ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

A physics-constrained sparse basis learning method for mixed noise suppression

Yongsheng Wang[1,2†], Deying Wang[3,4†], Kai Zhang[3], Wenqing Liu[4], Longjiang Kou[4], and Huailiang Li[1,2*]

[1]State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, University of Technology, Chengdu, Sichuan, China

[2]PetroChina Qinghai Oilfield Company, Dunhuang, Gansu, China

[3]School of Geosciences, China University of Petroleum (East China), Qingdao, Shandong, China

[4]Seismic Data Processing and Interpretation Center, Research Institute of Petroleum Exploration and Development-Northwest, PetroChina, Lanzhou, Gansu, China

(This article belongs to the *Special Issue: Advanced Artificial Intelligence Theories and Methods for Seismic Exploration*)

Abstract

Suppressing complex mixed noise in seismic data poses a significant challenge for conventional methods, which often cause signal damage or leave residual noise. While sparse basis learning is a promising approach for this task, traditional data-driven learning methods are often insensitive to the physical properties of seismic signals, leading to incomplete noise removal and compromised signal fidelity. To address this limitation, we propose a physics-constrained sparse basis learning method for mixed noise suppression. Our method integrates local dip attributes—estimated and iteratively refined by a plane-wave destructor filter—as a physical constraint within the dictionary learning framework. This constraint guides the learning process to achieve high-fidelity signal reconstruction while effectively suppressing multiple noise types. Tests on complex synthetic and real data demonstrate that the proposed method outperforms conventional techniques and industry-standard workflows in attenuating mixed noise, including strong anomalous amplitudes, ground roll, and random and coherent components, thereby significantly enhancing the signal-to-noise ratio and imaging quality.

Keywords: Multiple-type noise suppression; Dictionary learning; Physical constraint; Plane-wave destructor filter

†These authors contributed equally to this work.

***Corresponding author:**
Huailiang Li
(lihl@cdut.edu.cn)

Citation: Wang Y, Wang D, Zhang K, Liu W, Kou L, Li H. A physics-constrained sparse basis learning method for mixed noise suppression. *J Seismic Explor*. 2025;34(4):42-59.
doi: 10.36922/JSE025280034

Received: July 13, 2025

1st revised: August 25, 2025

2nd revised: September 6, 2025

Accepted: September 9, 2025

Published online: October 27, 2025

1. Introduction

The evolution of seismic data denoising techniques reflects a deepening understanding of signal and noise characteristics and the continuous refinement of processing methodologies.[1] Early methods primarily relied on fixed-basis transforms, such as the Fourier transform (F-K filtering) and the wavelet transform. These approaches operate on the assumption that effective signals and noise exhibit distinct characteristics in the transformed domain, allowing for their separation through filtering or thresholding.[2]

The underlying principle for many of these techniques is sparse representation, which aims to find the most compact signal representation within an overcomplete dictionary, thereby enabling effective compression, feature extraction, and denoising.[3,4] Transforms such as the curvelet and shearlet were developed to better represent the linear and curvilinear features common in seismic wavefields, offering improved performance over traditional wavelets in preserving edges and suppressing coherent noise.[5-7] However, the efficacy of these fixed-basis methods is inherently limited; they may introduce artifacts or damage signals when the characteristics of the signal and noise overlap in the transform domain.[8,9] To overcome the rigidity of fixed bases, adaptive dictionary learning methods such as K-singular value decomposition (K-SVD) and the method of optimal directions were introduced. These techniques learn the dictionary atoms directly from the data, allowing the basis to adapt to the specific morphological features of seismic signals.[10,11]

In recent years, deep learning (DL) has been widely applied to seismic data denoising due to its powerful nonlinear modeling and feature learning capabilities.[12] Initial supervised models, such as the denoising convolutional neural network, demonstrated state-of-the-art performance by using residual learning to focus on noise components.[13] However, their reliance on large volumes of paired clean and noisy data for training significantly increases the preprocessing workload and limits their application in scenarios where clean reference data is unavailable.[13] To address this, recent research has shifted toward more flexible DL paradigms. Self-supervised learning models, for instance, can be trained effectively on noisy data alone, eliminating the need for clean labels by leveraging the statistical properties of the data and noise.[14] Furthermore, physics-informed neural networks (PINNs) have emerged as a promising direction. By incorporating physical laws, such as the acoustic wave equation, directly into the network's loss function, PINNs ensure that the denoising process respects the underlying wave propagation physics, which enhances generalization and produces more physically plausible results.[15]

Despite these advancements, significant challenges remain. Data-driven dictionary learning, if unconstrained, is prone to learning non-physical features that mimic noise, leading to incomplete noise suppression and signal damage. DL models, while powerful, often lack interpretability, and their performance can be unreliable when applied to data with characteristics different from the training set.[9,16,17] To address these issues, this paper proposes a physics-constrained sparse basis learning method for mixed noise suppression. This method constructs a joint optimization model that introduces a dip regularization term, penalizing components in the reconstructed signal that do not conform to local coherence. By simultaneously imposing a smoothness constraint on the dictionary atoms, the learned basis is guided to be more physically meaningful. A plane-wave destructor (PWD) filter is used to iteratively estimate and update the local dip field, ensuring that the physical constraint adapts to the progressively refined signal. Tests on synthetic and real data demonstrate that our method outperforms conventional techniques in suppressing complex mixed noise while preserving the integrity of the effective signal.

2. Materials and methods

2.1. Dip-constrained and gradient-optimized learning framework

The core idea of the novel prestack seismic data joint denoising framework proposed in this study is to combine the signal representation capability of sparse transforms with the dip attributes of effective signals. This integration aimed to achieve high-fidelity, effective signal reconstruction while simultaneously performing multi-type noise suppression.[18]

Prestack seismic data $Y \in \mathbb{R}^{N_t \times N_z}$ (where N_t represents the number of time samples and N_z represents the number of traces) can be expressed as the sum of effective seismic signals $X \in \mathbb{R}^{N_t \times N_z}$ and multiple types of noise $N \in \mathbb{R}^{N_t \times N_z}$:

$$Y=X+N \tag{I}$$

The primary goal of denoising was to estimate the effective signal X from the raw data Y. Within the framework of basis learning, we assume that the effective signal X can be approximately represented by a dictionary (set of basis functions) $D \in \mathbb{R}^{N_t \times N_z}$ and its corresponding sparse coefficient matrix $A \in \mathbb{R}^{N_t \times N_z}$:

$$X=\approx DA \tag{II}$$

Here, the column vectors of $D = [d_1,d_2...,d_{Nk})]$, denoted as d_j, are referred to as atoms or basis functions, and N_k is the number of atoms. Each column a_i of the sparse coefficient matrix A represents the sparse representation of the corresponding trace x_i under the dictionary D. The core challenge is to learn a dictionary D that provides a compact representation of the effective signal features and to solve for the corresponding sparse coefficients A.

Conventional basis function learning is achieved by solving the following optimization problem:

$$\min_{D,A} \left\| Y - DA \right\|_F^2 + \lambda_s R_s(A), \tag{III}$$

where $\|Y - DA\|_F^2 = \sum_{i,j}\left(Y_{i,j} - (DA)_{i,j}\right)^2$ represents the data fidelity term, which measures the error between the reconstructed data DA and the Raw data Y. $R_s(A)$ is the regularization term for the sparse coefficients A, used to introduce a sparsity prior. The L_1-norm regularization, i.e., $\mathcal{R}_s(A) = \|A\|_1 = \sum_{i,j}\left\|A_{i,j}\right\|$, is commonly used to induce sparse solutions, meaning that the information for each seismic trace can be represented by a linear combination of a few atoms from the dictionary.[6] $\lambda_s > 0$ represents the weighting balance between the data fidelity term and the sparse regularization term. The selection of an appropriate value for the hyperparameter λ_s is critical to the success of the denoising task, as it governs the trade-off between fitting the data and enforcing sparsity. A very small λ_s would cause the optimization to prioritize the data fidelity term, leading the model to fit the noisy data Y too closely and fail to remove noise. Conversely, a very large λ_s would heavily penalize non-sparse solutions, forcing the coefficient matrix A to be extremely sparse, at the risk of over-smoothing the data and removing important features of the effective signal. Therefore, an optimal λ_s must be chosen to ensure that the sparsity constraint is strong enough to separate noise, while the data fidelity term preserves the integrity of the underlying signal. The optimal value is data-dependent, influenced by factors such as the noise level, and is typically determined empirically.

In prestack data, effective reflection signals typically exhibit good spatial coherence and predictable dips within local regions. For instance, in common midpoint gathers or common offset gathers, reflection events possess specific kinematic characteristics. This coherence is a key feature that distinguishes signals from various interferences such as random noise, linear noise, anomalous amplitudes, and ground roll. To make the basis learning framework more suitable for seismic data denoising and to enhance denoising performance by incorporating physical meaning, this study introduced local dip attributes as a physical constraint within the learning framework.[16]

This research presents a dip regularization term to penalize components in the reconstructed signal DA that does not conform to local coherence. This constraint was built upon the local dip $P = \{p_{i,j}\}$ (the local dip at data point (i,j)). First, we defined a linear operator L_p, which depends on the local dip field P and is used to enhance signal smoothness along the dip direction or to suppress different components. Ideally, if the signal DA is perfectly aligned along the dip P, then the value of $L_p(DA)$ will be close to zero. This constraint term can be expressed as:

$$\text{will be close to } R_{coh}(DA, P) = \left\|L_p(DA)\right\|_F^2, \tag{IV}$$

where the local dip field P can be estimated, computed, and updated during the iterative process based on the current reconstructed signal DA, allowing this constraint to adaptively match the local structural features of the data.

This paper posits that the basic building blocks of effective signals (atoms in the dictionary D) inherently possess certain physical properties. For example, they should exhibit smoothness and band-limited characteristics, rather than containing excessive high-frequency noise or irregular oscillations. To ensure that the learned atoms are more physically meaningful, this method imposes a smoothness constraint on the dictionary D itself by penalizing its gradient:

$$\min_{D,A,P} j(D, A, P) = \left\|Y - DA\right\|_F^2 + \lambda_s \left\|A\right\|_1 + \lambda_{coh}\left\|L_p(DA)\right\|_F^2 + \lambda_{atom}\left\|\nabla D\right\|_F^2 \tag{V}$$

where λ_s, λ_{coh}, and λ_{atom} are regularization parameters used to balance the weights of different constraints. The local dip field P, as part of the regularization term, reflects the model's adaptability to data characteristics. The objective function above, by jointly optimizing the dictionary D, sparse coefficients A, and physical parameter P, yields a solution that fits the effective signal while satisfying both sparsity and physical priors.

Given that the objective function $j(D, A, P)$ is non-convex with respect to D, A, and P, we employed an alternating iterative optimization strategy that decomposes the problem into the following four sub-steps:

Sub-step one: Initialization

(i) Initialize dictionary $D^{(0)}$: Randomly select data patches from the raw data Y or use Ricker wavelets for initialization

(ii) Initialize sparse coefficients $A^{(0)}$: Use a zero matrix or small random values

(iii) Initialize local dip field $P^{(0)}$: Estimate from the raw data Y using the PWD method

(iv) Set current iteration $t = 0$ and maximum iterations T_{max}.

Sub-step two: Updating sparse coefficients

(i) Fix $D^{(t)}$ and $P^{(t)}$ and establish the objective function for solving A:

$$A^{(t+1)} = \underset{A}{argmin}\left\|Y - D^{(t)}A\right\|_F^2 + \lambda_s\left\|A\right\|_1 + \lambda_{coh}\left\|L_{P^{(t)}}(D^{(t)}A)\right\|_F^2 \tag{VI}$$

(ii) The function above is an L_1-norm minimization problem with a quadratic regularization term. Assuming $L_{P(t)}$ is a linear operator, let $Q^{(t)} = \mathcal{L}_{P^{(t)}}D^{(t)}$. Then, this subproblem can be rewritten as:

$$A^{(t+1)} = \underset{A}{argmin}\left\|Y - D^{(t)}A\right\|_F^2 + \lambda_s\left\|A\right\|_1 + \lambda_{coh}\left\|Q^{(t)}A\right\|_F^2 \tag{VII}$$

This can be transformed into:

$$A^{(t+1)} = \underset{A}{argmin} \left\| \begin{pmatrix} Y \\ 0 \end{pmatrix} - \begin{pmatrix} D^{(t)} \\ \sqrt{\lambda_{coh}} Q^{(t)} \end{pmatrix} A \right\|_F^2 + \lambda_s \quad \text{(VIII)}$$

This problem can be efficiently solved using methods such as the fast iterative shrinkage-thresholding algorithm or the alternating direction method of multipliers.

Sub-step three: Updating the dictionary

(i) Fix $A^{(t+1)}$ and $P^{(t)}$ and solve the subproblem for D:

$$\mathrm{D}^{(t+1)} = \underset{\mathrm{D}}{\mathrm{argmin}} \left\| \mathrm{Y} - \mathrm{DA} \right\|_{\mathrm{F}}^{(t)2} + \lambda_{\mathrm{coh}} \left\| \mathcal{L}_{\mathrm{P}^{(t)}} (\mathrm{DA}^{(t+1)}) \right\|_{\mathrm{F}}^2 + \lambda_{\mathrm{atom}} \left\| \nabla \mathrm{D} \right\|_{\mathrm{F}}^2 \quad \text{(IX)}$$

(ii) The equation above is a quadratic programming problem with respect to D. If optimized column-by-column d_k, it can be simplified as:

$$\underset{D}{min} \left\| \begin{pmatrix} Y \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} I \otimes (A^{(t+1)})^T \\ \sqrt{\lambda_{coh}} M_{P^{(t)},\, A^{(t+1)}} \\ \sqrt{\lambda_{atom}} \mathrm{N} \end{pmatrix} vec(D) \right\|_2^2 \quad \text{(X)}$$

where M and N are matrix forms of expressing $\left\| \mathcal{L}_{P^{(t)}} (DA^{(t+1)}) \right\|_F^2$ and $\left\| \nabla D \right\|_F^2$ as quadratic forms with respect to $vec(D)$; $\otimes$ denotes the Kronecker product; and $vec(.)$ is the vectorization operator. This results in a large-scale least squares problem that can be solved using iterative methods such as gradient descent or the conjugate gradient method.[9]

Sub-step four: Local dip field update

Fix $D^{(t+1)}$ and $A^{(t+1)}$ to obtain the current effective signal estimate $X^{(t+1)}=D^{(t+1)}\, A^{(t+1)}$. Then, update the local dip field $X^{(t+1)}$ based on P:

$$P^{(t+1)} = \mathrm{PWD}(X^{(t+1)}) \quad \text{(XI)}$$

2.2. Plane-wave deconstruction filtering dip angle estimation

In the aforementioned constrained learning framework, the core of the physical constraint lies in the quantification and utilization of seismic signal local coherence. The PWD filter, proposed and developed by Sergey Fomel, cannot only be used to estimate the local dip field but also directly serve as a coherence constraint operator, providing strong support for this objective.[17]

The PWD theory assumes that, within a local time-space window, seismic data can be approximately viewed as a superposition of a series of plane waves. A 2D plane wave can be expressed as:

$$d(t,x) = f(t-\sigma x), \quad \text{(XII)}$$

where σ represents the local dip of the plane wave. PWD is essentially a steerable prediction-error filter. The prediction error at the filter's output is minimized when the correct local dip is applied. Any components that do not conform to this local plane wave model (such as various types of noise) cannot be effectively predicted and thus manifest as larger energy at the filter output. Assuming a 2D seismic data $d(t,x)$, the theory aims to predict the value of $d(t,x)$, based on information from neighboring traces. According to the plane wave assumption, the following differential relationship is derived:

$$\frac{\partial d}{\partial x} + \sigma \frac{\partial d}{\partial t} = 0 \quad \text{(XIII)}$$

The above equation indicates that the directional derivative along the plane wave direction (t,x) domain is zero. PWD is the discrete realization of this differential operator. A first-order PWD operator can be used to predict the value at a central point $d_{i,j.}$ Its predicted value $d_{Ei,j}$ is calculated from two neighboring points $d_{i,j-1}$ and $d_{i,j+1}$ in the x-direction. To introduce the dip σ into the prediction, a shift in the time direction needs to be considered:

$$d_{Ei,j} = \frac{1}{2}\left[d_{i-round(\sigma),j-1} + d_{i+round(\sigma),j+1} \right] \quad \text{(XV)}$$

To address the precision issue caused by the integer shifts in the above equation, Fomel proposed more accurate Taylor expansion and finite-difference methods:

$$e_{i,j} = d_{i,j} - [c_{-1}(\sigma) d_{i,j-1} + c_1(\sigma) d_{i,j+1}], \quad \text{(XV)}$$

where $e_{i,j}$ represents the prediction error of $d_{i,j}$. c_{-1} (σ) and c_1 (σ) are functions of the local dip σ, used to perform data interpolation or extrapolation along the dip direction. In practical applications, a separable approximation is commonly used, where a three-point PWD operator F_σ applied to a data point $d_{i,j}$ can be approximated as: can be approximated as

$$F_\sigma(d_{i,j}) \approx \frac{d_{i,j+1} - d_{i,j-1}}{2\Delta x} - \sigma_{i,j} D_t(d_{i,j}) \quad \text{(XVI)}$$

where D_t is a differential operator in the time direction, and $\sigma_{i,j}$ is the local dip at point (i,j). The output energy reflects the degree to which the data deviates from the local plane wave assumption. Conversely, this can be used to find a dip value σ that minimizes the output energy of the PWD filter. For each local window in the data, the optimal local dip σ is estimated by solving the following minimization problem:

$$\underset{\sigma}{min} \left\| \mathcal{F}_\sigma(d) \right\|_F^2, \quad \text{(XVII)}$$

where d represents the data within the local window, and F_σ is the PWD operator parameterized by the dip σ.

The PWD theory aligns well with the joint denoising framework proposed in this study, providing a concrete implementation for crucial steps of the algorithm. In our objective **Equation (VI)**, the local coherence constraint term is $\lambda_{coh}\left\|\mathcal{L}_P(DA)\right\|_F^2$. We define the PWD operator as L_P, so this constraint term becomes:

$$R_{coh}(DA,P)=\left\|F_{P\sigma}(DA)\right\|_F^2 \tag{XVIII}$$

where, $P\equiv\sigma(t,x)$ represents the local dip field required by the PWD, and F^{σ} is the PWD operator guided by this dip field. It constrains all components in the reconstructed signal DA that cannot be predicted by the local plane wave model.

Furthermore, for sub-step four predicted by the local plane **(Equation VI)**—this can be achieved by solving the PWD-based dip estimation problem:

$$P^{(t+1)}=\underset{P}{argmin}\left\|\mathcal{F}_P(X^{(t+1)})\right\|_F^2 \tag{XIX}$$

This process ensures that the dip field consistently aligns with the continuously improving signal estimation throughout the iterative process, thereby guiding the entire optimization toward clearer physical meaning and a more distinct signal structure.

2.3. Learning-based seismic data denoising framework

Unlike the training phase, where *D, A, and P* are optimized simultaneously, in the denoising phase, the dictionary D_{opt} and the dip field P_{opt} are treated as known optimal parameters. The objective function for solving the sparse coefficient matrix A_f is:

$$\underset{A}{min}\left\|Y-D_{opt}A\right\|_F^2+\lambda_s\left\|A\right\|_1+\lambda_{coh}\left\|L_{P_{opt}}(D_{opt}A)\right\|_F^2, \tag{XX}$$

where $\left\|Y-D_{opt}A\right\|_F^2$ is the data fidelity term, which ensures that the sparse representation, after reconstruction using the optimal dictionary D_{opt}, has minimal error with respect to the raw data Y, thus preserving the fidelity of the denoising process. The term $\lambda_s\left\|A\right\|_1$ is the sparse regularization term, encouraging the solution to be represented sparsely using only a few atoms from D_{opt}. $\lambda_{coh}\left\|\mathcal{L}_{P_{opt}}(D_{opt}A)\right\|_F^2$ is the dip constraint term, which uses the estimated dip field P_{opt} to enforce structural constraints on the denoised data, requiring that the final denoising result conforms to the local coherence structure defined by P_{opt}.

By solving the optimization problem in **Equation I**, we obtain the sparse coefficient matrix A_f. Combining it with D_{opt} yields the final denoising result:

$$X_f=D_{opt}A_f \tag{XXI}$$

The proposed method achieves a relative balance among data fidelity, sparse representation, and structural constraint. The estimated noise $N_e = Y–X_f$ includes interference components that are neither effectively represented by the dictionary nor conform to the local coherence constraint. The proposed method comprises two distinct phases within a single workflow, as illustrated in Figure 1: A learning phase and an application phase. The "iterative optimization" block constitutes the learning phase, during which the optimal dictionary (D_{opt}) and dip field (P_{opt}) are learned from the raw data. The subsequent steps form the application phase, in which these learned parameters are used to process the raw data once to obtain the final denoised result. Unlike the learning phase, where D, A, and P need to be optimized simultaneously, in the application phase, the dictionary D_{opt} and dip field P_{opt} are treated as known, optimal parameters.

3. Results

3.1. Synthetic data example

To validate the proposed method, we conducted comparative denoising experiments on three synthetic datasets (Blocks A, B, and C), derived from a complex physical model based on a block in Western China. The performance of our method was benchmarked against three techniques: shearlet transform, DL model (a classic supervised learning framework based on the denoising convolutional neural network), and traditional dictionary learning. The first dataset, Block A, was contaminated with strong anomalous amplitude interference, random noise, and coherent noise, as shown in the raw shot gather in Figure 2A. Figure 2 compares the denoising results, where the proposed method (Figure 2E) effectively removes vertical interference while preserving signal continuity, outperforming the shearlet (Figure 2B), DL (Figure 2C), and traditional dictionary learning (Figure 2D) methods, which exhibit residual noise or signal loss. The removed noise profiles are displayed in Figure 3. The results from the comparative methods show significant signal leakage (Figure 3A-C), whereas the noise removed by our method consists primarily of interference, with almost no effective signal components, demonstrating superior signal preservation (Figure 3D). Figure 4 presents the final constrained dip fields, where the result from our method (Figure 4E) exhibits weaker residual noise and better preservation of effective signal features compared to the raw data and other results (Figure 4A-D). Finally, the dictionary iteration process is shown in Figure 5. Compared to the initial dictionary (Figure 5A) and the traditional result (Figure 5B), the dictionary learned by the proposed method (Figure 5C) more effectively captures signal features while discarding noise elements.

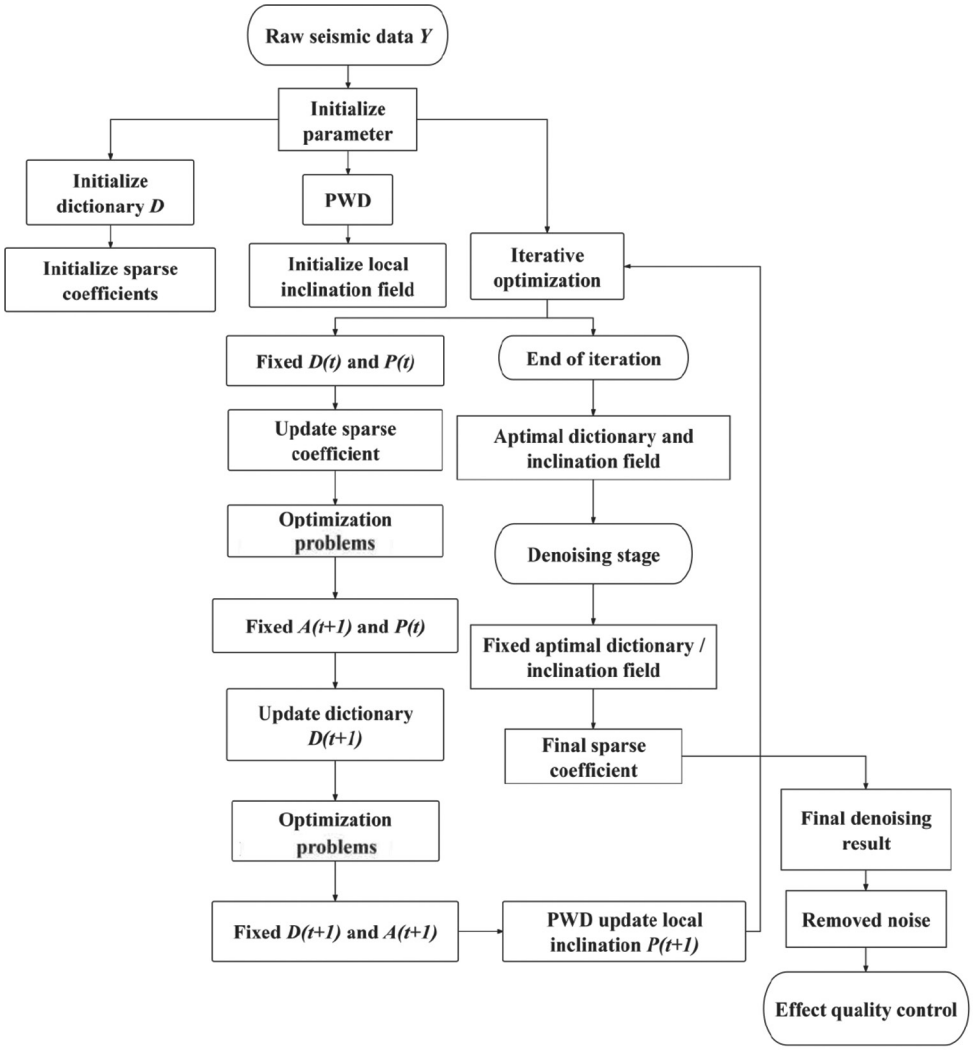

Figure 1. Flowchart of the proposed method
Abbreviation: PWD: Plane-wave destructor.

The second dataset, Block B, was characterized by strong ground roll, as depicted in Figure 6A. The results in Figure 6 demonstrate that while the benchmark methods (Figure 6B-D) struggled to suppress the ground roll, our proposed method (Figure 6E) achieved excellent multi-type noise removal while preserving the underlying signal. The removed noise sections in Figure 7 confirm this: While the other methods showed significant signal leakage (Figure 7A-C), our method successfully isolated the ground roll (Figure 7D). The corresponding dip fields and dictionary iterations are shown in Figures 8 and 9, respectively. The conclusions are consistent with those of the first experiment: In contrast to the dip fields of the raw data (Figures 8A), the shearlet result (Figures 8B), the deep learning result (Figures 8C), and the traditional dictionary learning result (Figures 8D), our method produced a much cleaner dip field (Figures 8E). Additionally, when compared with the initial dictionary (Figure 9A) and the result from traditional dictionary learning (Figure 9B), our method yielded a dictionary more representative of the true signal structure (Figure 9C). The third dataset, Block C, contained a complex mix of strong noise, including intermixed vertical amplitudes and coherent acquisition noise (Figure 10A). As illustrated in Figure 10, the comparative methods (Figure 10B-D) had a minimal effect on this complex noise, while the proposed method (Figure 10E) effectively resolved the issue. Figure 11 further depicts that the other techniques showed a mixture of noise and signal in the removed components (Figure 11A-C), whereas our method cleanly separated the complex noise structures (Figure 11D). These findings are further validated in Figure 12 and Figure 13. In contrast to the dip fields of the raw data (Figure 12A), the shearlet result (Figure 12B), the deep learning result (Figure 12C),

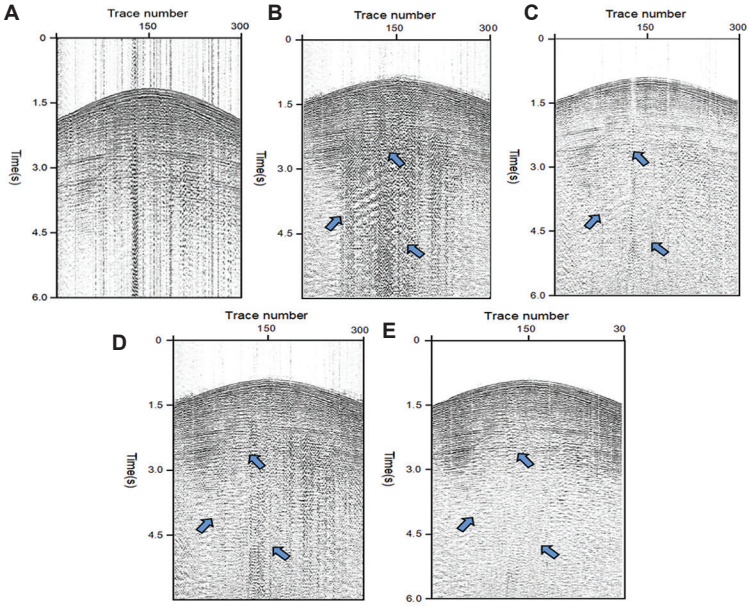

Figure 2. Raw data and denoised data of Block A. (A) Raw data. (B) Result using shearlet. (C) Result using deep learning. (D) Result using traditional dictionary learning. (E) Result using the proposed method.

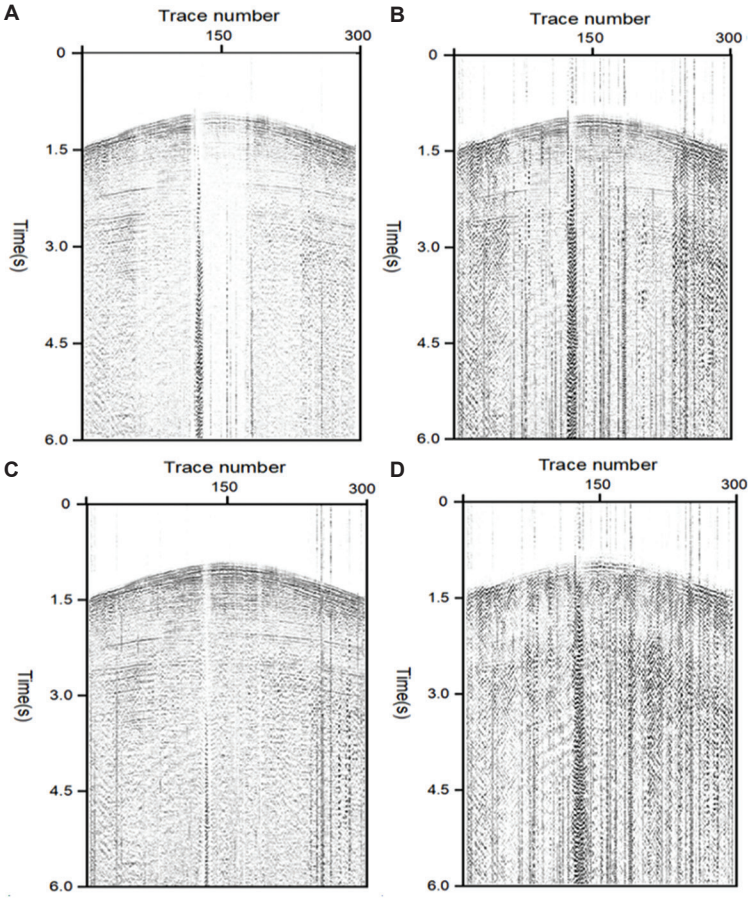

Figure 3. Removed noise using different methods for Block A. (A) Removed noise using shearlet. (B) Removed noise using deep learning. (C) Removed noise using traditional dictionary learning. (D) Removed noise using the proposed method.

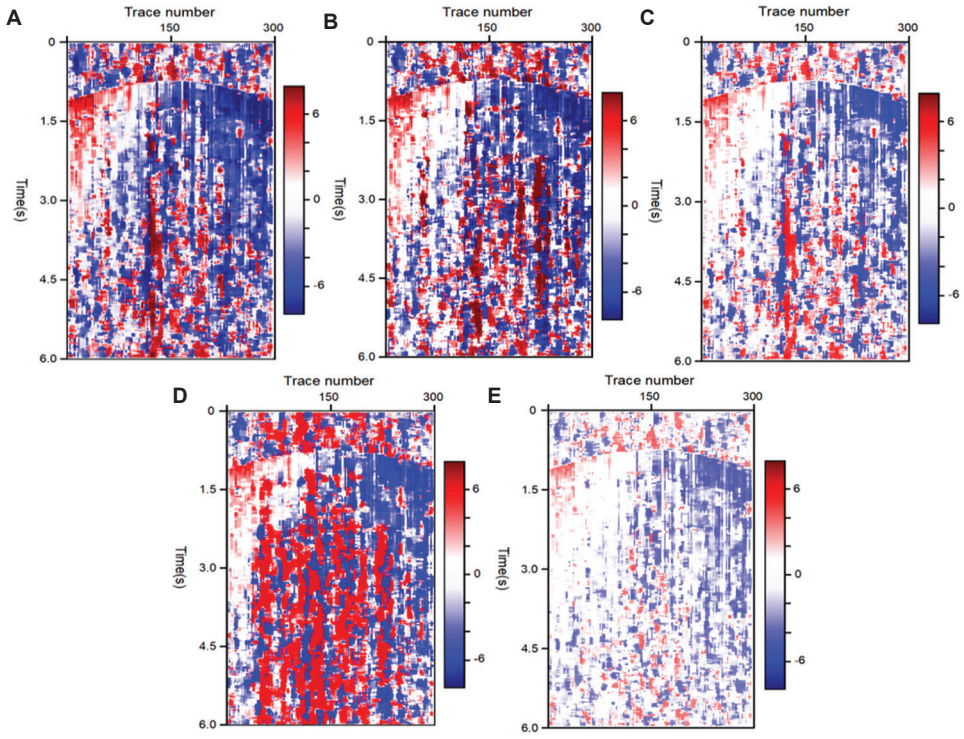

Figure 4. Dip fields of raw data and denoised data for Block A. (A) Dip field of raw data. (B) Dip field of result using shearlet. (C) Dip field of result using deep learning. (D) Dip field of result using traditional dictionary learning. (E) Dip field of result using the proposed method.

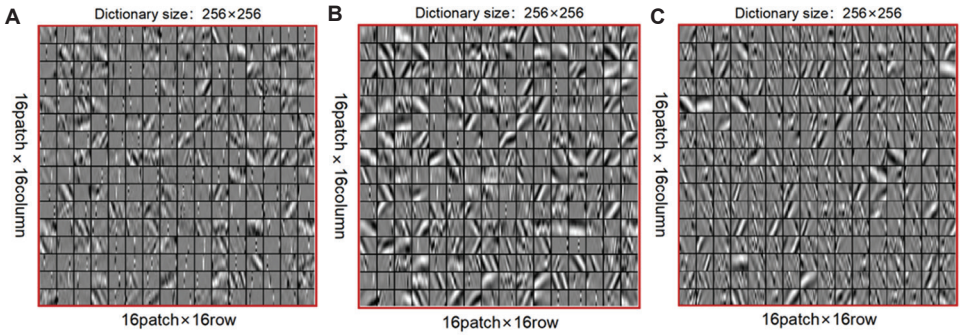

Figure 5. Initial dictionary and final learned dictionary of Block A. (A) Initial dictionary. (B) Result using traditional dictionary learning. (C) Result using the proposed method.

and the traditional dictionary learning result (Figure 12D), our method yields a cleaner final dip field (Figure 12E). Similarly, when compared with the initial dictionary (Figure 13A) and the result from traditional dictionary learning (Figure 13B),our method produces a more signal-focused dictionary (Figure 13C). Finally, the stacked sections for all three blocks are presented. For Block A (Figure 14), Block B (Figure 15), and Block C (Figure 16), the stacks processed by our method consistently demonstrated significant improvements in signal-to-noise ratio and continuity of geological events compared to the raw data and the results from the benchmark methods. In all cases, weak signals previously masked by strong noise were effectively recovered, highlighting the practical applicability of the proposed approach.

3.2. Real data example

To further validate the effectiveness and applicability of our method, we processed real seismic data from a work area in Western China. The performance was benchmarked against a DL method and a conventional industrial workflow.

A raw shot gathered from the dataset is shown in Figure 17A, which is heavily contaminated by severe ground

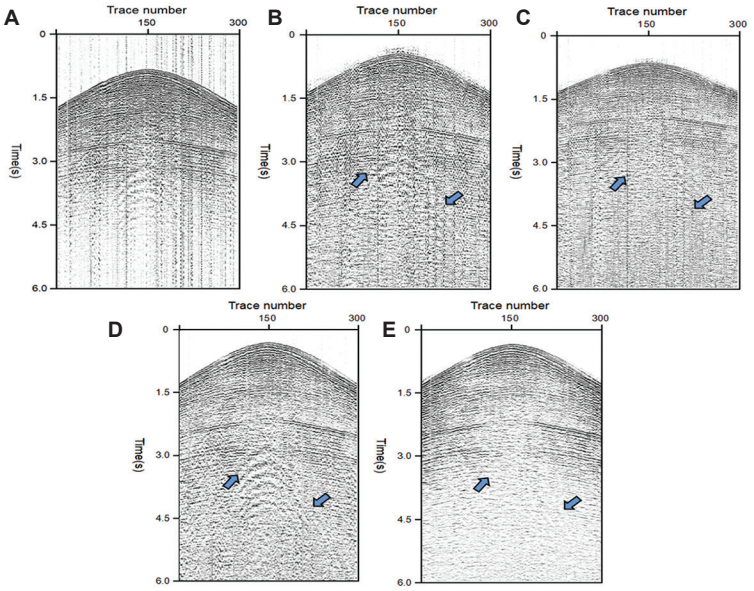

Figure 6. Raw data and denoised data for Block B. (A) Raw data. (B) Result using shearlet. (C) Result using deep learning. (D) Result using traditional dictionary learning. (E) Result using the proposed method.

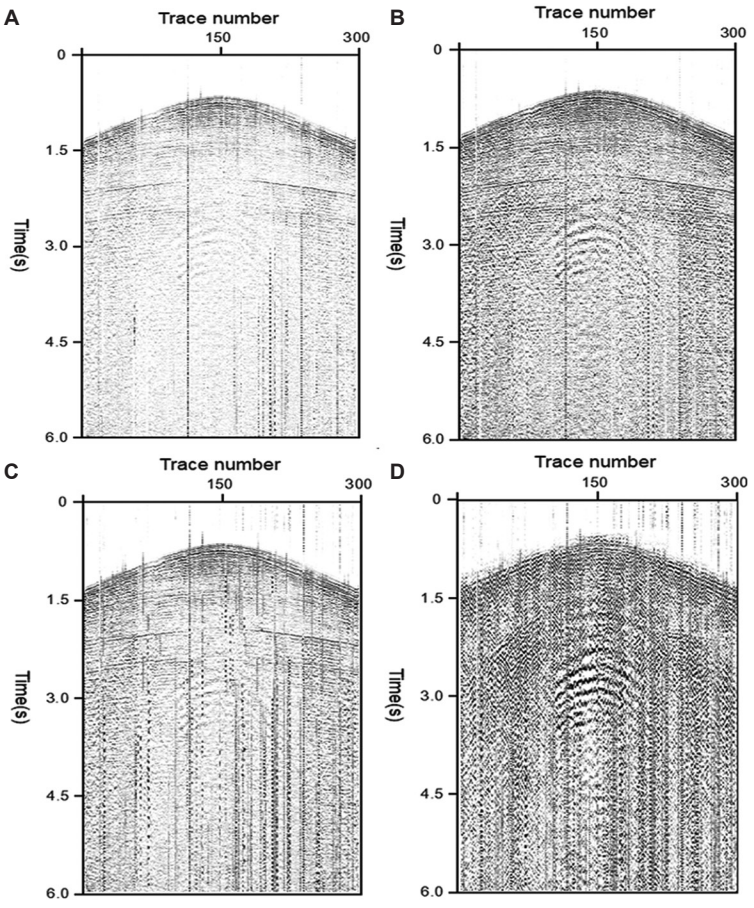

Figure 7. Removed noise using different methods for Block B. (A) Removed noise using shearlet. (B) Removed noise using deep learning. (C) Removed noise using traditional dictionary learning. (D) Removed noise using the proposed method.

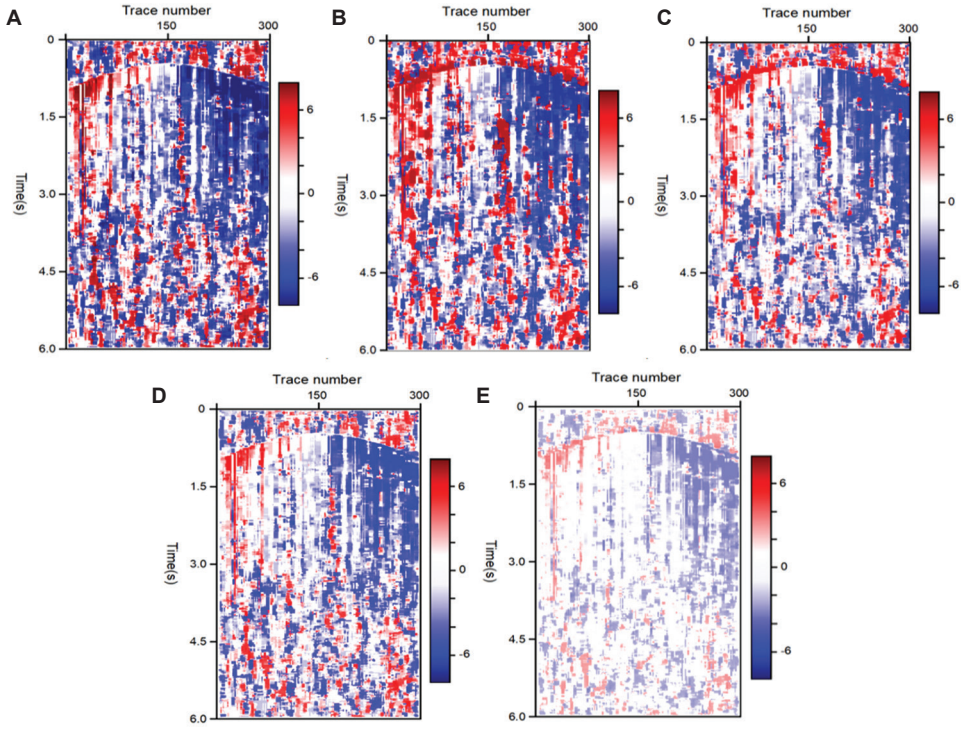

Figure 8. Dip fields of raw and denoised data from Block B. (A) Dip field of raw data. (B) Dip field of result using shearlet. (C) Dip field of result using deep learning. (D) Dip field of result using traditional dictionary learning. (E) Dip field of result using the proposed method.

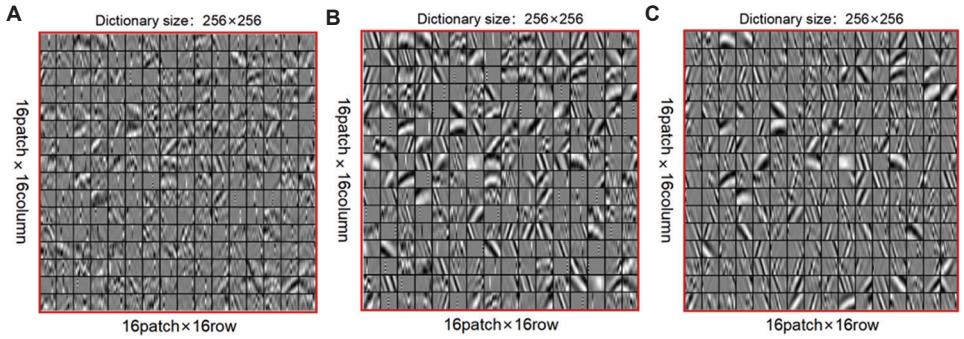

Figure 9. Initial dictionary and final learned dictionary of Block B. (A) Initial dictionary. (B) Result using traditional dictionary learning. (C) Result using the proposed method.

roll and random noise. This results in a low signal-to-noise ratio, where effective signals are obscured. Figure 17 presents the denoising results, displaying that all three methods removed a substantial amount of noise (Figure 17B-D). For a more detailed comparison of signal preservation, a partial enlargement is provided in Figure 18. In contrast to the raw data (Figure 18A), the DL method produced a cleaner result but with subtle smearing along the reflections (Figure 18B), and the conventional workflow left noticeable residual noise and compromised the continuity of reflection events (Figure 18C). The result from our proposed method (Figure 18D), however, shows superior noise removal while preserving signal integrity. The noise profiles for each method are depicted in Figure 19A-C. The coherence plots[19-21] for the DL method (Figure 19D) and the conventional workflow (Figure 19E) exhibit higher coherence values along noise and main reflection events, indicating weaker denoising and poorer signal preservation. The plot for our method (Figure 19F) demonstrates significantly lower correlation between the removed noise and the denoised result, confirming higher-fidelity separation of signal from noise. Finally, we evaluated the impact of denoising on seismic imaging by comparing stacked sections for two sub-regions. For region A, shown in Figure 20, the stacked

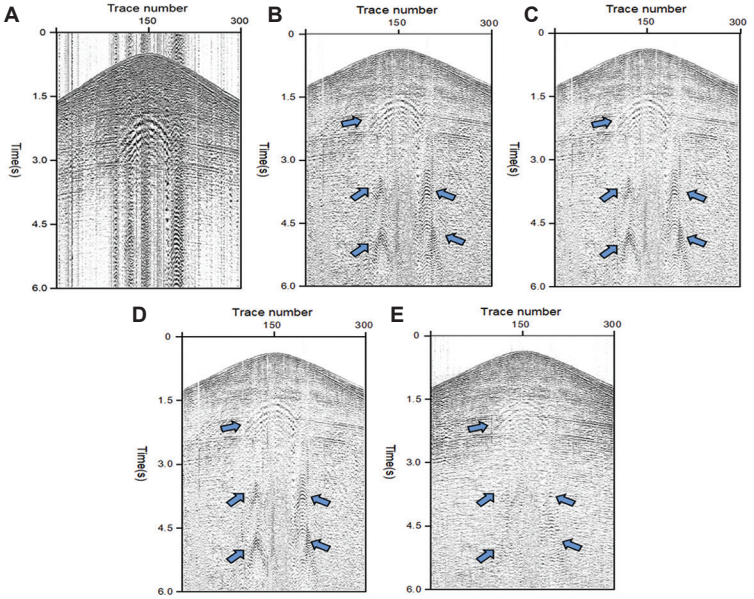

Figure 10. Raw data and denoised data of Block C. (A) Raw data. (B) Result using shearlet. (C) Result using deep learning. (D) Result using traditional dictionary learning. (E) Result using the proposed method.

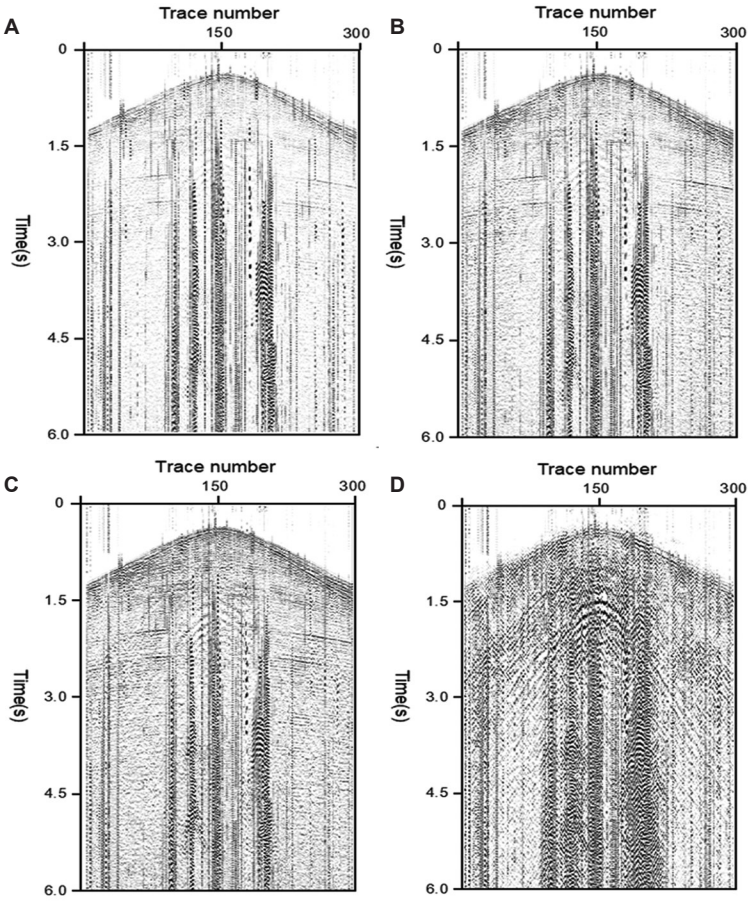

Figure 11. Removed noise using different methods for Block C. (A) Removed noise using shearlet. (B) Removed noise using deep learning. (C) Removed noise using traditional dictionary learning. (D) Removed noise using the proposed method.

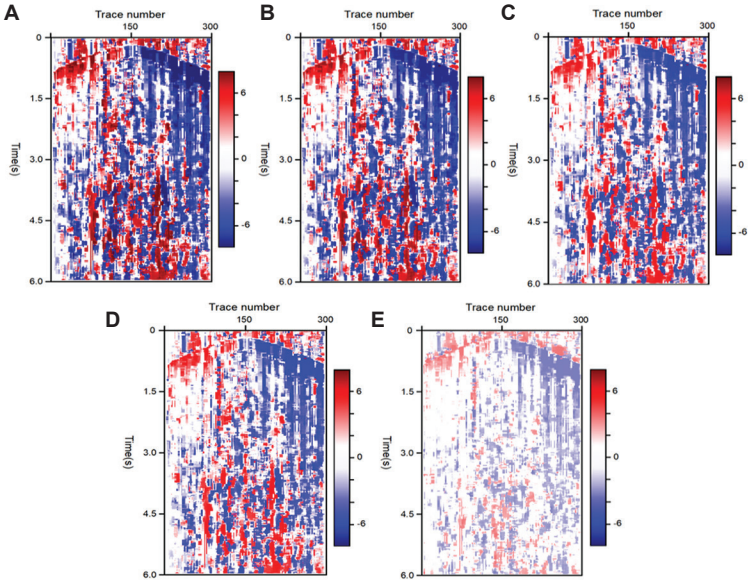

Figure 12. Dip fields of raw data and denoised data from Block C. (A) Dip field of raw data. (B) Dip field of result using shearlet. (C) Dip field of result using deep learning. (D) Dip field of result using traditional dictionary learning. (E) Dip field of result using the proposed method.

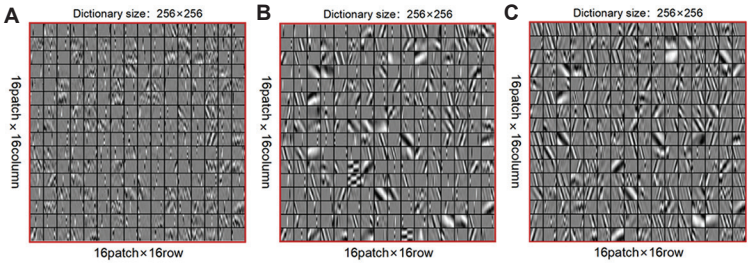

Figure 13. Initial dictionary and final learned dictionary for Block C. (A) Initial dictionary. (B) Result using traditional dictionary learning. (C) Result using the proposed method.

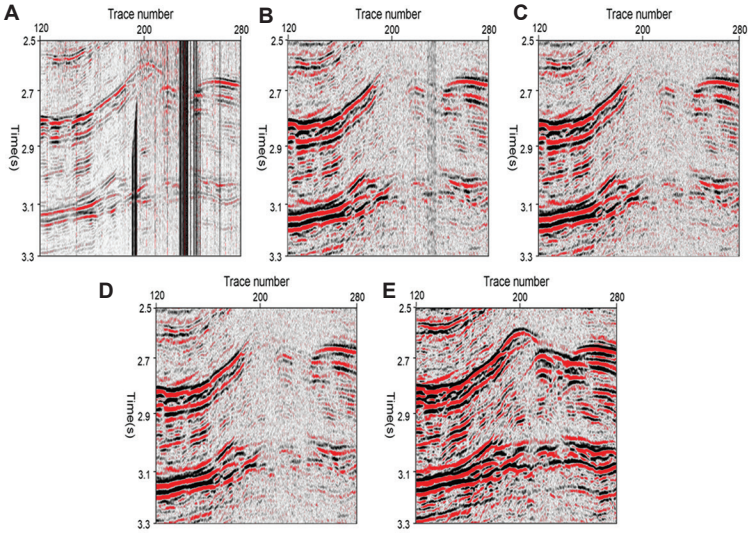

Figure 14. Stacked sections comparing raw data, traditional method results, and the proposed method for Block A. (A) Stack of raw data. (B) Stack of shearlet-denoised data. (C) Stack of deep learning-denoised data. (D) Stack of dictionary learning-denoised data. (E) Stack of proposed method-denoised data.

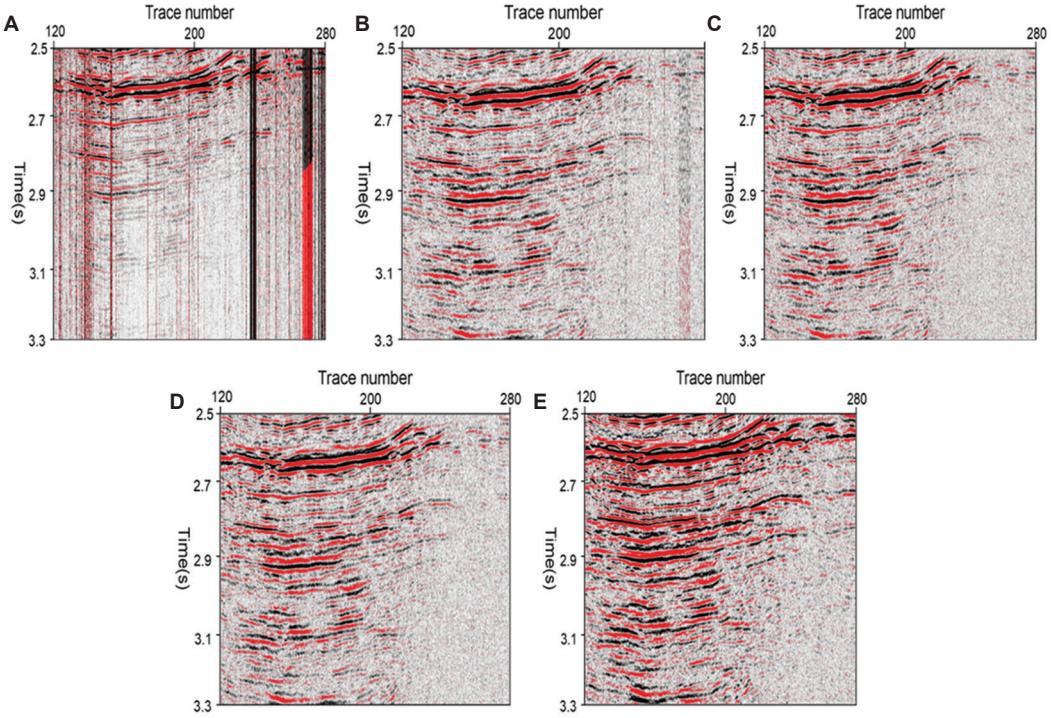

Figure 15. Stacked sections comparing raw data, traditional method, and the proposed method for Block B. (A) Stack of raw data. (B) Stack of shearlet-denoised data. (C) Stack of deep learning-denoised data. (D) Stack of dictionary learning-denoised data. (E) Stack of proposed method-denoised data.

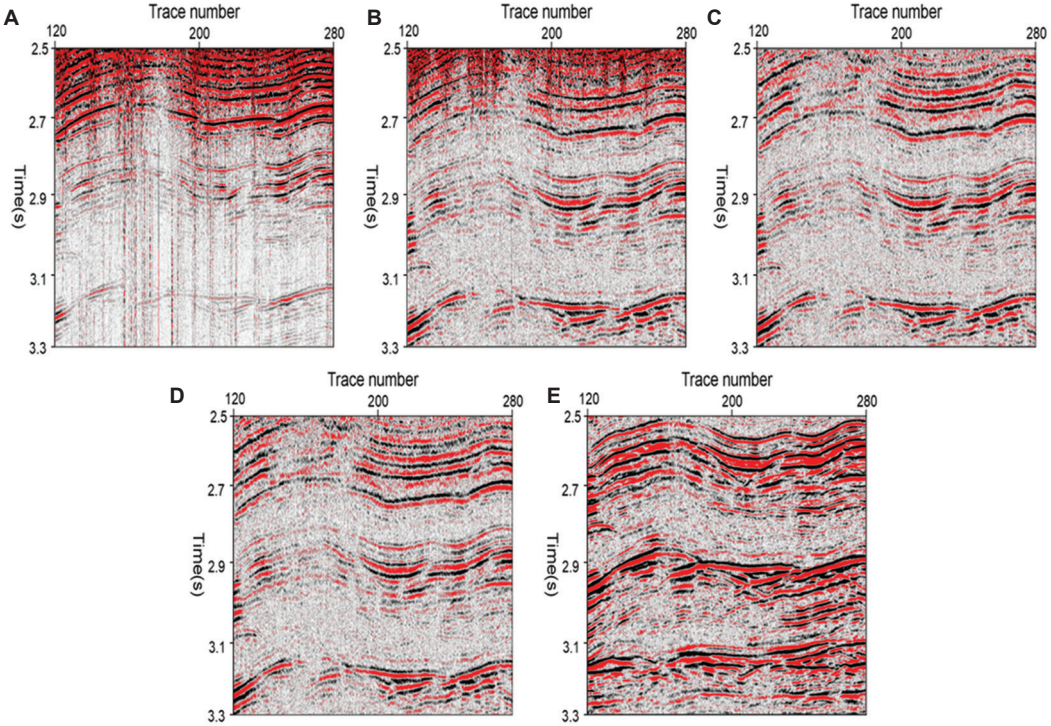

Figure 16. Stacked sections comparing raw data, traditional method, and the proposed method for Block C. (A) Stack of raw data. (B) Stack of shearlet-denoised data. (C) Stack of deep learning-denoised data. (D) Stack of dictionary learning-denoised data. (E) Stack of proposed method-denoised data.

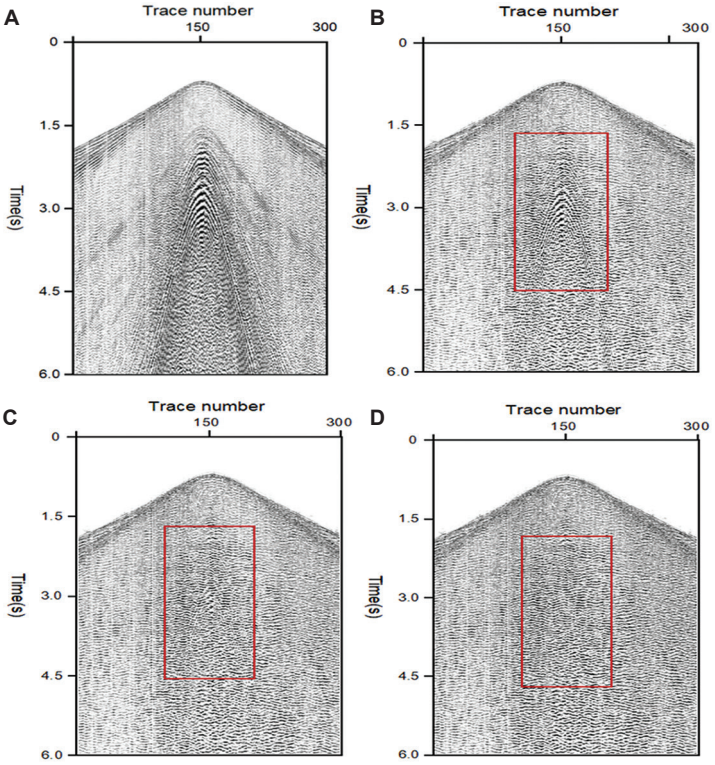

Figure 17. Raw data and denoised results. (A) Raw data. (B) Result using deep learning. (C) Result using conventional industrial workflow. (D) Result using the proposed method.

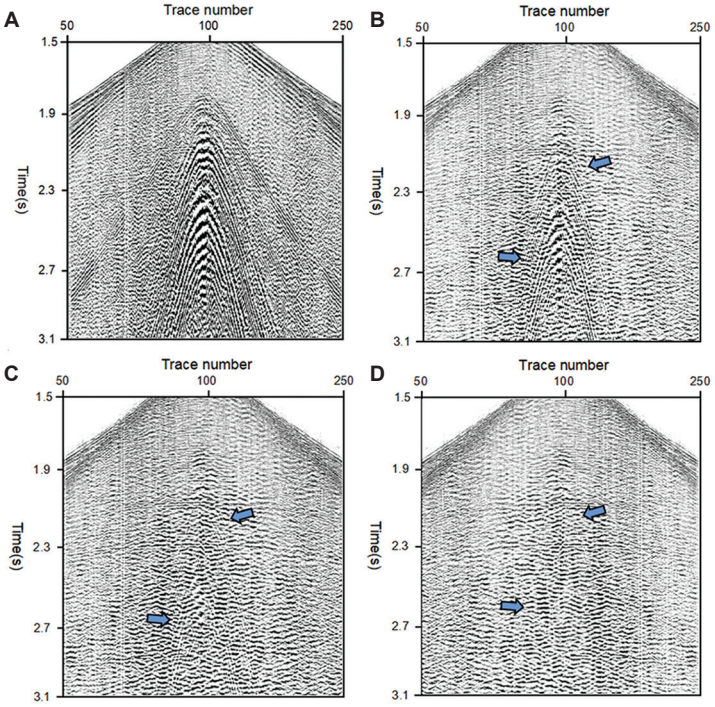

Figure 18. Partial enlarged image of raw data and denoised data. (A) Raw data. (B) Result using deep learning. (C) Result using conventional industrial workflow. (D) Result using the proposed method.

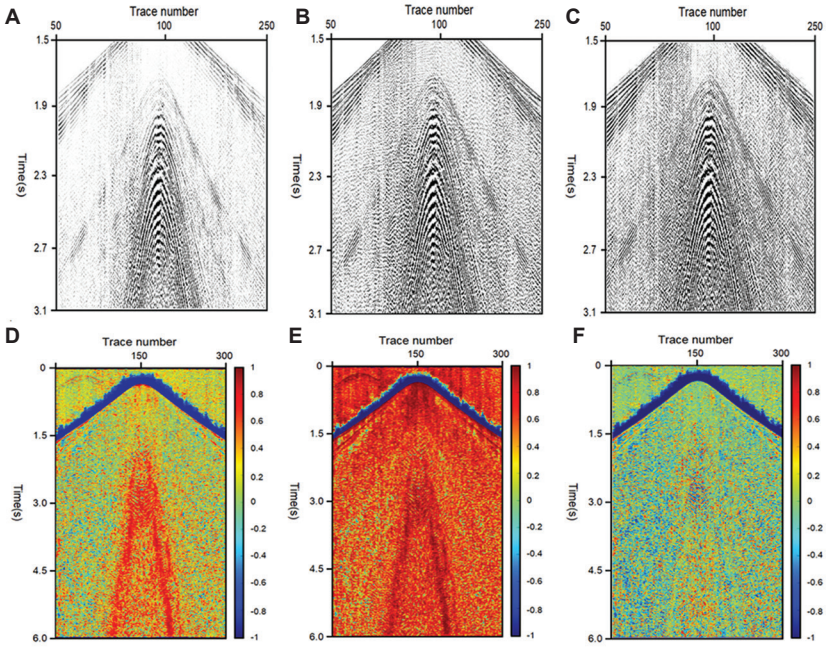

Figure 19. Removed noise and local coherence analysis. (A) Removed noise using deep learning. (B) Removed noise using a conventional industrial workflow. (C) Removed noise using the proposed method. (D) Local coherence between the removed noise and the denoised result using deep learning. (E) Local coherence between the removed noise and the denoised result using a conventional industrial workflow. (F) Local coherence between the removed noise and the denoised result using the proposed method and raw data.

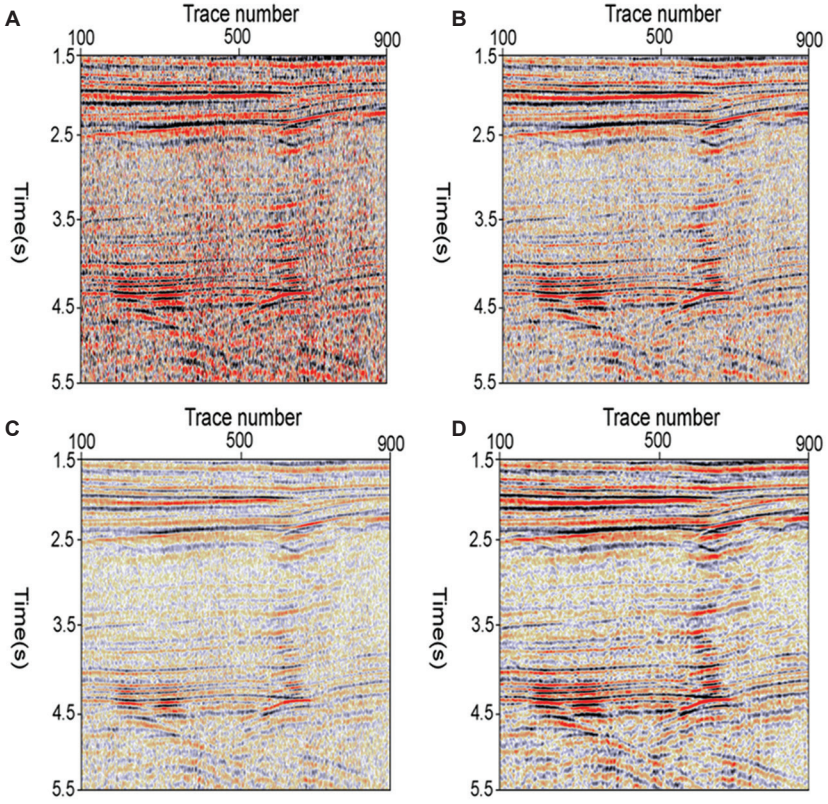

Figure 20. Stacked sections of raw data and denoised data in Region A. (A) Raw data. (B) Result using deep learning. (C) Result using conventional industrial workflow. (D) Result using the proposed method.

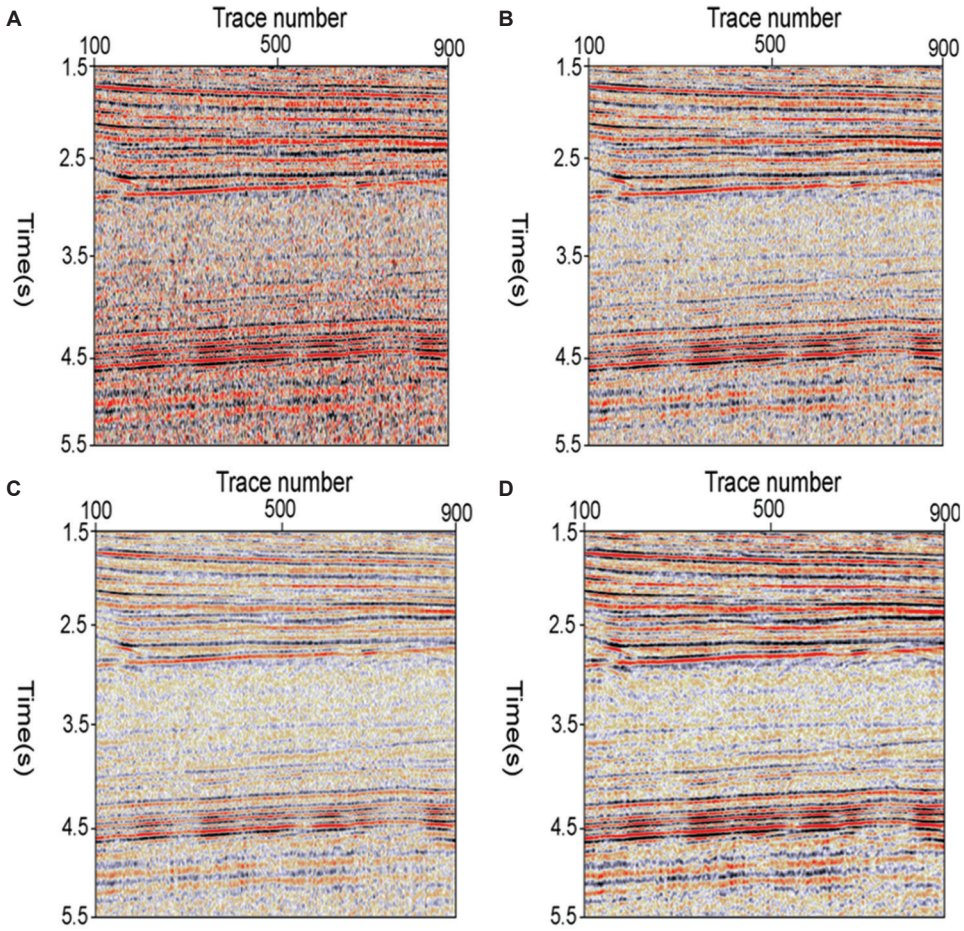

Figure 21. Stacked section of raw data and denoised data in Region B. (A) Raw data. (B) Result using deep learning. (C) Result using conventional industrial workflow. (D) Result using the proposed method.

section from the raw data (Figure 20A) suffers from low SNR and poor reflector continuity. While the results from deep learning (Figure 20B) and the conventional industrial workflow (Figure 20C) offer improvements, the result from the proposed method (Figure 20D) demonstrates the most significant enhancement, with clearer, more continuous reflectors and more prominent structural features such as faults and pinch-outs. A similar conclusion is drawn from the stacked results for Region B, presented in Figure 21. Compared to the raw data stack (Figure 21A) and the results from both deep learning (Figure 21B) and the conventional industrial workflow (Figure 21C), the proposed method's result (Figure 21D) again exhibits substantial improvement. In both regions, our method effectively recovered weak signals previously masked by strong noise, confirming its superior capability and practical value.

4. Discussion

The physics-constrained sparse basis learning approach for seismic data processing holds significant potential for future research. Future research will focus on exploring more advanced methods of incorporating physical attributes, such as geological models, velocity fields, or wavefield propagation theories. These additions could further enhance the recognition and preservation of valid seismic signals. Another promising direction involves integrating the powerful feature extraction capabilities of DL with the theoretical strengths of sparse representation. Hybrid models that combine these elements could lead to more efficient and higher-fidelity adaptive seismic data processing while maintaining physical interpretability.

5. Conclusion

In this study, we proposed a physics-constrained sparse basis learning method to address the critical challenge of suppressing complex, mixed noise in seismic data without damaging effective signals. The primary advantage of our method lies in the integration of local dip information, derived from a PWD filter, as a physical constraint within the dictionary learning framework. This innovation

effectively overcomes a key limitation of traditional data-driven approaches by preventing the learned basis from incorporating non-physical, noise-like features, thereby ensuring high-fidelity signal preservation. Our extensive experiments on both synthetic and real data demonstrated that this approach provides superior suppression of mixed noise—including anomalous amplitudes, ground roll, random, and coherent noise—compared to conventional techniques and other learning-based techniques. Ultimately, the enhanced clarity and continuity of reflectors in the final seismic images confirm the practical value of our method for improving the delineation of geological structures and recovering weak signals.

Acknowledgments

None.

Funding

This work was supported by the China National Petroleum Corporation Science and Technology Special Project "Research on Risk Exploration Targets and Engineering Technology Breakthroughs in the Qaidam Basin, Including Field Trials" (2023YQX10108).

Conflict of interest

Huailiang Li is an Editorial Board Member of this journal but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. The authors declare that they have no competing interests.

Author contributions

Conceptualization: Deying Wang, Huailiang Li
Formal analysis: Huailiang Li, Longjiang Kou
Funding acquisition: Yongsheng Wang
Investigation: Kai Zhang, Wenqing Liu
Methodology: Deying Wang, Longjiang Kou
Validation: Yongsheng Wang
Writing–original draft: Deying Wang, Huailiang Li
Writing–review & editing: Kai Zhang, Wenqing Liu

Availability of data

The data presented in this study are available from the corresponding author upon reasonable request.

References

1. Zhang H, Chen, X, Li H, *et al*. 3D seismic data denoising approach based on curvelet transform. *Oil Geophys Prospect.* 2017;52(2):226-232.

2. Liu Y, Fomel S, Liu C, Wang D, Liu G, Feng X. High-order seislet transform and its application of random noise attenuation. *Chin J Geophys.* 2009;52(8):2142-2151.

   doi: 10.3969/j.issn.0001-5733.2009.08.024

3. Zhang H, Chen X, Yang H. Optimistic wavelet basis selection in seismic signal noise elimination. *Oil Geophys Prospect.* 2011;46(1):70-75.

4. Wrinch D, Jeffreys H. On certain fundamental principles of scientific inquiry. *Philos Mag.* 1921;42(250):369-390.

   doi: 10.1080/14786442108633773

5. Häuser M, Ma J. Interpolating seismic data via the POCS method based on shearlet transform. *J Geophys Eng.* 2012;9(3):852-861.

6. Herrmann FJ. Curvelet-based seismic data processing: A multiscale and nonlinear approach. *Geophysics*. 2006;71(4):E11-E25.

7. Herrmann FJ, Hennenfent G, Moghaddam PP. Compressive seismic data acquisition: An overview. *Geophysics*. 2008;73(1):R23-R33.

8. Huang X. A simulation of acquisition design and data processing for offshore compressive sensing seismic. *Oil Geophys Prospect.* 2020;55(2):248-256.

9. Starck JL, Candès EJ, Donoho DL. The curvelet transform for image denoising. *IEEE Trans Image Process.* 2002;11(6):670-684.

   doi: 10.1109/tip.2002.1014998

10. Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process.* 2006;54(11):4311-4322.

    doi: 10.1109/TSP.2006.881199

11. Engan K, Aase SO, Husoy JH. Frame Based Signal Compression using Method of Optimal Directions (MOD). In: *1999 IEEE International Symposium on Circuits and Systems, ISCAS'99*. Vol. 4. United States: IEEE; 1999. p. 244-247.

    doi: 10.1109/ISCAS.1999.779928

12. Liu L, Liu Y, Liu C, Zheng Z. Iterative seismic random noise suppression method based on compressive sensing. *Chin J Geophys.* 2021;64(12):4629-4643.

    doi: 10.6038/cjg2021P0045

13. Zhang K, Zuo W, Chen Y, Meng D, Lei L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans Image Process*. 2017; 26(7):3142-3155.

    doi: 10.1109/TIP.2017.2662206

14. Kaur R, Waldeland AU, Alaei R, Solberg AHS. Self-supervised learning for seismic denoising. *IEEE Trans Geosci Remote Sens.* 2022;60:1-12.

15. Song C, Alkhalifah T, Riyanti F. A physics-informed deep learning approach for denoising seismic data. *Geophysics*. 2022;87(5):V239-V250.

16. Sacchi MD, Ulrych TJ, Walker CJ. Interpolation and extrapolation using a high-resolution discrete Fourier transform. *IEEE Trans Signal Process.* 1998;46(1):31-38.

doi: 10.1109/78.651165

17. Yao Z, Sun C, Li H, Yang A. Time-variant wavelet extraction and seismic reflectivity inversion based on basis pursuit. *Oil Geophys Prospect.* 2019;54(1):137-144.

18. Candès EJ, Donoho DL. Curvelets-a new multiresolution representation for visual information. In: *Wavelet Applications in Signal and Image Processing VIII.* Vol. 4119. United States: SPIE; 2000. p. 248-262.

19. Chen Y, Ma J, Fomel S. Seismic noise attenuation using a deep-learning approach with a local-correlation-based structural constraint. *IEEE Trans Geosci Remote Sens.* 2020;58(7):4668-4679.

20. Zhang L, AlRegib G. SSIM-based quality control for seismic processing. *Geophysics*. 2021;86(3):V235-V248.

21. Amini H, Javaherian A. A review of seismic attributes. *J Appl Geophys.* 2022;205:104764.