IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 23, 2020, accepted August 10, 2020, date of publication August 18, 2020, date of current version August 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3017690

Sinusoidal Seismic Noise Suppression Using Randomized Principal Component Analysis

HANG WANG[1], HUA ZHANG[2], AND YANGKANG CHEN[1], (Member, IEEE)
[1]School of Earth Sciences, Zhejiang University, Hangzhou 310027, China
[2]State Key Laboratory of Nuclear Resources and Environment, East China University of Technology, Nanchang 330013, China

Corresponding authors: Hua Zhang (zhhua1979@163.com) and Yangkang Chen (chenyk2016@gmail.com).

This work was supported in part by the National Natural Science Foundation of China under Grant 41664006 and Grant 41874126, and in part by the Natural Science Foundation of Jiangxi Province under Grant 20181BAB203028.

ABSTRACT Sinusoidal interference is a common type of noise caused by power facilities. It usually has the fixed frequency and large amplitude which will have a negative impact on seismic data processing. We propose a new approach for suppressing sinusoidal noise, which is called randomized principal component analysis. It takes advantage of the characteristics of noisy record itself and does not need to calculate the amplitude or phase of the sinusoidal noise. The key algorithm steps of this method can be illustrated as follows: first, interpolation will be applied to get the improved time accuracy. Second, frequency search and narrow-band-pass filter are used to separate the noise-dominated component and determine its precise period. Then, for one of the periods, we use several periods around it to form a ''section''. The randomization is then implemented along the lateral direction to disturb the reflected signal in this ''section''. Besides, principal component analysis is applied to extract sinusoidal noise which still has strong correlation, and to filter the signal which has been converted into random noise. Last, the separated noise is subtracted from the record. Via the synthetic model and field examples, this method is compared to the notch filter to demonstrate its superiority.

INDEX TERMS Principal component analysis, randomization, sinusoidal noise attenuation.

I. INTRODUCTION

Sinusoidal noise and its harmonics have always been a problem in reflected seismic data due to the presence of power systems in the acquisition area. They often have the fundamental frequency of 60 Hz in North America and 50 Hz in some other countries such as China and Japan. Sinusoidal noise with fundamental frequency recorded by geophone cable usually has large amplitude compared to reflected signal, especially in the deep parts of seismic section. Its existence causes much trouble in the subsequent processing. Note that the ''harmonic'' here is not the land vibroseis harmonics, but the sinusoidal noise with the fundamental frequency and its multiples.

According to the investigation in four places in Canada by [1], except the fundamental frequency with strongest noise amplitude, there is still large interference energy existing in the first few odd harmonics. Because of the requirement for the frequency bandwidth of raw data in seismic exploration, simply notching many harmonics by analog filter in the field is not practical. Therefore, finding another way to solve this problem has become necessary.

Seismic denoising methods have been developed by many scholars. For random noise attenuation, prediction-based methods [2]–[5], sparse-transform-based methods [6], [7], rank-reduction-based methods [8]–[10], machine-learning-based methods [11]–[14] and orthogonalization [15] are most common used algorithms. For the removal of harmonic noise [16], [17], there are also many different methods presented besides the analog notch filter. Among them, real-time suppression methods are widely used in seismic and electromagnetic field. But this kind of methods may not be so practical because of the location requirement of acquisition system. Therefore, suppressing power-line interference (sinusoidal noise) by subsequent processing has been an agreement the norm. Digital notch filter is the most commonly used method because of its simple procedures. But, just like the analog one mentioned above, when we notch the fundamental frequency and its harmonics, the cost is that the frequency components of signal near it will be also eliminated simultaneously. As a

The associate editor coordinating the review of this manuscript and approving it for publication was Jiju Poovvancheri.

result, a distortion of signal takes place due to above operations. The spiking deconvolution is also used to suppress the harmonic noise by flattening the spectrum. But the method does not seem to perform well because of the presence of the strong spectral peaks [18]. In order to overcome the disadvantages of notch filter and the spiking deconvolution, other approaches have been developed. Among these methods, estimation-subtraction is the most popular route developed by many scholars. Reference [19] regarded the noise as a stationary noise with a fixed frequency, amplitude and phase. They make an initial guess frequency of the sinusoidal noise, and the precise values of these three parameters are then determined by an iterative process. Afterwards, the estimated sinusoidal noise will be subtracted from the contaminated data. Reference [18] designed a Wiener filter which takes the corrupted trace as desired output and a reference sinusoidal trace as initial input. The frequency of reference trace can be determined by an automatic search algorithm. At last, a filter is solved, which makes the error between its convolutions with the reference trace and the corrupted trace is least. Then, the result of convolution will be subtracted from the corrupted data.

Reference [20] presented a block subtraction method which takes advantage of the fact that sinusoidal in different cycles have similar shape. A portion of the signal trace which contains negligible reflected signal (such as several cycles before the first break) is extracted and subtracted in different shifting position. Reference [21] proposed a method based on damped least square Levinson-Marquardt (L-M) inversion, which can simultaneously estimate the frequency and phase of different noise components. Singular value decomposition (SVD) is introduced into this method to simplify the L-M solution. The shallow seismic data contaminated by harmonics of 60 Hz is used to test the validity of this method. Reference [1] presented a new algorithm to estimate stationary harmonics of one or more fundamental frequencies in a time series by seeking a linear combination of them. Field data examples have proven the effectiveness of this method. Reference [22] used a suitable linear combination of Nyman and Gaiser estimation (NGE) frequency estimates to produce a more accurate estimate of the fundamental frequency, which is faster than least-squares estimation and more accurate than NGE. They applied the method to magnetotelluric data and suppressed the power-line noise effectively.

Different from the above methods, we try to find another route to solve the problem by randomization and principle component analysis (PCA), which avoids the process of estimating amplitude and phase. PCA is a very effective tool to extract signal from noisy data and is widely applied in image denoising and enhancement [23]–[25]. Reference [26] provided a method that attenuates coherent noise and random noise simultaneously by using randomization operator and multichannel singular spectrum analysis (MSSA), which is also a kind of PCA. Randomization operator is used to preserve signal (after NMO) and disturb the coherent noise which can be treated as random noise later. Then, MSSA is applied to suppress the random noise. Inspired by this algorithm, we present a similar approach to extract fundamental sinusoidal noise and its harmonics. In other words, we form a ''seismic section'' with horizontal events by using several randomized cycles and separate out the sinusoidal noise which is treated as signal.

In common cases, the frequency leakage is inevitable and will have big influence on the frequency components near noise frequency, especially when the energy of noise is very large. For conventional notch filter, it just simply suppresses the noise spike in an small fixed frequency range and can not eliminate the leaking energy. If we set a big range of notching frequencies, the signal frequencies near it may be harmed inevitably. However, the randomization and principle component analysis (RPCA) can get a suitable estimation of noise in both time and frequency domain and overcome this defect. Note that the ''RPCA'' here does not stand for the robust PCA but the randomized PCA.

Though the fundamental frequency of power-line interference in North American may have a deviation of up to 0.03 Hz according to [27], most of the current methods are based on the assumption that the harmonic noise is stationary over the whole record [1]. Our method is also derived under this assumption. Fortunately, the validity of this stationary model has been confirmed by many scholars, and based on this assumption, the field example we used also produces a good result.

In this article, we first give the detailed description of this algorithm which includes preprocess, randomization, Principal Component Analysis (PCA) and stack. Then we test the effect of this method on the synthetic traces and field data. The results show that our method can remove the sinusoidal noise effectively and has obvious advantage over conventional notch filter. At last, we give a brief discussion of this method and conclude.

II. METHODOLOGY

A. PREPROCESS

The sample interval in time domain is always limited in field seismic acquisition (such as 2ms and 4ms), which makes it difficult to precisely estimate the period of sinusoidal noise. Therefore, we first consider interpolation of the contaminated traces by a factor n, which is the number of interpolation points at each interval. There are many interpolation methods in practical engineering application, such as linear interpolation, cubic interpolation and spline interpolation. Considering the smoothness of interpolation results, we choose spline interpolation as the first step in this article. Note that the spline interpolation is not the only suitable way to interpolate the seismic traces. Other methods such as sinc interpolation can also do this job well. For a seismic data trace $x(t)$ with power-line interference, after the spline interpolation, we can get the interpolated trace $x_p(t)$ which has N_p sample points.

The basic principal of spline interpolation can be described as follows: for a seismic trace $x(t)$ (N_t points) with power-line

interference, $t = 1, 2, 3, \ldots, N_t$, we try to find a piecewise function:

$$S(t) = \begin{cases} S_1(t), & t \in [1, 2] \\ S_2(t), & t \in [2, 3] \\ \ldots \\ S_3(t), & t \in [N_t - 1, N_t], \end{cases} \tag{1}$$

which meets the requirements that:

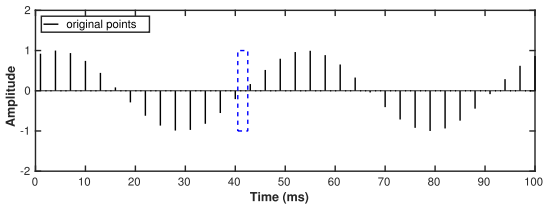
$$\begin{aligned} S(t) &= x(t) \quad t = 1, 2, 3, \ldots, N_t, \\ S_i - 1(t) &= S_i(t) \quad t, i = 2, 3, \ldots, N_t - 1, \\ S_i' - 1(t) &= S_i'(t) \quad t, i = 2, 3, \ldots, N_t - 1, \\ S_i'' - 1(t) &= S_i''(t) \quad t, i = 2, 3, \ldots, N_t - 1. \end{aligned} \tag{2}$$

With the known boundary condition $S'(1)$, $S'(N_t)$ or $S''(1)$, $S''(N_t)$, We can get the analytic expression of the function. Then, according to it, we calculate the function value of interpolation points between two original sample points. The original seismic record $x(t)$ will be extended to $x_p(t)$ with N_p points. N_p has the following relationship with N_t:

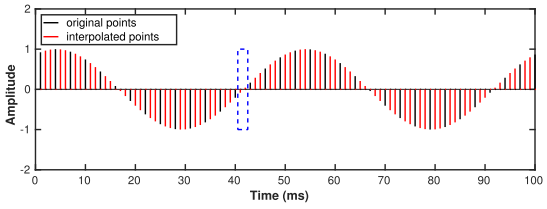
$$N_p = (N_t - 1) \times n + 1, \tag{3}$$

where n is the number of interpolation points at each interval.

As shown in Figs. 1a and 1b, in a sinusoidal waveform, we insert two points (red lines) in each interval. It is very clear that, after the interpolation, the period of sinusoidal noise can be estimated more precisely because there are more sampling points within the same duration (See the blue dash rectangles in these two diagrams).



(a)



(b)

FIGURE 1. The diagram of interpolation. (a) Sinusoidal noise before the interpolation. (b) Sinusoidal noise after the interpolation. The rectangles indicate that there are more sampling points to help determine the noise period in (b).

Assume $x_p(t)$ contains several single-frequency harmonics $(f_1, \ldots, f_{ns})$ which have large values in the amplitude spectrum. Then, in order to generate the noise-dominated series which are more favorable for the extraction of sinusoidal in the subsequent processing, we design some zero-phase filters that just give pass bands of these harmonics. Fig. 4a is the amplitude spectrum of a zero-phase filter for the 50 Hz noise, which is equal to 1 at [40, 60] Hz and 0 at other places (Because there are zero values in the spectrum, we use the original amplitude instead of the log amplitude in the vertical axis). The center frequency 50 Hz can be roughly found by selecting the maximum value of the current amplitude spectrum. Note that the range 40-60 is just an example. In the other noisy data with different noise frequencies, one can adjust the range accordingly to make it fully cover the noise frequencies. Fig. 4b is the corresponding phase spectrum. The values of it are all zeros. The filtering process is to multiply the spectrum of noisy data by the filter spectrum, point by point. Fig. 5a shows the amplitude spectrum of the noisy data before and after the process of narrow-band filter. Fig. 5b shows the corresponding phase spectrum. It is clear that this simple filter only cuts part of the noisy spectrum between 40-60 Hz, which covers the location of the 50 Hz noise.

We mark one of these separated components with certain sinusoidal noise as $x_b(t)$ (Note that there is still some residual signal existing in $x_b(t)$). The reason why we separate them from $x_p(t)$ is that, since we aim to extract the sinusoidal noise instead of signals, the reflected signals will have the negative impact on the extraction of the noise. In the noise-dominated component $x_b(t)$, there is less residual signal compared to the $x_p(t)$, which will make the extraction more precisely.

For the narrow-band sinusoidal noise, we use a special analysis tool to get the precise frequency of it, which is same as the automatic search algorithm mentioned by [18]. The amplitude spectrum of Fourier transform can indicate the rough maximum frequency f_m that represents the sinusoidal noise. Then, within the scope of $[f_m - 1, f_m + 1]$ Hz, we divide the frequency into N_f parts with a refined interval Δf. Theoretically, the interval can be set as small as possible. For every discrete frequency within it, we calculate the scalar product:

$$F(k) = \sum_{i=1}^{N_p} x_b(i) e^{-j2\pi \cdot (f_m - 1 + k\Delta f) \cdot i \cdot dt}, \quad k = 0, 1, 2, \ldots, N_f. \tag{4}$$

Therefore, we can get an accurate noise frequency by searching the maximum value of $F(k)$, which is the basis of determining the period used in next step. Fig. 2 is an example of precise frequency analysis for a 60.03 Hz sinusoidal noise.

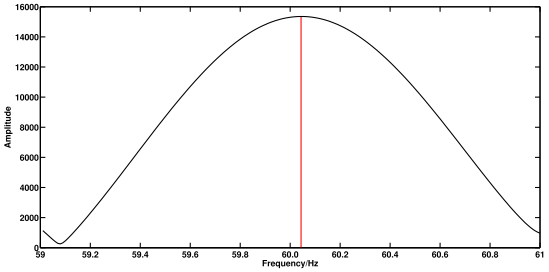


FIGURE 2. The result of precise frequency analysis for seismic trace with 60.03 Hz sinusoidal noise, which is sampled at 1*ms*, 1000 points. The red line indicates the maximum frequency which is also the noise frequency.

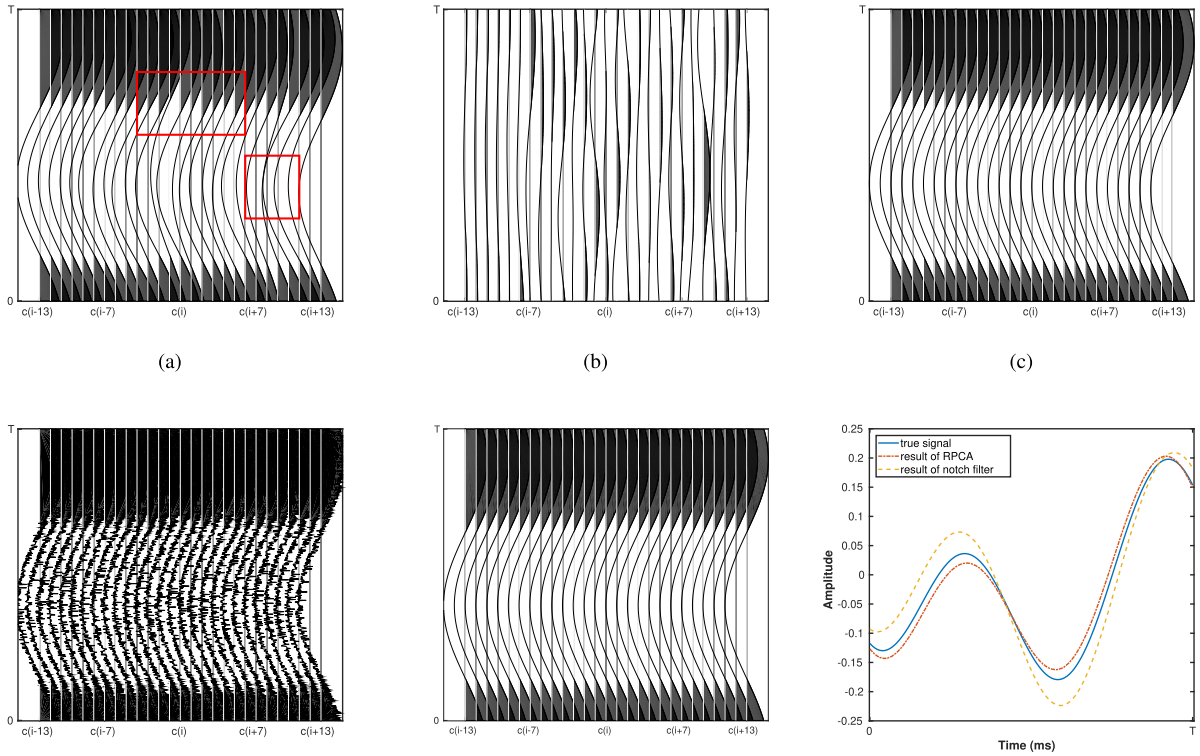


FIGURE 3. The diagram of randomization and PCA: list several cycles within the time window (w=13) near c_i in the horizontal direction. These cycles are picked from the synthetic data in Fig. 8a. (a) Noisy cycles. (b)Underlying signal in these cycles. (c) Cycles of sinusoidal noise. (d) Cycles after randomization alone the horizontal direction (reflected signal has been converted into random noise), (e) The left sinusoidal noise after picking the largest diagonal element of (d). It is similar to but not exactly the same as (c). (f) The comparison among underlying signal (blue line), denoised cycle of RPCA (red line) and denoised cycle of notch filter (yellow line) in the location of c_i. Note that this procedure aims to extract the signal only in c_i instead of all $2w + 1$ cycles. Therefore, the comparison only includes the cycle in the location of c_i.

B. RANDOMIZATION

Assuming the precise frequency derived from above procedures is f_0, the accurate period of the sinusoidal can be deduced by the reciprocal relationship between frequency and period. With the refined time sample interval dt_0 having been decided, we can get the number of points N_0 within one period.

$$N_0 = 1/f_0/dt_0. \tag{5}$$

Then we can break the seismic trace into different cycles depending on the period N_0. For a certain sinusoidal cycle c_i, in the time window of $2w + 1$ cycles near it ($[c_{i-w}, c_{i+w}]$), we place them in the horizontal direction and form a matrix C, which can be illustrated as follows:

$$\begin{bmatrix} c_{i-w} \\ \vdots \\ c_i \\ \vdots \\ c_{i+w} \end{bmatrix} \rightarrow \begin{bmatrix} c_{i-w} & \dots & c_i & \dots & c_{i+w} \end{bmatrix},$$

$$\downarrow$$

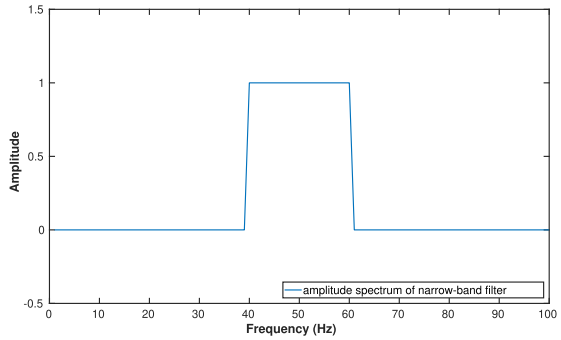
$$\begin{bmatrix} c_{i-w,1} & \dots & c_{i,1} & \dots & c_{i+w,1} \\ c_{i-w,2} & \dots & c_{i,2} & \dots & c_{i+w,2} \\ \vdots & \dots & \vdots & \dots & \vdots \\ c_{i-w,N_0} & \dots & c_{i,N_0} & \dots & c_{i+w,N_0} \end{bmatrix} = C. \tag{6}$$

Within the matrix C, along the row, we exchange the values in the same time point (with the same row subscript). This procedure is called randomization and is to preserve the shape of sinusoidal noise while disturbing reflected signal. After doing this exchange, we can get a "section" C_e made of sinusoidal cycles and random noise (i.e., disturbed reflected signal).

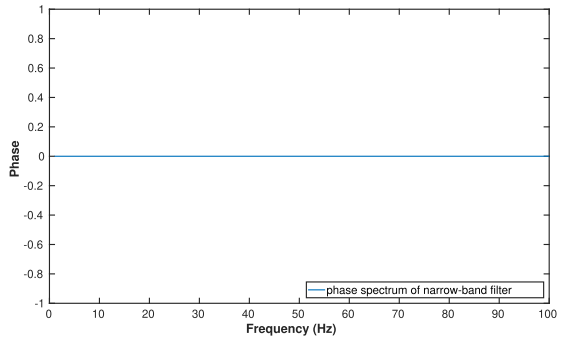
$$C = \begin{bmatrix} c_{i-w,1} & \dots & c_{i,1} & \dots & c_{i+w,1} \\ c_{i-w,2} & \dots & c_{i,2} & \dots & c_{i+w,2} \\ \vdots & \dots & \vdots & \dots & \vdots \\ c_{i-w,N_0} & \dots & c_{i,N_0} & \dots & c_{i+w,N_0} \end{bmatrix},$$

$$\downarrow$$

$$C_e = \begin{bmatrix} c_{i,1} & \dots & c_{i-w,1} & \dots & c_{i+w,1} \\ c_{i+w,2} & \dots & c_{i,2} & \dots & c_{i-w,2} \\ \vdots & \dots & \vdots & \dots & \vdots \\ c_{i-w,N_0} & \dots & c_{i+w,N_0} & \dots & c_{i,N_0} \end{bmatrix}. \tag{7}$$

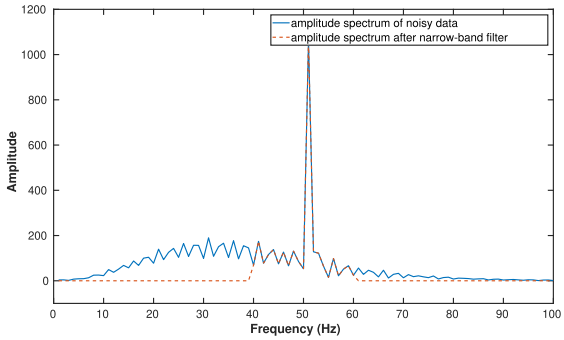


(a)

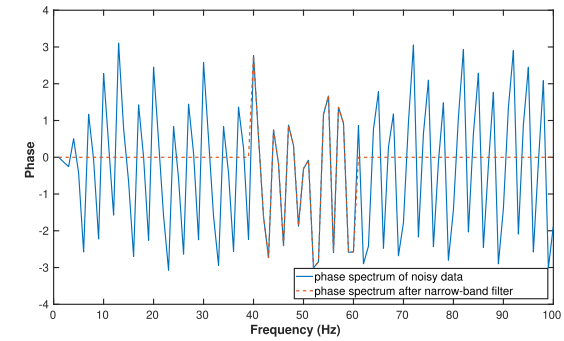


(b)

FIGURE 4. (a) The amplitude spectrum of a zero-phase filter for the 50 Hz noise. (b) The corresponding phase spectrum.



(a)



(b)

FIGURE 5. (a) The amplitude spectrum of the noisy data before and after the process of narrow-band filter. (b) The corresponding phase spectrum.

C. PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is always a good tool for extracting signal from seismic section, especially for the horizontal events. In details, for the matrix C_e with size of $N_0 \times M(M = 2w+1)$, we assume that every column of it has been subtracted from the mean of this column and divided by $\sqrt{N_0 - 1}$, the covariance matrix can be written as:

$$C_e^T C_e = Q \Sigma Q^T, \tag{8}$$

where Q with size of $M \times M$ is an orthogonal matrix, Σ is a diagonal matrix and its diagonal elements θ_i is a descending sequence $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_M$. At the same time, the C_e can also be decomposed as:

$$C_e = UDV^T, \tag{9}$$

where U with size of $N_0 \times N_0$ and V with size of $M \times M$ are orthogonal matrices, D is a diagonal matrix with size of $N_0 \times M$ and its diagonal elements σ_i is also a descending sequence $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{N_0}$. Further, we can have the equation:

$$C_e^T C_e = VD^T U^T UDV^T = VD^T DV^T, \tag{10}$$

From the equations above, we find that:

$$\begin{aligned} V &= Q, \\ \theta_i &= \sigma_i^2, \\ \Sigma &= D^T D. \end{aligned} \tag{11}$$

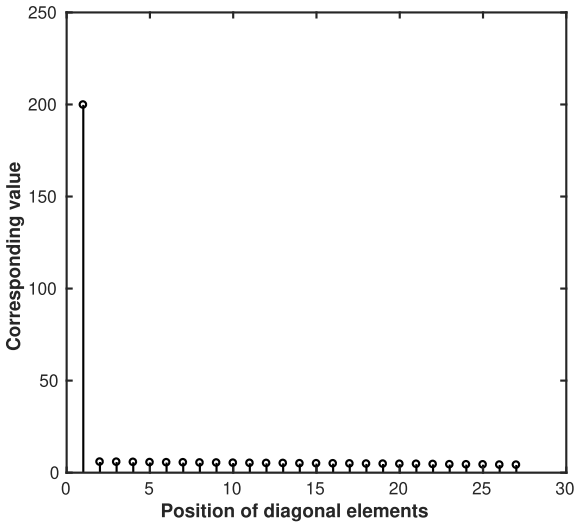


FIGURE 6. The distribution of diagonal elements in Fig. 3d. The first one corresponds to the sinusoidal noise.

The large values of σ_i in the upper left of D indicate the most correlative part of the matrix which can also be thought as the sinusoidal cycles.

When we form a matrix of cycles after randomization, the θ_i and Q can be obtained by computing eigenvalues of $C_e^T C_e$. Then, the σ_i and V will be calculated by equation 11 (U can be derived from $C_e C_e^T$). Sinusoidal noise will be represented by the first large value σ_1. So we use the matrix D'

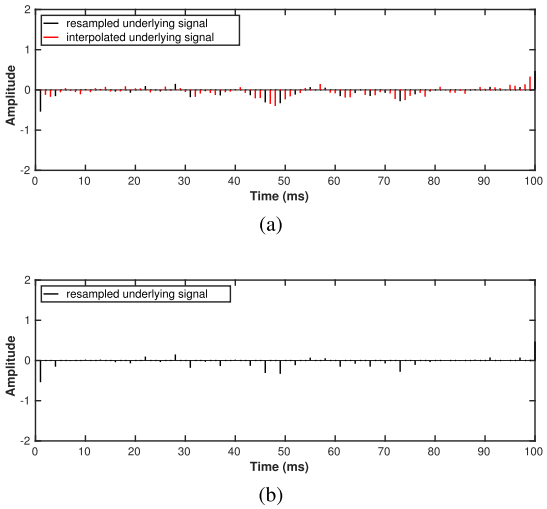


FIGURE 7. The diagram of resampling. (a) Estimated underlying signal before the resampling. (b) Estimated underlying signal after the resampling.

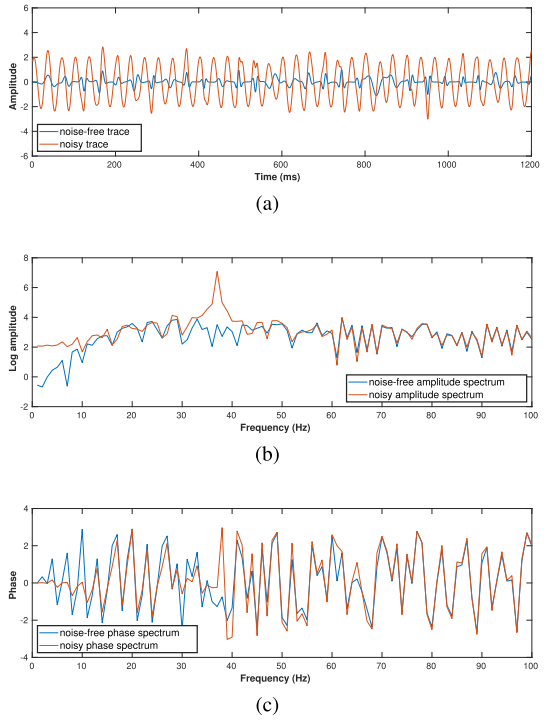


FIGURE 8. Example of synthetic trace. (a) Noisy data and original noise-free data. (b) Log amplitude spectrum of the noisy data and the noise-free data. (c) Phase spectrum of the noisy data and the noise-free data.

(only contains the first diagonal element of D and the other elements are zero) to get the estimation of sinusoidal noise:

$$C' = UD'V^T,$$
$$D' = \begin{bmatrix} \sigma_1 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix}. \quad (12)$$

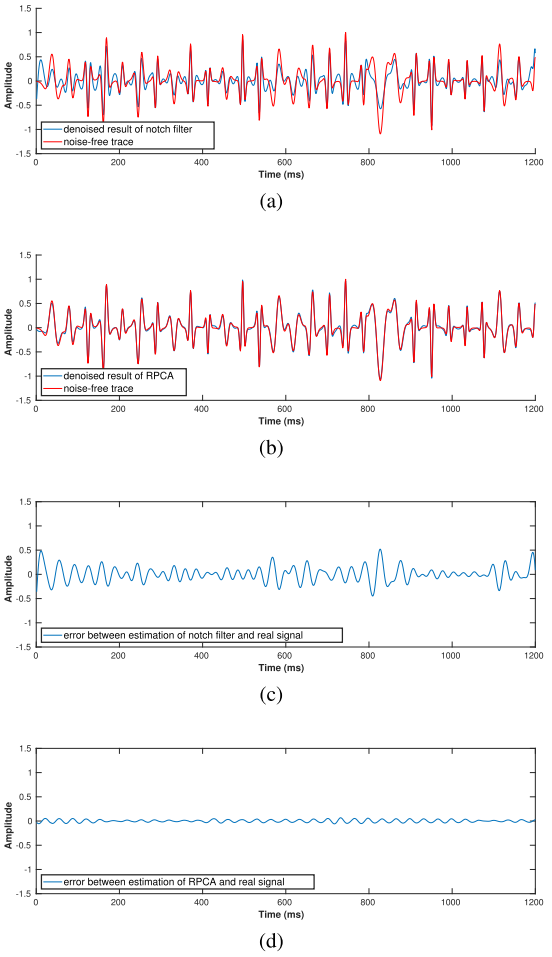


FIGURE 9. Example of synthetic trace. (a) Denoised data by notch filter and the original data. (b) Denoised data by RPCA and the original data. (c) Error between estimation of notch filter and the original signal. (d) Error between estimation of RPCA and the original signal.

In this case, we use PCA to keep the sinusoidal noise which has the stable shape and filter the reflected signal which is treated as the random noise. Then, we can get the matrix C' which stands for the sinusoidal noise.

$$C_e = \begin{bmatrix} c_{i,1} & \dots & c_{i-w,1} & \dots & c_{i+w,1} \\ c_{i+w,2} & \dots & c_{i,2} & \dots & c_{i-w,2} \\ \vdots & \dots & \vdots & \dots & \vdots \\ c_{i-w,N_0} & \dots & c_{i+w,N_0} & \dots & c_{i,N_0} \end{bmatrix},$$
$$\downarrow$$
$$C' = \begin{bmatrix} c'_{i-w,1} & \dots & c'_{i,1} & \dots & c'_{i+w,1} \\ c'_{i-w,2} & \dots & c'_{i,2} & \dots & c'_{i+w,2} \\ \vdots & \dots & \vdots & \dots & \vdots \\ c'_{i-w,N_0} & \dots & c'_{i,N_0} & \dots & c'_{i+w,N_0} \end{bmatrix}. \quad (13)$$

The procedure can be illustrated by the Figs. 3a-3f: we can see that the waveforms of different cycles in Fig. 3a are not so consistent because of the existence of the reflected signal (see the red rectangles). The underlying signal and pure sinusoidal noise of Fig. 3a are shown in the Figs. 3b and 3c. After

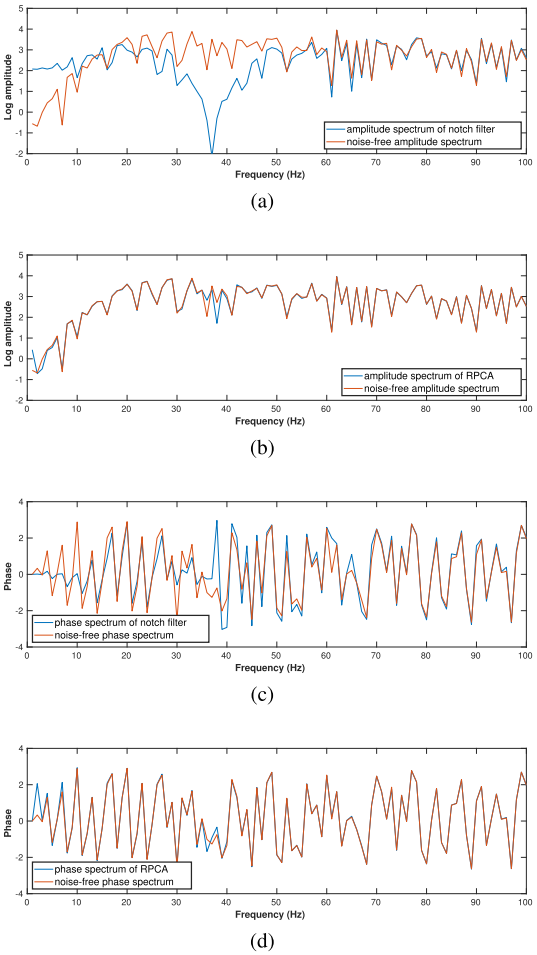


FIGURE 10. Example of synthetic trace. (a) Log amplitude spectrum of notch filter result and the original data. (b) Log amplitude spectrum of RPCA result and the original data. (c) Phase spectrum of notch filter result and the original data. (d) Phase spectrum of RPCA result and the original data.

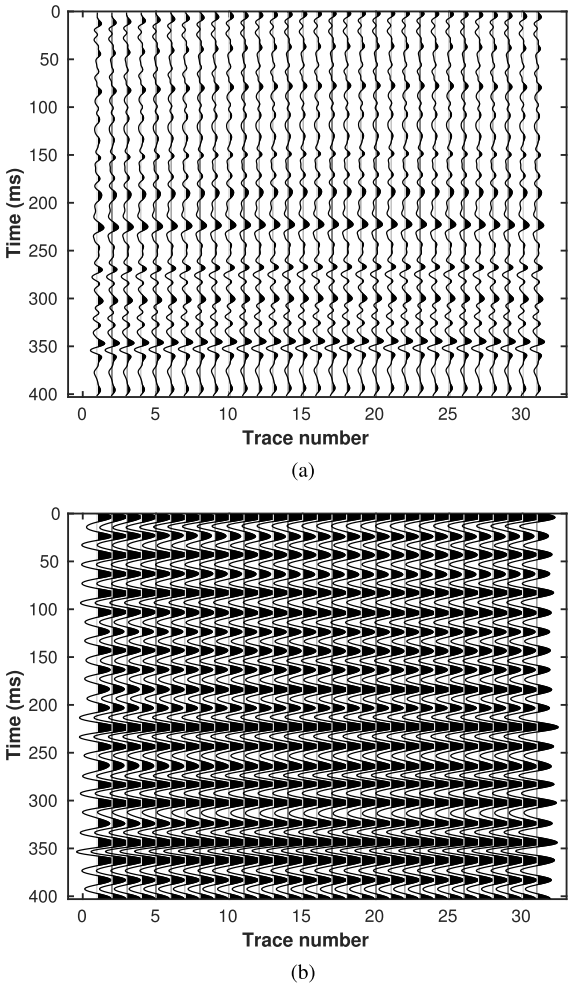


FIGURE 11. Example of synthetic section. (a) Original noise-free data. (b) Noisy data.

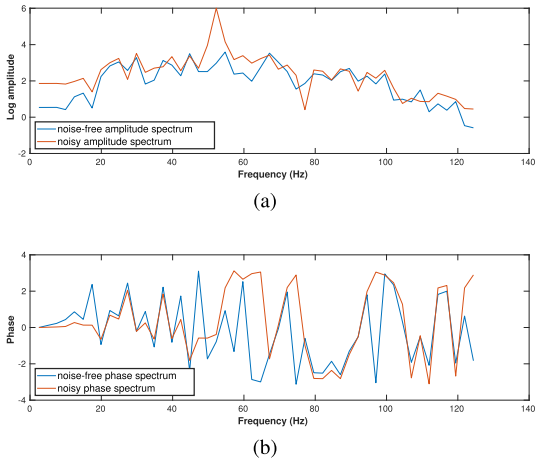


FIGURE 12. Example of synthetic section. (a) Log amplitude spectrum of the noisy data and the noise-free data in 10th trace. (b) Phase spectrum of the noisy data and the noise-free data in 10th trace.

horizontal exchanging, there are many little spikes in the white area of Fig. 3d which is the disturbed reflected signal. If taking a look at the σ_i of Fig. 3d, we find that the first value is far larger than the others (see Fig. 6). Therefore, we can just pick the first one σ_1 and get the estimation of the noise (Fig. 3e).

Next, the noise cycles in the Fig. 3e separated by PCA are averaged to form a new signal-free cycle as the final noise estimate which will be put in the location of the cycle c_i mentioned above. By repeating these procedures, we can calculate signal-free cycles $\tilde{c}_i$ in every place of the seismic trace. At the last step, we subtract the formed noise trace $n(t)$ from original contaminated trace and resample the result to make sure that the number of sample points is the same as the original data. The final denoised result is $s(t)$.

Fig. 3f is the comparison of the true signal (blue line), denoised cycle of RPCA (red line) and denoised cycle of notch filter (yellow line) in the location of c_i. It is worth mentioning again that although there are $2w+1$ cycles in Figs. 3a-3e, they are used to estimate the sinusoidal noise and signal only in c_i. In Fig. 3f, we find that, compared to the result of notch filter, the result of RPCA is closer to the shape of the true underlying signal, which demonstrates the superiority of the RPCA.

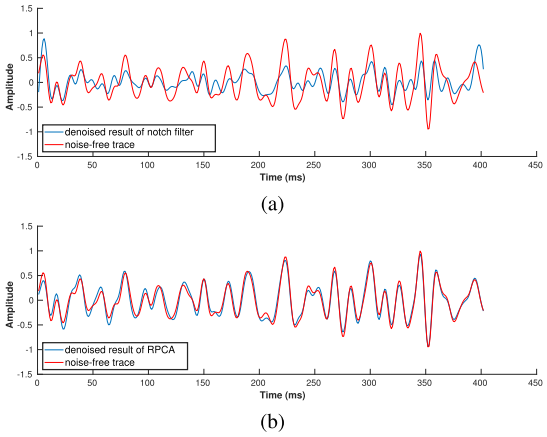


FIGURE 13. Example of synthetic section. (a) Denoised data by notch filter and the original data in 10th trace. (b) Denoised data by RPCA and the original data in 10th trace.

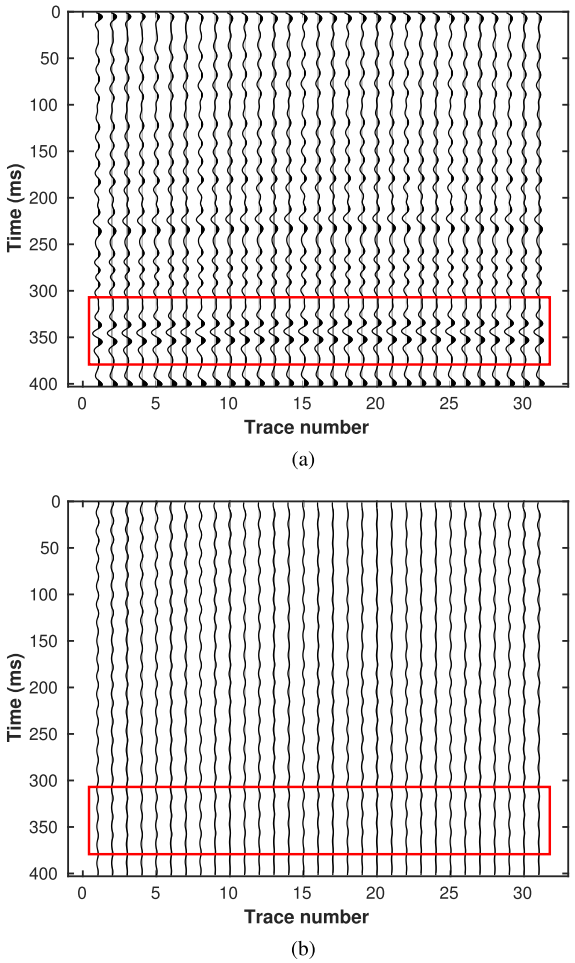


FIGURE 14. Example of synthetic section. (a) Error between estimation of notch filter and the original signal. (b) Error between estimation of RPCA and the original signal.

As for the resampling process, we use Figs. 7a and 7b to illustrate it. When we get the estimation of underlying signal which contains the interpolated parts (red lines) in Fig. 7a,

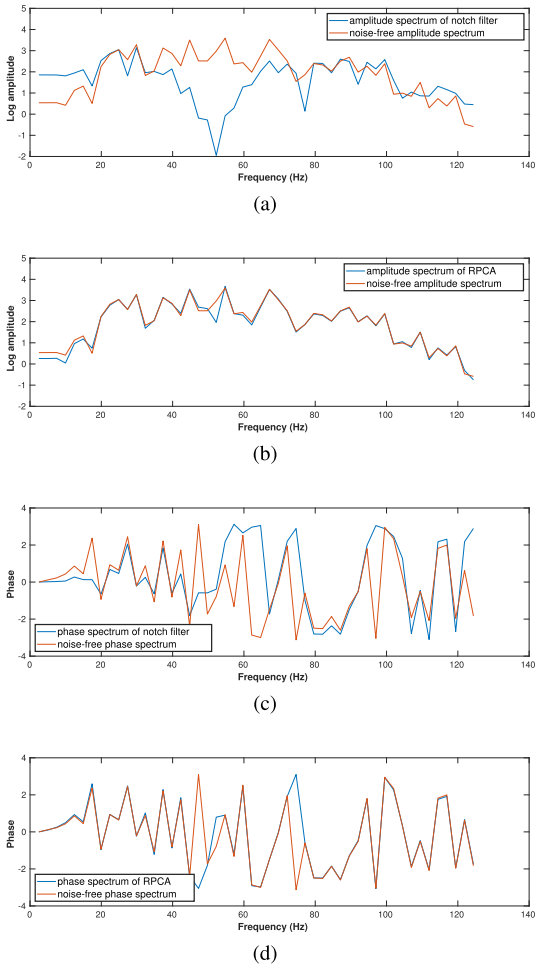


FIGURE 15. Example of synthetic section. (a) Log amplitude spectrum of notch filter result and the original data in 10th trace. (b) Log amplitude spectrum of RPCA result and the original data in 10th trace. (c) Phase spectrum of notch filter result and the original data in 10th trace. (d) Phase spectrum of RPCA result and the original data in 10th trace.

resampling is to just discard the interpolated parts and obtain the resampled underlying signal. The Algorithm 1 shows the procedures of the proposed method.

III. EXAMPLES

A. SYNTHETIC SEISMIC DATA

Figs. 8a-10d show the difference between filtered result from notch filter and that from RPCA for the same original data (the blue line in Fig. 8a). It is added a 36.12 Hz sinusoidal noise (See the red line in Fig. 8a). The log amplitude spectrum and phase spectrum of them are shown in Figs. 8b and 8c. We can clearly find that the notch filter result (the blue line in Fig. 9a) has the larger distortion compared to the result obtained by the RPCA (the blue line in Fig. 9b), which almost coincides with the original data. The Figs. 9c and 9d are the errors between their results and the real noise-free data. The error of RPCA is obviously smaller than that of notch filter. The signal to noise (S/N) ratios for the input noisy data and two denoised results are listed in Table 1. It is a proof that the

Algorithm 1 Proposed Sinusoidal Noise Attenuation Framework

Require: An input sequence $x(t)$ denoting the noisy trace.

for $f_m = f_1, \ldots, f_{ns}$ **do**
  spline interpolation (SP): $x_p(t) = SP(x(t))$
  narrow-band filter (NBF): $x_b(t) = NBF(x_p(t))$
  frequency analysis (FA): $f_0 = FA(x_b(t))$
  get precise period: $N_0 = 1/f_0/dt_0$
  form cycles: separate trace into cycles c_i
  for $i = 1, \ldots, n$ **do**
    RPCA: $\tilde{c}_i = RPCA(c_i)$
  end for
  estimate noise: $n(t) = [\tilde{c}_1^T, \ldots, \tilde{c}_n^T]$
  subtract and resample (RS): $s(t) = RS(x_p(t) - n(t))$
  set new noisy trace: $x(t) = s(t)$
end for

TABLE 1. Comparison of S/N ratios for the original data and two denoised results obtained by notch filter and the RPCA in Figs. 9a and 9b.

Data type	Original	Notch filtering	RPCA
S/N (dB)	-14.1049	3.6646	20.8079

RPCA can generate a more clear result than the conventional method. Figs. 10a and 10b show the comparison of the log amplitude spectrum of original noise-free data, the results of notch filter and RPCA. We can see that the notch filtering provides a log spectrum which has bigger differences with the original one. While, the spectrum obtained by the RPCA are very similar to the original one with only little difference. This phenomenon can also be observed in the phase spectrum in Figs. 10c and 10d. Because the notch filter is a zero-phase filter, the phase spectrum of it is nearly unchanged compared to the phase spectrum of the noisy data.

Fig. 11a is a synthetic model which is masked by around 50 Hz sinusoidal noise (Fig. 11b). Figs. 12a and 12b are the log amplitude spectrum and phase spectrum of them in the 10th trace. The filtered result of the 10th trace shown in Fig. 13b demonstrates that the RPCA can suppress the sinusoidal noise effectively and preserve the signals well even if the frequency is not precise 50 Hz. Respectively, Fig. 13a is the denoised result of the 10th trace by notch filter. It is obviously different from the noise-free data. Their error sections are shown in Figs. 14a and 14b. It seems that notch filter will do more harm to the original signal. There are some leaked events in the red rectangle in Fig. 14a. The S/N ratio for the noisy data and two denoised results are listed in Table 2. Through the S/N ratio table, the validity of the RPCA can be confirmed again. Figs. 15a and 15b are the log amplitude spectrum of the results of these two methods and the noise-free data. The RPCA spectrum is still similar to the true signal while the notch filter is not. In the phase spectrum of them (Figs. 15c and 15d), the error of RPCA is also far smaller than the notch filter, which shows that the RPCA has a good performance on keeping the fidelity of signal.

TABLE 2. Comparison of S/N ratios for the original section and two denoised results from notch filter and RPCA.

Data type	Original	Notch filtering	RPCA
S/N (dB)	-12.9889	1.6098	13.1612

In order to test the effectiveness of RPCA for different noise strength, we set the initial S/N ratios as a descending sequence (which corresponds to the noise amplitude being 0.2,0.5,1,10,50 and 100 times of the signal maximum amplitude) and calculate the denoised results respectively. Table 3 and Fig. 16 are the comparison of output S/N ratios obtained

TABLE 3. The S/N ratios of denoised results by these two methods under different original S/N ratios.

Original (dB)	2.76	-5.18	-11.21	-31.21	-45.19	-51.21
RPCA (dB)	39.04	38.59	37.74	35.55	21.06	14.67
Nortch filtering (dB)	18.54	12.54	7.15	-13.03	-26.68	-33.03

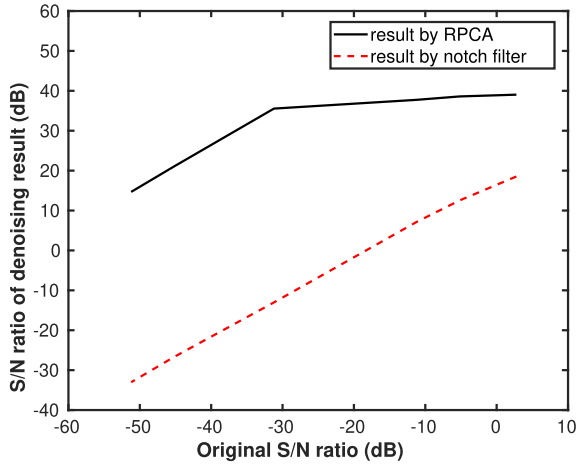


FIGURE 16. The S/N ratio comparison of the denoised results by the two methods under different noise strengths.

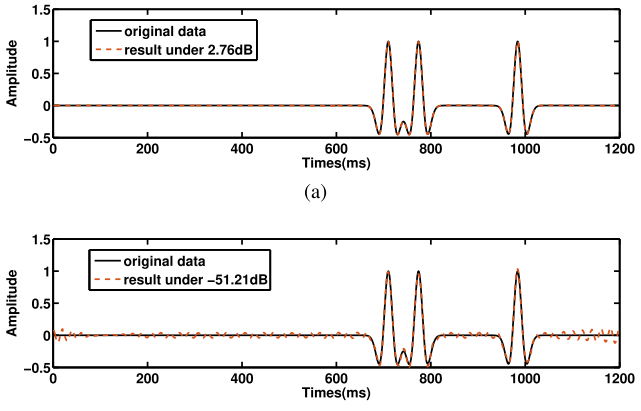


FIGURE 17. The denoised waveforms of RPCA corresponding to the 2.79 dB and -51.21 dB input S/N ratios.

by the two different methods. It is obvious that a low initial S/N ratio will have a negative effect on the performance of both two methods, and the RPCA is more stable than the notch filter under all circumstances. Figs. 17a and 17b are the denoised waveforms obtained by the RPCA corresponding to the 0.2 and 100 times. Note that the low input S/N ratio will cause distortion in Fig. 17b.

B. SEISMIC FIELD DATA

Fig. 18a shows a pre-stack field data example. The 15-21 traces are masked by the 50 Hz sinusoidal noise, and the reflected signals are much weaker than the noise, which makes them not so clear. The noisy log amplitude spectrum of 17th trace is shown in Fig. 18b. We can find an apparent noisy spike near the 50 Hz. After applying the RPCA, we can get a noise-free data (Fig. 19a) with the signal appearing and sinusoidal noise suppressed. As a comparison, we also calculate the notch filtering result (Fig. 19b) and put their log amplitude spectrum (17th trace) together (Fig. 21). In real

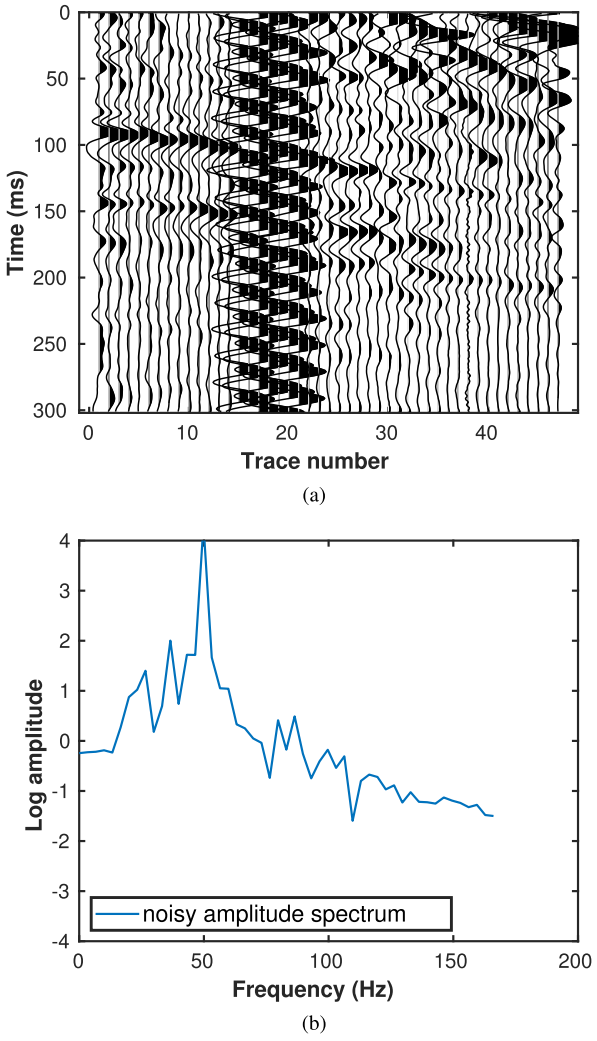


FIGURE 18. Field data example with 50 Hz noise. (a) Original noisy data. (b) Log amplitude spectrum of 17th trace.

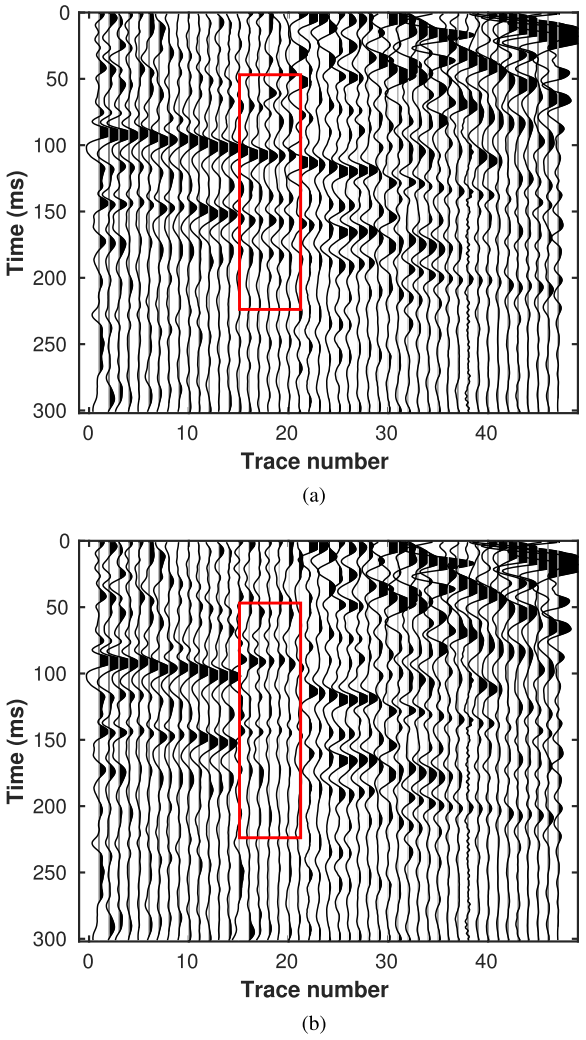


FIGURE 19. Field data example with 50 Hz noise. (a) The denoised result of the RPCA. (b) The denoised result of notch filter.

examples, because we do not have a true noise-free signal as a reference, the phase is not so meaningful to illustrate the superiority of RPCA, so we just show the amplitude spectrum. From the above results, we can see that, though notch filter and the RPCA both can suppress the sinusoidal noise, the notch filter does more harm to the original signal (the frequencies near 50 Hz in Fig. 21). The amplitude of notch filter result in the red box in Fig. 19b is weaker than that of RPCA, which indicates the loss of information. The positions of events in the red box in Fig. 19b are not consistent with those in the adjacent traces. However, the events in Fig. 19a are perfectly recovered and have good consistency. Besides, the removed noise in the 17th trace is displayed in the Figs. 20a and 20b. The noise removed by the RPCA is very close to the standard sinusoidal waveform while the one removed by the notch filter is not.

Fig. 22a is another pre-stack field data example. The 29-31 traces of this section are masked by the 60 Hz sinusoidal noise. The log amplitude spectrum of 30th trace is

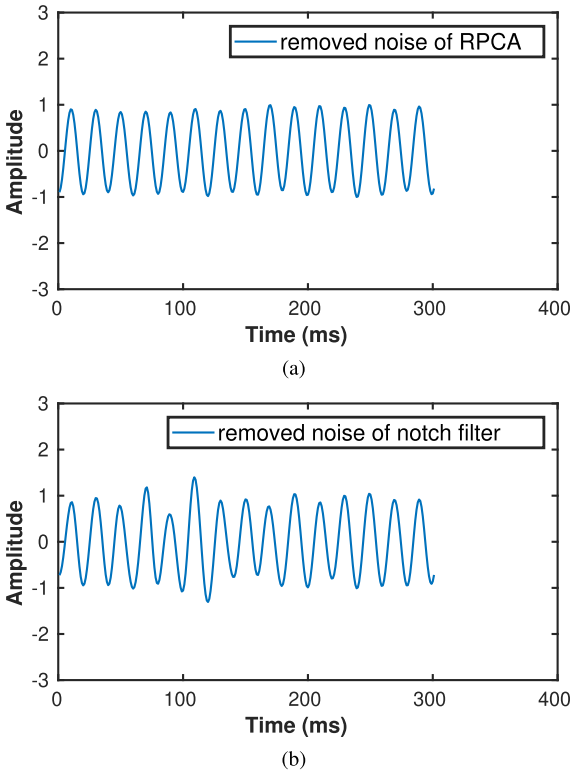


FIGURE 20. Field data example with 50 Hz noise. (a) The removed noise of the RPCA in 17th trace. (b) The removed noise of notch filter in 17th trace.

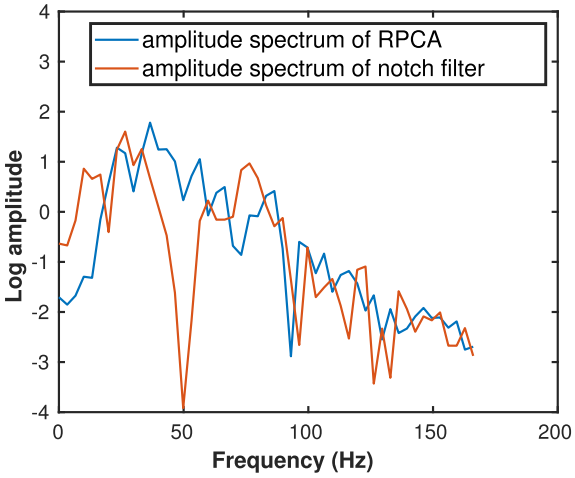


FIGURE 21. Field data example with 50 Hz noise. Denoised log amplitude spectrum of 17th trace from notch filter and RPCA.

shown in Fig. 22b. We use the proposed method mentioned above to filter them and recover the noise-free traces (Fig. 23a). We can see that the most energy of noise has been suppressed effectively, and the reflected signals become dominant again. The events in the red boxes are clear and consistent. However, the notch filter dose not perform well in the red boxes in Fig. 23b. The energy of events is weak and not consistent. Their difference is also shown in the log spectrum of the 30th trace (Fig. 25). The notch filter

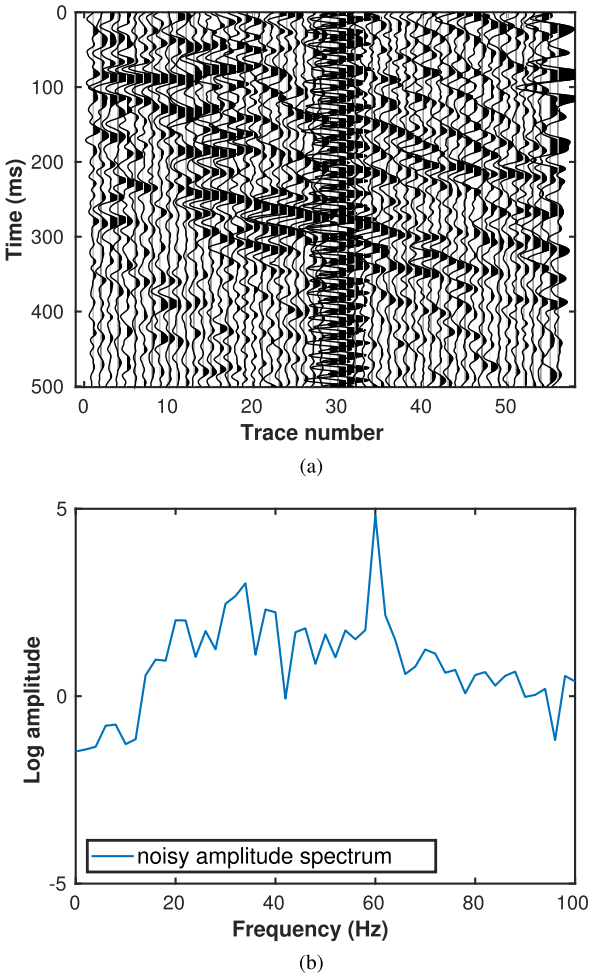


FIGURE 22. Field data example with 60 Hz noise. (a) Original noisy data. (b) Log amplitude spectrum of 30th trace.

obviously causes more damage to signal frequencies than the RPCA, which illustrates that the RPCA outperforms the notch filter. Additionally, Figs. 24a and 24b are the removed noise in the 30th trace by these two methods. It is obvious that the removed noise by notch filter differs greatly from the standard sinusoidal noise, which shows that it cannot suppress this kind of noise as accurately as the RPCA.

IV. DISCUSSION

Sinusoidal noise has been always a common problem in seismic data acquisition. Its large amplitude and the spectral leakage it causes will mask the true features of signal. Suppressing this kind of noise is very necessary. Of course, if there is only one trace contaminated by the sinusoidal noise, the simple stacking or some median filters will suppress the interference. However, the sinusoidal noise sometimes covers a large area due to the powerline system, which invalidates the stacking method and the median filters because they are only good at the single noise trace instead of continuous noisy traces. The proposed method here can be applied to such noisy traces one trace at a time and alleviates their impact on final results.

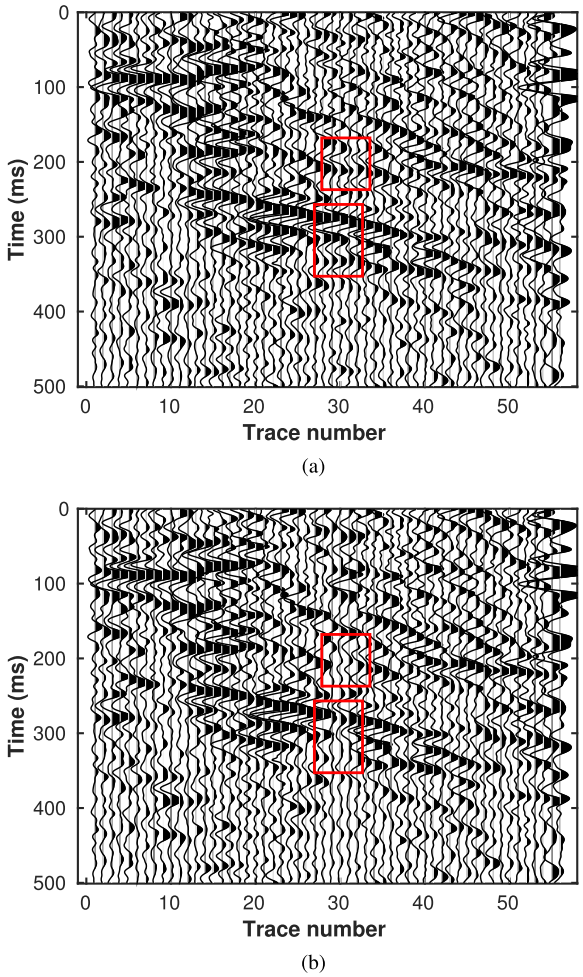


FIGURE 23. Field data example with 60 Hz noise. (a) The denoised result of the RPCA. (b) The denoised result of notch filter.

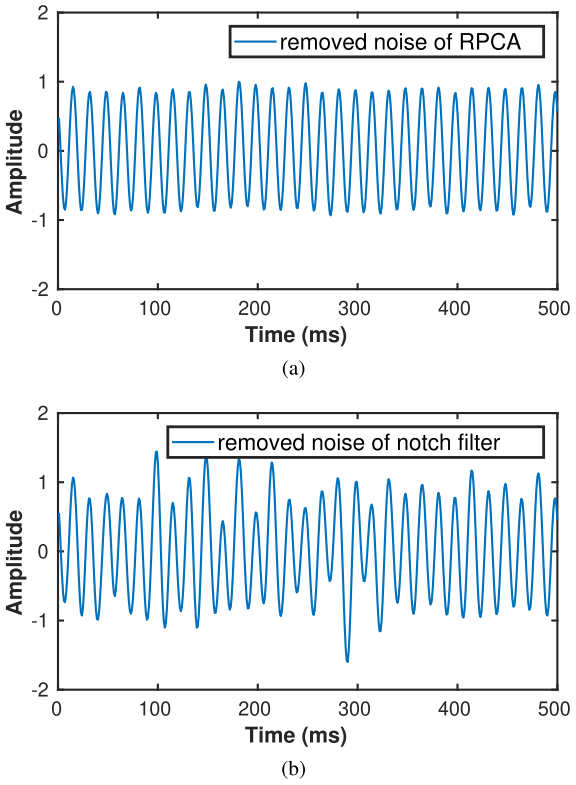


FIGURE 24. Field data example with 60 Hz noise. (a) The removed noise of the RPCA in 30th trace. (b) The removed noise of notch filter in 30th trace.

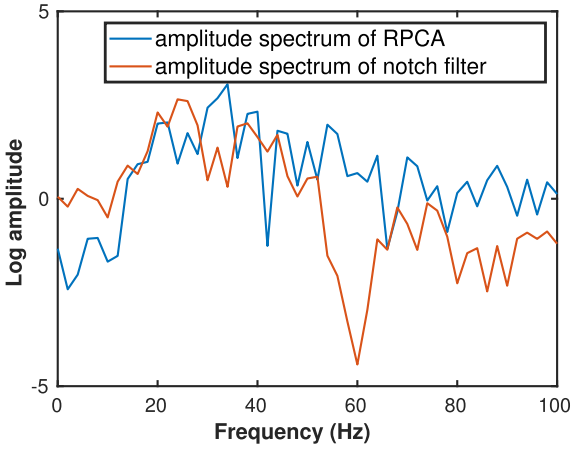


FIGURE 25. Field data example with 60 Hz noise. Denoised log amplitude spectrum of 30th trace from notch filter and RPCA.

The poor performance of a notch filter is mainly caused by the frequency leakage. A tapered notch filter can mitigate the effect but have no ability to eliminate it completely. Additionally, simply notching the noisy amplitude spectrum will undoubtedly harm the reflected signal. Compared to the tapered notch filter, the RPCA can estimate the sinusoidal noise precisely in both time and frequency domains, and can handle the frequency leakage well. As a result, this method can preserve the signal and get a better performance than the notch filter.

In the RPCA, the number of interpolation factor n is a key parameter which can affect the precision of noise estimation. In detail, the smaller interpolation interval is, the more precise denoised result is, but the computational cost is higher. It is also the reason why the results degrade with the stronger noise in Fig. 16. Although the interpolation can improve the precision of noise period estimation, there is still little error between the estimated period and the real period. When noise become stronger, the error will be enlarged more seriously, and the result will become worse (Fig. 17b).

Because of the computational complexity of the RPCA, its computing cost is larger than that of notch filter and will increase significantly when interpolation factor n is large. Table 4 is the comparison of calculation time of these two methods (interpolation factor n is 21,$w = 13$). We also give the calculation costs of different interpolation factor n in Table 5 ($w = 13$). When the n increases, the execution time grows obviously. But note that, in the common seismic data, by our many experiments, we find that a proper n that is smaller than 80 can mostly meet the precision requirement of the seismic industry. For example, the parameters n are

TABLE 4. Comparison of the computing cost (interpolation factor n is 21, $w = 13$). The RPCA is much slower than the notch filter.

Method	Notch filtering	RPCA
Computing time (s)	0.74	1.41

TABLE 5. The computing costs of different interpolation factors n (when $w = 13$).

n	10	20	40	80
Computing time (s)	0.72	1.41	3.12	6.47

TABLE 6. The computing costs of different parameters w (when $n = 21$).

w	10	15	20
Computing time (s)	1.26	1.45	1.93

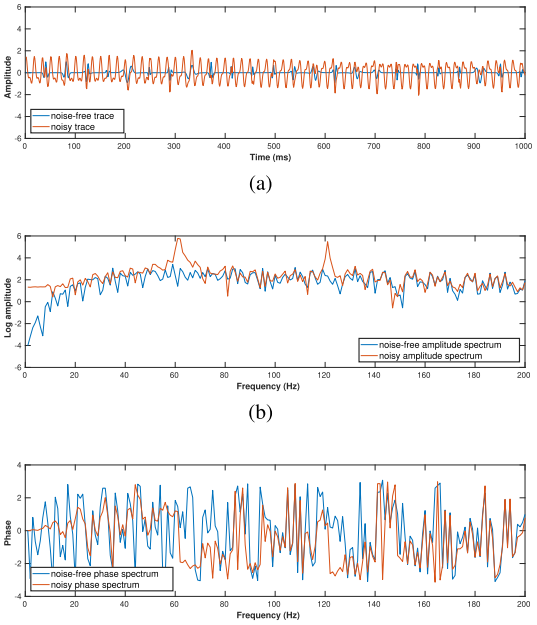


FIGURE 26. Example of synthetic trace with 60 and 120 Hz noise. (a) Noisy data and original noise-free data. (b) Log amplitude spectrum of the noisy data and the noise-free data. (c) Phase spectrum of the noisy data and the noise-free data.

set to 21 and 61 in this article. Additionally, the temporal window parameter w is the other factor that affects the size and computing speed of SVD input matrix. However, due to the strong lateral coherency of the sinusoidal noise itself, the w does not need to be very big and can be set to the range 10-20 by our experience. In this article, all examples share the same w ($w = 13$). Table 6 is the computing time of different w ($n = 21$). It shows that the cost does not change much in this range. Thus, the parameter n is the main factor that has the impact on the computing costs. We will focus on accelerating this algorithm in the future.

When there are many harmonics in the noisy data, and their frequencies are exact multiples of the fundamental frequency, we can suppress them simultaneously (by skipping the procedure of narrow-band-pass filter) because they share a common period. But when the assumption about frequency multiplicity is not valid, it is necessary to suppress them separately. Fig. 26a is an example that contains not only 60Hz but also 120Hz sinusoidal noise. Figs. 26b and 26c are the corresponding log amplitude and phase spectrum. The two noise spikes are obvious in the log amplitude spectrum, and the phase spectrum has the big distortion due to the existence of the sinusoidal noise. Figs. 27a and 27b are the denoised results of the notch filter and RPCA. Figs. 27c and 27d are their corresponding error sequences. These results show that the proposed method can separate the sinusoidal noise more precisely than the notch filter since the denoised results of it contain less distortion and its error energy is smaller than the later one. Figs. 28a-28d are the corresponding log amplitude and phase spectrum of the denoised results from the notch filter and the proposed method. We can find that, in the Fourier domain, the denoised results of the proposed method are much closer to the true signals than the notch filter. The later one is greatly different from the true signal

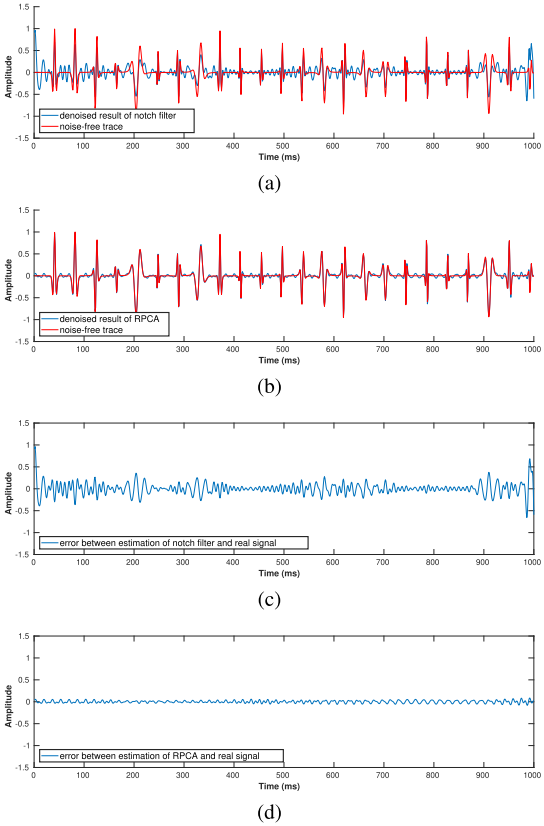


FIGURE 27. Example of synthetic trace with 60 and 120 Hz noise. (a) Denoised data by notch filter and the original data. (b) Denoised data by RPCA and the original data. (c) Error between estimation of notch filter and the original signal. (d) Error between estimation of RPCA and the original signal.

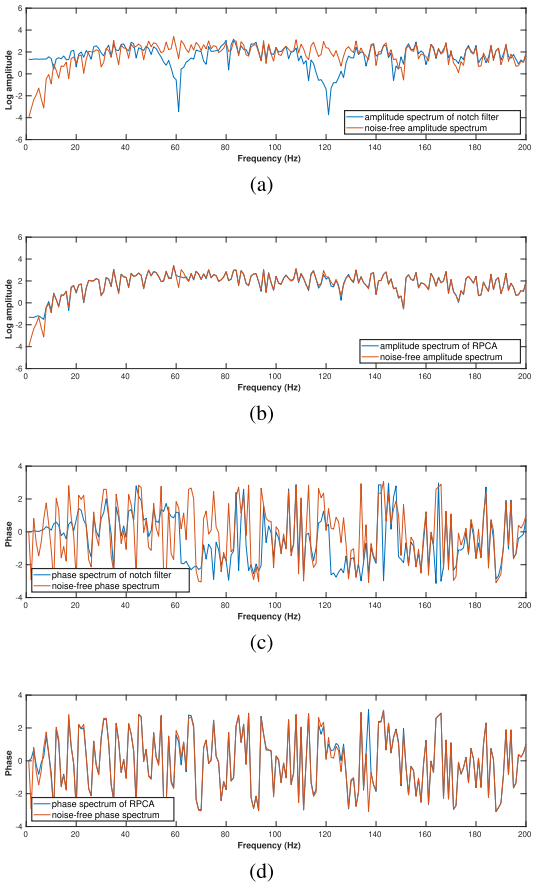


FIGURE 28. Example of synthetic trace with 60 and 120 Hz noise. (a) Log amplitude spectrum of notch filter result and the original data. (b) Log amplitude spectrum of RPCA result and the original data. (c) Phase spectrum of notch filter result and the original data. (d) Phase spectrum of RPCA result and the original data.

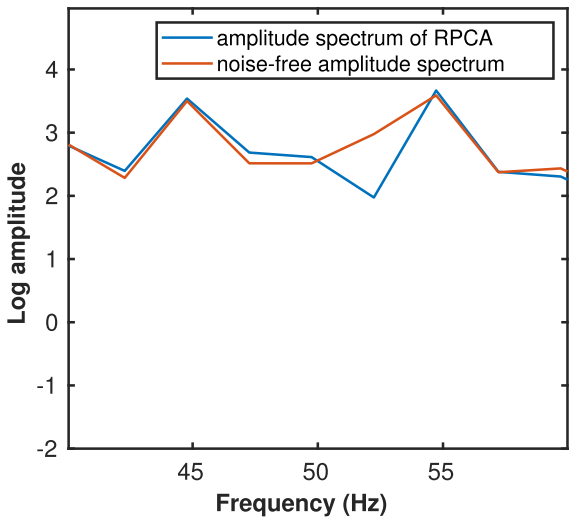


FIGURE 29. The details of Fig. 15b.

at 60 and 120 Hz in the log amplitude spectrum, and its phase spectrum is still seriously influenced by the noise. All the results shown above demonstrate that the proposed method can also handle the multiple harmonics well.

Note that the input sinusoidal noisy traces are not manually picked. We have an another work to automatically pick these noisy traces, and these noisy traces will be the input of the RPCA.

It is also worth mentioning that, although the RPCA has advantage over the notch filter, there are still some differences between the signal spectrum and spectrum of the denoised result (see Fig. 29, the detail of Fig. 15b). In the future research, we will concentrate on fixing this problem.

V. CONCLUSION

We propose a new strategy for sinusoidal noise attenuation which can automatically estimate the noise waveform and avoid complicate estimation of noise amplitude and phase by using the procedures of randomization and PCA. In detail, we first refine the time sample interval by using spline interpolation. A narrow-band-pass filter will be used to separate the rough noise out in the frequency domain. Analyzing its precise frequency is also needed by the subsequent procedure for constructing the suitable cycles. After the preprocessing, we use randomization and PCA to obtain the estimation of noise trace and subtract it from the original trace. In the synthetic and real examples, we find that this method can get better and less distorted results than the notch filter since it will estimate the noise properly and solve the frequency leakage problem. When the S/N ratio is higher than -30 dB, this method keeps stable and can generate a clear trace. Application on synthetic model and real pre-stack data indicates that this method can output a seismic section with a high S/N ratio and causes no damages to the reflected signals. In the future research, we will focus on speeding up this method to make it more affordable for the sinusoidal noise suppression.

REFERENCES

[1] K. E. Butler and R. D. Russell, "Cancellation of multiple harmonic noise series in geophysical records," *Geophysics*, vol. 68, no. 3, pp. 1083–1090, May 2003.

[2] L. L. Canales, "Random noise reduction," in *Proc. SEG Tech. Program Expanded Abstr.*, Jan. 1984, pp. 525–527.

[3] M. Galbraith, "Random noise attenuation by f-x prediction," in *Proc. 61th Annu. Int. Meeting, SEG, Expanded Abstr.*, 1991, pp. 1428–1431.

[4] G. Liu, X. Chen, J. Du, and K. Wu, "Random noise attenuation using f-x regularized nonstationary autoregression," *Geophysics*, vol. 77, no. 2, pp. 61–69, 2012.

[5] B. Bahia and M. D. Sacchi, "Widely linear denoising of multicomponent seismic data," *Geophys. Prospecting*, vol. 68, no. 2, pp. 431–445, Feb. 2020.

[6] A. Latif and W. A. Mousa, "An efficient undersampled high-resolution radon transform for exploration seismic data processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 1010–1024, Feb. 2017.

[7] D. Gemechu, J. Ma, and X. Yong, "A compound method for random noise attenuation," *Geophys. Prospecting*, vol. 66, no. 8, pp. 1548–1567, Oct. 2018.

[8] W. Huang, R. Wang, Y. Chen, H. Li, and S. Gan, "Damped multichannel singular spectrum analysis for 3D random noise attenuation," *Geophysics*, vol. 81, no. 4, pp. V261–V270, 2016.

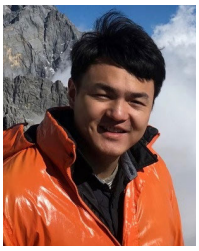
[9] M. A. Nazari Siahsar, S. Gholtashi, E. Olyaei Torshizi, W. Chen, and Y. Chen, "Simultaneous denoising and interpolation of 3-D seismic data via damped data-driven optimal singular value shrinkage," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1086–1090, Jul. 2017.

[10] R. Rekapalli, R. K. Tiwari, M. K. Sen, and N. Vedanti, "3D seismic data denoising and reconstruction using multichannel time slice singular spectrum analysis," *J. Appl. Geophys.*, vol. 140, pp. 145–153, May 2017.
[11] M. Zhang, Y. Liu, M. Bai, and Y. Chen, "Seismic noise attenuation using unsupervised sparse feature learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9709–9723, Dec. 2019.
[12] S. Zu, H. Zhou, R. Wu, W. Mao, and Y. Chen, "Hybrid-sparsity constrained dictionary learning for iterative deblending of extremely noisy simultaneous-source data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2249–2262, Apr. 2019.
[13] F. Wang and S. Chen, "Residual learning of deep convolutional neural network for seismic random noise attenuation," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1314–1318, Aug. 2019.
[14] J. Sun, S. Slang, T. Elboth, T. Larsen Greiner, S. McDonald, and L.-J. Gelius, "Attenuation of marine seismic interference noise employing a customized U-Net," *Geophys. Prospecting*, vol. 68, no. 3, pp. 845–871, 2020.
[15] Y. Chen and S. Fomel, "Random noise attenuation using local signal-and-noise orthogonalization," *Geophysics*, vol. 80, no. 6, pp. WD1–WD9, Nov. 2015.
[16] H. Karsli and D. Dondurur, "A mean-based filter to remove power line harmonic noise from seismic reflection data," *J. Appl. Geophys.*, vol. 153, pp. 90–99, Jun. 2018.
[17] E. Shoshitaishvili, L. S. Sorenson, and R. A. Johnson, "Data improvement by subtraction of high-amplitude harmonics from the 2D land vertical- and multi-component seismic data acquired over the Cheyenne Belt in SE Wyoming," *71st Annu. Int. Meeting, SEG, Expanded Abstr.*, 2001, pp. 2021–2024.
[18] F. Linville and R. A. Meek, "Canceling stationary sinusoidal noise," *Geophysics*, vol. 57, no. 11, pp. 1493–1501, 1992.
[19] D. Nyman and J. E. Gaiser, "Adaptive rejection of high-line contamination," in *Proc. 53rd Annu. Int. Meeting*, 1983, pp. 1270–1274.
[20] K. E. Butler and R. D. Russell, "Subtraction of powerline harmonics from geophysical records," *Geophysics*, vol. 58, no. 6, pp. 898–903, Jun. 1993.
[21] J. Xia and R. D. Miller, "Design of a hum filter for suppressing power-line noise in seismic data," *J. Environ. Eng. Geophysics*, vol. 5, no. 2, pp. 31–38, Jun. 2000.
[22] A. Saucier, M. Marchant, and M. Chouteau, "A fast and accurate frequency estimation method for canceling harmonic noise in geophysical records," *Geophysics*, vol. 71, no. 1, pp. 7–18, 2006.
[23] C. Yang, L. Lu, H. Lin, R. Guan, X. Shi, and Y. Liang, "A Fuzzy-Statistics-Based principal component analysis (FS-PCA) method for multispectral image enhancement and display," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3937–3947, Nov. 2008.
[24] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 973–980, Mar. 2011.
[25] L. Xu, J. Li, Y. Shu, and J. Peng, "SAR image denoising via clustering-based principal component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6858–6869, Nov. 2014.
[26] S. K. Chiu, "Coherent and random noise attenuation via multichannel singular spectrum analysis in the randomized domain," *Geophys. Prospecting*, vol. 61, pp. 1–9, Jun. 2013.
[27] R. Adams, J. McIntyre, and F. Symonds, "Characteristics of the eastern interconnection line frequency," *IEEE Trans. Power App. Syst.*, vol. PAS-101, no. 12, pp. 4542–4547, Dec. 1982.

HANG WANG received the master's degree in geological resources and geological engineering from the China University of Petroleum, Beijing. He is currently pursuing the Ph.D. degree in geophysics with Zhejiang University. His research interests include seismic signal processing, machine learning, and FWI.

HUA ZHANG received the bachelor's and the master's degrees in geophysics from the China University of Mining and Technology, in 2004 and 2007, respectively, and the Ph.D. degree in geological resources and geological engineering from the China University of Petroleum, Beijing, in 2013.

His research interests include seismic data interpolation and denoising, simultaneous source separation, and seismic data processing.



YANGKANG CHEN (Member, IEEE) received the B.S. degree in exploration geophysics from the China University of Petroleum, Beijing, China, in 2012, and the Ph.D. degree in geophysics from The University of Texas at Austin, Austin, TX, USA, under the supervision of Prof. S. Fomel, in 2015.

From 2016 to 2018, he was a Distinguished Postdoctoral Research Associate with the Oak Ridge National Laboratory, Oak Ridge, TN, USA. He is currently an Assistant Professor with Zhejiang University, China. He has authored more than 100 internationally renowned journal articles and more than 80 international conference papers. His current research interests include seismic signal analysis, seismic modeling and inversion, simultaneous-source data deblending and imaging, global adjoint tomography, and high-performance computing.