Deep learning for prestack strong scattered noise suppression

Dawei Liu[*1,2], Xiaohai Yang[3], Xiaokai Wang[1], Haibo Mao[3], Mauricio D. Sacchi[2], Wenchao Chen[1]*
[1]Xi'an Jiaotong University [2]University of Alberta [3]Institute of Geophysics, Xinjiang Oilfield Company Ltd.

SUMMARY

As a robust feature extraction method, deep learning has made significant progress in attenuating noise from seismic datasets. One critical assumption of deep learning for prediction is that test and training data should arise from the same distribution. Poststack data from a given survey can approximately meet this assumption because of subsurface structure similarities. Also, poststack data has a high signal-to-noise ratio (SNR) and a moderate amplitude variation, which makes the network training relatively easy. Therefore, denoising poststack seismic data with deep learning techniques appears to be a solved problem. However, noise is often prevalent in prestack data. The noticeable amplitude decay and significant waveform changes of reflections make network training unstable. More than that, the strong near-surface scattered noise on land data, which often overlaps with valuable signals in the $t-x$ domain and $f-k$ domain, poses a severe challenge to the conventional suppression strategy in common shot and common receiver gathers. Supervised deep learning methods, which require access to realistic learning samples, have failed to make a breakthrough on prestack seismic data denoising. Some unsupervised deep learning methods have made progress in areas with weak scattered noise, but further work is required to meet the industry's requirements. To deal with the problem above and make deep learning more generalized and tractable for processing prestack denoising, we propose to train a denoising network on the offset vector tile (OVT) domain. OVT is a particular prestack seismic gather type that can faithfully represent a continuous wavefield; hence, it is an excellent domain to extract seismic data features. We use a 3D survey containing 1260 OVT volumes to illustrate the validity of the proposed methods. Only two OVT volumes with the same azimuth value are used to train the network, and the rest of the 1258 OVT volumes are adopted as test data. The results show that our method can effectively attenuate random and scattered noise. Moreover, our approach ameliorates the poor denoising performance on the boundary of OVT gathers compared to the conventional method used to construct training samples. It is worth mentioning that our calculation time for an OVT volume ($200 \times 200 \times 3001$) is only about 6 minutes, which is about one-tenth of the comparison conventional method. This work focuses on the following aspects: We study the suppression of intense prestack scattered noise based on deep learning, we also examine which is the best domain to perform deep learning denoising.

INTRODUCTION

Improving noise removal in seismic data processing leads to a more confident interpretation. In the past, researchers introduced a wide range of methods for seismic noise attenuation. Each category of these methods utilizes a particular prior knowledge of the physical where the signal is represented. Typically, filtering methods, based on the predictable property of the seismic signal, are most frequently used, including *f-k* filtering (Embree et al., 1963; Treitel et al., 1967), *f-x* deconvolution (Gulunay, 1986), and multichannel singular spectrum analysis (Oropeza and Sacchi, 2011). A further category of seismic denoising techniques considers the the low-rank property, including low-rank factorization (Trickett, 2003) and nuclear norm minimization (Kreimer and Sacchi, 2012). An important category of seismic denoising techniques uses the signal or noise sparsity in a certain transform domain, such as wavelet transform domain (Deighan and Watts, 1997; Chen et al., 2017) and curvelet transform (Naghizadeh and Sacchi, 2018).

Techniques also exist in attenuating seismic noise in the OVT domain. OVT technology applies to wide-azimuth seismic exploration (Vermeer, 2007). Each OVT is built from a limited range of shots along the shot line and a limited range of receivers along the receiver line, thereby having a limited range of offset and azimuth. Hence, spatial continuity of prestack wavefield improves, which is beneficial to noise attenuation. Li et al. (2015) applied volume $\tau-p$ transform to denoise a low SNR seismic data on the OVT domain. Duan et al. (2016) combined both 5D interpolation and migration in the OVT domain and showed that footprint can be effectively suppressed from the source. Ling and Hu (2019) sort data into the OVT domain and found that such OVT gathers are suitable for internal multiples attenuation. Sun et al. (2019) utilized curvelet transform to denoise OVT gathers with a low SNR. Li et al. (2019) sorted data into OVT and utilized a unified learning-based framework to improve conventional denoising methods. Performing OVT processing and preserving azimuth information requires handling massively increased data volumes, putting forward high processing speed requirements.

In recent years, we have witnessed a growing academic interest in deep learning. Unlike conventional methods, deep learning automatically learns feature extraction ability from training data instead of adopting handcrafted filters, which mainly depends on a priori knowledge of designers and fails to consider the benefits of big data. The more significant part of the literature on noise attenuation with deep learning focuses on the supervised learning method. Zhang et al. (2010) first applied BP neural network to seismic random noise reduction. Song et al. (2020) used a deep convolutional autoencoder neural network to eliminate random seismic noise. Liu et al. (2020) proposed a 3D-DnCNN to denoise poststack seismic data and reported that deep learning can outperform the conventional method in suppressing arc-like imaging noise. Sun et al. (2020) employed a customized network with element-wise summation to marine seismic interference noise and pointed out that its processing speed is significantly faster than any existing industry denoising algorithm. Overall, these studies highlight deep learning results with powerful feature extraction ability

10.1190/segam2021-3583393.1

and super fast computing speed due to easy access to GPU.

We propose a new prestack seismic noise attenuation method combining deep learning with the OVT partitioning method. In learning theory, the training and test sets are assumed to be drawn from the same probability distribution. For poststack noise attenuation, this assumption can be followed by training a subvolume generated from the whole survey because the subsurface structures are similar. However, the SNR of prestack data is much lower, and the amplitude decays due to wave propagation are large, making distribution changes. Deep learning often suffers severe performance degradation when the distribution of test data mismatches the training data. Therefore, deep learning for prestack data denoising is rarely applied on a large scale in practical applications, significantly when contaminated by strong scattered noise. To make deep learning more generalized in prestack denoising, we sort the seismic data into OVT before training the network. The prestack wavefield of OVT is smooth and continuous. Moreover, there is minor spatial and temple amplitude variation in OVT. The sound characteristics of OVT domain data offer a favourable data foundation for network learning. In return, the massive data of OVT can give full play to the advantages of deep learning computing efficiency.

METHOD

Model formulation

Noise attenuation, which recovers useful signals $\mathbf{x} \in \mathcal{X} = \mathbb{R}^m$ from noisy data $\mathbf{y} \in \mathcal{Y} = \mathbb{R}^m$, is an important subfield of statistical data analysis. We model the noisy prestack seismic data

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{y}$ is contaminated with scattered and random noise $\mathbf{n}$. The supervised deep learning approaches have been developed for suppressing seismic noise. Deep neural networks interpret the image denoising problems as a regression problem and learn to map the noisy seismic data to clean reflections. This happens by training a convolutional neural network (CNN) with a large number of training sample pairs containing noisy input and clean labels

$$(\mathrm{y}_i, \mathrm{x}_i) \sim (\mathrm{Y}, \mathrm{X}) = (\mathrm{X} + \mathrm{N}, \mathrm{X}), \quad i = 1, \ldots, N \tag{2}$$

where Y, X, and N are random variables taking values in $\mathcal{Y}$, $\mathcal{X}$, and $\mathcal{N}$, respectively. By network training, deep learning aim to find the regression function

$$h^* = \underset{h}{\operatorname{argmin}}\, \mathbb{E}_{\mathrm{X,N}} \left\{ L\left(h(\mathrm{X}+\mathrm{N}), \mathrm{X}\right) \right\}, \tag{3}$$

which is an expected risk minimization task. Here, we chose the most commonly used loss function, that is, pixel-wise mean square error

$$L\left(h(\mathrm{X}+\mathrm{N}), \mathrm{X}\right) = \|h(\mathrm{X}+\mathrm{N}) - \mathrm{X}\|_2^2. \tag{4}$$

Since joint distribution function $P(\mathrm{Y}, \mathrm{X})$ is unknown, equation (4) is usually intractable. To circumvent it, the expectation is estimated by the empirical risk, which is the sample mean over the training dataset. The empirical risk minimization task is minimized over the CNN parameterized mappings $f_\theta : \mathcal{Y} \to \mathcal{X}$ with parameters θ. The network training is then equivalent to find the optimal parameters

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} \|f_\theta(\mathrm{y}_i) - \mathrm{x}_i\|_2^2, \tag{5}$$

which minimizes the loss on the training sample pairs. After training, the network f_{θ^*} is applied to unseen noisy seismic data to obtain useful signals. It is worth noting that the training data and test data should independent and identically distributed. Otherwise, the test data is not reliable, and the dataset bias leads to incorrect predictions.

Training samples

Denoising prestack seismic data faces more challenges compared to denoise poststack data. It has to confront the amplitude decay and waveform distortion due to wave-propagation effects. Moreover, prestack seismic data are typically embedded with a lot of noise. A high degree of scattered noise masks the primary reflection data and is very difficult to remove without damaging reflections. There are many kinds of prestack seismic data, such as common shot , common receiver, and common midpoint gathers. Good selection of data domain can reduce the adverse impact mentioned above. To make the network easier to learn the distribution of valuable signals, we choose the appropriate data type, namely OVT. OVT is single-fold coverage of the entire survey area with similar offsets and azimuths, thereby reducing the spatial discontinuity and laying the data foundation for network learning.

Deep learning needs a large training dataset containing noise-free data, which are not available in practice. Instead, we use the denoising results of a conventional method and approximately consider them as the training labels. After sorting the data into OVT, a noise attenuation method based on 3-D continuous wavelet transform (3D-CWT) is applied to denoise the prestack seismic data (Wang et al., 2020). 3D-CWT can make full use of the spatial structure information of seismic data, so it is more suitable for processing 3D seismic data. The definition of 3D-CWT is

$$\begin{aligned} CWT(f;\mathbf{b},a,\rho,\varphi) &= \left\langle f, \psi_{\mathbf{b},a,(\rho,\varphi)} \right\rangle \\ &= \frac{1}{a^3} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\mathbf{m}) \psi^* \left(\frac{1}{a} \Re_{\rho,\varphi} (\mathbf{m} - \mathbf{b}) \right) dxdydz \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \hat{f}(\mathbf{k}) \hat{\psi}^* \left(a \Re_{\rho,\varphi}(\mathbf{k}) \right) e^{j\mathbf{bk}} dk_x dk_y dk_z \end{aligned} \tag{6}$$

where a is the dilation parameter, $\mathbf{m} = (x,y,z)^T$ and $\mathbf{k} = (k_x, k_y, k_z)^T$ are 3D vectors, $\mathbf{b} = (b_x, b_y, b_z)^T$ is a 3D vector of translation operation, and (ρ, φ) is the rotation parameter. $\Re_{\rho,\varphi}$ is a rotation operation and indicates the rotation of a vector by angle ρ along the dip direction and angel φ along the azimuth direction. $\hat{f}(\mathbf{k})$ and $\hat{\psi}(\mathbf{k})$ is wavenumber domain of $f(\mathbf{x})$ and $\psi(\mathbf{x})$, respectively. $\psi(\mathbf{x})$ is 3D-CWT and here we choose 3D Morlet wavelet because it has orientation-selective properties.

CWT is computationally intensive, requiring to divide the space into very fine-grained grids to ensure that wavelets could extract the complete signal. To speed up and make 3D-CWT more efficient, we use a fast 3D-CWT with fast Fourier transform to get noisy seismic data coefficients. For each parameter $(a, \rho, \phi, \mathbf{b})$, we calculate the wavelet subband and apply a threshold filter to the subband. Finally, all filtered subbands are utilized to reconstruct the useful signals $\hat{\mathbf{x}}$. In this way, we build N training sample pairs $\{(y_i, \hat{x}_i)\}_N^{i=1}$ and then feed them to the network.

Network architecture

Our network architecture is a feedforward neural network, as shown in Figure 1. The input layer contains a convolution operation followed by an activation function to increase the nonlinearity of the neural network. ReLU is the most used activation function because a model that uses it is easier to train and often achieves better practice performance. Hence, we set ReLU as the default activation function. In the middle is 15 hidden layers. Unlike the input layer, the batch normalization (BN) is added before each activation function to accelerate deep network training. The last layer is the convolutional output layer. We use the same padding to produce an output of the same size as the input. It is worthwhile to mention that all the convolution operations are 3D to fully utilize the spatial structure information of seismic data.

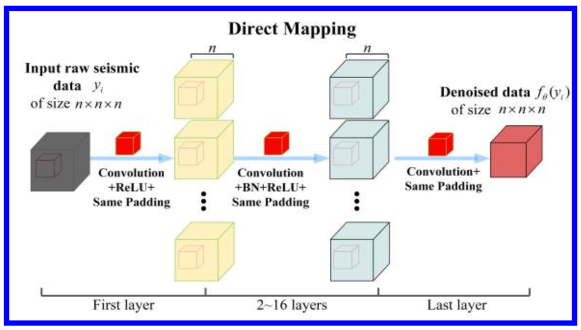

Figure 1: The network architecture used in this abstract

The biggest difference in our network architecture is that we use the strategy of direct mapping instead of residual learning, which was frequently used in previous networks. Indeed, Residual learning is famous for removing additive white Gaussian noise. This is partly because it learns a map from noisy data to noise where the noise has a fixed probability distribution. The learning target is stable, so it has achieved good performance in noise removal. However, the amplitude changes in prestack seismic data, and so does noise distribution. Thus, learning a map to noise may not be a good choice. At the same time, we notice that the valuable signals in the OVT domain are spatially continuous, and the distributions are similar. More importantly, the underground structures of different OVT volumes are the same. Therefore, in denoising in the OVT domain, it is easier to learn a mapping towards proper signals than noise, so we choose the direct mapping strategy.

EXAMPLES

We use a 3D wide azimuth survey in Western China to illustrate the proposed method and its validity. Figure 2 displays the offset-azimuth locations of OVT volumes in polar coordi-

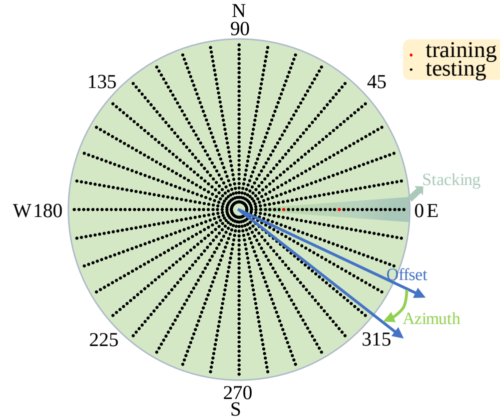

Figure 2: Locations of OVT volumes in offset-azimuth polar coordinates.

nates. We see that this survey has 1260 OVT volumes, composed of 35 offsets by 36 azimuths. As indicated by the red points in Figure 2, two OVT volumes with the same azimuth are filtered by 3D CWT and then fed to the network for training. After that, the well-trained network is applied to denoise the whole survey in the OVT domain.

Figure 3a plots a noisy inline section of OVT volumes, which is severely contaminated with random noise and near-surface scattered noise. Figure 3b and Figure 3c are the denoised data and removed noise by 3D CWT, respectively. Despite the low SNR of original data, the coherent noise in the denoised data is significantly weakened, and the reflections become much clearer. This proves that the OVT domain is suitable for noise suppression because of a continuous wavefield. Moreover, It can be seen from Figure 3c that the noise distribution, especially the near-surface scattered noise, is nonuniform. In this case, using the network to learn a map to noise should not be a good choice. Figure 3d and Figure 3e are our results, which utilize the network to learn a direct mapping to useful signals. From the figures above, we can see that the SNR has been significantly improved, and no obvious damage to the useful signals can be found. Closer inspection of crossline results in Figure 4 and time slice results in Figure 5 can draw the same conclusion. The most surprising aspect of our method is that the network can learn the distribution of useful signals by learning minimal amounts of OVT volumes, and then it can suppress the noise of all other OVT volumes.

To further compare the denoising results, we stack the OVT volumes with the stack range shown in Figure 2. Figure 6a displays the stack results without noise attenuation, and a lot of noise can still be observed. Both stack results of 3D CWT and our proposed methods can enhance the weak signal and obtain clear events. Figure 7 reveal that our approach is similar to 3D CWT in general, which proves that our learning strategy is successful. Besides, our boundary is better processed.

Finally, we sort the data used for stacking into common midpoint gathers. The scattered noise in Figure 8a severely masks the primary reflections. We see that the denoising results in Figure 8 also accord with earlier observations, which shows that our method has successfully removed the scattered noise.

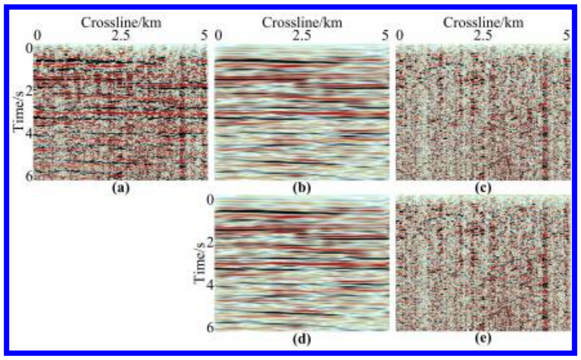

Figure 3: Inline section results of a OVT volume. (a) noisy. (b) and (c) are results by 3D CWT. (d) and (e) are our reuslts.

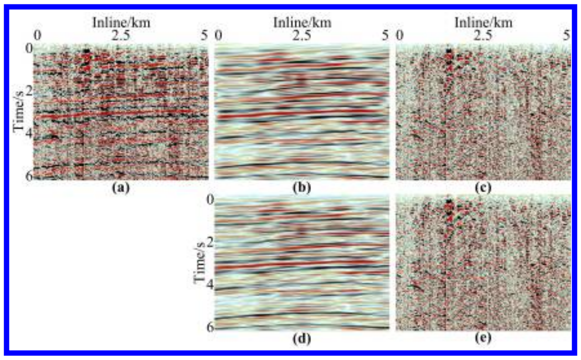

Figure 4: Crossline section results of a OVT volume. (a) noisy. (b) and (c) are results by 3D CWT. (d) and (e) are our reuslts.

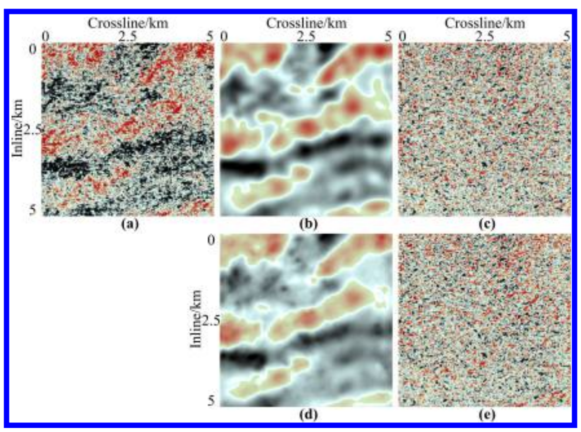

Figure 5: Time slice results of a OVT volume. (a) noisy. (b) and (c) are results by 3D CWT. (d) and (e) are our reuslts.

CONCLUSION

We propose a prestack denoising method by combining the merits of deep learning and OVT partitioning techniques. The wavefield continuity in the OVT domain provides good learning conditions for the network. The field results demonstrate that only a minimal number of OVT volumes can make the network obtain the ability to suppress the whole survey's noise. This is very promising in practice because the prestack data is massive, and deep learning can save ten times the processing time compared with the conventional method used in this abstract.

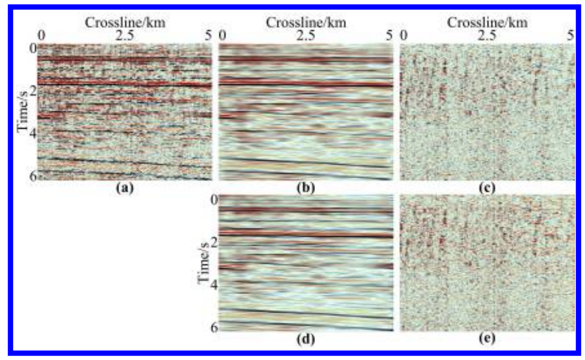

Figure 6: Inline section results of stacked data. (a) noisy. (b) and (c) are results by 3D CWT. (d) and (e) are our reuslts.

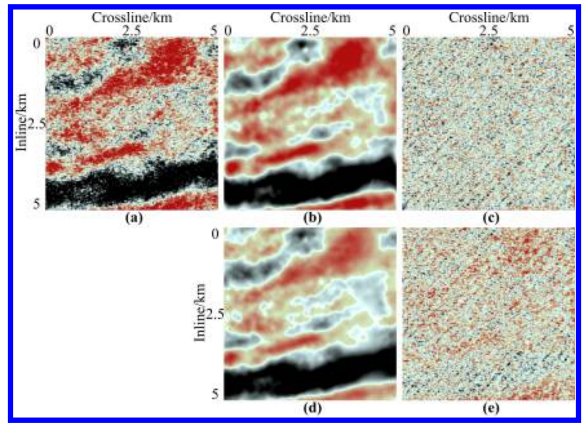

Figure 7: Time slice results of stacked data. (a) noisy. (b) and (c) are results by 3D CWT. (d) and (e) are our reuslts.

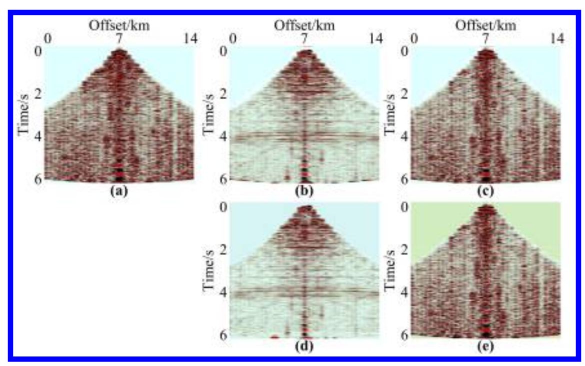

Figure 8: Results of common mid-point gathers. (a) noisy. (b) and (c) are results by 3D CWT. (d) and (e) are our reuslts.

ACKNOWLEDGMENTS

The first author acknowledges the China Scholarship Council. We acknowledge support from the Science Foundation of China (grants 41774135 and 41974131). We also thank Xinjiang Oilfield for providing the field data.

REFERENCES

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: https://doi.org/10.1093/gjihttps://doi.org//ggx332.

Deighan, A. J., and D. R. Watts, 1997, Ground-roll suppression using the wavelet transform: Geophysics, **62**, 1896–1903, doi: https://doi.org/10.1190/1.1444290.

Duan, W., J. Pei, F. Li, S. Jianhua, L. Zhengwen, and S. Wenjie, 2016, Acquisition footprints supersession with source and receiver line interpolation in the OVT domain: Oil Geophysical Prospecting, **51**, 40–48.

Embree, P., J. P. Burg, and M. M. Backus, 1963, Wide-band velocity filtering–the pie-slice process: Geophysics, **28**, 948–974, doi: https://doi.org/10.1190/1.1439310.

Gulunay, N., 1986, FXDECON and complex Wiener prediction filter: 56th Annual International Meeting, SEG, Expanded Abstracts, 279–281, doi: https://doi.org/10.1190/1.1893128.

Kreimer, N., and M. D. Sacchi, 2012, A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation: Geophysics, **77**, no. 3, V113–V122, doi: https://doi.org/10.1190/geo2011-0399.1.

Li, C., Y. Zhang, and C. C. Mosher, 2019, A hybrid learning based framework for seismic denoising: The Leading Edge, **38**, 542–549, doi: https://doi.org/10.1190/tle38070542.1.

Li, F., W. Duan, R. Zhao, L. Luo, and Q. Dang, 2015, Seismic signal to noise ratio improvement with volume tau-p transform in OVT domain: Oil Geophysical Prospecting, **50**, 418–423.

Ling, Y., and S. Hu, 2019, Internal multiple attenuation by OVT migration and its application in carbonate rock environment: 79th Annual International Conference and Exhibition, EAGE, Extended Abstracts, 1–5, doi: https://doi.org/10.3997/2214-4609.201901021.

Liu, D., W. Wang, X. Wang, C. Wang, J. Pei, and W. Chen, 2020, Poststack seismic data denoising based on 3-D convolutional neural network: IEEE Transactions on Geoscience and Remote Sensing, **58**, 1598–1629, doi: https://doi.org/10.1109/TGRS.2019.2947149.

Naghizadeh, M., and M. Sacchi, 2018, Ground-roll attenuation using curvelet downscaling: Geophysics, **83**, no. 3, V185–V195, doi: https://doi.org/10.1190/geo2017-0562.1.

Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, no. 3, V25–V32, doi: https://doi.org/10.1190/1.3552706.

Song, H., Y. Gao, W. Chen, Y.-J. Xue, H. Zhang, and X. Zhang, 2020, Seismic random noise suppression using deep convolutional autoencoder neural network: Journal of Applied Geophysics, **178**, 104071, doi: https://doi.org/10.1016/j.jappgeo.2020.104071.

Sun, J., S. Slang, T. Elboth, T. L. Greiner, S. McDonald, and L.-J. Gelius, 2020, Attenuation of marine seismic interference noise employing a customized U-Net: Geophysical Prospecting, **68**, 845–871, doi: https://doi.org/10.1111/1365-2478.12893.

Sun, M., Z. Li, Y. Qu, Z. Li, L. Li, J. Zhao, and B. Ning, 2019, A seismic denoising method based on curvelet transform with sparse constraint in OVT domain: Geophysical Prospecting for Petroleum, **58**, 208–218.

Treitel, S., J. L. Shanks, and C. W. Frasier, 1967, Some aspects of fan filtering: Geophysics, **32**, 789–800, doi: https://doi.org/10.1190/1.1439889.

Trickett, S. R., 2003, F-xy eigenimage noise suppression: Geophysics, **68**, 751–759, doi: https://doi.org/10.1190/1.1567245.

Vermeer, G. J. O., 2007, Reciprocal offset-vector tiles in various acquisition geometries: 77th Annual International Meeting, SEG, Expanded Abstracts, 61–65, doi: https://doi.org/10.1190/1.2792382.

Wang, X., W. Chen, D. Liu, and W. Xu, 2020, The deep prestack seismic data denoising based on high dimensional continuous wavelet transform: Annual Meeting of Chinese Geoscience Union.

Zhang, Y., X. Tian, X. Deng, and Y. Cao, 2010, Seismic denoising based on modified BP neural network: 6th International Conference on Natural Computation, 1825–1829.