Unsupervised Diffusion Model for Seismic Deconvolution

Hongzhi Yu, Wenchao Chen, Xiaokai Wang, *Member, IEEE*, and Dawei Liu

***Abstract*— Seismic data deconvolution is vital for enhancing resolution and accurate subsurface interpretation. Traditional methods heavily rely on predefined assumptions that limit their robustness to noisy data. As state-of-the-art generative models, diffusion models excel in capturing accurate prior distributions, which are beneficial to inversion. Moreover, diffusion models inherently resist noise due to their training in reverse noisy processes. Building on this foundation, we introduce an unsupervised diffusion model for seismic deconvolution, leveraging diffusion posterior sampling (DPS) to incorporate observed seismic data into the sampling process to guide high-accuracy reflectivity generation. Unlike traditional single-trace approaches, our method performs deconvolution across entire 2-D profiles, effectively capturing spatial continuity. Though solely trained on synthetic data, our method exhibits satisfactory performance when applied to synthetic and field datasets, demonstrating strong noise resistance and remarkable generalization capabilities.**

***Index Terms*— Diffusion model, high-resolution seismic data, seismic deconvolution, unsupervised.**

I. Introduction

THE high-resolution seismic data are crucial for the accurate interpretation of geological structures, and deconvolution methods are among the effective approaches to enhancing resolution. The method primarily depends on the convolution model [1], where seismic records are represented as the convolution of seismic wavelets and subsurface reflection coefficients. Deconvolution techniques improve seismic resolution by compressing the seismic wavelet and broadening the frequency spectrum. Several deconvolution methods have been developed based on the least-squares principle, including predictive deconvolution [2] and minimum-phase deconvolution [3]. However, these traditional deconvolution methods assume that the reflection coefficients are statistically white noise with Gaussian distribution, and the seismic wavelet follows a minimum-phase characteristic. Although these simplified assumptions allow for stable and causal wavelet inverse filters, their Gaussianity yields band-limited output, making it difficult to sharply resolve closely spaced reflectors. Subsequently, several researchers proposed methods to address these limitations, such as minimum entropy deconvolution, sparse-spike deconvolution, and l_p sparsity constraints. The sparse prior assumes that reflection coefficients are composed of a few isolated spikes. However, this assumption is suitable only when the seismic records are dominated by a limited number of strong reflections. For these traditional deconvolution methods, inversion is typically performed trace-by-trace (single-trace inversion), which makes it challenging to consider the spatial correlations in seismic data, resulting in outputs lacking spatial structure and poor lateral continuity. Although several researchers have proposed multitrace deconvolution methods, these methods generally rely on adjacent traces for deconvolution. While this strategy can improve lateral continuity to some extent, it remains insufficient for capturing long-range contextual dependencies effectively. Moreover, field noisy seismic data put challenges for traditional deconvolution to recover fine details accurately. In contrast, most intelligent deconvolution methods are data-driven, relying on large amounts of training data to learn the mapping relationship between inputs and outputs. In addition, the network structure contains implicit regularizations, which improves its robustness to noise.

In recent years, with the development of deep learning, there has been widespread attention on applying deep learning to seismic data resolution enhancement. Compared to traditional deconvolution methods, deep learning methods show greater robustness to noise because they learn hierarchical, nonlinear representations of the data that can inherently distinguish signal from noise. During training, deep learning models are exposed to large datasets (often with varying noise levels), which allows them to capture the underlying coherent patterns of the seismic signals while filtering out localized noise. Their nonlinear modeling capabilities further enable them to approximate complex relationships between seismic data and reflectivity, overcoming the limitations of linear assumptions in classical methods. In addition, regularization techniques such as dropout and batch normalization act as implicit regularizers, preventing overfitting noise and steering solutions toward geologically plausible results. In practical applications, Chai et al. [4] used convolutional neural networks (CNNs) for deconvolution, achieving promising results. Building on a generalized convolution model, Gao et al. [5] applied deep learning techniques for seismic data deconvolution, improving its vertical resolution. Though more effective than traditional methods, they are supervised and encounter generalization

Received 16 January 2025; revised 20 February 2025; accepted 3 March 2025. Date of publication 7 March 2025; date of current version 1 April 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFA0716904, in part by the National Natural Science Foundation of China under Grant 42374135, in part by the Fundamental Research Funds for the Central Universities under Grant xzy012023073, and in part by the Innovation Capability Support Program of Shaanxi under Grant 2021TD-08. *(Corresponding author: Dawei Liu.)*

Hongzhi Yu is with the School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China (e-mail: yhz129i@stu.xjtu.edu.cn).

Wenchao Chen and Xiaokai Wang are with the School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China, and also with Shaanxi Smart Networks and Ubiquitous Access Research Center, Xi'an, Shaanxi 710049, China (e-mail: wencchen@xjtu.edu.cn; xkwang@xjtu.edu.cn).

Dawei Liu is with the School of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China, and also with the Department of Physics, University of Alberta, Edmonton, AB T6G 2C7, Canada (e-mail: 409791715@qq.com).

Digital Object Identifier 10.1109/LGRS.2025.3549055

issues. These methods typically require retraining for different datasets [6], [7], [8]. However, real data often lack paired labels, which poses a challenge to their practicality.

The diffusion model [9] is a generative model that has attracted considerable attention due to its remarkable success in generating high-quality outputs across diverse data types, including images, audio, and text. In contrast to the aforementioned deep learning methods, diffusion models learn to approximate the probability distribution of the target data and exhibit desirable properties, including fixed training objectives and scalability. The fixed training objective is primarily reflected in the fact that regardless of the task, the goal remains focused on learning noise prediction and optimizing the variational lower bound. Scalability is demonstrated through the flexible adjustment of model size, training computation, and sampling steps, allowing the model's performance to improve smoothly with increased resources while maintaining practicality. Diffusion models effectively capture prior information by modeling the distribution of the target data, providing significant advantages over traditional regularization methods when solving inverse problems. This advantage is particularly evident when processing noisy data. Unlike simplified priors utilized in traditional approaches, diffusion models benefit from the complex prior information they have learned. Moreover, this approach of directly modeling the target data distribution allows training to be conducted using only clean reflection coefficients, eliminating the need for paired data-reflectivity training samples. Furthermore, compared to generative models such as generative adversarial networks (GANs), diffusion models are not dependent on adversarial training, thus avoiding training instability issues such as mode collapse. In addition, their optimization process, based on maximizing likelihood estimation and training a denoiser, enables more stable convergence while rendering them inherently robust to noise when solving inverse problems. This architecture allows diffusion models to effectively handle large and complex datasets, especially intricate seismic data.

We explored the potential application of diffusion models to seismic deconvolution for achieving high-resolution results. Since diffusion models are unconditional generative models [10], directly applying them to deconvolution could yield reflection coefficients that do not match the observed data. To address these issues and adapt diffusion models for seismic data deconvolution, we introduced diffusion posterior sampling (DPS) [11], incorporating observed data as conditional information at each step of the sampling process to guide the reverse generation process. We trained the model on synthetic data and achieved promising results across multiple datasets, demonstrating satisfactory generalization capability and robustness to noise. In Section II, we introduce the diffusion model and DPS. In Section III, we use synthetic and field seismic data to prove the effectiveness of our method. Finally, we conclude this article in Section IV.

II. Method

A. Score-Based Diffusion Model

The diffusion model is a generative model designed to reconstruct the data following a desired distribution from Gaussian noise data. It defines a forward process, where noise is progressively added to the original data and represents the generation process as the reverse of this learned noise-adding procedure. This concept was initially introduced in [9] and later expanded upon in [10] with the proposal of a score-based diffusion model that utilizes stochastic differential equations (SDE) to define the forward process.

In the forward process, given target data $\boldsymbol{x}_0 \sim p_0$, noise is progressively added over time steps T, gradually transforming the data into a simple standard Gaussian distribution $\boldsymbol{x}_T \sim \mathcal{N}(0, \mathbf{I})$. The forward process is defined as follows:

$$\mathrm{d}\boldsymbol{x} = -\frac{\beta_t}{2}\boldsymbol{x}\mathrm{d}t + \sqrt{\beta_t}\mathrm{d}\boldsymbol{w} \tag{1}$$

where β_t is the noise schedule of the process. According to [9], β_t is typically taken to be a linear function that increases monotonically of t and $\boldsymbol{w}$ is the standard Wiener process. The reverse-time SDE of (1) can be expressed as follows:

$$\mathrm{d}\boldsymbol{x} = \left[-\frac{\beta_t}{2}\boldsymbol{x} - \beta_t \nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t)\right]\mathrm{d}t + \sqrt{\beta_t}\mathrm{d}\overline{\boldsymbol{w}} \tag{2}$$

where $\mathrm{d}t$ corresponds to time running backward and $\mathrm{d}\overline{\boldsymbol{w}}$ corresponds to the standard Wiener process running backward.

Once given a time-dependent score function $\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t)$, the reverse-time stochastic differential equation for the generation process can be solved. The score of a distribution can be estimated by training a score-based model on samples with score matching. To estimate $\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t)$, we use a neural network $\boldsymbol{s}_\theta$ trained via denoising score matching (DSM) to approximate

$$\theta^* = \arg\min_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{x}_t|\boldsymbol{x}_0,\boldsymbol{x}_0}\left[\left\|\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) - \nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{x}_t|\boldsymbol{x}_0)\right\|_2^2\right]. \tag{3}$$

By substituting $\boldsymbol{s}_{\theta^*}(\boldsymbol{x}_t, t)$ for $\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t)$ in (2), the sampling results can be obtained recursively.

B. Diffusion Posterior Sampling

DPS [11] constrains the sampling process by integrating observed data into the sampling iterations. The general form of the forward model can be stated as

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{x}) + \boldsymbol{\epsilon} \tag{4}$$

where $\mathcal{A}(\cdot)$ represents the forward measurement operator and $\boldsymbol{\epsilon}$ denotes the measurement noise. According to Bayes' theorem, we can derive

$$p(\boldsymbol{x}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})/p(\boldsymbol{y}) \tag{5}$$

where the term $\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t)$ in (2) can be rewritten in the form of sampling from the posterior distribution as follows:

$$\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t|\boldsymbol{y}) = \nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t) + \nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{y}|\boldsymbol{x}_t). \tag{6}$$

The first term, $\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t)$, is the unconditional score term, which depends only on the training data distribution and can be replaced with the pretrained $\boldsymbol{s}_{\theta^*}$. However, the second term $\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{y}|\boldsymbol{x}_t)$ is difficult to obtain directly due to its dependence on time t, whereas we only have a direct dependence of $\boldsymbol{y}$ on $\boldsymbol{x}_0$. To circumvent this problem, Chung et al. [11] propose an approximation with a theoretically guaranteed upper bound on the approximation error

$$p(\boldsymbol{y}|\boldsymbol{x}_t) \simeq p(\boldsymbol{y}|\hat{\boldsymbol{x}}_0) \tag{7}$$

and $\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{y}|\boldsymbol{x}_t)$ can be further expressed as follows:

$$\nabla_{\boldsymbol{x}_t} \log p(\boldsymbol{y}|\boldsymbol{x}_t) \simeq \zeta_t \nabla_{\boldsymbol{x}_t} \left\| \boldsymbol{y} - \mathcal{A}(\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)) \right\|_2^2 \tag{8}$$

where $\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t)$ emphasizes that $\hat{\boldsymbol{x}}_0$ is a function of $\boldsymbol{x}_t$. The step size is given by $\zeta_t = P/\|\boldsymbol{y} - \mathcal{A}(\hat{\boldsymbol{x}}_0)\|$, where P is a hyperparameter that controls the step size, allowing for adjustments in the iterative process. Finally, (6) is reformulated as follows:

$$\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t|\boldsymbol{y}) \simeq \boldsymbol{s}_{\theta^*}(\boldsymbol{x}_t, t) - \zeta_t \nabla_{\boldsymbol{x}_t} \left\| \boldsymbol{y} - \mathcal{A}(\hat{\boldsymbol{x}}_0) \right\|_2^2. \tag{9}$$

Following [9], we predefine $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=0}^{t} \alpha_s$ to simplify calculations, where β_t is typically set to linearly increase from 1^{-4} to 0.02.

C. Seismic Deconvolution With DPS

Seismic deconvolution is the convolution of reflection coefficients and seismic wavelets, which is formally consistent with (4). Specifically, $\boldsymbol{x}$ represents the reflection coefficients, $\boldsymbol{y}$ is the observed seismic data, $\boldsymbol{\epsilon}$ denotes an unknown data error represented as additive noise, and $\mathcal{A}(\cdot)$ is also the forward operator.

In the convolution model, no specific assumption is made about the seismic wavelet, which may either be time-invariant or time-varying. In this letter, the seismic wavelet is assumed to be time-invariant, and the forward operator $\mathcal{A}(\cdot)$ is represented as a Toeplitz matrix formed by the seismic wavelet, where A_{ij} is given as follows:

$$A_{ij} = \begin{cases} s_{i-j+1}, & i \geq j \\ 0, & i < j \end{cases} \tag{10}$$

where s_{i-j+1} refers to the $(i - j + 1)$th sample point of the seismic wavelet. Therefore, we must estimate the seismic wavelet in advance. The algorithm is referred to as DeconDPS, and its algorithm flowchart is presented in Algorithm 1.

Algorithm 1 DeconDPS

Require: $T, \boldsymbol{y}, \mathcal{A}, \{\zeta_t\}_{t=1}^N, \{\boldsymbol{\sigma}_t\}_{t=1}^N$

1: $\boldsymbol{x}_T \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\quad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ **if** $t > 1$, **else** $\boldsymbol{z} = 0$
4: $\quad \hat{\boldsymbol{x}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x}_t + (1 - \bar{\alpha}_t)\boldsymbol{s}_\theta(\boldsymbol{x}_t, t)\right)$
5: $\quad \boldsymbol{x}'_{t-1} \leftarrow \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\boldsymbol{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\boldsymbol{x}}_0 + \boldsymbol{\sigma}_t \boldsymbol{z}$
6: $\quad \boldsymbol{x}_{t-1} \leftarrow \boldsymbol{x}'_{t-1} - \zeta_t \nabla_{\boldsymbol{x}_t} \|\boldsymbol{y} - \mathcal{A}(\hat{\boldsymbol{x}}_0)\|_2^2$
7: **end for**
8: **return** $\hat{\boldsymbol{x}}_0$

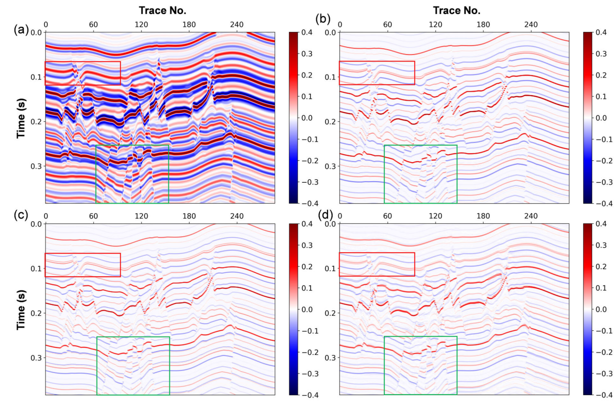

Fig. 1. Deconvolution of synthetic data. (a) Synthetic data. (c) True reflection coefficient. (b) and (d) Results of FISTA and DeconDPS, respectively.

III. Experiments

We first generated a 2-D reflection coefficient dataset of 5000 data; each reflection coefficient data size is 300 × 896, following the style shown in Fig. 1(c). This dataset captures features of real geological structures, including fault structures, and horizontal and curved strata; 80% of the dataset is selected for training. Before training, every data are randomly cropped to 256 × 256 patches. The remaining 20% are reserved for testing. We set T, the number of iterations in the forward process, to 1000. The Adam optimizer [12] is selected as the optimization algorithm with a learning rate of 1e-4. The training phase is conducted on an NVIDIA GeForce RTX 3090 GPU for a total of 1 000 000 steps. All subsequent tests on synthetic and real data are conducted based on this model. For comparative analysis, we selected the sparse spike deconvolution method, implemented using the fast iterative shrinkage thresholding algorithm (FISTA) [13].

A. Synthetic Data Examples

We selected a synthetic reflection coefficient with 288 traces, each containing 384 time points. A 30-Hz Ricker wavelet is convolved with the reflection coefficient to generate clean synthetic seismic data for testing. For the DeconDPS method, the hyperparameter P is set to 0.3. In the FISTA method, the regularization parameter $\mu = 0.05$, with two conditions for stopping the iteration: either the maximum iteration K_{iter} of 500 is reached, or the convergence tolerance falls below 10^{-5}.

Fig. 1 shows the results of the deconvolution. For noise-free input seismic data, both DeconDPS and FISTA accurately reconstruct the reflectivity coefficients, achieving high detail recovery and strong lateral continuity. As highlighted by the green box in Fig. 1, both methods clearly display the fault structure. In the thin-layer structure shown in the red box, FISTA performs relatively better in the noise-free data. Fig. 2 illustrates the normalized average amplitude spectra across multiple traces. The figure shows that the DeconDPS method closely matches the true reflectivity coefficients at low frequencies. Although there is a slight decline at higher frequencies, the overall performance remains robust, further

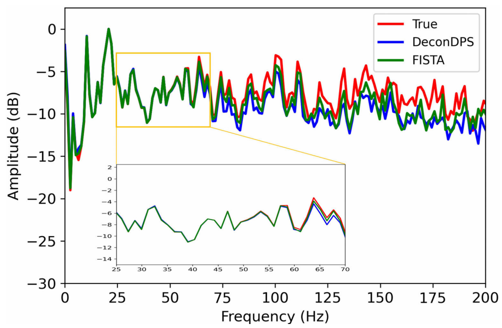

Fig. 2. Multitrace normalized average amplitude spectrum.

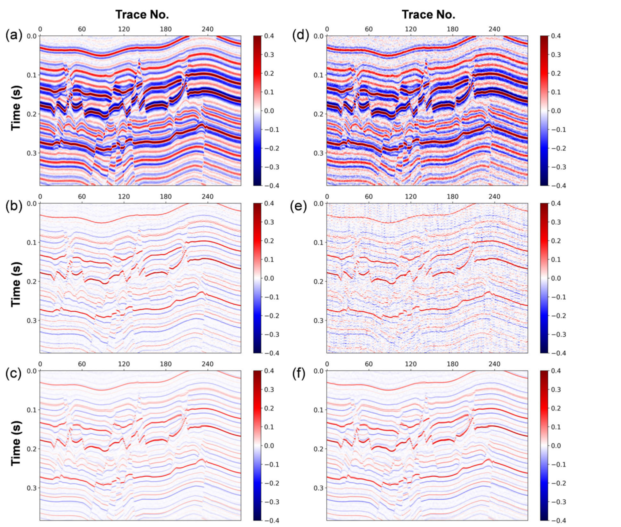

Fig. 3. Deconvolution results under different noise levels. (a) and (d) Represent seismic data with SNRs of 18 and 9 dB, respectively. (b) and (e) Corresponding FISTA results, while (c) and (f) present the corresponding DeconDPS results.

validating the effectiveness of the DeconDPS method for deconvolution.

Although the performance is satisfactory in the absence of noise, real seismic data are inevitably distorted by noise. During testing with noisy data, clean synthetic data are augmented with noise at a fixed SNR; we adopt the definition

$$\mathrm{SNR} = \frac{||\boldsymbol{y}||_2^2}{\alpha^2 ||\boldsymbol{n}||_2^2} \tag{11}$$

where $\boldsymbol{y}$ is clean signal, α is a scalar that determines the desired SNR, and $\boldsymbol{n}$ is sampled from a standard normal distribution.

Seismic data with SNRs of 18 and 9 dB are introduced for further testing, as illustrated in Fig. 3(a) and (d). For noisy data, we increase the step size in the DeconDPS method by setting the hyperparameter P to 0.3, Meanwhile, in the FISTA method, we increase the regularization term to promote sparse solutions by setting μ to 0.1. Subsequently, the FISTA method was applied for deconvolution, and the results are presented in Fig. 3(b) and (e). At lower noise levels, FISTA reconstructs relatively accurate reflectivity coefficients. However, the lateral continuity of the results is noticeably degraded due to noise. As noise levels increase, only the stronger components are reconstructed, while finer details are largely obscured by noise.

TABLE I

IMPACT OF DIFFERENT SNRS ON ACCURACY (dB)

SNR (dB)	18	15	12	9	6	3
FISTA	3.23	2.87	2.27	0.77	-1.02	-1.75
DeconDPS	3.41	3.21	2.89	2.53	2.33	1.80

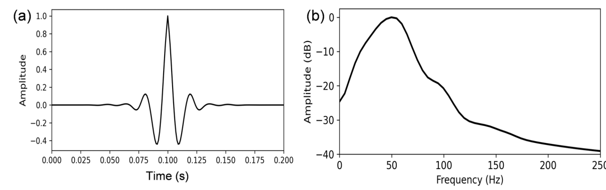

Fig. 4. (a) Extracted source wavelet from the seismic data. (b) Amplitude spectrum.

In contrast, the DeconDPS method, as illustrated in Fig. 3(c) and (f), performs consistently well across various noise levels, demonstrating strong robustness to noise.

To further analyze the effectiveness of the proposed method, we performed an accuracy analysis of the sampling result. Noisy synthetic data with different SNR values are selected and compared with the FISTA method. To quantitatively evaluate the results, we calculate the reconstruction accuracy

$$\text{Accuracy (dB)} = 10 \log_{10} \frac{||\boldsymbol{x}||_2^2}{||\boldsymbol{x} - \boldsymbol{x}^*||_2^2} \tag{12}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}^*$ are the true and inverted reflectivity models, respectively. The results are shown in Table I.

As shown in Table I, at high SNR, the performance difference between the two methods is minimal, with the proposed method slightly outperforming FISTA. However, as noise increases, FISTA's performance deteriorates significantly, while the DeconDPS method demonstrates superior robustness to noise. Even at 3 dB, the accuracy remains relatively high with the DeconDPS method.

B. Field Data Examples

The effectiveness of the methods was subsequently validated on a real 2-D dataset. This 2-D dataset comprised 384 traces, each with 320 time sampling. For the DeconDPS method, the hyperparameter P is set to 0.2, as it exhibited the best performance after multiple tests. Considering the complexity of real data and its distortion by noise, the regularization parameter μ in the FISTA method is set to 0.1 to promote sparse solutions, while all other parameters remain consistent with previous experiments. The statistically estimated wavelet and frequency spectrum are illustrated in Fig. 4.

The results are illustrated in Fig. 5, where (a) depicts the real seismic data, and Fig. 5(b) and (c) displays the results of DeconDPS and FISTA, respectively. From an overall perspective, the DeconDPS method demonstrates significantly better lateral continuity because FISTA performs trace-by-trace deconvolution, whereas DeconDPS conducts deconvolution across the entire 2-D profile. As highlighted by the green box in Fig. 5, the fault is clearly visible in the original data. However, the FISTA result fails to reconstruct the fault effectively. As highlighted in the red box, channel

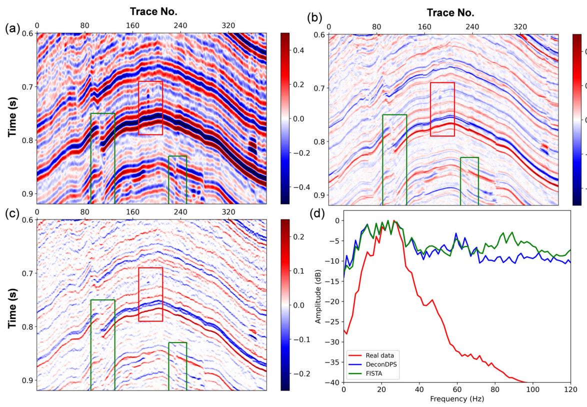

Fig. 5. Deconvolution results of real data. (a) Real seismic data. (b) Results of DeconDPS. (c) Results of FISTA. (d) Multitrace normalized average amplitude spectrum.

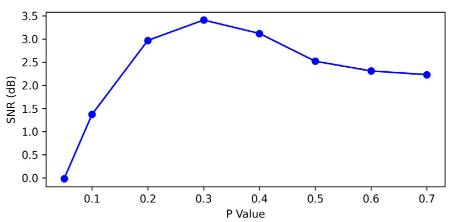

Fig. 6. SNR of the results sampled with different P values.

responses are visible in the original data. DeconDPS preserves these features well, outperforming FISTA in recovering detailed information. This demonstrates that DeconDPS outperforms FISTA in recovering detailed information. Fig. 5(d) illustrates the normalized amplitude spectra across multiple traces. DeconDPS significantly broadens the frequency range of the data, confirming high-resolution results.

These results on real seismic data validate the effectiveness of the DeconDPS method and underscore its potential for broader applications in deconvolution techniques. The success of DeconDPS in processing both synthetic and real data, despite being trained exclusively on synthetic examples, highlights its strong generalization capability and potential for broader applications in seismic data processing.

C. Discussion

1) Different Hyperparameters: As shown in (8), the proposed method has only one hyperparameter, P, making its selection a key factor influencing the sampling results. To further illustrate this, we performed a sensitivity analysis on the hyperparameter P using noisy data with an SNR of 18 dB. The result is presented in Fig. 6. It shows that there is an optimal value P that yields the best performance. In practice, we empirically select a value for P, typically set to 0.3, and fine-tune it during testing to achieve optimal performance.

2) Computational Efficiency: We performed a single sampling on 288×384 data using an NVIDIA GeForce RTX 3090 GPU and compared the runtime and GPU memory usage with those of another deep learning method (DDNM [14]). The FISTA method is executed on the CPU. The results are shown in Table II. Due to the gradient calculation required at each step, DeconDPS consumes more memory and time than DDNM. However, because of the inherently high computational cost of diffusion models, the runtime for both methods significantly exceeds that of the FISTA method even though the latter is executed on the CPU. Therefore, our future research will focus on accelerating the sampling process to reduce inference time.

TABLE II
COMPARISON OF INFERENCE TIME AND MEMORY USAGE

Method	DeconDPS	DDNM	FISTA
Inference Time (s)	98.51	90.26	23.48
GPU Memory Usage (G)	3.91	2.01	NA

IV. CONCLUSION

We reconstruct the reflection coefficients using a diffusion model. To address the issue of **nonuniqueness** in the generation of unconditional diffusion models, we introduced the DPS method that integrates prior information by sampling from the posterior distribution, thereby guiding and constraining the generated results. This approach does not require paired training data; instead, it only requires synthetic reflectivity coefficients that incorporate as much realistic subsurface information as possible. Both synthetic and real datasets demonstrate the effectiveness of the proposed method, highlighting its strong robustness to noise and remarkable generalization capabilities. These results also demonstrate that the DeconDPS method requires only a single pretraining step and the tuning of one hyperparameter to achieve robust performance across different datasets.

REFERENCES

[1] E. A. Robinson, "Predictive decomposition of seismic traces," *Geophysics*, vol. 22, no. 4, pp. 767–778, Oct. 1957.
[2] K. L. Peacock and S. Treitel, "Predictive deconvolution: Theory and practice," *Geophysics*, vol. 34, no. 2, pp. 155–169, Apr. 1969.
[3] A. J. Berkhout, "Least-squares inverse filtering and wavelet deconvolution," *Geophysics*, vol. 42, no. 7, pp. 1369–1383, Dec. 1977.
[4] X. Chai et al., "Deep learning for multitrace sparse-spike deconvolution," *Geophysics*, vol. 86, no. 3, pp. V207–V218, May 2021.
[5] Z. Gao et al., "A deep-learning-based generalized convolutional model for seismic data and its application in seismic deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4503117.
[6] N. Wang, Y. Shi, J. Ni, J. Fang, and B. Yu, "Enhanced seismic attenuation compensation: Integrating attention mechanisms with residual learning in neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5927011.
[7] D. Liu et al., "From shallow to deep: Enhancing seismic resolution with weak supervision," *Geophysics*, vol. 90, no. 3, pp. 1–108, 2025.
[8] D. Liu et al., "Enhancing ground-penetrating radar (GPR) data resolution through weakly supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5107513.
[9] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.
[10] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," 2020, *arXiv:2011.13456*.
[11] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. Chul Ye, "Diffusion posterior sampling for general noisy inverse problems," 2022, *arXiv:2209.14687*.
[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[13] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.
[14] Y. Wang, J. Yu, and J. Zhang, "Zero-shot image restoration using denoising diffusion null-space model," 2022, *arXiv:2212.00490*.