Efficient and scalable posterior surrogate for seismic inversion via wavelet score-based generative models

Ege Cirakman[*1], *Huseyin Tuna Erdinc*[*2], *Felix J. Herrmann*[2]
[1]*Istanbul Technical University,* [2]*Georgia Institute of Technology*

SUMMARY

Seismic inversion poses significant computational challenges due to its high dimensionality and non-unique solutions. We propose a novel method integrating the Wavelet Score-Based Generative Model (WSGM) with Simulation-Based Inference (SBI) to enable efficient posterior sampling for full-waveform inference. Our approach reduces memory requirements ($\approx$ 50%) and significantly decreases sampling time ($\approx$ 73%) compared to standard score-based diffusion models, while preserving accuracy. Furthermore, WSGM naturally supports the generation of velocity models at multiple resolutions, leveraging its hierarchical structure. Experimental results on pairs of synthetic seismic images and velocity models demonstrate that our method enables posterior sampling for large-scale 2D geophysical problems and facilitates the assessment of uncertainties relevant to subsurface characterization.

INTRODUCTION

Accurate subsurface characterization remains a fundamental challenge in geophysical exploration, with seismic inversion serving as the primary tool for reconstructing subsurface properties, such as the acoustic wave-speed. The inverse problem of estimating velocity models from seismic observations is inherently ill-posed due to its high dimensionality, non-uniqueness and sensitivity to noise (Virieux and Operto (2009); Tarantola (2005)). As noted in the literature, traditional methods such as full-waveform inversion (FWI), that rely on point estimates, fail to capture the full uncertainty of the problem and do not produce posterior distributions, which is essential for informed decision-making in reservoir characterization and management (Siahkoohi et al. (2022); Fichtner et al. (2013); Xiao et al. (2025)).

Recent advances in machine learning have introduced promising algorithms to develop neural surrogates for posterior distributions. Specifically, SBI can facilitate posterior approximation of posterior $p(\mathbf{x} \mid \mathbf{y})$ in Bayesian inference without explicit evaluation of the costly likelihood/simulator, in our case directly related to the creation of subsurface images (Cranmer et al. (2020)). SBI can be implemented using various types of generative models, such as conditional normalizing flows (Dinh et al. (2017)) or score-based generative models(Song et al. (2020)), each with its own strengths and weaknesses. However, we observe that most of the existing generative modeling approaches often overlook the multiscale structure and long-range spatial correlations inherent in subsurface velocity models (Rizzuti and Vasconcelos (2024)) .

Based on this insight, we propose a conditional variation of WSGM(Guth et al. (2022)) within the SBI framework to perform posterior estimation for velocity inversion from seismic images. Our approach leverages Daubechies (db2) wavelets to decompose the posterior across multiple resolution scales (32×32 to 256×256), enabling hierarchical modeling and formulation of scale-specific score functions. This multi-scale factorization maintains consistency between observations and velocity estimates across resolutions, while reducing computational requirements compared to standard diffusion methods.

The key contributions of this paper are as follows: (i) the introduction of conditional WSGM for posterior sampling in seismic inversion, (ii) a cascaded network architecture designed to reduce memory consumption, (iii) comprehensive experiments on synthetic datasets demonstrating superior performance in reconstructing complex velocity structures, and (iv) the generation of uncertainty estimates that correlate well with errors. The remainder of the paper is organized as follows: we present the theory, describe the experimental setup, and discuss the results, establishing WSGM as an efficient and scalable approach for probabilistic seismic inversion.

THEORY

Seismic imaging

Seismic imaging aims to reconstruct a velocity model $\mathbf{x} \in \mathbb{R}^n$ from seismic observations $\mathbf{y} \in \mathbb{R}^m$ recorded at the surface, governed by the forward model $\mathbf{y} = \mathscr{F}(\mathbf{x}) + \boldsymbol{\varepsilon}$, where $\mathscr{F} : \mathbb{R}^n \to \mathbb{R}^m$ is a nonlinear operator solving the wave equation and $\boldsymbol{\varepsilon}$ represents noise (Virieux and Operto (2009)). The inverse problem is ill-posed, non-uniqueness (e.g., $\mathscr{F}(\mathbf{x}_1) \approx \mathscr{F}(\mathbf{x}_2) \approx \mathbf{y}$) and computational expensive due to its high dimensionality. Traditional full-waveform inversion (FWI) minimizes $\|\mathscr{F}(\mathbf{x}) - \mathbf{y}\|_2^2$ misfit objective to estimate $\mathbf{x}$, but it provides only point estimates without systematic uncertainty quantification (Virieux and Operto (2009)). In contrast, our study targets estimation of the posterior $p(\mathbf{x} \mid \mathbf{y})$ in the Bayesian inference setting using the WSGM with SBI.

SBI for posterior estimation

SBI proposes to directly estimate posterior $p_\theta(\mathbf{x} \mid \mathbf{y})$ using simulated data pairs $\mathscr{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, where $\mathbf{x}_i$'s are generated via the forward model, and train conditional generative models without explicit likelihood $p(\mathbf{y} \mid \mathbf{x})$ computation, which can be costly or physically impossible in many scientific settings (Cranmer et al. (2020)). A common generative model, normalizing flows can perform this task; yet, the inherent invertible structure may cause limitations in its performance (Rizzuti and Vasconcelos (2024)). In this work, we instead adopt a Conditional Score-Based Generative Model—specifically, WSGM—within the SBI framework, enabling efficient posterior estimation across multiple scales. Notably, in our formulation $\mathbf{y}$ represents RTM

images, which serve as summary statistics extracted from seismic observational data(Deans (2002); Yin et al. (2024)).

Score-based generative models (SGMs) and WSGM

SGMs learn the gradient of the log-density (score function) $\nabla_{\mathbf{x}} \log p(\mathbf{x})$ using a neural network $s_\theta(\mathbf{x})$, typically trained via a denoising score-matching objective (Song et al. (2020)). Sampling proceeds via Langevin dynamics (Hyvärinen (2005)): $\mathbf{x}_{t+1} = \mathbf{x}_t + \eta s_\theta(\mathbf{x}_t) + \sqrt{2\eta}\mathbf{n}_t$, where $\mathbf{n}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and η is the step size. While SGMs have shown impressive results in image synthesis tasks (Ho et al. (2020)), their application in scientific domains such as geophysics poses additional challenges. In these settings, score functions can be highly ill-conditioned due to long-range spatial correlations in the data, which result in poorly conditioned covariance structures. This makes both the training and sampling procedures significantly slower and more memory-intensive.

WSGM addresses these challenges through a multi-scale decomposition. WSGM proposes to decompose data into scaling coefficients, $\mathbf{x}_j = \gamma_j^{-1}\mathbf{G}\mathbf{x}_{j-1}$, and detail coefficients ,$\bar{\mathbf{x}}_j = \gamma_j^{-1}\bar{\mathbf{G}}\mathbf{x}_{j-1}$, using orthonormal wavelet filters $\mathbf{G}$ and $\bar{\mathbf{G}}$ where j and γ_j denote scale and scale dependent normalization, respectively. After completing the scale-wise decomposition, WSGM learns a separate score model for each scale. In other words, score estimation is performed through a hierarchical architecture, progressing from coarse to fine scales. Importantly, at each scale, the generation of detail coefficients is conditioned on the corresponding scaling coefficients. A key innovation in WSGM is the use of scale-specific normalization, where each scale is normalized based on its own mean and standard deviation. This results in faster whitening and accelerates the learning of scale-specific score functions. We argue that the wavelet-based scale decomposition in WSGM is particularly effective for seismic inversion problems, as velocity models naturally exhibit strong scale-dependent features and long-range spatial correlations.

Training objective and conditional WSGM

To enable posterior estimation in seismic inversion, we extend SGM and WSGM to learn the conditional score $\nabla_{\mathbf{x}} \log p(\mathbf{x} \mid \mathbf{y})$. For SGM, this involves training a network $s_\theta(\mathbf{x}, \mathbf{y}, \sigma(t))$ to approximate the score via a denoising objective conditioned on $\mathbf{y}$ (Batzolis et al. (2021); Song et al. (2024)). Given paired data $(\mathbf{x}, \mathbf{y})$, the objective becomes:

$$\widehat{\theta}_{\text{SGM}} = \arg\min_\theta \mathbb{E}_{\mathbf{y},\mathbf{x},\mathbf{n}} \|s_\theta(\mathbf{x}+\mathbf{n}, \mathbf{y}, \sigma(t)) - \mathbf{x}\|_2^2$$

where $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma(t)^2\mathbf{I})$ and $\sigma(t)$ follows a noise schedule (Karras et al. (2022)).

For WSGM (Guth et al. (2022)), the multi-scale structure enables hierarchical conditioning and modeling. The posterior density can be expressed by hierarchical factorization as follows:

$$p(\mathbf{x} \mid \mathbf{y}) = p(\mathbf{x}_J \mid \mathbf{y}_J) \prod_{j=1}^{J} p(\bar{\mathbf{x}}_j \mid \mathbf{x}_j, \bar{\mathbf{y}}_j)$$

where $\mathbf{x}_j$ is the velocity approximation at scale j, formed via normalized wavelet transform (WT) as $\text{WT}(\mathbf{x}_j) = (\mathbf{x}_{j+1}, \bar{\mathbf{x}}_{j+1})$ with $\mathbf{x}_j$ and $\bar{\mathbf{x}}_j$ representing scaling and detail coefficients at scale j and $j = 1$ corresponding to the finest scale. We can reverse the process with the inverse wavelet transform (IWT) and make similar arguments for $\mathbf{y}_j$.

With this factorization, we have divided the learning process to different cascaded models. The learning at the coarsest scale ($j = J$) can be expressed with the objective of SGM. However, for finer scales the score network learns $s_{\theta_j}(\bar{\mathbf{x}}_j, \mathbf{x}_j, \bar{\mathbf{y}}_j, \sigma(t))$. The loss at scale j integrates these dependencies and the objective becomes:

$$\widehat{\theta}_{\text{WSGM}} = \arg\min_\theta \mathbb{E}_{\bar{\mathbf{y}}_j, \mathbf{x}_j, \bar{\mathbf{x}}_j, \mathbf{n}} \left\| s_\theta(\bar{\mathbf{x}}_j + \mathbf{n}, \mathbf{x}_j, \bar{\mathbf{y}}_j, \sigma(t)) - \bar{\mathbf{x}}_j \right\|_2^2$$

Posterior sampling proceeds by solving the reverse-time SDE conditioned on unseen $\mathbf{y}^{\text{obs}}$. For WSGM, this process occurs sequentially: the coarsest-scale velocity $\mathbf{x}_J$ is sampled first, followed by detail coefficients $\bar{\mathbf{x}}_J$, conditioned on $\mathbf{x}_J$ and $\bar{\mathbf{y}}_J$. Then inverse wavelet transform of aggregated scaling and detail coefficients are used to proceed with finer scale and this process is repeated up to the original scale of inputs.

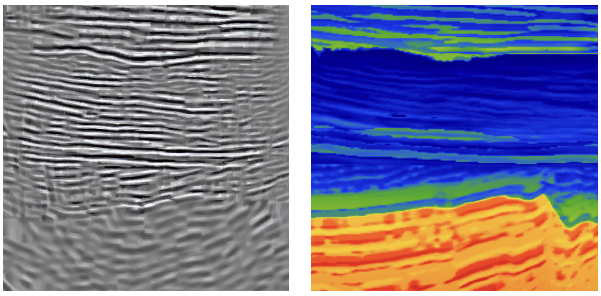
EXPERIMENTS AND RESULTS

Dataset creation

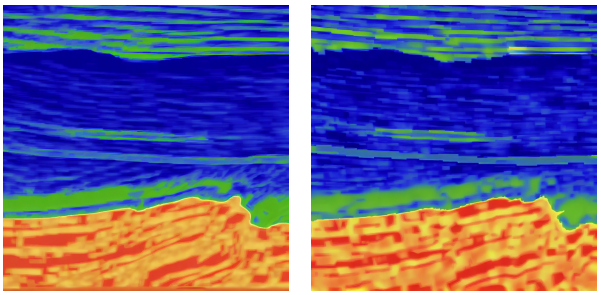
To assess the proposed methodology, we utilize a synthetic 3D Earth model derived from the Compass model as a representative of geological formations in the North Sea region (E. Jones et al. (2012)). The training dataset pairs consisting of the 2D velocities sliced through the 3D synthetic model and reverse-time migration (RTM) pairs. The total number of training samples is 3000 and the computational grid/resolution is 256x256 with a spatial resolution of 6.25 m, each sample covering an area of 3.2km x 3.2km. Seismic wave data is generated with 16 sources and 256 receivers with a dominant frequency of 15 Hz and a recording duration of 1.8 seconds. To simulate real-world conditions, 10 dB SNR colored Gaussian noise is added to the shot records before migration. The migration process for RTM is preformed with a Gaussian severely smoothed 2D background model. Wave simulations and imaging are performed using the open-source package JUDI (Louboutin et al. (2023)).

Inference results

We trained both WSGM and SGM using identical hyperparameters within a 16GB GPU memory constraint. Following training, we conducted a comparative evaluation of their performance on posterior estimation tasks in seismic inversion. Figure 1 presents the ground-truth velocity model and the corresponding RTM image, alongside posterior samples generated by WSGM and SGM using the same initial noise seed. We observe that both models are able to capture the layered structure and long-range spatial correlations present in the velocity model. WSGM demonstrates higher reconstruction quality compared to SGM across multiple evaluation metrics: SSIM (0.77 vs. 0.69), PSNR (24.5.dB vs. 19.97.dB), and RMSE (0.1018 vs. 0.1078), where the first value corresponds to WSGM and the second to SGM. The Structural Similarity Index (SSIM) measures perceptual image quality, with values closer to 1 indicating greater similarity



(a) RTM (b) GT



(c) WSGM (d) SGM

Figure 1: Posterior sampling results showing: (a) initial RTM condition on which samples of posterior are conditioned, (b) ground-truth (GT) velocity model, (c) WSGM posterior sample, and (d) SGM posterior sample. The WSGM result shows superior preservation of fine-scale details and reduced noise.

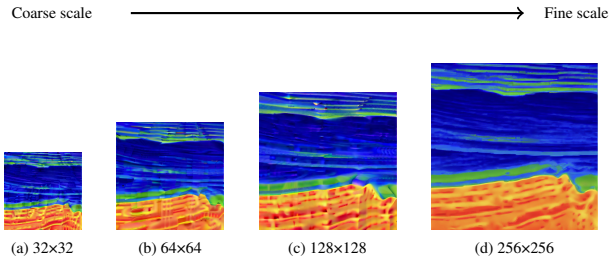
to the ground truth. Peak Signal-to-Noise Ratio (PSNR), expressed in decibels, quantifies the ratio between the maximum possible signal and the level of noise—higher PSNR values indicate better reconstruction fidelity. Root Mean Squared Error (RMSE) measures the average magnitude of error between the predicted and true velocity models, with lower values indicating higher accuracy. Collectively, these metrics confirm that WSGM produces more accurate and perceptually faithful posterior samples than SGM.

Despite having identical training durations and configurations, SGM's training at full-resolution introduces noticeable residual noise, especially at finer scales in the mid-section of the velocity. In contrast, WSGM's hierarchical, multi-scale architecture facilitates higher quality denoising at all levels of resolution. As summarized in Table 1, WSGM achieves these improvements while using approximately 50% less GPU memory (7.89GB vs. 15.7GB). These results highlight that our wavelet-based approach not only enhances reconstruction fidelity but also improves computational efficiency, essential for large-scale seismic inversion tasks.

Method	Batch Size	GPU Usage	Training Time	PSNR (dB)	SSIM
WSGM	4	7.89 GB	46 hours	24.5	0.77
SGM	4	15.7 GB	46 hours	19.97	0.69

Table 1: Comparison of model performance between WSGM and SGM after each was trained for the same duration.

Multi-scale decomposition analysis



(a) 32×32 (b) 64×64 (c) 128×128 (d) 256×256

Figure 2: Multi-scale visualization showing the progression from coarse to fine resolution in wavelet space. The progression represents the hierarchical nature of the wavelet decomposition, with each level containing increasingly detailed geological structures.

One of the advantages of our formulation is its ability to generate velocity models at multiple resolutions. Figure 2 illustrates the progressive refinement of velocity estimates across scales, demonstrating a hierarchical reconstruction process from coarse (32×32) to fine (256×256) resolution. This multi-scale strategy decomposes the ill-posed inverse problem into a sequence of simpler, more tractable sub-problems, facilitating both more stable learning and improved reconstruction fidelity at higher resolutions.

Uncertainty quantification

Uncertainty estimates play a crucial role in assessing the reliability of posterior samples. Well-calibrated networks can produce uncertainty estimates that correlate with the inherent errors in the model—a valuable feature in settings where ground truth data is unavailable or can never be observed. Figure 3 presents the posterior means, uncertainty estimates (standard deviations) and pixel-wise RMSE for both WSGM and SGM. The posterior means, conditioned on RTM, closely approximate the ground-truth velocity model. However, WSGM demonstrates superior performance by producing more sharper reflectors and preserving fine-scale velocity contrasts. This is quantitatively supported by WSGM's lower RMSE (0.1018 vs. 0.1078) and higher SSIM (0.77 vs. 0.69). The pixel-wise standard deviations in Figure 3 (g) and (h) reveal that both models yield uncertainty estimates that align with subsurface layering. Notably, high uncertainty values are observed along geological boundaries and in regions with complex structural features—areas typically associated with limited seismic illumination. These spatially coherent uncertainty patterns provide valuable insights for risk-aware decision-making in exploration and reservoir characterization.

Our posterior means and standard deviations are computed using 32 samples generated with the same initial noise seed. We highlight that sample generation with WSGM is significantly faster than its SGM counterpart. Specifically, WSGM reduces the time required for sampling by approximately 73% compared to SGM. This computational efficiency stems from the better-conditioned score functions in WSGM, which enables faster convergence of the reverse diffusion process(Guth et al. (2022)).

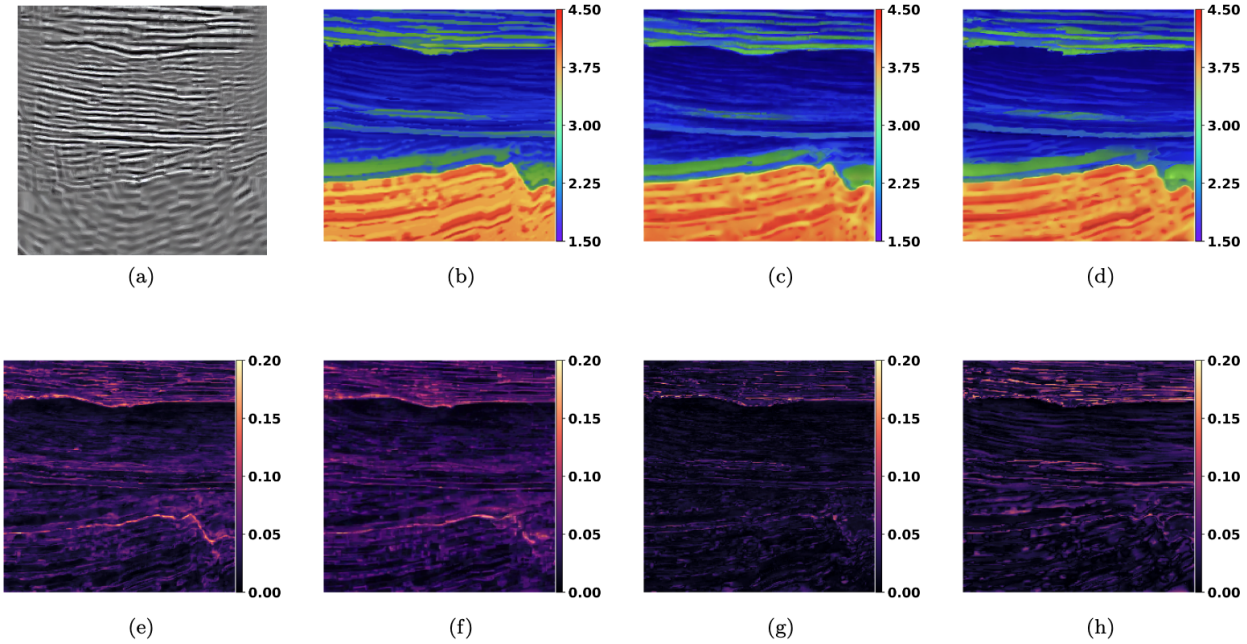


Figure 3: Performance analysis and uncertainty quantification and : (a) RTM condition, (b) ground-truth velocity model (GT) velocity, conditional mean posterior samples from (c) WSGM and (d) SGM, the standard deviation of (e) WSGM and (f) SGM, and RMSE for (g) WSGM and (h) SGM. Note the correlation between higher uncertainty and error regions.

Previously, we presented examples conditioned on the same RTM input. To further evaluate performance, we now showcase results on multiple unseen test RTMs and compare the performance of WSGM and SGM. As illustrated in Figure 4, WSGM consistently captures long-range spatial correlations in the velocity models across all three examples and produces samples that are visually more coherent and geologically plausible than those generated by SGM. One observed limitation is that WSGM occasionally over smooths certain regions, leading to a slight loss of sharpness in some layer boundaries. However, considering that natural images such as velocity models are dominated by low-frequency components, this smoothing effect does not appear to result in geologically implausible or meaningless reconstructions. Instead, it reflects a trade-off that favors stability and realism in the posterior samples.

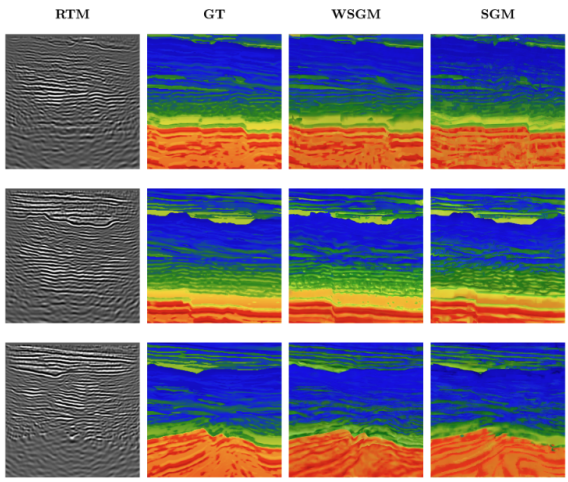


Figure 4: Comparison of GT, WSGM, and SGM. Posterior samples for WSGM and SGM were generated starting from the same initial noise

CONCLUSION

We have introduced a novel conditional variant of WSGM within the SBI framework for efficient posterior sampling in 2D seismic inversion. By decomposing the ill-posed score estimation problem into better-conditioned subproblems across multiple scales—via normalized wavelet coefficients—our method significantly improves computational efficiency, reducing memory usage by approximately 50% and accelerating posterior sampling by 73%. This approach also achieves superior reconstruction quality as measured by SSIM, PSNR, and RMSE metrics. Future work will investigate the use of curvelet transforms to further enhance directional feature representation, with potential application to 3D seismic volumes.

ACKNOWLEDGEMENT

This research was carried out with the support of Georgia Research Alliance, a partners of the ML4Seismic Center. During the preparation of this work, the authors used ChatGPT to refine sentence structures and improve readability. 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: arXiv preprint arXiv:2111.09217.

Cranmer, K., J. Brehmer, and G. Louppe, 2020, The frontier of simulation-based inference: Proceedings of the National Academy of Sciences, **117**, 30055–30062.

Deans, M. C., 2002, Maximally informative statistics for localization and mapping: Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292), IEEE, 1824–1829.

Dinh, L., J. Sohl-Dickstein, and S. Bengio, 2017, Density estimation using real nvp: International Conference on Learning Representations.

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.

Fichtner, A., J. Trampert, P. Cupillard, E. Saygin, T. Taymaz, Y. Capdeville, and A. Villaseñor, 2013, Multiscale full waveform inversion: Geophysical Journal International, **194**, 534–556.

Guth, F., S. Coste, V. De Bortoli, and S. Mallat, 2022, Wavelet score-based generative modeling: Advances in neural information processing systems, **35**, 478–491.

Ho, J., A. Jain, and P. Abbeel, 2020, Denoising diffusion probabilistic models: Advances in neural information processing systems, **33**, 6840–6851.

Hyvärinen, A., 2005, Estimation of non-normalized statistical models by score matching: Journal of Machine Learning Research, **6**, 695–709.

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, **35**, 26565–26577.

Louboutin, M., P. Witte, Z. Yin, H. Modzelewski, Kerim, C. da Costa, and P. Nogueira, 2023, slimgroup/judi.jl: v3.2.3.

Rizzuti, G., and I. Vasconcelos, 2024, Multiscale uncertainty quantification for post-stack seismic inversion with wavelet flows: 85th EAGE Annual Conference & Exhibition, European Association of Geoscientists & Engineers, 1–5.

Siahkoohi, A., G. Rizzuti, and F. J. Herrmann, 2022, Deep bayesian inference for seismic imaging with tasks: Geophysics, **87**, S281–S302.

Song, Y., O. Lopez, X. Yang, and M. Ravasi, 2024, Score-based generative modeling for full waveform inversion with an application to nonlinear ultrasound imaging: arXiv preprint arXiv:2411.06651.

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.

Tarantola, A., 2005, Inverse problem theory: Methods for data fitting and model parameter estimation: SIAM: Society for Industrial and Applied Mathematics.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Xiao, Z., J. Rao, S. Eisenträger, K.-V. Yuen, and S. A. Hadigheh, 2025, Generative adversarial network-based ultrasonic full waveform inversion for high-density polyethylene structures: Mechanical Systems and Signal Processing, **224**, 112160.

Yin, Z., R. Orozco, M. Louboutin, and F. J. Herrmann, 2024, Wise: Full-waveform variational inference via subsurface extensions: Geophysics, **89**, A23–A28.