

# Random noise attenuation using local signal-and-noise orthogonalization<sup>a</sup>

<sup>a</sup>Published in Geophysics, 80, no. 6, WD1-WD9, (2015)

*Yangkang Chen and Sergey Fomel*

## ABSTRACT

We propose a novel approach to attenuate random noise based on local signal-and-noise orthogonalization. In this approach, we first remove noise using one of the conventional denoising operators, and then apply a weighting operator to the initially denoised section in order to predict the signal-leakage energy and retrieve it from the initial noise section. The weighting operator is obtained by solving a least-squares minimization problem via shaping regularization with a smoothness constraint. Next, the initially denoised section and the retrieved signal are combined to form the final denoised section. The proposed denoising approach corresponds to orthogonalizing the initially denoised signal and noise in a local manner. We evaluate denoising performance by using local similarity. In order to test the orthogonalization property of the estimated signal and noise, we calculate the local similarity map between the denoised signal section and removed noise section. Low values of local similarity indicate a good orthogonalization and thus a good denoising performance. Synthetic and field data examples demonstrate the effectiveness of the proposed approach in applications to noise attenuation for both conventional and simultaneous-source seismic data.

## INTRODUCTION

Most random noise attenuation techniques necessitate a control of the trade-off between preservation of useful signal and removal of random noise.  $F$ - $x$  deconvolution and projection (Canales, 1984; Soubaras, 1995), also known as  $f$ - $x$  predictive filtering (Bekara and van der Baan, 2009; Chen and Ma, 2014), is a common technique for random noise attenuation thanks to its convenient implementation and high efficiency. Commonly a prediction filter length is modified in order to increase or decrease the ability of  $f$ - $x$  deconvolution to remove random noise, at the expense of decreasing or increasing the preservation of useful signal. When the data character is complex, the filter length should be long enough to capture certain amounts of useful signal, but at the same time to attenuate a limited amount of noise. Otherwise, a portion of useful signal can be lost because of a large number of dip components in the data (Chen and Ma, 2014). For certain specific types of random noise, such as blending noise in simultaneous-source seismic data (Beasley et al., 1998; Berkhout, 2008; Abma et al.,

2010; Chen et al., 2014a), median filtering (MF) can be particularly effective (Huo et al., 2012; Chen, 2014; Chen et al., 2014b). However, the ability of MF to remove spike-like blending noise strongly depends on the length of the filtering window. A longer filter leads to a more powerful attenuation for blending noise, but at the same time causes a possible loss of useful energy. In contrast, when the length of the filtering window is small, MF may eliminate less useful energy but also will be less able to remove spiky noise. Removing spiky noise without losing too much useful signal becomes a difficult trade-off.

Suppose that the seismic data  $\mathbf{d}$  is composed of signal  $\mathbf{s}_{true}$  and noise  $\mathbf{n}_{true}$ , a successful random noise attenuation will give accurate signal and noise estimation:  $\mathbf{s}_0 \approx \mathbf{s}_{true}$  and  $\mathbf{n}_0 \approx \mathbf{n}_{true}$ , where  $\mathbf{s}_0$  and  $\mathbf{n}_0$  are the estimated signal and noise. However, the ideal situation may not occur in practice for two main possible reasons: incorrect parameter selection or inadequacy of denoising assumptions. When those assumptions are not met, the conventional denoising approaches may not achieve the optimal performance. For most noise attenuation approaches, the noise section will contain a certain amount of useful signal, which can be called *signal-leakage energy*. Sometimes, the leakage energy is negligible, whereas in other cases, this loss of energy may result in a decrease in resolution.

Local seismic attributes measure seismic signal characteristics neither instantaneously, at each signal point, nor globally, across a data window, but locally in the neighborhood of each point (Fomel, 2007a). One of the most useful local attributes is *local similarity*, which has found numerous successful applications in different areas of seismic data processing: multicomponent image registration (Fomel et al., 2005; Fomel, 2007a), time-lapse registration (Fomel and Jin, 2009; Zhang et al., 2013), time-frequency analysis (Liu et al., 2011b), structure-enhancing filtering (Liu et al., 2010), phase estimation (Fomel and van der Baan, 2014), etc. By using a weighted stacking of seismic data according to local similarity to the reference trace, a seismic image with an increased signal-to-noise ratio can be obtained (Liu et al., 2009, 2011a). In this paper, we introduce a new local seismic attribute: *local orthogonalization weight* (LOW) in order to perform local orthogonalization. LOW appeared previously in the definition of local similarity and can be obtained by solving a minimization problem using shaping regularization with a smoothness constraint.

In order to compensate for the loss of useful signal in traditional noise attenuation approaches, because of incorrect parameter selection or inadequacy of denoising assumptions, we employ a weighting operator on the initially denoised section for the retrieval of useful signal from the initial noise section. The new denoising process corresponds to local orthogonalization of signal and noise based on the assumption that final estimated signal and noise should be orthogonal to each other in the time-space domain. The orthogonality assumption is similar as assuming that the signal and noise do not correlate with each other. Thus, the orthogonality assumption is assumed to be valid for all kinds of noise that do not correlate with the useful signal, e.g., random noise. The proposed local-orthogonalization approach can be considered as a specific case of previously proposed nonstationary matching filtering (Fomel,

2009) for one-point filter length. The proposed approach is not very sensitive to the parameter selection, and thus is robust in practice. We use two synthetic examples and three field data examples to demonstrate successful performance of the proposed approach in applications to both conventional and simultaneous-source seismic data.

## METHOD

### Compensating for the signal-leakage energy by weighting

For many random noise attenuation approaches, the leakage energy is not negligible. We can attempt to retrieve the leaking signal from the noise section by applying a simple nonstationary weighting operator to the initially denoised signal assuming that the leakage energy can be predicted by weighting the useful signal:

$$\mathbf{s}_1 = \mathbf{w} \circ \mathbf{s}_0. \quad (1)$$

Here  $\mathbf{s}_1$  is the retrieved signal,  $\mathbf{w} \circ \mathbf{s}_0 = \text{diag}(\mathbf{w})\mathbf{s}_0 = \text{diag}(\mathbf{s}_0)\mathbf{w}$ , which denotes Hadamard (or Schur) product, and  $\text{diag}(\cdot)$  denotes the diagonal matrix composed of an input vector. The weighting vector  $\mathbf{w}$  can be estimated by solving the following minimization problem:

$$\min_{\mathbf{w}} \left\| \overbrace{\mathbf{d} - \mathbf{P}[\mathbf{d}]}^{\mathbf{n}_0} - \mathbf{w} \circ \overbrace{\mathbf{P}[\mathbf{d}]}^{\mathbf{s}_0} \right\|_2^2, \quad (2)$$

where  $\mathbf{d}$  denotes the observed noisy data, and  $\mathbf{P}$  denotes the initial random noise attenuation operator. Equation 2 uses a weighted (scaled) signal  $\mathbf{s}_0$  to match the leakage energy in the initial noise section ( $\mathbf{n}_0$ ) in a least-squares sense. In the next section, we will introduce an approach to calculate the weighting vector  $\mathbf{w}$  using local orthogonalization.

### Local orthogonalization

Based on the assumption that final estimated signal  $\mathbf{s}$  and noise  $\mathbf{n}$  should be orthogonal to each other, we can orthogonalize the estimated signal and estimated noise by

$$\hat{\mathbf{n}} = \mathbf{n}_0 - w\mathbf{s}_0, \quad (3)$$

$$\hat{\mathbf{s}} = \mathbf{s}_0 + w\mathbf{s}_0, \quad (4)$$

where  $w$  is the global orthogonalization weight (GOW) defined by

$$w = \frac{\mathbf{n}_0^T \mathbf{s}_0}{\mathbf{s}_0^T \mathbf{s}_0}. \quad (5)$$

Here  $[\cdot]^T$  denotes transpose. Appendix A provides a demonstration and proof for the global orthogonalization process as denoted by equations 3, 4 and 5. The orthogonality assumption is similar as assuming that the signal and noise do not correlate

with each other, which implies the kinds of noise that do not correlate with the useful signal, e.g., random noise. The orthogonality assumption is assumed to be valid in the original time-space domain, but also has the possibility of being applied in other transformed domains. Instead of using GOW, we propose to use nonstationary local orthogonalization weight (LOW). One possible definition of LOW is:

$$w_m(t) = \frac{\sum_{i=t-m/2}^{t+m/2} s_0(i)n_0(i)}{\sum_{i=t-m/2}^{t+m/2} s_0^2(i)}, \quad (6)$$

where  $w_m(t)$  denotes the LOW for each temporal point  $t$  with a local window length  $m$ .  $s_0(t)$  and  $n_0(t)$  here denotes the initially estimated signal and noise for each point  $t$ .

## Solving LOW by shaping regularization

In order to better control the locality and smoothness of LOW, we follow the local-attribute scheme introduced by Fomel (2007a). Simplifying equation 2, and adding a local-smooth constraint to the optimization problem, we obtain:

$$\mathbf{w} = \arg \min_{\mathbf{w}} \|\mathbf{n}_0 - \mathbf{S}_0 \mathbf{w}\|_2^2 + \mathbf{R}(\mathbf{w}). \quad (7)$$

Here,  $\mathbf{w}$  is the LOW,  $\mathbf{S}_0$  is a diagonal matrix composed of the initial estimated signal  $\mathbf{s}_0$ :  $\mathbf{S}_0 = \text{diag}(\mathbf{s}_0)$ .  $\mathbf{R}$  denotes a smoothing regularization operator. Then we solve the least-squares problem 7 with the help of shaping regularization (Fomel, 2007b) using a local-smoothness constraint:

$$\mathbf{w} = [\lambda^2 \mathbf{I} + \mathcal{T}(\mathbf{S}_0^T \mathbf{S}_0 - \lambda^2 \mathbf{I})]^{-1} \mathcal{T} \mathbf{S}_0^T \mathbf{n}_0, \quad (8)$$

where  $\mathcal{T}$  is a triangle smoothing operator and  $\lambda$  is a scaling parameter set as  $\lambda = \|\mathbf{S}_0^T \mathbf{S}_0\|_2$ . Inserting  $\mathbf{w}$  into equation 1, we can obtain a new denoising approach:

$$\mathbf{s} = \mathbf{s}_0 + \mathbf{s}_1 = \mathbf{P}[\mathbf{d}] + \mathbf{w} \circ \mathbf{P}[\mathbf{d}] = (\mathbf{I} + \text{diag}(\mathbf{w}))\mathbf{P}[\mathbf{d}], \quad (9)$$

$$\mathbf{n} = \mathbf{n}_0 - \mathbf{s}_1 = \mathbf{d} - \mathbf{P}[\mathbf{d}] - \mathbf{w} \circ \mathbf{P}[\mathbf{d}] = \mathbf{d} - (\mathbf{I} + \text{diag}(\mathbf{w}))\mathbf{P}[\mathbf{d}]. \quad (10)$$

Here,  $\mathbf{s}$  and  $\mathbf{n}$  are the final estimated signal and noise.

The controlling parameter for calculating LOW is the width of the triangle smoother, or the smoothing radius (SR). Figure 1 shows the calculated LOWs that correspond to the second synthetic example (Figure 3) with different smoothing radii. The LOW are calculated by solving the equation 7 with the initial denoised signal and initial removed noise set as Figures 3d and 3e, respectively, using the approach shown in equation 8. As we can see from the comparison, LOW is robust with respect to

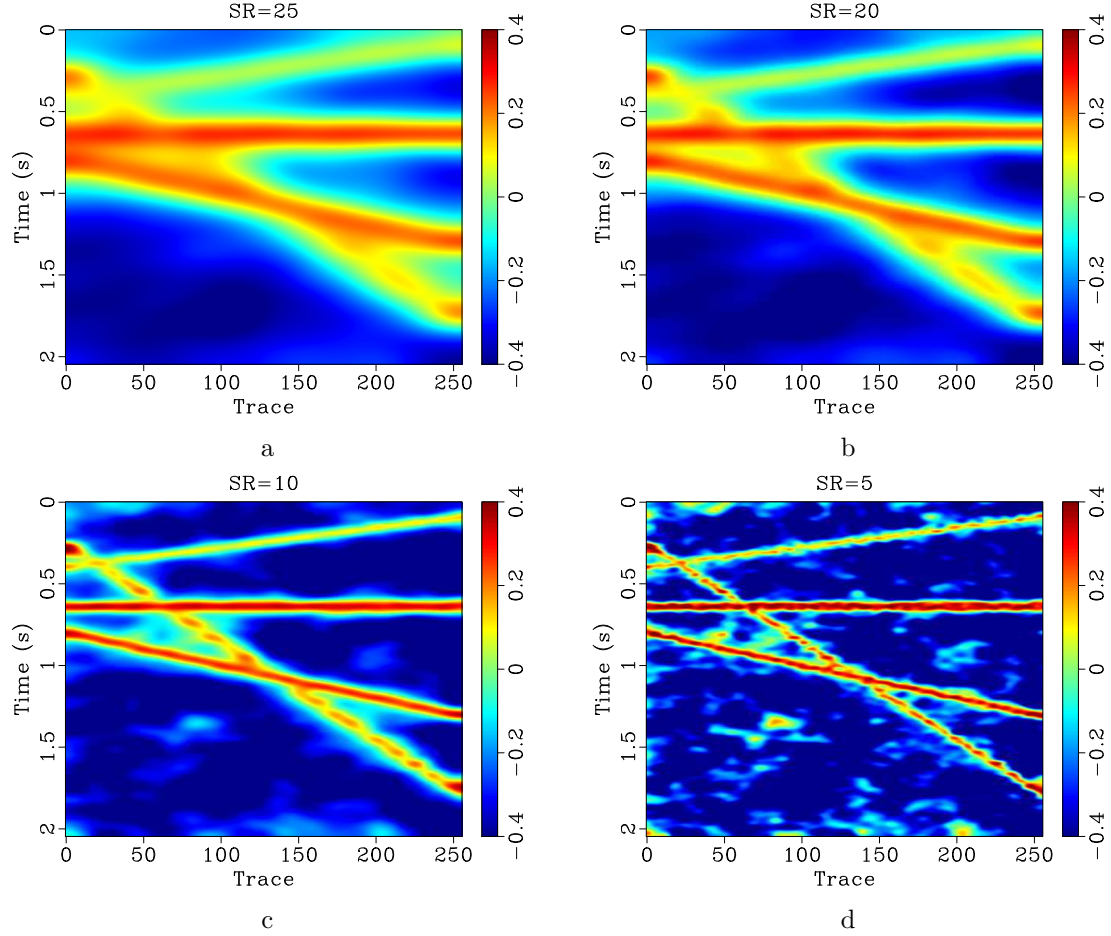


Figure 1: LOWs using different smoothing radii for the example as shown in Figure 3. (a) Vertical and lateral smoothing radii are 25 samples. (b) Vertical and lateral smoothing radii are 20 samples. (c) Vertical and lateral smoothing radii are 10 samples. (d) Vertical and lateral smoothing radii are 5 samples.

the choice of the smoothing radius (e.g., from 25 samples to 20 samples, or from 10 samples to 5 samples), but may differ a lot for significantly different smoothing radii (e.g., from 25 samples to 5 samples). We can also conclude from the comparison that as the smoothing radius increases, the temporal and spatial resolutions decrease, but the anti-noise ability increases. Note that the orthogonalization equations 2 and 7 can also be understood as one-point nonstationary matching filtering (Fomel, 2009).

## SYNTHETIC EXAMPLES

We use two synthetic examples to test the performance of the proposed denoising algorithm. For consistency, we use *denoising* to denote random noise attenuation for regular seismic data, and *deblending* to denote blending noise attenuation for simultaneous-source seismic data. The first example is a deblending test, similar to the one used by Huo et al. (2012). The second example is a denoising test that contains four crossing linear events. In order to test the improvement in the signal-to-noise ratio of these two synthetic examples, we utilize the following measurement:

$$SNR = 10 \log_{10} \frac{\|\mathbf{s}_{true}\|_2^2}{\|\mathbf{s}_{true} - \mathbf{s}\|_2^2}. \quad (11)$$

In addition to SNR, we also propose to use local similarity (Fomel, 2007a) as a convenient measure to evaluate denoising performance. Appendix B gives a brief review of local similarity. After the local similarity map between the denoised data and removed noise is calculated, we can judge from the local similarity if there is leakage energy in the noise. When the clean data is unknown for field data denoising tests, the SNR based evaluation becomes unavailable. Besides, the SNR is not always the best measurement for denoising performance because it does not measure the signal leakage. However, the local similarity map can always be used to evaluate the denoising performance.

Figure 2 shows the deblending performance using a conventional MF and the proposed denoising approach. The original clean, unblended data are shown in Figure 2a. After blending using a simple random-dithering method (Chen et al., 2014a), we obtained the noisy data containing spike-like blending noise (2c). Figure 2d demonstrates a denoised section after using MF. The noise section corresponding to MF is shown in Figure 2e and contains a certain amount of signal leakage in the form of linear events. Using the proposed denoising approach, we obtained a denoised section with stronger-amplitude linear events. In the corresponding noise section, any coherent linear events are barely noticeable, suggesting a nearly perfect deblending. In this example, because the noise are spike-like blending noise, we did smooth too much during shaping regularization. In this example, the vertical and lateral smoothing radii are both 2 samples. Figures 2f and 2i show local similarity of the noise section to the denoised section before and after using the proposed approach. After applying the proposed method, the noise section exhibits low similarity with the denoised section. Although structure-oriented median filtering can handle the dipping reflector, they

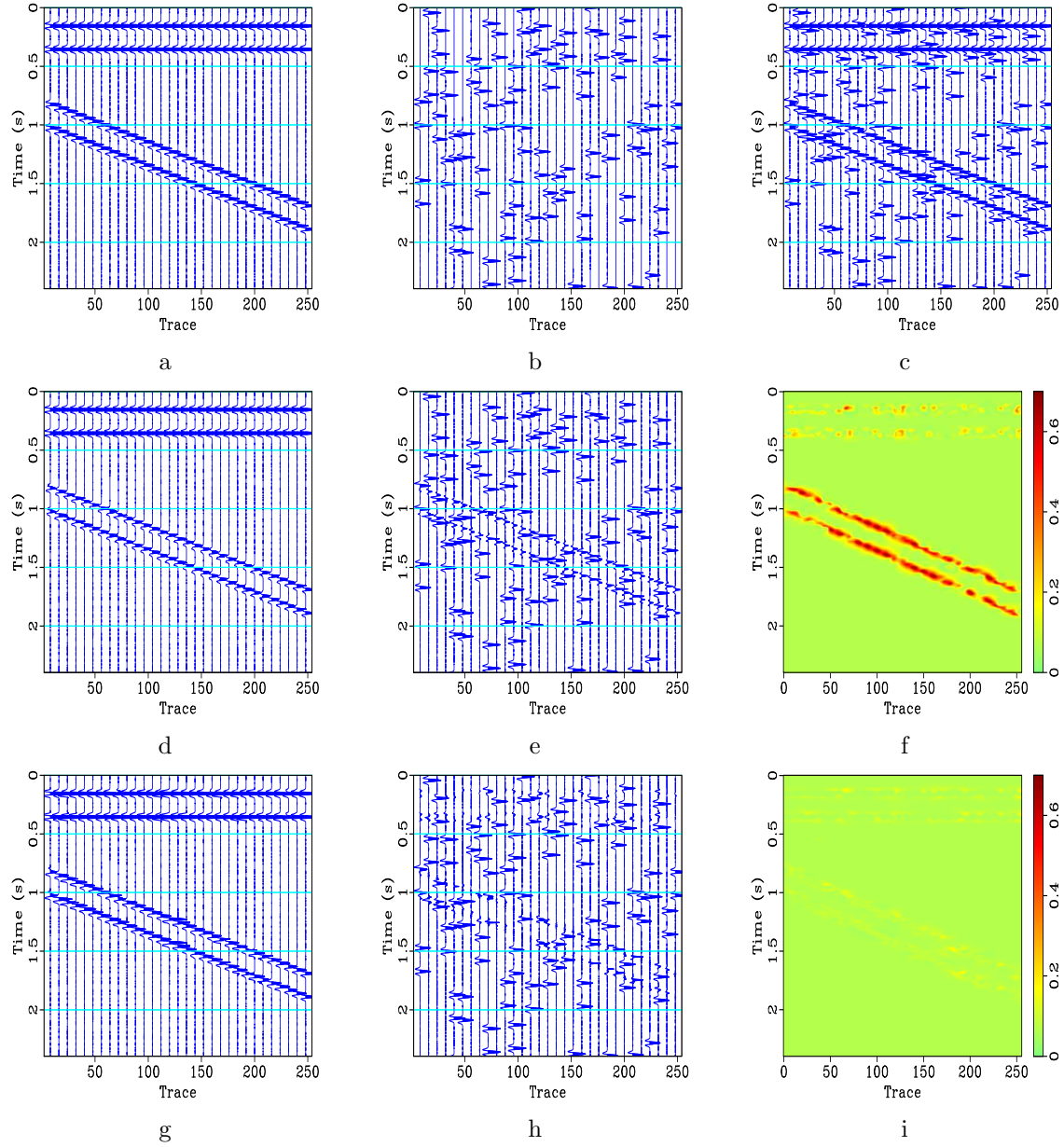


Figure 2: Comparison of using the local-orthogonalization-based random noise attenuation approach for deblending test before and after. (a) Unblended section. (b) Blending noise section. (c) Blended section. (d) Deblended result using MF. (e) Noise section using MF. (f) Local similarity map between (d) and (e). (g) Deblended result using the proposed method. (h) Noise section using the proposed method. (i) Local similarity map between (g) and (h).

can only work well for relative simpler structures where the dip estimation can be precisely obtained, in which case there will be no need to use the proposed approach. However, when the dip estimation is not accurate, or there are conflicting dips in a processing window, the structure-oriented median filtering may not work well, in which case we can use the proposed approach for retrieving the leaked useful energy.

Figure 3 shows the denoising performance based on the conventional  $f$ - $x$  deconvolution (Canales, 1984). The clean data is shown in Figure 3a and the noise section, corresponding to  $f$ - $x$  deconvolution, is shown in Figure 3f; it contains some horizontal and low-dip-angle signals. The coherent leakage energy is also visible in the local similarity map (Figure 3f). Applying the proposed denoising approach, we obtained a denoised section with stronger-amplitude horizontal events. The corresponding noise section does not contain any coherent signal. In this example, the vertical and lateral smoothing radii are both 25 samples. Figures 3f and 3i denote local similarity of the noise section to the denoised section before and after using the proposed approach. The similarity decreases significantly after applying local orthogonalization.

The SNR comparison before and after using local-orthogonalization approach is summarized in Table 1.

Test	Deblending	denoising
Original (dB)	1.173	-1.72
Initially denoised (dB)	17.65	21.21
Orthogonalized (dB)	20.95	25.30

Table 1: Comparison of SNR before and after using the proposed approach.

## FIELD EXAMPLES

In this section, we use three field data sets to evaluate the proposed method. The first example is a simulated deblending test and the latter two examples are common denoising tests.

Figure 4a shows the unblended data for the first test, which is a deep-water common receiver section containing events that are nearly flat. After numerical blending using the same blending approach as in the first synthetic example, the blended data (shown in Figure 4c) demonstrate strong interference. Figure 4d shows the deblended data after the use of  $f$ - $x$  deconvolution. When checking the noise section in Figure 4e, we found many coherent signals, suggesting a significant leakage of useful signal energy. The denoised section after using the proposed approach is shown in Figure 4g. The noise section no longer contains coherent useful signals while the removed blending noise appears nearly unchanged. In this example, we set the vertical and lateral smoothing radii to 25 samples. The similarity maps before and after application of the proposed method are shown in Figure 4f and 4h. The SNR using the previous



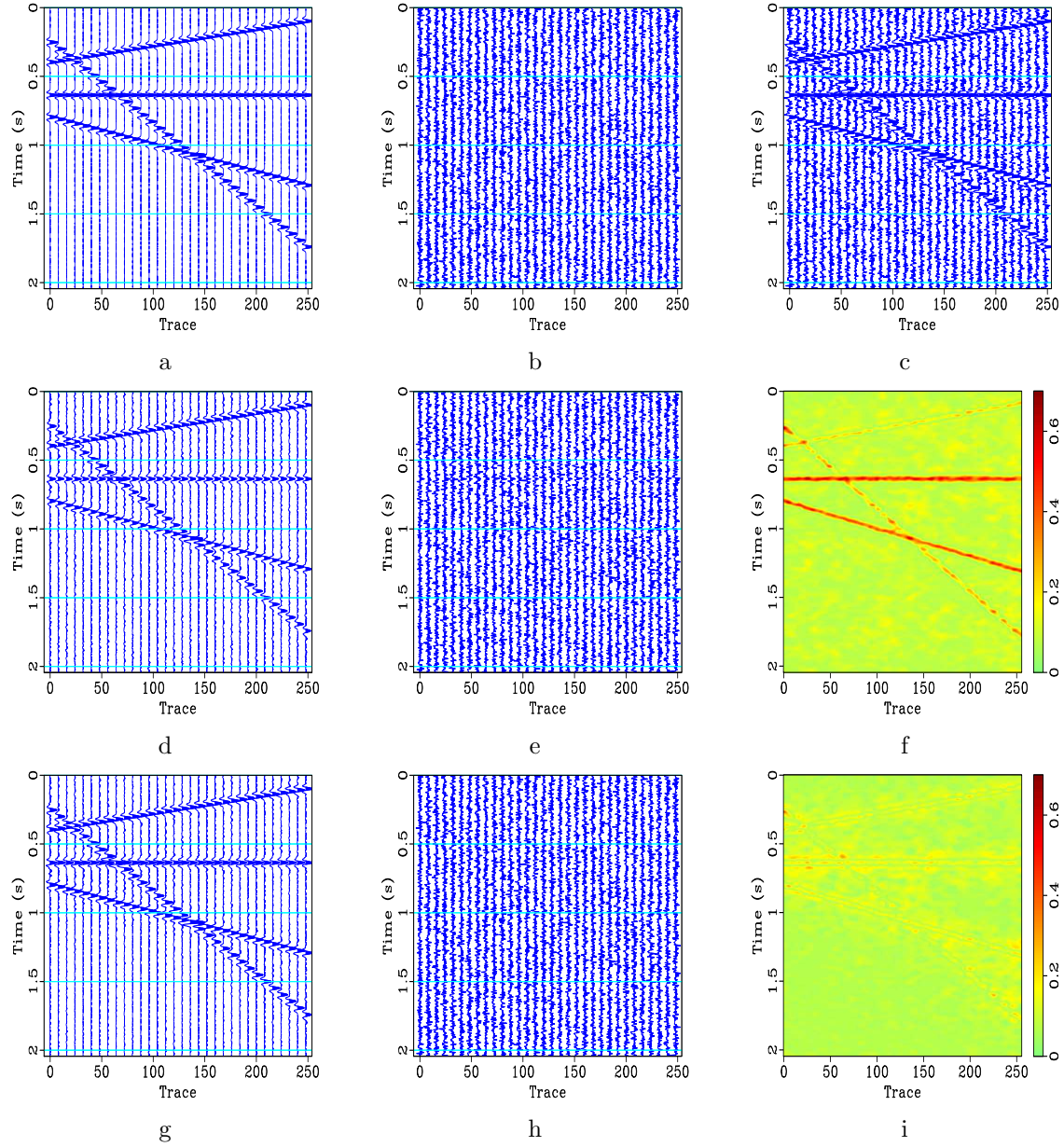


Figure 3: Comparison of using the local-orthogonalization-based random noise attenuation approach for denoising test before and after. (a) Clean section. (b) Random noise section. (c) Noisy section. (d) Denoised result using  $f$ - $x$  deconvolution. (e) Noise section using  $f$ - $x$  deconvolution. (f) Local similarity map between (d) and (e). (g) Denoised result using the proposed method. (h) Noise section using the proposed method. (i) Local similarity map between (g) and (h).

criteria of the blended data is 0.04. After the initial deblending using  $f$ - $x$  deconvolution, SNR improves to 7.44. Further, after the proposed retrieving-signal process, SNR improves to 10.08. It is worth to be mentioned that the preservation of useful signal is of significant importance in deblending, and thus the proposed approach is particularly attractive in its application to deblending.

The second field data test is for a land seismic section, which was used previously by Fomel (2007a) and Liu and Fomel (2012). The input image is shown in Figure 5. The shallow part of the section is noisy, which may make interpretation difficult. After a conventional  $f$ - $x$  deconvolution, a much cleaner section is obtained (Figure 6a). Some coherent signal is left in the noise section shown in Figure 6d, especially around 0.75s and the 350th trace. After local signal-and-noise orthogonalization, we obtained a clean denoised section but with strong amplitude in some regions and almost no loss of useful component in the noise section (Figure 6d). In this example, the vertical and lateral smoothing radii are both 2 samples. Unlike the synthetic data test and the previous field data test, for this field data test, the clean signal is unknown, so the SNR measurement cannot be applied in this case. The two corresponding similarity maps are shown in Figure 7. The local similarity after applying the proposed approach is nearly zero along the whole seismic section.

The third field data example is borrowed from Liu et al. (2012). The noisy data is shown in Figure 8. The initial signal and noise models are prepared using  $f - x$  regularized nonstationary autoregression (RNAR). Following the approach of Liu et al. (2012), the initially denoised data are shown in Figure 9a. The corresponding noise section is shown in Figure 9c. The general denoising approach shows good performance considering the clean denoised image and random removed noise. However, we can find small useful signals from the noise section, as indicated by the frame boxes in Figure 9c. The denoised image using the proposed approach is shown in Figure 9b. After using local orthogonalization to retrieve useful signal, the coherent signal disappears in the final noise section (Figure 9d), as indicated by the frame boxes. The zoomed sections corresponding to frame boxes A and B in Figure 9 are shown in Figure 10. In this example, we chose the vertical and lateral smoothing radius as 5 samples. Before using the proposed approach, the local similarity map between the denoised data and removed noise (Figure 11a) showed some high-similarity anomalies. After applying the proposed approach, the similarity map is nearly zero everywhere across the whole section.

## CONCLUSIONS

Conventional methods for noise attenuation may cause some leakage of signal energy in the noise section as a result of incorrect parameter selection or inadequacy of denoising assumptions. We have shown that it is possible to retrieve the leakage energy by applying a weighting operator to the initial signal and adding the retrieved signal energy to the initially denoised data to obtain the final denoised data. In order to design an optimized weighting operator, we introduce a new local attribute,

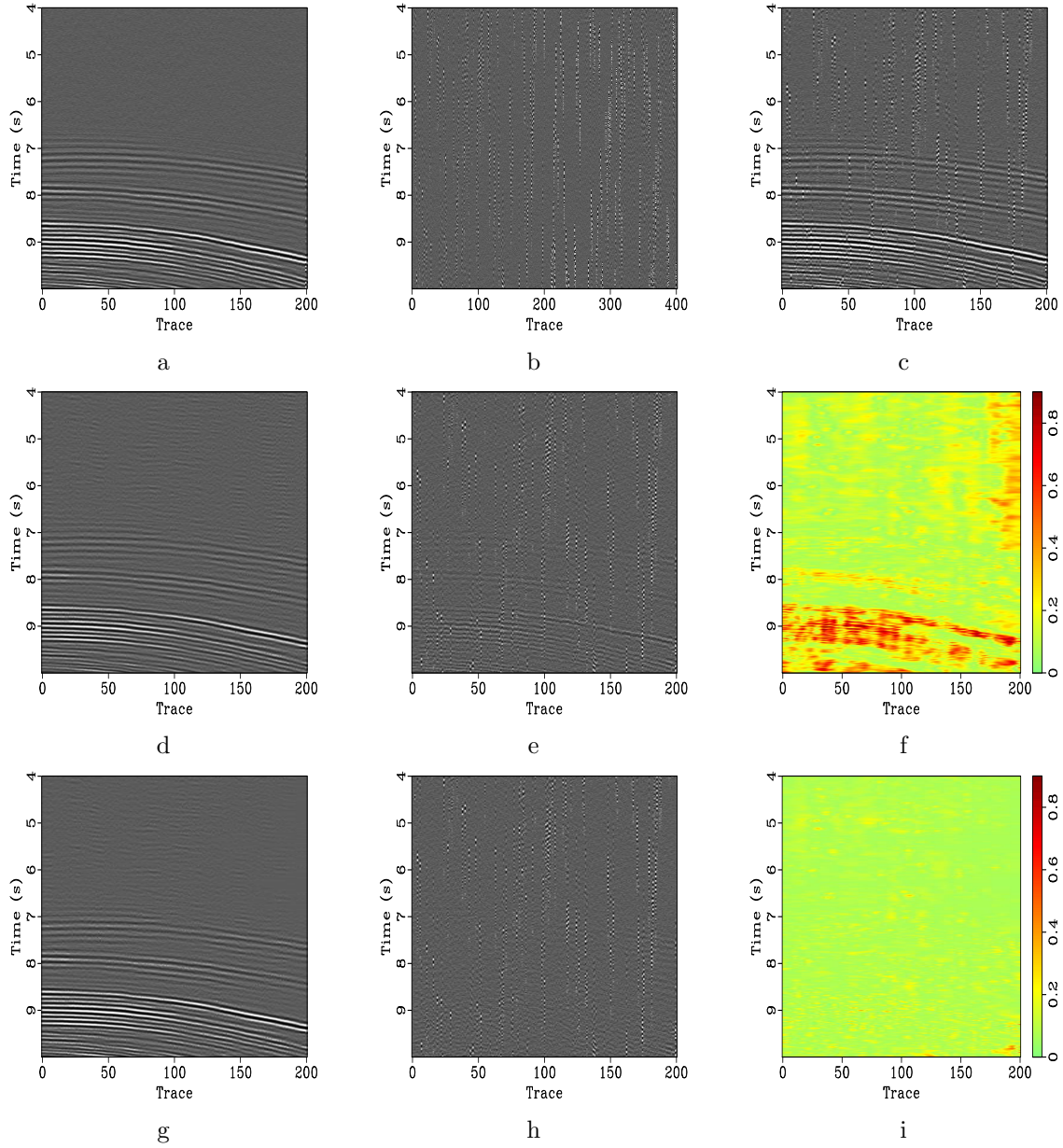


Figure 4: Comparison of using the local-orthogonalization-based random noise attenuation approach for deblending test before and after. (a) Unblended section (not clean, also containing random noise). (b) Blending noise section. (c) Blended section. (d) Deblended result using  $f - x$  deconvolution. (e) Noise section using  $f - x$  deconvolution. (f) Local similarity map between (d) and (e). (g) Deblended result using the proposed method. (h) Noise section using the proposed method. (i) Local similarity map between (g) and (h).

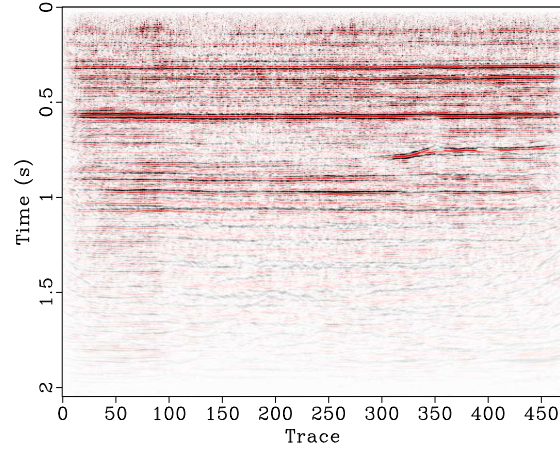
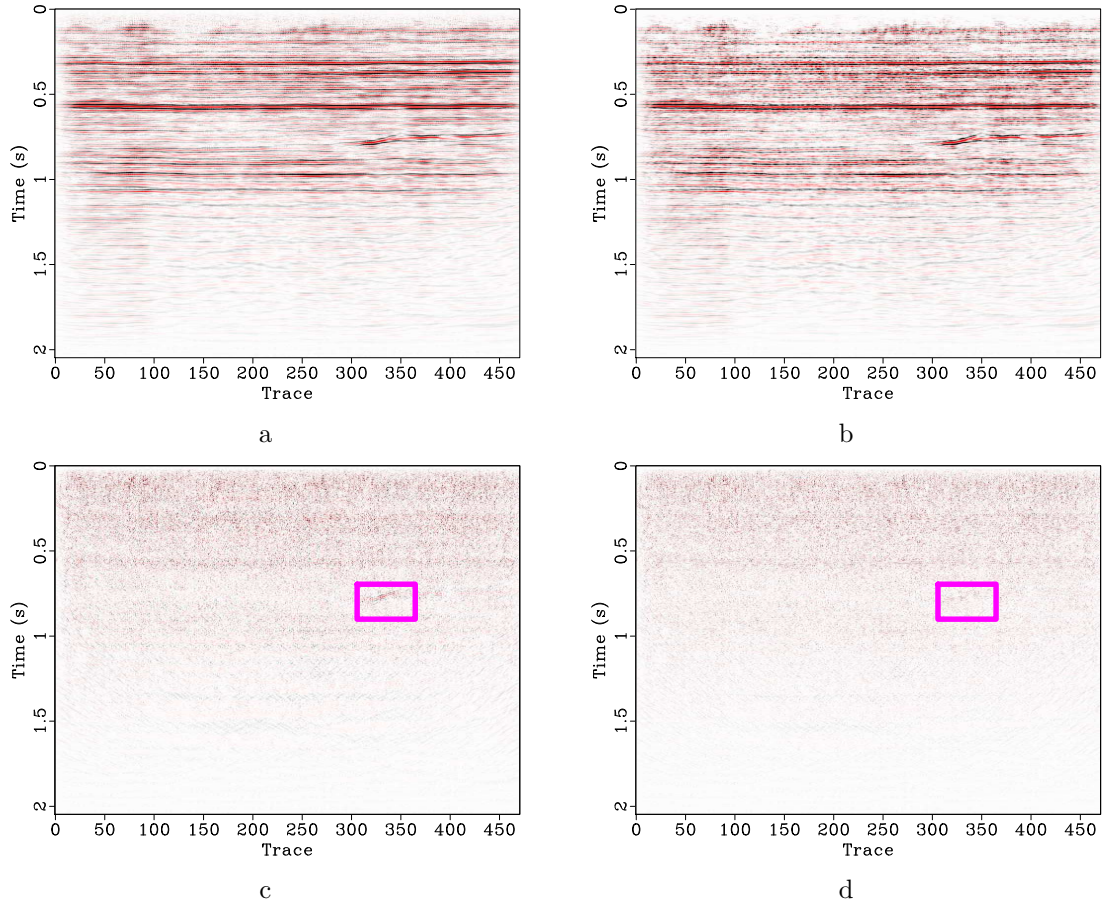


Figure 5: Noisy post-stack land field data (Fomel, 2013).

Figure 6: (a) Denoised section using  $f-x$  deconvolution. (b) Denoised section using the proposed approach. (c) Noise section using  $f-x$  deconvolution. (d) Noise section using the proposed approach.



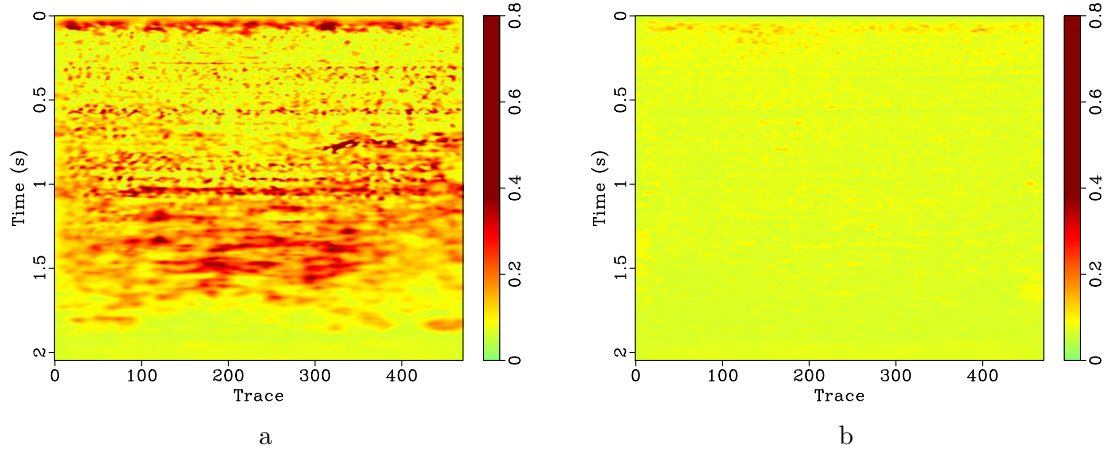


Figure 7: Comparison of similarity between noise section and denoised section with and without using the proposed random noise attenuation approach. (a) Similarity map using  $f - x$  deconvolution. (b) Similarity map using the proposed method.

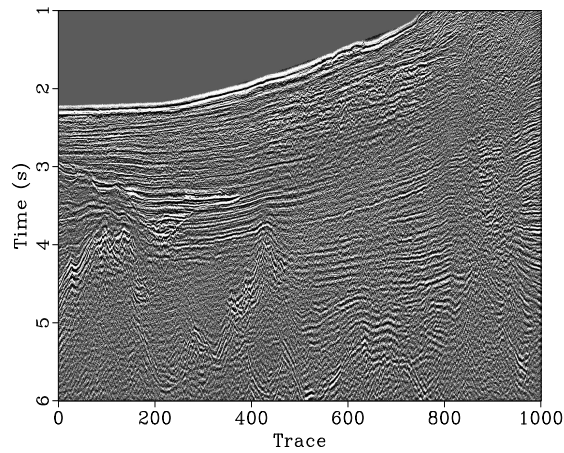


Figure 8: Noisy post-stack field data (Liu et al., 2012).

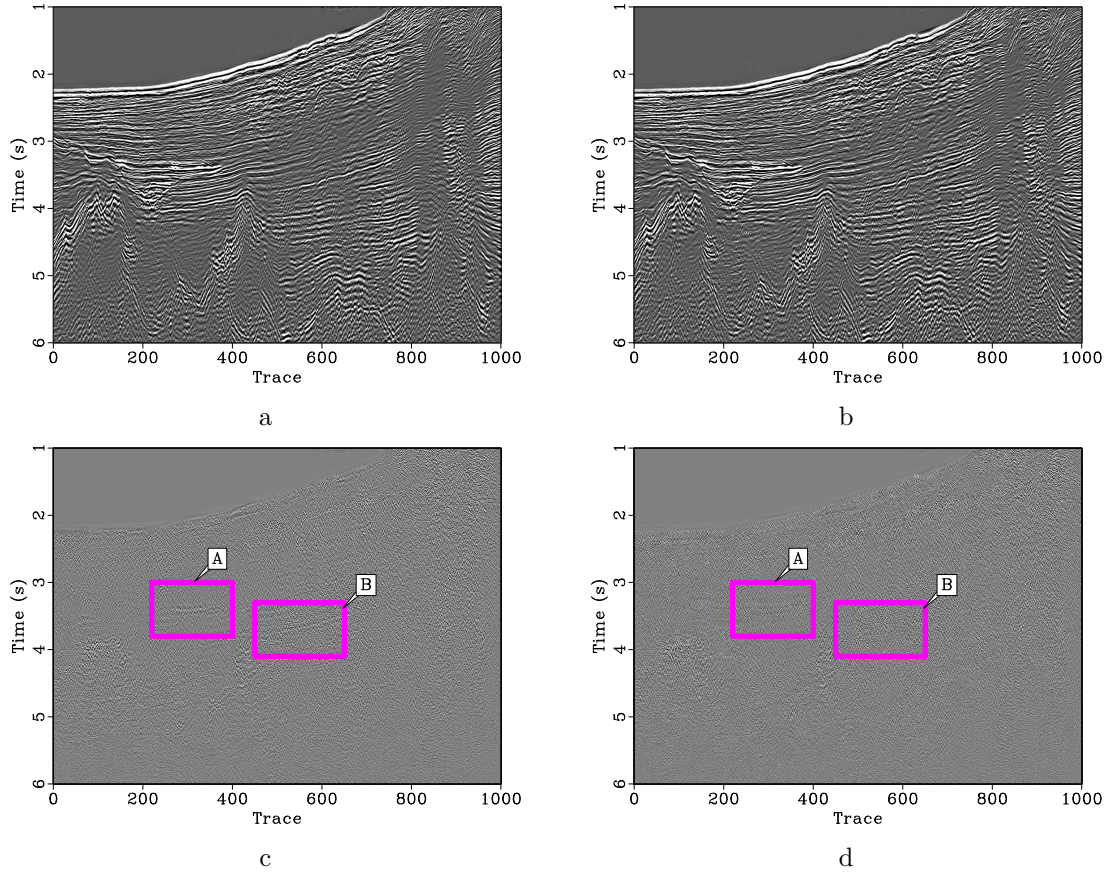


Figure 9: (a) Denoised section using  $f-x$  RNAR. (b) Denoised section using the proposed approach. (c) Noise section using  $f-x$  RNAR. (d) Noise section using the proposed approach.



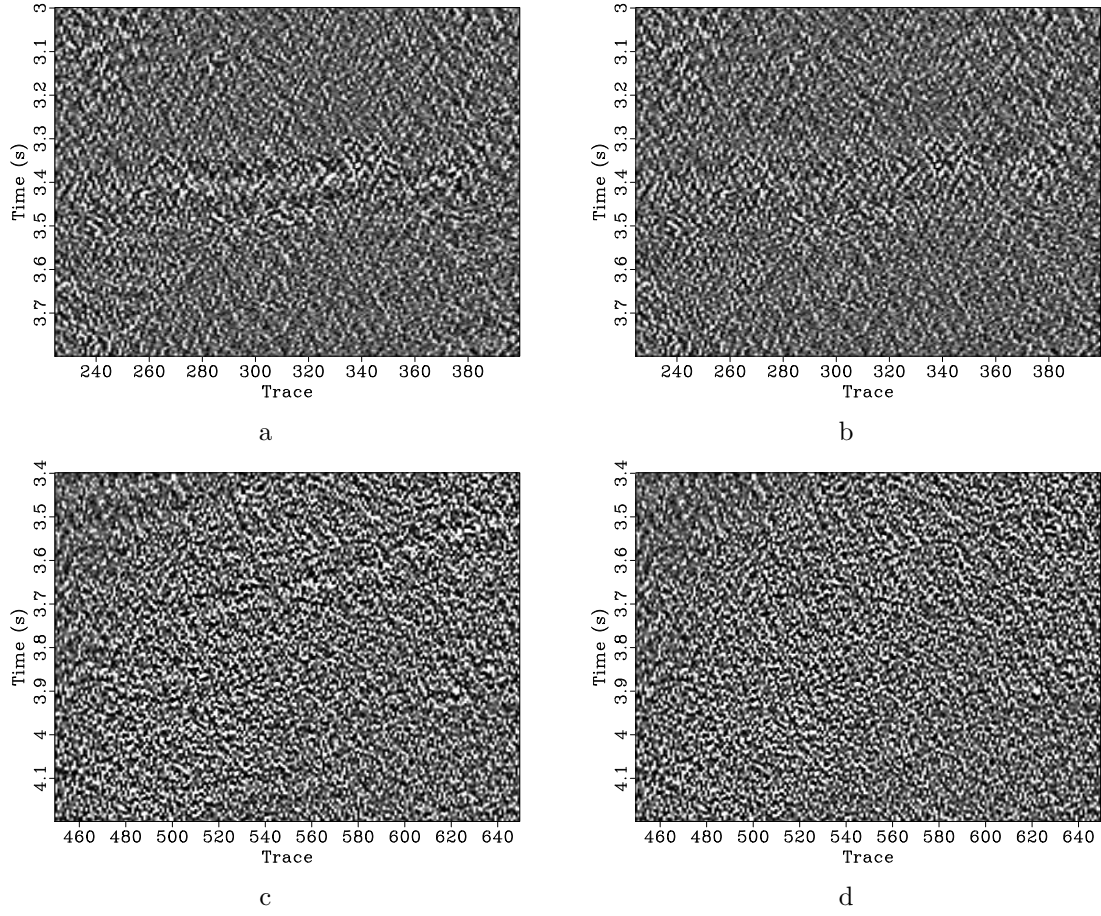


Figure 10: (a) Zoomed section using  $f$ - $x$  RNAR (frame box A shown in Figure 9). (b) Zoomed section using the proposed approach (frame box A shown in Figure 9). (c) Zoomed section using  $f$ - $x$  RNAR (frame box B shown in Figure 9). (d) Zoomed section using the proposed approach (frame box B shown in Figure 9).

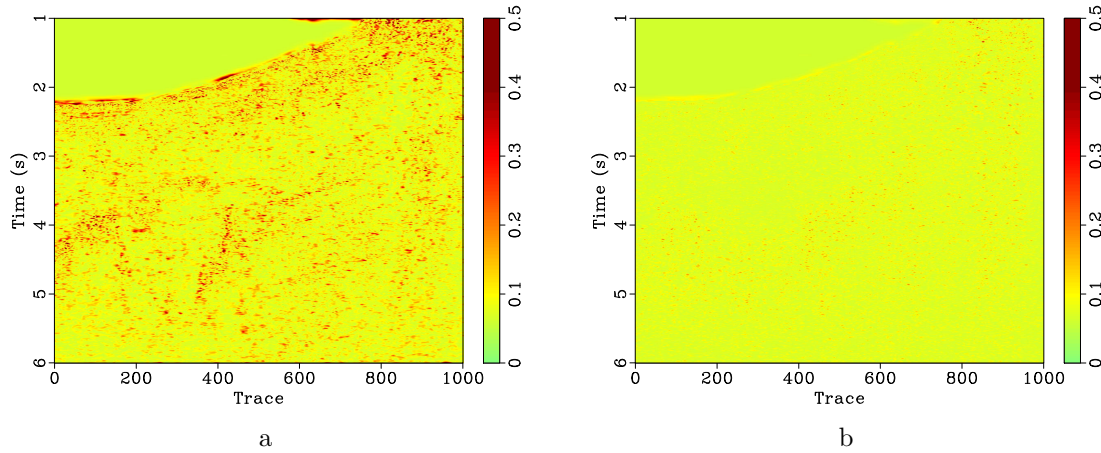


Figure 11: Comparison of similarity between noise section and denoised section with and without using the proposed random noise attenuation approach. (a) Similarity map using  $f$ - $x$  RNAR. (b) Similarity map using the proposed method.

called local orthogonalization weight (LOW). LOW can be obtained by solving a least-square minimization problem using shaping regularization with a smoothness constraint. The proposed denoising approach corresponds to locally orthogonalizing signal and noise. Once the initial signal and noise models are given, the proposed approach can help retrieve the residual signal in the noise section. The local orthogonalization approach to random noise attenuation is applicable to processing of blended simultaneous-source seismic data, in which the preservation of useful signal is particularly important. Although the examples were 2-D, the method applies equally well in 3-D and in any other number of dimensions.

## ACKNOWLEDGMENTS

We thank FairfieldNodal for the data used in the first field data example. We also thank Luke Decker for helpful discussions, Bill Harlan and two anonymous reviewers for constructive suggestions that made the manuscript better. The paper was prepared in the Madagascar open-source software environment (Fomel et al., 2013).

## APPENDIX A: SIGNAL-AND-NOISE ORTHOGONALIZATION

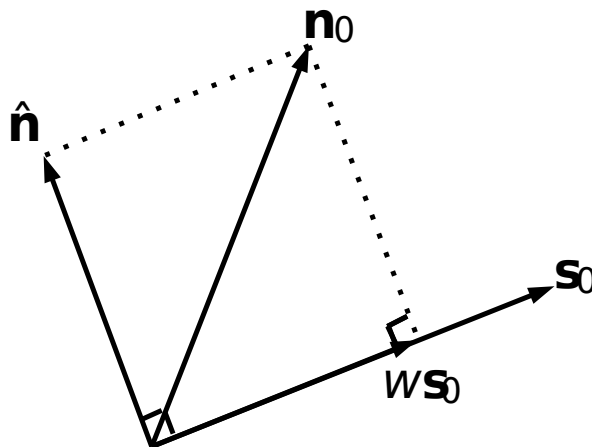


Figure A-1: Demonstration of signal-and-noise orthogonalization.

As shown schematically in Figure A-1, the initially estimated signal and noise are denoted by  $s_0$  and  $n_0$ , respectively. By projecting  $n_0$  to the direction of  $s_0$ , we can



get the projection  $w\mathbf{s}_0$ . The other component of  $\mathbf{n}_0$  is the final estimated noise, as shown in equation 3. The final estimated signal is thus the summation of the initially estimated signal  $\mathbf{s}_0$  and the projection component  $w\mathbf{s}_0$ .

When  $w = \frac{\mathbf{n}_0^T \mathbf{s}_0}{\mathbf{s}_0^T \mathbf{s}_0}$ , the following equation holds:

$$\begin{aligned} \hat{\mathbf{n}}^T \hat{\mathbf{s}} &= (\mathbf{n}_0 - w\mathbf{s}_0)^T (\mathbf{s}_0 + w\mathbf{s}_0) = \mathbf{n}_0^T \mathbf{s}_0 - w\mathbf{s}_0^T \mathbf{s}_0 + w\mathbf{n}_0^T \mathbf{s}_0 - w^2 \mathbf{s}_0^T \mathbf{s}_0 \\ &= \mathbf{n}_0^T \mathbf{s}_0 - \frac{\mathbf{n}_0^T \mathbf{s}_0}{\mathbf{s}_0^T \mathbf{s}_0} \mathbf{s}_0^T \mathbf{s}_0 + \frac{\mathbf{n}_0^T \mathbf{s}_0}{\mathbf{s}_0^T \mathbf{s}_0} \mathbf{n}_0^T \mathbf{s}_0 - \left( \frac{\mathbf{n}_0^T \mathbf{s}_0}{\mathbf{s}_0^T \mathbf{s}_0} \right)^2 \mathbf{s}_0^T \mathbf{s}_0 = 0 \end{aligned} \quad (\text{A-1})$$

Here,  $\hat{\mathbf{s}}$  and  $\hat{\mathbf{n}}$  denote the final estimated signal and noise, respectively, and appear orthogonal to each other. This orthogonalization approach is also known as Gram-Schmidt orthogonalization (Hazewinkel, 2001). Note that  $w$  defined above can be obtained by solving the least-squares optimization problem:

$$\min_w \|w\mathbf{s}_0 - \mathbf{n}_0\|_2^2, \quad (\text{A-2})$$

which we extend to equation 7 in the main text.

## APPENDIX B: REVIEW OF LOCAL SIMILARITY

Fomel (2007a) defined local similarity between vectors  $\mathbf{a}$  and  $\mathbf{b}$  as:

$$\mathbf{c} = \sqrt{\mathbf{c}_1^T \mathbf{c}_2} \quad (\text{B-1})$$

where  $\mathbf{c}_1$  and  $\mathbf{c}_2$  come from two least-squares minimization problems:

$$\mathbf{c}_1 = \arg \min_{\mathbf{c}_1} \|\mathbf{a} - \mathbf{B}\mathbf{c}_1\|_2^2, \quad (\text{B-2})$$

$$\mathbf{c}_2 = \arg \min_{\mathbf{c}_2} \|\mathbf{b} - \mathbf{A}\mathbf{c}_2\|_2^2, \quad (\text{B-3})$$

where  $\mathbf{A}$  is a diagonal operator composed from the elements of  $\mathbf{a}$ :  $\mathbf{A} = \text{diag}(\mathbf{a})$  and  $\mathbf{B}$  is a diagonal operator composed from the elements of  $\mathbf{b}$ :  $\mathbf{B} = \text{diag}(\mathbf{b})$ . Least-squares problems B-2 and B-3 can be solved with the help of shaping regularization with a smoothness constraint:

$$\mathbf{c}_1 = [\lambda_1^2 \mathbf{I} + \mathcal{T}(\mathbf{B}^T \mathbf{B} - \lambda_1^2 \mathbf{I})]^{-1} \mathcal{T} \mathbf{B}^T \mathbf{a}, \quad (\text{B-4})$$

$$\mathbf{c}_2 = [\lambda_2^2 \mathbf{I} + \mathcal{T}(\mathbf{A}^T \mathbf{A} - \lambda_2^2 \mathbf{I})]^{-1} \mathcal{T} \mathbf{A}^T \mathbf{b}, \quad (\text{B-5})$$

where  $\mathcal{T}$  is a smoothing operator, and  $\lambda_1$  and  $\lambda_2$  are two parameters controlling the physical dimensionality and enabling fast convergence when inversion is implemented iteratively. These two parameters can be chosen as  $\lambda_1 = \|\mathbf{B}^T \mathbf{B}\|_2$  and  $\lambda_2 = \|\mathbf{A}^T \mathbf{A}\|_2$  (Fomel, 2007a). The definition of  $\mathbf{c}_1$  and  $\mathbf{c}_2$  are equivalent to definition of LOW in this paper.

## REFERENCES

- Abma, R. L., T. Manning, M. Tanis, J. Yu, and M. Foster, 2010, High quality separation of simultaneous sources by sparse inversion: 72nd Annual International Conference and Exhibition, EAGE, Extended Abstracts.
- Beasley, C. J., R. E. Chambers, and Z. Jiang, 1998, A new look at simultaneous sources: 68th Annual International Meeting, SEG, Expanded Abstracts, 133–135.
- Bekara, M., and M. van der Baan, 2009, Random and coherent noise attenuation by empirical mode decomposition: *Geophysics*, **74**, V89–V98.
- Berkhout, A. J., 2008, Changing the mindset in seismic data acquisition: *The Leading Edge*, **27**, 924–938.
- Canales, L., 1984, Random noise reduction: SEG expanded abstracts: 54th Annual international meeting, 525–527.
- Chen, Y., 2014, Deblending using a space-varying median filter: *Exploration Geophysics*, doi:<http://dx.doi.org/10.1071/EG14051>.
- Chen, Y., S. Fomel, and J. Hu, 2014a, Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization: *Geophysics*, **79**, V179–V189.
- Chen, Y., and J. Ma, 2014, Random noise attenuation by  $f$ - $x$  empirical mode decomposition predictive filtering: *Geophysics*, **79**, V81–V91.
- Chen, Y., J. Yuan, Z. Jin, K. Chen, and L. Zhang, 2014b, Deblending using normal moveout and median filtering in common-midpoint gathers: *Journal of Geophysics and Engineering*, **11**, 045012.
- Fomel, S., 2007a, Local seismic attributes: *Geophysics*, **72**, A29–A33.
- , 2007b, Shaping regularization in geophysical-estimation problems: *Geophysics*, **72**, R29–R36.
- , 2009, Adaptive multiple subtraction using regularized nonstationary regression: *Geophysics*, **74**, V25–V33.
- , 2013, Seismic data decomposition into spectral components using regularized nonstationary autoregression: *Geophysics*, **78**, O69–O76.
- Fomel, S., M. Backus, K. Fouad, B. Hardage, and G. Winters, 2005, A multistep approach to multicomponent seismic image registration with application to a West Texas carbonate reservoir study: 75th Annual International Meeting, SEG, Expanded Abstracts, 1018–1021.
- Fomel, S., and L. Jin, 2009, Time-lapse image registration using the local similarity attribute: *Geophysics*, **74**, A7–A11.
- Fomel, S., P. Sava, I. Vlad, Y. Liu, and V. Bashkardin, 2013, Madagascar open-source software project: *Journal of Open Research Software*, **1**, e8.
- Fomel, S., and M. van der Baan, 2014, Local skewness attribute as a seismic phase detector: *Interpretation*, **2**, SA49–SA56.
- Hazewinkel, M., 2001, *Encyclopedia of mathematics*: Springer.
- Huo, S., Y. Luo, and P. G. Kelamis, 2012, Simultaneous sources separation via multidirectional vector-median filtering: *Geophysics*, **77**, V123–V131.
- Liu, G., X. Chen, J. Du, and K. Wu, 2012, Random noise attenuation using  $f$ - $x$  regularized nonstationary autoregression: *Geophysics*, **77**, V61–V69.

- Liu, G., S. Fomel, and X. Chen, 2011a, Stacking angle-domain common-image gathers for normalization of illumination: *Geophysical Prospecting*, **59**, 244–255.
- , 2011b, Time-frequency analysis of seismic data using local attributes: *Geophysics*, **76**, P23–P34.
- Liu, G., S. Fomel, L. Jin, and X. Chen, 2009, Stacking seismic data using local correlation: *Geophysics*, **74**, V43–V48.
- Liu, Y., and S. Fomel, 2012, Seismic data analysis using local time-frequency decomposition: *Geophysical Prospecting*, **60**, 1–10.
- Liu, Y., S. Fomel, and G. Liu, 2010, Nonlinear structure-enhancing filtering using plane-wave prediction: *Geophysical Prospecting*, **58**, 415–427.
- Soubaras, R., 1995, Prestack random and impulsive noise attenuation by f-x projection filtering: 65th Annual International Meeting, SEG, Expanded Abstracts, 711–714.
- Zhang, R., X. Song, S. Fomel, M. K. Sen, and S. Srinivasan, 2013, Time-lapse seismic data registration and inversion for CO<sub>2</sub> sequestration study at Cranfield: *Geophysics*, **78**, B329–B338.