

# Full-waveform variational inference with full common-image gathers and diffusion network

Yunlin Zeng, Huseyin Tuna Erdinc, Rafael Orozco, and Felix J. Herrmann

Georgia Institute of Technology

## SUMMARY

Accurate seismic imaging and velocity estimation are essential for subsurface characterization. Conventional inversion techniques, such as full-waveform inversion, remain computationally expensive and sensitive to initial velocity models. To address these challenges, we propose a simulation-based inference framework with conditional elucidated diffusion models for posterior velocity-model sampling. Our approach incorporates both horizontal and vertical subsurface offset common-image gathers to capture a broader range of reflector geometries, including gently dipping structures and steep dipping layers. Additionally, we introduce the background-velocity model as an input condition to enhance generalization across varying geological settings. We evaluate our method on the SEAM dataset, which features complex salt geometries, using a patch-based training approach. Experimental results demonstrate that adding the background-velocity model as an additional conditioning variable significantly enhances performance, improving SSIM from 0.717 to 0.733 and reducing RMSE from 0.381 km/s to 0.274 km/s. Furthermore, uncertainty quantification analysis shows that our proposed approach yields better-calibrated uncertainty estimates, reducing uncertainty calibration error from 6.68 km/s to 3.91 km/s. These results show robust amortized seismic inversion with uncertainty quantification.

## INTRODUCTION

Accurate seismic imaging and reliable velocity-model estimation are critical for understanding subsurface geological structures. Conventional full-waveform inversion (FWI) is a powerful method for subsurface imaging, but it often suffers from strong non-linearity, high computational costs, and sensitivity to the initial velocity model (Virieux and Operto, 2009; Warner et al., 2013). Recently, simulation-based inference (SBI) methods (Cranmer et al., 2020)—which leverage pairs of simulated seismic observations and their corresponding ground-truth velocity models  $\{\mathbf{x}^{(i)}, \mathbf{y}^{(i)}\}_{i=1}^N$ —have emerged as promising alternatives to traditional inversion techniques. These methods utilize machine learning models to amortize the inversion process, enabling efficient inference of subsurface velocity from new observational data, while also providing systematic Bayesian uncertainty quantification (Zhang and Curtis, 2023).

One notable SBI framework is Wavefield Inversion using conditional normalizing flows (WISE), which samples posterior velocity models conditioned on seismic common-image gathers (CIGs) (Yin et al., 2023). By leveraging CIGs, the WISE framework efficiently extracts physics-informed summary statistics from seismic shot data and enables inversion with uncertainty quantification. However, horizontal CIGs alone may be

insufficient for imaging complex geological settings—particularly those involving steeply dipping or vertical reflectors—due to geometric limitations associated with horizontal offsets (Shan and Biondi, 2008; van Leeuwen et al., 2015).

We propose an SBI approach that incorporates both horizontal and vertical subsurface offset image gathers to include full information of migrated wavefields. The addition of vertical offsets extends the imaging condition to better capture steeply dipping reflectors by correlating wavefields at vertically shifted depths (Biondi, 2006; Sava and Vasconcelos, 2011; van Leeuwen et al., 2015). Furthermore, instead of using conditional normalizing flows as in Yin et al. (2023), we adopt score-based diffusion generative networks. Recent studies have demonstrated that score-based methods (Ho et al., 2020; Song et al., 2020; Karras et al., 2022) offer superior sample quality compared to CNFs, particularly in capturing multimodal distributions and generating realistic posterior samples for complex seismic inverse problems (Orozco et al., 2024).

In this study, we extend the SBI framework by integrating both horizontal and vertical CIGs into a score-based diffusion generative network (Karras et al., 2022). Additionally, we propose using the background-velocity model as an extra conditioning variable alongside the CIGs. We argue that including the background model not only broadens the amortization capacity of the network—enabling generalization across diverse geological settings—but also improves sample quality (Elsemlüller et al., 2024). To validate our approach, we present inference results on the SEAM dataset, which features complex salt dome structures (Fehler and Keliher, 2011).

## METHOD

Our proposed method employs a two-stage process involving: (1) generation of extended CIGs (both horizontal and vertical), and (2) posterior velocity-model inference using the score-based network conditioned on these extended gathers.

### Extended Imaging Conditions: Horizontal and Vertical CIGs

Seismic imaging typically involves correlating the forward-propagated source wavefield  $\mathbf{u}$  and backward-propagated receiver wavefield  $\mathbf{v}$ . The zero-offset imaging condition, while common, does not capture velocity-model inaccuracies or steeply dipping reflectors effectively (Nolet, 1986). Extended imaging conditions address this by introducing non-zero spatial offsets into the wavefield correlation, resulting in extended-image gathers that directly encode velocity-related information. When the velocity model is correct, energy focuses at zero offset; deviations from zero indicate velocity errors (Sava and Fomel, 2005; Shan and Biondi, 2008; Biondi and Symes, 2004).

Horizontal CIGs correlate wavefields horizontally displaced

around the imaging point and are defined as:

$$I_x(x, z, h_x) = \iint \mathbf{u}(s_x, x - h_x, z, t) \mathbf{v}(s_x, x + h_x, z, t) dt ds_x, \quad (1)$$

where  $h_x$  is horizontal offset,  $t$  is propagation time, and  $s_x$  is the horizontal source location (Biondi, 2006). Horizontal offsets capture reflection curvature indicative of velocity inaccuracies and work best for reflectors with gentle to moderate dips (Rickett and Sava, 2002).

In contrast, vertical CIGs correlate wavefields at vertically displaced imaging points:

$$I_z(x, z, h_z) = \iint \mathbf{u}(s_x, x, z - h_z, t) \mathbf{v}(s_x, x, z + h_z, t) dt ds_x, \quad (2)$$

where  $h_z$  is the vertical offset (Shan and Biondi, 2008). Vertical subsurface offsets work best at imaging steeply dipping or vertical reflectors, which require impractically large horizontal offsets if only horizontal CIGs are used (Biondi and Shan, 2005). By combining both horizontal and vertical offsets, our method generates comprehensive subsurface images capable of accurately representing complex geological structures and robustly informing velocity updates.

### Conditional Diffusion for Seismic Inference

Diffusion generative networks are neural density estimation algorithms that approximate the data distribution by learning the score function, defined as the gradient of the log-likelihood  $\nabla_{\mathbf{x}} \log p(\mathbf{x})$  (Song et al., 2020). These networks operate by progressively perturbing data samples into a Gaussian distribution through a stochastic process and then learning to reverse this transformation to generate new samples from the target distribution.

A key component in diffusion networks is the noise schedule, denoted by  $\sigma(t)$ , which governs the gradual corruption of data over time. As the time variable  $t$  increases, the data distribution becomes increasingly smoothed (the PDF becomes more noisy), eventually converging to a standard Gaussian. The network is trained to estimate the score function  $\nabla_{\mathbf{x}} \log p(\mathbf{x}, \sigma(t))$  at each time step, enabling it to reconstruct the original distribution during inference.

Once trained, it can perform likelihood estimation and generate new samples by numerically solving a reverse-time stochastic differential equation. In our seismic inversion work, we employ elucidated diffusion models (EDM), which provide a well-conditioned and computationally efficient framework for both training and sampling (Karras et al., 2022). EDM simplify the standard score-matching objective and introduce an optimized noise schedule, leading to improved sample quality and reduced computational cost.

### Conditional Diffusion for Seismic Imaging

In seismic inversion, we are interested in sampling from the posterior distribution  $p(\mathbf{x} | \mathbf{y})$ , where  $\mathbf{x}$  represents the velocity model and  $\mathbf{y}$  denotes the CIGs obtained from migrated seismic shot data. To achieve this, we modify the standard diffusion framework to learn the conditional score function  $\nabla_{\mathbf{x}} \log p(\mathbf{x} | \mathbf{y})$ .

Following prior work on conditional diffusion (Batzolis et al., 2021; Orozco et al., 2024), we incorporate the observation  $\mathbf{y}$  as an additional conditioning variable in the score-matching objective:

$$\hat{\theta} = \operatorname{argmin}_{\theta} \mathbb{E}_{\mathbf{y} \sim p(\mathbf{y} | \mathbf{x})} \mathbb{E}_{\mathbf{n} \sim \mathcal{N}(0, t^2 I)} \|s_{\theta}(\mathbf{x} + \mathbf{n}, \mathbf{y}, t) - \mathbf{x}\|_2^2, \quad (3)$$

where  $s_{\theta}(\mathbf{x}, \mathbf{y}, t)$  represents the learned score function, and  $\mathbf{n} \sim \mathcal{N}(0, t^2 I)$  denotes the Gaussian noise perturbation.

To train the model, we construct a dataset  $\mathcal{D} = \{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=0}^N$ , where each sample pair consists of a ground-truth velocity model and the corresponding CIGs obtained through seismic migration. The training process involves minimizing the following loss:

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{i=0}^N \mathbb{E}_{\mathbf{n} \sim \mathcal{N}(0, t^2 I)} \|s_{\theta}(\mathbf{x}^{(i)} + \mathbf{n}, \mathbf{y}^{(i)}, t) - \mathbf{x}^{(i)}\|_2^2. \quad (4)$$

After training, posterior samples can be generated by conditioning the model on new observations  $\mathbf{y}_{\text{obs}}$  and solving the reverse SDE, analogous to standard unconditional diffusion models.

### Incorporating Vertical and Horizontal CIGs into Diffusion Model

Our approach conditions the diffusion model on both horizontal and vertical offset CIGs to enhance posterior velocity sampling, which introduces depth shifts in wavefield correlations and captures reflector geometries that horizontal offsets alone cannot resolve (Shan and Biondi, 2008; van Leeuwen et al., 2015).

In our model, we use 50 channels of horizontal offsets and 50 channels of vertical offsets. In addition, we include a background-velocity model as a separate channel, which serves as a reference condition. Including the background as an additional input channel expands the amortization scope of our network and generalizes across different velocity backgrounds during inference.

## EXPERIMENT SETUP

We conduct experiments using 2D slices extracted from the 3D SEAM model (Fehler and Keliher, 2011), which features complex salt geometries characteristic of the Gulf of Mexico. To create a robust training dataset while ensuring diversity, we select a continuous range of crosslines for training and enforce a minimum separation of 2 km between the nearest training and test slices to reduce redundancy.

For seismic data simulation, we generate narrow-offset 2D seismic lines with a spatial grid of  $1744 \times 512$  and a grid spacing of 20 m. The shot interval is set to 1000 m to optimize computational efficiency, while receivers are densely sampled at 20 m intervals with a maximum offset of 6 km. The recorded shot gathers have a total duration of 9 s, and Gaussian noise is added to achieve a signal-to-noise ratio of 25 dB within the source frequency band. Each velocity model contains 32 shot

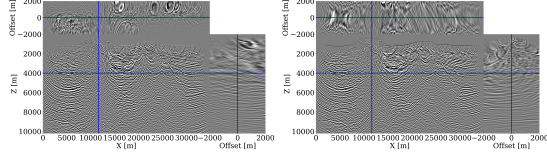


Figure 1: Common-image gather (CIGs) plots. *left*: Horizontal subsurface offset CIGs. *right*: Vertical subsurface offset CIGs.

records and assumes a constant density with spatially varying velocity.

To create initial migration-velocity models, we preprocess the ground-truth velocity models by removing salt bodies and smoothing the result using a Gaussian kernel with a grid size of 25. The subsurface-offset migration is performed with 50 evenly distributed offsets between  $-2000$  m and  $2000$  m. After migration, subsets of the full seismic lines are extracted to form training pairs, using sliding windows of  $512 \times 512$  grids, resulting in 770 training samples.

We adopt a patch-based training strategy for our conditional diffusion model (Orozco et al., 2024), training on smaller velocity patches while performing inference on full-grid resolution ( $512 \times 1744$ ). This is made possible by the convolutional architecture of the network, which enables flexible evaluation on larger spatial domains without requiring retraining. As a result, the model scales effectively to high-resolution domains. Training each model requires approximately 12 GPU hours per iteration.

## RESULTS

### Quality of Posterior Samples

Using EDM conditioned on both horizontal (Fig. 1 left) and vertical (Fig. 1 right) CIGs migrated from the background in Fig. 3, we further evaluate the impact of incorporating a background-velocity model as an additional input channel. Compared to the baseline model that conditions only on horizontal and vertical CIGs (Fig. 4 and Fig. 6), the inclusion of the background-velocity model significantly improves reconstruction quality (Fig. 5 and Fig. 7). Specifically, the structural similarity index measure (SSIM) increases from 0.717 to 0.733, and the root-mean-square error (RMSE) decreases from 0.381 km/s to 0.274 km/s. These metrics are computed by comparing the posterior mean velocity models to the ground truth velocity model shown in Fig. 2.

The improvement in SSIM and RMSE shows the benefit of using background-velocity information to help the network’s ability to generalize across different subsurface conditions and resolve velocity ambiguities in seismic inversion. The additional background input leads to sharper geological layer reconstructions, particularly in the deeper regions beneath the salt structure, and the bottom left part of the salt.

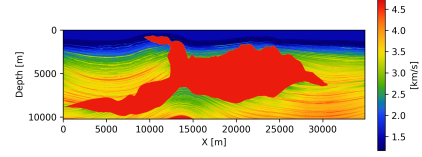


Figure 2: One ground truth velocity model in the test set.

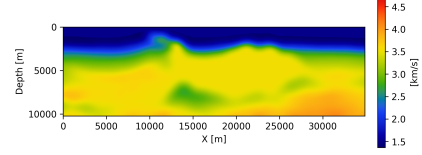


Figure 3: The migration-velocity model.

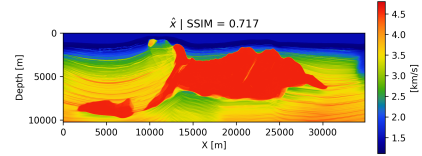


Figure 4: Posterior mean velocity model inferred using horizontal and vertical CIGs without incorporating the background as an additional input.

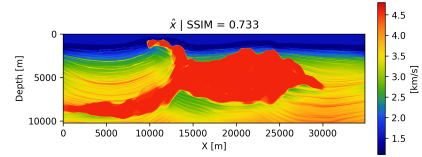


Figure 5: Posterior mean velocity model inferred using horizontal and vertical CIGs with incorporating the background as an additional input.

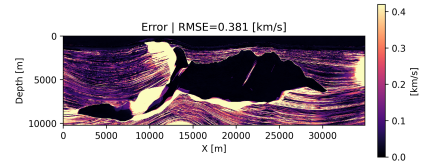


Figure 6: Absolute error between the posterior mean in Fig. 4 and the ground truth velocity model.

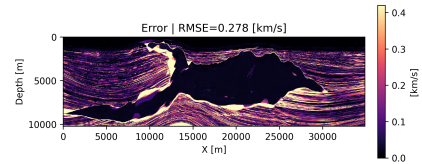


Figure 7: Absolute error between the posterior mean in Fig. 5 and the ground truth velocity model.

## UNCERTAINTY QUANTIFICATION

Beyond sample quality, accurately quantifying uncertainty is critical for reliable seismic inversion. We use calibration of uncertainty to evaluate the quality of uncertainty quantification.

### Calibration of Uncertainty

The calibration of uncertainty evaluates how well the predicted uncertainties align with observed errors (Guo et al., 2017; Laves et al., 2020). Ideally, the predicted uncertainty ( $\hat{\sigma}$ ) should match the observed error magnitude exactly. Formally, perfect calibration satisfies:

$$\mathbb{E}_{\mathbf{x}_{\text{gt}}, \mathbf{y}} [|\mathbf{x}_{\text{gt}} - \hat{\mathbf{x}}| \mid \hat{\sigma} = \sigma] = \sigma, \quad \forall \sigma \geq 0. \quad (5)$$

To quantify calibration, predicted uncertainties ( $\hat{\sigma}$ ) and observed errors ( $|\mathbf{x}_{\text{gt}} - \hat{\mathbf{x}}|$ ) are binned based on uncertainty magnitude. For each bin  $B_k$ , we compute average uncertainty in bin  $B_k$

$$\text{UQ}(B_k) = \frac{1}{|B_k|} \sum_{i \in B_k} \hat{\sigma}_i, \quad (6)$$

and average error in bin  $B_k$

$$\text{Error}(B_k) = \frac{1}{|B_k|} \sum_{i \in B_k} |\mathbf{x}_{\text{gt},i} - \hat{\mathbf{x}}_i|. \quad (7)$$

The averages of uncertainty and error in each bin are plotted against each other, with ideal calibration represented by points falling on the 45° line. The deviation from this line is quantified by the Uncertainty Calibration Error (UCE), defined as the area between the calibration curve and the 45° line:

$$\text{UCE} = \sum_k |B_k| \cdot |\text{UQ}(B_k) - \text{Error}(B_k)|. \quad (8)$$

Ideally, uncertainty calibration curves should align with the 45° line for a perfect agreement. In Fig. 8, we compare the calibration performance of diffusion models conditioned on different CIG with and without background input. Both curves exhibit a positive correlation between predicted uncertainty and actual error. However, incorporating the background as an additional input improves calibration, as seen in the orange curve aligning more closely with the optimal line than the baseline (blue curve). To quantify this, we compute the uncertainty calibration error (UCE), which measures the deviation between predicted uncertainty and observed error across all bins. Adding the background channel reduces UCE from 6.68 km/s to 3.91 km/s.

## CONCLUSION AND DISCUSSION

In this work, we introduce a conditional score-based generative model for seismic velocity inversion, using the EDM

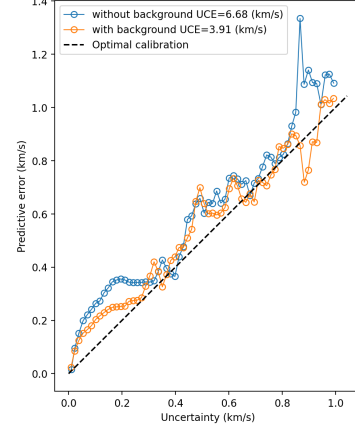


Figure 8: Comparison of uncertainty calibration between the diffusion network without (blue) and with (orange) a background channel. The orange curve aligns more closely with the optimal 45° calibration line, and the UCE is reduced.

network. By incorporating horizontal and vertical CIGs, our method achieves decent inversion performance. Additionally, we demonstrate that including the background-velocity model as a conditioning variable significantly improves inference performance. Specifically, the SSIM improves from 0.717 to 0.733, the RMSE decreases from 0.381 km/s to 0.274 km/s, and the UCE is reduced from 6.68 km/s to 3.91 km/s, indicating improved uncertainty estimation reliability.

Future work will explore extending this framework to 3D seismic inversion. Additionally, we aim to further enhance the diffusion network by training separate networks conditioned individually on horizontal and vertical CIGs and subsequently aligning these representations in the latent space. This latent-space alignment, inspired by twin-network and multi-view learning approaches (Chen et al., 2020; Tian et al., 2020), has the potential to more effectively extract information from both types of gathers. Overall, our findings demonstrate that conditional score-based generative models, when properly conditioned on extended imaging gathers and background velocity, offer a powerful alternative to traditional seismic inversion methods.

## ACKNOWLEDGMENT

This research was carried out with the support of Georgia Research Alliance and partners of the ML4Seismic Center.

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

## REFERENCES

- Batzolis, G., J. Stanczuk, C.-B. Schönlieb, and C. Etmann, 2021, Conditional image generation with score-based diffusion models.
- Biondi, B., and G. Shan, 2005, *in* Prestack imaging of over-turned reflections by reverse time migration: 1284–1287.
- Biondi, B., and W. W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: *GEOPHYSICS*, **69**, 1283–1298.
- Biondi, B. L., 2006, 3d seismic imaging: Society of Exploration Geophysicists.
- Chen, T., S. Kornblith, M. Norouzi, and G. Hinton, 2020, A simple framework for contrastive learning of visual representations.
- Cranmer, K., J. Brehmer, and G. Louppe, 2020, The frontier of simulation-based inference: Proceedings of the National Academy of Sciences, **117**, 30055–30062.
- Else Müller, L., H. Olischläger, M. Schmitt, P.-C. Bürkner, U. Koethe, and S. T. Radev, 2024, Sensitivity-aware amortized bayesian inference: Transactions on Machine Learning Research.
- Fehler, M., and P. J. Keliher, 2011, Model development, *in* SEAM Phase 1: Challenges of Subsalt Imaging in Tertiary Basins, with Emphasis on Deepwater Gulf of Mexico: Society of Exploration Geophysicists, 15–50.
- Guo, C., G. Pleiss, Y. Sun, and K. Q. Weinberger, 2017, On calibration of modern neural networks.
- Ho, J., A. Jain, and P. Abbeel, 2020, Denoising diffusion probabilistic models: Advances in neural information processing systems, **33**, 6840–6851.
- Karras, T., M. Aittala, T. Aila, and S. Laine, 2022, Elucidating the design space of diffusion-based generative models.
- Laves, M.-H., S. Ihler, J. F. Fast, L. A. Kahrs, and T. Ortmaier, 2020, Well-calibrated regression uncertainty in medical imaging with deep learning: Proceedings of the Third Conference on Medical Imaging with Deep Learning, PMLR, 393–412.
- Nolet, G., 1986, Imaging the earth’s interior, j. f. claerbout, blackwell scientific publications, oxford, 1985 414pp., £42.00 and fundamentals of geophysical data processing, 2nd edn j. f. claerbout, blackwell scientific publications, oxford, 1985 274 pp., £42.00: Geophysical Journal International, **86**, 217–219.
- Orozco, R., H. T. Erdinc, Y. Zeng, M. Louboutin, and F. J. Herrmann, 2024, Machine learning-enabled velocity model building with uncertainty quantification.
- Rickett, J., and P. Sava, 2002, Offset and angle-domain common image-point gathers for shot-profile migration: Geophysics, **67**, 883–889.
- Sava, P., and S. Fomel, 2005, *in* Time-shift imaging condition: 1850–1853.
- Sava, P., and I. Vasconcelos, 2011, Extended imaging conditions for wave-equation migration: Geophysical Prospecting, **59**, 35–55.
- Shan, G., and B. Biondi, 2008, *in* Angle-domain common-image gathers for steep reflectors: 3068–3072.
- Song, Y., J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, 2020, Score-based generative modeling through stochastic differential equations: arXiv preprint arXiv:2011.13456.
- Tian, Y., D. Krishnan, and P. Isola, 2020, Contrastive multi-view coding.
- van Leeuwen, T., R. Kumar, and F. J. Herrmann, 2015, Affordable full subsurface image volume—an application to wemva: Presented at the EAGE Annual Conference Proceedings. ((EAGE Workshop on Wave Equation based Migration Velocity Analysis, Madrid)).
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *GEOPHYSICS*, **74**, WCC1–WCC26.
- Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Štekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3d full-waveform inversion: *GEOPHYSICS*, **78**, R59–R80.
- Yin, Z., R. Orozco, M. Louboutin, and F. J. Herrmann, 2023, Wise: full-waveform variational inference via subsurface extensions.
- Zhang, X., and A. Curtis, 2023, Bayesian variational time-lapse full-waveform inversion: arXiv preprint arXiv:2308.08805.