

# A dictionary learning method with atom splitting for seismic footprint suppression

Dawei Liu<sup>1</sup>, Lei Gao<sup>1</sup>, Xiaokai Wang<sup>1</sup>, and Wenchao Chen<sup>1</sup>

## ABSTRACT

The acquisition footprint causes serious interference with seismic attribute analysis, which severely hinders accurate reservoir characterization. Therefore, acquisition footprint suppression has become increasingly important in industry and academia. We have assumed that the time slice of 3D poststack migration seismic data mainly comprises two components: useful signals and the acquisition footprint. Useful signals describe the spatial distributions of geologic structures with local piecewise smooth morphological features. However, the acquisition footprint often behaves as periodic artifacts in the time-slice domain. In particular, the local morphological features of the acquisition footprint in marine seismic acquisition appear as stripes. Because useful signals and the acquisition footprint have different morphological features, we can train an adaptive dictionary and divide the atoms of the dictionary into two subdictionaries to reconstruct these two

components. We have devised an adaptive dictionary learning method for acquisition footprint suppression in the time slice of 3D poststack migration seismic data. To obtain an adaptive dictionary, we use the K-singular value decomposition algorithm to sparsely represent the patches in the time slice of 3D poststack migration seismic data. Each atom of the trained dictionary represents certain local morphological features of the time slice. According to the difference in the variation level between the horizontal and vertical directions, the atoms of the trained dictionary are divided into two types. One type significantly represents the local morphological features of the acquisition footprint, whereas the other type represents the local morphological features of useful signals. Then, these two components are reconstructed using morphological component analysis based on different types of atoms, respectively. Synthetic and field data examples indicate that our method can effectively suppress the acquisition footprint with fidelity to the original data.

## INTRODUCTION

Acquisition footprint refers to the noise component in 3D seismic data that is highly correlated to seismic acquisition and processing (Marfurt et al., 1998). This component tends to “mirror” parts of the acquisition geometry in the lateral deployment of sources and receivers, which are used to acquire seismic survey data (Chopra and Larsen, 2000) and manifest inline or crossline striation in a given time slice. The acquisition footprint masks real amplitude anomalies for stratigraphic interpretation, amplitude-variation-with-offset (AVO) analysis, and reservoir attribute studies (Marfurt et al., 1998), thereby making data interpretation problematic.

The causes of acquisition footprints vary and can generally be categorized into two groups (Drummond et al., 2000). One type

of footprint is related to the design of the acquisition geometry, including the distribution of source and receiver lines. Irregular offset and azimuth distribution belong to this type of footprint problem, and the resulting acquisition footprint is usually a periodic coherent noise. The other type of footprint arises from signal processing problems such as aliased noise leakage when attenuating ground roll or multiples, incorrect velocities when applying normal moveout, and AVO effects.

An essential method for suppressing footprint is selecting appropriate field parameters and designing optimal acquisition geometry. However, carrying out these approaches is typically costly in practice. The data volumes of the poststack data are greatly reduced compared with the originally recorded data from the acquisition step (prestack data). Many researchers aim to suppress the acquisition footprint when processing poststack data.

Manuscript received by the Editor 14 September 2020; revised manuscript received 19 July 2021; published ahead of production 19 August 2021; published online 19 October 2021.

<sup>1</sup>Xi'an Jiaotong University, School of Information and Communications Engineering, Xi'an 710049, China. Email: 409791715@qq.com; 280436373@qq.com; xkwang@xjtu.edu.cn; wenchchen@xjtu.edu.cn (corresponding author).

© 2021 Society of Exploration Geophysicists. All rights reserved.

Filtering methods are the most commonly used and can be classified into two categories: filtering along with slices of seismic data and structure-oriented filtering (Davogustto and Marfurt, 2011). In the first class, the  $k_x$ - $k_y$  filter (Falconer and Marfurt, 2008) and the truncated singular value decomposition (TSVD) filter (Al-Bannagi et al., 2005) have been applied to suppress the acquisition footprint in time slices. However, the  $k_x$ - $k_y$  domain filtering method would cause some distortion to the useful signals, and the spatial coherence of the acquisition footprint makes the singular value decomposition (SVD)-based method less effective. In the other class, scholars used a smoothing operation parallel to the reflector dip and azimuth unless a significant discontinuity exists (Fehmers and Höcker, 2003), thus suppressing the acquisition footprint effectively as it changes with depth. Gómez and Velis (2019) further examine the attenuation performance of structure-oriented filtering in the frequency domain and show that it improves the lateral continuity of events. In addition, the wavelet transform (Alali et al., 2018) has also been used to suppress the acquisition footprint. Yu et al. (2017) use a 3D complex wavelet transform to suppress the acquisition footprint and other random noise on prestack gathers. However, a single method rarely yields a good result, and a combination of methods is recommended (Davogustto and Marfurt, 2011).

Starck et al. (2004, 2005) present the theory of morphological component analysis (MCA) and decompose images into textural and natural additive ingredients. MCA provides a complete representation of the content of an image and has been applied in a wide range of fields such as image inpainting (Elad et al., 2005), image superresolution (Shen et al., 2010), and image denoising (Yong et al., 2009). In addition, the MCA method has been widely applied in seismic data processing because seismic data can be described as a linear superposition of multiple basic waveforms with different morphological features (Chen et al., 2017). Here, the term “morphology” refers to the geometric structures of seismic signals in the spatial dimension. A fundamental consideration in using the MCA method is the choice of overcomplete complex waveform dictionaries, which can be used to sparsely represent signals under different practical problems. The overcomplete dictionaries can be divided into two main categories: predetermined dictionaries and adaptive dictionaries. Many linear transforms have been chosen as the predetermined dictionaries and have facilitated rapid numerical implementation for sparse seismic data. Yarham et al. (2006) use various methods to estimate the ground roll and separate the ground roll from the recorded signals by exploiting the curvelet transform domain, which represents the ground roll sparsely while preserving reflector information. Wang et al. (2011) construct an overcomplete dictionary in MCA with the 2D undecimated wavelet transform and curvelet transform to extract sedimentary features from seismic data. Based on the coherency of reflective events, Wang et al. (2012) propose a data-adaptive ground-roll attenuation method using the stationary wavelet transform (SWT), which can play the role of a discriminant between ground-roll and reflection waveforms. Based on the morphological divergence between ground-roll and body-wave signals, Chen et al. (2013) present a ground-roll noise attenuation method with 1D SWT for the ground roll and local discrete cosine transforms for the body waves. To suppress ground roll, Chen et al. (2017) construct dictionaries that can represent body waves and ground-roll

sparingly using tunable  $Q$ -factor wavelet transforms with a low  $Q$ -factor and a high  $Q$ -factor. Each predetermined dictionary has its own merits, but none can represent all features completely. Adaptive dictionaries based on learning comprise a different route for designing dictionaries and are optimally adapted for sparsely representing signals. Olshausen and Field (1996) propose a construction dictionary strategy with an unsupervised learning algorithm subject to maximizing sparseness coding for natural images. The K-singular value decomposition (K-SVD) is an important dictionary learning algorithm for creating a dictionary for sparse representations. Rubinstein et al. (2010) use an efficient K-SVD-like algorithm to learn sparse dictionaries, and the algorithm provided better generalization. Turquais et al. (2017) extract a redundant dictionary by learning the morphological diversity using the K-SVD algorithm. Then, they divide the dictionaries into noise and signal subdictionaries according to a statistical classification to suppress coherent noise in the raw shot gather from a marine seismic data set via MCA. However, they require a noise-only model or a pure signal model to segregate the atoms, and the widespread spatial distribution of the footprints cannot meet this requirement. In a follow-up study, Gómez and Velis (2020) use an atom filtering method to classify atoms in the dictionary, hence suppressing the footprint. To summarize, with the purpose of using adaptive dictionaries to remove coherent noise, the basis of this approach is to divide dictionaries into two appropriate subdictionaries: one for the representation of coherent noise and the other for the useful signals.

In this paper, we present a local MCA method with an adaptive dictionary to suppress acquisition footprint in the time slices of 3D poststack migration seismic data. In the proposed method, we divide the time slice into many patches and train an adaptive dictionary to decompose each patch linearly. The adaptive dictionary aims at finding a sparse representation of the input data. In other words, the minimum number of atoms used to linearly approximate the patches is very small. In practice, solving the sparse dictionary learning problem can be performed using the K-SVD algorithm (Aharon et al., 2006). After the K-SVD algorithm, the dictionary is a matrix comprising atoms that represent different morphological features.

In addition, we propose a method for classifying the atoms of the sparse dictionary into two classes by examining their morphological features. It is necessary to add that there are additional options for suppressing other kinds of noise by third or more atom classification. Here, two classes are mainly concerned with the suppression of the footprint. Once we obtain atoms representing the footprint patterns, we can use them to reconstruct the footprint component and obtain the denoised result by subtracting the footprint from the input. The local morphological features of the acquisition footprint and the useful signals are significantly different. In the time slices, the local morphological features of the acquisition footprint appear as thin stripes extending approximately along the inline or crossline direction. However, the local morphological features of the useful signals appear as curve-like structures (that describe the geologic body or geologic structure such as a channel or fault) or piecewise continuous (that represents stable sedimentary strata). Each atom of the dictionary represents one kind of local morphological feature of the time-slice data. Therefore, we assume that the atoms representing the local morphological features of the acquisition footprint vary slowly along the horizontal or vertical direction

and vary rapidly along the other orthogonal direction. However, the atoms for representing the local morphological features of useful signals vary slowly in both directions.

The time slice of 3D poststack migration seismic data can be modeled as a superposition of the acquisition footprint component, useful data component, and random noise. Here, random noise indicates a component that has little or no correlation to the atoms and cannot be represented sparsely by the dictionary. To date, several studies have investigated the textural attributes for seismic data classification (Vinther et al., 1995; Vinther, 1997; West et al., 2002; Gao, 2003). Subsequently, textural attributes are used to classify MCA atoms (Turquais et al., 2017). In this study, we classify the atoms in a similar manner. Based on these differences in directional variation among the atoms, a parameter of the atoms called the directional variation difference (DVD) is presented and used to classify the atoms into the two categories. With a threshold of DVD, the sparse dictionary can be split into two subdictionaries: one consisting of atoms representing local morphological features of the acquisition footprint, and the other consisting of atoms representing local morphological features of the useful signals. According to local MCA theory (Elad, 2010), the atoms in the former subdictionary can be used to reconstruct the acquisition footprint component of each patch, and the atoms in the latter subdictionary can be used to reconstruct the useful signal component of each patch. The error of the linear approximation of the patch can express random noise because random noise does not have regular local morphological features.

The remainder of this paper is organized as follows. First, the theory and methodology are presented. Then, two applications of our theory are discussed. The acquisition footprint noise is removed from the time slices of a synthetic data set and a 3D poststack migration seismic data set from a marine acquisition geometry.

## METHODS

The proposed method is composed of three steps: learning dictionary construction, atom classification, and reconstruction. The three steps are presented in the following subsections.

### Learning dictionary construction

Following Gao et al. (2016), a time slice  $\mathbf{d}$  of 3D poststack migration seismic data (e.g., shown in Figure 1) is modeled as a superposition of useful signal components  $\mathbf{s}$ , acquisition footprint components  $\mathbf{f}$ , and random noise  $\mathbf{n}$  (to facilitate mathematical expression, all 2D slices of seismic data are reorganized as column vectors in this section) as follows:

$$\mathbf{d} = \mathbf{f} + \mathbf{s} + \mathbf{n}. \quad (1)$$

One patch of  $\mathbf{d}$  is denoted by  $\mathbf{p}_l$ , where  $l$  represents the location of the patch. Each patch is a local part of  $\mathbf{d}$  and can be obtained by

$$\mathbf{p}_l = \mathbf{R}_l \mathbf{d}, \quad (2)$$

where the matrix  $\mathbf{R}_l$  is the operator to split out the patch at the  $l$ th location of the corresponding data. The patches are analyzed with local MCA (Elad, 2010), which is an application of the adaptive dictionary-based sparse representation theory.

We denote matrix  $\mathbf{A}$  as the dictionary, which can sparsely linearly approximate each patch of the time slice data, and each column of  $\mathbf{A}$  is an atom of the dictionary  $\mathbf{A}$ . The term  $\mathbf{x}_l$  denotes the coefficient vector to linearly approximate the patch at the  $l$ th location of the time-slice data with the dictionary  $\mathbf{A}$ . According to the definition,  $\mathbf{A}$  and  $\mathbf{x}_l$  can be solved by optimizing the dictionary  $\mathbf{A}$  and coefficient vector  $\mathbf{x}$  using the K-SVD algorithm (Aharon et al., 2006):

$$\begin{aligned} \{\mathbf{A}, \{\mathbf{x}_l\}_{l=1}^L\} &= \arg \min_{\mathbf{A}, \{\mathbf{x}_l\}_{l=1}^L} \sum_{l=1}^L \|\mathbf{R}_l \mathbf{d} - \mathbf{A} \mathbf{x}_l\|_2^2 \\ \text{s.t. } \|\mathbf{x}_l\|_0 &\leq T_0, \quad 1 \leq l \leq L, \end{aligned} \quad (3)$$

where  $L$  is the total number of the patch location;  $\|\cdot\|_0$  is the  $\ell_0$ -norm, which counts the total number of nonzero elements of a vector;  $\|\mathbf{x}_l\|_0$  is the number of the nonzero elements of  $\mathbf{x}_l$ ; and  $T_0$  is the sparsity level, which is the upper bound of the atom number to linearly approximate each patch. Equation 3 indicates that each patch of the time slice data  $\mathbf{d}$  can be linearly approximated by relatively few of the dictionary atoms  $\mathbf{A}$ . In other words, equation 3 imposes a constraint on each atom to represent a certain basic morphological feature across different patches. Thus, all atoms form a collection of local morphological features of the time slice data  $\mathbf{d}$ , which are mixtures of useful signals and footprint patterns.

Equation 3 is nonconvex with regard to  $\{\mathbf{A}, \{\mathbf{x}_l\}_{l=1}^L\}$ . To stably solve this problem, an optimization method based on the block-coordinate descent has been proposed to split the dictionary learning problem into two simpler subproblems. Each subproblem can be easily solved with respect to a single variable by assuming that the other parameters are known.

We first assume that the dictionary  $\mathbf{A}$  is known. The subproblem of finding the sparse coding coefficients  $\{\mathbf{x}_l\}_{l=1}^L$  is a classic problem:

$$\mathbf{x}_l = \arg \min_{\mathbf{x}_l} \|\mathbf{R}_l \mathbf{d} - \mathbf{A} \mathbf{x}_l\|_2^2 \quad \text{s.t. } \|\mathbf{x}_l\|_0 \leq T_0, \quad 1 \leq l \leq L, \quad (4)$$

where  $\mathbf{R}_l$  is the sampling operator used to extract and vectorize the patch from  $\mathbf{d}$ . Many algorithms, such as iterative soft thresholding (Daubechies et al., 2004) and the interior-point solver (Chen et al., 1998), can be used to solve the problem. We use the orthogonal matching pursuit method (Pati et al., 1993) to approximate the solution because it achieves suitable calculation speeds with high accuracy. In addition, choosing a suitable initial dictionary helps obtain the desired solution. We chose the 2D redundant discrete cosine dictionary (DCT) as the initial dictionary in our method, as shown in Figure 2a. The atom size is  $16 \times 16$ , and the redundancy, which is the ratio of the total number of atoms to the length of the atom vector, is four.

The second step is to update the dictionary  $\mathbf{A}$  under the condition that all coefficients are known. The dictionary  $\mathbf{A}$  is updated column by column by minimizing the following objective function:

$$\mathbf{A} = \arg \min_{\mathbf{A}} \sum_{l=1}^L \|\mathbf{R}_l \mathbf{d} - \mathbf{A} \mathbf{x}_l\|_2^2 \quad \text{s.t. } \|\mathbf{a}_j\|_2 \leq 1, \quad (5)$$

where  $\mathbf{a}_j$  indicates the columns of the dictionary  $\mathbf{A}$ . The constraint in equation 5 can prevent the slow convergence rate caused by an arbitrarily large atom amplitude. We use SVD to solve this problem according to the K-SVD method because it is quick and straightforward in the calculation.

### Atom classification method

As observed in the time slice of 3D poststack migration seismic data in Figure 1, there exists significant morphological diversity between the acquisition footprint and the useful signals. The useful signal  $\mathbf{s}$  represents the underground geology, whereas the acquisition footprint  $\mathbf{f}$  mimics parts of the acquisition geometry. Therefore, the useful signal  $\mathbf{s}$  shows a slow variation and local linear singularity patterns, whereas the acquisition footprint  $\mathbf{f}$  shows a stripe-like pattern approximately perpendicular to the crossline direction. In other words, the useful signal  $\mathbf{s}$  appears piecewise smooth, whereas the acquisition footprint  $\mathbf{f}$  is characterized by texture.

Different atoms in the dictionary  $\mathbf{A}$  also exhibit significant morphological diversity. In the initial dictionary shown in Figure 2a, some atoms have stripe-like features along the vertical direction (such as elliptical markings) and have approximately morphological features of the acquisition footprint  $\mathbf{f}$ . The other part of the atoms does not have noticeable stripe features (such as rectangular markings) and have approximately the morphological features of useful signals  $\mathbf{s}$ . Let the sparsity level  $T_0$  be 10, and after K-SVD training, a dictionary of sparse representation of the original data patch is obtained, as shown in Figure 2b. We found that after training, the patterns of atoms marked by the black ellipse are closer to the morphological features of the acquisition footprint  $\mathbf{f}$  compared to the initial dictionary. Similarly, the atoms marked by the black rectangle are transformed into atoms that have closer morphological features of the useful signals  $\mathbf{s}$ .

Based on the preceding analysis, the atoms of  $\mathbf{A}$  can be classified into two categories: the atoms representing the local morphological features of the useful signals  $\mathbf{s}$  and those representing the acquisition footprint  $\mathbf{f}$ . Random noise  $\mathbf{n}$ , which does not have

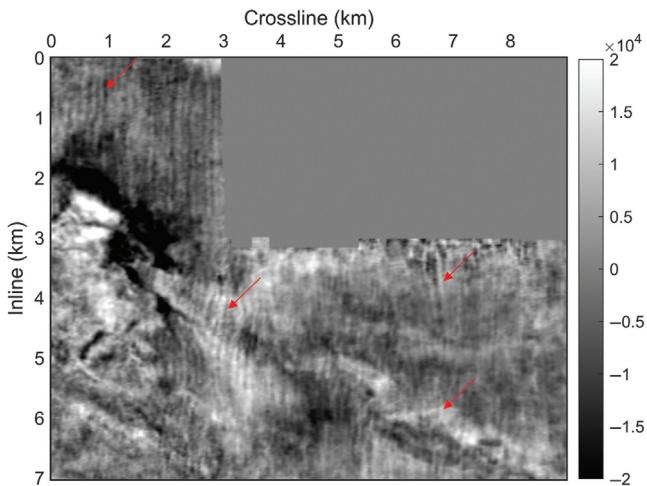


Figure 1. A time slice at 1700 ms in 3D seismic data acquired in the East China Sea (Gao et al., 2016). This footprint pattern is evident in the time slice as indicated by the red arrows; therefore, we used it for the parameter selection of our method.

any regular local morphological features, is handled as a third component. Note that random noise here is a by-product of noise attenuation with dictionary learning and does not require intentional suppression.

How to group atoms is a key issue when suppressing coherent noise with dictionary learning. In contrast to the one-class classifier (Turquais et al., 2017) or atom filtering (Gómez and Velis, 2020), we present a novel atom classification method according to atom difference in the variation level along the horizontal and vertical directions, which correspond to the crossline and inline directions of seismic data in Figure 1. In the time slice, the local patterns of the acquisition footprint appear as thin stripes extending approximately along the inline or crossline direction; therefore, each of the atoms for representing local morphological features of the acquisition footprint vary slowly along one direction and rapidly along the other direction. The local patterns of the useful signals can be divided into two types: One represents the geologic body or geologic structure such as the channel or fault, appearing as curve-like or point-like features, and the other represents the stable sedimentary strata and appears piecewise continuous. Therefore, the variation level of the atoms for representing the local morphological features of the useful signals does not show a significant difference between the inline and crossline directions. The difference function can measure the variation level. According to the difference function, the variation level along the inline and crossline directions, respectively, of the atom  $\mathbf{a} = (a_{i,j})_{i,j=1,2,\dots,N}$  with the size  $N \times N$  is defined as follows:

$$V_{\text{inline}}(\mathbf{a}) = \sum_{j=1}^N \sum_{i=1}^{N-1} |a_{i+1,j} - a_{i,j}|, \quad (6)$$

$$V_{\text{xline}}(\mathbf{a}) = \sum_{i=1}^N \sum_{j=1}^{N-1} |a_{i,j+1} - a_{i,j}|. \quad (7)$$

Then, the DVD, which is the difference of the variation level along the inline and along the crossline direction, can be expressed as

$$\text{DVD}(\mathbf{a}) = |V_{\text{inline}} - V_{\text{xline}}|. \quad (8)$$

If  $\text{DVD}(\mathbf{a})$  is greater than a threshold  $\text{DVD}_{\text{thr}}$ , the atom  $\mathbf{a}$  has horizontal or vertical stripe features similar to the morphological features of the acquisition footprint so that it can be identified as an atom of the dictionary  $\mathbf{A}^f$  to represent the acquisition footprint. Otherwise, the atom  $\mathbf{a}$  does not have apparent stripe features and can be identified as an atom of dictionary  $\mathbf{A}^s$  to represent the useful signals.

### Footprint suppression through reconstruction

Let the dictionary  $\mathbf{A}^s$  be the set of atoms representing the local morphological feature of the useful signals  $\mathbf{s}$ , and let  $\mathbf{x}_i^s$  be the vector consisting of the corresponding coefficients. Let the dictionary  $\mathbf{A}^f$  be the set of atoms for representing the acquisition footprint  $\mathbf{f}$ , and let  $\mathbf{x}_i^f$  be the vector consisting of the corresponding coefficients.

Then, the useful signal component  $\mathbf{p}_l^s$  and the acquisition footprint component  $\mathbf{p}_l^f$  of each patch  $\mathbf{p}_l$  (for  $l = 1, 2, \dots, L$ ) can be represented as follows:

$$\mathbf{p}_l^s = \mathbf{A}^s \mathbf{x}_l^s, \quad (9)$$

$$\mathbf{p}_l^f = \mathbf{A}^f \mathbf{x}_l^f. \quad (10)$$

To stably obtain  $\mathbf{p}_l^s$  and  $\mathbf{p}_l^f$ , we add a constraint that  $\mathbf{x}_l^s$  and  $\mathbf{x}_l^f$  are sparse, as follows:

$$\min_{\mathbf{x}_l^s, \mathbf{x}_l^f} \|\mathbf{p}_l - \mathbf{A}^s \mathbf{x}_l^s - \mathbf{A}^f \mathbf{x}_l^f\|_2^2 + \lambda_s \|\mathbf{x}_l^s\|_1 + \lambda_f \|\mathbf{x}_l^f\|_1, \quad (11)$$

where  $\lambda_s$  and  $\lambda_f$  are regularization parameters. We solve equation 11 based on the MCA method through iterative calculation of the following equations:

$$\hat{\mathbf{x}}_l^s = T_\alpha(\mathbf{A}^{s\dagger}(\mathbf{p}_l - \mathbf{A}^f \mathbf{x}_l^f)), \quad (12)$$

$$\hat{\mathbf{x}}_l^f = T_\alpha(\mathbf{A}^{f\dagger}(\mathbf{p}_l - \mathbf{A}^s \mathbf{x}_l^s)), \quad (13)$$

where  $T_\alpha$  is a soft thresholding function with threshold  $\alpha$  and  $\mathbf{A}^{s\dagger}$  and  $\mathbf{A}^{f\dagger}$  are the pseudoinverse matrices of  $\mathbf{A}^s$  and  $\mathbf{A}^f$ , respectively. Finally, the useful signal component  $\mathbf{s}$  and the acquisition footprint component  $\mathbf{f}$  can be calculated by combining the corresponding component of each patch as follows:

$$\mathbf{s} = \left( \sum_{l=1}^L \mathbf{R}_l^T \mathbf{R}_l \right)^{-1} \sum_{l=1}^L \mathbf{R}_l^T \mathbf{p}_l^s, \quad (14)$$

$$\mathbf{f} = \left( \sum_{l=1}^L \mathbf{R}_l^T \mathbf{R}_l \right)^{-1} \sum_{l=1}^L \mathbf{R}_l^T \mathbf{p}_l^f. \quad (15)$$

## EXAMPLES

We illustrate in this section the efficiency of the proposed method on a synthetic data set and a marine-migrated stack data set from China. After the synthetic data example, we conducted two experiments to select the suitable parameters for the field example. In addition, we compare the results against a wavenumber filtering and TSVD method (Al-Bannagi et al., 2005) in the field data application.

### Synthetic data application

We first use the synthetic data to demonstrate the validity of the proposed method. Figure 3a displays the clean 3D cube, which consists of  $300 \times 300 \times 300$  samples. The time-sampling interval is 1 ms, and the total recording time is 300 ms. From the time slice in Figure 3a, it is apparent that the synthetic data contain two intersecting faults. Accordingly, the surrounding events show a sharp bending deformation in the inline section and crossline section, which poses a challenge to fault preservation. We use the textural pattern that decays with time to mimic the morphological component of footprint noise and add it to the clean cube, as shown in Figure 3. It can be seen from the color bar that the amplitude range of added noise is between -1 and 1, the same as clean data. The amplitude of the noise is relatively strong, which seriously distorts the useful signals. Next, the proposed method is applied to suppress the acquisition footprint with the DVD threshold of 2.4. Figure 3c and 3d presents the separated useful signals and the removed footprint noise, respectively. The results show that our method can effectively suppress the footprint noise and that the fault structure is well preserved. It is noticeable that the dictionary used for the synthetic data example is just the initial DCT dictionary because the synthetic data are relatively simple compared to the field data, which also proves the rationality of our choice of the initial dictionary.

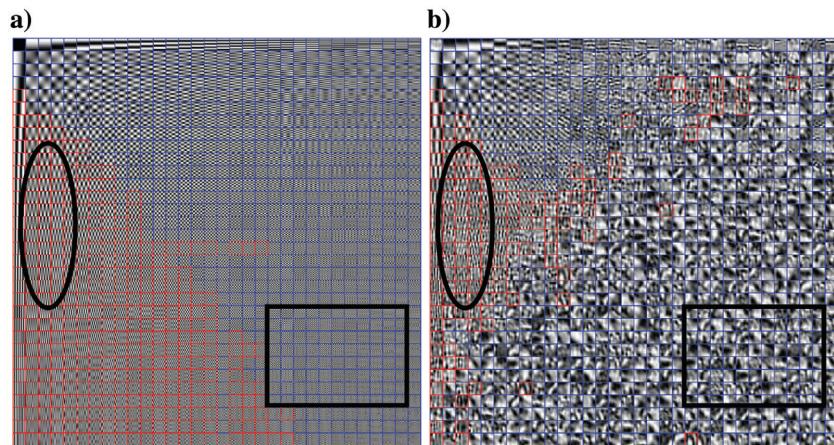


Figure 2. Training dictionary using the K-SVD method. (a) The initial dictionary. The atoms can be divided into two categories according to their DVD values. The red boxes point out the atoms whose values are greater than the DVD threshold. They have stripe features along the vertical direction and have approximately morphological features of the acquisition footprint. The other atoms do not have noticeable stripe features and exhibit morphological features of useful signals. (b) The training results. After training the dictionary, the patterns of red atoms marked by the black ellipse are closer to the morphological features of the acquisition footprint compared to that of the initial dictionary. Similarly, the blue atoms marked by the black box are transformed into atoms with closer morphological features of the useful signals.

### Parameter selection

The dictionary learning problem with the nonconvex constraint shown in equation 3 is a nondeterministic polynomial-time hard problem. Therefore, it is intractable for field seismic data, and we can only obtain a series of approximate solutions. Selecting different training dictionary parameters results in diverse atoms that satisfy the constraints of equation 3. We categorize the atoms into two types and obtain two subdictionaries mainly representing the local morphological characteristics of the acquisition footprint and useful signals,

respectively. Once the atoms of the well-trained dictionary are easier to classify, the useful signals and acquisition footprint noise can be separated more accurately. In addition to the training parameters, the classification threshold of DVD, used to classify the atoms in the dictionary, is also a crucial parameter. A common concern is how to choose these parameters correctly. With the aid of the time slice shown in Figure 1, we investigate the behavior of the proposed method for different values of these parameters.

The training dictionary parameters include the size of the atom, the redundancy of the dictionary, and the sparsity level. A time

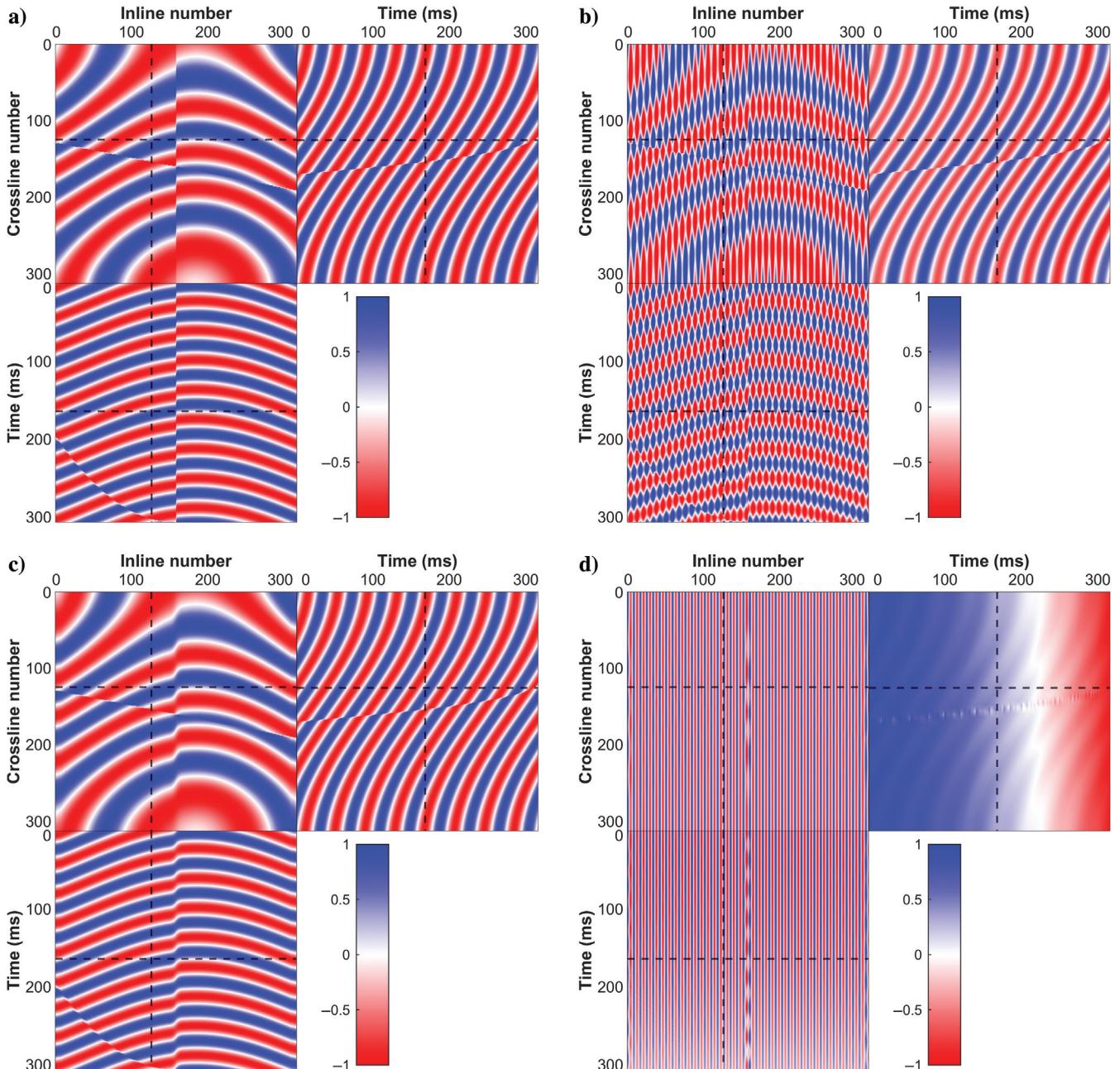


Figure 3. Synthetic data example. (a) Clean 3D cube. (b) Noisy 3D cube contaminated by coherent noise with textural features. (c) The result after coherent noise reduction. (d) Residual.

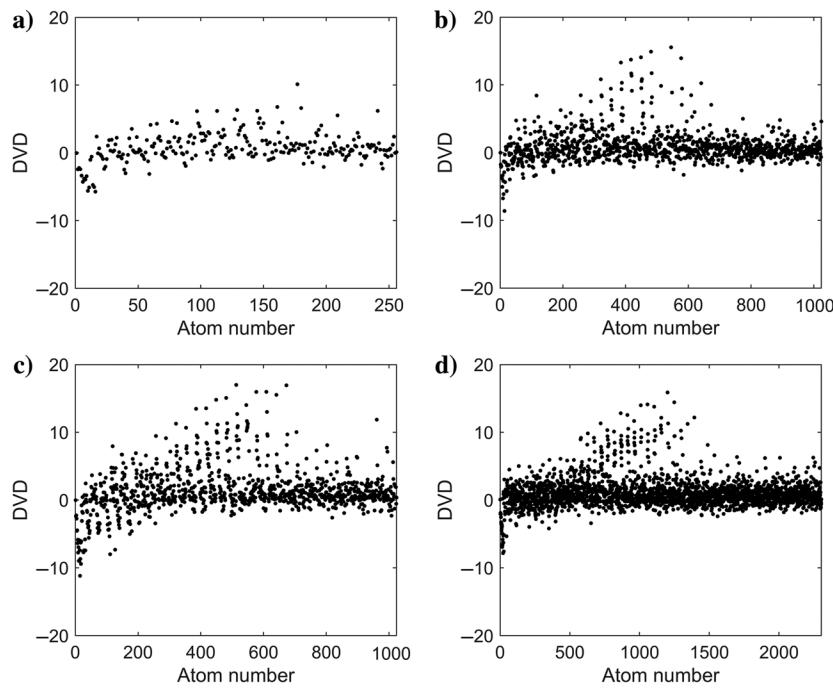


Figure 4. Parameter selection for dictionary training from the aspect of DVD distribution. The parameters include atom size, redundancy, and sparsity level. When the atom size is  $16 \times 16$ , the difference in the parameter DVD of each atom is larger than that when the atom size is  $8 \times 8$ , and it is easier to classify the atoms by the parameter DVD. (a) The atom size is  $8 \times 8$ , redundancy is four, and the sparsity level is four. (b) The atom size is  $16 \times 16$ , redundancy is four, and the sparsity level is four. (c) The atom size is  $16 \times 16$ , redundancy is four, and the sparsity level is 10. (d) The atom size is  $16 \times 16$ , redundancy is nine, and the sparsity level is four.

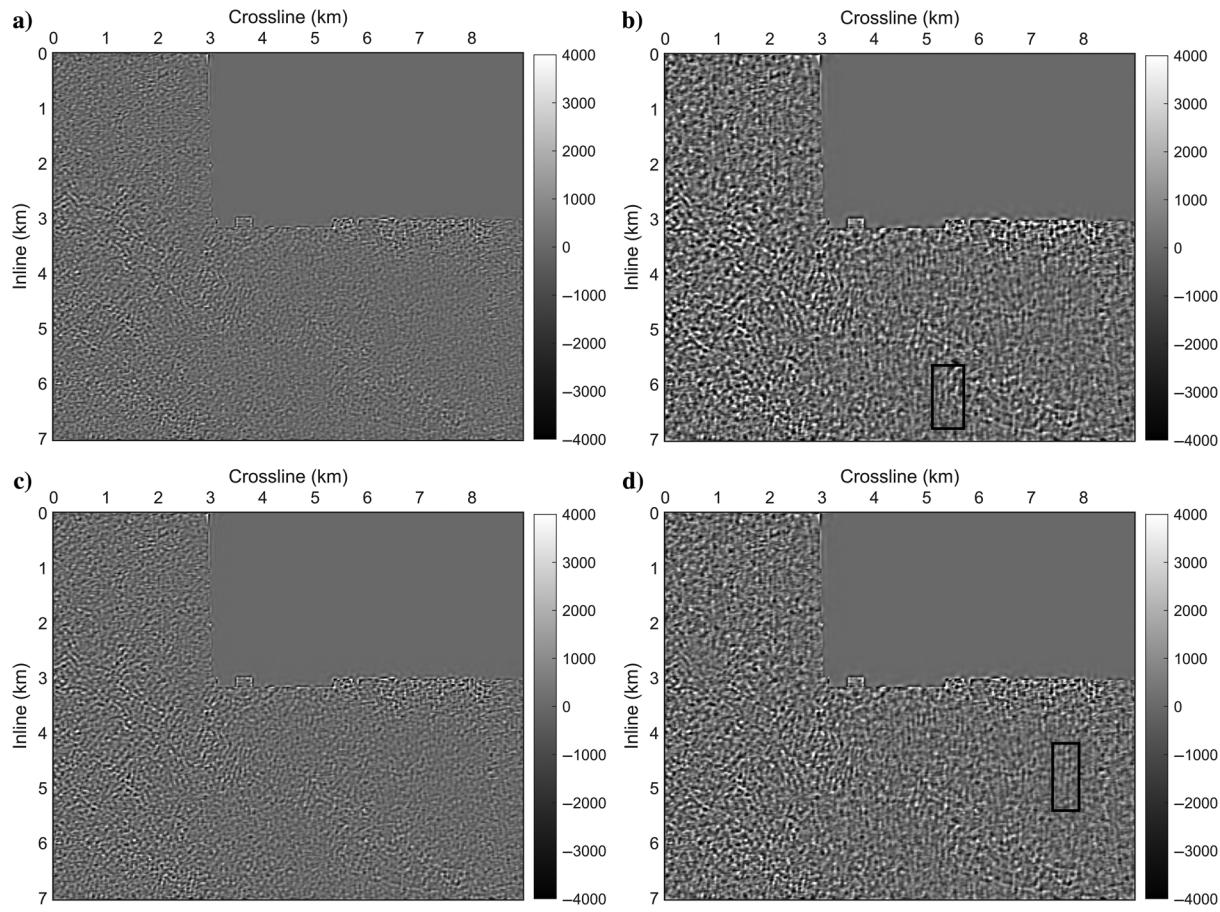


Figure 5. Parameter selection for dictionary training from the aspect of reconstruction time-slicing data error. The parameters include atom size, redundancy, and sparsity level. A significant acquisition footprint (e.g., rectangular regions) is observed in (b and d), and no significant acquisition footprint is observed in (c). (a) The atom size is  $8 \times 8$ , redundancy is four, and the sparsity level is four. (b) The atom size is  $16 \times 16$ , redundancy is four, and the sparsity level is four. (c) The atom size is  $16 \times 16$ , redundancy is four, and the sparsity level is 10. (d) The atom size is  $16 \times 16$ , redundancy is nine, and the sparsity level is four.

slice **d** is divided into a group of data patches of the same size as in equation 2, each of which is represented sparsely by the atoms of the dictionary. The size of the atom, which determines the dimensions of the linear space described by the dictionary, is the same

size as the patch. For example, if the size of an atom has  $16 \times 16$  samples, the dimension of the linear space described in the dictionary is  $16 \times 16 = 256$ . The redundancy of the dictionary is the total number of atoms in the dictionary divided by the size

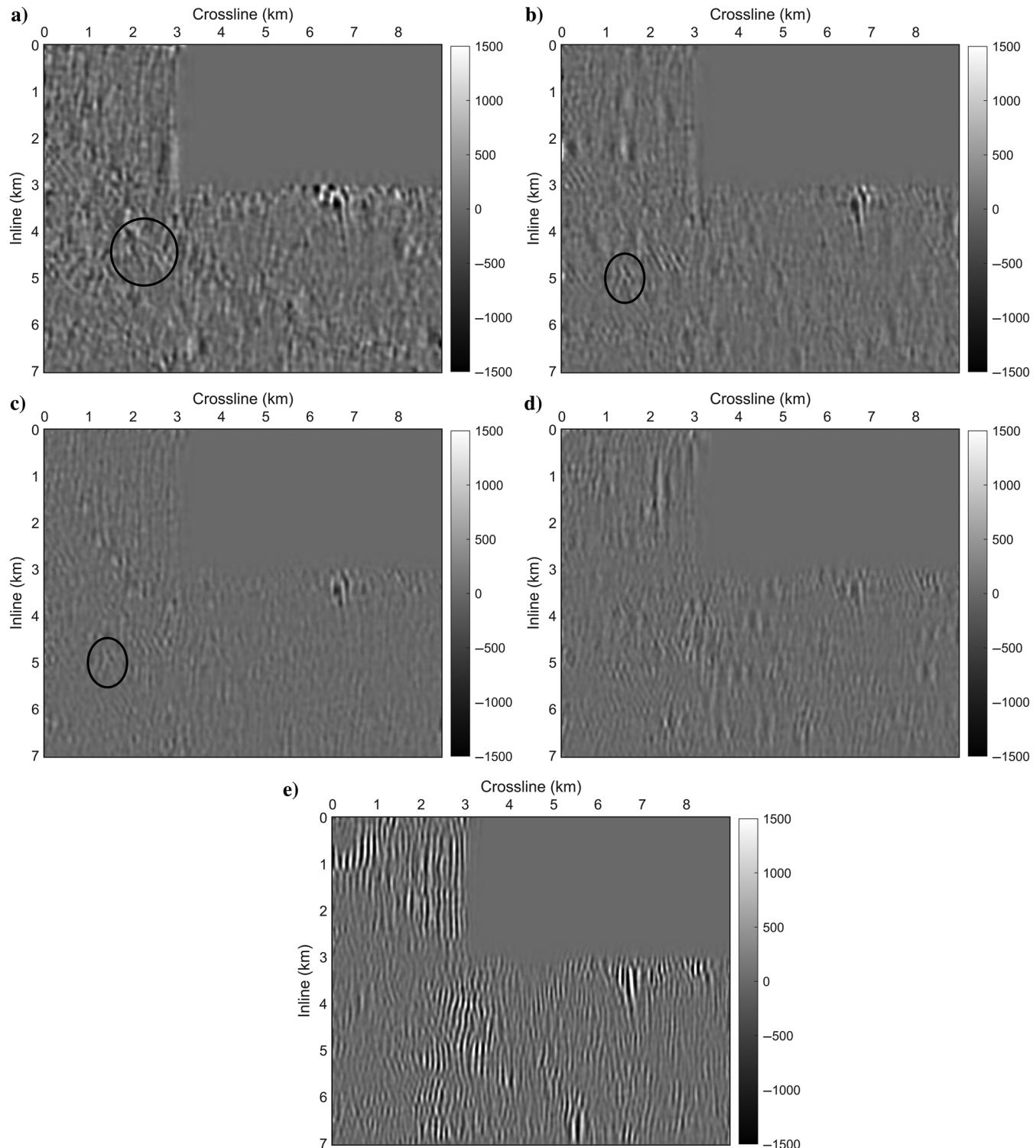


Figure 6. Reconstructing time-slicing data in different ranges of DVD. When using the  $\text{DVD} < 6$  atoms, there is some signal leakage, for example, elliptical regions: (a)  $3 \leq \text{DVD} < 4$ ; (b)  $4 \leq \text{DVD} < 5$ ; (c)  $5 \leq \text{DVD} < 6$ ; (d)  $6 \leq \text{DVD} < 7$ ; and (e)  $\text{DVD} \geq 7$ .

of one atom. For example, if the size of an atom is  $16 \times 16$  and the number of atoms is  $16 \times 16 \times 4 = 1024$ , the redundancy of the dictionary is four.

Figures 4 and 5 show the DVD calculated from atoms of the four dictionaries with different parameters and the error of reconstructing time-slicing data with each dictionary, respectively. The larger the size of the atom, the more effective the representation of the local morphological feature corresponding to the time slice. Therefore, it is easier to classify atoms according to their morphological features. Comparing Figure 4a and 4b, we can see that when the atom size is  $16 \times 16$ , the difference in the DVD of each atom is larger than that when the atom size is  $8 \times 8$ , and it is easier to classify the atoms by DVD. Comparing Figure 4b–4d, there is no apparent change in DVD in the dictionary obtained by changing the redundancy and sparsity level. Figure 5 shows the corresponding reconstruction errors of the different parameters in Figure 4. Figure 5b and 5d contains significant acquisition footprints (e.g., rectangular regions), and no significant acquisition footprint is observed in Figure 5c. This example shows that increasing the redundancy of the dictionary cannot effectively reduce the acquisition footprint in the reconstruction error, whereas increasing the sparsity level can effectively reduce the acquisition footprint in the reconstruction error. Reconstructing the useful signals directly and reconstructing the acquisition footprint first before subtracting it from the input are two methods of suppressing the acquisition footprint. Moreover, if we reconstruct the footprint and subtract it from the input, the sparsity level is a sensitive parameter that must be large enough to capture complete footprint patterns. No clear structural information of useful signals is observed in Figure 5b–5d, indicating that the learned dictionaries under these redundancy and sparsity levels contain all signal patterns. In other words, the subdictionary of useful signals can reconstruct the complete useful signals and accomplish the objective of suppressing the acquisition footprint. In addition, because we directly obtain useful signals, random noise also can be attenuated as a by-product.

Figure 6 shows the reconstruction results of the time-slicing data at different ranges of DVD. The atomic reconstruction results contain useful signals in the atoms of  $3 \leq \text{DVD} < 4$ ,  $4 \leq \text{DVD} < 5$ , and  $5 \leq \text{DVD} < 6$  (e.g., elliptical regions), whereas there are no obvious useful signals in the reconstruction results of  $6 \leq \text{DVD} < 7$  and  $\text{DVD} \geq 7$  atoms. Therefore, we can suppress the acquisition footprint under the premise of the maximum to preserve the structure of the useful signals when using the  $\text{DVD} < 6$  atoms and  $\text{DVD} \geq 6$  atoms to reconstruct the useful signals and the acquisition footprint, respectively.

In summary, compared with the choice of atom size as the  $8 \times 8$  samples, the size of the atom with the  $16 \times 16$  samples can more effectively suppress the acquisition footprint noise because a larger atom size can represent more complex morphological components, which reduces the atom correlation and makes it easier to classify atoms. However, a large atom size is computationally intensive. Therefore, we make a trade-off and set the atom size as  $16 \times 16$ . Dictionary redundancy and the sparsity level are not sensitive parameters for directly separating useful signals. Their values over a wide range can ensure that there is little signal leakage in the reconstruction error. However, one of our comparison methods is a  $k_x \cdot k_y$  filter, which cannot attenuate random noise. Therefore, for fairness, we separate the footprints from the input data.

That is, the reconstruction error, which is the third component of our method, is added back to the separated useful signals. To better suppress footprint noise, the reconstruction error cannot contain the footprint. From the perspective of reconstruction error in Figure 5, a sparsity level of 10 is better than four. There is no obvious acquisition footprint in the reconstruction error when taken as 10, whereas there is a significant acquisition footprint structure when the sparsity level is taken as four. Therefore, we select the following parameters to suppress the acquisition footprint: The atom size is  $16 \times 16$ , the redundancy is four, the sparsity level is 10, and the initial dictionary is a redundant 2D DCT. The dictionary trained by the K-SVD method can be obtained, as shown in Figure 2b. Although the results presented subsequently are satisfactory, the parameters are empirically selected, and a trade-off is made between denoising performance and computational cost. Several questions remain regarding selecting parameters and the relationship between these parameters and the size and distance of the footprint stripes.

#### Field data application

To examine the effectiveness of our proposed method, we show the suppression results of 3D poststack migration field data in Figure 7. This volume is a marine data set acquired from the East China Sea, consisting of 300 traces and 564 lines. The spatial sampling intervals in the inline and crossline directions are 12.5 and 30.0 m, respectively. In the time direction, its range is from 1000 to 1700 ms with a 2 ms time-sampling interval. As shown in Figure 6, this data set suffers from an acquisition footprint problem due to insufficient sampling intervals between towed streamers.

The proposed method is applied to the presented data, following the parameters in the previous section. We randomly selected 10 time slices and performed 10 iterations to train the dictionary. One baseline

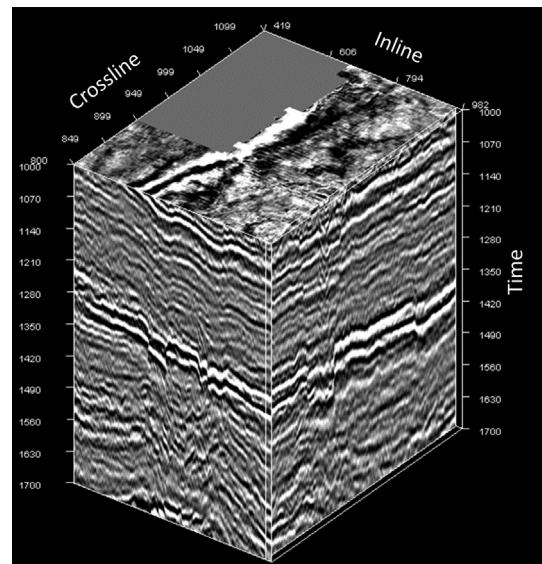


Figure 7. A 3D cube of a marine data set acquired in the East China Sea. We use these data to examine the effectiveness of the proposed method.

method is the  $k_x$ - $k_y$  filter. Figure 8a shows the wavenumber domain of the original time slice at 1700 ms. It is challenging to select the appropriate filtering region, and improper filtering usually causes significant damage to the original data. In this case, we apply a 2D filter in the wavenumber domain using commercial software, which can smooth the filtering boundary between the pass and stop bands to reduce signal distortion. We use our results (shown in Figure 8b and 8c) as an initial guide for the filtering region of two circular bands, and then we fine-tuned until the footprint patterns were well removed. Figure 8d and 8e displays the wavenumber domain of the filtered footprint and useful signals, respectively. The wavenumber domain of random noise is outside the filtering regions; hence, our proposed method adds the residual into separated useful signals for an impartial comparison. We also compare the suppression results with the TSVD method proposed by Al-Bannagi et al. (2005). Time

slices are cropped into overlapped patches of size  $21 \times 21$ , and they overlap on 15 samples in both dimensions. The first six ranks are reserved as useful signals. It should be stated that the three methods maintain the same parameter settings for processing all time slices.

We first compared the three methods in a time slice at 1040 ms, as shown in Figure 9. Figure 9a shows that this time slice is contaminated by footprint noise. Figure 9b and 9c shows the results suppressed by the  $k_x$ - $k_y$  filter and the residual. We observe that the acquisition footprint noise is suppressed effectively because we have carefully selected the filter parameters for a fair comparison. Even so, the  $k_x$ - $k_y$  filter smears the useful signals and causes some damage to the dipping structures, as indicated by the black circles. TSVD is also a powerful tool for suppressing acquisition footprint artifacts, but it causes too much damage to the useful signals, as shown in the ellipse regions in Figure 9d. Figure 9e displays the

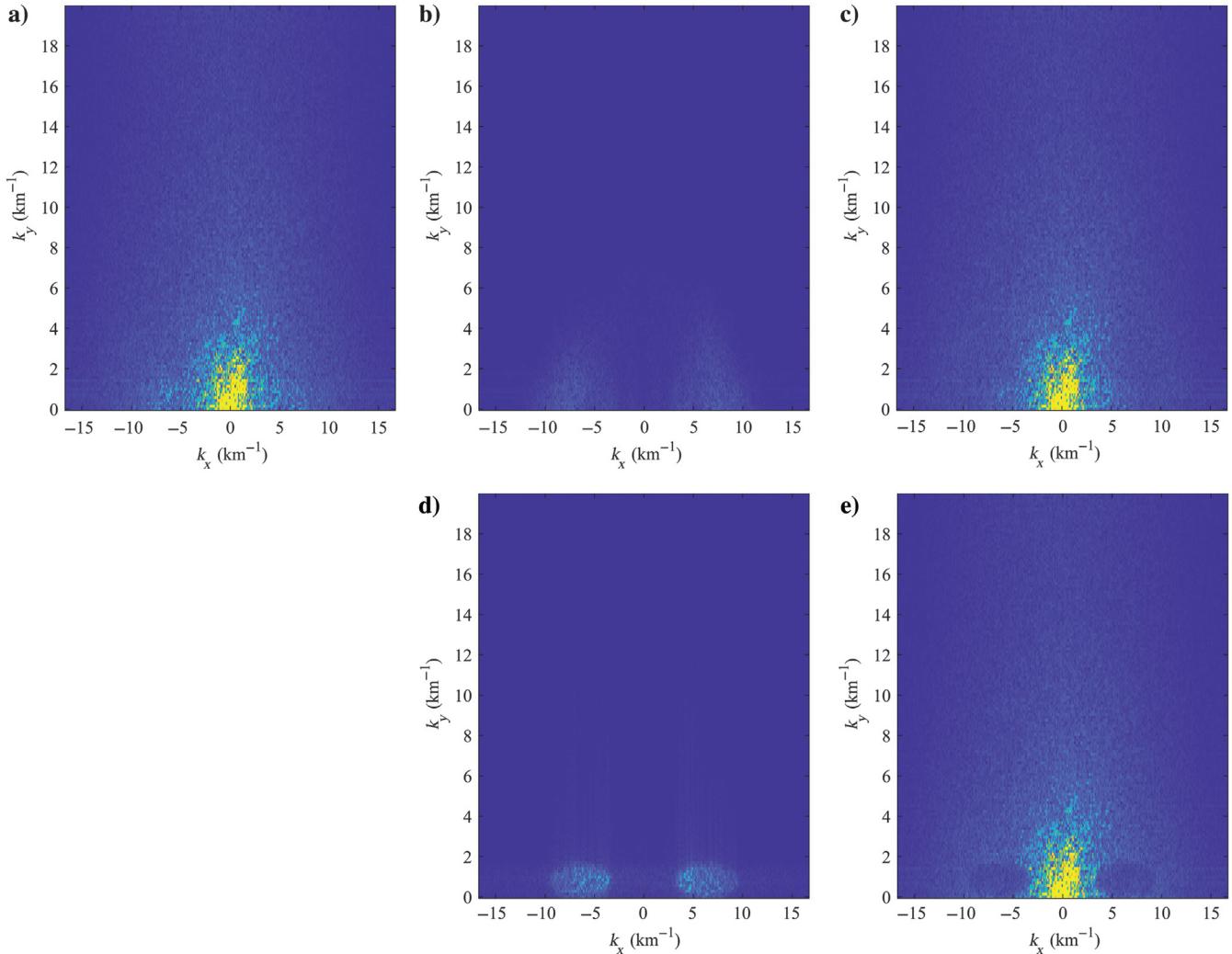


Figure 8. A 2D fast Fourier transform of the time slice at 1700 ms. We apply a well-designed  $k_x$ - $k_y$  to suppress the footprint. The initial filtering region of two circular bands is guided by our method in (b and c). (a) The wavenumber domain of the original time slice. (b) The wavenumber domain of the separated footprint with our method. (c) The wavenumber domain of the separated useful signals with our method. (d) The wavenumber domain of the separated footprint with a  $k_x$ - $k_y$  filter. (e) The wavenumber domain of the separated useful signals with a  $k_x$ - $k_y$  filter.

acquisition footprint noise removed by the TSVD method, and the figure clearly indicates leakage of useful signals. The results obtained from our method are presented in Figure 9f and 9g. We observed that the proposed method actualizes an effective footprint suppression with no harm to the structures.

The same conclusion can be obtained from the time slice at 1678 ms in Figure 10. The time slice is blurred after  $k_x$ - $k_y$  filtering, as displayed in Figure 10b. Because the acquisition footprint noise of the low wavenumber overlaps with the useful signals, the  $k_x$ - $k_y$  filtering method may harm the useful signals. Especially in the region indicated by the black circles, some artifacts are introduced, whereas our method preserves the structure of useful signals. An inspection of the white circles in Figure 10d and 10e reveals that signal leakage caused by TSVD significantly affects the sharpness of the filtered data. Our method effectively suppresses the footprint

while preserving the structure of useful signals. It is important to point out that the two time slices in this example are not used to train the dictionary, which indicates that our method has good generalization ability.

To further compare the denoising results, two lines in Figure 10a are selected for section inspections. Figure 11a shows an inline section at 4.4875 km, which contains footprint patterns. Figure 11b and 11c shows the filtered section by  $k_x$ - $k_y$  filtering, and we see that the footprint suppression result is acceptable. However, some signal patterns are distorted where the acquisition footprint overlaps signals, as indicated by the black ellipses. Looking at Figure 11d and 11e, it is apparent that many footprint patterns are left after TSVD. The white circle regions indicate that some signal structure leaks into the residual. From Figure 11f and 11g, we can see that the data structures and amplitudes are well preserved with our proposed

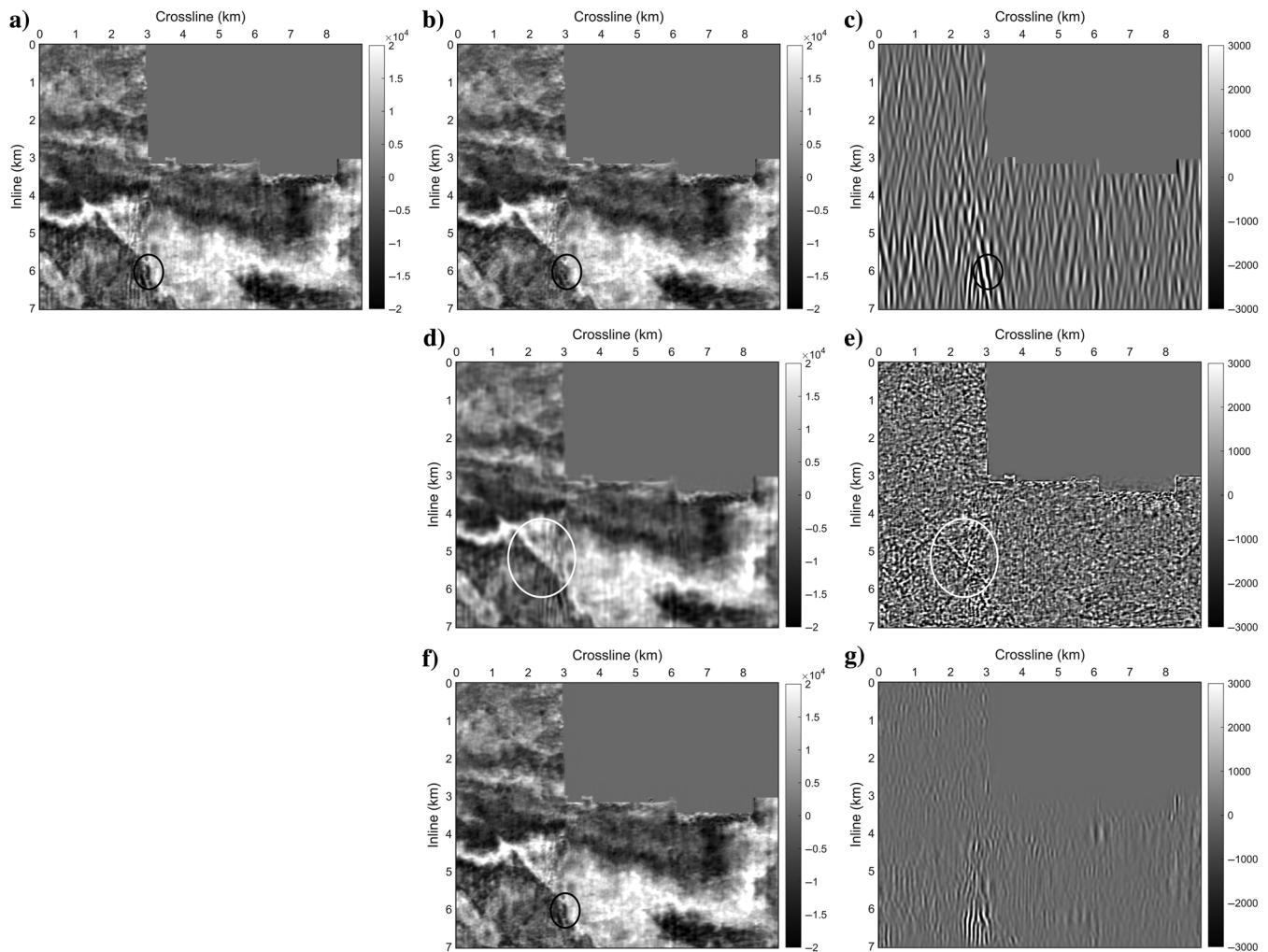


Figure 9. Results of the time slice at 1040 ms. The  $k_x$ - $k_y$  filter smears the useful signals and causes some damage to the dipping structures, as indicated by the black circles. TSVD causes too much damage to the useful signals, as shown in the ellipse regions in (d and e). The proposed method in (f and g) actualizes an effective footprint suppression with no harm to the structures. (a) Original time slice. (b) Result after footprint noise suppression by a  $k_x$ - $k_y$  filter. (c) Residual after footprint noise suppression by a  $k_x$ - $k_y$  filter. (d) Result after footprint noise suppression by TSVD. (e) Residual after footprint noise suppression by TSVD. (f) Result after footprint noise suppression by the proposed method. (g) Residual after footprint noise suppression by the proposed method.

method, whereas footprint noise is removed. Compared with the preceding figures, it can be seen that, although the removed noise energy of our method is slightly weaker than that of TSVD and the  $k_x$ - $k_y$  filtering, the filtered events of our method have better continuity, especially at approximately 1.35 s. Therefore, our method has better fidelity.

Figure 12a shows a crossline section at 1.74 km. The filtered results of the  $k_x$ - $k_y$  filtering are presented in Figure 12b and 12c, which yields acceptable footprint suppression. We found that the damage is obvious in the upper right part of the section, as illustrated by the black circles, perhaps because we keep the filtering parameters fixed and they are not appropriate for those patterns. This also reveals that the filter is poorly adaptive and that the parameter selection is complicated. For the suppression results of TSVD shown in Figure 12d and 12e, significant signal leakage was observed in the white ellipse

regions. Compared to the  $k_x$ - $k_y$  filtering and TSVD, our proposed method can accomplish this task without signal leakage and does not significantly affect the sharpness of the separated useful signals. In accordance with the inline results, our method can be adapted to different structures because all time slices have been effectively separated into two corresponding components even when they change over time.

## DISCUSSION

Following the MCA theory (Elad et al., 2005), we assume that the original seismic data are a mixture of useful signals, footprint noise, and residual noise. We trained an adaptive dictionary to learn the morphological features of the original data. The components of useful signals and footprint noise in each patch can be

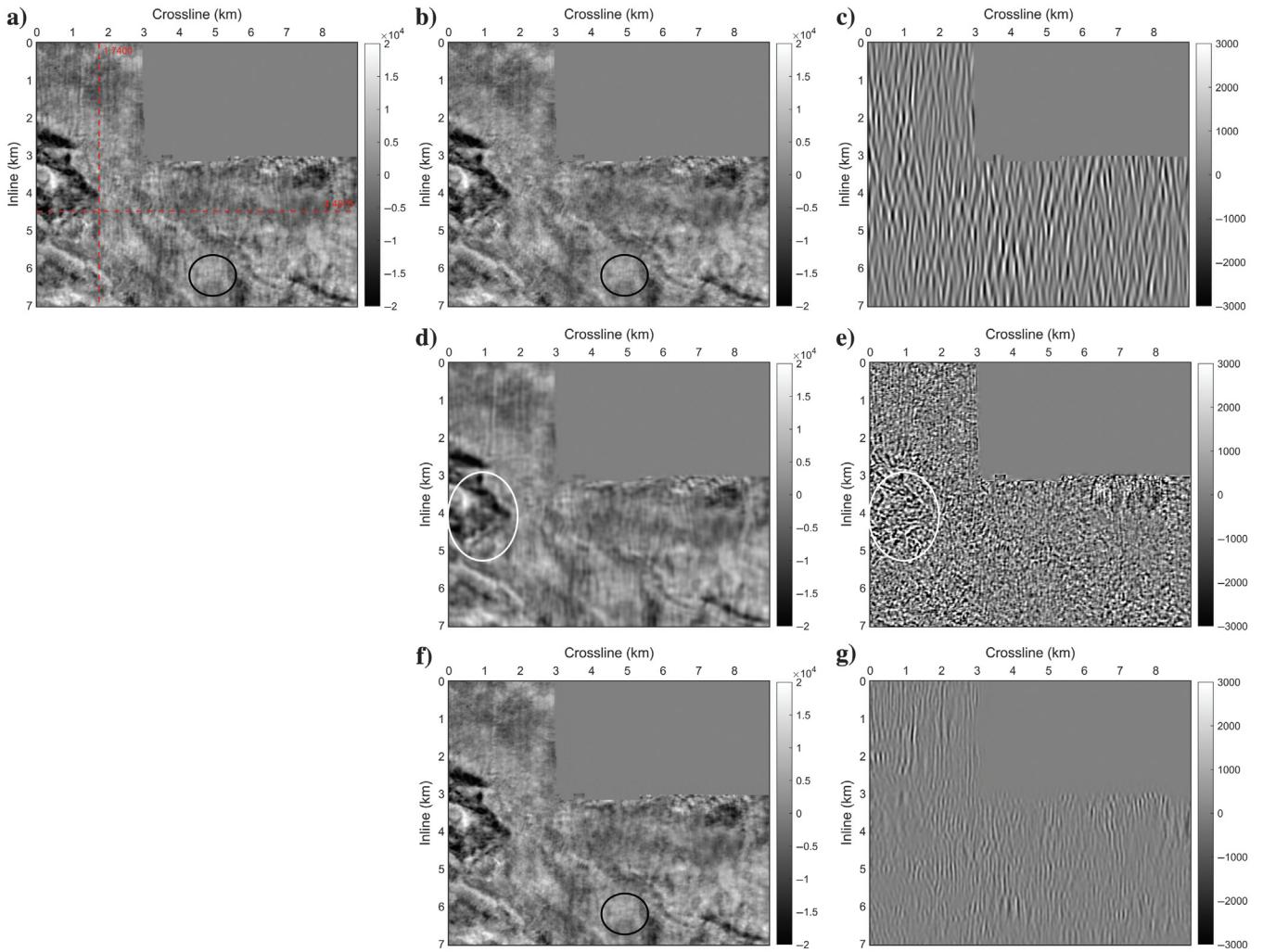


Figure 10. Results of the time slice at 1678 ms. The black circles indicate that the  $k_x$ - $k_y$  filtering method introduces some artifacts, whereas our method preserves the structure of useful signals. The white circles in (d and e) reveal that signal leakage caused by TSVD significantly affects the sharpness of the filtered data. Our method in (f and g) effectively suppresses the footprint while preserving the structure of useful signals. The two red lines in (a) denote the position of the sections for subsequent comparison. (a) Original time slice. (b) Result after footprint noise suppression by a  $k_x$ - $k_y$  filter. (c) Residual after footprint noise suppression by a  $k_x$ - $k_y$  filter. (d) Result after footprint noise suppression by TSVD. (e) Residual after footprint noise suppression by TSVD. (f) Result after footprint noise suppression by the proposed method. (g) Residual after footprint noise suppression by the proposed method.

separated by reconstructing them using two subdictionaries. It is undeniable that constructing appropriate subdictionaries is the key to the success of our method. We follow two principles to obtain the two subdictionaries. On the one hand, the morphological differences between the atoms of the two dictionaries should be large enough to distinguish them properly; to achieve this, we selected a larger atom size, but this is a trade-off because it increases the computation time. On the other hand, the number of atoms and atom morphological diversity in each dictionary should be large enough to ensure a sparse representation. In addition to the redundancy and sparsity level, the initial dictionary also needs to be selected carefully. The redundant 2D DCT exhibits significant morphological diversity because their atoms have separate stripe patterns and piecewise-smooth patterns. After training, there is

still a considerable number of atoms with stripe morphological features. An alternative method for determining the initial dictionary is to use original data themselves (Elad and Aharon, 2006), that is, randomly selecting some patches from the training data set. However, the patterns of the acquisition footprint are mixed with the useful signals in almost every atom of this initial dictionary, and it is difficult to separate them after training. Therefore, there are not enough atoms in the acquisition footprint subdictionary to represent the footprint noise sparsely and the suppression performance is inferior to our method. Because the subdictionaries contain sufficient morphological features of useful signals and collected footprints, respectively, the transfer learning capability of our method is suitable even if the parameters remain the same for different time slices.

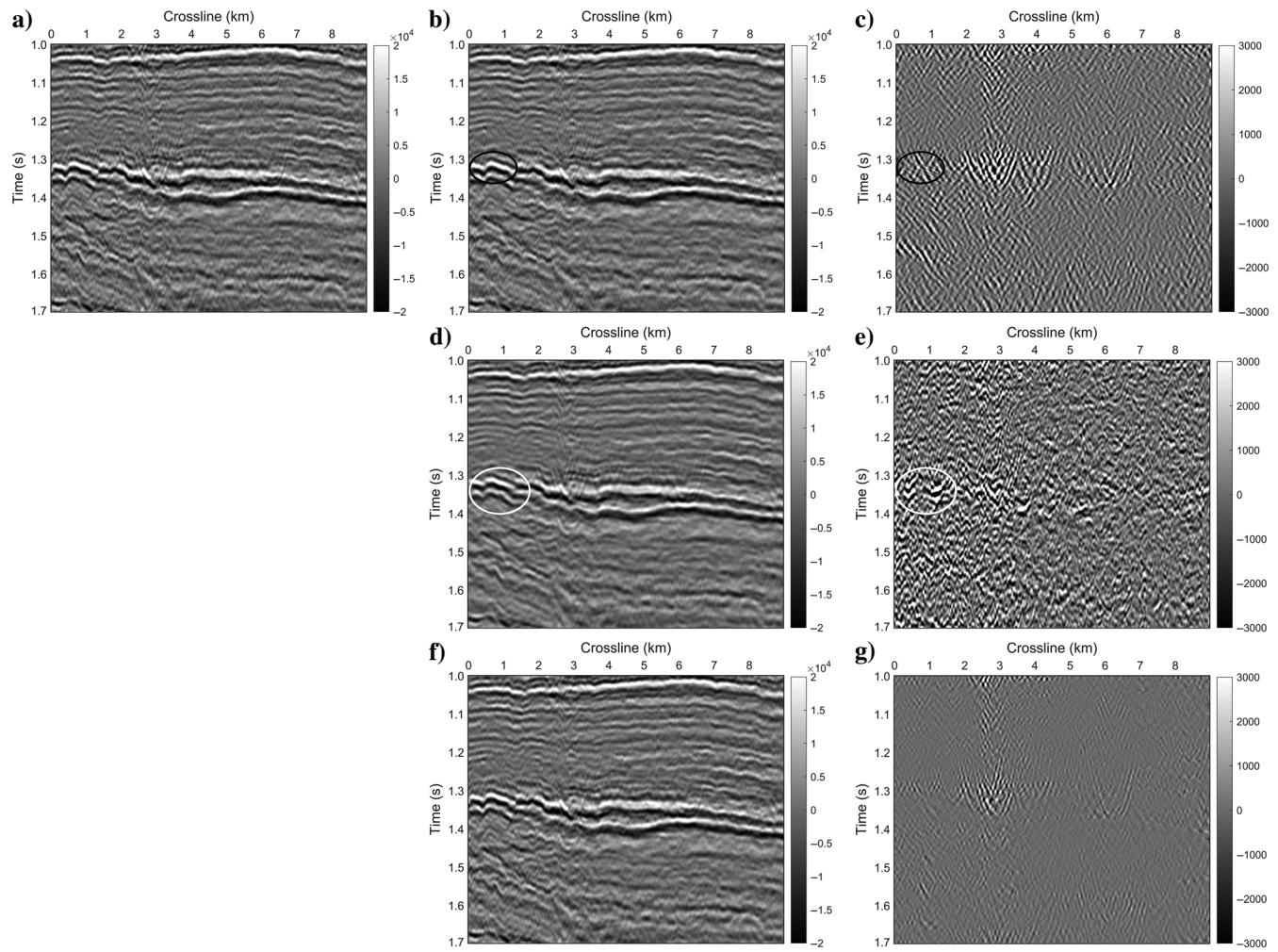


Figure 11. Results of the inline section at 4.4875 km. The black circles in (b and c) indicate that the  $k_x-k_y$  filtering method introduces some artifacts. The white circles in (d and e) reveal that there is some signal leakage caused by TSVD. Our method in (f and g) effectively suppresses the footprint while preserving the structure of useful signals. (a) Original inline section. (b) Result after footprint noise suppression by a  $k_x-k_y$  filter. (c) Residual after footprint noise suppression by a  $k_x-k_y$  filter. (d) Result after footprint noise suppression by TSVD. (e) Residual after footprint noise suppression by TSVD. (f) Result after footprint noise suppression by the proposed method. (g) Residual after footprint noise suppression by the proposed method.

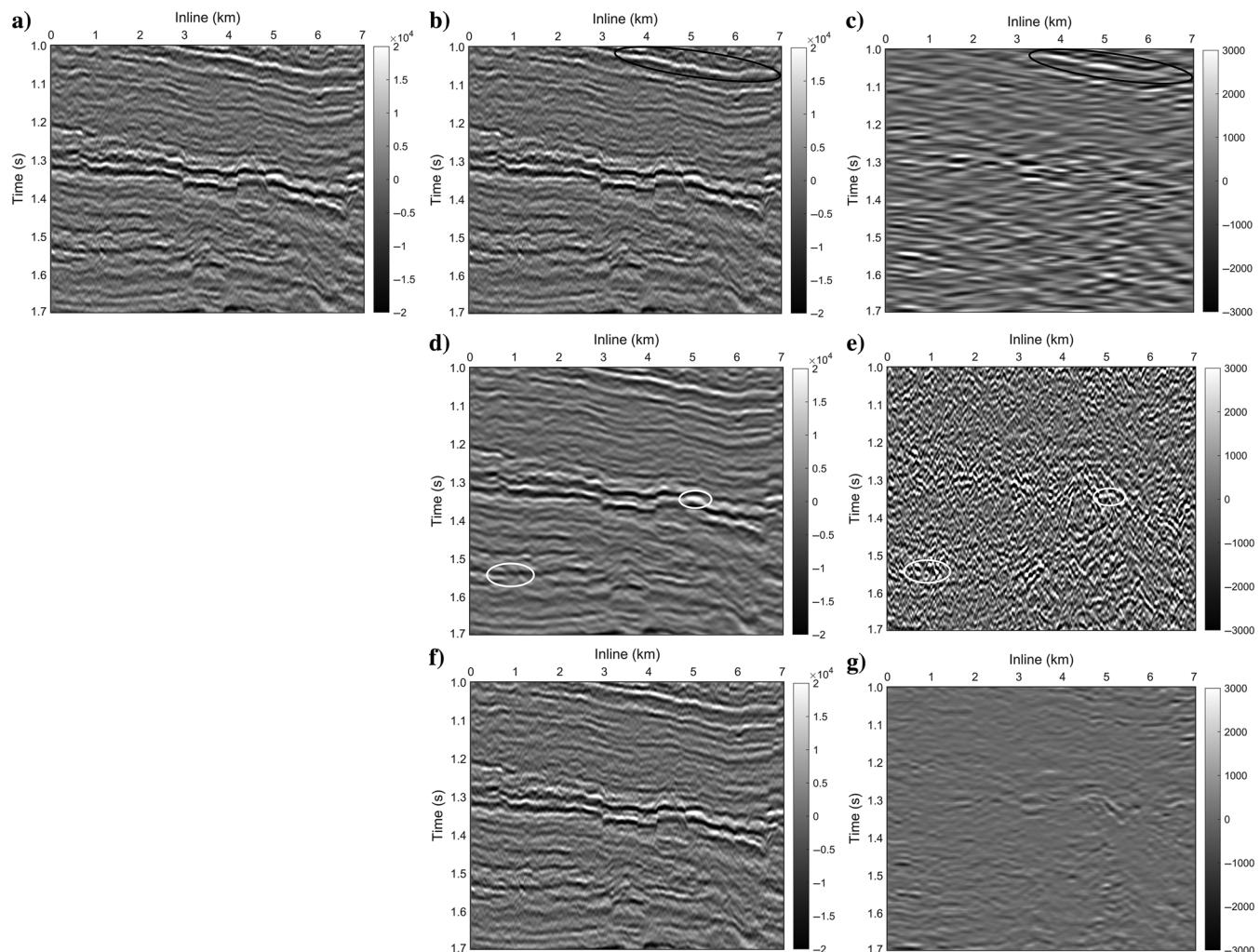


Figure 12. Results of the crossline section at 1.74 km. The black circles indicate that some signal patterns leak into the residual with the  $k_x$ - $k_y$  filtering method. From the white circles in (d and e), we find that signal leakage caused by TSVD significantly affects the sharpness of the filtered data. Our method pertains to the structure of useful signals well. (a) Original crossline section. (b) Result after footprint noise suppression by a  $k_x$ - $k_y$  filter. (c) Residual after footprint noise suppression by a  $k_x$ - $k_y$  filter. (d) Result after footprint noise suppression by TSVD. (e) Residual after footprint noise suppression by TSVD. (f) Result after footprint noise suppression by the proposed method. (g) Residual after footprint noise suppression by the proposed method.

## CONCLUSION

We have presented an acquisition footprint suppression method based on local MCA and adaptive dictionary sparse representation of time slices. The adaptive dictionary can sparsely represent each patch of the time slice data with atoms representing the local morphological features of the time slice. Furthermore, we presented a dictionary-splitting method based on DVD, which divides the adaptive dictionary into two subdictionaries. One subdictionary consists of atoms representing the local morphological features of the acquisition footprint, whereas the other subdictionary consists of atoms representing the local morphological features of the useful signals. Thus, the footprint suppression problem can be solved by reconstructing useful signals with the corresponding dictionary. We used synthetic and field data examples to test the performance of our method. The results indicate that the proposed method can effectively suppress the acquisition footprint with strong robustness

and adaptability. Compared to  $k_x$ - $k_y$  filtering and TSVD, our proposed method shows higher fidelity in useful signal reconstruction with less signal leakage.

## ACKNOWLEDGMENTS

This research is sponsored by National Natural Science Foundation of China (41774135, 41974131). Many thanks go to the assistant editor E. Gasperikova, and the anonymous associate editor and reviewers for their insightful suggestions.

## DATA AND MATERIALS AVAILABILITY

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

## REFERENCES

- Aharon, M., M. Elad, and A. Bruckstein, 2006, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation: *IEEE Transactions on Signal Processing*, **54**, 4311–4322, doi: [10.1109/TSP.2006.881199](https://doi.org/10.1109/TSP.2006.881199).
- Alali, A., G. Machado, and K. J. Marfurt, 2018, Attribute-assisted footprint suppression using a 2D continuous wavelet transform: *Interpretation*, **6**, no. 2, T457–T470, doi: [10.1190/INT-2017-0175.1](https://doi.org/10.1190/INT-2017-0175.1).
- Al-Bannagi, M. S., K. Fang, P. G. Kelamis, and G. S. Douglass, 2005, Acquisition footprint suppression via the truncated SVD technique: Case studies from Saudi Arabia: *The Leading Edge*, **24**, 832–834, doi: [10.1190/1.2032259](https://doi.org/10.1190/1.2032259).
- Chen, S. S., D. L. Donoho, and M. A. Saunders, 1998, Atomic decomposition by basis pursuit: *SIAM Journal on Scientific Computing*, **20**, 33–61, doi: [10.1137/S1064827596304010](https://doi.org/10.1137/S1064827596304010).
- Chen, W., W. Wang, J. Gao, C. Jiang, and J. Lei, 2013, Sparsity optimized separation of ground-roll noise based on morphological diversity of seismic waveform components: *Chinese Journal of Geophysics*, **56**, 2771–2782.
- Chen, X., W. Chen, X. Wang, and W. Wang, 2017, 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).
- Chopra, S., and G. Larsen, 2000, Acquisition footprint — Its detection and removal: *CSEG Recorder*, **25**, 16–20.
- Daubechies, I., M. Defrise, and C. De Mol, 2004, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint: *Communications on Pure and Applied Mathematics*, **57**, 1413–1457, doi: [10.1002/cpa.20042](https://doi.org/10.1002/cpa.20042).
- Davogusitto, O., and K. J. Marfurt, 2011, Removing acquisition footprint from legacy data volumes: 81st Annual International Meeting, SEG, Expanded Abstracts, 1025–1029, doi: [10.1190/1.3627379](https://doi.org/10.1190/1.3627379).
- Drummond, J. M., A. J. L. Budd, and J. W. Ryan, 2000, Adapting to noisy 3D data — Attenuating the acquisition footprint: 70th Annual International Meeting, SEG, Expanded Abstracts, 9–12, doi: [10.1190/1.1816247](https://doi.org/10.1190/1.1816247).
- Elad, M., 2010, Sparse and redundant representations: From theory to applications in signal and image processing: Springer Publishing Company.
- Elad, M., and M. Aharon, 2006, Image denoising via sparse and redundant representations over learned dictionaries: *IEEE Transactions on Image Processing*, **15**, 3736–3745, doi: [10.1109/TIP.2006.881969](https://doi.org/10.1109/TIP.2006.881969).
- Elad, M., J. L. Starck, P. Querre, and D. L. Donoho, 2005, Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA): *Applied & Computational Harmonic Analysis*, **19**, 340–358, doi: [10.1016/j.acha.2005.03.005](https://doi.org/10.1016/j.acha.2005.03.005).
- Falconer, S., and K. J. Marfurt, 2008, Attribute-driven footprint suppression: 78th Annual International Meeting, SEG, Expanded Abstracts, 2667–2671, doi: [10.1190/1.3063897](https://doi.org/10.1190/1.3063897).
- Fehmers, G. C., and C. F. W. Höcker, 2003, Fast structural interpretation with structure-oriented filtering: *Geophysics*, **68**, 1286–1293, doi: [10.1190/1.1598121](https://doi.org/10.1190/1.1598121).
- Gao, D., 2003, Volume texture extraction for 3D seismic visualization and interpretation: *Geophysics*, **68**, 1294–1302, doi: [10.1190/1.1598122](https://doi.org/10.1190/1.1598122).
- Gao, L., X. Wang, W. Chen, and C. Wang, 2016, A local morphological-component analysis-based acquisition footprint suppression method: 86th Annual International Meeting, SEG, Expanded Abstracts, 4710–4714, doi: [10.1190/segam2016-13856781.1](https://doi.org/10.1190/segam2016-13856781.1).
- Gómez, J. L., and D. R. Velis, 2019, Spectral structure-oriented filtering of seismic data with self-adaptive paths: *Geophysics*, **84**, no. 5, V271–V280, doi: [10.1190/geo2018-0788.1](https://doi.org/10.1190/geo2018-0788.1).
- Gómez, J. L., and D. R. Velis, 2020, Footprint removal from seismic data with residual dictionary learning: *Geophysics*, **85**, no. 4, V355–V365, doi: [10.1190/geo2019-0482.1](https://doi.org/10.1190/geo2019-0482.1).
- Marfurt, K. J., R. M. Scheet, J. A. Sharp, and M. G. Harpel, 1998, Suppression of the acquisition footprint for seismic sequence attribute mapping: *Geophysics*, **63**, 1024–1035, doi: [10.1190/1.1444380](https://doi.org/10.1190/1.1444380).
- Olshausen, B. A., and D. J. Field, 1996, Emergence of simple-cell receptive field properties by learning a sparse code for natural images: *Nature*, **381**, 607–609, doi: [10.1038/381607a0](https://doi.org/10.1038/381607a0).
- Pati, Y., R. Rezaifar, Y. Rezaifar, and P. Krishnaprasad, 1993, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition: Presented at 27th Annual Asilomar Conference on Signals, Systems, and Computers, 40–44.
- Rubinstein, R., M. Zibulevsky, and M. Elad, 2010, Double sparsity: Learning sparse dictionaries for sparse signal approximation: *IEEE Transactions on Signal Processing*, **58**, 1553–1564, doi: [10.1109/TSP.2009.2036477](https://doi.org/10.1109/TSP.2009.2036477).
- Shen, L., Z. Xiao, and H. Han, 2010, Image super-resolution based on MCA and wavelet-domain HMT: Proceedings of International Forum on Information Technology and Applications, 264–269.
- Starck, J. L., M. Elad, and D. Donoho, 2004, Redundant multiscale transforms and their application for morphological component separation: *Advances in Imaging & Electron Physics*, **132**, 287–348, doi: [10.1016/S1076-5670\(04\)32006-9](https://doi.org/10.1016/S1076-5670(04)32006-9).
- Starck, J. L., M. Elad, and D. Donoho, 2005, Image decomposition via the combination of sparse representations and a variational approach: *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, **14**, 1570–1582, doi: [10.1109/TIP.2005.852206](https://doi.org/10.1109/TIP.2005.852206).
- Turquais, P., E. G. Asgedom, and W. Söellner, 2017, Coherent noise suppression by learning and analyzing the morphology of the data: *Geophysics*, **82**, no. 6, V397–V411, doi: [10.1190/geo2017-0092.1](https://doi.org/10.1190/geo2017-0092.1).
- Vinther, R., 1997, Seismic texture classification applied to processed 2D and 3D seismic data: 67th Annual International Meeting, SEG, Expanded Abstracts, 721–724, doi: [10.1190/1.1886112](https://doi.org/10.1190/1.1886112).
- Vinther, R., K. Mosegaard, K. Kierkegaard, I. Abatzis, C. Andersen, O. V. Vejbaek, F. If, and P. H. Nielsen, 1995, Seismic texture classification: A computer aided approach to stratigraphic analysis: 65th Annual International Meeting, SEG, Expanded Abstracts, 153–155, doi: [10.1190/1.1887260](https://doi.org/10.1190/1.1887260).
- 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).
- West, B. P., S. R. May, J. E. Eastwood, and C. Rossen, 2002, Interactive seismic facies classification using textural attributes and neural networks: *The Leading Edge*, **21**, 1042–1049, doi: [10.1190/1.1518444](https://doi.org/10.1190/1.1518444).
- Yarham, C., U. Boeniger, and F. Herrmann, 2006, Curvelet-based ground roll removal: 76th Annual International Meeting, SEG, Expanded Abstracts, 2777–2782, doi: [10.1190/1.2370101](https://doi.org/10.1190/1.2370101).
- Yong, X., R. K. Ward, and G. E. Birch, 2009, Artifact removal in EEG using morphological component analysis: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 345–348.
- Yu, Z., R. Abma, J. Etgen, and C. Sullivan, 2017, Attenuation of noise and simultaneous source interference using wavelet denoising: *Geophysics*, **82**, no. 3, V179–V190, doi: [10.1190/geo2016-0240.1](https://doi.org/10.1190/geo2016-0240.1).

Biographies and photographs of the authors are not available.