

Bayesian Joint Velocity and Impedance Inversion via Diffusion Models Conditioned on Common Image Gatherers

Yunlin Zeng, Huseyin Tuna Erdinc, Felix J. Herrmann

SUMMARY

We present a simulation-based inference framework for joint Bayesian recovery of subsurface velocity and acoustic impedance from seismic data. A score-based diffusion model is conditioned on two complementary Common Image Gatherers (CIGs): an inverse-scattering CIG encoding reflectivity amplitude and an anti-ISIC CIG encoding kinematic velocity errors. The model simultaneously samples the posterior distributions of both parameters. Training labels are deliberately decoupled to prevent the model from exploiting the Gardner coupling: velocity targets are lightly smoothed to match the long-wavelength content of the anti-ISIC CIG, while impedance targets retain the unsmoothed ground truth. On the 1000-model Compass benchmark the model achieves velocity SSIM 0.967 (RMSE 0.050 km/s) and impedance SSIM 0.867 (RMSE 0.279 km/s · g/cm³), with velocity quality confirmed by CIG focusing.

INTRODUCTION

Recovering subsurface P-wave velocity and acoustic impedance simultaneously from seismic data is central to exploration geophysics. Velocity controls wave propagation kinematics and is the primary target of conventional Full Waveform Inversion (FWI) (Virieux and Operto, 2009), while acoustic impedance $Z = \rho v$ encodes rock and fluid properties through amplitude-variation-with-offset (AVO) responses (Avseth et al., 2010) and is the primary target of multi-parameter seismic inversion. Obtaining both parameters jointly, together with calibrated uncertainty estimates, remains a challenge. Conventional FWI yields a single point estimate without uncertainty. Simulation-based inference (SBI) (Cranmer et al., 2020) addresses this by training a generative surrogate on paired physics simulations to enable fast posterior sampling. Score-based diffusion models are particularly effective for SBI due to their expressiveness in capturing non-Gaussian posteriors (Karras et al., 2022; Wu et al., 2024). The WISE framework (Yin et al., 2024) demonstrated accurate velocity posteriors by conditioning an Elucidated Diffusion Model (EDM) on subsurface offset Common Image Gatherers (CIGs); deep learning approaches have also used CIG volumes for velocity-model building with uncertainty quantification (Geng et al., 2022; Siahkoobi et al., 2022; Orozco et al., 2024). Our prior work extended this to joint vertical and horizontal CIGs (Zeng et al., 2025).

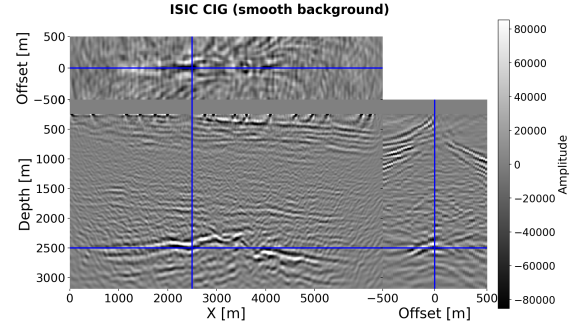


Figure 1: ISIC CIG for a held-out test model from the Compass benchmark. Top panel: horizontal slice at fixed depth (offset vs. X); bottom-left: zero-offset RTM image; bottom-right: gather at $x = 2500$ m (depth vs. offset). The ISIC CIG captures broadband reflectivity amplitude, providing a short-wavelength impedance constraint.

Here we extend to simultaneous Bayesian recovery of velocity and impedance by conditioning on two imaging conditions with fundamentally different physics. The FWI-type CIG cross-correlates forward and back-propagated wavefields: reflectors appear at non-zero offset when the migration velocity is incorrect, with offset move-out encoding the velocity error, making it the natural input for tomographic updates. The Inverse Scattering Imaging Condition (ISIC) CIG applies the imaging condition (Sava and Vasconcelos, 2011; Whitmore and Crawley, 2012; Albano et al., 2022) to recover reflectivity amplitudes, making it sensitive to short-wavelength impedance contrasts and a natural input for AVO analysis (Kumar2013, 2013). Together, these two CIG types provide the joint information content needed to simultaneously constrain both parameters. Figure 1 and Figure 2 illustrate their complementary character on a held-out test model from the Compass benchmark.

A key challenge is the Gardner coupling (Gardner et al., 1974) ($\rho = 0.31(1000v)^{0.25}$, $Z = \rho v$): a network could predict impedance trivially via $Z \propto v^{1.25}$ without learning ISIC amplitude physics. We prevent this by decoupling training labels: velocity targets are lightly smoothed (defined precisely in the Method section below) to match what the anti-ISIC CIG can resolve kinematically, while impedance targets retain the unsmoothed ground truth via Gardner’s law applied to the raw velocity, preserving sharp reflectors. The forward simulation uses Gardner density applied to the lightly smoothed velocity for kinematic self-consistency with the velocity label.

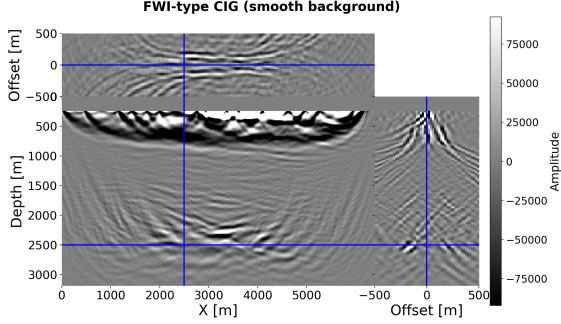


Figure 2: Anti-ISIC CIG for the same test model. The gather at $x = 2500$ m (bottom-right) shows strong offset-dependent moveout that encodes the kinematic velocity error between the smooth background and the true velocity field.

METHOD

Given observed shot data $\mathbf{d}_{\text{obs}}$ and a smooth background $\mathbf{x}_0$ (Gaussian-smoothed slowness, $\sigma=20$ grid samples), two $N_h=51$ -offset CIGs are computed:

$$\mathbf{y}_{\text{ISIC}} = \nabla \mathcal{F}_{\text{ISIC}}[\mathbf{x}_0]^\top \Delta \mathbf{d}, \quad \mathbf{y}_{\text{aISIC}} = \nabla \mathcal{F}_{\text{aISIC}}[\mathbf{x}_0]^\top \Delta \mathbf{d},$$

where $\Delta \mathbf{d} = \mathcal{F}(\mathbf{x}_0) - \mathbf{d}_{\text{obs}}$ is the data residual, $\mathcal{F}(\mathbf{x}_0)$ is the acoustic wave-equation forward operator evaluated at the background model, and $\nabla \mathcal{F}_{\text{ISIC/aISIC}}[\mathbf{x}_0]$ is the extended Born operator for the ISIC and anti-ISIC imaging conditions respectively. Both CIGs are concatenated into a 102-channel conditioning tensor $\mathbf{y} \in \mathbb{R}^{102 \times N_z \times N_x}$ ($N_z=256$, $N_x=512$), normalized by per-channel statistics estimated on the training set.

A lightweight summary UNet h_ϕ (encoder-decoder with skip connections) compresses $\mathbf{y}$ from 102 channels to 18 summary channels, discarding CIG features uninformative for inversion while preserving the complementary velocity and impedance information. An EDM-preconditioned SongUNet D_θ (Karras et al., 2022) then jointly denoises the two-channel target $\mathbf{x} = (\mathbf{v}, \mathbf{Z})$ conditioned on the summary:

$$\mathcal{L}(\theta, \phi) = \mathbb{E}_{\mathbf{x}, \mathbf{y}, \mathbf{n}} \|D_\theta([\mathbf{x} + \mathbf{n}, h_\phi(\mathbf{y})], \sigma_n) - \mathbf{x}\|_2^2, \quad \mathbf{n} \sim \mathcal{N}(0, \sigma_n^2 I).$$

Here, the symbol σ_n denotes the noise level in the EDM framework (distinct from the spatial Gaussian smoothing parameter used for the velocity targets). Velocity-model training targets are obtained by applying a Gaussian filter with standard deviation $\sigma=2$ grid samples (each sample is 12.5 m, so $\sigma=25$ m) to the raw ground-truth velocity model. At this level of smoothing, long-wavelength structure is preserved while short-wavelength reflectivity is suppressed, matching the kinematic resolution achievable from the anti-ISIC CIG. We refer to these lightly smoothed models as ground-truth velocity throughout this paper. The joint two-channel output enables the model to learn the full multi-parameter velocity-impedance posterior, including residual coupling

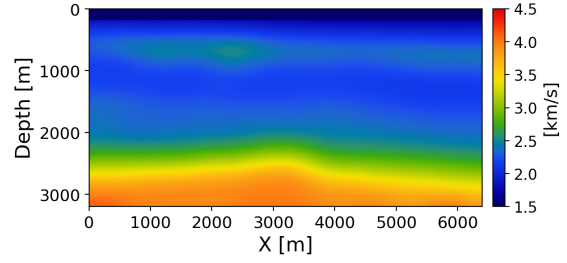


Figure 3: Background velocity model used as migration input for CIG computation. This model is obtained by applying a Gaussian filter with standard deviation $\sigma=20$ grid samples to the slowness field.

not captured by the Gardner relation. At inference, 16 posterior samples are drawn via the EDM probability-flow ODE, providing both a posterior mean estimate and a pixel-wise uncertainty map.

The summary UNet uses 4 encoder and 4 decoder blocks with channel progression [64, 128, 256, 512] and a 1024-channel bottleneck. The EDM SongUNet operates at three resolution levels (256, 128, 64 pixels) with 4 residual blocks per level, base channel width 64, and self-attention at the 16×16 resolution. The combined model totals approximately 54 million trainable parameters.

EXPERIMENT SETUP

We use the 1000-model Compass benchmark (E. Jones et al., 2012) (512×256 grid, $\Delta x=12.5$ m) with North Sea geology including complex structural interfaces and faults. Observed data are generated from the lightly smoothed ground-truth velocity ($N_{\text{src}}=16$ jittered sources, Ricker wavelet, SNR=12 dB), ensuring CIG residuals are kinematically consistent with the velocity labels. Impedance labels use the unsmoothed raw ground-truth impedance via Gardner’s law applied to the raw velocity, decoupled from the velocity label smoothing. This forced decoupling requires the network to extract velocity from anti-ISIC CIG kinematics and impedance from ISIC CIG amplitude independently. Training uses model indices 1-800; validation uses 801-1000; training runs for 100 epochs with batch size 2 and Adam optimizer ($\text{lr}=10^{-4}$).

RESULTS

We evaluate on a held-out test model from the Compass benchmark using 16 posterior samples. Figure 3 shows the background velocity used as migration input and Figure 4 shows the lightly smoothed ground-truth velocity. The background is spectrally smooth and lacks the structural detail present in the ground truth. The model must recover these features from the data residuals used to form the CIGs alone.

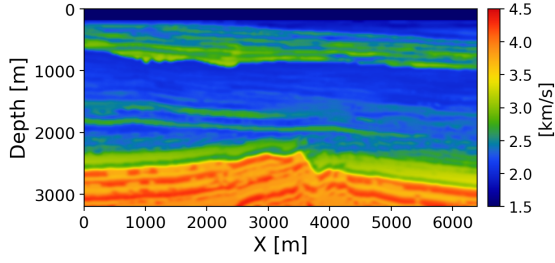


Figure 4: Ground-truth (lightly smoothed) velocity for a held-out test model.

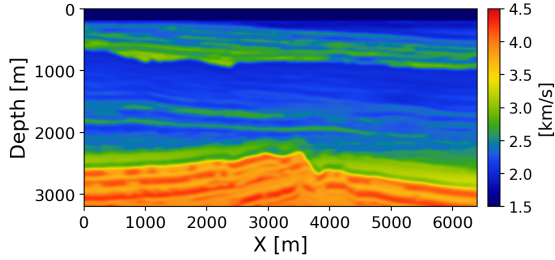


Figure 5: Posterior mean velocity from 16 samples. SSIM=0.967, RMSE=0.050 km/s.

Figure 5 shows the posterior mean velocity (SSIM=0.967, RMSE=0.050 km/s). Comparing with Figure 4, the model successfully recovers the water layer, fine-scale sedimentary layering, and deep structural interfaces from only 16 jittered sources, recovering structure entirely absent from the smooth background.

Figure 6 shows the posterior mean acoustic impedance (SSIM=0.867, RMSE=0.279 km/s $\cdot$ g/cm³). Impedance is recovered at substantially higher spatial resolution than velocity, resolving sharp sedimentary reflectors absent in the smooth velocity. This confirms the model exploits ISIC CIG amplitude information (Sava and Vasconcelos, 2011; Whitmore and Crawley, 2012; Albano et al., 2022) rather than inferring impedance from velocity via the Gardner relation.

Figure 7 and Figure 8 show pixel-wise posterior standard deviation for velocity and impedance respectively.

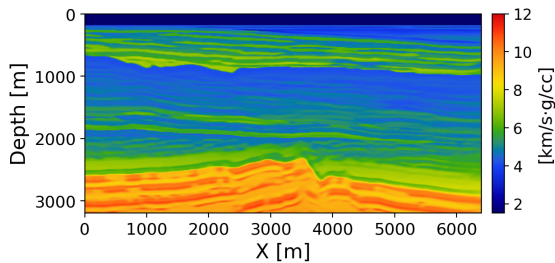


Figure 6: Posterior mean acoustic impedance, SSIM=0.867, RMSE=0.279 km/s $\cdot$ g/cm³.

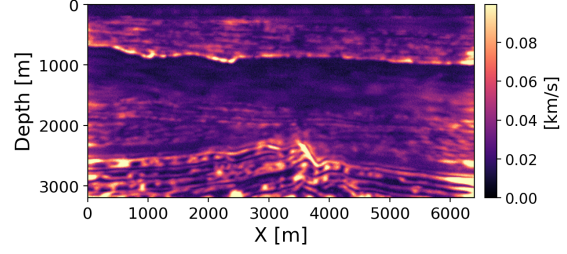


Figure 7: Posterior standard deviation for velocity from 16 samples.

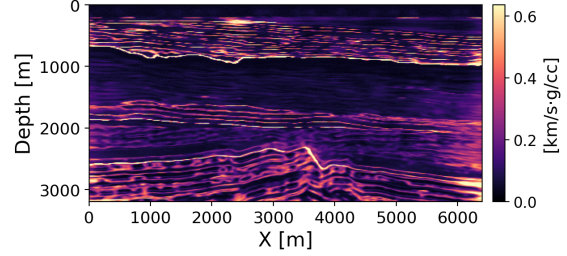


Figure 8: Posterior standard deviation for impedance from 16 samples.

Velocity uncertainty is highest at structural interfaces with strong velocity contrasts and increases with depth where sparse 16-source acquisition limits illumination. Impedance uncertainty, by contrast, concentrates along reflective interfaces (water bottom, intra-sediment reflectors) rather than in broad depth zones, consistent with the ISIC CIG sensitivity to short-wavelength amplitude contrasts (Sava and Vasconcelos, 2011; Kumar2013, 2013).

Beyond statistical metrics, we validate the velocity quality through CIG focusing. Using the same lightly smoothed ground-truth velocity as the true model, we compute new ISIC CIGs with two migration backgrounds: the original smooth background ($\sigma=20$, same as Figure 1) and the EDM posterior mean (conditioned on both ISIC and anti-ISIC CIGs). Figure 1 already shows the result with the smooth background: the gather at $x=2500$ m exhibits curved offset moveout indicating incorrect migration kinematics. Figure 9 shows the result with the posterior mean background: events are substantially more focused around zero offset and reflectors appear at their correct depths, confirming the posterior mean velocity correctly predicts wave propagation kinematics. This physics-based validation confirms the posterior mean is not merely statistically accurate but constitutes a geophysically meaningful velocity model suitable for seismic imaging.

We further compare the full model (ISIC + anti-ISIC, 18 summary channels) against an ISIC-only ablation (9 summary channels). Figure 10 shows vertical traces at $x=1600$ m for both configurations. For velocity, condi-

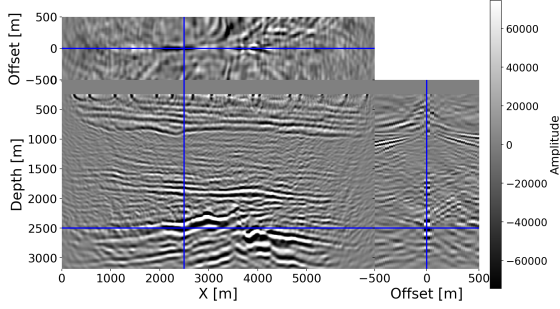


Figure 9: ISIC CIG computed using the EDM posterior mean velocity (conditioned on both ISIC and anti-ISIC CIGs) as migration background.

tioning on both CIG types improves alignment with the ground truth particularly at depths ≥ 1000 m, where the anti-ISIC CIG provides additional kinematic constraints. For impedance, the full model yields better ground-truth alignment overall at this position, suggesting the anti-ISIC CIG contributes complementary information that benefits impedance recovery as well.

CONCLUSIONS

We presented a joint multi-parameter Bayesian inversion framework that simultaneously recovers P-wave velocity and acoustic impedance by conditioning an EDM diffusion model on complementary ISIC and anti-ISIC Common Image Gatherers. The two CIG types supply fundamentally different information: the anti-ISIC CIG constrains the long-wavelength velocity field through kinematic offset moveout, while the ISIC CIG constrains short-wavelength impedance contrasts through broadband reflectivity amplitude. Deliberate training label decoupling prevents the model from bypassing amplitude physics via the Gardner relation.

On the 1000-model Compass benchmark, the posterior mean velocity achieves $\text{SSIM}=0.967$ and $\text{RMSE}=0.050$ km/s, and the posterior mean impedance achieves $\text{SSIM}=0.867$ and $\text{RMSE}=0.279$ km/s $\cdot$ g/cm³. Velocity quality is confirmed by CIG focusing: replacing the smooth migration background with the EDM posterior mean produces substantially more focused events in the offset gather, demonstrating that the inferred velocity is not only statistically accurate but geophysically meaningful as a migration input. The joint posterior further provides calibrated, spatially structured uncertainty: velocity uncertainty concentrates at structural interfaces and at depth where 16-source acquisition limits illumination, while impedance uncertainty localizes at reflective interfaces rather than in broad depth zones.

These results suggest that conditioning diffusion models on complementary CIG types provides a principled and computationally efficient path toward joint parameter recovery with uncertainty quantification in seismic

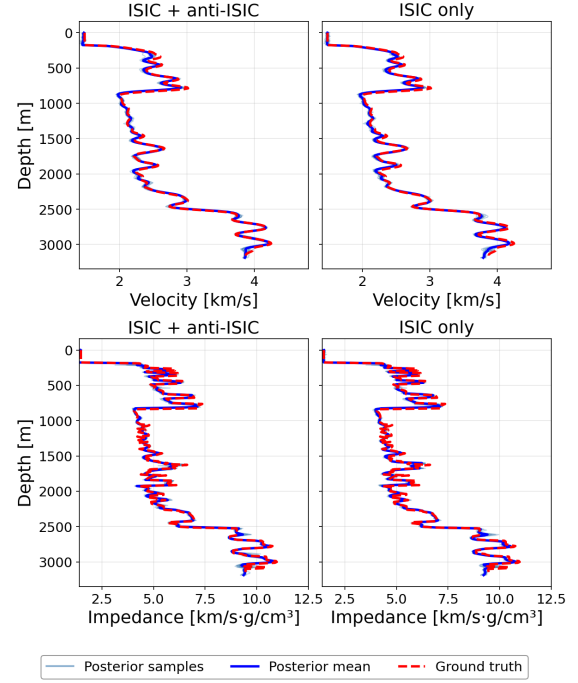


Figure 10: Velocity (top row) and impedance (bottom row) traces at $x=1600$ m. Left column: full model (ISIC + anti-ISIC). Right column: ISIC-only ablation. Ground truth (red dashed), posterior mean (blue solid), individual samples (light blue).

exploration. Several directions remain open. First, the current framework uses 16 sources per model; evaluating accuracy as a function of source count will clarify the acquisition requirements for reliable joint inversion. Second, extending the evaluation to the full 200-sample validation set and to a range of geological scenarios will establish the statistical robustness of the posteriors. Third, the summary network compression ratio (102 to 18 channels) represents a design choice whose impact on posterior quality warrants systematic ablation. Finally, adapting the framework to field data requires bridging the simulation-to-real gap through domain randomization or transfer learning strategies, which we identify as the most critical next step toward practical deployment.

ACKNOWLEDGEMENT

This research was carried out with the support of Georgia Research Alliance and partners of the ML4Seismic Center. The authors acknowledge the use of Claude and Claude Code (Anthropic) for writing assistance and code execution.

REFERENCES

- Albano, N. A., J. C. Costa, J. Schleicher, and A. Novais, 2022, On the estimation of reflectivity in reverse time migration: Implementational forms of the inverse-scattering imaging condition: *Geophysics*, 88, S1–S16.
- Avseth, P., T. Mukerji, and G. Mavko, 2010, *Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk*: Cambridge university press.
- Cranmer, K., J. Brehmer, and G. Louppe, 2020, The frontier of simulation-based inference: *Proceedings of the National Academy of Sciences*, 117, 30055–30062.
- E. Jones, C., J. A. Edgar, J. I. Selvage, and H. Crook, 2012, Building complex synthetic models to evaluate acquisition geometries and velocity inversion technologies: In 74th EAGE Conference and Exhibition Incorporating EUROPEC 2012, cp–293.
- Gardner, G. H. F., L. W. Gardner, and A. R. Gregory, 1974, Formation velocity and density—the diagnostic basics for stratigraphic traps: *Geophysics*, 39, 770–780.
- Geng, Z., Z. Zhao, Y. Shi, X. Wu, S. Fomel, and M. Sen, 2022, Deep learning for velocity model building with common-image gather volumes: *Geophysical Journal International*, 228, 1054–1070.
- Karras, T., M. Aittala, T. Aila, and S. Laine, 2022, Elucidating the design space of diffusion-based generative models: *Advances in Neural Information Processing Systems*, 26565–26577.
- Kumar2013 2013, Ava analysis and geological dip estimation via two-way wave-equation based extended images: volume SEG Technical Program Expanded Abstracts 2013 of SEG Technical Program Expanded Abstracts.
- Orozco, R., H. T. Erdinc, Y. Zeng, M. Louboutin, and F. J. Herrmann, 2024, Machine learning-enabled velocity model building with uncertainty quantification.
- Sava, P., and I. Vasconcelos, 2011, Wave-equation migration velocity analysis using extended images: *Geophysics*, 76, WA65–WA76. (Subsurface-offset extended imaging; CIGs for MVA and AVA).
- Siahkoohi, A., G. Rizzuti, and F. J. Herrmann, 2022, Deep Bayesian inference for seismic imaging with tasks: *Geophysics*, 87, S281–S302.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, 74, WCC1–WCC26.
- Whitmore, N. D., and S. Crawley, 2012, Applications of rtm inverse scattering imaging conditions: SEG Technical Program Expanded Abstracts 2012, 1–6.
- Wu, Z., Y. Sun, Y. Chen, B. Zhang, Y. Yue, and K. L. Bouman, 2024, Principled probabilistic imaging using diffusion models as plug-and-play priors: *arXiv preprint arXiv:2405.18782*.
- Yin, Z., R. Orozco, M. Louboutin, and F. J. Herrmann, 2024, Wise: full-waveform variational inference via subsurface extensions: *Geophysics*, 89, 1–31.
- Zeng, Y., H. T. Erdinc, R. Orozco, and F. J. Herrmann, 2025, Full-waveform variational inference with full common-image gathers and diffusion network: *arXiv preprint arXiv:2504.15289*.