

Application of principal component analysis in weighted stacking of seismic data^a

^aPublished in IEEE Geoscience and Remote Sensing Letters, 14, no. 8, 1213-1217, (2017)

*Jianyong Xie**, *Wei Chen[†]*, *Dong Zhang**, *Shaohuan Zu**, and *Yangkang Chen[‡]*

ABSTRACT

Optimal stacking of multiple datasets plays a significant role in many scientific domains. The quality of stacking will affect the signal-to-noise ratio (SNR) and amplitude fidelity of the stacked image. In seismic data processing, the similarity-weighted stacking makes use of the local similarity between each trace and a reference trace as the weight to stack the flattened prestack seismic data after normal moveout (NMO) correction. The traditional reference trace is an approximated zero-offset trace that is calculated from a direct arithmetic mean of the data matrix along the spatial direction. However, in the case that the data matrix contains abnormal mis-aligned trace, erratic and non-gaussian random noise, the accuracy of the approximated zero-offset trace would be greatly affected, thereby further influence the quality of stacking. We propose a novel weighted stacking method that is based on principal component analysis (PCA). The principal components of the data matrix, namely the useful signals, are extracted based on a low-rank decomposition method by solving an optimization problem with a low-rank constraint. The optimization problem is solved via a common singular value decomposition algorithm. The low-rank decomposition of the data matrix will alleviate the influence of abnormal trace, erratic and non-gaussian random noise, thus will be more robust than the traditional alternatives. We use both synthetic and field data examples to show the successful performance of the proposed approach.

INTRODUCTION

Stacking of multiple datasets to output a final smoothed data is a common research subject in many scientific fields. The stacking can be viewed as the simplest form of random noise attenuation for enhancing multi-dimensional remote sensing datasets (Tian et al., 2014; Zhuang et al., 2015). In seismic data processing, stacking simply means the summation of a collection of seismic traces from different records into a single trace. The quality of stacking greatly affect the performance of many seismic data processing tasks. It can be considered as the simplest way for improving the SNR (Elad and Aharon, 2006; Candès et al., 2006) in prestack seismic data processing. It can quickly obtain a meaningful post-stack seismic image without wavefield

continuation. The three steps in the stacking process are normal moveout (NMO) velocity analysis, NMO correction and trace summation.

The original weighted stacking used an SNR-based weighted stacking strategy to minimize random noise. Schoenberger (1996) proposed a different weighted stacking approach that can suppress the multiples effectively by solving a set of optimization equations in order to determine the stacking weights. Neelamani et al. (2006) took the signal structures into consideration and proposed a simultaneous stacking and denoising approach (SAD). Based on the statistics theory, Trickett (2007) proposed to use a maximum-likelihood estimator for weighted stacking by estimating the probability distribution of random noise. Tang (2007) calculated the stacking weights as functions of angle and azimuth and proposed a selective stacking approach.

Li and Gao (2014) proposed a novel method for stacking seismic data in time-frequency domain (Liu et al., 2016; Lin et al., 2015). Liu et al. (2009) proposed a similarity-weighted stacking approach that designs the weights of each trace by calculating the local similarity between each trace and a reference trace, and the method was demonstrated to be superior to the state-of-the-art weighted stacking approaches. The reference trace in the traditional similarity-weighted stacking method is an approximated zero-offset trace directly calculated from the spatial arithmetic mean of data matrix (Rubinstein et al., 2010). When the data matrix contains mis-aligned trace, erratic and non-gaussian random noise, the spatial arithmetic mean of the data matrix is of low fidelity to approximate the zero-offset trace. In this letter, we propose a novel principal component analysis (PCA) (Farrell and Mersereau, 2005; Du and Fowler, 2007) based weighted stacking method. Considering the complicated situations of field seismic data as mentioned above, we propose to extract the principal components of seismic data to approximate a highly accurate zero-offset trace. The principal components of the data matrix are extracted via solving an optimization problem with low-rank constraint. A singular value decomposition can be used to efficiently solve the optimization problem and then the low-rank approximation of the data matrix, which has a high SNR and is close to the ideal NMO-corrected common midpoint (CMP) gather, can be easily obtained. The new stacking method is easy to implement and can obtain significantly better stacked profile with cleaner geological structures. We first use a simple synthetic example to show the principle and then use a real pre-stack field data example to further demonstrate the tremendous improvement over the traditional approaches.

METHOD

Equal-weight stacking

In order to enhance the SNR of the zero-offset trace, all traces in a prestack CMP gather are summed. The traditional equal-weight stacking is the average of all traces

in a prestack CMP gather.

$$\hat{s}(t) = \frac{1}{N} \sum_{i=1}^N s_i(t), \quad (1)$$

where t is time, $s_i(t)$ denotes the i th trace. $\hat{s}(t)$ is the stacked zero-offset trace, and N is the number of traces.

Weighted stacking

However, the quality of each trace is different and each trace should contribute differently to the final stacked trace. A better stacking approach is to weight each trace before stacking based on certain criteria. The weighted stacking process can be formulated as

$$\hat{s}_w(t) = \frac{1}{\sum_{i=1}^N w_i(t)} \sum_{i=1}^N w_i(t) \cdot s_i(t) \quad (2)$$

where $w_i(t)$ is the weight applied to trace i and time t in a CMP gather. $\hat{s}_w(t)$ is the stacked trace after weighting.

Different methods have been proposed to apply weights according to different criteria. For example, the smart stacking proposed by Rashed (2008) is based on sign difference between sample point and the alpha-trimmed mean to remove frequency distortions. Neelamani et al. (2006) uses an iterative algorithm called *leave me out* (LMO) to estimate noise variances from data. The desired signal is assumed to be flat with constant amplitude across all the traces within a gather in the LMO method.

Similarity-weighted stacking

The similarity-weighted stacking method was proposed by Liu et al. (2009) to weight each trace according to the time-variant local similarity between each trace and the mean trace.

To implement the similarity-weighted stacking, we first apply the equal-weight stack to the NMO-corrected CMP gather to obtain the reference trace. Then we compute the local similarity (Chen et al., 2015) between the reference trace and the NMO-corrected CMP gather and apply soft thresholding (Donoho, 1995) to all local similarity values. Finally, we apply the weighted stack to the CMP gather using local similarity based weights to get the final stacked trace. The local similarity based weighting criteria is defined as:

$$w_i(t) = \begin{cases} \eta_i(t) - \epsilon, & \eta_i > \epsilon \\ 0 & \eta_i \leq \epsilon \end{cases}, \quad (3)$$

where ϵ is the threshold value, $\eta_i(t)$ is the local similarity between i th prestack trace and the reference trace:

$$\eta_i(t) = \mathcal{S}(s_i(t), \hat{s}(t)). \quad (4)$$

where $\mathcal{S}(a, b)$ denotes the local similarity between traces a and b . $\hat{s}(t)$ is the arithmetic mean as introduced in (1). ϵ can be intelligently chosen by defining a percentage. The percentage is used to preserve or reject the values during thresholding. For example, a percentage of 10 means that we preserve 10% largest value of the local similarity η . The result is not very sensitive to ϵ . We usually keep a 50% of largest η values to obtain the stacking result.

Low-rank approximation of the data matrix

In the similarity-weighted stacking approaches (2)-(4), $\hat{s}(t) = \frac{1}{N} \sum_{i=1}^N s_i(t)$ is not an appropriate zero-offset approximation. The spatial arithmetic mean of the data matrix is the true zero-offset trace only if the random noise is statistically white, and all traces after NMO correction are aligned well. Furthermore, no existing abnormal traces should exist in the data matrix. These requirements are seldom met due to the extremely complicated features of real seismic data and seismic data are always contaminated with different types of noise, e.g. erratic noise and colored noise.

A better way for calculating the approximation of the zero-offset trace is to calculate the spatial arithmetic mean of a low-rank approximated data matrix using principal component analysis (PCA). PCA is an important tool for multivariate analysis in statistics. The idea is to reduce the dimensionality of a data set while preserving as much variability of data variables as possible (Jolliffe, 2010).

Suppose the data matrix \mathbf{D} is composed of signal component \mathbf{S} , random noise \mathbf{N} , erratic noise \mathbf{E} , and mis-aligned data components \mathbf{M} :

$$\mathbf{D} = \mathbf{S} + \mathbf{N} + \mathbf{E} + \mathbf{M}. \quad (5)$$

For seismic stacking in this paper, \mathbf{D} is simply a common midpoint gather. If we assume the error components $\mathbf{N} + \mathbf{E} + \mathbf{M}$ are composed of small random perturbations, an optimal estimate of \mathbf{S} can be acquired via the following optimization problem:

$$\begin{aligned} \min \quad & \|\mathbf{N} + \mathbf{E} + \mathbf{M}\|_F^2, \\ \text{s.t.} \quad & \text{rank}(\mathbf{S}) = k, \quad \mathbf{D} = \mathbf{S} + \mathbf{N} + \mathbf{E} + \mathbf{M}, \end{aligned} \quad (6)$$

where k denotes the rank constraint applied to the target signal components. The problem can be efficiently solved via singular value decomposition (SVD). The observed data matrix \mathbf{D} can be decomposed into a group of eigen-images via the SVD. The low-rank component \mathbf{S} can be described with a few eigen-images that are associated with the largest singular values. The other noise items \mathbf{N} , \mathbf{E} , \mathbf{M} , however, will have energy spread over all the eigen-images.

Weighted stacking based on PCA

After the low-rank approximation of the data matrix $\hat{\mathbf{S}}$ is obtained, a better zero-offset reference trace can be obtained via calculating the arithmetic mean of the data

matrix along the spatial direction:

$$\hat{s}_{lr}(t) = \hat{S}\mathbf{a}, \quad (7)$$

where \mathbf{a} is an averaging column vector $[\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}]_{1 \times N}^T$. Here $[\cdot]$ denotes the transpose of the input matrix/vector. Substituting $\hat{s}(t)$ in (4) with $\hat{s}_{lr}(t)$, we can obtain a new weighting criteria:

$$\omega_i(t) = \begin{cases} \beta_i(t) - \epsilon, & \beta_i > \epsilon \\ 0 & \beta_i \leq \epsilon \end{cases}, \quad (8)$$

where $\beta_i(t)$ is the local similarity between i th prestack trace and the low-rank approximated reference trace:

$$\beta_i(t) = \mathcal{S}(s_i(t), \hat{s}_{lr}(t)). \quad (9)$$

Inserting the new weighting criteria, as shown in (8), into (2), we obtain the new PCA-based weighted stacking approach. The detailed algorithm workflow of the proposed weighted stacking approach can be expressed as:

1. Calculate the SVD of data matrix \mathbf{D} :

$$[\mathbf{U}, \Sigma, \mathbf{V}] = \text{SVD}(\mathbf{D}) \quad (10)$$

2. Calculate the low-rank approximated singular value matrix by selecting the k largest diagonal elements and setting others zero:

$$\hat{\Sigma} = \Sigma(1:k, 1:k) \quad (11)$$

3. Calculate the low-rank approximated data matrix

$$\hat{\mathbf{S}} = \mathbf{U}\hat{\Sigma}\mathbf{V}^T. \quad (12)$$

4. Calculate the arithmetic mean of the low-rank approximated data matrix according to (7).
5. Calculate the local similarity between each trace and the low-rank approximated zero-offset reference trace.
6. Calculate the PCA-based weighting function $w_i(t)$ according to (8).
7. Stack the CMP gather using the calculated weighting function according to (2).

For more complicated cases, the nonlinear equivalent of standard PCA (NPCA) can be used to potentially obtain even better performance. The NPCA reduces the observed variables to a number of uncorrelated principal components. The most important advantages of nonlinear over linear PCA are that it incorporates nominal and ordinal variables, and that it can handle and discover nonlinear relationships between variables, which may indicate that the traces may not need to be exactly flattened before stacking.

EXAMPLES

We first use a simple synthetic example to show the logic of the proposed approach. Fig. 1 shows an NMO-corrected CMP gather on the left. It is salient that there is an abnormal trace in the CMP gather. Besides, there is a significant amount of random noise in the data. For this data set, it is not plausible to directly take the arithmetic mean of the traces to stand for the zero-offset trace, since the abnormal trace will greatly degrade the fidelity of the equal-weight stacked trace and the non-gaussian random noise will cause a lot of residual noise in the stacked trace. However, the low-rank approximated data matrix (as shown in the right subfigure in Fig. 1) that alleviates the influence of the abnormal trace and non-Gaussian random noise is much cleaner and close to the ideal NMO-corrected CMP gather.

The second to fifth subfigures (from left to right) in Fig. 2 show the stacked traces using traditional stacking, smart stacking (Rashed, 2008), SNR stacking (Neelamani et al., 2006) and similarity weighted stacking (Liu et al., 2009), respectively. Despite the existence of the abnormal first trace, the similarity-weighted trace obtains the best stacking performance as shown in Liu et al. (2009). We zoom in the last figure in Fig. 2 and show it in the middle of Fig. 3. In addition, we show the clean signal on the left of Fig. 3 and the stacked signal using the proposed method on the right of Fig. 3. It is obvious that the abnormal trace makes the stacked trace based on local similarity deviate from the true signal a lot around 0.175s, as pointed out by the arrow. The proposed method, however, obtains a successful stacked signal. The SNR of the similarity based method is 8.85 dB and the SNR of the PCA based method is 9.38 dB. We can conclude from the synthetic example that the proposed method can not only help improve the SNR, but also can help better preserve the amplitude of the useful signals.

The number of stacking signals may be large, e.g., more than 50. The number of abnormal traces can also be larger than 1. Actually, when the number of stacking signals becomes larger, the method becomes more appealing, since more signals also implies more complicated signal features, more noise, and more abnormal traces. The superiority of the proposed method will become more practically beneficial when the data size and structure becomes more complicated, e.g., the field data example discussed next.

We then apply the proposed method to a field data example, as shown in Fig. 4. The second field data is a 2-D line from the Blake Outer Ridge area offshore Florida and Georgia. Fig. 5 shows a demonstration of flattening one CMP gather (Fig. 5a) using NMO velocity analysis (Fig. 5b) and NMO correction (Fig. 5c). The white line overlaying Fig. 5b denotes the automatically picked NMO velocity used for flattening the gather. Fig. 6 shows the weight calculated using the proposed method (equations 8 and 9). The stacked seismic images using different methods are shown in Fig. 7.

As shown in Fig. 7, the three sections (a),(b),(c) correspond to the stacked images of this field data example using three different methods: equal-weight stacking, similarity-weighted stacking, and PCA-based stacking. The proposed approach ob-

tains a more satisfying stacking performance, as shown in Fig. 7c. The amplitude of seismic reflections becomes obviously stronger using the proposed approach (Fig. 7c). Fig. 8 shows zoomed sections that correspond to the pink frame boxes shown in Fig. 7. It further confirms that the proposed method can help obtain cleaner stacked image and better preserve the amplitude of the useful signals. In order to quantitatively compare the stacking performance of different methods, we plot the average spectrum (1D FFT along the time axis) of all the traces of the stacked images for three methods and show them in Fig. 9. The black line denotes the average spectrum of equal-weight stacking method. The green line corresponds to similarity stacking method. The red line corresponds to the proposed approach. It is very obvious that the energy of the proposed algorithm is stronger than that of the equal-weight stacking method and similarity stacking method, indicating that the proposed algorithm helps best preserve the amplitude of seismic reflection events. We also calculate the maximum frequency energy (MFE), i.e., the maximum amplitude of the average frequency spectrum, for three methods. MFE is 231.83 for the equal-weight method, 415.24 for the similarity method, and 504.77 for the proposed method. The proposed method improve the equal-weight method by 117.33% and improve the similarity method by 21.56% in terms of MFE.

It is worth mentioning that there is no 100% reliable way to compare the stacking performance of real seismic data in the literature, since there is no ground truth for real data. When judging the stacking performance by human observation, we care about the "large amplitude/energy" of those key reflectors since stronger amplitude/energy makes us see clearer the reflectors, as highlighted by the ellipses and arrows in Fig. 8. When interpreting the seismic profiles, the No 1 principle is to detect those critical reflectors, the No 2 principle is the coherency of those reflectors since it is related to the horizon tracking in seismic interpretation. The "accuracy" or "fidelity" of amplitude is not very important in interpretation of time-migrated seismic images. However, the modern seismic imaging techniques such as true-amplitude least-squares migration (Chen et al., 2017), seek to obtain true-amplitude subsurface image where the amplitude correctness is of vital importance. However, this is beyond the scope of the letter For judging the stacking performance of field data, we highly recommend to judge by "observing" and "interpreting".

CONCLUSION

The similarity-weighted stacking approach can obtain a much improved stacking result than the equal-weight stacking considering the increased SNR, however, will still cause energy damage when an inappropriate reference trace is used to calculate the similarity based weights. We proposed a new weighted stacking method that is based on principal component analysis (PCA). The principal of the method is to prepare an ideal NMO-corrected data matrix via low-rank approximation. The low-rank approximation, or in other words the principal components, is obtained via solving a low-rank constrained optimization problem via singular value decomposition (SVD). The pro-

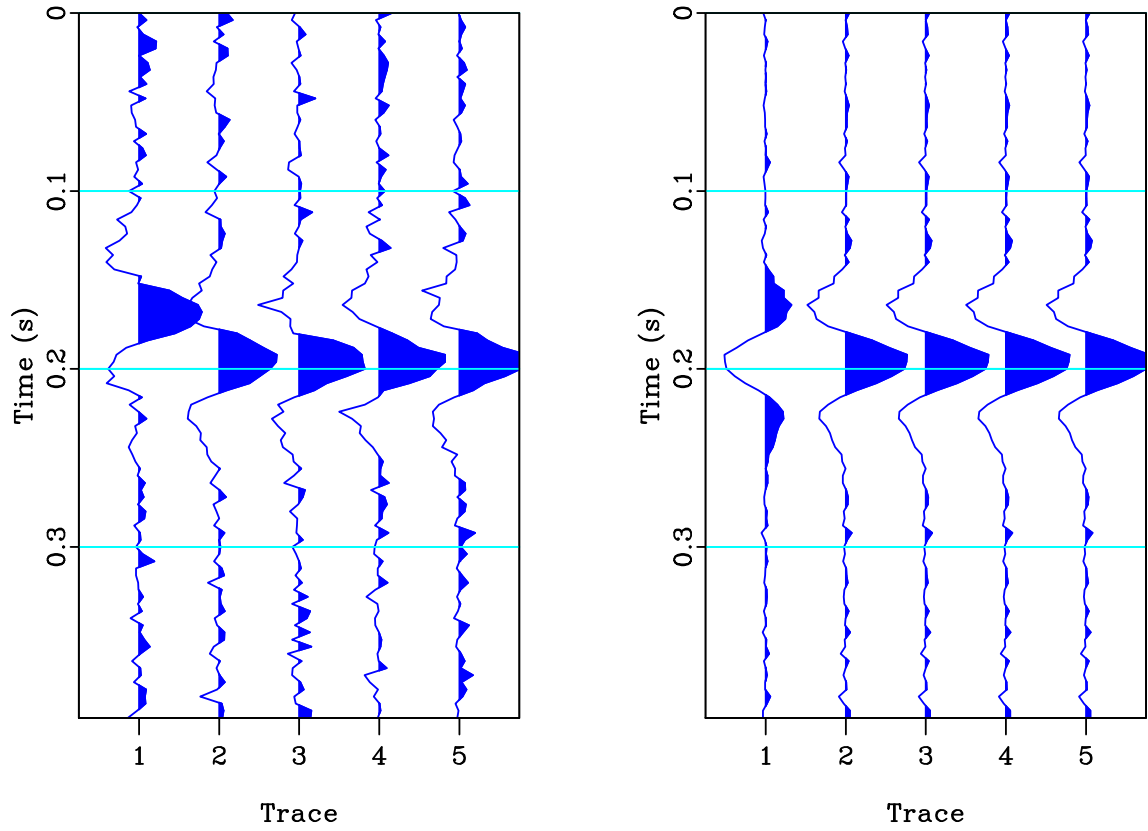


Figure 1: Left: Original NMO-corrected data. Right: Low-rank approximated data matrix.

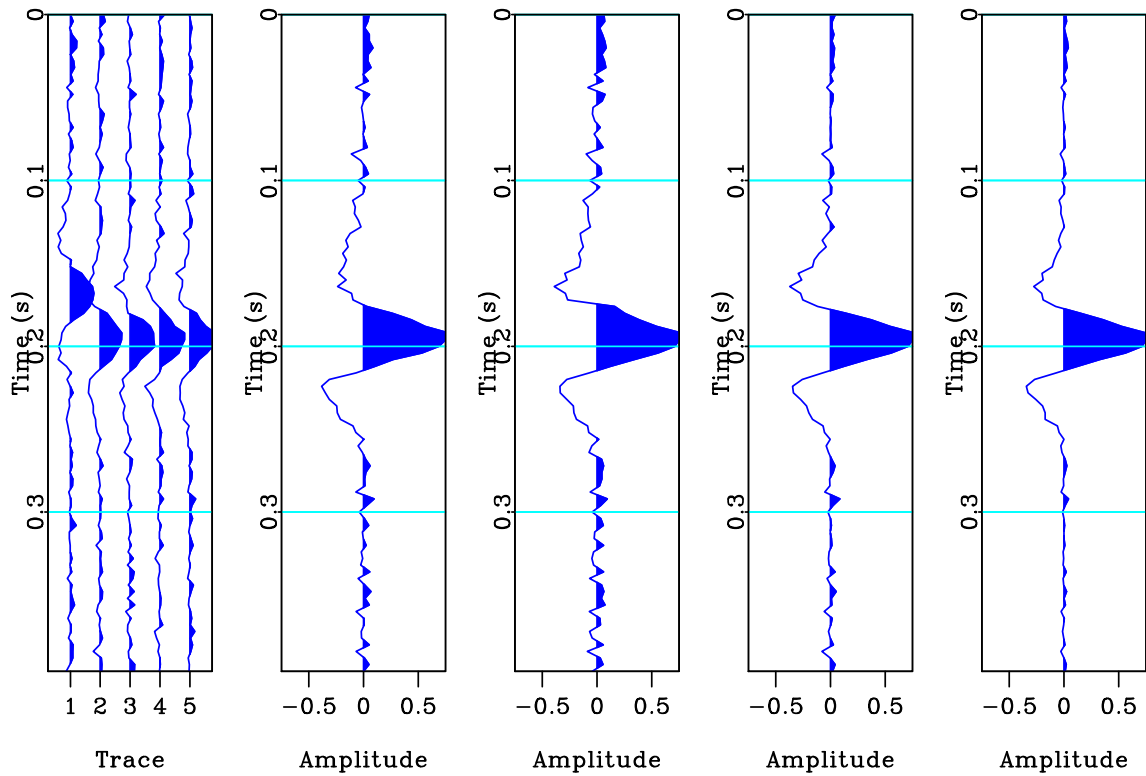


Figure 2: Rashed08

, SNR stack (Neelamani et al., 2006), and similarity stack (Liu et al., 2009).] From left to right: NMO-corrected data, conventional stack, smart stack (Rashed, 2008), SNR stack (Neelamani et al., 2006), and similarity stack (Liu et al., 2009).

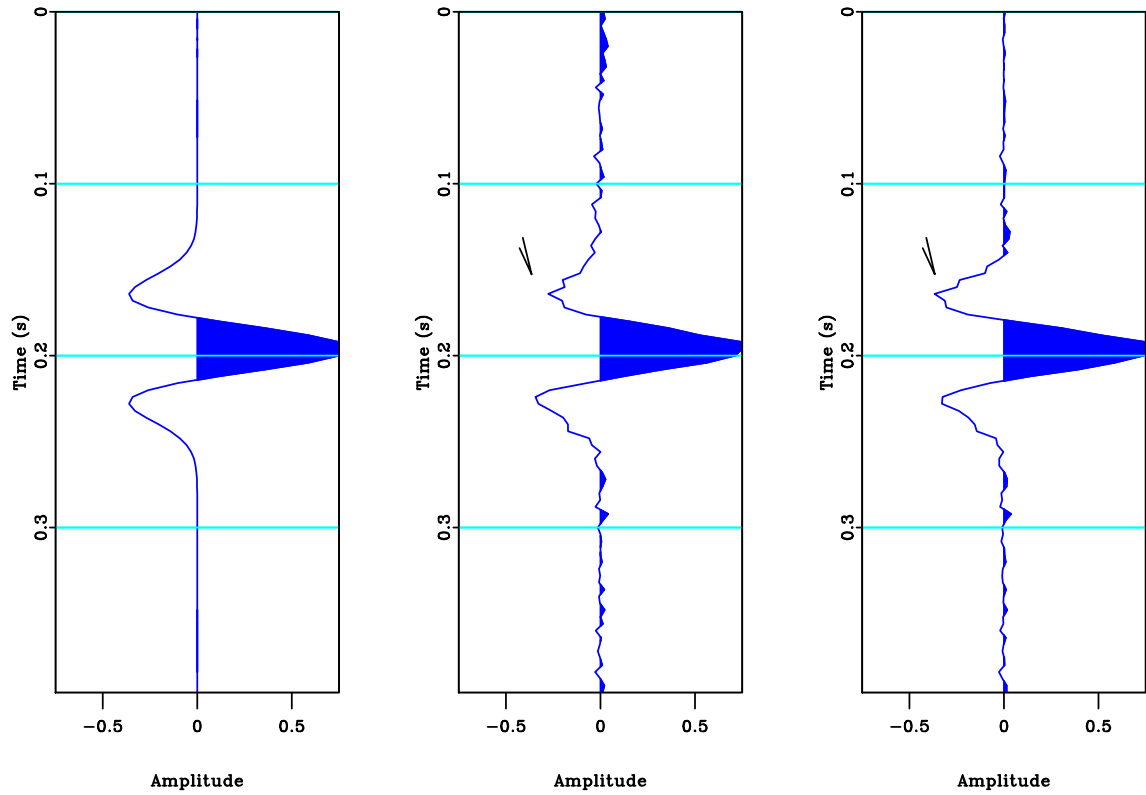


Figure 3: From left to right: true signal, stacked signal using the similarity-weighted stacking (SNR=8.85 dB), and stacked signal based on the PCA method SNR=9.38 dB.

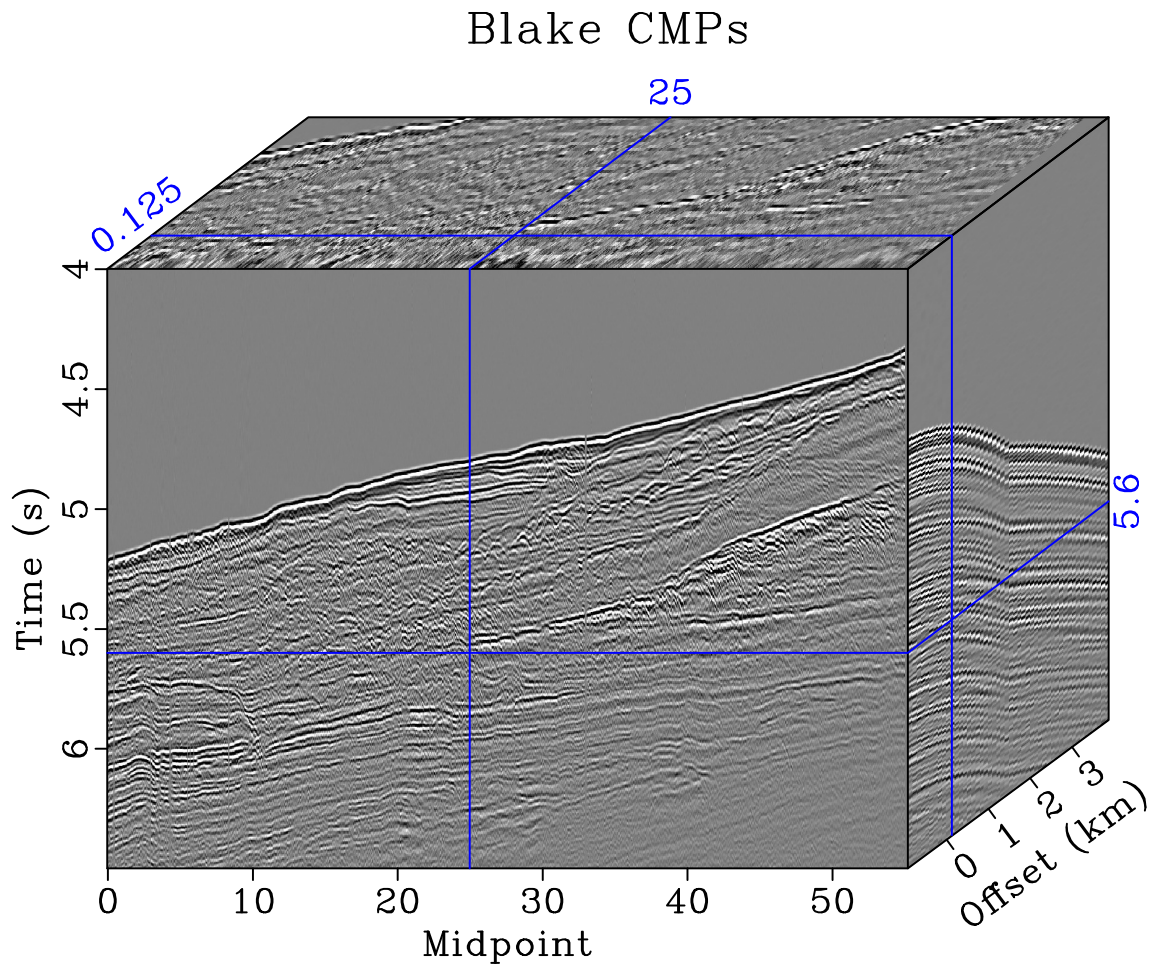


Figure 4: Field data example.

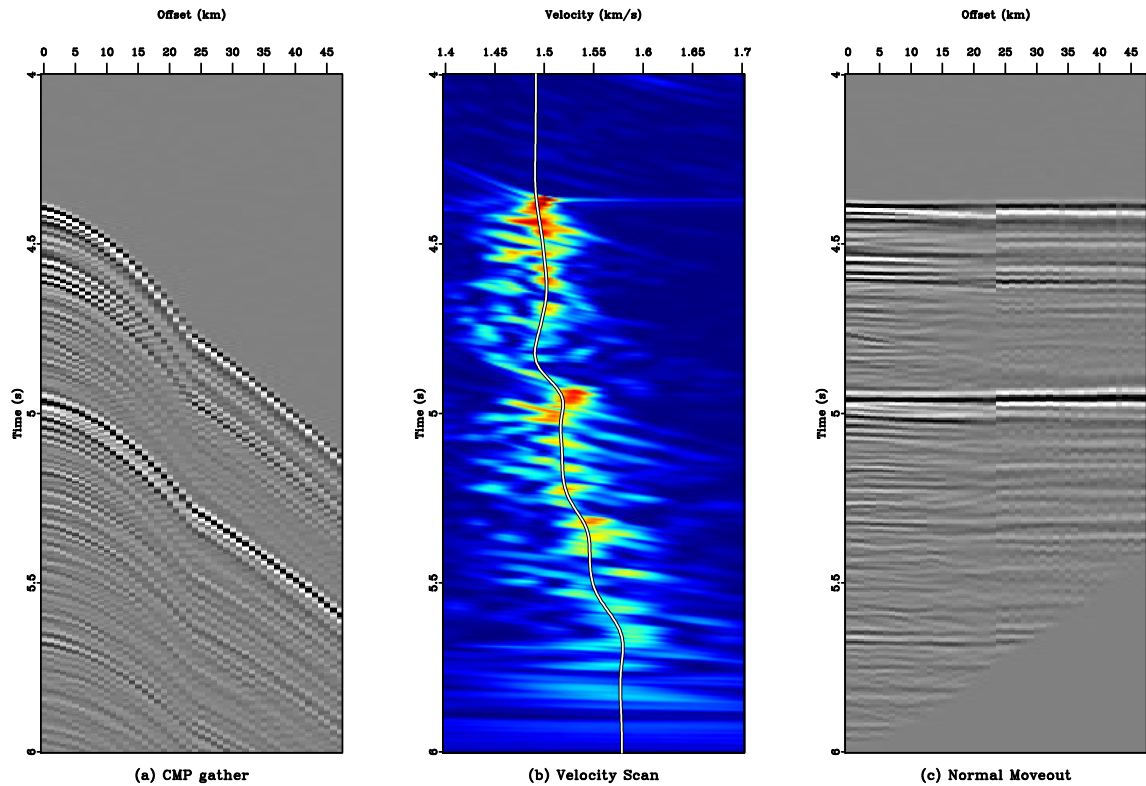


Figure 5: Demonstration of NMO correction for the field data example.

posed PCA-based stacking method can help alleviate the negative effects caused from abnormal trace, erratic and non-gaussian random noise existing in the data matrix, and thus is robust in field data processing. The proposed approach is tested via a synthetic CMP gather and a field data example, which shows very promising performance. Future research topics include substituting the current PCA framework with more sophisticated algorithms to make the obtained components statistically as independent as possible, such as Independent Component Analysis (ICA), where higher-order statistics rather than second-order moments are used to determine basic vectors and is proven to be stronger than PCA (Hyvärinen et al., 2001).

ACKNOWLEDGMENTS

This research is supported by Supported by State Key Laboratory of Geodesy and Earth's Dynamics (Institute of Geodesy and Geophysics, CAS, Grant No. SKLGED2017-4-3-E) and Sinopec Key Laboratory of Geophysics (Grant No. 33550006-15-FW2099-0017). J. Xie is supported by China Scholarship Council (201606440091). J. Xie also thanks the Shell Ph.D Scholarship to support excellence in geophysical research.

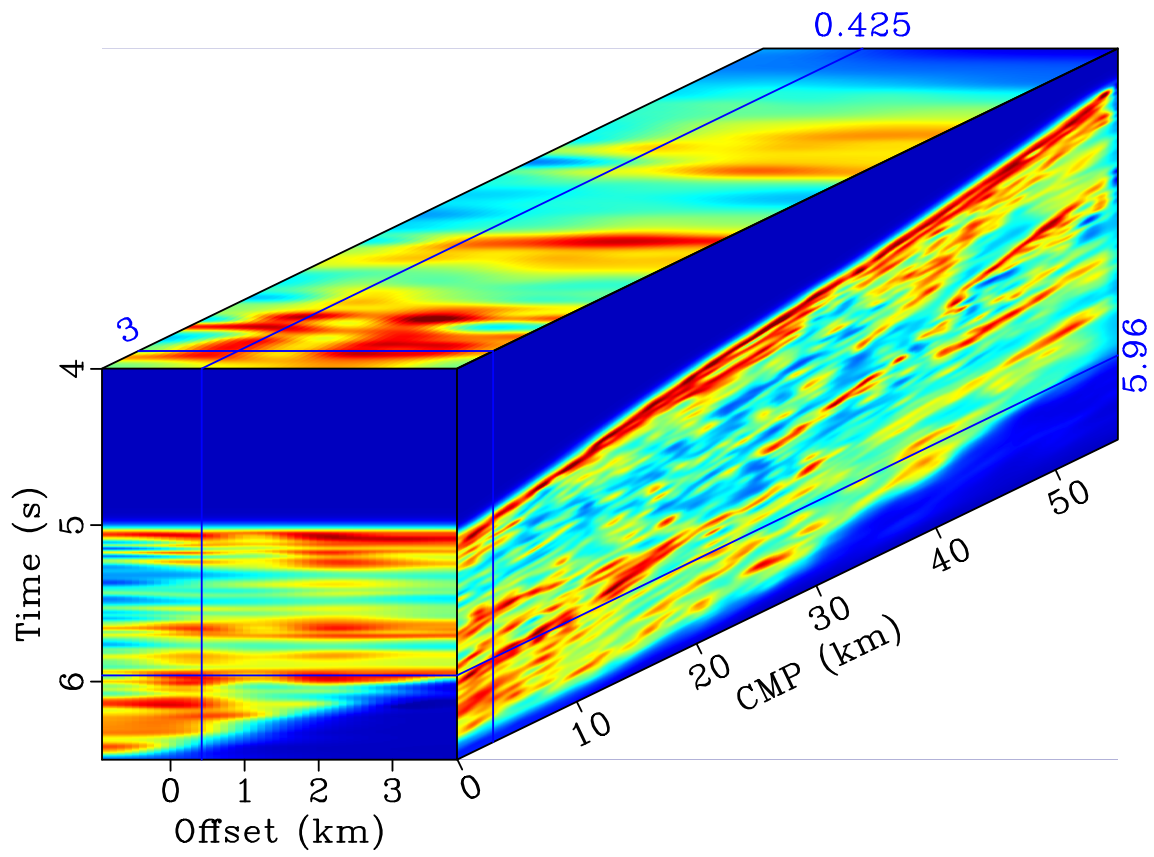


Figure 6: PCA-based weights for each time and space indices.

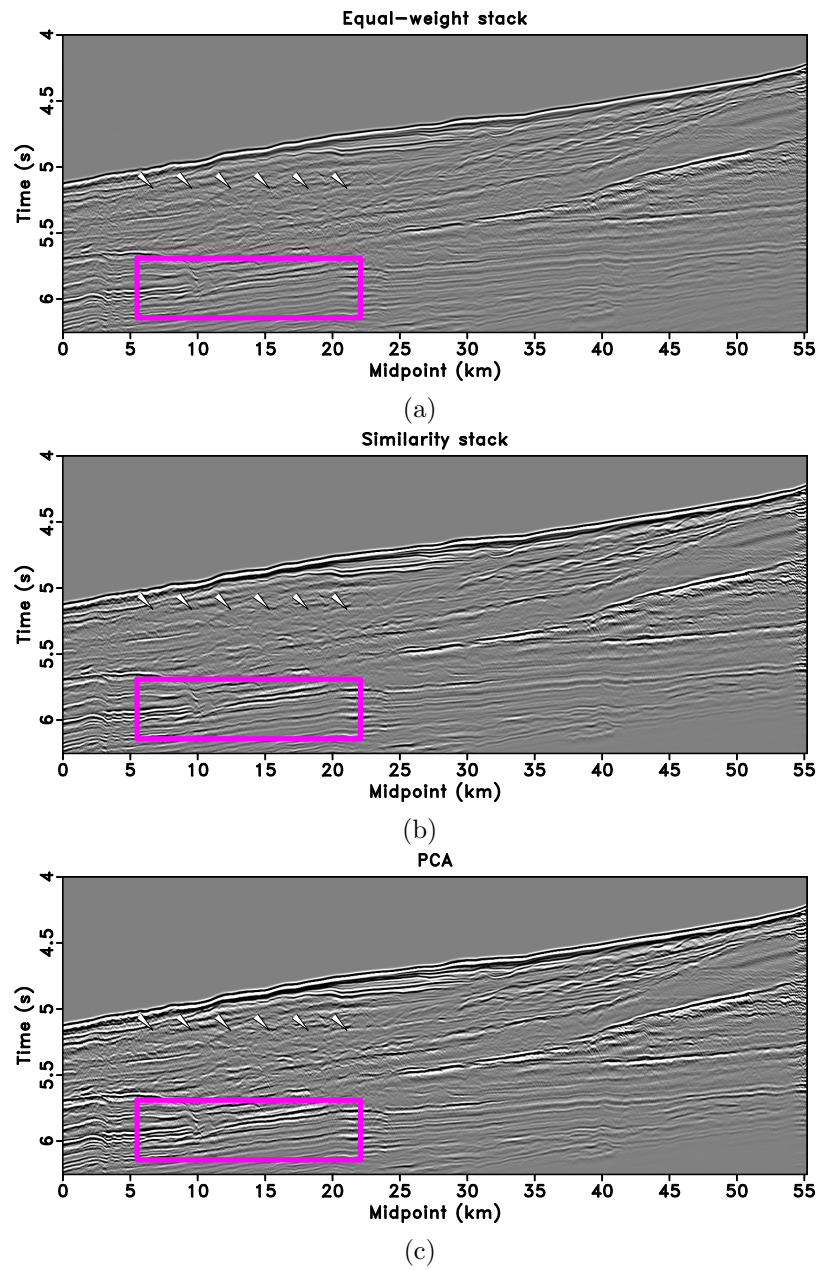


Figure 7: (a) Stacked result using equal-weight method. (b) Stacked result using local similarity based weight. (c) Stacked result using the proposed method.

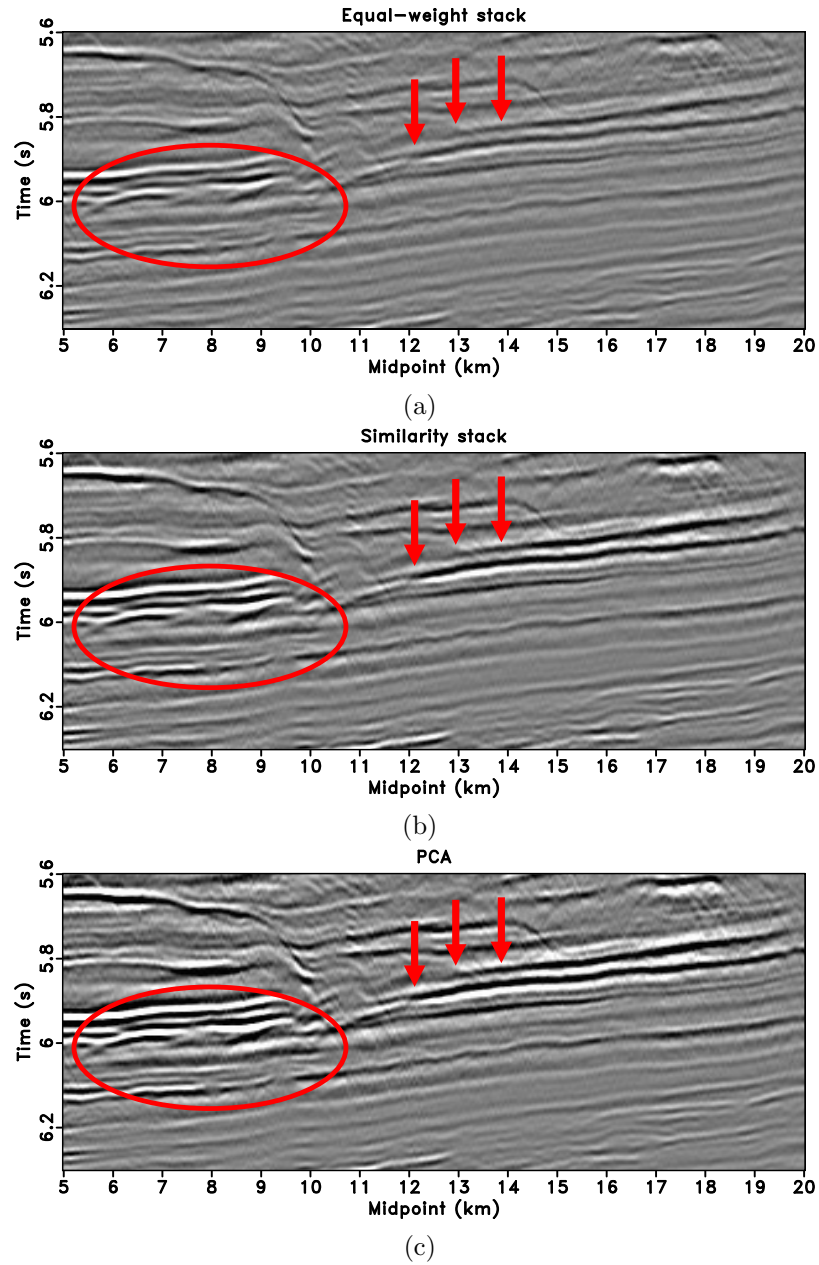


Figure 8: Zoomed sections from Fig. 7. (a) Stacked result using equal-weight method. (b) Stacked result using local similarity based weight. (c) Stacked result using the proposed method.

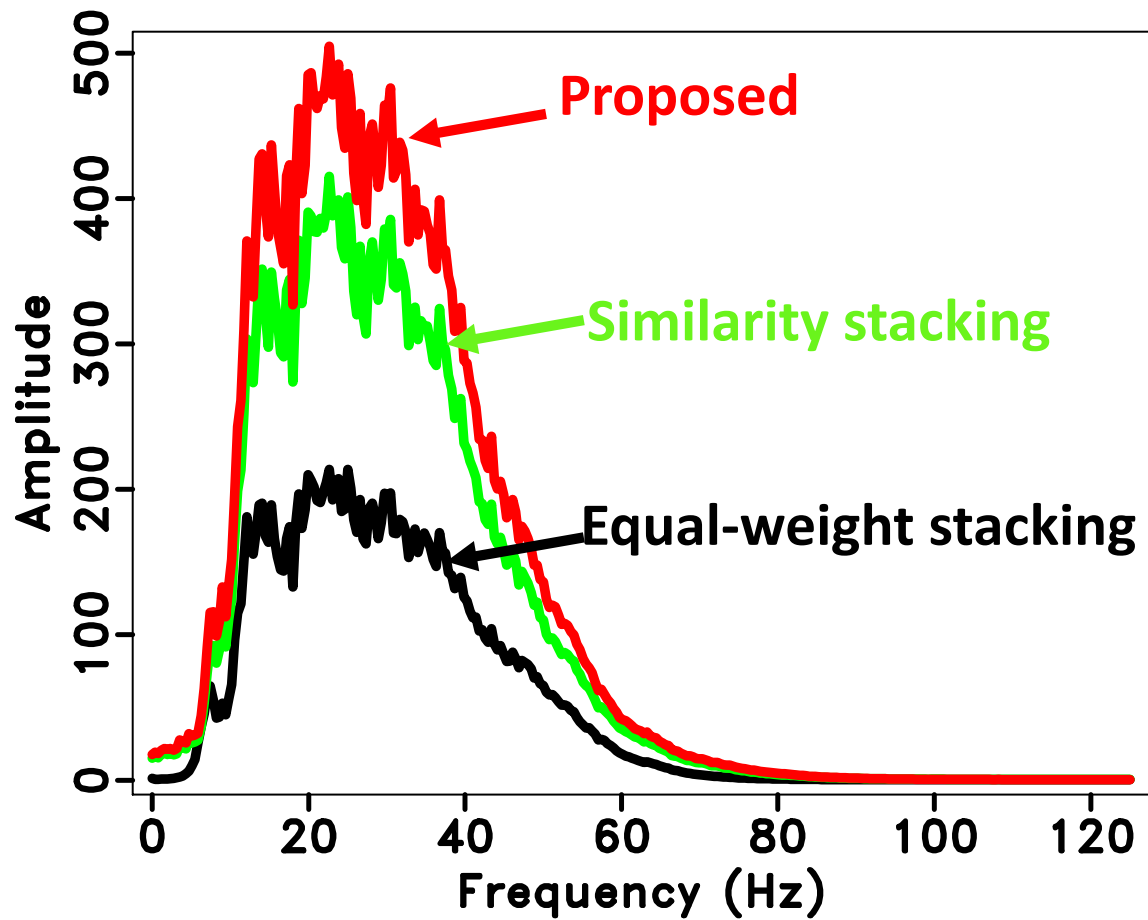


Figure 9: Comparisons of the average spectrum of all the traces. The black line denotes the average spectrum of equal-weight stacking method. The green line corresponds to similarity stacking method. The red line corresponds to the proposed approach. Note that energy of the proposed algorithm is obviously stronger than the other two methods.

REFERENCES

- Candès, E. J., J. Romberg, and T. Tao, 2006, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information: *IEEE Transactions on Information Theory*, **52**, 489–509.
- Chen, Y., H. Chen, K. Xiang, and X. Chen, 2017, Preserving the discontinuities in least-squares reverse time migration of simultaneous-source data: *Geophysics*, **82**, no. 3, S185–S196.
- Chen, Y., S. Jiao, J. Ma, H. Chen, Y. Zhou, and S. Gan, 2015, Ground-roll noise attenuation using a simple and effective approach based on local band-limited orthogonalization: *IEEE Geoscience and remote sensing letters*, **12**, 2316–2320.
- Donoho, D. L., 1995, De-noising by soft-thresholding: *IEEE transactions on information theory*, **41**, 613–627.
- Du, Q., and J. Fowler, 2007, Hyperspectral image compression using jpeg2000 and principal component analysis: *IEEE Geoscience and Remote Sensing Letters*, **4**, 201–205.
- Elad, M., and M. Aharon, 2006, Image denoising via sparse and redundant representations over learned dictionaries: *IEEE Transactions on Image Processing*, **15**, 3736–3745.
- Farrell, M., and R. M. Mersereau, 2005, On the impact of pca dimension reduction for hyperspectral detection of difficult targets: *IEEE Geoscience and Remote Sensing Letters*, **2**, 192–195.
- Hyvärinen, A., J. Karhunen, and E. Oja, 2001, *Independent component analysis*: New York: John Wiley & Sons.
- Jolliffe, I. T., 2010, *Principal component analysis*, 2nd ed.: Springer.
- Li, Q., and J. Gao, 2014, Application of seismic data stacking in time-frequency domain: *IEEE Geoscience and Remote Sensing Letters*, **11**, 1484–1488.
- Lin, H., Y. Li, H. Ma, B. Yang, and J. Dai, 2015, Matching-pursuit-based spatial-trace time-frequency peak filtering for seismic random noise attenuation: *IEEE Geoscience and Remote Sensing Letters*, **12**, 394–398.
- Liu, G., S. Fomel, L. Jin, and X. Chen, 2009, Stacking seismic data using local correlation: *Geophysics*, **74**, V43–V48.
- Liu, W., S. Cao, and Y. Chen, 2016, Seismic time-frequency analysis via empirical wavelet transform: *IEEE Geoscience and Remote Sensing Letters*, **13**, 28–32.
- Neelamani, R., T. A. Dickens, and M. Deffenbaugh, 2006, Stack-and-denoise: A new method to stack seismic datasets: 76th Annual International Meeting, SEG, Expanded Abstracts, 2827–2831.
- Rashed, M., 2008, Smart stacking: A new CMP stacking technique for seismic data: *The Leading Edge*, **27**, 462–467.
- Rubinstein, R., M. Zibulevsky, and M. Elad, 2010, Double sparsity: learning sparse dictionaries for sparse signal approximation: *IEEE Transaction on signal processing*, **58**, 1553–1564.
- Schoenberger, M., 1996, Optimum weighted stack for multiple suppression: *Geophysics*, **61**, 891–901.
- Tang, Y., 2007, Selective stacking in the reflection-angle and azimuth domain: 77th

- Annual International Meeting, SEG, Expanded Abstracts, 2320–2324.
- Tian, Y., Y. Li, and B. Yang, 2014, Variable-eccentricity hyperbolic-trace TFPF for seismic random noise attenuation: *IEEE Transactions on Geoscience and Remote Sensing*, **52**, 6449–6458.
- Trickett, S., 2007, Maximum-likelihood-estimation stacking: 77th Annual International Meeting, SEG, Expanded Abstracts, 2640–2643.
- Zhuang, G., Y. Li, Y. Liu, H. Lin, H. Ma, and N. Wu, 2015, Varying-window-length TFPF in high-resolution radon domain for seismic random noise attenuation: *IEEE Geoscience and Remote Sensing Letters*, **12**, 404–408.