

Improving sparse representation with deep learning: A workflow for separating strong background interference

Dawei Liu¹, Wei Wang², Xiaokai Wang³, Zhensheng Shi³, Mauricio D. Sacchi⁴, and Wenchao Chen³

ABSTRACT

Revealing hidden reservoirs that are severely shielded by strong background interference (SBI) is critical to subsequent refined interpretation. To enhance the characterization of these reservoirs, current interpretation workflows merge multiple attribute information, necessitating intensive human expertise. As an alternative, we regard SBI suppression as a signal separation problem and develop a workflow to suppress SBI by cascading a sparse representation method and deep learning. SBI has coherent morphological characteristics in seismic sections; reservoir seismic responses, such as channels and karst caves, have a narrow spatial distribution, exhibiting abrupt morphological characteristics. As their morphologies differ, we select two 2D sparse representation

dictionaries to identify their individual components. Through the morphological component analysis (MCA) technique, we can obtain adequate SBI separation results. However, the MCA separation is inevitably limited because 2D dictionaries cannot adequately represent 3D structures, but 3D dictionaries are not viable due to computing constraints. As an extension, we use 3D deep learning to improve the separation results based on the 2D MCA results. Specifically, the network is fed with training samples from a region with better SBI suppression results obtained by the MCA method. After learning a direct mapping from noisy data to SBI, the network can improve the separation results and remove more SBI than the previous conventional method. Field data experiments demonstrate that our separation workflow successfully enhances reservoir structures after removing SBI.

INTRODUCTION

Reservoir prediction plays a significant role in oil field development planning because it provides an estimate of the remaining oil and gas reserves for production forecasts. However, strong reflection events, often existing in actual seismic profiles, obscure potential reservoir signatures. Strong reflections are formed by large wave impedance differences between the upper and lower strata of the reflection interface. For example, shale underlies thin channel sand reservoirs (Zhang et al., 2013), oil shale overlies a beach-bar sandstone reservoir (Liu et al., 2014), high-velocity hard layers sandwich softer coal seams (Zhang et al., 2015), etc. The amplitude of strong reflection layers is overwhelming in comparison with

nearby target reflections. Accordingly, strong reflection events shield reservoir responses manifested by weak reflection events. As a result of this shielding effect, reservoir prediction accuracy is greatly reduced. Such strong reflections can be identified as strong background interference (SBI). Therefore, separating SBI and enhancing weak reflections attract considerable critical attention.

Model-driven studies, which use handcrafted priors to separate strong-amplitude background interference, represent a growing field. They fall into five main classes: multiwavelet decomposition (Zhang et al., 2012; Guan et al., 2016), matching pursuit algorithms (Liu et al., 2014; Xu et al., 2019; Zhu et al., 2019), empirical mode decomposition (Wang et al., 2016; Chen et al., 2017a; Jiang et al., 2020, 2021), phase decomposition (Zhang et al., 2021b), and

Manuscript received by the Editor 23 March 2022; revised manuscript received 2 September 2022; published ahead of production 27 September 2022.

¹Xi'an Jiaotong University, School of Information and Communications Engineering, Xi'an, China and University of Alberta, Department of Physics, Edmonton, Alberta, Canada. E-mail: 409791715@qq.com.

²Shanghai Luoshu Investments Co. Ltd., Shanghai, China. E-mail: wwangcharles@qq.com.

³Xi'an Jiaotong University, School of Information and Communications Engineering, Xi'an, China. E-mail: xkwang@xjtu.edu.cn; 1303118188@qq.com; wenchchen@xjtu.edu.cn (corresponding author).

⁴University of Alberta, Department of Physics, Edmonton, Alberta, Canada. E-mail: msacchi@ualberta.ca.

© 2023 Society of Exploration Geophysicists. All rights reserved.

inverse Q filtering (Guo and Wang, 2004). Although these model-driven methods have achieved remarkable results in SBI elimination, the following three limitations have not yet been fully addressed. First, SBI removal quality heavily depends on the consistency between the adopted prior and real data distribution. In other words, we can obtain satisfactory results in the areas where the distribution of seismic data to be processed is consistent with our prior assumptions. However, seismic data are inherently nonstationary, which means that seismic data distribution in certain areas may deviate from our assumptions, thereby failing to remove SBI effectively. Second, model-driven approaches rely on appropriate parameter selection to achieve high-quality SBI separation. It is common in practice to preset fixed parameters for the sake of simplicity. However, considering the nonstationary characteristics of seismic data, setting a constant parameter may produce suboptimal separation results in some particular regions. Moreover, most model-driven methods suffer from highly variable sensitivity, further aggravating the problem. Third, the signal decomposition algorithms are computationally intensive, making it impractical to implement them for 3D SBI removal. In general, a 3D volume-based denoising method is superior to corresponding 1D or 2D denoising methods. If we could introduce a 3D SBI removal method with an acceptable computational cost, it would be beneficial to further improve the SBI separation performance in practice.

As a trending field in recent years, data-driven deep learning has been proposed as the solution to the preceding challenges. To date, deep learning has been successfully applied to many seismic data tasks, such as interpolation (Wang et al., 2019; Fang et al., 2021), denoising (Yu et al., 2019; Saad et al., 2021), velocity estimation (Park and Sacchi, 2020), and fault segmentation (Wu et al., 2019). As researchers dig deeper, it becomes more apparent that deep learning is not omnipotent and infallible. It has its own limitations, such as the black-box problem (Castelvecchi, 2016), model drift (Zhang et al., 2020b), and vulnerabilities to adversarial examples (Yuan et al., 2019). Therefore, recent attention has focused on the prospect of deploying deep learning algorithms in real-world scenarios. In other words, it is crucial to investigate how to maximize the advantages of deep learning in practical applications, while avoiding the flaws outlined previously. To our understanding, deep learning comprises two types of prospecting practical applications at present. The first category is accelerating conventional methods (Liu and Grana, 2019; Di et al., 2020; Liu et al., 2020b, 2021b; Wang et al., 2021). The second category pertains to improving conventional methods with deep learning. For example, integrating a deep learning denoiser with a conventional method, such as the plug-and-play method, can lead to enhanced interpolation results (Park et al., 2020; Zhang et al., 2020a).

Training sample selection is another reasonable way to improve a conventional method with deep learning. As previously discussed, a conventional method may result in inconsistent processing performance across the entire work area. Deep learning directly acquires prior knowledge from training samples. Accordingly, training a network using all labels generated by a conventional method enables similar processing performance. Furthermore, by selecting high-quality labels generated by a conventional method, the network can gain processing performance and then generalize this improvement to previously unseen regions. Therefore, training sample selection can facilitate the network surpassing the conventional method in the remaining unlabeled regions. For example, to eliminate seismic arc-

like imaging noise, Liu et al. (2020a) develop a fault confidence metric using the structural gradient tensor to exclude labels containing this type of noise. After training it with a set of rationally selected training samples, the network outperforms the conventional method for generating labels in arc-like imaging noise suppression. It confirms that the data processing capability of the network is learned from the training samples rather than by merely replicating the conventional method. This allows us to further refine the conventional method, which is used for producing training labels, by training a network using selected high-quality training samples.

Inspired by the preceding fact, we propose a novel SBI separation workflow by integrating a classic signal decomposition method and deep learning. Morphological component analysis (MCA) is a versatile signal decomposition technique based on sparse representation, which excels at simultaneously separating multiple waveforms pertaining to significant geologic targets from observed seismic data (Starck et al., 2005). Previous studies have proven the effectiveness of MCA in ground-roll attenuation (Xu et al., 2016; Chen et al., 2017b), linear noise attenuation (Guo et al., 2021), footprint noise suppression (Liu et al., 2021a), and so forth. On this foundation, we select the curvelet transform and the 2D stationary wavelet transform (2D-SWT) as sparse representation dictionaries of SBI and weak reflections, respectively, according to their waveform differences. After MCA separation, we can obtain preliminary but acceptable SBI separation results. Then, we move forward to further improving these results using deep learning. Specifically, shallow-layer seismic data, which are slightly shielded by SBI, are handpicked to build a training data set. The reason is that weak reflections are more evident in the shallow layer, and the first-step MCA separation can obtain better results than those in deep layers. Therefore, their features can be more easily identified by the network. In addition, as weak reflections and SBI exhibit different profile morphological features, the MCA separation model mentioned previously is implemented on 2D vertical sections. Undoubtedly, it loses 3D spatial information. We introduce a 3D network that can fully exploit the 3D spatial geologic structure of SBI and potential channel sand bodies. After feeding these selected training samples, the network can further improve the SBI suppression results. The field data experiment demonstrates that the proposed workflow is capable of accurately recovering weak-energy reflections and removing SBI with high accuracy.

THEORY

The proposed workflow for SBI separation is illustrated in detail in Figure 1. This section first introduces the MCA theory, including the dictionaries chosen for sparsely representing the SBI and weak useful reflections. Next, a training sample selection strategy and network architecture are presented.

Morphological component analysis

Following the sparse decomposition technique (Chen et al., 2001), Starck et al. (2005) propose the MCA algorithm for the separation of K components $\mathbf{s}_k \in \mathbb{R}^{N \times 1}$, $k = 1, \dots, K$. Each component \mathbf{s}_k exhibits distinct morphological characteristics and adding these components together synthesizes the signal:

$$\mathbf{s} = \mathbf{s}_1 + \dots + \mathbf{s}_K. \quad (1)$$

MCA separation is achieved by using several mathematical transforms as overcomplete dictionaries $\mathbf{D}_k \in \mathbb{R}^{N \times L_k}$ (typically $L_k \gg N$)

and then combining $\mathbf{D}_k (k = 1, \dots, K)$ to produce a combined dictionary \mathbf{D} for sparsely representing the original signal. Each dictionary \mathbf{D}_k can identify a specific morphological component \mathbf{s}_k of the signal, but \mathbf{s}_k and $\mathbf{D}_j (j \neq k)$ should be mutually incoherent. Concretely speaking, there are two additional technical prerequisites that must be fulfilled for the successful separation of various components.

Prerequisite 1: Any given component \mathbf{s}_k can be represented by a proper dictionary \mathbf{D}_k in a highly sparse manner:

$$\mathbf{x}_k^{\text{opt}} = \arg \min_{\mathbf{x}_k} \|\mathbf{x}_k\|_0 \quad \text{subject to } \mathbf{s}_k = \mathbf{D}_k \mathbf{x}_k \quad \text{for } k = 1, \dots, K, \quad (2)$$

where $\mathbf{x}_k \in \mathbb{R}^{L_k}$ denotes the coefficients computed by \mathbf{D}_k . Similarly, $\mathbf{x}_k^{\text{opt}}$ represents the sparse solution, i.e., $\|\mathbf{x}_k^{\text{opt}}\|_0 < N$. The l_0 norm counts the number of nonzero elements in a vector.

Prerequisite 2: Any other dictionary \mathbf{D}_j results in a nonsparse representation of $\mathbf{s}_k (j \neq k)$:

$$\forall \{j, k\} \in \{1, \dots, K\}, \quad \mathbf{x}_{jk}^{\text{opt}} = \arg \min_{\mathbf{x}_{jk}} \|\mathbf{x}_{jk}\|_0 \quad \text{subject to } \mathbf{s}_k = \mathbf{D}_j \mathbf{x}_{jk} (j \neq k), \quad (3)$$

where the l_0 norm of the coefficients $\|\mathbf{x}_{jk}^{\text{opt}}\|_0 > \|\mathbf{x}_k^{\text{opt}}\|_0$, requiring that the dictionary \mathbf{D}_j cannot represent \mathbf{s}_k in a sparse manner.

Following the preceding prerequisites, MCA aims to determine the sparsest representations of the signal \mathbf{s} with respect to the augmented dictionary composed of $\{\mathbf{D}_k\}_{k=1, \dots, K}$ (Elad et al., 2005). As a result, we need to address the following problem:

$$\begin{aligned} \{\mathbf{x}_1^{\text{opt}}, \dots, \mathbf{x}_K^{\text{opt}}\} &= \arg \min_{\{\mathbf{x}_1, \dots, \mathbf{x}_K\}} \sum_{k=1}^K \|\mathbf{x}_k\|_0 \\ \text{subject to } \mathbf{s} &= \sum_{k=1}^K \mathbf{D}_k \mathbf{x}_k. \end{aligned} \quad (4)$$

Solving the optimization task in equation 4 obtains the representation coefficients $\{\mathbf{x}_k\}_{k=1, \dots, K}$, resulting in a successful separation of various components $\{\mathbf{s}_k\}_{k=1, \dots, K}$. Unfortunately, as a nonconvex and

non-deterministic polynomial (NP)-hard optimization problem, there is an exponential increase in computational complexity when the number of columns in the augmented dictionary grows. To make it tractable, the l_0 -norm constraint is replaced with an l_1 -norm constraint. Together with a relaxation of the equality constraint, the model becomes an unconstrained convex optimization problem (Starck et al., 2005):

$$\{\mathbf{x}_1^{\text{opt}}, \dots, \mathbf{x}_K^{\text{opt}}\} = \arg \min_{\{\mathbf{x}_1, \dots, \mathbf{x}_K\}} \left\| \mathbf{s} - \sum_{k=1}^K \mathbf{D}_k \mathbf{x}_k \right\|_2^2 + \lambda \sum_{k=1}^K \|\mathbf{x}_k\|_1, \quad (5)$$

where λ indicates the Lagrange multiplier for adjusting a ratio between misfit and constraint terms. The block coordinate relaxation method (Sardy et al., 2000) is an effective way to solve equation 5 by decoupling it into K subproblems, each of which equals a typical $l_2 - l_1$ sparse optimization form:

$$\begin{aligned} \mathbf{x}_k &= \arg \min_{\mathbf{x}_k} \left\| \mathbf{s} - \mathbf{D}_k \mathbf{x}_k - \sum_{j \neq k} \mathbf{D}_j \mathbf{x}_j \right\|_2^2 + \lambda \|\mathbf{x}_k\|_1, \\ \text{for } k &= 1, \dots, K. \end{aligned} \quad (6)$$

Among those subproblems, $K - 1$ coefficient terms corresponding to distinct $\mathbf{D}_j (j \neq k)$ are assumed fixed, and one — the k th coefficient term — remains in question. Each subproblem can be solved by the soft-thresholding algorithm (Donoho and Johnstone, 1994) in an iterative manner:

$$\mathbf{x}_k^{l+1} = S_\lambda \left(\mathbf{D}_k^\dagger \left(\mathbf{s} - \sum_{j \neq k} \mathbf{D}_j \mathbf{x}_j^l \right) \right), \quad (7)$$

where \mathbf{D}_k^\dagger indicates Moore-Penrose pseudoinverse of \mathbf{D}_k . Similarly, \mathbf{x}^l represents the solution at the l th iteration. The soft-thresholding operator S_λ with a threshold λ can be written as

$$S_\lambda(a) = \begin{cases} a - \lambda \text{sign}(a), & |a| \geq \lambda \\ 0, & \text{others} \end{cases}. \quad (8)$$

Upon completion of the iteration, the separated components can be determined by

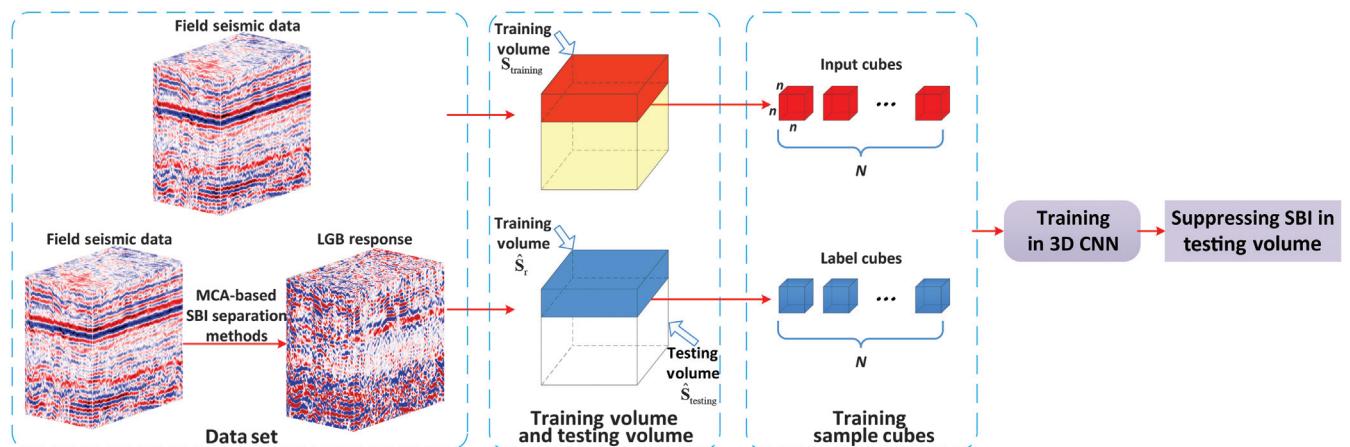


Figure 1. The proposed workflow for SBI separation.

$$\mathbf{s}_k = \mathbf{D}_k \mathbf{x}_k^{\text{opt}}. \quad (9)$$

To adapt MCA to SBI separation, we assume that the original seismic data are a linear superposition of SBI called \mathbf{s}_b , the weak-energy target reflection called \mathbf{s}_r , and random noise called \mathbf{s}_n (Wang et al., 2011). Accordingly, we can formulate inline or crossline vertical sections in 3D seismic data as

$$\mathbf{s} = \mathbf{s}_b + \mathbf{s}_r + \mathbf{s}_n. \quad (10)$$

Incorporating a noise component \mathbf{s}_n is helpful to increase the applicability of our model, making it not only effective for seismic data disturbed by random noise but also for actual data containing other waveform components that slightly violate model assumptions.

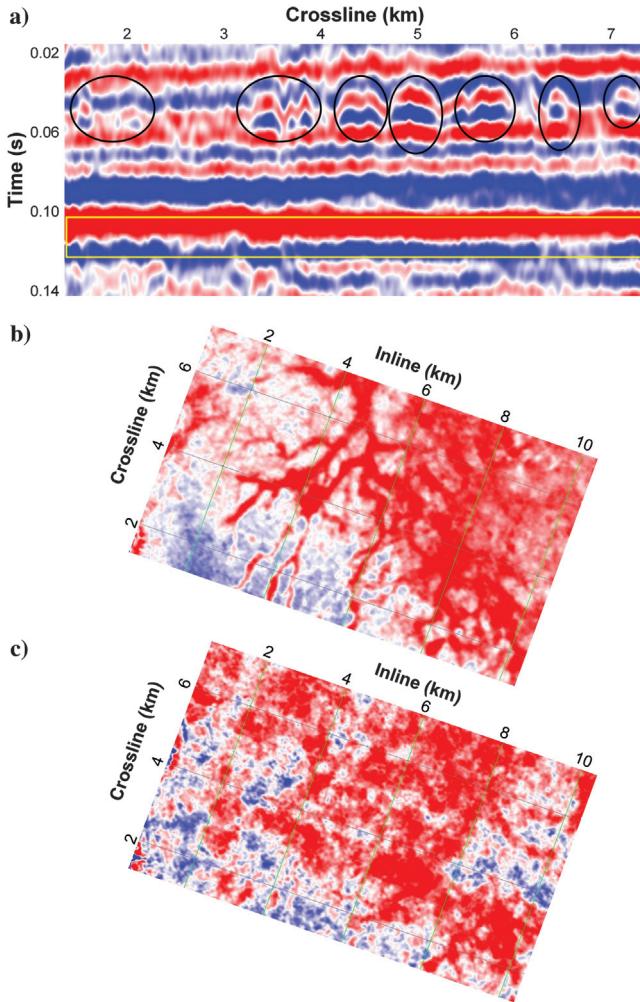


Figure 2. Comparative analysis of different morphological characteristics. (a) Raw seismic section. As indicated by the black ellipses, the channel structure in the shallow layer is apparent. However, the deep layer data are covered by the strong reflection layer as outlined by the yellow box, making it challenging to clearly visualize the weak-energy LGB waveform responses. (b) A time-slice example (50 ms) from the shallow data with obvious channel structures. Accordingly, we select the shallow layer data for training the network. (c) A deep time-slice example (100 ms). Due to the influences on deep layer data from SBI, weak-energy LGB waveform responses are almost invisible.

Figure 2a presents a raw original seismic section contaminated by SBI. In consonance with the MCA theory, \mathbf{s}_b and \mathbf{s}_r should exhibit noticeable morphological differences. Fortunately, the waveform responses of \mathbf{s}_b are stable, thus producing continuous and coherent events. On the contrary, weak-energy target reflections map to the waveform response of laterally inhomogeneous geologic bodies (LGB), which manifests as abrupt features in the seismic section. This apparent difference allows us to potentially distinguish between these two components.

It is important to note that MCA makes a distinction between the various components of the original signal based on the incoherence degree across the dictionaries. Furthermore, as the representation coefficients become more sparse, the MCA separation effect becomes more pronounced. To this end, choosing appropriate dictionaries, which have unique atomic characteristics to provide a highly sparse characterization of each component, is crucial to the successful application of MCA. Ideally, once we find suitable dictionaries, the SBI separation could be accomplished by solving

$$\{\mathbf{x}_b^{\text{opt}}, \mathbf{x}_r^{\text{opt}}\} = \arg \min_{\{\mathbf{x}_b, \mathbf{x}_r\}} \|\mathbf{s} - \mathbf{D}_b \mathbf{x}_b - \mathbf{D}_r \mathbf{x}_r\|_2^2 + \lambda (\|\mathbf{x}_b\|_1 + \|\mathbf{x}_r\|_1), \quad (11)$$

where \mathbf{x}_b and \mathbf{x}_r are the sparse representation coefficients under the overcomplete dictionaries \mathbf{D}_b and \mathbf{D}_r with regard to \mathbf{s}_b and \mathbf{s}_r , respectively.

2D-SWT for sparse representation of LGB responses

Target reservoirs, such as channel sand-body sediment, exhibit LGB seismic responses, which can be generally characterized as an abrupt waveform component. As shown by the black ellipses in Figure 2a, it presents a point distribution structure, thus leading to a noticeable point singularity feature. This observation is corroborated by the fact that conventional interpretations of channel sand-body sediments also rely on amplitude anomalies in seismic profiles.

Wavelet transforms are well known for capturing objects with point singularities. A typical example is the 2D-SWT, which possesses multiscale and redundant properties. Prior studies about acquisition footprints (Cvetkovic et al., 2007) and ground-roll attenuation (Wang et al., 2012) have convincingly shown that 2D-SWT is helpful for seismic component decomposition. On the one hand, 2D-SWT offers better translation-invariance properties as opposed to 2D discrete wavelet transforms. On the other hand, Gibbs phenomena are substantially reduced by 2D-SWT. Due to these factors, 2D-SWT is chosen as the dictionary for capturing abrupt point features. More details on 2D-SWT are presented in Appendix A.

The 2D-SWT atoms with three different scales are displayed in Figure 3a. Each scale consists of three directions: horizontal, vertical, and diagonal. A waveform dictionary can sparsely represent a signal component if the atoms in it accurately match the structural properties of the signal component. It can be apparently observed that these multiscale and isotropic atoms precisely match the abrupt feature of weak reflections. Consequently, we use 2D-SWT to extract the point features of weak-energy target reflections because it has superior sparse representation capabilities for point singularity targets.

Curvelet transform for sparse representation of SBI

SBI, also known as stable sedimentary stratum, generates coherent waveform responses. As illustrated by the yellow box in Figure 2a,

SBI presents a smooth curve distribution structure in the seismic section, showing evident curved singularity with a wide spatial distribution. In this regard, we need to select a dictionary with the ability to depict curved singularities effectively. Most wavelet transforms are inadequate at representing curved geometry regularity because their atoms are solely generated by isotropic dilation. This isotropic dilation enables these atoms to be well localized, but only dispersed across multiple locations and scales. To express anisotropic regularity in multivariate functions, e.g., edges in images, the basic elements of the dictionary need correspondingly to be distributed in multiple directions. In this manner, the optimal representation coefficients are sparse because multidirectional atoms can more accurately characterize continuously changing curve features.

Currently, several multidimensional transformations, such as the directional wavelet transform, the curvelet transform, and the contourlet transform, have been developed to handle this problem. Specifically, the curvelet transform introduced by Candès and Donoho (2000) is a popular technique for multiscale geometric analysis. For most of the 2D functions that exhibit a smooth pattern except for discontinuities along smooth curves, the curvelet transform can provide a nearly optimal nonadaptive sparse representation (Ma and Plonka, 2010). Recall that SBI is characterized by curved features in seismic sections. As a result, we choose the curvelet transform as the sparse representation dictionary \mathbf{D}_b . More details on the curvelet transform are provided in Appendix B.

It can be clearly observed from Figure 3b that curvelet atoms possess favorable geometric properties of multiscale, anisotropy, and multidirectionality. Equivalently, this observation indicates that the curvelet transform is efficient in capturing curve-like singularities.

Note that, because the profile morphological features of SBI and LGB responses differ noticeably, the proposed MCA separation algorithm is performed on 2D seismic sections. Moreover, direct implementation of 3D MCA is not feasible due to the computationally intensive nature of 3D dictionaries (see “Discussion” section). To allow access to 3D seismic data, we first remove the SBI component section by section along the inline direction. In this manner, a great part of SBI can be attenuated; however, a minimal amount of signal leakage may occur. Then, the fidelity is improved by reapplying the MCA algorithm along the crossline direction to the earlier removed SBI, thereby allowing us to retrieve the leaked signals. Summing up the extracted abrupt components from these two separation steps produces the final weak-energy target reflections, as illustrated in Figure 4.

Data preparation and network training

During the training phase, neural networks directly acquire prior knowledge about how to separate SBI from training samples, which has a crucial impact on the separation performance of the network. In other words, it brings additional benefits when we construct training samples through elaborate processing procedures or artificially select some satisfied training samples according to certain criteria. The network can automatically incorporate the prior knowledge behind such procedures or criteria. It is especially advantageous when these procedures are time-consuming and labor-intensive or when mathematical formulas cannot explicitly express these selection criteria. In these situations, deep learning is superior to traditional methods in either processing efficiency or flexibility.

Taking advantage of this specific property, we prepare high-quality training samples from two perspectives. On the one hand, we

obtain satisfactory SBI removal results by a sophisticated compressive sensing algorithm, i.e., MCA. However, the preceding separation model is a pseudo-3D processing strategy due to computational limitations. Specifically, it considers that channel sand bodies disrupt reflection continuity in seismic sections, but ignores the fact that their identification is easier in time or depth slices of 3D seismic data sets. To incorporate more spatial information and enhance separation precision, we directly cut the original 3D seismic data set and the corresponding SBI extracted by MCA into small 3D cubes, and then pair them together to construct training samples. On the other hand, we observe that the severity levels shielded by SBI are varied. For instance, weak-energy abrupt waveform responses are evident in shallow-layer seismic sections, as indicated by the black ellipses in Figure 2a. Such clear visibility of channel structures also is confirmed in the time slice of 50 ms, as presented in Figure 2b. Continuing to examine the deep-layer reflections as shown by the yellow box in Figure 2a, the sandstone reservoirs underlie sand-shale. Unfortunately, owing to the strong reflection interface of sand-shale, it is extremely challenging to accurately visualize the weak-energy LGB waveform responses. A time slice of 100 ms provides an example of this situation, as shown in Figure 2c.

For efficient and accurate separation of LGB waveform response features, we select shallow-layer seismic data with clearly visible LGB waveform responses as training data. Subsequently, the well-trained network is used to identify LGB waveform responses in deep layers, where the LGB waveform responses are fully submerged in SBI. A significant advantage of this sample construction strategy lies in obtaining high-quality labels because conventional methods can produce superior results with pronounced LGB waveform responses in the shallow-layer data. With these training data, the network can more easily understand how to accurately distinguish between SBI with coherent features and weak-energy targets with abrupt features. Moreover, adjacent regions have a similar underground structure, as does the stratigraphic structure causing SBI. Therefore, although the amplitude varies, the coherent features for characterizing SBI are nearly identical across different time slices. In other words, neighboring time slices do not fluctuate dramatically, supporting the validity of our sample construction strategy.

Having finished the data preparation, we feed these training samples into a 3D network for training. The input of the network is the original 3D seismic cube \mathbf{s} , which consists of SBI \mathbf{s}_b , target reflec-

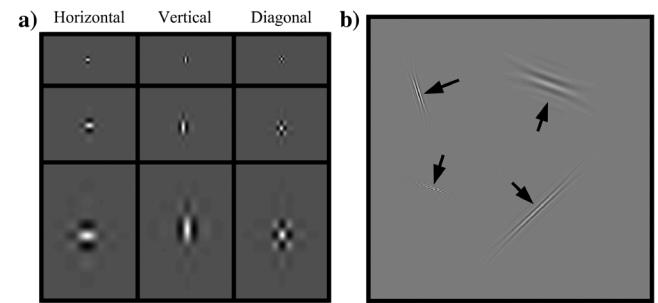


Figure 3. Comparison of 2D-SWT atoms and curvelet transform atoms. (a) The 2D-SWT atoms. The 2D-SWT atoms are isotropic and can capture point features; thus, they are selected as the dictionary to sparsely represent the weak-energy target reflections. (b) Curvelet transform atoms. Curvelet transform atoms are isotropic and can capture curve singularity features, and thus they are selected as the dictionary to sparsely represent SBI.

tions \mathbf{s}_r , and random noise \mathbf{s}_n . The network output targets, either $\hat{\mathbf{s}}_b$ or $\hat{\mathbf{s}}_r$, are the corresponding labels constructed by MCA. Motivated by the success of denoising convolutional neural network (DnCNN) (Zhang et al., 2017) and similar SBI features, we adopt a residual learning strategy $\Re(\mathbf{s}) \approx \hat{\mathbf{s}}_b$ to train the network, mapping the input raw seismic data to coherent waveform responses. Correspondingly, the network can provide the estimated LGB waveform responses by $\hat{\mathbf{s}}_r = \mathbf{s} - \Re(\mathbf{s})$. We omit \mathbf{s}_n here because it is merged into \mathbf{s}_b . Using the averaged mean square to measure the error between network outputs and labels, we can formulate the loss function as

$$l(\Theta) = \frac{1}{2N} \sum_{i=1}^N \|\Re(\mathbf{s}^i; \Theta) - (\mathbf{s}^i - \hat{\mathbf{s}}_r^i)\|_F^2, \quad (12)$$

where N indicates the total amount of cubes used for training. A locally optimal solution to trainable parameters Θ is determined

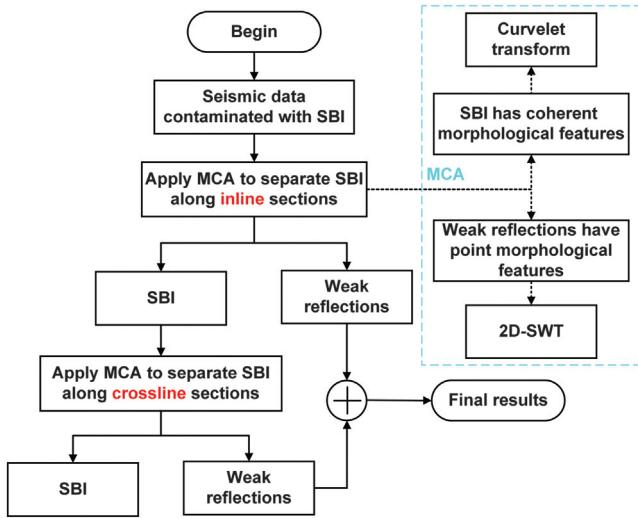


Figure 4. A flowchart illustrating the MCA process for SBI separation of 3D seismic data.

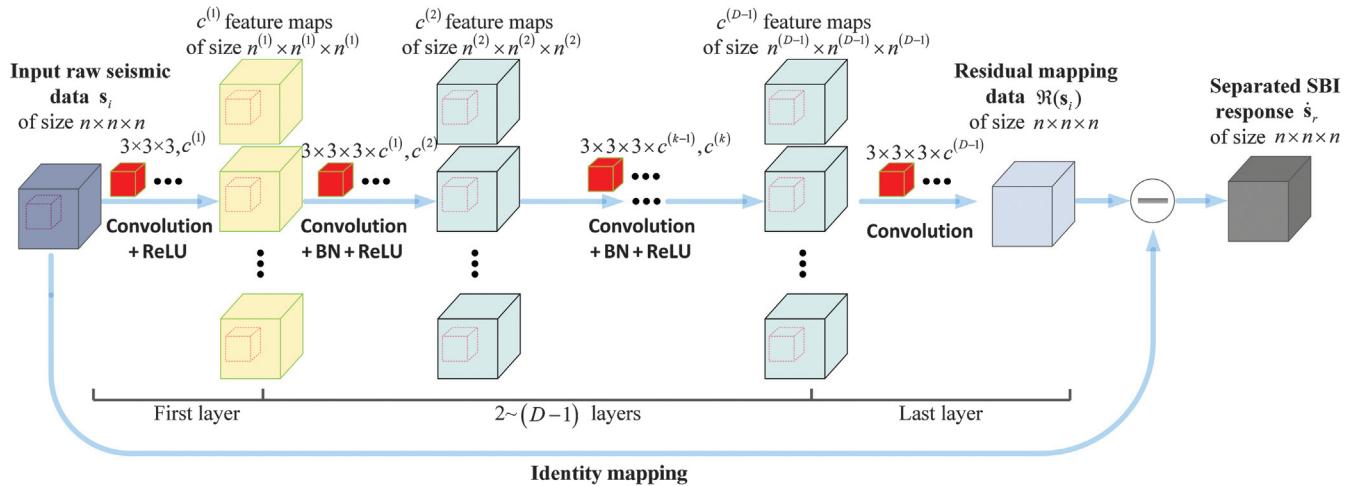


Figure 5. The network architecture of the proposed 3D CNN for SBI separation. The number of convolution channels is denoted by c . This figure is slightly revised from Liu et al. (2019).

by the classic back-propagation algorithm (Rumelhart et al., 1986) after iterative training.

Figure 5 illustrates the architecture of the proposed 3D convolutional neural network (CNN) for separating SBI in detail. Three stages constitute the proposed network regression model. It begins with one convolutional layer. With reference to VGGNet (Simonyan and Zisserman, 2015), we follow their experimental parameter setting and assign the convolution kernel size to $3 \times 3 \times 3$. In addition, we eliminate all pooling layers to ensure that the network output is the same shape as the input. The middle stage primarily comprises 15 convolutional layers, each with batch normalization (BN) layer between them. In general, BN improves the performance of CNNs on two fronts: reducing covariance shift and increasing training convergence (Ioffe and Szegedy, 2015). Afterward, one convolutional layer constitutes the final stage.

A standard convolution calculation shrinks voxels near the boundary. As such, we should select a padding strategy among the existing padding strategies that can address this issue. According to our testing, a basic zero padding satisfies our demands and does not cause additional artifacts in the results. It is obvious that our network architecture and training process are straightforward and concise. With such a simple network, it is easier to verify that our training sample selection strategy is effective, even if introducing more regularization terms would enhance our results even further.

EXAMPLES

Synthetic data example

The first step is to examine the effectiveness of the aforementioned MCA-based SBI separation method on 2D synthetic data. Figure 6a illustrates the 2D velocity model used to generate the synthetic data. It can be observed that a 100 m wide and 10 m thick channel is embedded at the interface between the fourth and fifth strata. Then, this velocity model is sampled along the horizontal and vertical dimensions at intervals of 2.5 m. A Ricker wavelet with a dominant frequency of 60 Hz is used in this example. Figure 6b depicts the migrated profile obtained through wave equation-based prestack depth migration. As can be seen from the synthetic seismic profile, the reflection waveform associated with channel sedimen-

tation appears as a bright spot, exhibiting a narrow distribution with abrupt point structures. Regrettably, it substantially merges with the reflection waveform of stable sedimentary strata, making separation difficult.

Data are thereafter separated using the proposed MCA algorithm. We obtain the abrupt waveform component section in Figure 6c and the coherent waveform component section in Figure 6d, respectively. Comparing the separated results with the original data profile, it can be seen that the reflection waveform of channel sedimentation is completely stripped from SBI. In addition, the coherent waveform component virtually does not contain any residual reflection waveform structure associated with the channels, indicating the high fidelity of our method. Figure 7 shows the results of the 500th trace, which is indicated by the green line in Figure 6. The original seismic trace is contaminated by the SBI, shielding the LGB responses near 0.35 km as shown by the red box in Figure 7a. Examining the separated results shown in Figure 7b, we can find that high-fidelity weak-energy signals are successfully recovered, and SBI distortion is effectively reduced, as indicated by the green arrows. In conclusion, the proposed method succeeds in removing SBI from the synthetic data.

Field data example

In this section, we evaluate the proposed workflow on a 3D seismic data set from eastern China. The raw data quality is relatively high because it is collected in winter and there are few interferences. Accordingly, the signal-to-noise ratio (S/N) of the migrated data is satisfactorily high if SBI is temporarily overlooked. After extracting the time slices near the target layer, we obtain the data set that has 256 time sample points with a sampling interval of 1 ms. Crossline direction and inline direction have 401 and 681 samples, respectively, with a 20 m spatial interval. A complex fluvial depositional system exists in this work area, with obvious marker beds in the middle. Because channels are viewed as potential reservoirs, channel interpretation has been regarded as one of the most challenging tasks in the interpretation of data from this region. Figure 8a illustrates a crossline section flattened along the shielding layer. Due to the strong covering effect of the marker beds, it is quite difficult to clearly identify weak-energy sedimentary characteristics. Apparently, interpretations and attribute analyses are difficult to perform on the original seismic profile. In the first stage, MCA is used to remove SBI. Figure 8b and 8c, respectively, depicts the abrupt waveform component and coherent waveform component separated from the preceding original section. Comparing Figure 8b with Figure 8a, it is evident that the reflection waveform structure associated with channel sand-body deposition in the original profile has been effectively separated, particularly the areas indicated by the solid green line box that was severely distorted by SBI. MCA also makes channel boundaries clear so that we can identify obvious channel structures from the areas marked by the green box with a dashed line. Moreover, as illustrated in Figure 8c, the interface of SBI can be clearly detected

after separation, facilitating understanding of the sedimentary process.

As a further convincing demonstration of the proposed MCA separation, we compare results on the 200 ms time slice, which is one of the most seriously affected time slices by SBI. Figure 9a shows the relative geologic time slice of the original data, and we can generally discern strong-energy fault structures. Due to the strong

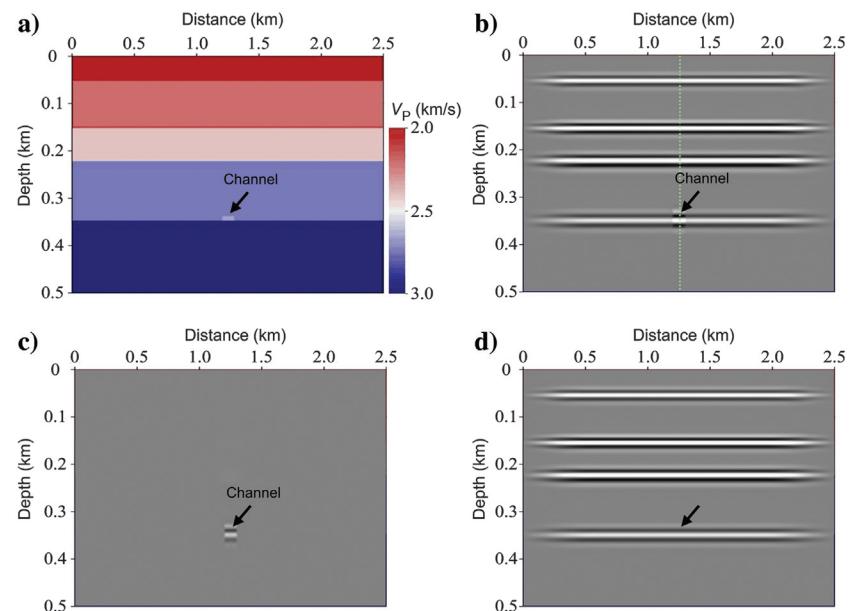


Figure 6. SBI suppression on synthetic data based on MCA. (a) A 2D velocity model with an embedded channel. (b) Prestack depth migration seismic profile. The green line denotes the position of a single-trace comparison in the following figure. (c) Recovered LGB waveform component by MCA. (d) Separated SBI waveform component by MCA.

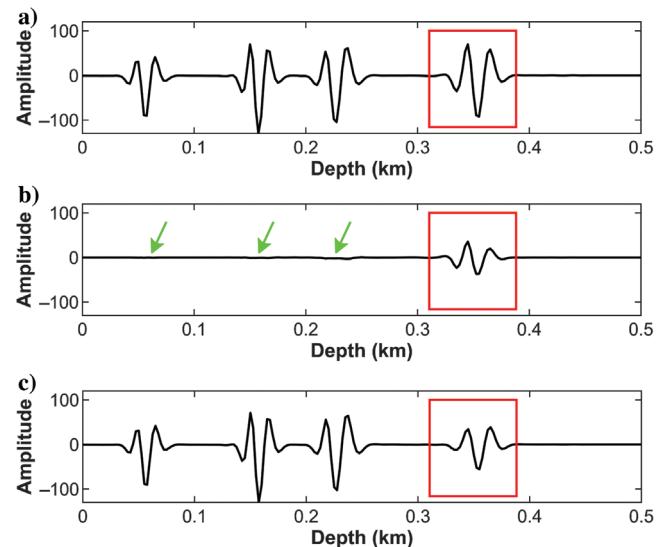


Figure 7. Comparison of the SBI separation results with the 500th trace in the synthetic profile. (a) Synthetic data contaminated by SBI. The red box indicates the location of the channel structure. (b) Recovered LGB waveform component by MCA. By observing the green arrows, we can conclude that SBI is effectively suppressed. (c) Separated SBI waveform component by MCA.

shielding effect of the marker beds, many weak-energy LGB responses are covered and therefore difficult to observe. From the time slice of the abrupt waveform component obtained by the MCA method in Figure 9b, it can be seen that fluvial deposit structures, distributed in a scattered pattern along the crossline section, are largely stripped from SBI. As well as being able to visualize some partially observed channel sandstone in the original data more clearly, sedimentary structural features that were not visible in the raw data also are recovered as indicated by the green rectangle. More importantly, channel sandstone bodies cannot be observed in the removed SBI in Figure 9c, demonstrating the high fidelity of this method. Figure 10a and 10b displays the attribute slices derived by the semblance-based coherence algorithm (Marfurt et al., 1998), which provides additional evidence that the fluvial sandstone bodies are more apparent after MCA separation than in the original slice. The original data can merely highlight the position of fault structures. After the MCA separation, some detailed fluvial sedimentary structures buried in the background of stable sedimentation strata can be viewed more clearly.

In the second stage, we train the network on shallow layer data and investigate its improvement on SBI separation, especially for deeper-layer data. Specifically, the training data set is constructed with time intervals ranging from 0 to 80 ms based on the results

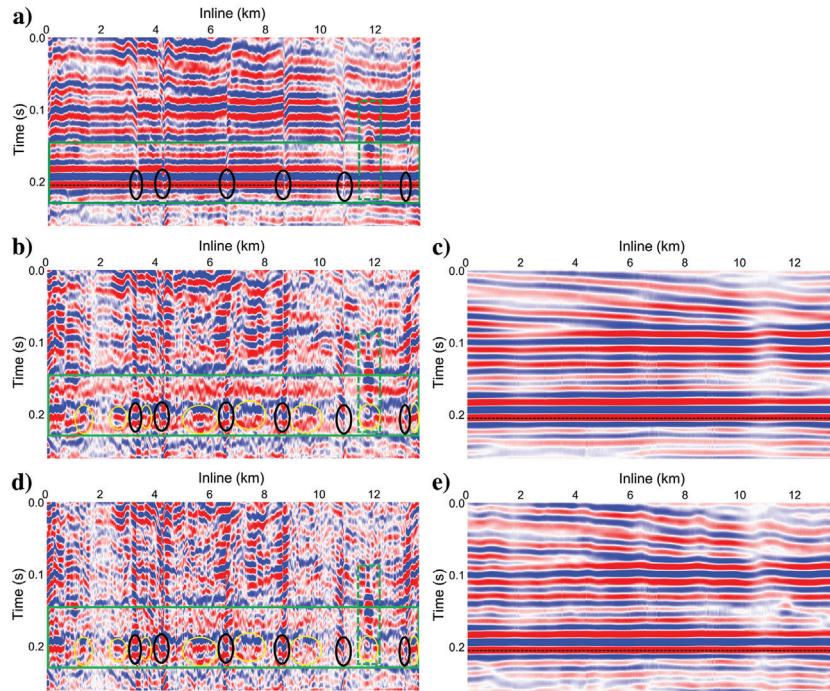


Figure 8. Crossline section results of SBI separation at 6.40 km. (a) A raw seismic crossline section flattened along the shielding layer, which is polluted by SBI. The dashed black line indicates the position for subsequent time slice comparison. The green box with a dashed line marks an apparent channel position. (b) The recovered LGB waveform component by MCA. The channel boundary in spatial and temporal directions is clearer than in the original section, as indicated by the green box with a dashed line. Moreover, as indicated by the solid green line box, channel structures are much clearer, but this section still contains a few SBI residues. (c) The separated SBI waveform component by MCA. (d) The recovered LGB waveform component by the network. By comparing the solid green line box, we observe that the network recovers more LGB waveform responses than the first-stage MCA, especially in the areas marked by the yellow circles. Furthermore, the black circles indicate more prominent fault structures than the MCA results. (e) The separated SBI waveform component by the network.

from the previous stage. Only one parameter, the learning rate, should be prudently predefined for the network training; we set it to 10^{-4} after several tests. A more detailed description of the experiment setting can be found in Table 1. Figure 8d and 8e shows the second-stage section results, and we can easily observe that they have significant improvement compared with the results of the former stage. The weak-energy reflections contain more LGB response waveform features, especially the 150–220 ms regions that SBI severely distorts, as marked by the green rectangle. Specifically, it is evident from the yellow circles that channel sandstone deposits are more prominent, which is consistent with the subsequent time slice results. Moreover, the fault structures indicated by the black circles are more apparent than those in the MCA results. Looking at the time slice results of the network in Figure 9d and 9e, we can draw the same conclusion. Although MCA has effectively removed most SBI, some SBI remains visible because the SBI energy of this time slice is extremely strong, as indicated by the yellow arrows. Obviously, 2D dictionaries are inadequate for completely capturing 3D underground structures. In coincidence with the section results, the network further suppresses SBI in the time slice, allowing more channel sand bodies to emerge, particularly around the yellow arrows. Moreover, the network alleviates the seismic imaging anomalies caused by fault boundaries compared with the first-stage MCA

results. These findings indicate that well-chosen training samples contribute considerably to improving SBI separation results. In addition, adding knowledge of 3D data structures also is one of the reasons that the network achieves a performance improvement.

For a more intuitive analysis, we compare the coherence attribute between the MCA results and network results. The MCA parameters are fixed for simplicity, and therefore a small amount of SBI residues remain, which slightly obscure the channel structure in Figure 10b. The coherence attribute of the network results is shown in Figure 10c. The previous graph shows that our network further achieves remarkable SBI suppression, and more prominent channel sandstone deposits can be identified on this time slice. Moreover, the overall fault structures remain unchanged after SBI is eliminated, demonstrating the reliability of our network. Clearly, our network generalizes the satisfactory results obtained by MCA in shallow layers to deep layers due to the similarity of subsurface structures. This also verifies that our sample selection is reasonable because the network has acquired the desired skills to extract SBI features after learning the selective training samples. In addition, the network can automatically gain knowledge regarding 3D data structures, which in turn results in substantial performance improvements.

We now pay attention to the inline results presented in Figure 11. The original section shown in Figure 11a, located at 4.74 km, has complex waveform features due to SBI distortions, making it difficult to clearly characterize the channel characteristics of some weak energies. Figure 11b shows

the section results with MCA separation, and we find that the strong-amplitude SBI has been substantially reduced. This allows the relative amplitude of reservoir seismic response to describe its actual development more accurately. Furthermore, as shown in Figure 11c, the MCA algorithm does not influence the stratum succession, proving its reliability. Unfortunately, the MCA results contain a few SBI residues because the fluvial sandstone energy is too weak to meet our prerequisite 1. Turning now to the network results in Figure 11d, it is apparent that the LGB waveform responses are more apparent than the MCA results, as weak reflections are considerably enhanced. Although solely trained on shallow layer data, our network shows satisfactory generalization capabilities to process adjacent seismic data. Moreover, the fact that our results in Figure 11e are free of the edge artifacts at patch borders confirms the validity of our network.

In summary, our workflow successfully removes SBI and unveils channel structures. In addition, with the abrupt waveform components separated by our workflow, we can qualitatively analyze and characterize the fluvial facies sedimentary system. Moreover, the extracted LGB reflection waveforms of channel sand bodies can be used to quantitatively analyze associated seismic properties, such as reservoir thickness, physical properties, and oil and gas properties, so as to facilitate reservoir modeling, etc.

DISCUSSION

2D dictionaries versus 3D dictionaries

The morphological difference in seismic sections between SBI (coherent morphological features) and LGB waveform responses (abrupt morphological features) is the key to the successful implementation of our MCA method. In seismic sections, as channels become broader in time slices, their abrupt features also become wider in space correspondingly. If we use a relatively small window, the wide channel region also may exhibit coherent features, which violates the MCA prerequisite requiring different morphological features. In this regard, the curvelet transform and 2D-SWT are unable to distinguish them effectively, leading to unsatisfactory separation results. A viable solution is to select a relatively big window where channels still have different morphological features from SBI. However, a larger window will result in a significant increase in computational costs. After several trial experiments, we select a compromise solution and set the window size to 512×512 to fully exploit their morphological feature differences.

The 3D multiscale transforms, such as the 3D curvelet transform (Ying et al., 2005) and

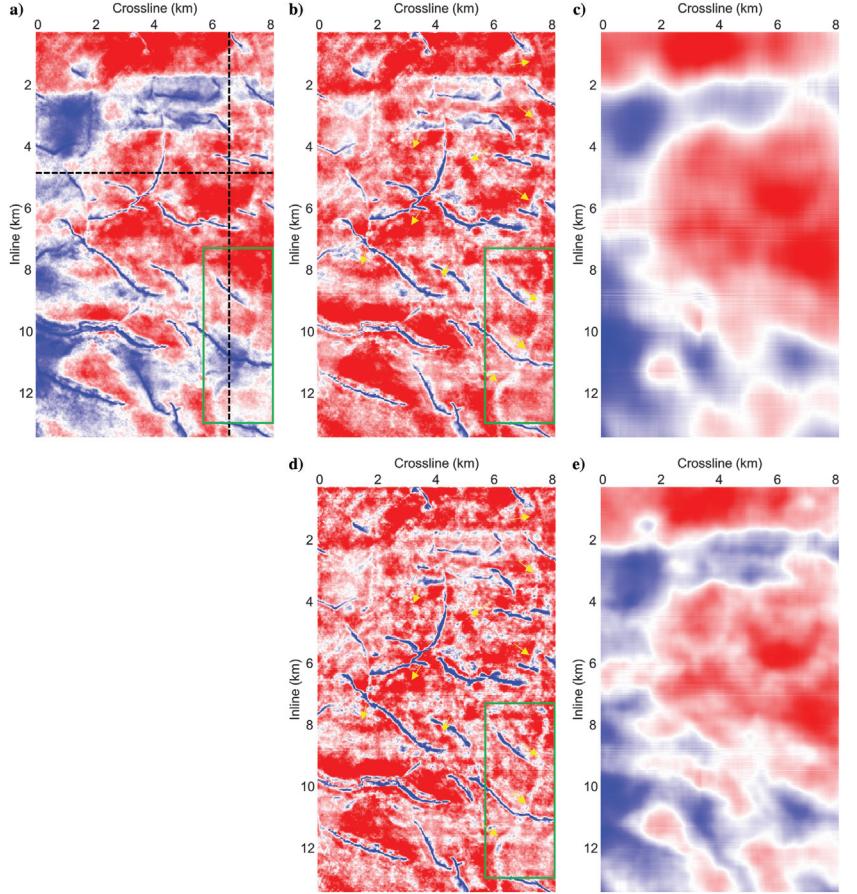


Figure 9. The results of SBI separation on the 200 ms time slice. (a) The raw time slice. Many finer channel structures are not clearly visible due to the shielding effect of SBI. Two dashed black lines denote the position for section comparison. (b) The recovered LGB waveform component by MCA. The channel structures are considerably improved, as shown by the green box. (c) The separated SBI waveform component by MCA. (d) The recovered LGB waveform component by the network. The yellow arrows indicate the network further reveals more channel sand bodies. (e) The separated SBI waveform component by the network.

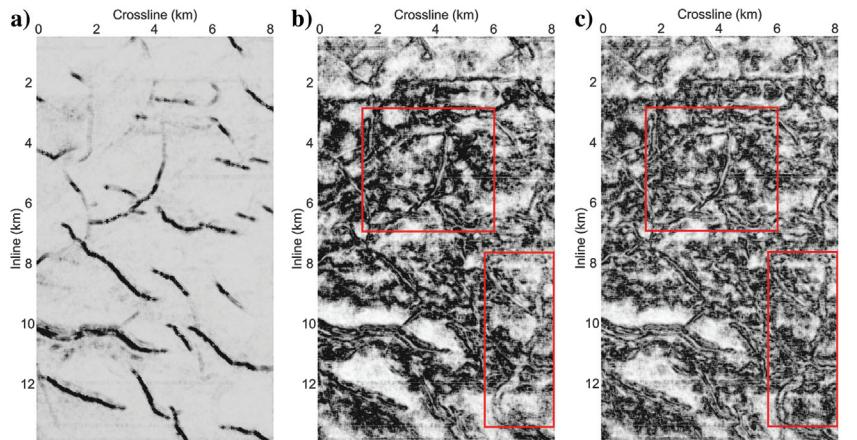


Figure 10. A comparison between the coherence attribute calculated on (a) raw seismic data, (b) MCA results, and (c) the network results. The red boxes show that the network reveals more prominent channel sandstone deposits. The fact that the main fault structures remain unchanged indicates that the network achieves reliable SBI suppression.

Table 1. Summary of experimental parameters for deep learning method.

Parameter type	Description	Value
Network	Convolution kernels (time \times crossline \times inline \times input channels \times output channels)	First layer $3 \times 3 \times 3 \times 1 \times 64$ Last layer $3 \times 3 \times 3 \times 64 \times 1$ Other layers $3 \times 3 \times 3 \times 64 \times 64$
Training	Volume size (time \times crossline \times inline)	$80 \times 360 \times 640$
	Patch size (time \times crossline \times inline)	$40 \times 40 \times 40$
	Patch overlapping (time \times crossline \times inline)	$20\% \times 20\% \times 20\%$
	Total patches	15,000
	Batch size	6
	Epoch	40
	Run time (four GPUs)	18.5 h
Testing	Volume size (time \times crossline \times inline)	$256 \times 401 \times 681$
	Patch size (time \times crossline \times inline)	$180 \times 180 \times 180$
	Patch overlapping (time \times crossline \times inline)	$44.44\% \times 44.44\% \times 44.44\%$
	Total patches	48
	Run time (one GPU)	168 s

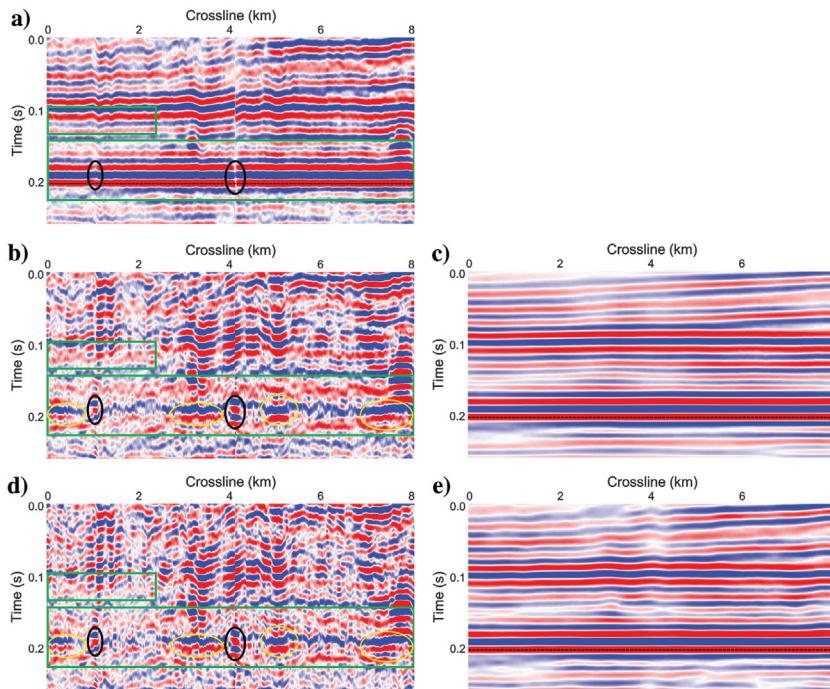


Figure 11. In-line section results of SBI separation at 4.74 km. (a) A raw seismic inline section contaminated by SBI. The dashed black line indicates the position for the previous time-slice comparison. Black circles mark fault positions. (b) The recovered LGB waveform component by MCA. Channel structures are much clearer, but the green boxes indicate that a few waveform responses of stable sedimentation can still be observed. (c) The separated SBI waveform component by MCA. (d) The recovered LGB waveform component by the network. The green boxes indicate that the network suppresses higher levels of SBI and recovers more refined LGB waveform responses than MCA, especially in the yellow circles. (e) The separated SBI waveform component by the network.

3D-SWT (Chen et al., 2012), have more flexible directional selectivity than 2D transforms. Ideally, 3D dictionaries would achieve better SBI separation with minimal signal damage compared with 2D dictionaries. Nevertheless, as 2D data transition to 3D data, we still need to take into account the change in morphological features. When we use a 3D window, SBI is characterized by curved surface features, whereas LGB waveform responses exhibit curved line features. Both of these features can be sparsely represented by the 3D curvelet transform. To comply with the MCA framework, it is necessary to filter out partial curvelet coefficients to discriminate between curved surface and curved line features. However, selecting these regions in the curvelet domain is labor-intensive and reliant upon expertise. Moreover, similar to the 2D window, a relatively large 3D window should be used. Unfortunately, as the window size grows, the computational time and storage costs of 3D dictionaries increase drastically. The solver of MCA requires repeated iterations, which exacerbates the computational problem of 3D dictionaries. Based on the preceding limitations of 3D dictionaries, we finally select 2D dictionaries. Even so, we envision routinely using the 3D curvelet transform in the near future, allowing us to improve our results further.

Which is better for learning using the network: SBI or useful reflections?

The network is used to learn a nonlinear mapping from the original data to target separation labels in a supervised manner. Selecting a suitable separation target with relatively consistent features can facilitate the network learning process. We train the network in shallow-layer areas with relatively low background noise. On the one hand, the channel structure is relatively obvious, satisfying the MCA prerequisites. Therefore, MCA provides high-quality training samples in this area. More importantly, the geologic conditions do not change significantly within a small time window. Accordingly, there is no obvious feature variation in adjacent stable sedimentation strata, and the network is easy to capture the SBI features. In addition, the original seismic data have been flattened to decrease their spatial fluctuations and increase the similarity. In contrast, channel deposits have varying signatures, which are quite distinct from different time slices. For example, the field data have a meandering river system in the shallow layer, whereas the deep layer develops braided rivers. Therefore, using the network to learn a mapping to LGB waveform responses, i.e., $\mathfrak{R}(\mathbf{s}) \approx \mathbf{s}_r$, is more challenging. Figure 12 shows the results of the mapping to channels, and it is apparent that there is chan-

nel leakage due to insufficient learning of the varying channel features. As a result, we decide to learn a mapping to extract SBI using the network rather than extracting the channel structures.

Can deep learning outperform traditional methods used for constructing labels?

The denoising ability for supervised deep learning is learned from training samples. Therefore, it comes naturally that the labels limit the network denoising performance. Although the training data are constructed by the traditional method, we believe that deep learning has the potential to exceed the labels due to the following reasons.

First, most conventional denoising methods, including the MCA method used in this paper, denoise the whole 3D seismic data set with 2D or 3D sliding windows. In general, the denoising process of different patches is independent. As we all know, seismic data in the same work area have similarities in useful signals and noise distribution. Therefore, these conventional methods are characterized by local filtering methods, which cannot fully exploit the self-similarity across seismic patches. On the contrary, the CNN-based denoising method is somewhat similar to a nonlocal denoising method. Its loss function minimizes the average errors between network outputs and the labels. Although CNNs are locally perceived in a single patch, they share network parameters during the entire training process. As a result, the network learns many repetitive features in the training process and benefits from the self-similarity of seismic data. Moreover, we use 3D convolutional kernels, which can better explore the 3D spatial structure relationship of seismic data than conventional 2D methods. Intuitively, 3D DnCNN can discriminate useful signals and noises from a global perspective, thereby eliminating noise more effectively.

Second, considering the complexity of seismic data, different patches vary in reflection structure and noise energy. Accordingly, even the same conventional method has varying denoising capabilities for different patches. It is desirable to modify the parameters of conventional methods accordingly to adapt to these structural and energy changes. However, practical processing typically uses the same parameters for an entire seismic data set to simplify the procedure. This aggravates variations in the denoising effect across different patches. We select relatively high S/N patches to build our training samples, in which conventional denoising methods are more effective. With a higher quality of training samples, the network can identify the useful signal more accurately and extend the satisfactory denoising performance of the conventional methods from selected samples to the entire work area. Consequently, training sample selection manually provides additional a priori knowledge, which improves denoising performance.

Third, the achievement of a successful denoising effect depends on a well-suited regularization method. Similar to the commonly used sparse and low-rank regularization, the network architecture itself is an implicit regularization method (Ulyanov et al., 2018). The network weights serve as a parameterization of the network output, which aims to produce smooth and continuous results. By exploiting this network constraint, unsupervised deep learning methods have exceeded traditional methods in certain denoising tasks (Saad et al., 2021). Because the network can produce reliable results even without training labels, it is reasonable to expect that the proper use of training samples could exceed the performance of the traditional approach.

Finally, the rational use of well-trained networks improves processing results. For example, natural images and seismic data possess different characteristics. Nevertheless, a network trained on natural images can serve seismic interpolation (Zhang et al., 2020a). This method is effective because the network has been inserted into an inverse problem framework, which leverages the impressive capabilities of the existing deep learning denoiser. In a broader sense, our proposed two-stage method can be regarded as the following inverse problem:

$$\{\mathbf{x}_b^{\text{opt}}, \mathbf{x}_r^{\text{opt}}\} = \arg \min_{\{\mathbf{x}_b, \mathbf{x}_r\}} \|\mathbf{s} - \mathbf{D}_b \mathbf{x}_b - \mathbf{D}_r \mathbf{x}_r\|_2^2 + \lambda(\|\mathbf{x}_b\|_1 + \|\mathbf{x}_r\|_1),$$

subject to $\mathbf{D}_b \mathbf{x}_b = \mathfrak{R}(\mathbf{s}; \Theta)$, (13)

which plugs the network into an inverse problem framework. We can reformulate equation 13 as an unconstrained optimization problem:

$$\begin{aligned} \{\mathbf{x}_b^{\text{opt}}, \mathbf{x}_r^{\text{opt}}\} &= \arg \min_{\{\mathbf{x}_b, \mathbf{x}_r\}} \|\mathbf{s} - \mathbf{D}_b \mathbf{x}_b - \mathbf{D}_r \mathbf{x}_r\|_2^2 \\ &\quad + \lambda(\|\mathbf{x}_b\|_1 + \|\mathbf{x}_r\|_1) + \lambda_\phi \Phi(\mathbf{D}_b \mathbf{x}_b), \end{aligned} \quad (14)$$

which can be solved under the plug-and-play algorithm (Zhang et al., 2021a). Here, $\Phi(\bullet)$ is an implicit regularizer whose related subproblem will be solved by the well-trained CNN denoiser $\mathfrak{R}(\mathbf{s}^i; \Theta)$. The inverse problem framework provides flexibility for the implementation of networks so that they can be applied to improve conventional methods for various situations.

Based on the preceding analysis, deep learning is theoretically able to improve the conventional method. The results of the field data example in the manuscript also prove that the network has revealed more LGB response waveform features. Therefore, we report that the network has improved the conventional method.

As a further verification that the network can outperform the conventional method, we conduct a straightforward experiment. Figure 13a shows an original time slice from a field data set in eastern China, which contains 330 inline samples with 20 m intervals, 1000 crossline samples with 10 m intervals, and 126 time samples with

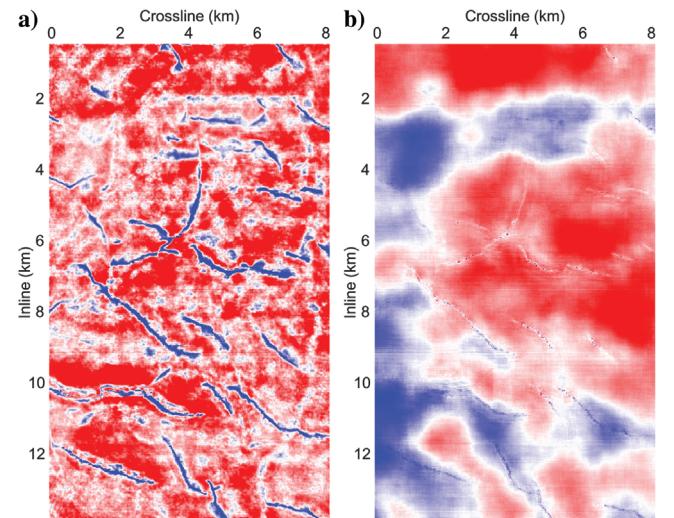


Figure 12. Results of using the network to map the LGB waveform responses. (a) The recovered LGB waveform component on the 200 ms time slice. (b) The separated SBI waveform component. Some channel structure leakage can be observed.

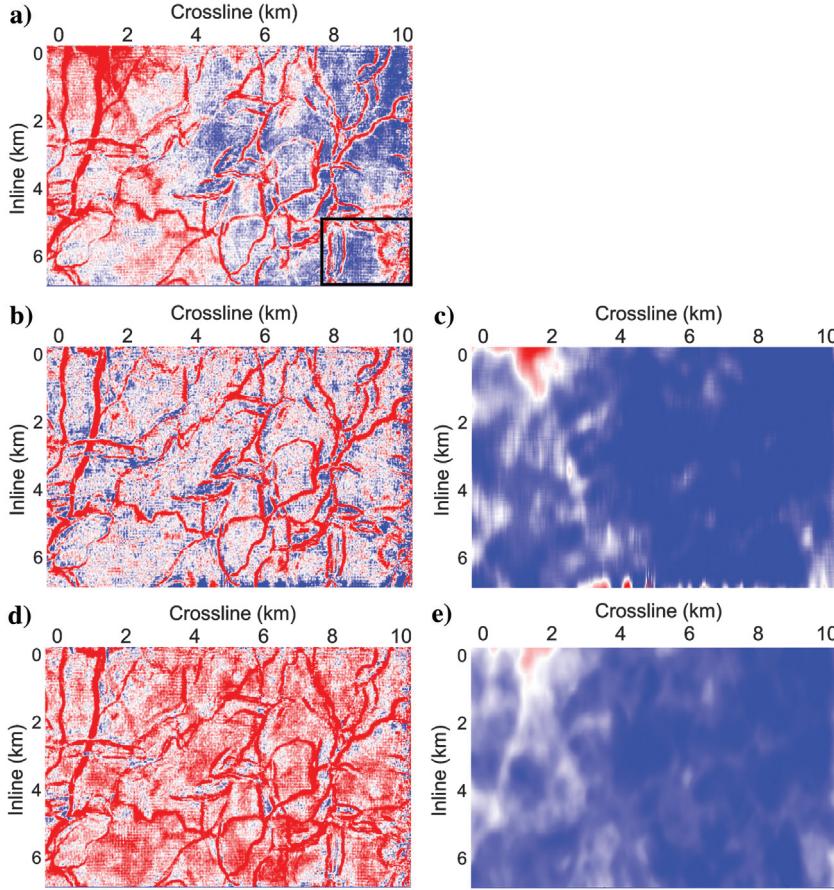


Figure 13. Results of 240 ms time slice from second field data set. (a) The raw time slice. The black rectangle indicates the training region for the network. (b) The recovered LGB waveform component by MCA. (c) The separated SBI waveform component by MCA. (d) The recovered LGB waveform component by the network. More LGB waveform responses are revealed compared with the MCA method. (e) The separated SBI waveform component by the network. The network reduces the criss-cross artifacts caused by MCA.

2 ms intervals. Similar to the training process in the previous field data example, we choose a relatively high S/N region for training, as indicated by the black rectangle. The network is tested directly on the original data rather than the MCA results. Figure 13d shows the recovered LGB waveform component by the network, and it can be observed that SBI has been effectively removed. Compared with Figure 13b obtained by MCA, channel sand-body contours in the network results are more consistent and complete. In addition, the network reveals more subtle sand-body responses. By comparing the separated SBI component in Figure 13c and 13e, we can find that SBI removed by the network has a greater continuity and regularity. Moreover, the 3D network improves the criss-cross artifacts generated by a 2D section-by-section operation with MCA.

CONCLUSION

We propose a two-stage SBI separation method. First, we suppress SBI using the MCA theory with appropriate overcomplete dictionaries. SBI has a coherent waveform response, whereas weak-energy target reflections exhibit LGB waveform responses. In this regard, the curvelet transform, which is a multiscale and multidirectional trans-

formation, is adopted as a dictionary for extracting the SBI component. The 2D-SWT is suited to analyzing signals that contain discontinuities or sharp spikes, serving as a dictionary to identify the weak-energy abrupt component. With the block coordinate relaxation method, we can effectively suppress SBI as demonstrated by the synthetic data. Then, the MCA results serve as a basis for training sample construction. To improve the MCA results further by fully exploiting 3D spatial information, we train a 3D network to extract SBI with a rational selection of training samples. Specifically, shallow-layer data with noticeable LGB features are selected for training. One field data set demonstrates that our network recovers more channel details based on the acceptable results of MCA. These findings provide insight into the potential for enhancing a conventional method with deep learning.

ACKNOWLEDGMENTS

This research is sponsored in part by the National Key R&D Program of China (2021YFA0716904) and in part by the National Natural Science Foundation of China (41974131 and 41774135). The work of D. Liu was supported by the China Scholarship Council. Many thanks go to the assistant editor C. Farquharson and other anonymous reviewers for their insightful suggestions. We also thank Daqing Oilfield Limited Co. of China for providing the real dataset.

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.

APPENDIX A

2D-SWT WAVELET TRANSFORM

The definition of 2D-SWT is given by a tensor product of the 1D-SWT scaling function $\varphi(t)$ and the wavelet function $\psi(t)$:

$$\begin{aligned} \varphi(x, y) &= \varphi(x)\varphi(y), \\ \psi^H(x, y) &= \varphi(x)\psi(y), \\ \psi^V(x, y) &= \psi(x)\varphi(y), \\ \psi^D(x, y) &= \psi(x)\psi(y), \end{aligned} \quad (\text{A-1})$$

where $\varphi(x, y)$ denotes a 2D scaling function. Three oriented wavelets, $\psi^H(x, y)$, $\psi^V(x, y)$, and $\psi^D(x, y)$, stand for horizontal, vertical, and diagonal directions, respectively. The Symmlet wavelet is adopted as the wavelet basis function with a vanishing moment of four. Subsequently, the à trous algorithm (Shensa, 1992) is used to construct wavelet filter banks H and G for extracting information from a 2D signal $f(x, y)$:

$$\begin{aligned} A_{j+1}[u, v] &= \sum_x \sum_y H_j[x - 2u] H_j[y - 2v] A_j[x, y], \\ D_{j+1}^H[u, v] &= \sum_x \sum_y H_j[x - 2u] G_j[y - 2v] A_j[x, y], \\ D_{j+1}^V[u, v] &= \sum_x \sum_y G_j[x - 2u] H_j[y - 2v] A_j[x, y], \\ D_{j+1}^D[u, v] &= \sum_x \sum_y G_j[x - 2u] G_j[y - 2v] A_j[x, y], \end{aligned} \quad (\text{A-2})$$

where A_j stands for the approximation coefficients at scale j . Horizontal, vertical, and diagonal detail coefficients are denoted by D_j^H , D_j^V , and D_j^D , respectively. Actually, the size of these coefficients is identical to that of the original signal $f(x, y)$. By inserting 2^{j-1} zeros among each coefficient of H and G , we can generate the j th scale decomposition filter banks H_j and G_j , respectively. Once MCA separation is complete, the extracted signal component can be reconstructed using the inverse 2D-SWT with the preceding coefficients:

$$\begin{aligned} A_j[x, y] &= \frac{1}{4} \sum_{i=0}^3 \left\{ \sum_u \sum_v \tilde{H}_j[x - 2u - i] \tilde{H}_j[y - 2v - i] A_{j+1}[u, v] \right. \\ &\quad + \sum_u \sum_v \tilde{H}_j[x - 2u - i] \tilde{G}_j[y - 2v - i] D_{j+1}^H[u, v] \\ &\quad + \sum_u \sum_v \tilde{G}_j[x - 2u - i] \tilde{H}_j[y - 2v - i] D_{j+1}^V[u, v] \\ &\quad \left. + \sum_u \sum_v \tilde{G}_j[x - 2u - i] \tilde{G}_j[y - 2v - i] D_{j+1}^D[u, v] \right\}, \end{aligned} \quad (\text{A-3})$$

where \tilde{H} and \tilde{G} refer to synthesis filter banks.

APPENDIX B CURVELET TRANSFORM

Suppose that $\xi = (\xi_1, \xi_2)^\top$ represents a location variable in the frequency domain. Then, the polar coordinates in the frequency domain can be calculated by $r = \sqrt{\xi_1^2 + \xi_2^2}$ and $\omega = \arctan(\xi_1/\xi_2)$. The Fourier transform of the mother curvelet $\phi_{j,0,0}(\mathbf{x})$ at scale 2^{-j} is limited in a trapezoid by

$$\hat{\phi}_{j,0,0}(r, \omega) := 2^{-3j/4} W(2^{-j}r) V_{N_j}(\omega), j \in \mathbb{N}_0, \quad (\text{B-1})$$

where $N_j = 4 \cdot 2^{\lfloor j/2 \rfloor}$. Here, $\mathbf{x} = (x_1, x_2)^\top$ denotes a location variable in the spatial domain. Window functions W and V_N need to have compact support and satisfy the following admissibility condition (Candès and Donoho, 2005; Candès et al., 2006):

$$\sum_{j=-\infty}^{\infty} |W(2^j r)|^2 = 1 \quad \text{for } r \in (3/4, 3/2), \quad (\text{B-2})$$

$$\sum_{l=0}^{N-1} V_N^2(\omega - \frac{2\pi l}{N}) = 1 \quad \text{for all } \omega \in [0, 2\pi), \quad (\text{B-3})$$

where N is an arbitrary positive integer and represents the number of wedges in a circular ring. Please refer to Ma and Plonka (2010) for details on how to construct W and V explicitly.

Curvelets at the scale 2^{-j} are obtained by dilating, rotating, and shifting the mother curvelet:

$$\phi_{j,\mathbf{k},l}(\mathbf{x}) := \phi_{j,0,0}(\mathbf{S}_{\theta_{j,l}}^\top(\mathbf{x} - \mathbf{b}_{\mathbf{k}}^{j,l})), \quad j, l \in \mathbb{Z}, \mathbf{k} \in \mathbb{Z}^2, \quad (\text{B-4})$$

where \mathbf{S}_θ is the rotation matrix $\mathbf{S}_\theta = \begin{bmatrix} 1 & 0 \\ -\tan \theta & 1 \end{bmatrix}$ used to change the orientation of $\phi_{j,0,0}$ by θ degrees. Here, \top is the transpose operation, and we have $\mathbf{S}_\theta^{-1} = \mathbf{S}_\theta^\top = \mathbf{S}_{-\theta}$. The term $\mathbf{b}_{\mathbf{k}}^{j,l} := \mathbf{S}_{\theta_{j,l}}^{-\top}(k_1 2^{-j}, k_2 2^{-\lfloor j/2 \rfloor}) = \mathbf{S}_{\theta_{j,l}}^{-\top} \mathbf{k}_j$ signifies a translation parameter that modifies the spatial position of $\phi_{j,0,0}$. Correspondingly, a curvelet coefficient is computed by the inner product:

$$\begin{aligned} C_\phi f(j, \mathbf{k}, l) &= \langle f, \phi_{j,\mathbf{k},l} \rangle = \int_{\mathbb{R}^2} f(\mathbf{x}) \overline{\phi_{j,\mathbf{k},l}(\mathbf{x})} d\mathbf{x} \\ &= \int_{\mathbb{R}^2} \hat{f}(\xi) \overline{\hat{\phi}_{j,\mathbf{k},l}(\xi)} d\xi \\ &= \int_{\mathbb{R}^2} \hat{f}(\xi) \hat{\phi}_{j,0,0}(\mathbf{S}_{\theta_{j,l}} \xi) e^{i \langle \mathbf{b}_{\mathbf{k}}^{j,l}, \xi \rangle} d\xi. \end{aligned} \quad (\text{B-5})$$

REFERENCES

- Candès, E., L. Demanet, D. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: Multiscale Modelling and Simulation, **5**, 861–899, doi: [10.1137/05064182X](https://doi.org/10.1137/05064182X).
- Candès, E. J., and D. L. Donoho, 2000, Curvelets, multiresolution representation, and scaling laws: Wavelet Applications in Signal and Image Processing VIII, International Society for Optics and Photonics, 1–12.
- Candès, E. J., and D. L. Donoho, 2005, Continuous curvelet transform: I. Resolution of the wavefront set: Applied and Computational Harmonic Analysis, **19**, 162–197, doi: [10.1016/j.acha.2005.02.003](https://doi.org/10.1016/j.acha.2005.02.003).
- Castelvecchi, D., 2016, Can we open the black box of AI? Nature News, **538**, 20, doi: [10.1038/538020a](https://doi.org/10.1038/538020a).
- Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: SIAM Review, **43**, 129–159, doi: [10.1137/S003614450037906X](https://doi.org/10.1137/S003614450037906X).
- Chen, W., Y. Chen, and Z. Cheng, 2017a, Seismic time-frequency analysis using an improved empirical mode decomposition algorithm: Journal of Seismic Exploration, **26**, 367–380.
- Chen, X., W. Chen, X. Wang, and W. Wang, 2017b, Sparsity-optimized separation of body waves and ground-roll by constructing dictionaries using tunable Q -factor wavelet transforms with different Q -factors: Geophysical Journal International, **211**, 621–636, doi: [10.1093/gji/ggx332](https://doi.org/10.1093/gji/ggx332).
- Chen, X., W. Yang, Z. He, and W. Zhong, 2012, Adaptive acquisition footprint suppression based on a 3D stationary wavelet transform: A case study from China: Journal of Applied Geophysics, **77**, 1–6, doi: [10.1016/j.jappgeo.2011.11.004](https://doi.org/10.1016/j.jappgeo.2011.11.004).
- Cvetkovic, M., S. Falconer, K. J. Marfurt, and S. Chavez-Perez, 2007, 2D stationary-wavelet transform-based acquisition footprint suppression: 77th Annual International Meeting, SEG, Expanded Abstracts, 2590–2594, doi: [10.1190/1.2793005](https://doi.org/10.1190/1.2793005).
- Di, H., L. Truelove, C. Li, and A. Abubakar, 2020, Accelerating seismic fault and stratigraphy interpretation with deep CNNs: A case study of the Taranaki Basin, New Zealand: The Leading Edge, **39**, 727–733, doi: [10.1190/tle39100727.1](https://doi.org/10.1190/tle39100727.1).
- Donoho, D. L., and J. M. Johnstone, 1994, Ideal spatial adaptation by wavelet shrinkage: Biometrika, **81**, 425–455, doi: [10.1093/biomet/81.3.425](https://doi.org/10.1093/biomet/81.3.425).
- Elad, M., J.-L. Starck, P. Querre, and D. Donoho, 2005, Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA): Applied and Computational Harmonic Analysis, **19**, 340–358, doi: [10.1016/j.acha.2005.03.005](https://doi.org/10.1016/j.acha.2005.03.005).
- Fang, W., L. Fu, S. Liu, and H. Li, 2021, Dealiased seismic data interpolation using a deep-learning-based prediction-error filter: Geophysics, **86**, no. 4, V317–V328, doi: [10.1190/geo2020-0487.1](https://doi.org/10.1190/geo2020-0487.1).

- Guan, X., C. Xie, X. Feng, and Y. Jin, 2016, Application of wavelet decomposition technology in thin sand identification under strong reflection: International Geophysical Conference, SPG/SEG, 417–420, doi: [10.1190/IGCBeijing2016-123](https://doi.org/10.1190/IGCBeijing2016-123).
- Guo, J., and Y. Wang, 2004, Recovery of a target reflection underneath coal seams: *Journal of Geophysics and Engineering*, **1**, 46–50, doi: [10.1088/1742-2132/1/1/005](https://doi.org/10.1088/1742-2132/1/1/005).
- Guo, Y., S. Guo, K. Guo, and H. Zhou, 2021, Seismic data denoising under the morphological component analysis framework by dictionary learning: *International Journal of Earth Sciences*, **110**, 963–978, doi: [10.1007/s00531-021-02001-3](https://doi.org/10.1007/s00531-021-02001-3).
- Ioffe, S., and C. Szegedy, 2015, Batch normalization: Accelerating deep network training by reducing internal covariate shift: *Proceedings of the 32nd International Conference on Machine Learning*, PMLR, 448–456.
- Jiang, X., J. Cao, S. Zu, H. Xu, and J. Wang, 2021, Detection of hidden reservoirs under strong shielding based on bi-dimensional empirical mode decomposition and the Teager-Kaiser operator: *Geophysical Prospecting*, **69**, 1086–1101, doi: [10.1111/1365-2478.13073](https://doi.org/10.1111/1365-2478.13073).
- Jiang, Y., J. Zhang, H. Han, D. Feng, and J. Yu, 2020, Elimination of strong reflection influence based on optimized variational mode decomposition method: A case study of the target processing of beach bar sand of Es4 in Dongying Sag: *Oil Geophysical Prospecting*, **55**, 147–152.
- Liu, D., L. Gao, X. Wang, and W. Chen, 2021a, A dictionary learning method with atom splitting for seismic footprint suppression: *Geophysics*, **86**, no. 6, V509–V523, doi: [10.1190/geo2020-0681.1](https://doi.org/10.1190/geo2020-0681.1).
- Liu, D., W. Wang, X. Wang, C. Wang, J. Pei, and W. Chen, 2020a, Poststack seismic data denoising based on 3-D convolutional neural network: *IEEE Transactions on Geoscience and Remote Sensing*, **58**, 1598–1629, doi: [10.1109/TGRS.2019.2947149](https://doi.org/10.1109/TGRS.2019.2947149).
- Liu, D., X. Wang, W. Chen, Y. Zhou, W. Wang, Z. Shi, C. Wang, and C. Xie, 2019, 3D seismic waveform of channels extraction by artificial intelligence: 89th Annual International Meeting, SEG, Expanded Abstracts, 2518–2522, doi: [10.1190/segam2019-3216216.1](https://doi.org/10.1190/segam2019-3216216.1).
- Liu, D., X. Yang, X. Wang, H. Mao, M. D. Sacchi, and W. Chen, 2021b, Deep learning for prestack strong scattered noise suppression: First International Meeting for Applied Geoscience & Energy, Expanded Abstracts, 1601–1605, doi: [10.1190/segam2021-3583393.1](https://doi.org/10.1190/segam2021-3583393.1).
- Liu, M., and D. Grana, 2019, Accelerating geostatistical seismic inversion using TensorFlow: A heterogeneous distributed deep learning framework: *Computers & Geosciences*, **124**, 37–45, doi: [10.1016/j.cageo.2018.12.007](https://doi.org/10.1016/j.cageo.2018.12.007).
- Liu, P., G. Huang, J. Zhang, L. Ban, and L. Liu, 2014, Stripping technology of beach bar sandstone strong reflection cover based on MP method: Beijing 2014 International Geophysical Conference & Exposition, 884–886, doi: [10.1190/IGCBeijing2014-224](https://doi.org/10.1190/IGCBeijing2014-224).
- Liu, W., Q. Cheng, L. Liu, Y. Wang, and J. Zhang, 2020b, Accelerating high-resolution seismic imaging by using deep learning: *Applied Sciences*, **10**, 2502, doi: [10.3390/app10072502](https://doi.org/10.3390/app10072502).
- Ma, J., and G. Plonka, 2010, The curvelet transform: *IEEE Signal Processing Magazine*, **27**, 118–133, doi: [10.1109/MSP.2009.935453](https://doi.org/10.1109/MSP.2009.935453).
- Marfurt, K. J., R. L. Kirlin, S. L. Farmer, and M. S. Bahorich, 1998, 3-D seismic attributes using a semblance-based coherency algorithm: *Geophysics*, **63**, 1150–1165, doi: [10.1190/1.1444415](https://doi.org/10.1190/1.1444415).
- Park, M. J., J. Jennings, B. Clapp, and B. Biondi, 2020, Seismic data interpolation using a POCS-guided deep image prior: 90th Annual International Meeting, SEG, Expanded Abstracts, 3154–3158, doi: [10.1190/segam2020-3427320.1](https://doi.org/10.1190/segam2020-3427320.1).
- Park, M. J., and M. D. Sacchi, 2020, Automatic velocity analysis using convolutional neural network and transfer learning: *Geophysics*, **85**, no. 1, V33–V43, doi: [10.1190/geo2018-0870.1](https://doi.org/10.1190/geo2018-0870.1).
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986, Learning representations by back-propagating errors: *Nature*, **323**, 533–536, doi: [10.1038/323533a0](https://doi.org/10.1038/323533a0).
- Saad, O. M., Y. A. S. I. Oboue, M. Bai, L. Samy, L. Yang, and Y. Chen, 2021, Self-attention deep image prior network for unsupervised 3-D seismic data enhancement: *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1–14, doi: [10.1109/TGRS.2021.3108515](https://doi.org/10.1109/TGRS.2021.3108515).
- Sardy, S., A. G. Bruce, and P. Tseng, 2000, Block coordinate relaxation methods for nonparametric wavelet denoising: *Journal of Computational and Graphical Statistics*, **9**, 361–379, doi: [10.1080/10618600.2000.1047485](https://doi.org/10.1080/10618600.2000.1047485).
- Shensa, M. J., 1992, The discrete wavelet transform: Wedding the a trous and Mallat algorithms: *IEEE Transactions on Signal Processing*, **40**, 2464–2482, doi: [10.1109/78.157290](https://doi.org/10.1109/78.157290).
- Simonyan, K., and A. Zisserman, 2015, Very deep convolutional networks for large-scale image recognition: arXiv preprint, arXiv:1409.1556.
- Starck, J., M. Elad, and D. L. Donoho, 2005, Image decomposition via the combination of sparse representations and a variational approach: *IEEE Transactions on Image Processing*, **14**, 1570–1582, doi: [10.1109/TIP.2005.852206](https://doi.org/10.1109/TIP.2005.852206).
- Ulyanov, D., A. Vedaldi, and V. Lempitsky, 2018, Deep image prior: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Wang, B., N. Zhang, W. Lu, and J. Wang, 2019, Deep-learning-based seismic data interpolation: A preliminary result: *Geophysics*, **84**, no. 1, V11–V20, doi: [10.1190/geo2017-0495.1](https://doi.org/10.1190/geo2017-0495.1).
- Wang, D., Y. Wang, Y. Zhao, R. Tian, and Y. Xue, 2016, A strong seismic reflection amplitude suppressing method applied in the Ordos Basin: International Geophysical Conference, SPG/SEG, 389–391, doi: [10.1190/IGCBeijing2016-115](https://doi.org/10.1190/IGCBeijing2016-115).
- Wang, W., W. Chen, W. Liu, J. Xu, and J. Gao, 2011, Abrupt feature extraction via the combination of sparse representations: 81st Annual International Meeting, SEG, Expanded Abstracts, 1019–1024, doi: [10.1190/1.3627378](https://doi.org/10.1190/1.3627378).
- Wang, W., J. Gao, W. Chen, and J. Xu, 2012, Data adaptive ground-roll attenuation via sparsity promotion: *Journal of Applied Geophysics*, **83**, 19–28, doi: [10.1016/j.jappgeo.2012.04.004](https://doi.org/10.1016/j.jappgeo.2012.04.004).
- Wang, X., D. Liu, and W. Chen, 2021, Accelerating seismic dip estimation with deep learning: *IEEE Geoscience and Remote Sensing Letters*, **19**, 750505, doi: [10.1109/LGRS.2021.3120315](https://doi.org/10.1109/LGRS.2021.3120315).
- Wu, X., Y. Shi, S. Fomel, L. Liang, Q. Zhang, and A. Z. Yusifov, 2019, FaultNet3D: Predicting fault probabilities, strikes, and dips with a single convolutional neural network: *IEEE Geoscience and Remote Sensing Letters*, **57**, 9138–9155, doi: [10.1109/TGRS.2019.2925003](https://doi.org/10.1109/TGRS.2019.2925003).
- Xu, L., X. Wu, M. Zhang, X. Yin, and Z. Zong, 2019, Strong reflection identification and separation based on the local-frequency-constrained dynamic matching pursuit: *Oil Geophysical Prospecting*, **54**, 587–593.
- Xu, X., G. Qu, Y. Zhang, Y. Bi, and J. Wang, 2016, Ground-roll separation of seismic data based on morphological component analysis in two-dimensional domain: *Applied Geophysics*, **13**, 116–126, doi: [10.1007/s11770-016-0546-0](https://doi.org/10.1007/s11770-016-0546-0).
- Ying, L., L. Demanet, and E. Candes, 2005, 3D discrete curvelet transform: *Proceedings of SPIE*, **5914**, 591413, doi: [10.1117/12.616205](https://doi.org/10.1117/12.616205).
- Yu, S., J. Ma, and W. Wang, 2019, Deep learning for denoising: *Geophysics*, **84**, no. 6, V333–V350, doi: [10.1190/geo2018-0668.1](https://doi.org/10.1190/geo2018-0668.1).
- Yuan, X., P. He, Q. Zhu, and X. Li, 2019, Adversarial examples: Attacks and defenses for deep learning: *IEEE Transactions on Neural Networks and Learning Systems*, **30**, 2805–2824, doi: [10.1109/TNNLS.2018.2886017](https://doi.org/10.1109/TNNLS.2018.2886017).
- Zhang, H., X. Yang, and J. Ma, 2020a, Can learning from natural image denoising be used for seismic data interpolation? *Geophysics*, **85**, no. 4, WA115–WA136, doi: [10.1190/geo2019-0243.1](https://doi.org/10.1190/geo2019-0243.1).
- Zhang, J., Z. Liu, B. Liu, B. Zhu, and D. Feng, 2012, Analysis and identification of reservoir characteristics of weak reflectors under strong shielding layer: *Special Oil and Gas Reservoirs*, **19**, 23–26.
- Zhang, K., Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, 2021a, Plug-and-play image restoration with deep denoiser prior: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 6360–6376, doi: [10.1109/TPAMI.2021.3088914](https://doi.org/10.1109/TPAMI.2021.3088914).
- Zhang, K., W. Zuo, Y. Chen, D. Meng, and L. Zhang, 2017, Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising: *IEEE Transactions on Image Processing*, **26**, 3142–3155, doi: [10.1109/TIP.2017.2662206](https://doi.org/10.1109/TIP.2017.2662206).
- Zhang, M., X. Sun, and X. Yu, 2013, Case studies on removing strong reflections in carbonate and sand reservoir prediction using volume-based seismic waveform decomposition: 83rd Annual International Meeting, SEG, Expanded Abstracts, 1575–1579, doi: [10.1190/segam2013-1282.1](https://doi.org/10.1190/segam2013-1282.1).
- Zhang, S., Z. Zhang, G. Xiao, and X. Duan, 2021b, Strong reflection separation based on seismic phase decomposition and its application in reservoir prediction: First International Meeting for Applied Geoscience & Energy, Expanded Abstracts, 2094–2098, doi: [10.1190/segam2021-3583251.1](https://doi.org/10.1190/segam2021-3583251.1).
- Zhang, Y., X. Gao, Z. Chen, H. Zhong, L. Li, C. Yan, and T. Shen, 2020b, Learning salient features to prevent model drift for correlation tracking: *Neurocomputing*, **418**, 1–10, doi: [10.1016/j.neucom.2019.12.006](https://doi.org/10.1016/j.neucom.2019.12.006).
- Zhang, Z., J. Zhang, X. Gong, H. Zhang, and N. Dong, 2015, The method research of stripping strong reflection of coal seams: 77th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.201412753](https://doi.org/10.3997/2214-4609.201412753).
- Zhu, B., X. Xiang, H. Gao, Q. Ran, and M. Wang, 2019, Application of strong seismic reflection separation based on matching pursuit algorithm in Tarim Basin, China: 89th Annual International Meeting, SEG, Expanded Abstracts, 3379–3383, doi: [10.1190/segam2019-3215310.1](https://doi.org/10.1190/segam2019-3215310.1).

Biographies and photographs of the authors are not available.