

# Noniterative $f$ - $x$ - $y$ streaming prediction filtering for random noise attenuation on seismic data<sup>a</sup>

<sup>a</sup>Published in IEEE Transactions on Geoscience and Remote Sensing, 60, 1-9, Early Access, (2022)

*Yang Liu\**, *Zhisheng Zheng\**

## ABSTRACT

Random noise is unavoidable in seismic exploration, especially under complex-surface conditions and in deep-exploration environments. The current problems in random noise attenuation include preserving the nonstationary characteristics of the signal and reducing computational cost of broadband, wide-azimuth, and high-density data acquisition. To obtain high-quality images, traditional prediction filters (PFs) have proved effective for random noise attenuation, but these methods typically assume that the signal is stationary. Most nonstationary PFs use an iterative strategy to calculate the coefficients, which leads to high computational costs. In this study, we extended the streaming prediction theory to the frequency domain and proposed the  $f$ - $x$ - $y$  streaming prediction filter (SPF) to attenuate random noise. Instead of using the iterative optimization algorithm, we formulated a constraint least-squares problem to calculate the SPF and derived an analytical solution to this problem. The multi-dimensional streaming constraints are used to increase the accuracy of the SPF. We also modified the recursive algorithm to update the SPF with the snaky processing path, which takes full advantage of the streaming structure to improve the effectiveness of the SPF in high dimensions. In comparison with 2D  $f$ - $x$  SPF and 3D  $f$ - $x$ - $y$  regularized nonstationary autoregression (RNA), we tested the practicality of the proposed method in attenuating random noise. Numerical experiments show that the 3D  $f$ - $x$ - $y$  SPF is suitable for large-scale seismic data with the advantages of low computational cost, reasonable nonstationary signal protection, and effective random noise attenuation.

## INTRODUCTION

In seismic exploration, random noise is unavoidable because it is composed of environmental noise, the interference of wind motion, and the noise from geophones Yilmaz (2001). Meanwhile, complex subsurface media typically cause the energy loss experienced by seismic signals, which show low amplitude in deep exploration conditions. These factors result in low-quality data and make signal-to-noise ratio (SNR) gradually low. In obtaining high-quality seismic images and improving the

SNR, one key problem is the nonstationary characteristics of the signal. Seismic data are time-varying in nature, and the nonstationary properties of seismic data display that energy, track, time-frequency spectra of seismic events, and statistical characteristics of random noise change with time and space. The denoising methods that consider such nonstationary features can better preserve the valid signals. The other problem with denoising is the increasing computational costs, although broadband, wide-azimuth, and high-density data acquisition can lead to high-resolution and high-fidelity images. Many authors have proposed methods for random noise attenuation based on different theories. The mean filter Bonar and Sachhi (2012) and median filter Liu et al. (2009); Wu et al. (2018) are effective denoising methods for images, but they may somewhat smear seismic signals when complicated structures and low SNR are encountered. Mathematical transforms such as wavelet transforms Berkner and Wells (1998); Yu et al. (2007); Langston and Mousavi (2019) and seislet transforms Fomel and Liu (2010); Liu and Fomel (2010); Liu et al. (2015) can characterize the nonstationary properties of seismic signals and provide reasonable signal and noise separation based on their compression ability. Recently, deep learning or machine learning techniques Djarfour et al. (2014); Kimiaefar et al. (2016); Yu et al. (2019); Zhu et al. (2019) have also been proposed to suppress random noise; however, the initialization of neural networks requires a large number of samples and high computational cost. For supervised and semi-supervised learning, the preparation of the training set may need the help of traditional methods to generate denoised results as a reference.

Prediction filters (PFs) have proved effective for random noise attenuation, and they can be implemented in the time-space or frequency-space domain. When seismic events have varying slope, the configuration of the filter size influences the filtering results, especially the filter size of  $t$ - $x$  PFs along the time axis. There are few impacts on  $f$ - $x$  PFs because they only estimate data along the spatial directions. Besides, seismic events with different dominant frequencies are overlapped in the  $t$ - $x$  domain, and they can be naturally separated in the  $f$ - $x$  domain. Abma and Claerbout Abma and Claerbout (1995) discussed the differences in PFs in the  $f$ - $x$  and  $t$ - $x$  domains. The  $f$ - $x$  prediction filter for denoising was first introduced by Canales Canales (1984), and further developed by Güllünay Güllünay (1986) to a standard industry method known as FXDECON, which is equivalent to a  $t$ - $x$  domain prediction filter, selecting the entire trace along the time direction. Liu *et al.* Liu et al. (2012) developed the  $f$ - $x$  adaptive prediction filters to suppress random noise by using regularized nonstationary autoregression (RNA); the regularization term was used to limit the global smoothness of the filter coefficients. The method was further extended from a two-dimensional (2D) to three-dimensional (3D) case for random noise attenuation Liu and Chen (2013). The  $f$ - $x$ - $y$  RNA provides preferable adaptive features because it uses an iterative algorithm to calculate the frequency-space-varying filter coefficients, which leads to a large storage and high computational time, especially in large-scale data processing.

Local similarity constraints have been proposed to directly calculate the adaptive prediction filter without iterations, which can save the computational resources.

Starting with the prediction equation for a certain data point, Sacchi and Naghizadeh Sacchi and Naghizadeh (2009) transformed the ill-posed problem of adaptive prediction filter into a local smoothing problem, and introduced a quadratic regularization term to stabilize the solution of the local prediction filter. Fomel and Claerbout Fomel and Claerbout (2016) proposed the concept of streaming prediction error filter (SPEF) to update the filter as each new data value arrives in the time-space domain. This method combines the prediction equation with locally similar constraints to solve the overdetermined linear system. Arising from different starting points, these two methods share the same least-squares solution and significantly reduce the computational cost by avoiding the iterative algorithm. Liu and Li Liu and Li (2018) further proposed a streaming orthogonal prediction filter (SOPF) in the  $t$ - $x$  domain, which applies signal-and-noise orthogonalization based on the streaming prediction theory and provides a fast solution for the adaptive prediction filter to suppress random noise. Guo *et al.* Guo et al. (2020) attempted to eliminate seismic random noise by using the  $f$ - $x$  SPF only with 1D spatial constraint.

In this study, we derived the theory of the new  $f$ - $x$ - $y$  streaming prediction filter based on the local smoothness constraints in high dimensions. The multi-dimensional constraints make the filter involve the property of local similarity not only along the spatial directions (space  $x$  and space  $y$ ), but also the frequency direction. It is not a simple 3D extension with more spatial axis from 2D SPF Guo et al. (2020), we took advantage of 3D data with space  $y$  constraint and suppressed oscillation with frequency constraint, meanwhile, a special filter-updating path was developed to help the SPF solve the random noise attenuation problem in higher dimensions. We compared the feasibility of the 3D  $f$ - $x$ - $y$  SPF in attenuating random noise with the 2D  $f$ - $x$  SPF and the 3D  $f$ - $x$ - $y$  RNA on two synthetic models. The field data example confirms that the 3D  $f$ - $x$ - $y$  SPF with the matching processing path has a reasonable denoising ability and a low computational cost in practice.

## THEORY

### 3D $f$ - $x$ - $y$ streaming prediction filtering

A 2D seismic section  $s(t, x)$  containing linear events can be described as a plane wave function in the  $t$ - $x$  domain. In the  $f$ - $x$  domain, the linear events in seismic section  $\tilde{S}(f, x)$  are decomposed into a series of sinusoids. These sinusoids are superimposed and become harmonics at each frequency, which shows the prediction relationship of seismic traces in a frequency slice:

$$\sum_{p=1}^P a_{m,p} \tilde{S}_{m,n-p} = \tilde{S}_{m,n}, \quad (1)$$

where  $m \in [1, M]$  and  $n \in [1, N]$  are the indices of the seismic sample along the  $f$  axis and  $x$  axis, respectively.  $p \in [1, P]$  is the index of filter coefficients along

the  $x$  direction.  $\tilde{S}_{m,n}$  denotes the data point in  $\tilde{S}(f, x)$  and  $a_{m,p}$  indicates the filter coefficient in the  $f$ - $x$  domain. When curve events or amplitude-varying wavelets are shown in the seismic data, filter coefficients change from one data point to the next, which help to manage the nonstationary case:

$$\sum_{p=1}^P a_{m,n,p} \tilde{S}_{m,n-p} = \mathbf{S}_{m,n}^T \mathbf{A}_{m,n} = \tilde{S}_{m,n}, \quad (2)$$

where  $\{\mathbf{T}\}$  denotes the transpose operator,  $\mathbf{S}_{m,n} = [\tilde{S}_{m,n-1}, \tilde{S}_{m,n-2}, \dots, \tilde{S}_{m,n-P}]^T$  denotes the vector including the data points near  $\tilde{S}_{m,n}$ .  $\mathbf{A}_{m,n} = [a_{m,n,1}, a_{m,n,2}, \dots, a_{m,n,P}]^T$  is the vector of coefficients in a 2D adaptive prediction filter. Let  $P = 3$ , Fig. 1a illustrates how (2) works. Equation (2) denotes that the filter predicts data point along the spatial direction rather than the frequency direction. Therefore, an extension to the 3D  $f$ - $x$ - $y$  domain is straightforward:

$$\begin{aligned} & \sum_{p=-P}^P \sum_{q=-Q}^Q a_{m,n,l,p,q} \tilde{S}_{m,n-p,l-q} \\ &= \mathbf{S}_{m,n,l}^T \mathbf{A}_{m,n,l} = \tilde{S}_{m,n,l} \quad (|p| + |q| \neq 0), \end{aligned} \quad (3)$$

where  $l \in [1, L]$  is the index of the data sample along the  $y$  axis.  $p \in [-P, P]$  and  $q \in [-Q, Q]$  are the indices of filter coefficients in two spatial directions,  $\mathbf{A}_{m,n,l} = [a_{m,n,l,-P,-Q}, \dots, a_{m,n,l,p,q}, \dots, a_{m,n,l,P,Q}]^T$  is the vector of 3D filter coefficients. The adaptive prediction filter  $\mathbf{A}_{m,n,l}$  is defined as a space-noncausal structure and the filter size in the spatial direction is  $(2P + 1) \times (2Q + 1) - 1$ . The vector  $\mathbf{S}_{m,n,l} = [\tilde{S}_{m,n+P,l+Q}, \dots, \tilde{S}_{m,n-p,l-q}, \tilde{S}_{m,n-P,l-Q}]^T$  contains the data points near  $\tilde{S}_{m,n,l}$ . Let  $p = \{-2, -1, 0, 1, 2\}$  and  $q = \{-2, -1, 0, 1, 2\}$ , Fig. 2a demonstrates the distribution of the vectors  $\mathbf{S}_{m,n,l}$  and  $\mathbf{A}_{m,n,l}$ . Assuming that the contained noise is white Gaussian noise, the filter can be obtained by solving the minimization problem:

$$\min_{\mathbf{A}_{m,n,l}} \left\| \mathbf{S}_{m,n,l}^T \mathbf{A}_{m,n,l} - \tilde{S}_{m,n,l} \right\|_2^2, \quad (4)$$

equation (4) describes an ill-posed problem that the number of the unknown filter coefficients is greater than that of the known equations. Without any regularization, the equation will lead to an unstable solution:

$$\mathbf{A}_{LS} = (\mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T)^{-1} \mathbf{S}_{m,n,l}^* \tilde{S}_{m,n,l}, \quad (5)$$

where  $\{*\}$  denotes the conjugate operator.

To solve the underdetermined problem (4), constraint conditions based on local similarity/smoothness are used to stabilize the solution of (4). Assuming that the adaptive prediction filter at position  $(m, n, l)$  is similar to another one at position  $(m, n - 1, l)$  in the  $f$ - $x$ - $y$  domain,  $\lambda_x \mathbf{A}_{m,n,l} \approx \lambda_x \mathbf{A}_{m,n-1,l}$  can be treated as the constraint condition on the  $x$  axis. The autoregression equation can be expressed as



follows:

$$\begin{bmatrix} \tilde{S}_{m,n+P,l+Q} & \cdots & \tilde{S}_{m,n-p,l-q} & \cdots & \tilde{S}_{m,n-P,l-Q} \\ & \lambda_x & & & \\ & & \ddots & & \\ & & & \lambda_x & \\ & & & & \ddots \\ & & & & & \lambda_x \end{bmatrix} \begin{bmatrix} a_{m,n,l,-P,-Q} \\ \vdots \\ a_{m,n,l,p,q} \\ \vdots \\ a_{m,n,l,P,Q} \end{bmatrix} \approx \begin{bmatrix} \tilde{S}_{m,n,l} \\ \lambda_x a_{m,n-1,l,-P,-Q} \\ \vdots \\ \lambda_x a_{m,n-1,l,p,q} \\ \vdots \\ \lambda_x a_{m,n-1,l,P,Q} \end{bmatrix}, \quad (6)$$

and the simplified block matrix can be written as:

$$\begin{bmatrix} \mathbf{S}_{m,n,l}^T \\ \lambda_x \mathbf{I} \end{bmatrix} \mathbf{A}_{m,n,l} \approx \begin{bmatrix} \tilde{S}_{m,n,l} \\ \lambda_x \mathbf{A}_{m,n-1,l} \end{bmatrix}. \quad (7)$$

Equation (7) is solvable since there are  $(2P+1) * (2Q+1)$  equations with  $(2P+1) * (2Q+1) - 1$  unknown coefficients, which correspond to the following minimization problem:

$$\min_{\mathbf{A}_{m,n,l}} \left\| \mathbf{S}_{m,n,l}^T \mathbf{A}_{m,n,l} - \tilde{S}_{m,n,l} \right\|_2^2 + \lambda_x^2 \left\| \mathbf{A}_{m,n,l} - \mathbf{A}_{m,n-1,l} \right\|_2^2, \quad (8)$$

where  $\lambda_x$  is the constant weight for the regularization term along the  $x$  axis. In the frequency  $f$  direction, one can assume that the SPFs change smoothly and treat the irregular perturbations as the interference of noise. Meanwhile, the smoothness of the 3D SPFs also exists in different spatial directions and may change at different data point, therefore, we implemented local smoothness along  $f$ ,  $x$ , and  $y$  axis as constraints to calculate the  $f$ - $x$ - $y$  SPF. The block matrix form is:

$$\begin{bmatrix} \mathbf{S}_{m,n,l}^T \\ \lambda_f(m,n,l) \mathbf{I} \\ \lambda_x(m,n,l) \mathbf{I} \\ \lambda_y(m,n,l) \mathbf{I} \end{bmatrix} \mathbf{A}_{m,n,l} = \begin{bmatrix} \tilde{S}_{m,n,l} \\ \lambda_f(m,n,l) \mathbf{A}_{m-1,n,l} \\ \lambda_x(m,n,l) \mathbf{A}_{m,n-1,l} \\ \lambda_y(m,n,l) \mathbf{A}_{m,n,l-1} \end{bmatrix} \quad (9)$$

and the corresponding least-squares problem takes the following form:

$$\begin{aligned} \min_{\mathbf{A}_{m,n,l}} & \left\| \mathbf{S}_{m,n,l}^T \mathbf{A}_{m,n,l} - \tilde{S}_{m,n,l} \right\|_2^2 \\ & + \lambda_f^2(m,n,l) \left\| \mathbf{A}_{m,n,l} - \mathbf{A}_{m-1,n,l} \right\|_2^2 \\ & + \lambda_x^2(m,n,l) \left\| \mathbf{A}_{m,n,l} - \mathbf{A}_{m,n-1,l} \right\|_2^2 \\ & + \lambda_y^2(m,n,l) \left\| \mathbf{A}_{m,n,l} - \mathbf{A}_{m,n,l-1} \right\|_2^2, \end{aligned} \quad (10)$$

where  $\lambda_f(m,n,l)$ ,  $\lambda_x(m,n,l)$  and  $\lambda_y(m,n,l)$  denotes the variable weights of regularization terms along the frequency  $f$  axis, space  $x$  axis, and space  $y$  axis, respectively. They measure the variable similarity or closeness between the filter  $\mathbf{A}_{m,n,l}$  and the adjacent filters  $\mathbf{A}_{m-1,n,l}$ ,  $\mathbf{A}_{m,n-1,l}$ , and  $\mathbf{A}_{m,n,l-1}$ . Due to the prediction filter can characterize the energy spectrum of the input data Claerbout (1976), the adaptive filter shares analogical smoothness property with the 3D data, so the variation of the weights may consist with the smooth version of data. For simplicity, we select these

weights with constant value,  $\lambda_f(m, n, l) = \lambda_f$ ,  $\lambda_x(m, n, l) = \lambda_x$  and  $\lambda_y(m, n, l) = \lambda_y$ , to demonstrate the constrained relationship. The introduced regularization terms convert the ill-posed problem to the overdetermined inverse problem, and the least-squares solution of (9) and (10) is:

$$\begin{aligned} \mathbf{A}_{m,n,l} &= [(\lambda_f^2 + \lambda_x^2 + \lambda_y^2)\mathbf{I} + \mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T]^{-1} (\lambda_f^2 \mathbf{A}_{m-1,n,l} \\ &\quad + \lambda_x^2 \mathbf{A}_{m,n-1,l} + \lambda_y^2 \mathbf{A}_{m,n,l-1} + \tilde{S}_{m,n,l} \mathbf{S}_{m,n,l}^*) \\ &= (\lambda^2 \mathbf{I} + \mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T)^{-1} (\lambda^2 \hat{\mathbf{A}}_{m,n,l} + \tilde{S}_{m,n,l} \mathbf{S}_{m,n,l}^*). \end{aligned} \quad (11)$$

where

$$\begin{cases} \lambda^2 = \lambda_f^2 + \lambda_x^2 + \lambda_y^2, \\ \hat{\mathbf{A}}_{m,n,l} = (\lambda_f^2 \mathbf{A}_{m-1,n,l} + \lambda_x^2 \mathbf{A}_{m,n-1,l} + \lambda_y^2 \mathbf{A}_{m,n,l-1}) / \lambda^2. \end{cases} \quad (12)$$

The Sherman-Morrison formula is an analytic method for solving the inverse matrix Hager (1989):

$$(\lambda^2 \mathbf{I} + \mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T)^{-1} = \frac{1}{\lambda^2} \left( \mathbf{I} - \frac{\mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T}{\lambda^2 + \mathbf{S}_{m,n,l}^T \mathbf{S}_{m,n,l}^*} \right). \quad (13)$$

The derivation of the Sherman-Morrison formula in the complex space is described in Appendix . Elementary algebraic simplifications lead to the analytical solution:

$$\begin{aligned} \mathbf{A}_{m,n,l} &= (\lambda^2 \mathbf{I} + \mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T)^{-1} (\tilde{S}_{m,n,l} \mathbf{S}_{m,n,l}^* + \lambda^2 \hat{\mathbf{A}}_{m,n,l}) \\ &= \frac{1}{\lambda^2} \left( \mathbf{I} - \frac{\mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T}{\lambda^2 + \mathbf{S}_{m,n,l}^T \mathbf{S}_{m,n,l}^*} \right) (\tilde{S}_{m,n,l} \mathbf{S}_{m,n,l}^* + \lambda^2 \hat{\mathbf{A}}_{m,n,l}) \\ &= \hat{\mathbf{A}}_{m,n,l} + \frac{1}{\lambda^2} \tilde{S}_{m,n,l} \mathbf{S}_{m,n,l}^* - \frac{1}{\lambda^2} \frac{\mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T \tilde{S}_{m,n,l} \mathbf{S}_{m,n,l}^*}{\lambda^2 + \mathbf{S}_{m,n,l}^T \mathbf{S}_{m,n,l}^*} \\ &\quad - \frac{\mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T \hat{\mathbf{A}}_{m,n,l}}{\lambda^2 + \mathbf{S}_{m,n,l}^T \mathbf{S}_{m,n,l}^*} \\ &= \hat{\mathbf{A}}_{m,n,l} + \frac{1}{\lambda^2} \frac{\lambda^2 \tilde{S}_{m,n,l} \mathbf{S}_{m,n,l}^* + \tilde{S}_{m,n,l} \mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T \mathbf{S}_{m,n,l}^*}{\lambda^2 + \mathbf{S}_{m,n,l}^T \mathbf{S}_{m,n,l}^*} \\ &\quad - \frac{1}{\lambda^2} \frac{\mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T \tilde{S}_{m,n,l} \mathbf{S}_{m,n,l}^*}{\lambda^2 + \mathbf{S}_{m,n,l}^T \mathbf{S}_{m,n,l}^*} - \frac{\mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T \hat{\mathbf{A}}_{m,n,l}}{\lambda^2 + \mathbf{S}_{m,n,l}^T \mathbf{S}_{m,n,l}^*} \\ &= \hat{\mathbf{A}}_{m,n,l} + \frac{\tilde{S}_{m,n,l} - \mathbf{S}_{m,n,l}^T \hat{\mathbf{A}}_{m,n,l}}{\lambda^2 + \mathbf{S}_{m,n,l}^T \mathbf{S}_{m,n,l}^*} \mathbf{S}_{m,n,l}^*, \end{aligned} \quad (14)$$

Equation (14) is a recursion equation, which suggests that the filter  $\mathbf{A}_{m,n,l}$  recursively updates in a certain order. The residual can be written as:

$$r_{m,n,l} = \lambda^2 \frac{\tilde{S}_{m,n,l} - \mathbf{S}_{m,n,l}^T \hat{\mathbf{A}}_{m,n,l}}{\lambda^2 + \mathbf{S}_{m,n,l}^T \mathbf{S}_{m,n,l}^*}. \quad (15)$$

Once obtaining the solution of the 3D  $f$ - $x$ - $y$  SPF, one can compute the noise-free data  $\tilde{X}_{m,n,l}$  with the following equation:

$$\tilde{X}_{m,n,l} = \mathbf{S}_{m,n,l}^T \mathbf{A}_{m,n,l}. \quad (16)$$

Table 1: Cost comparison between  $f$ - $x$ - $y$  RNA and  $f$ - $x$ - $y$  SPF.

	Storage	Cost
$f$ - $x$ - $y$ RNA	$O(N_a N_f N_x N_y)$	$O(N_a N_f N_x N_y N_{iter})$
$f$ - $x$ - $y$ SPF	$O(N_a N_x N_y)$	$O(N_a N_f N_x N_y)$

$N_a$  is the filter size,  $N_f$  is the data size along frequency axis,  $N_x$  is the data size along  $x$  axis,  $N_y$  is the data size along  $y$  axis, and  $N_{iter}$  is the number of iterations.

The configuration of parameters  $\lambda_f$ ,  $\lambda_x$ , and  $\lambda_y$  is the basis for the proposed method. When the three parameters are 0, the corresponding regularization terms have no effect on restricting the inverse problem and the result of the  $f$ - $x$ - $y$  SPF becomes (5). By choosing  $\lambda_y = 0$  and removing the  $y$  axis, equation (14) is reduced to the solution of the 2D  $f$ - $x$  SPF. On the contrary, when the three parameters tend to infinity, more weight is applied on regularization terms. A large denominator in (14) indicates that the filter cannot receive any updates to maintain its adaptive and predictive properties. This denominator suggests that parameters  $\lambda_x^2$  and  $\lambda_y^2$  in (12) should have the same order of magnitude as  $\mathbf{S}_{m,n,l}^T \mathbf{S}_{m,n,l}^*$ , and the value of  $\lambda$  might be in the range of  $(0, 10 * \sqrt{\max(\mathbf{S}_{m,n,l}^T \mathbf{S}_{m,n,l}^*)})$ , which can balance the noise suppression and signal protection. Therefore, they can smoothly adjust the change of filters. Meanwhile, data distribution in the frequency axis may change sharply, which is not as smooth as those in the spatial directions; thus,  $\lambda_f$  should be smaller than  $\lambda_x$  and  $\lambda_y$ .

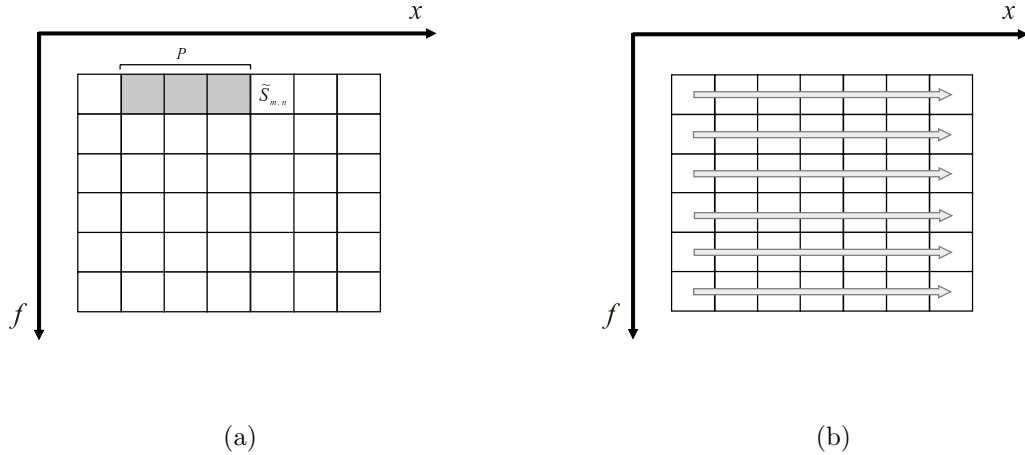


Figure 1: Schematic illustration of  $f$ - $x$  prediction filter (a) and filter processing path (b).

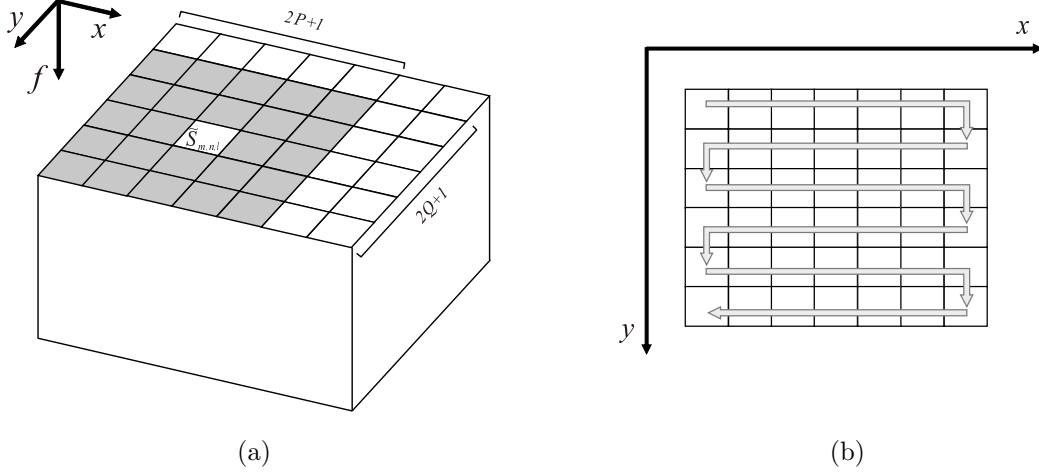


Figure 2: Schematic illustration of  $f$ - $x$ - $y$  prediction filter (a) and filter processing path in each frequency slice (b).

### Data processing path in 3D case

The processing path for the 2D  $f$ - $x$  SPF is shown in Fig. 1b, which initializes the filter at the beginning of each line and updates the filter until the end of the line along the spatial direction. We defined a new processing path for the 3D case; the 3D  $f$ - $x$ - $y$  SPF will process data from the beginning of a line to the end, and then from the end of the next line to the beginning in the spatial directions (Fig. 2b). This snaky processing path avoids the weakness that the filter should be initialized at the beginning of each line, which guarantees the effectiveness of initial filters.

For calculating the 3D  $f$ - $x$ - $y$  SPF, the neighboring filters  $\mathbf{A}_{m-1,n,l}$ ,  $\mathbf{A}_{m,n-1,l}$ , and  $\mathbf{A}_{m,n,l-1}$  need to be stored, and these filters will be used when the stream reaches the adjacent point. For easy program implementation, we designed the cache space to store the neighboring filters. The cache space for  $\mathbf{A}_{m-1,n,l}$ ,  $\mathbf{A}_{m,n,l-1}$ , and  $\mathbf{A}_{m,n-1,l}$  are  $N_a N_x N_y$ ,  $N_a N_x$ , and  $N_a$ , respectively, where  $N_a$  is the filter size,  $N_x$  is the data size along the  $x$  axis, and  $N_y$  is the data size along the  $y$  axis. Compared with the 3D  $f$ - $x$ - $y$  RNA Liu and Chen (2013), the proposed method calculates the filter coefficients without iterations, which reduces the requirement of computational resources (Table 1). Repeatedly processing the data with the SPF can further suppress the noise, but part of the signal will also be weakened. Computational cost increases with the number of repeated calculations, and a balance between computational cost and noise suppression effect is needed.

The key steps of using the 3D  $f$ - $x$ - $y$  SPF to attenuate the random noise are as follows:

1. Initializing filter coefficients  $\mathbf{A}_{m-1,n,l}$ ,  $\mathbf{A}_{m,n-1,l}$ , and  $\mathbf{A}_{m,n,l-1}$  with zeros.
2. Selecting reasonable parameters  $\lambda_f$ ,  $\lambda_x$ , and  $\lambda_y$  and computing  $\lambda^2 = \lambda_f^2 + \lambda_x^2 + \lambda_y^2$ .

3. Calculating  $\mathbf{S}^T_{m,n,l} \mathbf{S}^*_{m,n,l}$  and  $\mathbf{S}^T_{m,n,l} \hat{\mathbf{A}}_{m,n,l}$  in (14).
4. Computing residual  $r_{m,n,l}$  with (15) and updating filter  $\mathbf{A}_{m,n,l}$  with (14).
5. Estimating noise-free data  $\tilde{X}_{m,n,l}$  with (16).
6. Looping steps 2-5 in snake processing path until the entire process is completed.

## NUMERICAL EXAMPLES

### 3D synthetic qdome model

We start with the 3D qdome model Claerbout and Fomel (2008) containing curve events and faults (Fig. 3a) to evaluate the proposed method by handling the non-stationarity problem. The model size is 200 (time samples)  $\times$  150 (X traces)  $\times$  100 (Y traces). Fig. 3b displays the data with Gaussian noise added. We compared the 3D  $f$ - $x$ - $y$  SPF with the 2D  $f$ - $x$  SPF and the 3D  $f$ - $x$ - $y$  RNA Liu and Chen (2013) to test their ability for random noise attenuation. The filter length of the  $f$ - $x$  SPF is 5-sample ( $x$ ). We also selected the scale parameters with 0.008 ( $\lambda_f$ ) and 0.06 ( $\lambda_x$ ). Fig. 4a shows the denoised result obtained by using the  $f$ - $x$  SPF that eliminates most of the random noise. However, there is still an obvious signal in the noise section (Fig. 4b) because the 2D  $f$ - $x$  SPF has a low accuracy owing to the local similarity of filter coefficients only along the  $f$  and  $x$  directions. A more effective approach is to apply global smoothness. The denoised result obtained by using the 3D  $f$ - $x$ - $y$  RNA is shown in Fig. 4c. The filter size of the  $f$ - $x$ - $y$  RNA is 5-sample ( $x$ )  $\times$  5-sample ( $y$ ). The 3D  $f$ - $x$ - $y$  RNA has a better result than the 2D  $f$ - $x$  SPF, and it is visually difficult to detect the signal in the difference between the noisy data (Fig. 3b) and the denoised result (Fig. 4c). The global smoothness constraints along two spatial directions can help RNA to improve the result (Fig. 4d), but it also increases the computational cost because it iteratively solves the regularized least-squares problem (Table 1). We designed a 3D  $f$ - $x$ - $y$  SPF with 5-sample ( $x$ )  $\times$  5-sample ( $y$ ) coefficients for each sample and the scale parameters 0.008 ( $\lambda_f$ ), 0.06 ( $\lambda_x$ ), and 0.06 ( $\lambda_y$ ). The proposed method produces a reasonable result (Fig. 4e), where the geological structure is improved. It is also difficult to distinguish the signal in the removed noise (Fig. 4f), which is an indication of successful signal and noise separation. The signal-to-noise ratio (SNR) and time consumption were used to analyze the performance of each method (Table 2). The SNR is defined as:

$$SNR = 10 \log_{10} \frac{\|\mathbf{s}\|_2^2}{\|\mathbf{s} - \hat{\mathbf{s}}\|_2^2}, \quad (17)$$

where  $\mathbf{s}$  is the noise-free signal and  $\hat{\mathbf{s}}$  is the denoised signal. The computing platform uses Intel E5-2650 2.0GHz CPU and the displayed time consumption is the average of ten records. Although the denoised result obtained by using the  $f$ - $x$  SPF shows a relatively high SNR, the amplitude of the curved events is partly damaged, which is shown in Fig. 4b. The  $f$ - $x$ - $y$  RNA preserves a more detailed structure than the  $f$ - $x$  SPF at the cost of a lower SNR and longer computational time. In general, the  $f$ - $x$ - $y$

SPF provides a higher SNR while maintaining the computational cost at a feasible level.

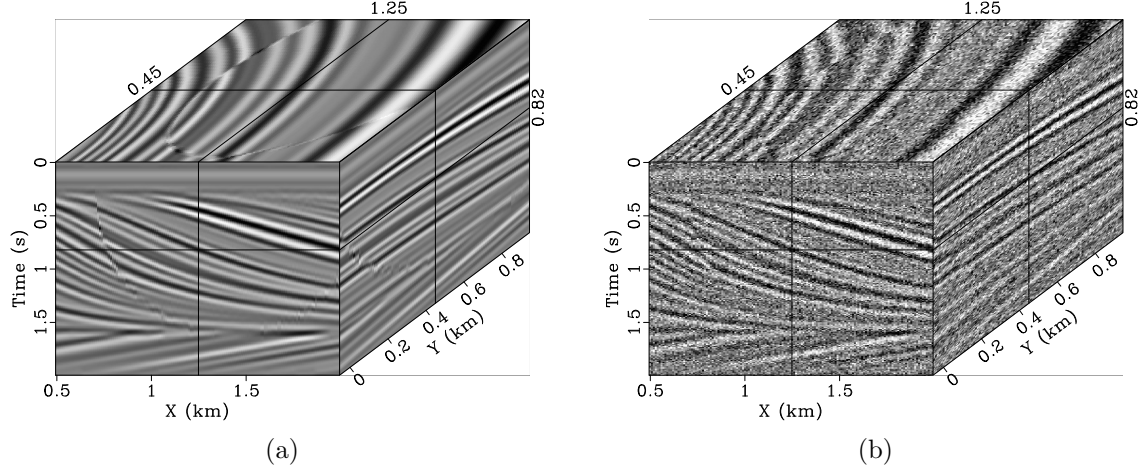


Figure 3: 3D synthetic model (a) and noisy data (b).

### 3D synthetic CMP gather

A 3D synthetic CMP gather created by Liu and Chen (2013) shows hyperbolic events (Fig. 5a). The data size is 126-sample (time)  $\times$  101-sample (X)  $\times$  101-sample (Y). The corresponding noisy data is shown in Fig. 5b. The challenge in this case is to account for both strong random noise and nonstationary events. Fig. 6a shows the denoised result using the 2D  $f$ - $x$  SPF with 5-sample ( $x$ ) filter size, which uses the scale parameters of 1.0 ( $\lambda_f$ ) and 4.5 ( $\lambda_x$ ). Note that the  $f$ - $x$  SPF can attain relatively high SNR (Table 2). However, it cannot provide sufficient protection for curve events; the events at far offsets can be destroyed (Fig. 6a) and part of the signals are left (Fig. 6b). Figure 6c is the denoised result by using the 3D curvelet transform, we adopted the percentage threshold with 10% to mute the random noise, and the 3D curvelet transform effectively suppresses the noise, but part of the signal energy leaks into the noise profile at  $-0.5 \sim 0.5$ km (X) (Figure 6d). We compared the proposed method with the 3D  $f$ - $x$ - $y$  RNA, and the filter size is 5-sample ( $x$ )  $\times$  5-sample ( $y$ ). The denoised result (Fig. 6e) and the removed noise (Fig. 6f) illustrate that the  $f$ - $x$ - $y$  RNA protects more signal by removing less noise. The parameters of the 3D  $f$ - $x$ - $y$  SPF are selected as 1.0 ( $\lambda_f$ ), 4.5 ( $\lambda_x$ ), and 4.5 ( $\lambda_y$ ), respectively. The filter size of the 3D SPF is the same as that of the 3D RNA. The proposed method recovers the curved events reasonable well (Fig. 6g), similar to the  $f$ - $x$ - $y$  RNA. However, the  $f$ - $x$ - $y$  SPF saves more computational resources and reveals a higher SNR than the  $f$ - $x$ - $y$  RNA (Table 2).

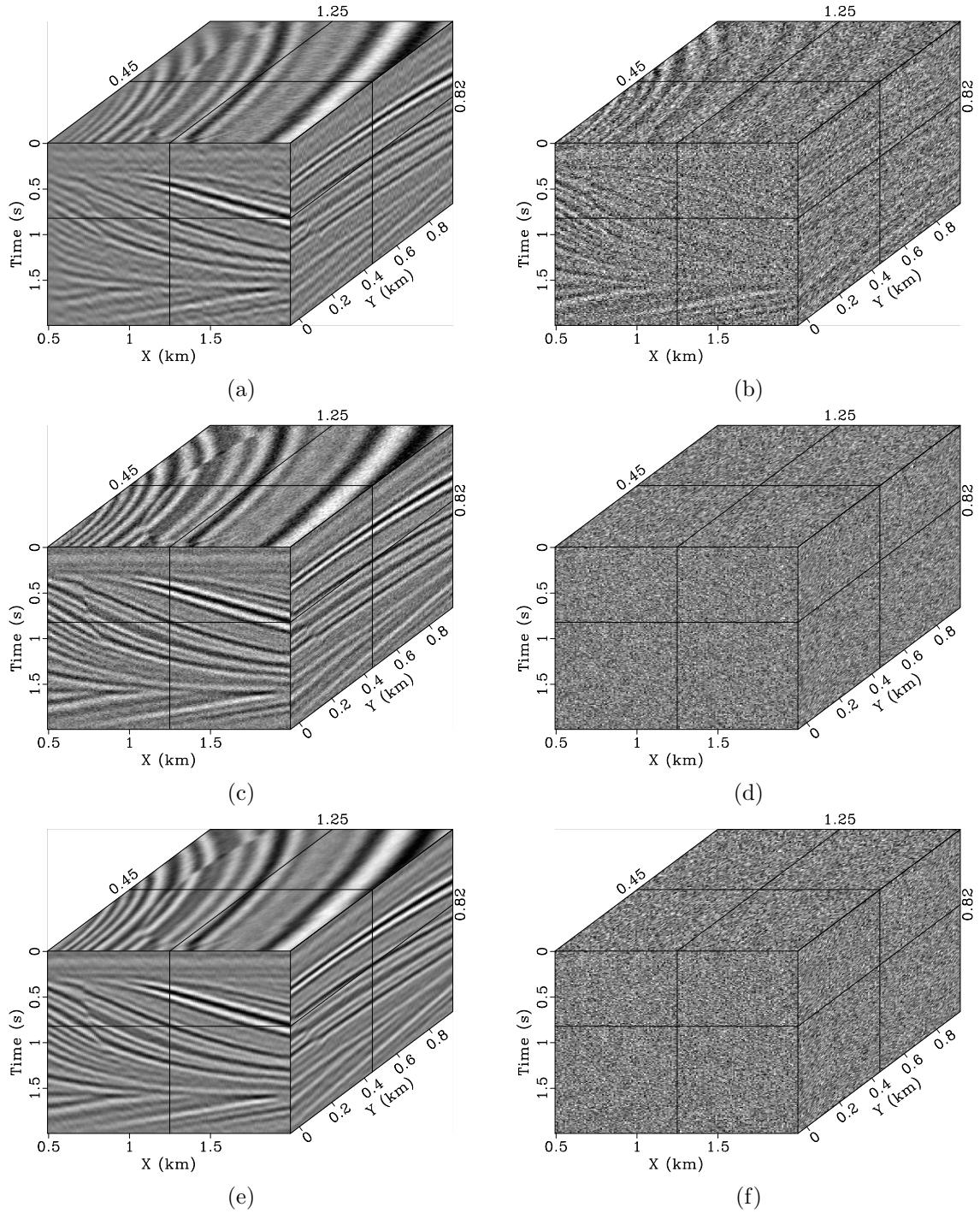


Figure 4: Denoised result by the  $f$ - $x$  SPF (a), noise removed by the  $f$ - $x$  SPF (b), denoised result by the  $f$ - $x$ - $y$  RNA (c), noise removed by the  $f$ - $x$ - $y$  RNA (d), denoised result by the  $f$ - $x$ - $y$  SPF (e), and noise removed by the  $f$ - $x$ - $y$  SPF (f).

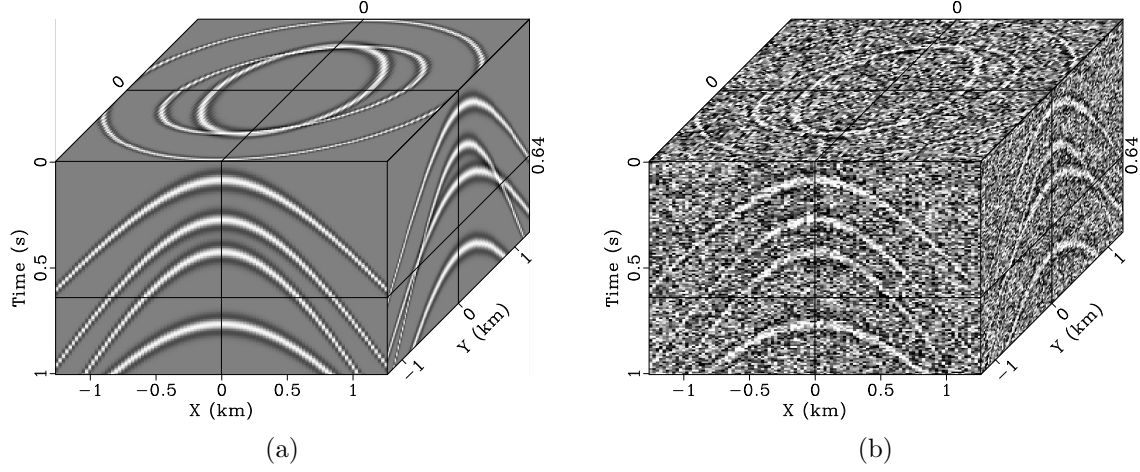


Figure 5: Synthetic 3D CMP gather (a) and noisy data (b).

Table 2: Comparison of the SNR and time consumption among different methods.

Model (size)	Method	$f$ - $x$ SPF		$f$ - $x$ - $y$ RNA		$f$ - $x$ - $y$ SPF	
		SNR (dB) <sup>1</sup>	Time (s) <sup>2</sup>	SNR (dB)	Time (s)	SNR (dB)	Time (s)
qdomo model (200*150*100)		6.284	$5.89 \times 10^0$	5.925	$8.23 \times 10^2$	14.16	$2.36 \times 10^1$
CMP gather (126*101*101)		2.422	$2.63 \times 10^0$	1.324	$1.53 \times 10^2$	2.544	$9.26 \times 10^0$
field data (700*266*310)		-	$8.79 \times 10^1$	-	$1.52 \times 10^3$	-	$1.37 \times 10^2$

<sup>1</sup> signal-noise-ratio of denoised result.<sup>2</sup> Time consumption is the average of ten records.

### 3D field data

For the field data test, we used a 3D time migration data to evaluate the effectiveness of the SPF (Fig. 7). The data size is 700-sample (time)  $\times$  266-sample (X)  $\times$  310-sample (Y). Strong random noise from the surface conditions contaminates both the simple layers at the shallow locations and complex structure at the deeper positions. We applied the 2D  $f$ - $x$  SPF with 5-sample ( $x$ ) to recover the events and selected 55 ( $\lambda_f$ ) and 280 ( $\lambda_x$ ) as the regularization terms for the improved  $f$ - $x$  SPF with frequency constraint. Fig. 8a and 8b show the denoised results and the removed noise at the same clip value, respectively. Both shallow plane events and deep dipping events show better lateral continuity, but the 2D SPF also removes a part of the events because it uses the local smoothness constraints without the  $y$  direction. Meanwhile, information is removed near  $x = 0$  and  $y = 0$  because of the inaccurate initial filter coefficients. We also compared the 2D SPF with the 2D  $f$ - $x$  EMD prediction filter Chen and Ma (2014) to test its ability for random noise attenuation. The denoised result of  $f$ - $x$  EMD prediction filter is shown in Fig. 9a. The 2D  $f$ - $x$  EMD prediction filter preserves signal better than the 2D SPF, but the difference (Fig. 9b) still shows obvious signal and the method is difficult to be implemented in 3D case. For comparison, we used the 3D  $f$ - $x$ - $y$  RNA to attenuate random noise. The  $f$ - $x$ - $y$  RNA with 5-sample (X)  $\times$  5-sample (Y)



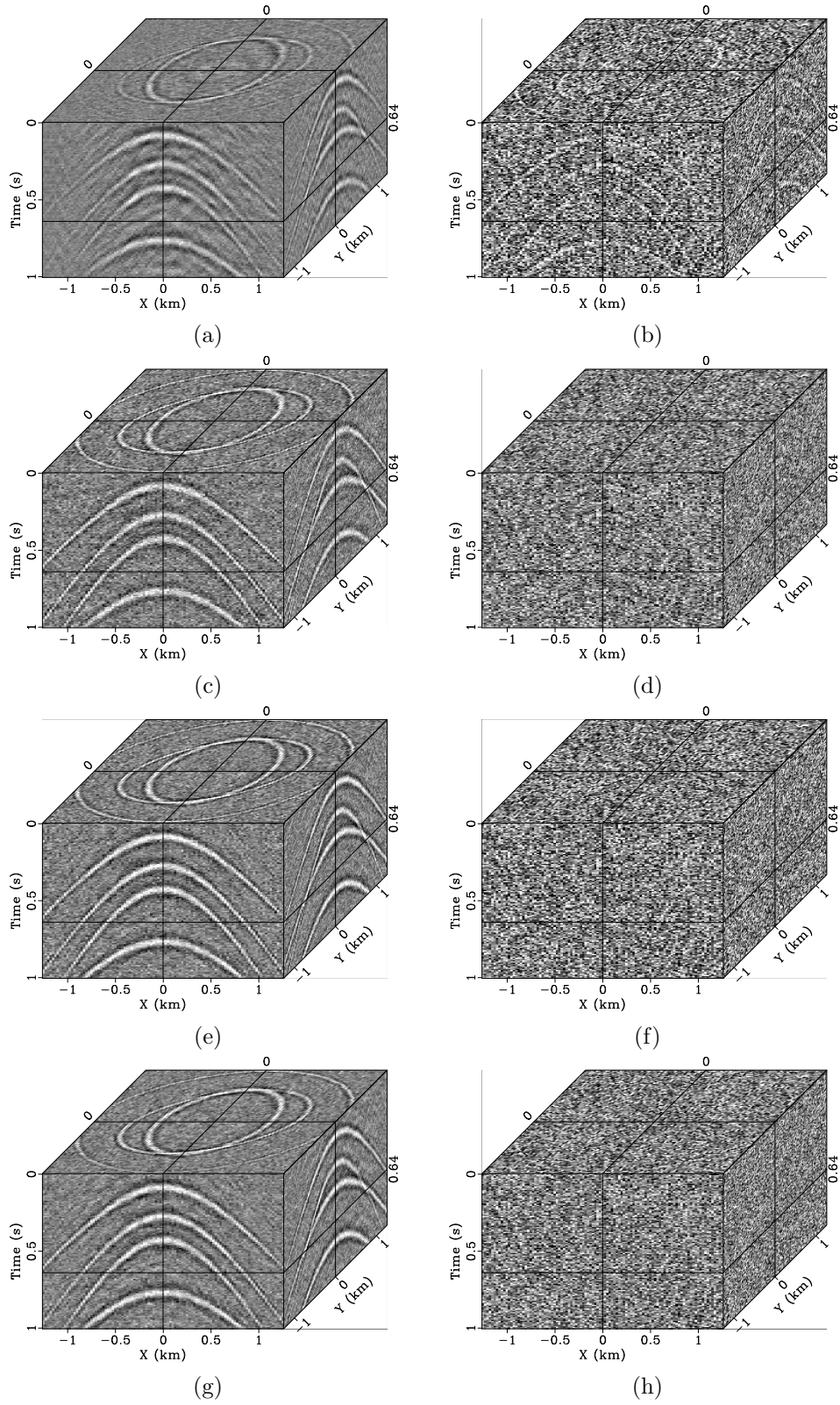


Figure 6: Denoised result by the  $f$ - $x$  SPF (a), noise removed by the  $f$ - $x$  SPF (b), denoised result by the 3D curvelet transform (c), noise removed by the 3D curvelet transform (d), denoised result by the  $f$ - $x$ - $y$  RNA (e), noise removed by the  $f$ - $x$ - $y$  RNA (f), denoised result by the  $f$ - $x$ - $y$  SPF (g), noise removed by the  $f$ - $x$ - $y$  SPF (h).

filter size outputs a smoother result and the lateral continuity is improved (Fig. 10a), where the reflection event in both shallow and deep parts becomes clearer than original field data. Fig. 10b displays that only parts of dipping events are slightly lost. The proposed  $f$ - $x$ - $y$  SPF can produce more feasible results than the 2D version because the 3D SPF has extra nonstationarity along the  $y$  axis (Fig. 11a), where the continuity of the events and the geological structure are enhanced. The scale parameters are selected as 90 ( $\lambda_f$ ), 600 ( $\lambda_x$ ), and 600 ( $\lambda_y$ ) and the same 5-sample (X)  $\times$  5-sample (Y) filter size as the 3D  $f$ - $x$ - $y$  RNA. The difference (Fig. 11b) between the noisy data (Fig. 7) and the denoised result (Fig. 11a) shows more uniformly-distributed random noise than the 3D  $f$ - $x$ - $y$  RNA, which demonstrates that the proposed filter is able to depict the variations in the nonstationary signals and provide an accurate estimation of complex wavefields even in the presence of strongly curved and conflicting events. Furthermore, the  $f$ - $x$ - $y$  SPF takes relatively low computational time (Table 2) that highlights its efficiency, especially in higher dimensions.

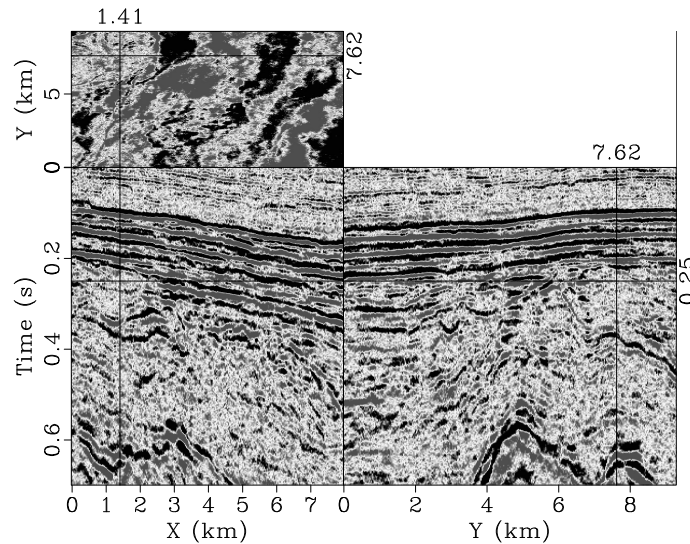


Figure 7: Three-dimensional field data.

## DISCUSSIONS

The frequency-space PFs shows different characteristics from the time-space PFs that depend on the sample rate along time direction. PFs in time-space domain cannot predict the seismic events from all directions (with all kinds of slopes) because the whole trace length is seldom chosen as the filter size along the time direction. The frequency method can avoid this problem, and it may get an accuracy prediction from the complex seismic data. However, computation in frequency domain may result in artificial effect, so we designed the local smoothness constraint to reduce the effect.

We highlighted the computational efficiency of the proposed  $f$ - $x$ - $y$  SPF because of not only the noniterative property but also the memory saving ability. Assuming

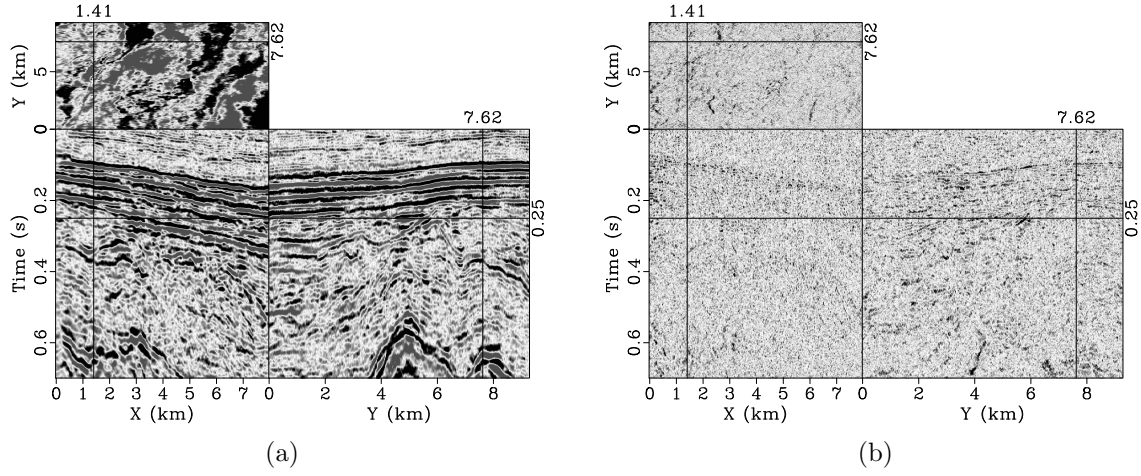


Figure 8: The denoised result by using the 2D  $f$ - $x$  SPF (a) and the difference between the noisy data (Fig. 7) and the denoised result by using the 2D  $f$ - $x$  SPF (Fig. 8a) (b).

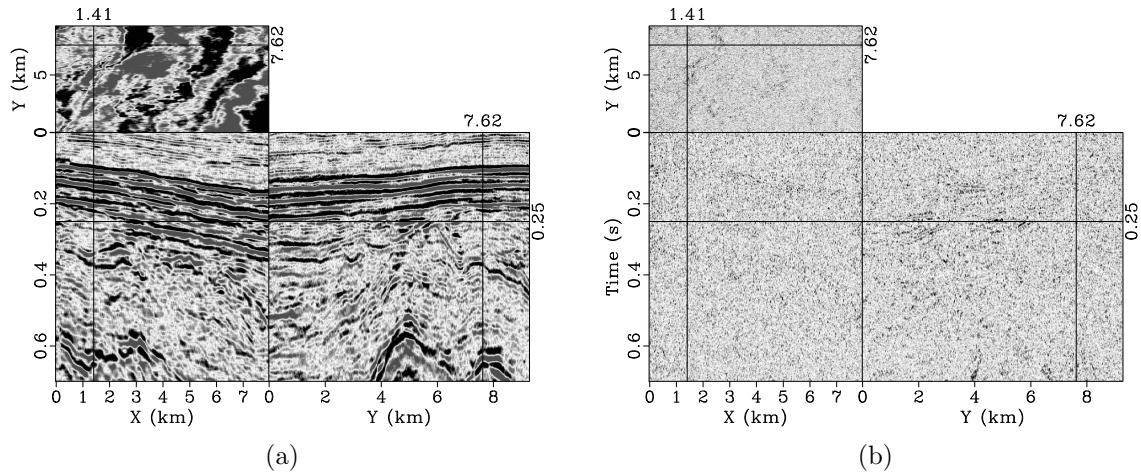


Figure 9: The denoised result by using the  $f$ - $x$  EMD prediction filter (a) and the difference between the noisy data (Fig. 7) and the denoised result by using the  $f$ - $x$  EMD prediction filter (Fig. 9a) (b).

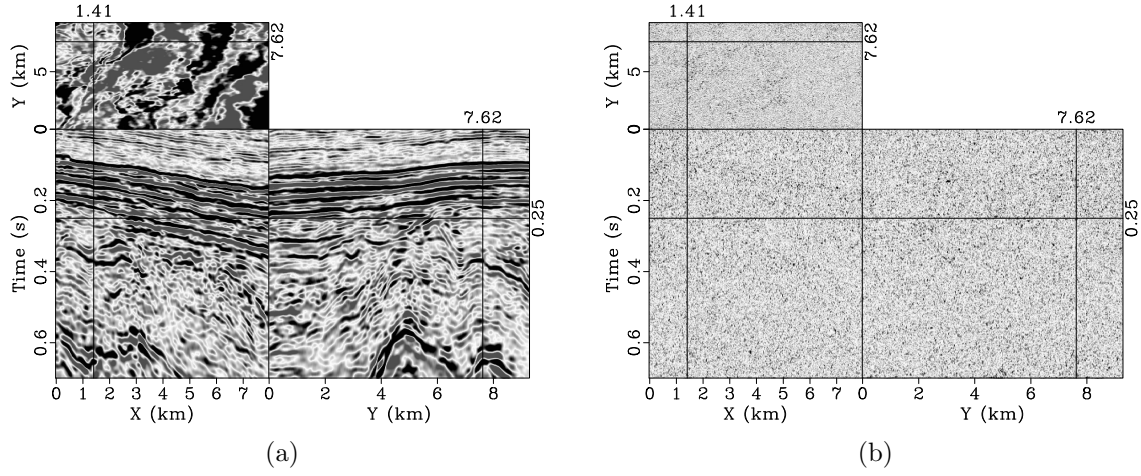


Figure 10: The denoised result by using the 3D  $f$ - $x$ - $y$  RNA (a) and the difference between the noisy data (Fig. 7) and the denoised result by using the 3D  $f$ - $x$ - $y$  RNA (Fig. 10a) (b).

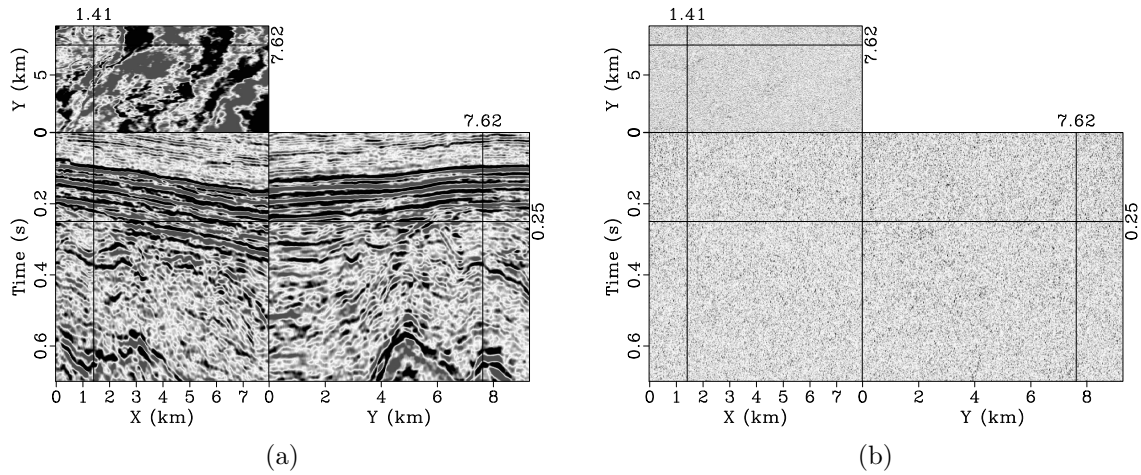


Figure 11: The denoised result by using the 3D  $f$ - $x$ - $y$  SPF (a) and the difference between the noisy data (Fig. 7) and the denoised result by using the 3D  $f$ - $x$ - $y$  SPF (Fig. 11a) (b).

that a conventional 3D data volume  $500$  (frequency)  $\times 500$  (space  $x$ )  $\times 500$  (space  $y$ ) is about 1 gigabyte with complex number format, then the  $f$ - $x$ - $y$  RNA with 5-sample (X)  $\times$  5-sample (Y) filter structure needs over 23 gigabytes memory space to cache the filter coefficients for data calculation, and the data exchange time is much more than that of iterative calculation. The proposed  $f$ - $x$ - $y$  SPF needs only about 50 megabytes to cache filter coefficients when calculating.

## CONCLUSIONS

In this study, we introduced a fast approach to nonstationary prediction filter for random noise attenuation in the 3D  $f$ - $x$ - $y$  domain. The proposed method employs a local similarity to constrain the autoregression equation for nonstationary prediction filter in the frequency-space domain, which belongs to the streaming prediction theory. Constrained conditions in the 3D frequency-space dimensions guarantee the accurate estimation of adaptive prediction filters and reasonable prediction of complex structures. Instead of using an iterative strategy, the new analytical solution in the frequency domain for the least-squares problem allows the proposed method to reduce computational complexity significantly. The matching snake processing path further improves the signal recovery ability of the three-dimensional SPF. Although the  $f$ - $x$ - $y$  SPF shows similar accuracy to the  $f$ - $x$ - $y$  RNA, the proposed method allows us to better balance the target event protection, random noise suppression, and computational efficiency. Numerical examples using synthetic models and field data show that the  $f$ - $x$ - $y$  SPF can effectively attenuate random noise and protect valid information in the nonstationary seismic data. The proposed method is superior in terms of its low computational cost even when analyzing large-scale seismic data.

## ACKNOWLEDGEMENTS

This work is supported by National Natural Science Foundation of China (grant nos. 41774127, 41974134 and 41522404) and National Key Research and Development Program of China (grant no. 2018YFC0603701).

## SHERMAN-MORRISON FORMULA IN THE COMPLEX SPACE

To get the inverse of  $(\lambda^2 \mathbf{I} + \mathbf{S}_{m,n,l}^* \mathbf{S}_{m,n,l}^T)$ , one can implement Sherman-Morrison formula to transform the inverse matrix as (13). Whereas  $\mathbf{S}_{m,n,l}^*$  and  $\mathbf{S}_{m,n,l}^T$  are

complex vectors. Here, we drop the subscript of vectors for a concise proof:

$$\begin{aligned}
& (\lambda^2 \mathbf{I} + \mathbf{S}^* \mathbf{S}^T) \frac{1}{\lambda^2} \left( \mathbf{I} - \frac{\mathbf{S}^* \mathbf{S}^T}{\lambda^2 + \mathbf{S}^T \mathbf{S}^*} \right) \\
&= \mathbf{I} - \frac{\mathbf{S}^* \mathbf{S}^T}{\lambda^2 + \mathbf{S}^T \mathbf{S}^*} + \frac{1}{\lambda^2} \mathbf{S}^* \mathbf{S}^T - \frac{1}{\lambda^2} \frac{\mathbf{S}^* \mathbf{S}^T \mathbf{S}^* \mathbf{S}^T}{\lambda^2 + \mathbf{S}^T \mathbf{S}^*} \\
&= \mathbf{I} + \frac{1}{\lambda^2} \frac{\lambda^2 \mathbf{S}^* \mathbf{S}^T + \mathbf{S}^T \mathbf{S}^* \mathbf{S}^* \mathbf{S}^T}{\lambda^2 + \mathbf{S}^T \mathbf{S}^*} - \frac{\mathbf{S}^* \mathbf{S}^T}{\lambda^2 + \mathbf{S}^T \mathbf{S}^*} \\
&\quad - \frac{1}{\lambda^2} \frac{\mathbf{S}^* \mathbf{S}^T \mathbf{S}^* \mathbf{S}^T}{\lambda^2 + \mathbf{S}^T \mathbf{S}^*} \\
&= \mathbf{I} + \frac{\mathbf{S}^* \mathbf{S}^T}{\lambda^2 + \mathbf{S}^T \mathbf{S}^*} + \frac{1}{\lambda^2} \frac{\mathbf{S}^T \mathbf{S}^* \mathbf{S}^* \mathbf{S}^T}{\lambda^2 + \mathbf{S}^T \mathbf{S}^*} - \frac{\mathbf{S}^* \mathbf{S}^T}{\lambda^2 + \mathbf{S}^T \mathbf{S}^*} \\
&\quad - \frac{1}{\lambda^2} \frac{\mathbf{S}^* \mathbf{S}^T \mathbf{S}^* \mathbf{S}^T}{\lambda^2 + \mathbf{S}^T \mathbf{S}^*} \\
&= \mathbf{I},
\end{aligned}$$

where  $\mathbf{I}$  denotes the identity matrix,  $\mathbf{S}$  is the complex column vector,  $\mathbf{S}^T$  is the transpose of  $\mathbf{S}$ ,  $\mathbf{S}^*$  is the conjugation of  $\mathbf{S}$ . Therefore,  $\mathbf{S}^* \mathbf{S}^T$  is a complex matrix, and  $\mathbf{S}^T \mathbf{S}^*$  is constant. Therefore, we prove that Sherman-Morrison formula can be applied in the complex space.

## REFERENCES

- Abma, R., and J. F. Claerbout, 1995, Lateral prediction for noise attenuation by t-x and f-x techniques: *Geophysics*, **60**, 1887–1896.
- Berkner, K., and R. O. Wells, 1998, Wavelet transforms and denoising algorithms: *Proc. Conf. Rec. 32nd Asilomar Conf. Signals Syst. Comput.*, 1639–1643.
- Bonar, D., and M. D. Sachhi, 2012, Denoising seismic data using the nonlocal means algorithm: *Geophysics*, **77**, A5–A8.
- Canales, L. L., 1984, Random noise reduction: *Proc. 53rd Annu. Int. Meet., SEG, Expanded Abstracts*, 525–527.
- Chen, Y., and J. Ma, 2014, Random noise attenuation by  $f$ - $x$  empirical mode decomposition predictive filtering: *Geophysics*, **79**, V81–V91.
- Claerbout, J. F., 1976, *Fundamentals of geophysical data processing*: McGraw-Hill.
- Claerbout, J. F., and S. Fomel, 2008, *Image estimation by example: Geophysical soundings image construction: Multidimensional autoregression*: Stanford University.
- Djarfour, N., J. Ferahtia, F. Babaia, K. Baddari, E. A. Said, and M. Farfour, 2014, Seismic noise filtering based on generalized regression neural networks: *Computers & Geosciences*, **69**, 1–9.
- Fomel, S., and J. Claerbout, 2016, Streaming prediction-error filters: *Proc. 85th Annu. Int. Meet., SEG, Expanded Abstracts*, 4787–4791.
- Fomel, S., and Y. Liu, 2010, Seislet transform and seislet frame: *Geophysics*, **75**, V25–V38.

- Gülünay, N., 1986, Fxdecon and complex wiener prediction filter: Proc. 56th Annu. Int. Meet., SEG, Expanded Abstracts, 279–281.
- Guo, L., C. Liu, Y. Liu, Z. Zheng, and Q. Wang, 2020, Seismic random noise attenuation based on streaming prediction filter in the f-x domain: Chinese Journal of Geophysics, **63**, 329–338.
- Hager, W. W., 1989, Updating the inverse of a matrix: SIAM review, **31**, 221–239.
- Kimiaefar, R., H. R. Siahkoobi, A. R. Hajian, and A. Kalhor, 2016, Seismic random noise attenuation using artificial neural network and wavelet packet analysis: Arabian Journal of Geosciences, **9**, 234.
- Langston, C. A., and S. M. Mousavi, 2019, Separating signal from noise and from other signal using nonlinear thresholding and scaletime windowing of continuous wavelet transforms: Bulletin of the Seismological Society of America, **109**, 1691–1700.
- Liu, G., and X. Chen, 2013, Noncausal f-x-y regularized nonstationary prediction filtering for random noise attenuation on 3D seismic data: Journal of Applied Geophysics, **93**, 60–66.
- Liu, G., X. Chen, J. Du, and K. Wu, 2012, Random noise attenuation using f-x regularized nonstationary autoregression: Geophysics, **77**, V61–V69.
- Liu, Y., and S. Fomel, 2010, OC-seislet: Seislet transform construction with differential offset continuation: Geophysics, **75**, WB235–WB245.
- Liu, Y., S. Fomel, and C. Liu, 2015, Signal and noise separation in prestack seismic data using velocity-dependent seislet transform: Geophysics, **80**, WD117–WD128.
- Liu, Y., and B. Li, 2018, Streaming orthogonal prediction filter in t-x domain for random noise attenuation: Geophysics, **83**, F41–F48.
- Liu, Y., C. Liu, and D. Wang, 2009, A 1D time-varying median filter for seismic random spike-like noise elimination: Geophysics, **74**, V17–V24.
- Sacchi, M. D., and M. Naghizadeh, 2009, Adaptive linear prediction filtering for random noise attenuation: Proc. 79th Annu. Int. Meet., SEG, Expanded Abstracts, 3347–3351.
- Wu, S., Y. Wang, Z. Di, and X. Chang, 2018, Random noise attenuation by 3D Multi-directional vector median filter: Journal of Applied Geophysics, **159**, 277–284.
- Yilmaz, Ö., 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: Society of Exploration Geophysicists.
- Yu, S., J. Ma, and W. Wang, 2019, Deep learning for denoising: Geophysics, **84**, V333–V350.
- Yu, Z., J. Ferguson, G. McMechan, and P. Anno, 2007, Wavelet-Radon domain dealiasing and interpolation of seismic data: Geophysics, **72**, V41–V49.
- Zhu, W., S. M. Mousavi, and G. C. Beroza, 2019, Seismic signal denoising and decomposition using deep neural networks: IEEE Transactions on Geoscience and Remote Sensing, **57**, 9476–9488.