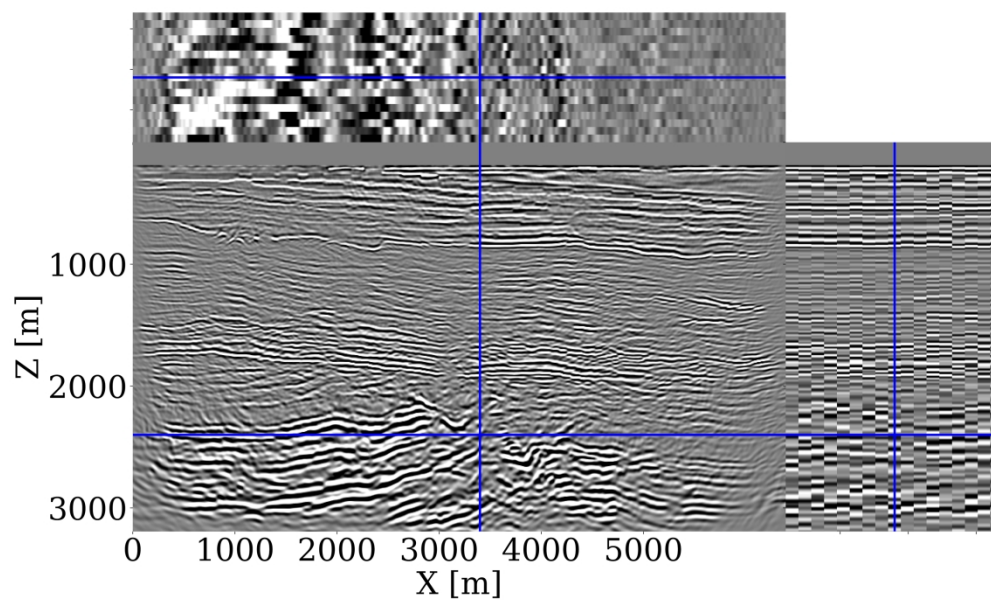


Figure 1g: Imaged reflectivity samples from WISE.

148x88mm (600 x 600 DPI)

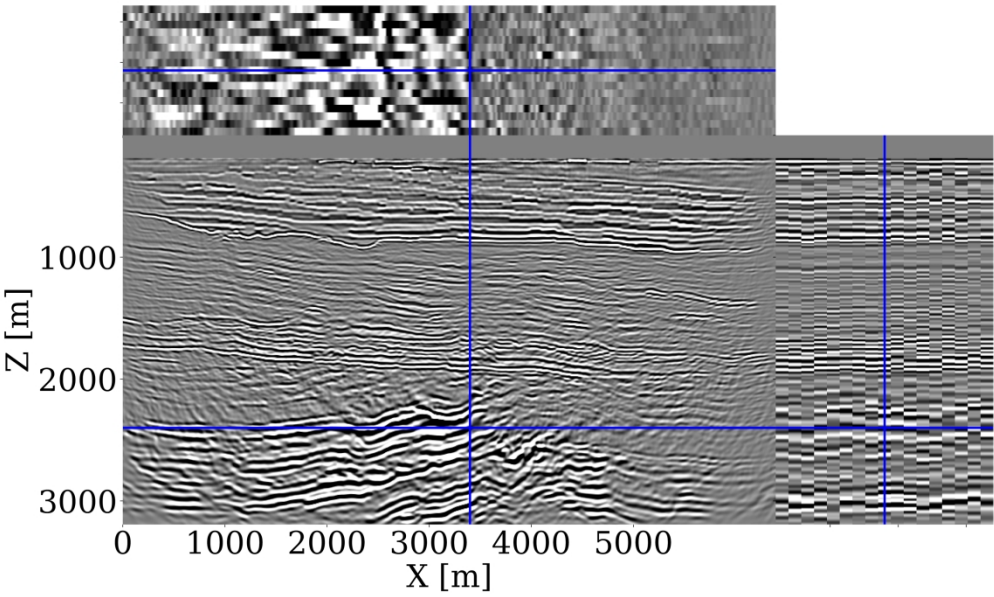


Figure 1h: Imaged reflectivity samples from WISER.

148x88mm (600 x 600 DPI)

Method	Offline PDE solves $\times n_{\text{OBN}}$	Online PDE solves $\times n_{\text{OBN}}$
WISE (Yin et al., 2024b)	$3 \times N$	2
WISER (proposed)	$3 \times N$	$2 + 2 \times \text{datapass}_1$
Non-amortized inference	0	$2 \times \text{datapass}_2$

Table 1: Number of PDE solves for several Bayesian inference methods for FWI. N represents number of training pairs. n_{OBN} represents number of OBNs in the seismic survey. datapass_1 and datapass_2 represent number of datapasses through the entire OBN dataset using WISER and a non-amortized inference method, such as the ones presented in Zhang et al. (2023).

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be accessed via the following URL:
Note: A digital object identifier (DOI) linking to the data in a general or discipline-specific data repository is strongly preferred.

Downloaded 12/06/24 to 143.215.16.69. Redistribution subject to SEG license or copyright; see Terms of Use at <http://library.seg.org/page/policies/terms>
DOI: 10.1190/geo2024-0483.1