Seismic data interpolation using streaming prediction filter in the frequency domain[a]

[a]Published in Geophysical Journal International, 229, 370-389, (2022)

Zhisheng Zheng, Yang Liu*, Cai Liu**

ABSTRACT

Surface conditions and economic factors restrict field geometries, so seismic data acquisition typically obtains field data with irregular spatial distribution, which can adversely affect the subsequent data processing and interpretation. Therefore, data interpolation techniques are used to convert field data into regularly distributed data and reconstruct the missing traces. Recently, the mainstream methods have implemented iterative algorithms to solve data interpolation problems, which require substantial computational resources and restrict their application in high dimensions. In this study, we proposed the f-x and f-x-y streaming prediction filters (SPFs) to reconstruct missing seismic traces without iterations. According to the streaming computation framework, we directly derived an analytic solution to the overdetermined least-squares problem with local smoothness constraints for estimating SPFs in the frequency domain. We introduced different processing paths and filter forms to reduce the interference of missing traces, which can improve the accuracy of filter coefficients. Meanwhile, we utilized a two-step interpolation strategy to guarantee the effective interpolation of the irregularly missing traces. Numerical examples show that the proposed methods effectively recover the missing traces in seismic data when compared with the traditional Fourier Projection Onto Convex Sets (POCS) method. In particular, the frequency domain SPFs are suitable for high-dimensional seismic data interpolation with the advantages of low computational cost and reasonable nonstationary signal reconstruction.

INTRODUCTION

With the limitation of field geometries, uniform data acquisition is rarely achieved in practice; instead, it always displays irregular or undersampled data distribution in spatial directions. However, many subsequent processing steps, such as multiple elimination and migration, are based on the prerequisite of regular data distribution. Seismic data interpolation has become a key technique in seismic data processing workflows. A failed interpolation method may create artifacts, which affect the accuracy of seismic imaging. Recently, high-density and wide-azimuth seismic acquisition can achieve large-scale field data; however, increasing computational costs

and nonstationary signal recovery have posed challenges to data interpolation. Many interpolation methods have been proposed to reconstruct missing data, which are based on signal processing or seismic kinematic/dynamic principles. Low-rank methods (Trickett et al., 2010; Gao et al., 2013; Chen et al., 2016; Gao et al., 2017) assume that seismic data are low rank in the mapping domain, and rank-reduction operators are used to recover the missing data. Plane wave decomposition (Fomel, 2002; Hellman and Boyer, 2016) reconstructs missing data by using local slope information. Compressive sensing framework with different sparse domains, e.g., Fourier transform (Abma and Kabir, 2006; Wang et al., 2010a) and seislet transform (Fomel and Liu, 2010; Liu and Fomel, 2010; Gao et al., 2015), iteratively recover the missing traces. Machine learning has become a popular research direction, and is also used for data interpolation (Jia and Ma, 2017; Oliveira et al., 2018; Mandelli et al., 2019; Kaur et al., 2019; Zhang et al., 2020). However, these methods always encounter high computational cost.

Prediction filter (PF) provide an important approach for seismic data interpolation. Spitz (1991) proved that high-frequency components can be estimated with the PF calculated from low-frequency components, and an f-x interpolation method beyond aliasing was proposed. Sacchi and Ulrych (1997) used an autoregression moving average (ARMA) model to calculate the PF and interpolated near-offset missing gaps in the f-x domain. Porsani (1999) proposed a half-step PF to efficiently interpolate missing data. Wang (2002) extended different kinds of PFs in high dimensions to implement the f-x interpolation algorithm. Abma and Kabir (2005) made comparison of several PF interpolation methods. Naghizadeh and Sacchi (2008) used an exponentially weighted recursive least squares (EWRLS) for the adaptive PF to interpolate data in the f-x domain. Wang et al. (2010b) created the virtual traces from marine seismic data and utilized the matching filter or nonstationary prediction error filter (PEF) to fill gaps. Liu and Fomel (2011) proposed an approach to interpolate aliased data based on adaptive PEF and regularized nonstationary autoregression (RNA) in the t-x-y domain. Li et al. (2017) proposed multidimensional adaptive PEF to reconstruct seismic data in the frequency domain. Liu and Chen (2018) introduced an efficient method based on the f-x RNA for regular and irregular missing data interpolation in the f-x domain. Liu et al. (2019) designed the multiscale and multidirectional PEF to improve the accuracy of filter coefficients while reconstructing seismic data.

It is difficult to balance the computational cost and the interpolation accuracy while dealing with large-scale data in most iterative methods. However, the framework of the streaming computation (Sacchi and Naghizadeh, 2009; Fomel and Claerbout, 2016) can directly solve the nonstationary autoregression problem. Liu and Li (2018) proposed an orthogonal t-x streaming PF (SOPF) for random noise attenuation, and Guo et al. (2020) provided an initial idea for the f-x SPF with a 1D spatial constraint and tried to efficiently eliminate seismic random noise. Therefore, we further improved the noniterative framework of the streaming computation in the frequency domain and used the SPF with new frequency-domain constraints for seismic data interpolation. The new adaptive PF can reduce the computational cost in

the interpolation problem while capturing details of nonstationary seismic data. In this study, we first discussed the two-step interpolation strategy for PF. With the two-step strategy, we derived the theory of the new f-x SPF and f-x-y SPF based on the streaming framework. The relevant interpolation algorithm and filter design were developed to help the SPF solve the problem of the nonstationary data reconstruction with a feasible computational cost. Numerical examples demonstrate the effectiveness and efficiency of the proposed methods.

THEORY

Two-step interpolation strategy

The classical method for restoring missing data is to ensure that the restored data, after specified filtering, have minimum energy (Claerbout, 1992). Data interpolation by prediction filtering can be divided into two problems (Liu and Fomel, 2011): filter estimation and data reconstruction with filters. The first step is to calculate the filter by minimizing the following autoregression problem:

$$\min_{\mathbf{F}} \|\mathbf{g} - \mathbf{G}^{\mathsf{T}}\mathbf{F}\|_2^2, \tag{1}$$

where $\{\bullet\}^{\mathsf{T}}$ denotes the transpose operator, $\mathbf{F}$ is adaptive PF, $\mathbf{g}$ is the data matrix, and $\mathbf{G}$ is the translation of $\mathbf{g}$. In the second step, the missing data are restored by minimizing the least-squares problem:

$$\min_{\hat{\mathbf{g}}} \|\hat{\mathbf{g}} - \mathbf{G}^{\mathsf{T}}\mathbf{F}\|_2^2, \tag{2}$$

where $\hat{\mathbf{g}}$ is the reconstructed data.

In the time-space or frequency-space domain, Eq. (1) and (2), established by the prediction filtering theory, are designed to solve different variables, and they are solved individually. The number of unknown filter coefficients in adaptive PF $\mathbf{F}$ is often greater than the number of known data in $\mathbf{g}$, in other words, the number of unknowns is greater than the number of equations. Eq. (1) usually presents an ill-posed problem, and different regularization terms have been introduced to stabilize the filter solution $\mathbf{F}$. Further discussion will be provided in the next section. Meanwhile, we extended the framework of the streaming computation in the frequency domain and proposed a local and multidimensional smoothness to constrain Eq. (1) in the frequency domain.

f-x streaming prediction filter

In the f-x domain, the PF predicts the data along the spatial direction, and the relationship between data and prediction filter can be summarized as

$$\begin{aligned} g_{m,n} &\approx \sum_p g_{m,n-p} f_{m,n,p} \\ &\approx \mathbf{G}_{m,n}^{\mathsf{T}} \mathbf{F}_{m,n}, \end{aligned} \tag{3}$$

where m and n are the indices of the seismic data sample $g_{m,n}$ along the frequency f axis and space x axis, respectively. Vector $\mathbf{F}_{m,n}$ is the group of filter coefficients $f_{m,n,p}$ in the adaptive prediction filter, and each group $\mathbf{F}_{m,n}$ corresponds to a data sample $g_{m,n}$. Vector $\mathbf{G}_{m,n}$ denotes several data points $g_{m,n-p}$ with spatial shift p near $g_{m,n}$. Spatial shift p is related to the filter size (the number of filter coefficients in $\mathbf{F}_{m,n}$), and the filter size should theoretically be larger than or equal to the number of seismic events contained in the local space window. When the causal filter structure is considered, as shown in Fig. 1a, the spatial shift is chosen as $p \in [1,3]$, vector $\mathbf{G}_{m,n}$ can be expressed as $\mathbf{G}_{m,n} = \{g_{m,n-1}, g_{m,n-2}, g_{m,n-3}\}$, and $\mathbf{F}_{m,n} = \{f_{m,n,1}, f_{m,n,2}, f_{m,n,3}\}$. For the non-causal filter structure (Fig. 1b), the spatial shift is $p \in [-3,-1] \cup [1,3]$, and vector $\mathbf{G}_{m,n}$ and $\mathbf{F}_{m,n}$ can be represented as $\mathbf{G}_{m,n} = \{g_{m,n+3}, g_{m,n+2}, g_{m,n+1}, g_{m,n-1}, g_{m,n-2}, g_{m,n-3}\}$, $\mathbf{F}_{m,n} = \{f_{m,n,-3}, f_{m,n,-2}, f_{m,n,-1}, f_{m,n,1}, f_{m,n,2}, f_{m,n,3}\}$.

With Eq. (3), we established the following minimization problem, like Eq. (1), to calculate filter $\mathbf{F}_{m,n}$:

$$\min_{\mathbf{F}_{m,n}} \|g_{m,n} - \mathbf{G}_{m,n}^{\mathsf{T}} \mathbf{F}_{m,n}\|_2^2. \tag{4}$$

According to the above explanation of vector $\mathbf{F}_{m,n}$, there are several unknown filter coefficients, yet we established only one equation. Eq. (4) describes an ill-posed problem, which requires constraints to obtain a stable solution. The framework of the streaming computation (Sacchi and Naghizadeh, 2009; Fomel and Claerbout, 2016; Liu and Li, 2018) establishes the constraint relationship by using local smoothness. We extended this method into the frequency domain, including the f-x domain and the f-x-y domain (the next section) to stabilize the filter coefficient solution. Additionally, we discussed in detail the effects of filter structure and processing path, and provided the corresponding interpolation algorithms. Here, multiple constraints based on local smoothness are used to constrain the solution of Eq. (4):

$$\begin{aligned} \min_{\mathbf{F}_{m,n}} \|g_{m,n} - \mathbf{G}_{m,n}^{\mathsf{T}} \mathbf{F}_{m,n}\|_2^2 + \lambda_f^2 \|\mathbf{F}_{m,n} - \mathbf{F}_{m-1,n}\|_2^2 \\ + \lambda_x^2 \|\mathbf{F}_{m,n} - \mathbf{F}_{m,n-1}\|_2^2, \end{aligned} \tag{5}$$

where λ_f and λ_x are the weights of the regularization terms along the frequency f and space x axes. $\lambda_f^2 \|\mathbf{F}_{m,n} - \mathbf{F}_{m-1,n}\|_2^2$ shows the local smoothness along the frequency direction ($\lambda_f \mathbf{F}_{m,n} \approx \lambda_f \mathbf{F}_{m-1,n}$). Likewise, $\lambda_x^2 \|\mathbf{F}_{m,n} - \mathbf{F}_{m,n-1}\|_2^2$ controls the local smoothness along the space direction as $\lambda_x \mathbf{F}_{m,n} \approx \lambda_x \mathbf{F}_{m,n-1}$. The block matrix Eq. (6) has the same solution as Eq. (5), and it demonstrates the effect of the local smoothness constraints; when $\mathbf{F}_{m-1,n}$ and $\mathbf{F}_{m,n-1}$ are considered known, the local smoothness conditions ($\lambda_f \mathbf{F}_{m,n} \approx \lambda_f \mathbf{F}_{m-1,n}$ and $\lambda_x \mathbf{F}_{m,n} \approx \lambda_x \mathbf{F}_{m,n-1}$), as newly added equations, can be used to stabilize the solution of $\mathbf{F}_{m,n}$:

$$\begin{bmatrix} \mathbf{G}_{m,n}^{\mathsf{T}} \\ \lambda_f \mathbf{I} \\ \lambda_x \mathbf{I} \end{bmatrix} \mathbf{F}_{m,n} \approx \begin{bmatrix} g_{m,n} \\ \lambda_f \mathbf{F}_{m-1,n} \\ \lambda_x \mathbf{F}_{m,n-1} \end{bmatrix}. \tag{6}$$

One can obtain the following least-squares solution of Eq. (5) and (6):

$$\begin{aligned}\mathbf{F}_{m,n} =&(\lambda_f^2\mathbf{I}+\lambda_x^2\mathbf{I}+\mathbf{G}_{m,n}^{*}\mathbf{G}_{m,n}^{\mathsf{T}})^{-1}\\ &(g_{m,n}\mathbf{G}_{m,n}^{*}+\lambda_f^2\mathbf{F}_{m-1,n}+\lambda_x^2\mathbf{F}_{m,n-1}),\end{aligned} \tag{7}$$

where $\{\bullet\}^{*}$ denotes the conjugate operator. Let

$$\begin{cases}\lambda^2=\lambda_f^2+\lambda_x^2\\ \lambda^2\tilde{\mathbf{F}}_{m,n}=\lambda_f^2\mathbf{F}_{m-1,n}+\lambda_x^2\mathbf{F}_{m,n-1}\end{cases}, \tag{8}$$

we get a simplified equation:

$$\mathbf{F}_{m,n}=(\lambda^2\mathbf{I}+\mathbf{G}_{m,n}^{*}\mathbf{G}_{m,n}^{\mathsf{T}})^{-1}(g_{m,n}\mathbf{G}_{m,n}^{*}+\lambda^2\tilde{\mathbf{F}}_{m,n}). \tag{9}$$

Meanwhile, $(\lambda^2\mathbf{I}+\mathbf{G}_{m,n}^{*}\mathbf{G}_{m,n}^{\mathsf{T}})$ has the analytical inversion $\frac{1}{\lambda^2}(\mathbf{I}-\frac{\mathbf{G}_{m,n}^{*}\mathbf{G}_{m,n}^{\mathsf{T}}}{\lambda^2+\mathbf{G}_{m,n}^{\mathsf{T}}\mathbf{G}_{m,n}^{*}})$ when one extends the Sherman-Morrison formula (Sherman and Morrison, 1950; Bartlett, 1951; Hager, 1989) to the complex space. We can obtain the analytical solution of the f-x SPF:

$$\begin{aligned}\mathbf{F}_{m,n}&=\frac{1}{\lambda^2}(\mathbf{I}-\frac{\mathbf{G}_{m,n}^{*}\mathbf{G}_{m,n}^{\mathsf{T}}}{\lambda^2+\mathbf{G}_{m,n}^{\mathsf{T}}\mathbf{G}_{m,n}^{*}})(g_{m,n}\mathbf{G}_{m,n}^{*}+\lambda^2\tilde{\mathbf{F}}_{m,n})\\ &=\tilde{\mathbf{F}}_{m,n}+\frac{g_{m,n}-\mathbf{G}_{m,n}^{\mathsf{T}}\tilde{\mathbf{F}}_{m,n}}{\lambda^2+\mathbf{G}_{m,n}^{\mathsf{T}}\mathbf{G}_{m,n}^{*}}\mathbf{G}_{m,n}^{*}.\end{aligned} \tag{10}$$

Eq. (10) shows a recursive relationship from the previous filters ($\mathbf{F}_{m-1,n}$ and $\mathbf{F}_{m,n-1}$) to current filter $\mathbf{F}_{m,n}$. For data interpolation, Eq. (2) becomes a well-posed problem, and the unknown data sample can be calculated by

$$\hat{g}_{m,n}=\mathbf{G}_{m,n}^{\mathsf{T}}\mathbf{F}_{m,n}. \tag{11}$$

In 2D data interpolation, we proposed the following Algorithm 1 to reconstruct the missing seismic data by using the 2D f-x SPF. To start, $\mathbf{F}_{m-1,n}$ and $\mathbf{F}_{m,n-1}$ are initialized to $\mathbf{0}$. When the processing path in Algorithm 1 is followed, both $\mathbf{F}_{m-1,n}$ and $\mathbf{F}_{m,n-1}$ are known. We designed a space-causal filter form (Fig. 1a) for the f-x SPF. Fig. 1b suggests that the space-noncausal form may involve the interference of the unknown data samples, but the causal one can avoid this problem.

f-x-y streaming prediction filter

The extension to the 3D f-x-y domain is straightforward as the f-x-y SPF can also efficiently perform data interpolation in high dimensions. In the f-x-y domain, the prediction relationship for seismic data in a certain frequency slice is expressed as

$$\begin{aligned}g_{m,n,l}&\approx\sum_p\sum_q g_{m,n-p,l-q}f_{m,n,l,p,q}\\ &\approx\mathbf{G}_{m,n,l}^{\mathsf{T}}\mathbf{F}_{m,n,l},\end{aligned} \tag{12}$$

Algorithm 1 f-x SPF interpolation algorithm

Input: $g_{m,n}$, $\mathbf{G}_{m,n}$, λ_f, and λ_x;
Output: $\hat{g}_{m,n}$;

```
F_{m-1,n} = F_{m,n-1} = 0
for loop over frequency direction do
    for loop over space direction do
        if encounter known data point then
            F_{m,n} = F̃_{m,n} + (g_{m,n} − G^T_{m,n} F̃_{m,n}) / (λ² + G^T_{m,n} G*_{m,n}) G*_{m,n}
        else if encounter missing data point then
            ĝ_{m,n} = G^T_{m,n} F_{m,n}
        end if
    end for
end for
```

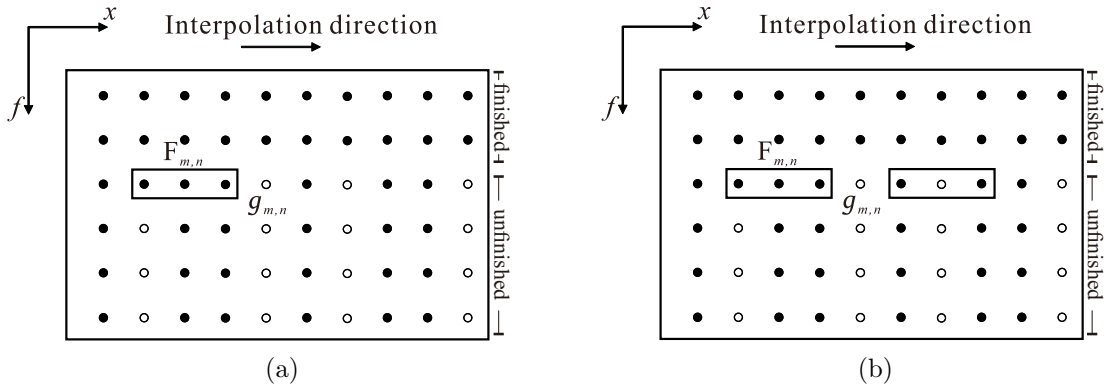


Figure 1: Space-causal filter form (a) and space-noncausal filter form (b) in the f-x domain. Solid circle denotes known sample, and hollow circle denotes unknown sample.

where m, n, and l denote the indices of data sample $g_{m,n,l}$ along the f, x, and y axes, respectively. Vector $\mathbf{F}_{m,n,l}$ is a group of filter coefficients in the f-x-y SPF, vector $\mathbf{G}_{m,n,l}$ contains the data points with spatial shifts p (x axis) and q (y axis) near $g_{m,n,l}$. Similar to the case of the f-x SPF, the range of spatial shifts p and q depends on the number of seismic events in the local window of space x and space y. As shown in Fig. 2, when $p \in \{-1, 1\}$ and $q \in \{1, 3\}$, vector $\mathbf{G}_{m,n,l}$ can be flattened as $\mathbf{G}_{m,n,l} = \{g_{m,n+1,l-1}, g_{m,n+1,l-2}, \cdots, g_{m,n-1,l-3}\}$, and $\mathbf{F}_{m,n,l} = \{f_{m,n,l,-1,1}, f_{m,n,l,-1,2}, \cdots, f_{m,n,l,1,3}\}$. For the space-noncausal case, $p \in \{-1, 1\}$, $q \in \{-3, 3\}$, and $|p| + |q| \neq 0$, $\mathbf{G}_{m,n,l}$ can be expressed as follows: $\mathbf{G}_{m,n,l} = \{g_{m,n+1,l+3}$, $g_{m,n+1,l+2}, \cdots, g_{m,n,l+1}, g_{m,n,l-1}, \cdots, g_{m,n-1,l-3}\}$, and $\mathbf{F}_{m,n,l} = \{f_{m,n,l,-1,-3}, f_{m,n,l,-1,-2}$, $\cdots, f_{m,n,l,0,-1}, f_{m,n,l,0,1}, \cdots, f_{m,n,l,1,3}\}$.

The cost function is defined as

$$J_{m,n,l} = \|g_{m,n,l} - \mathbf{G}_{m,n,l}^{\mathsf{T}} \mathbf{F}_{m,n,l}\|_2^2. \tag{13}$$

Assuming that the adaptive PF at position (m, n, l) is similar to another near positions $(m-1, n, l)$, $(m, n-1, l)$, and $(m, n, l-1)$ in the f-x-y domain, we can minimize the new cost function Eq. (14) to solve the overdetermined problem:

$$\begin{aligned} \hat{J}_{m,n,l} = J_{m,n,l} &+ \lambda_f^2 \|\mathbf{F}_{m,n,l} - \mathbf{F}_{m-1,n,l}\|_2^2 \\ &+ \lambda_x^2 \|\mathbf{F}_{m,n,l} - \mathbf{F}_{m,n-1,l}\|_2^2 \\ &+ \lambda_y^2 \|\mathbf{F}_{m,n,l} - \mathbf{F}_{m,n,l-1}\|_2^2, \end{aligned} \tag{14}$$

where λ_f, λ_x, and λ_y are the constant weights of the regularization terms, which describe the local smoothness constraints along different directions. The simplified block matrix can be written as

$$\begin{bmatrix} \mathbf{G}_{m,n,l}^{\mathsf{T}} \\ \lambda_f \mathbf{I} \\ \lambda_x \mathbf{I} \\ \lambda_y \mathbf{I} \end{bmatrix} \mathbf{F}_{m,n,l} \approx \begin{bmatrix} g_{m,n,l} \\ \lambda_f \mathbf{F}_{m-1,n,l} \\ \lambda_x \mathbf{F}_{m,n-1,l} \\ \lambda_y \mathbf{F}_{m,n,l-1} \end{bmatrix}. \tag{15}$$

Similar to the 2D case, the analytic solution of the f-x-y SPF is given by

$$\mathbf{F}_{m,n,l} = \tilde{\mathbf{F}}_{m,n,l} + \frac{g_{m,n,l} - \mathbf{G}_{m,n,l}^{\mathsf{T}} \tilde{\mathbf{F}}_{m,n,l}}{\lambda^2 + \mathbf{G}_{m,n,l}^{\mathsf{T}} \mathbf{G}_{m,n,l}^{*}} \mathbf{G}_{m,n,l}^{*}, \tag{16}$$

where

$$\begin{cases} \lambda^2 = \lambda_f^2 + \lambda_x^2 + \lambda_y^2 \\ \lambda^2 \tilde{\mathbf{F}}_{m,n,l} = \lambda_f^2 \mathbf{F}_{m-1,n,l} + \lambda_x^2 \mathbf{F}_{m,n-1,l} + \lambda_y^2 \mathbf{F}_{m,n,l-1} \end{cases}. \tag{17}$$

Furthermore, the unknown data sample in the 3D data cube can be calculated as

$$\hat{g}_{m,n,l} = \mathbf{G}_{m,n,l}^{\mathsf{T}} \mathbf{F}_{m,n,l}. \tag{18}$$

Because the f-x-y SPF predicts data along two spatial directions, we defined a new processing path in the space plane with zigzag shape (Fig. 2a) to prevent unnecessary

filter initialization. Meanwhile, the f-x-y SPF was assigned to the proposed filter form (Fig. 2a), which can better reduce the influence of unknown data samples than the noncausal filter in spatial directions (Fig. 2b). Following the processing path of Algorithm 2 for data processing, $\mathbf{F}_{m-1,n,l}$, $\mathbf{F}_{m,n-1,l}$, and $\mathbf{F}_{m,n,l-1}$ can be seen as known, requiring only calculating Eq. (16) and (18) to obtain the results.

Algorithm 2 f-x-y SPF interpolation algorithm

Input: $g_{m,n,l}$, $\mathbf{G}_{m,n,l}$, λ_f, λ_x, and λ_y;
Output: $\hat{g}_{m,n,l}$;

$\mathbf{F}_{m-1,n,l} = \mathbf{F}_{m,n-1,l} = \mathbf{F}_{m,n,l-1} = \mathbf{0}$
for loop over frequency direction **do**
  for loop over space y direction **do**
    if space y == even number **then**
      for forward loop over space x direction **do**
        if encounter known data point **then**
          $\mathbf{F}_{m,n,l} = \tilde{\mathbf{F}}_{m,n,l} + \frac{g_{m,n,l} - \mathbf{G}^{\mathsf{T}}_{m,n,l}\tilde{\mathbf{F}}_{m,n,l}}{\lambda^2 + \mathbf{G^T}_{m,n,l}\mathbf{G^*}_{m,n,l}}\mathbf{G}^*_{m,n,l}$
        else if encounter missing data point **then**
          $\hat{g}_{m,n,l} = \mathbf{G}^{\mathsf{T}}_{m,n,l}\mathbf{F}_{m,n,l}$
        end if
      end for
    else if space y == odd number **then**
      for backward loop over space x direction **do**
        if encounter known data point **then**
          $\mathbf{F}_{m,n,l} = \tilde{\mathbf{F}}_{m,n,l} + \frac{g_{m,n,l} - \mathbf{G}^{\mathsf{T}}_{m,n,l}\tilde{\mathbf{F}}_{m,n,l}}{\lambda^2 + \mathbf{G^T}_{m,n,l}\mathbf{G^*}_{m,n,l}}\mathbf{G}^*_{m,n,l}$
        else if encounter missing data point **then**
          $\hat{g}_{m,n,l} = \mathbf{G}^{\mathsf{T}}_{m,n,l}\mathbf{F}_{m,n,l}$
        end if
      end for
    end if
  end for
end for

NUMERICAL EXAMPLES

2D data interpolation using the f-x SPF

We generated a 2D synthetic model with one linear event and two curve events to evaluate the interpolation ability of the f-x SPF. The curve events bend in opposite directions (Fig. 3a and 3b), which challenges the adaptability of interpolation method. For a missing trace interpolation test (Fig. 3c), we removed 40% of the randomly selected traces, which caused the appearance of aliasing (Fig. 3d). For comparison, we used the 2D Fourier Project Onto Convex Sets (POCS) to recover the missing

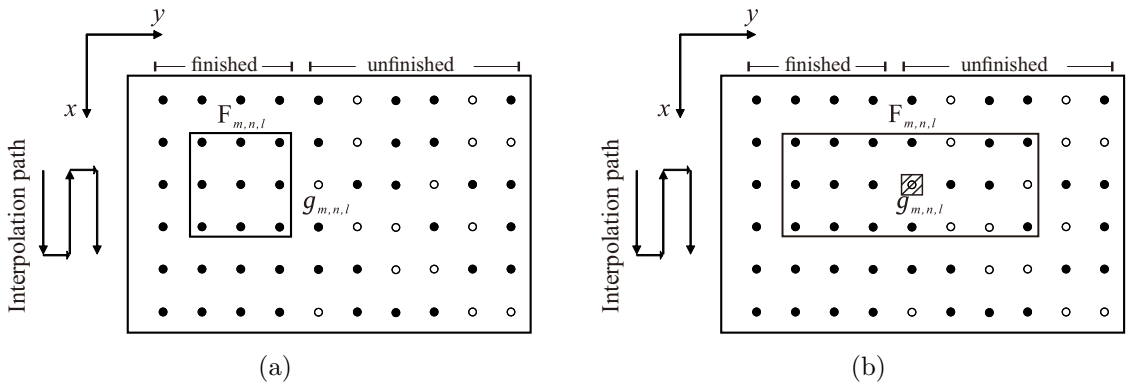


Figure 2: Proposed filter form (a) and space-noncausal filter form (b) in the f-x-y domain. Solid circle denotes known trace, and hollow circle denotes unknown trace.

traces (Fig. 4a). The interpolation result from the 2D Fourier POCS showed that the linear event was interpolated, but many discontinuities were present on the curve events. The interpolated error was slightly larger in the locations missing traces (Fig. 4b), which is caused by strongly variable slopes. The seislet POCS method (Gan et al., 2016) shows a clean interpolation result (Fig. 4c), but it produces interpolation errors where the events intersect (Fig. 4d). We designed the f-x SPF with $\lambda_f = 0.2$, $\lambda_x = 0.5$, and 30 (space) filter coefficients. The proposed method provided successful amplitude preservation, and the missing traces were interpolated reasonably well (Fig. 4e). The difference between the interpolated and original traces showed that most of the redundant fluctuations and artifacts were smaller than those created by the Fourier POCS (Fig. 4f). The F-K spectra (Fig. 5) showed that the f-x SPF recovered the missing data and successfully suppressed aliasing.

Fig. 6a shows a marine shot gather from a deepwater Gulf of Mexico survey with 40% of the data traces randomly removed (Fig. 6c). The F-K spectra (Fig. 6b and 6d) reflect the impact caused by missing data. Fig. 7a shows the interpolated result using the 2D Fourier POCS, which fails to recover steeply dipping events at the far-offset positions (Fig. 7b). In the interpolation result of seislet POCS, discontinuities of seismic events are present (Fig. 7c), and some energy remains in the interpolation error profile (Fig. 7d). We designed the f-x SPF with $\lambda_f = 0.05$, $\lambda_x = 0.1$, and 20 (space) coefficients. Fig. 7e and 7f show the interpolated result using the proposed method and the difference between the interpolated and original traces plotted at the same clip value. The reconstructed data naturally filled the broken events; meanwhile, the steeply dipping events and diffraction events were reasonably interpolated. The F-K spectrum of the interpolation result using the f-x SPF (Fig. 8c) is similar to that of the original data (Fig. 6b); it suppresses the low-frequency interference compared to the seislet POCS (Fig. 8b), and gives a cleaner spectrum than the Fourier POCS (Fig. 8a).

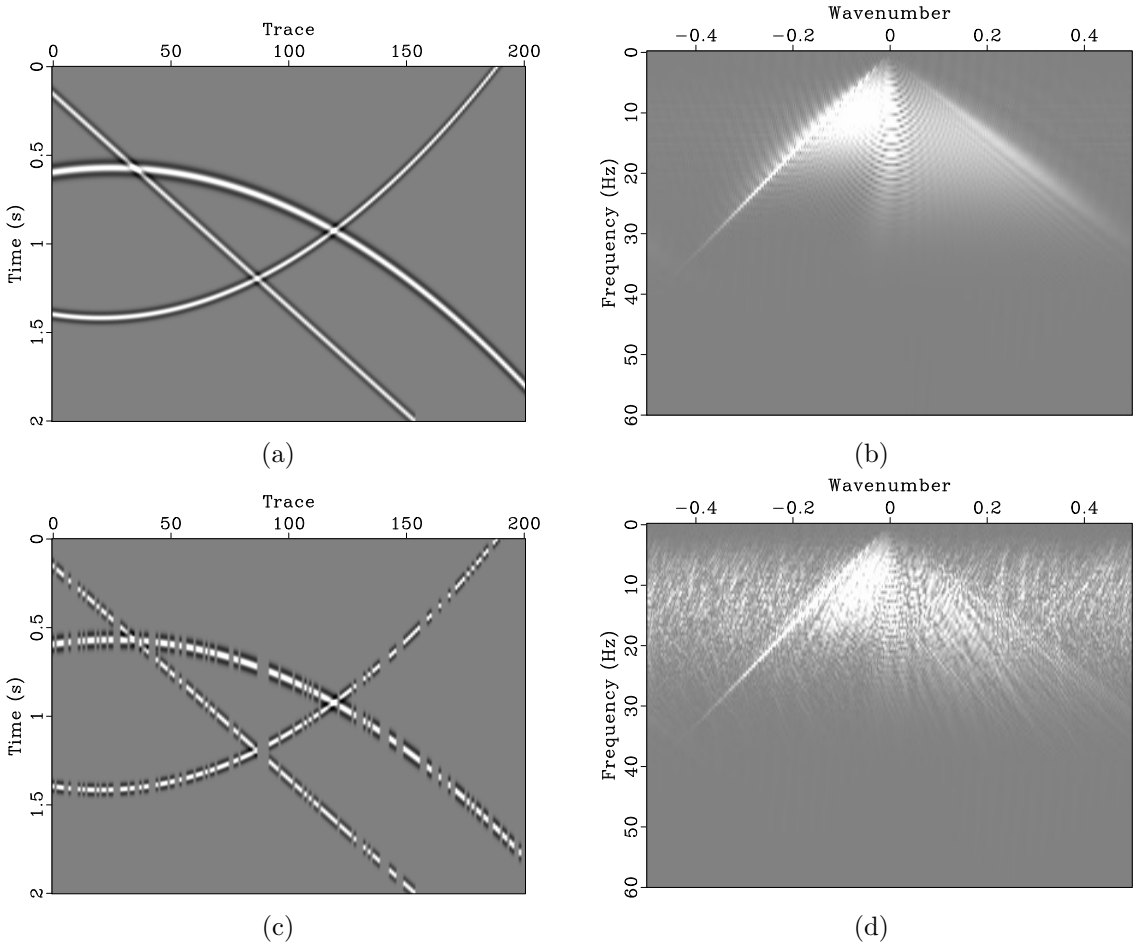


Figure 3: Synthetic model (a) and F-K spectrum (b). Model with 40% of the data traces randomly removed (c) and F-K spectrum (d).

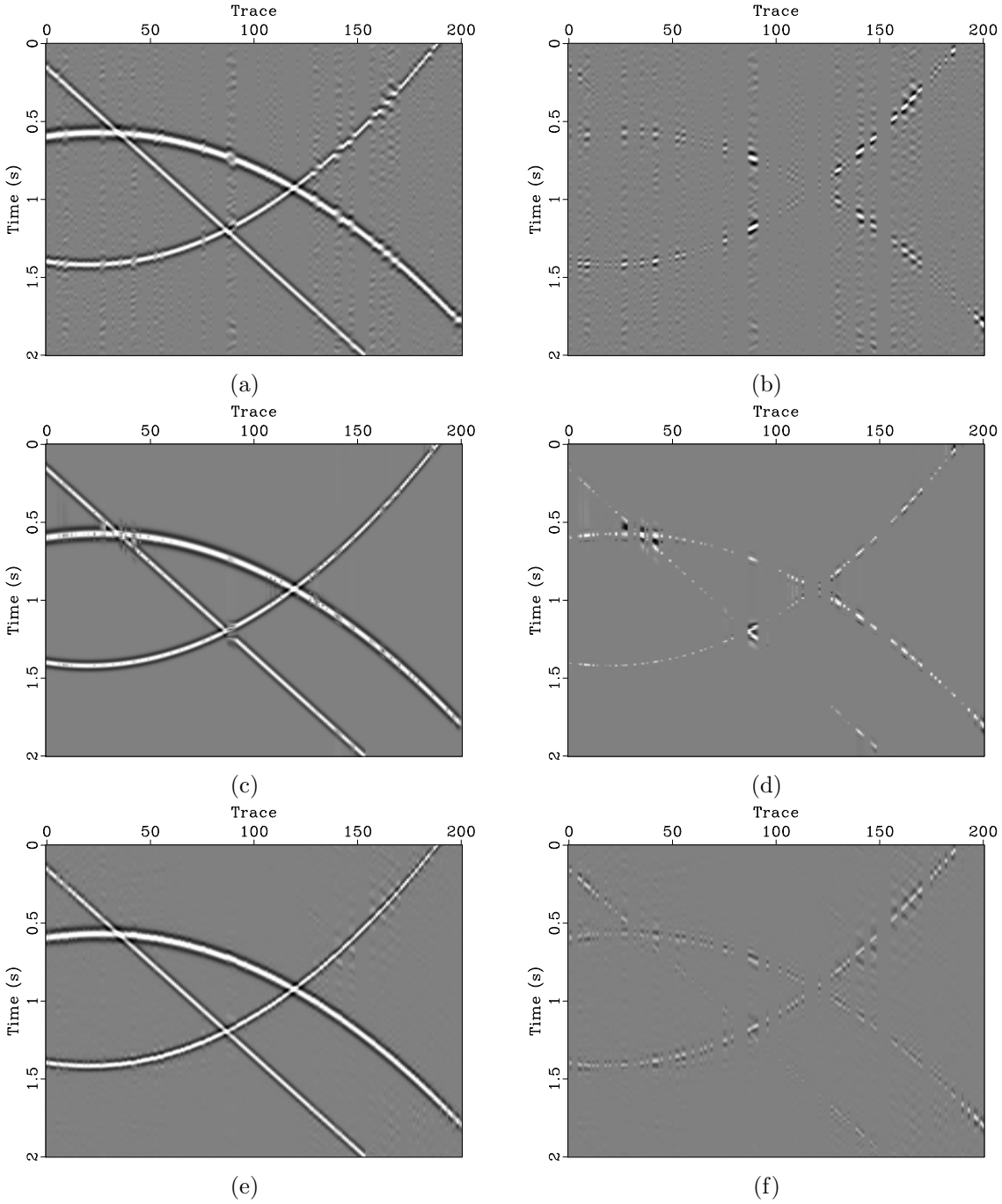


Figure 4: Reconstructed result (a) and interpolation error (b) using the 2D Fourier POCS, reconstructed result (c) and interpolation error (d) using the 2D seislet POCS, reconstructed result (e) and interpolation error (f) using the 2D f-x SPF.

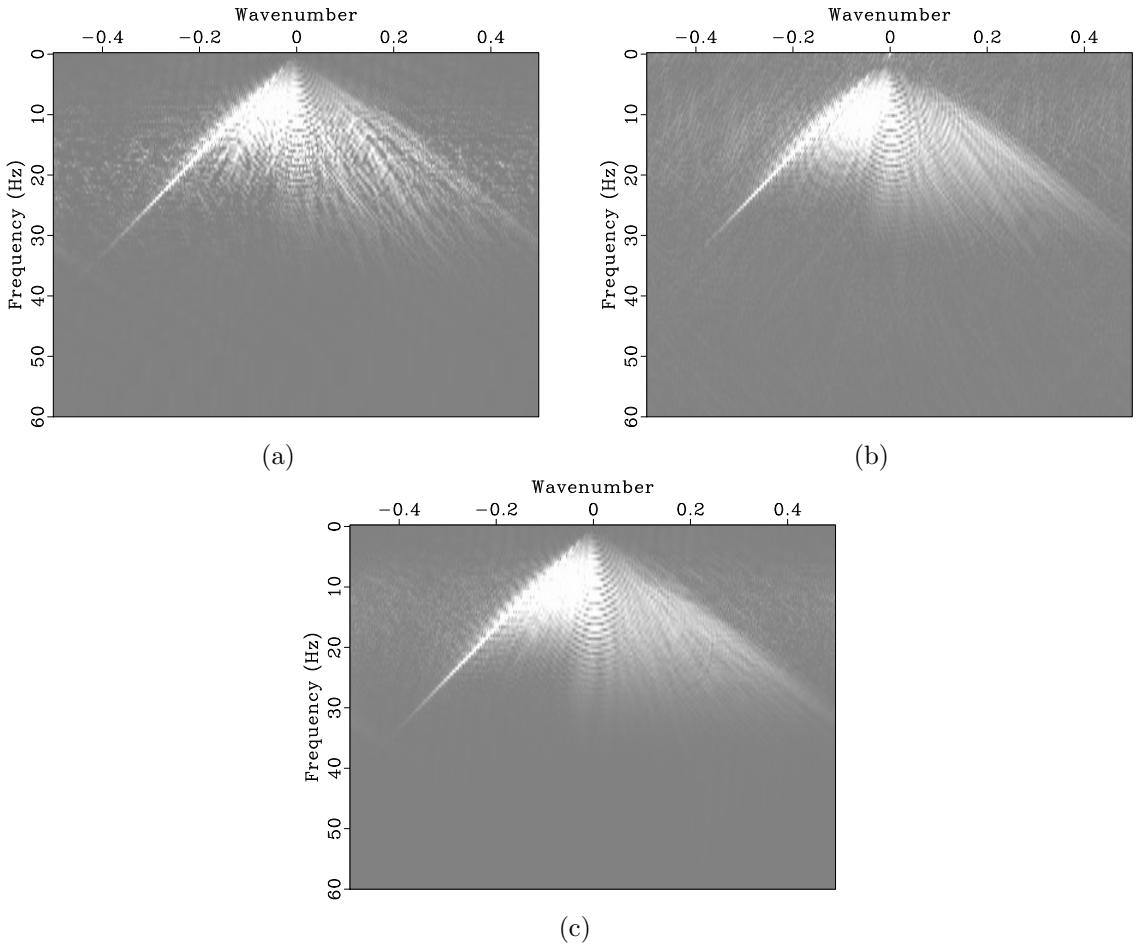


Figure 5: F-K spectrum of the interpolation result using the 2D Fourier POCS (a), F-K spectrum of the interpolation result using the 2D seislet POCS (b), F-K spectrum of the interpolation result using the 2D f-x SPF (c).

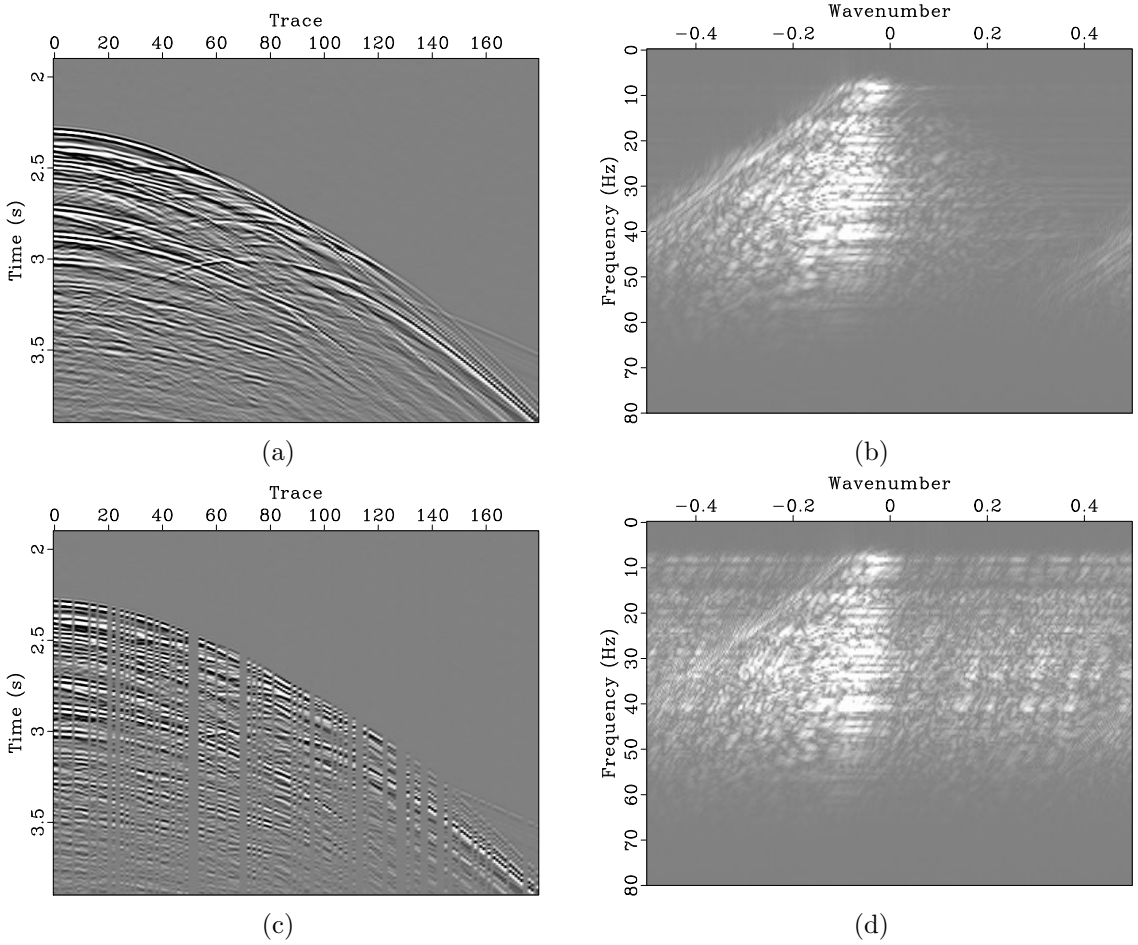


Figure 6: Field data (a) and F-K spectrum (b). Data with 40% of the seismic traces randomly removed (c) and F-K spectrum (d).

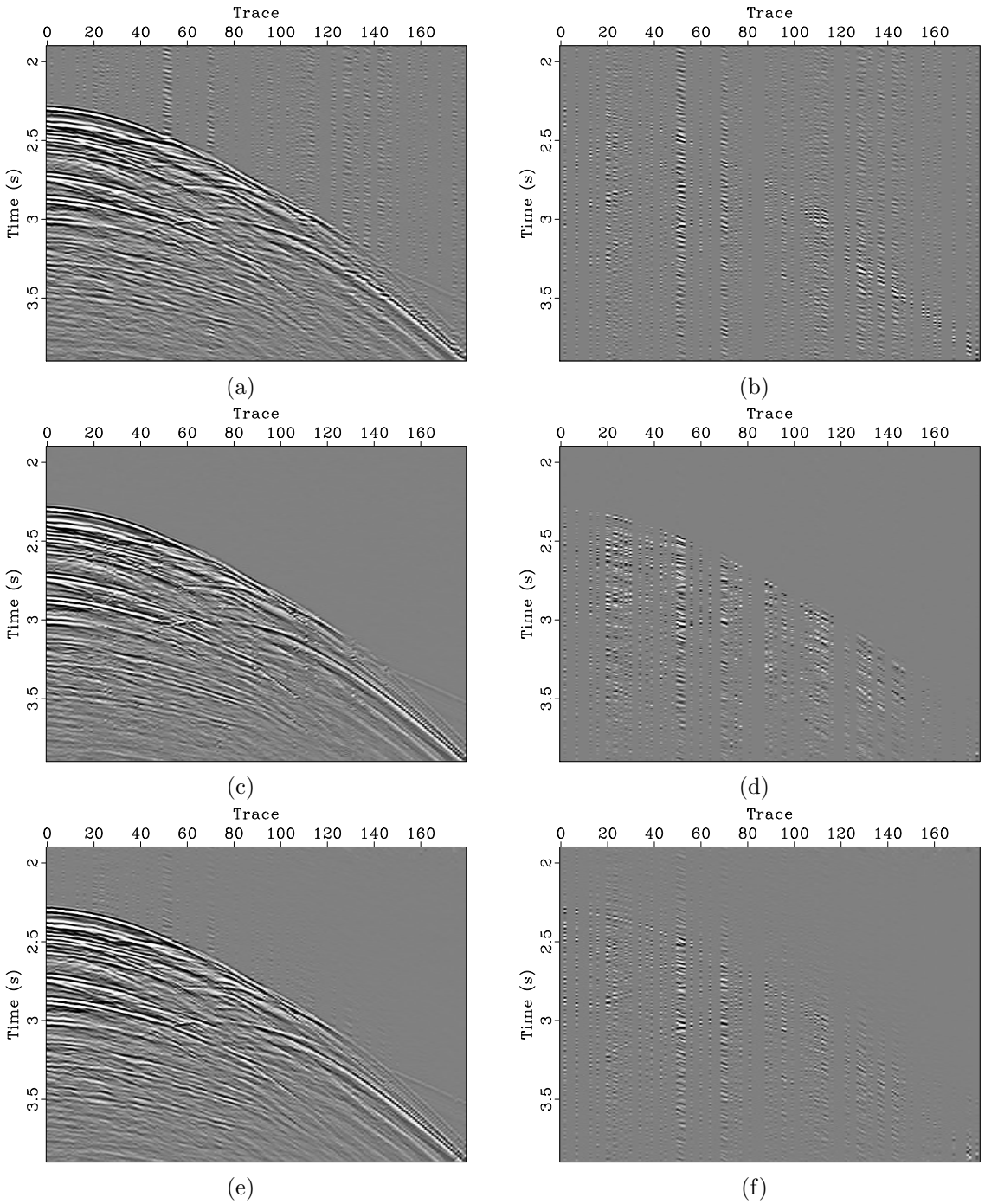


Figure 7: Reconstructed result (a) and interpolation error (b) using the 2D Fourier POCS, reconstructed result (c) and interpolation error (d) using the 2D seislet POCS, reconstructed result (e) and interpolation error (f) using the 2D f-x SPF.

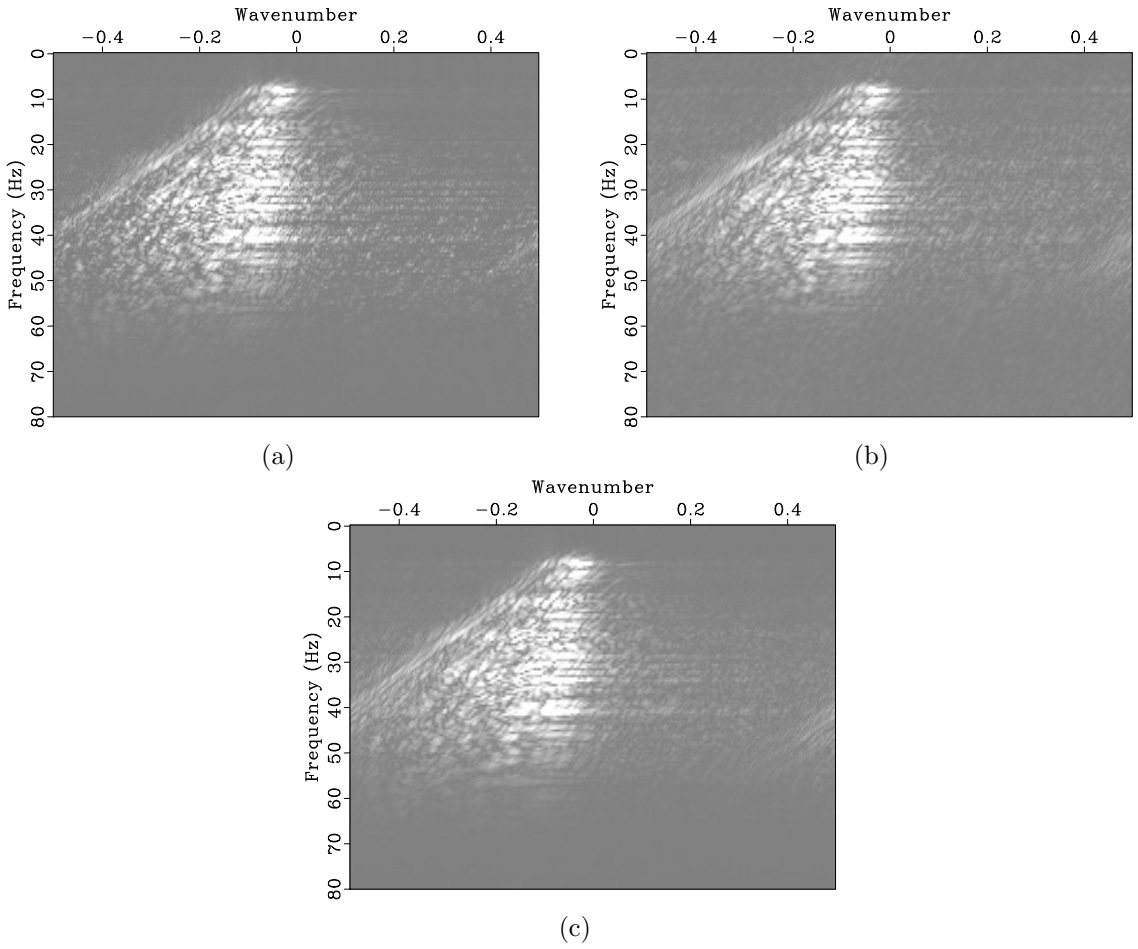


Figure 8: F-K spectrum of the interpolation result using the 2D Fourier POCS (a), F-K spectrum of the interpolation result using the 2D seislet POCS (b), F-K spectrum of the interpolation result using the 2D f-x SPF (c).

3D data interpolation using the f-x-y SPF

For the 3D data interpolation test, we selected a 3D synthetic model (Fig. 9a) containing curve events and faults, and the data cube randomly removed 70% of the seismic traces, where the faults were hard to distinguish (Fig. 9c). The F-K_x-K_y spectrum of the synthetic model and the missing data model are shown in Fig. 9b and 9d, respectively. Because currently seislet transform can only handle 2D datasets, here we evaluated the recovery ability of the f-x-y SPF by comparing with the 2D seislet POCS method and the 3D Fourier POCS method. The parameters of the f-x-y SPF are $\lambda_f = 0.0005$, $\lambda_x = 0.001$, $\lambda_y = 0.0008$, and 11 (space x) $\times$ 6 (space y) filter coefficients. The 2D seislet POCS does not reconstruct the missing data well (Fig. 10a), and generates large interpolation errors (Fig. 10b). The interpolated results (Fig. 10c and 10e) show that both the 3D Fourier POCS and the f-x-y SPF can recover the missing traces even when faults are present. However, the interpolation errors (Fig. 10d and 10f) display that the f-x-y SPF produces less error and preserves the amplitude of the events reasonably better than the 3D Fourier POCS. The comparison of the F-K_x-K_y spectra is shown in Fig. 11, the f-x-y SPF reduces the influence of aliasing. More importantly, the f-x-y SPF significantly reduces the computational cost by avoiding the iterative algorithm especially in higher dimensions. Compared with the Fourier POCS, the proposed methods solve the problem without iterations, which reduces the computational cost (Table 1). Table 2 shows the time consumption of each method, and the computation platform uses 2.0GHz E5-2650 CPU.

To further evaluate the interpolation ability of the proposed methods, we defined the signal-to-noise ratio (SNR) as a measurement:

$$SNR = 10 \log_{10} \left(\frac{\|\mathbf{D}\|_2^2}{\|\mathbf{D} - \hat{\mathbf{D}}\|_2^2} \right), \tag{19}$$

where $\mathbf{D}$ denotes original data, and $\hat{\mathbf{D}}$ denotes interpolation result. The traces in the 3D model (Fig. 9a) have been randomly removed from 5% to 95%. The interpolation results of the f-x-y SPF are shown in Fig. 12. The proposed SPF interpolation method in the frequency domain reconstructs data even under the severely degraded circumstance. Fig. 13 shows that the SPF in higher dimensions can effectively improve the SNR.

We used 3D field common mid-point (CMP) gathers after normal moveout (NMO) to further test the proposed method (Fig. 14a). Fig. 14b shows data binning, where the blank space denotes approximately 75% of the missing traces. A large amount of missing traces lead to the spatial aliasing artifacts as shown in Fig. 18a. The 2D seislet POCS works well with the horizontal events, but the data interpolation result is not reasonable where curve and horizontal events intersect (Fig. 15a). The 3D Fourier POCS can recover most horizontal events, however, some large gaps remain in Fig. 16a. We chose $\lambda_f = 6$, $\lambda_x = 60$, $\lambda_y = 15$, and 101 (space x) $\times$ 5 (space y) filter coefficients for the f-x-y SPF. The design of the filter coefficients is large, which

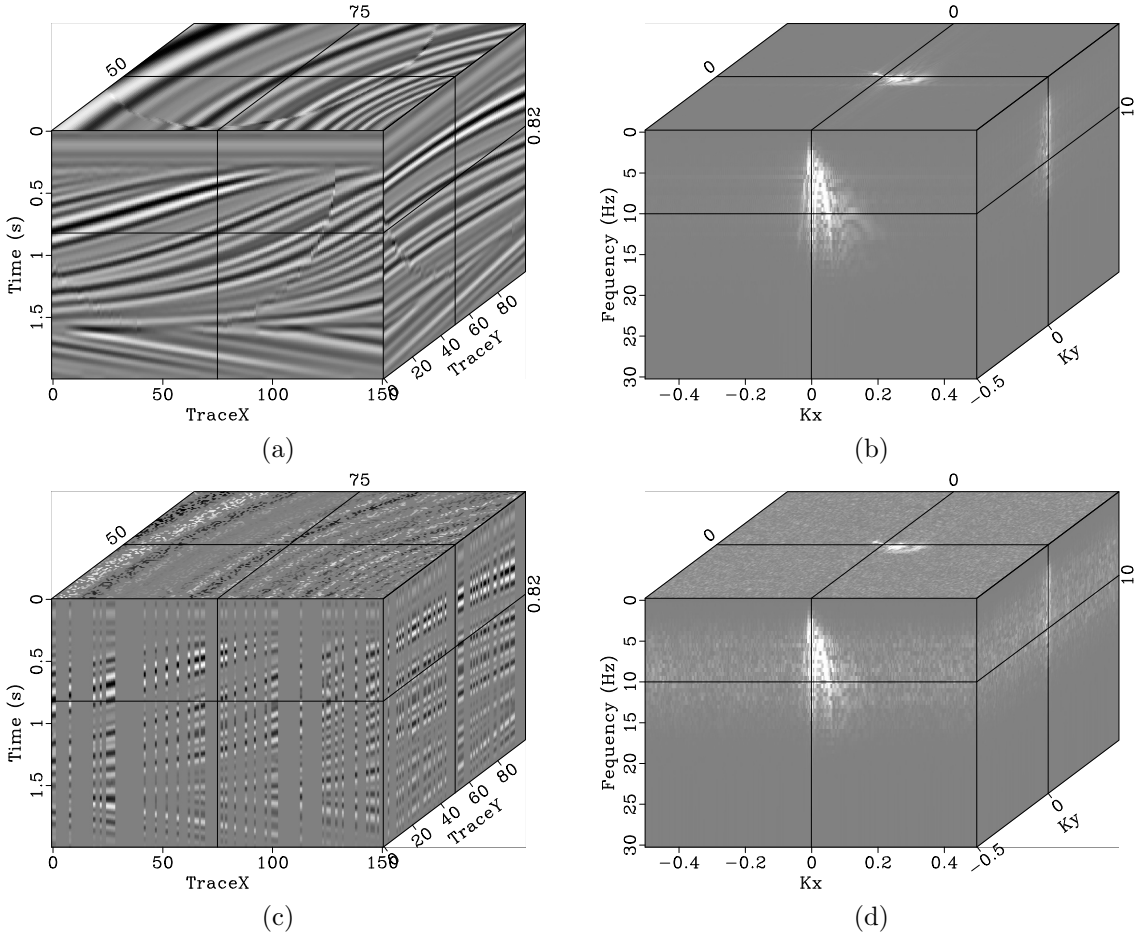


Figure 9: Synthetic 3D model (a) and F-K_x-K_y spectrum (b). Model with 70% of the data traces randomly removed (c) and F-K_x-K_y spectrum (d).

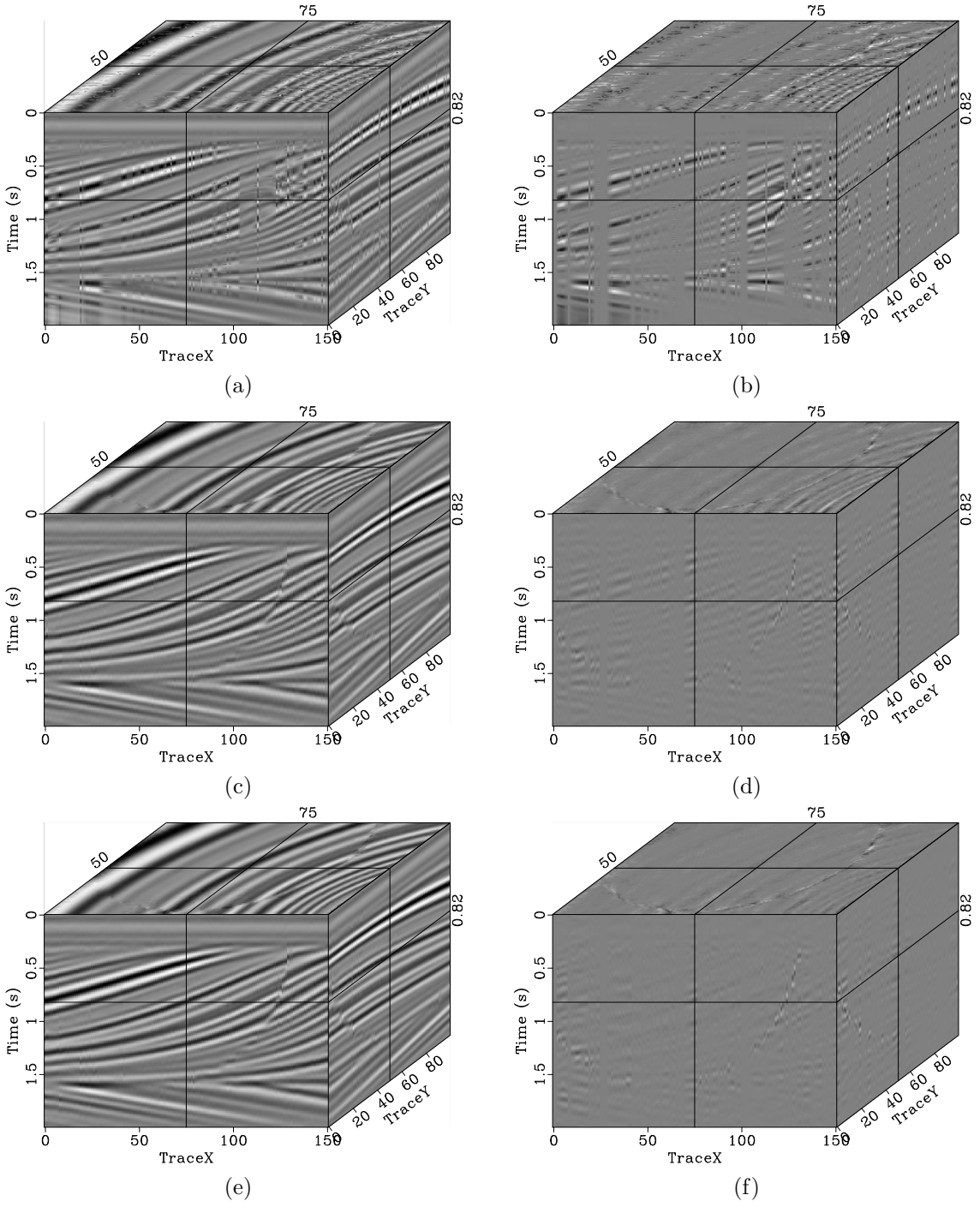


Figure 10: Reconstructed result (a) and interpolation error (b) using the 2D seislet POCS, reconstructed result (c) and interpolation error (d) using the 3D Fourier POCS, reconstructed result (e) and interpolation error (f) using the 3D f-x-y SPF.

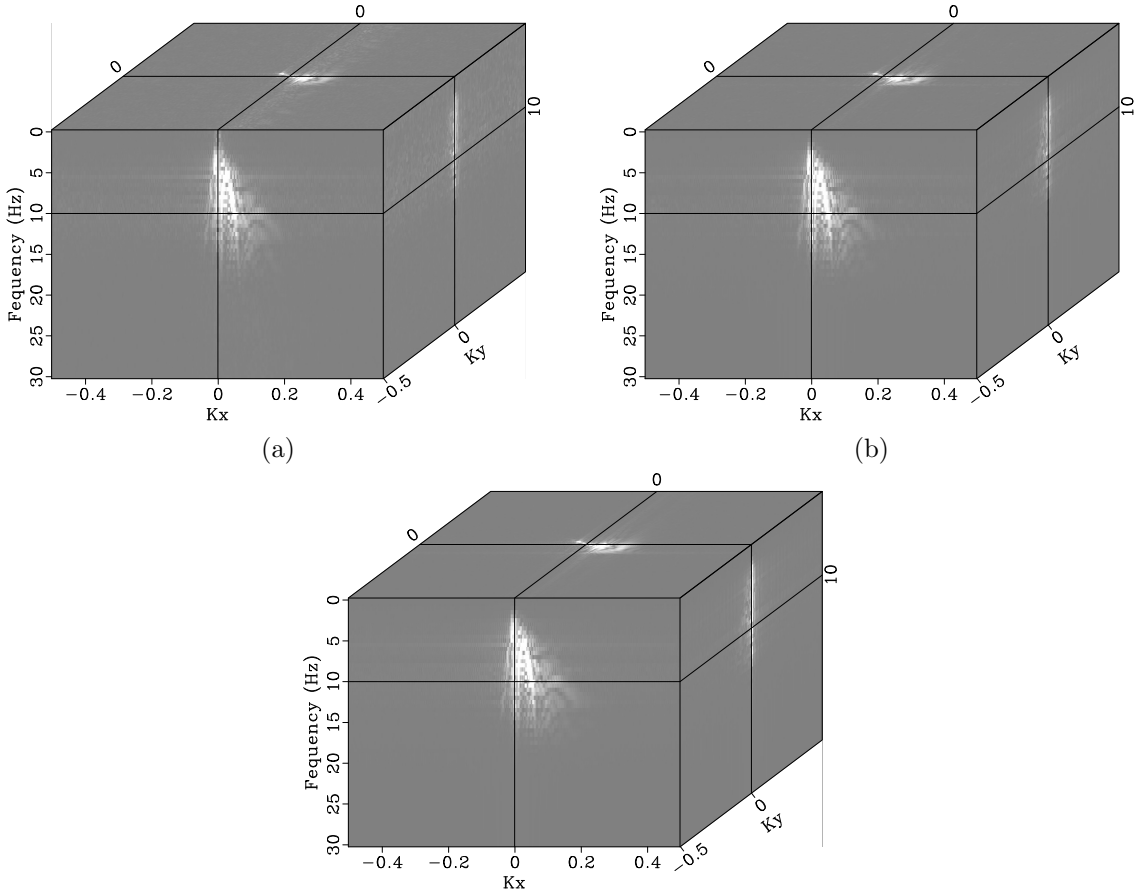


Figure 11: F-K_x-K_y spectrum of the interpolation result using the 2D seislet POCS (a), F-K_x-K_y spectrum of the interpolation result using the 3D Fourier POCS (b), F-K_x-K_y spectrum of the interpolation result using the 3D f-x-y SPF (c).

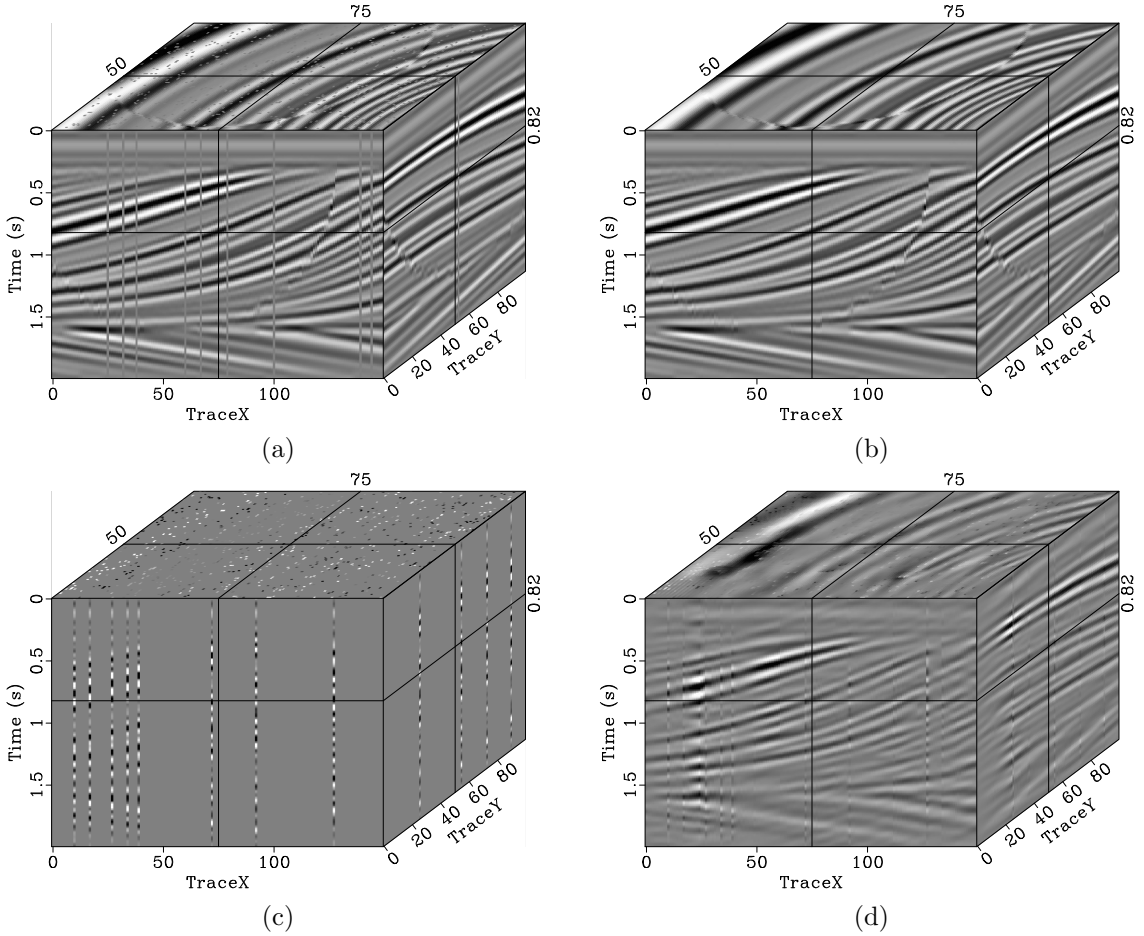


Figure 12: Model with 5% of the data traces randomly removed (a), interpolated data using the 3D f-x-y SPF corresponding to Fig. 12a (b), model with 95% of the data traces randomly removed (c), and interpolated data using the 3D f-x-y SPF corresponding to Fig. 12c (d).

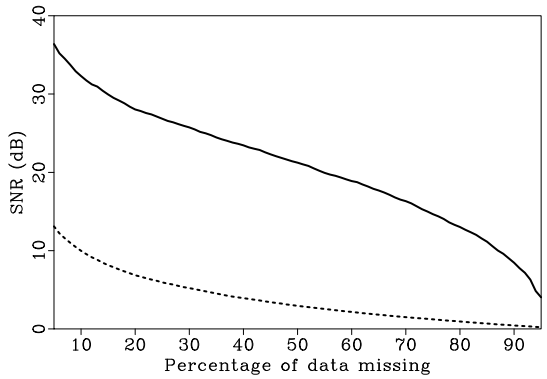


Figure 13: The SNR changing with different degree of trace missing (range from 5% to 95%). Dash line is the SNR of the model with trace missing and solid line is the SNR of interpolation result.

Table 1: Comparison of computational cost: the f-x SPF, the 2D Fourier POCS, the f-x-y SPF, and the 3D Fourier POCS.

	Storage space	Computational time
2D Fourier POCS	$O(N_f N_{k_x})$	$O(N_t N_x \log(N_t N_x) N_{iter})$
f-x SPF	$O(N_a N_x)$	$O(N_a N_f N_x)$
3D Fourier POCS	$O(N_f N_{k_x} N_{k_y})$	$O(N_t N_x N_y \log(N_t N_x N_y) N_{iter})$
f-x-y SPF	$O(N_a N_x N_y)$	$O(N_a N_f N_x N_y)$

N_t is the data size along the time axis, N_f is the data size along the frequency axis, N_x and N_y are the data size along the space x and y axis, respectively. N_{k_x} and N_{k_y} are the data size along the wavenumber k_x and k_y axis, respectively. N_a is the filter size, and N_{iter} is the number of iterations.

Table 2: Comparison of time consumption.

	2D synthetic model	2D field data	3D synthetic model	3D field data
Model size	501×201	500×180	$200 \times 151 \times 100$	$1000 \times 400 \times 16$
seislet POCS	4.23×10^1s	3.89×10^1s	1.81×10^3s	7.41×10^3s
Fourier POCS	3.15×10^0s	3.55×10^0s	1.03×10^2s	4.61×10^2s
SPF	4.00×10^{-1}s	2.68×10^{-1}s	3.38×10^1s	3.63×10^2s

increases the computational time-consumption of the f-x-y SPF. The interpolation result using the 3D f-x-y SPF (Fig. 17a) shows that the missing traces are reasonably reconstructed, the broken events are well recovered, and the continuity of both linear and curve events are interpolated well. A close-up comparison at TraceY=8 (Fig. 15b, 16b, and 17b) shows that the f-x-y SPF can handle more gaps and recover the amplitude of nonstationary events better than the 2D seislet POCS and the 3D Fourier POCS. The F-K_x-K_y spectra (Fig. 18) demonstrate that the f-x-y SPF can handle aliasing, and the energy is converged in the F-K_x-k_y spectrum (Fig. 18d).

DISCUSSION

We first discussed the value selection of the weights λ_f, λ_x, and λ_y. Currently, we considered them as empirical parameters, but there are still some ways to determine their value range. 1) In Eq. (10) and (16), λ^2 and $\mathbf{G}^{\mathsf{T}}_{m,n,l}\mathbf{G}^{*}_{m,n,l}$ (or $\mathbf{G}^{\mathsf{T}}_{m,n}\mathbf{G}^{*}_{m,n}$) are the denominator of the filter update term. It is expected that λ^2 can influence the filter update, so λ^2 and $\mathbf{G}^{\mathsf{T}}_{m,n,l}\mathbf{G}^{*}_{m,n,l}$ (or $\mathbf{G}^{\mathsf{T}}_{m,n}\mathbf{G}^{*}_{m,n}$) should have a similar order of magnitude, where $\lambda^2 = \lambda_f^2 + \lambda_x^2 + \lambda_y^2$ (or $\lambda^2 = \lambda_f^2 + \lambda_x^2$). 2) In terms of the prediction

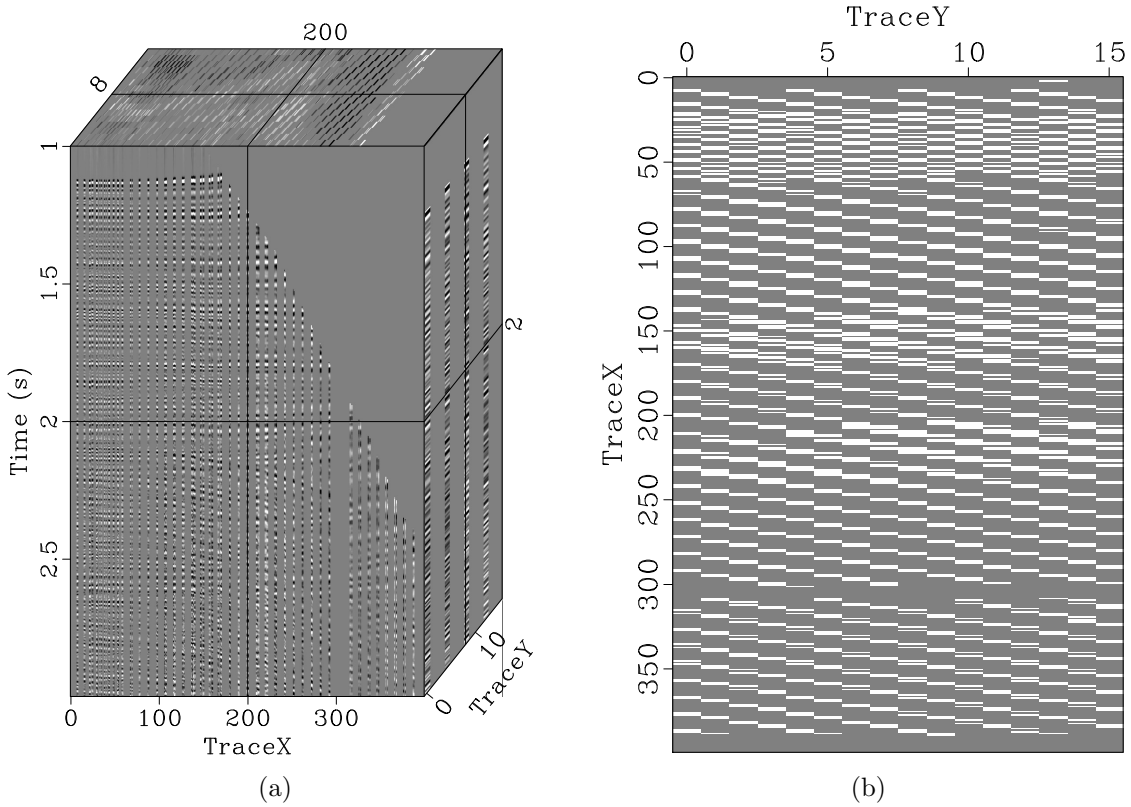
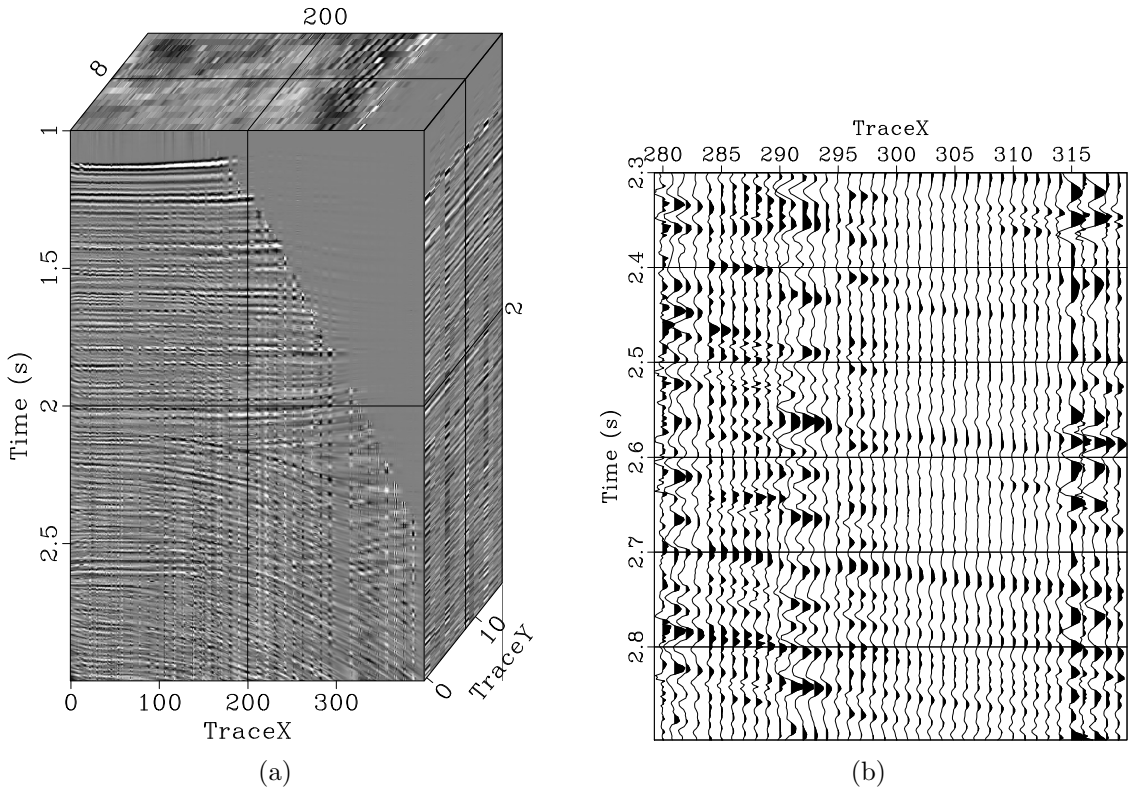
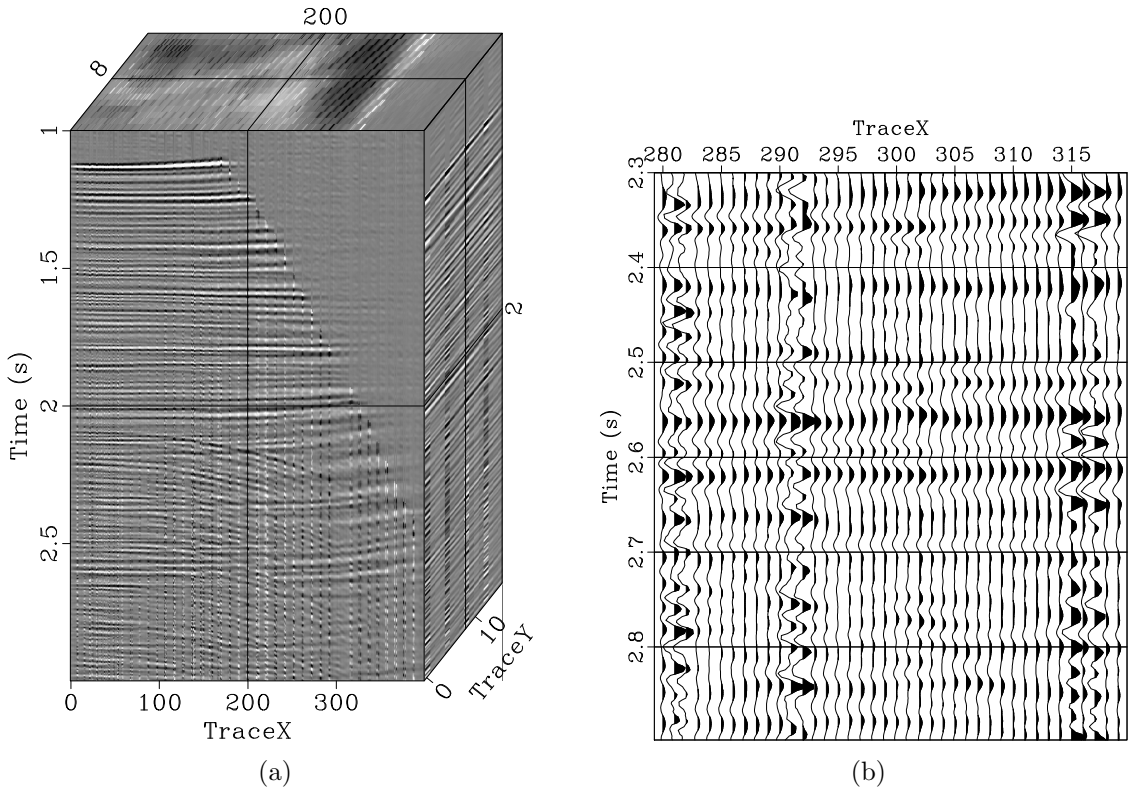


Figure 14: 3D field data with trace missing (a) and data binning of the field data (b).



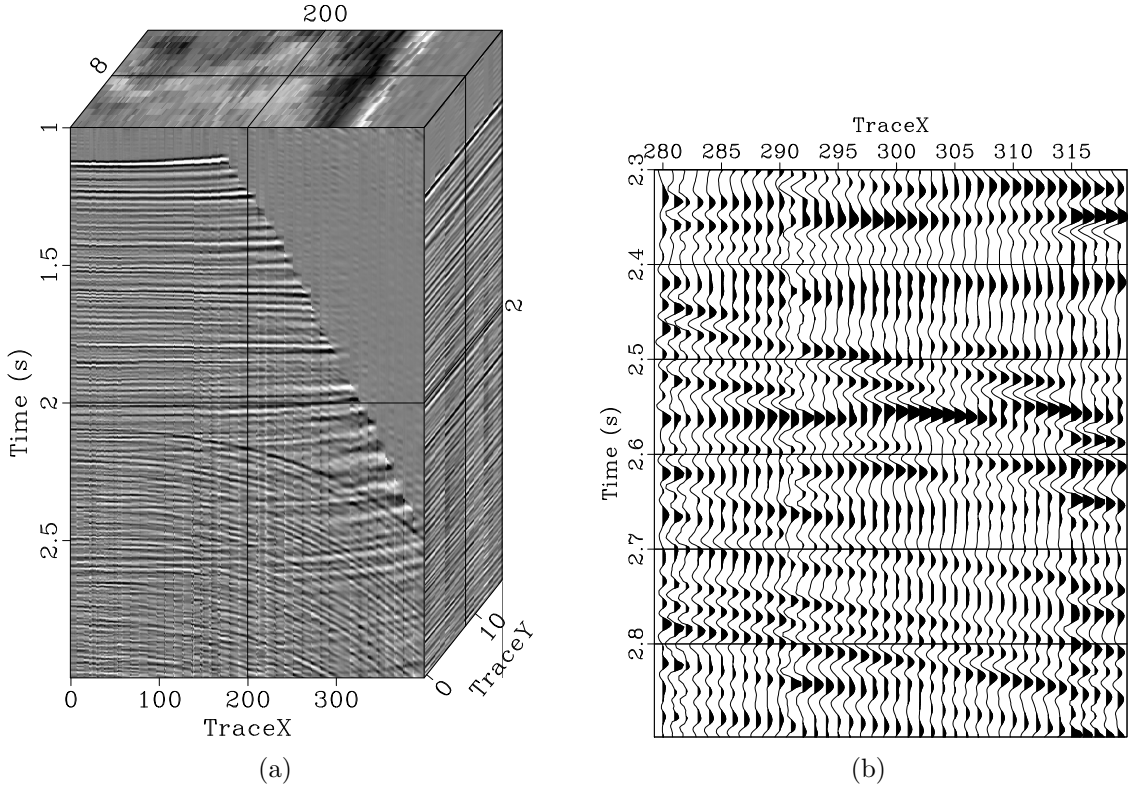
(a) (b)

Figure 15: Reconstructed result using the 2D seislet POCS (a), and close-up of the interpolated result (b).



(a) (b)

Figure 16: Reconstructed result using the 3D Fourier POCS (a), and close-up of the interpolated result (b).



(a) (b)

Figure 17: Reconstructed result using the 3D f-x-y SPF (a), and close-up of the interpolated result (b).

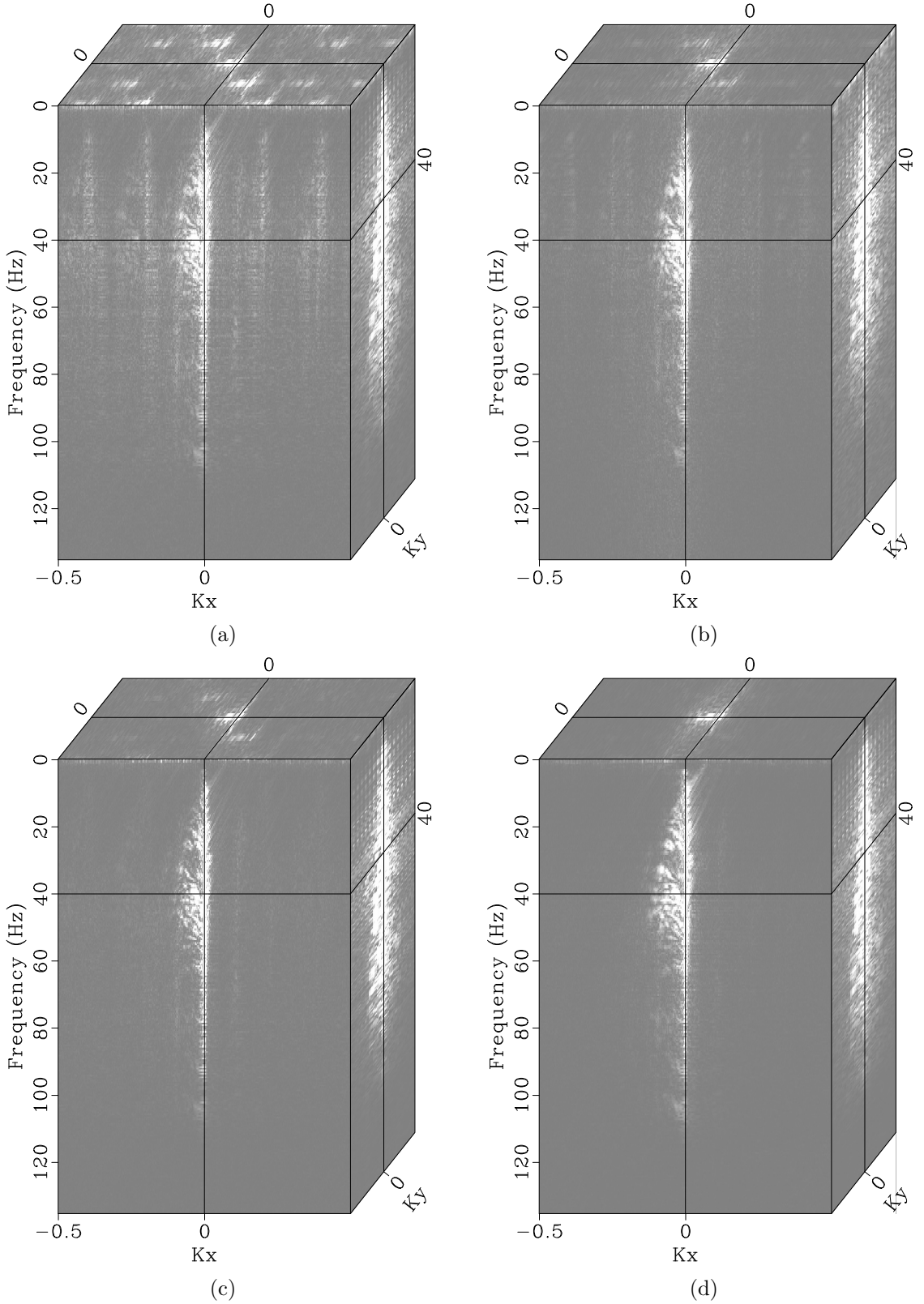


Figure 18: The F-K_x-K_y spectrum of the 3D field data (a), the F-K_x-K_y spectrum of the reconstructed result using the 2D seislet POCS (b), the F-K_x-K_y spectrum of the reconstructed result using the 3D Fourier POCS (c), the F-K_x-K_y spectrum of the reconstructed result using the 3D f-x-y SPF (d).

filtering theory in the frequency domain, the filter predicts the data sample along the spatial direction, not along the frequency direction. λ_x and λ_y in the spatial direction should be similar, and their values can be scaled according to the size of dataset. λ_f mainly stabilizes the filter in the frequency direction, and it generally smaller than λ_x and λ_y. 3) The regularization terms' weight affect the data interpolation result to a certain extent. The values of λ_f, λ_x, and λ_y can be set according to the previous method, and then they can be adjusted according to the interpolation effect.

For the case of regular decimation, we tested the synthetic 2D and 3D model in Fig. 19 and 20. In the 2D case, we used ten seismic traces to initialize the f-x SPF, in which the filter could reconstruct the missing data. As shown in Fig. 19b, the artifacts affect the data interpolation quality for traces with large slope differences between the seismic events. For the 3D data (Fig. 20), the f-x-y SPF can directly recover the missing traces. Although the regular decimation of the 3D model has strong spatial aliasing, we still obtained a reasonable interpolation result.

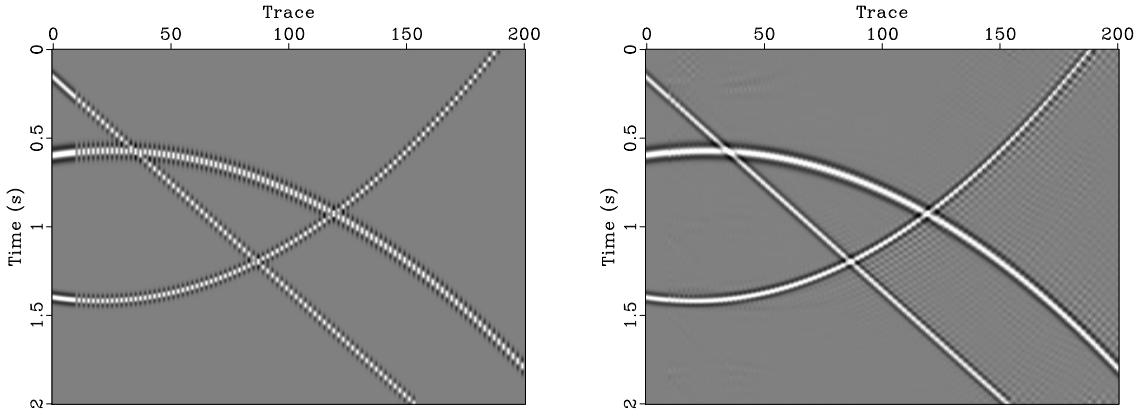


Figure 19: Regular decimation of synthetic 2D model (a), and interpolated result using the 2D f-x SPF (b).

We also tested the effectiveness of the proposed method in the case of low SNR. We added stronger noise to both the 2D and 3D synthetic model (Fig. 21a and 23a), which both had randomly decimated seismic traces. The strong random noise influenced the local slope calculation, which further affected the seislet transform. By using the 2D seislet POCS, the interpolation result of the 2D model (Fig. 22a and 22b) shows some smearing, and the 3D model cannot be reconstructed (Fig. 24a and 24b). For the 2D Fourier POCS method, there are some parts of the upward curve that are not recovered in the 2D model (Fig. 22c and 22d). Additionally, the 3D Fourier POCS method produces a reasonable reconstruction result (Fig. 24c), although leakage signal of the seismic events presents in the interpolation error profile (Fig. 24d). Because the PF can be used to attenuate random noise, and it may reduce the impact of random noise to some extent, our proposed methods (the f-x SPF and the f-x-y SPF) yield better reconstruction results (Fig. 22e, 22f, 24e, and 24f) under low SNR than other methods.

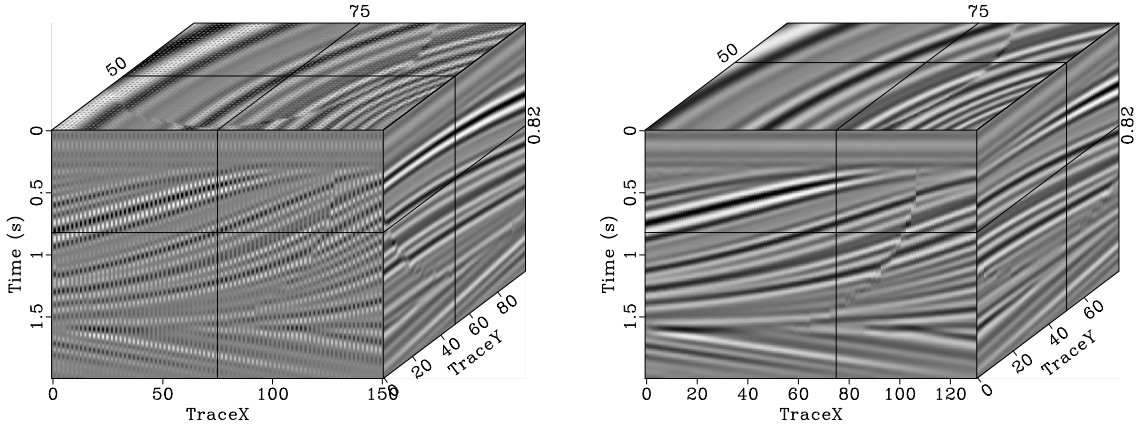


Figure 20: Regular decimation of synthetic 3D model (a), and interpolated result using the 3D f-x-y SPF (b).

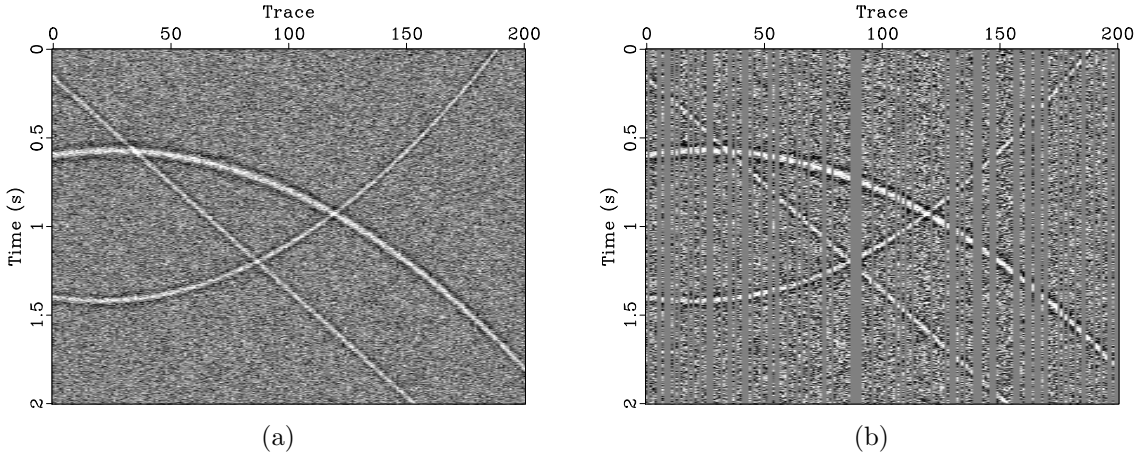


Figure 21: Synthetic model with random noise (a), model with 40% of the data traces randomly removed (b).

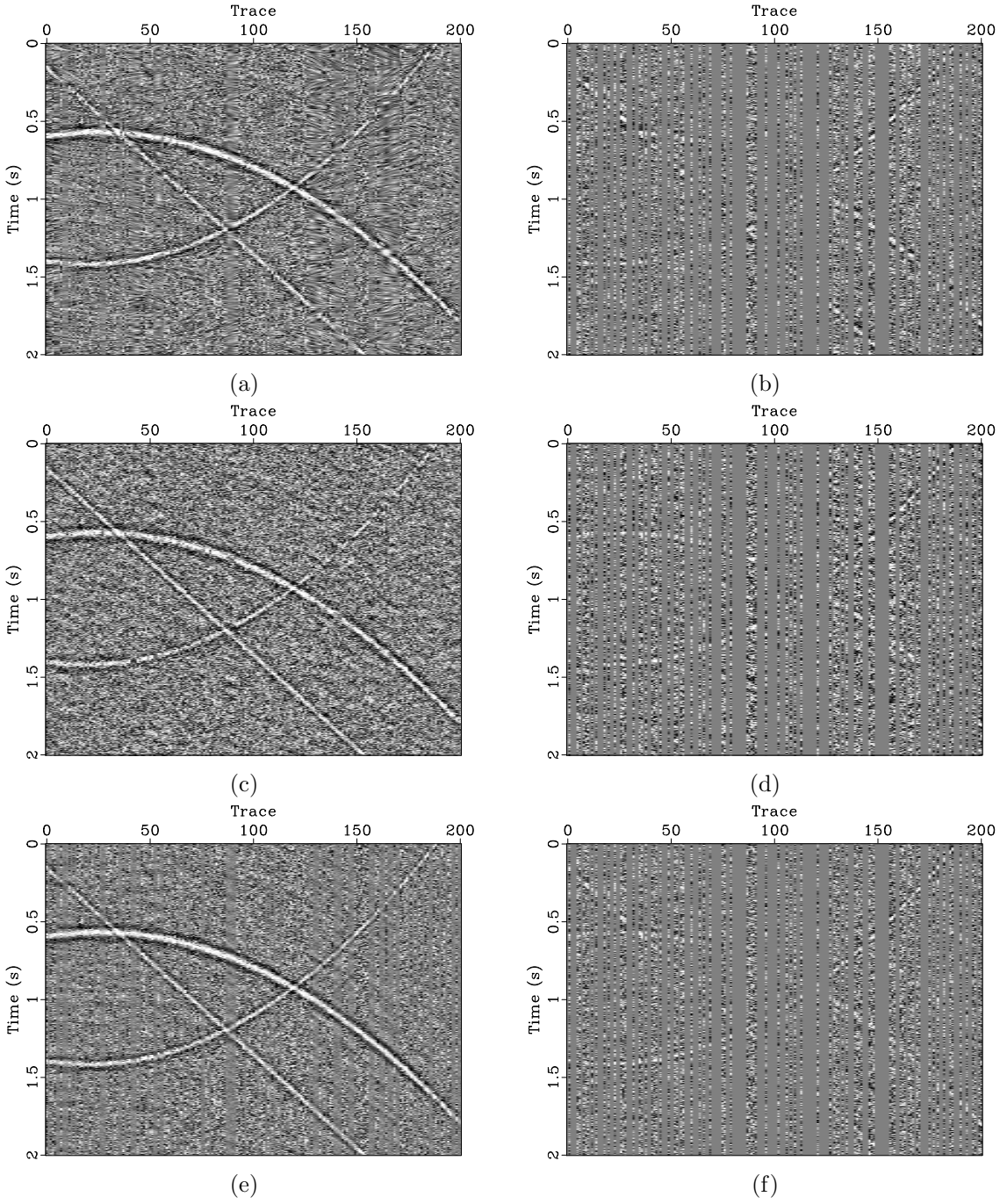


Figure 22: Reconstructed result (a) and interpolation error (b) using the 2D seislet POCS, Reconstructed result (c) and interpolation error (d) using the 2D Fourier POCS, reconstructed result (e) and interpolation error (f) using the 2D f-x SPF.

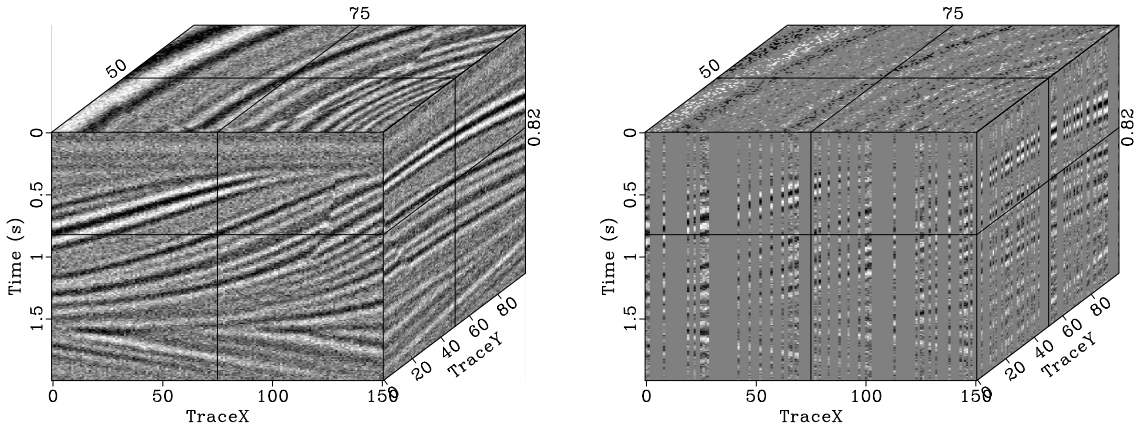


Figure 23: Synthetic 3D model with random noise (a), model with 70% of the data traces randomly removed (b).

CONCLUSION

We introduced a fast approach to adaptive PF for missing data interpolation in the frequency domain. Instead of using the iterative optimization algorithm, we proposed a two-step interpolation strategy with noniterative SPF in the f-x and f-x-y domains. The proposed method employs a local and multidimensional similarity to constrain the autoregression equations for adaptive PFs in the frequency domain, which are based on the streaming computation framework. The SPF in the frequency domain provides a fast and reasonably accurate estimation of nonstationary seismic data. To guarantee the interpolation results, we also designed the filter structure and the processing path according to the characteristics of the interpolation problem. The synthetic and field examples show that the proposed SPF in the frequency domain can depict nonstationary signal variation and provide a reliable description of complex wavefield with low computational cost even when analyzing large-scale seismic data. The properties are suitable for missing data interpolation in practice. Finally, we discussed the problems of parameter selection, interpolation of regularly decimated data, and interpolation of low SNR data; the proposed methods can cope with such problems.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (grant nos. 41974134 and 41774127) and the National key Research and Development Program of China (grant no. 2018YFC0603701).

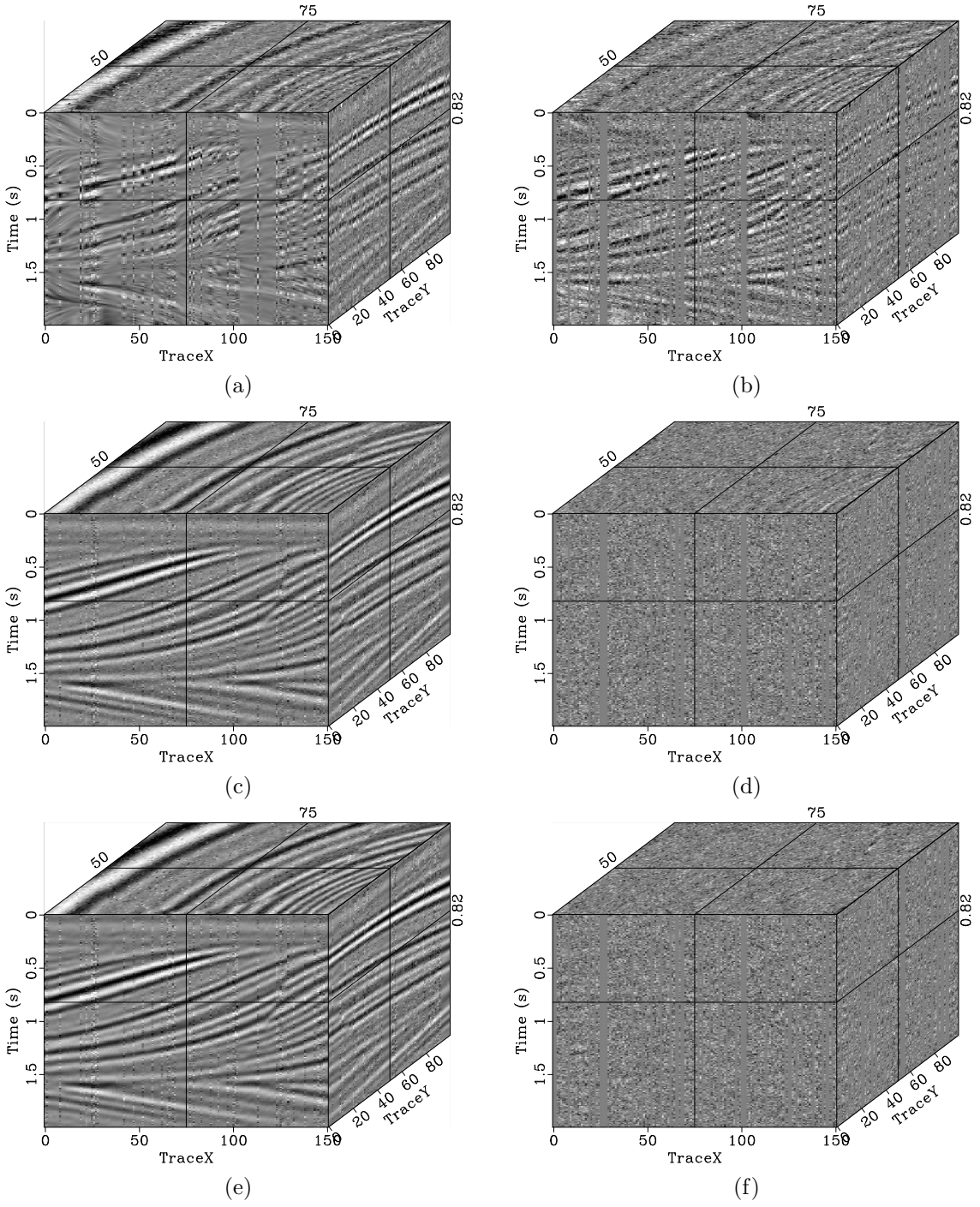


Figure 24: Reconstructed result (a) and interpolation error (b) using the 2D seislet POCS, reconstructed result (c) and interpolation error (d) using the 3D Fourier POCS, reconstructed result (e) and interpolation error (f) using the 3D f-x-y SPF.

REFERENCES

Abma, R., and N. Kabir, 2005, Comparisons of interpolation methods: The Leading Edge, **24**, 984–989.

———, 2006, 3D interpolation of irregular data with a POCS algorithm: Geophysics, **71**, E91–E97.

Bartlett, M. S., 1951, An inverse matrix adjustment arising in discriminant analysis: The Annals of Mathematical Statistics, **22**, 107–111.

Chen, Y., D. Zhang, X. Chen, S. Zu, W. Huang, and S. Gan, 2016, Simultaneous denoising and reconstruction of 5-D seismic data via damped rank-reduction method: Geophysical Journal International, **206**, 1695–1717.

Claerbout, J. F., 1992, 7, *in* Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publications, 164–180.

Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.

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.

Gan, S., S. Wang, Y. Chen, X. Chen, W. Huang, and H. Chen, 2016, Compressive sensing for seismic data reconstruction via fast projection onto convex sets based on seislet transform: Journal of Applied Geophysics, **130**, 194–208.

Gao, J., J. Cheng, and M. D. Sacchi, 2017, Five-dimensional seismic reconstruction using parallel square matrix factorization: IEEE Transactions on Geoscience and Remote Sensing, **55**, 2124–2135.

Gao, J., M. D. Sacchi, and X. Chen, 2013, A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions: Geophysics, **78**, V21–V30.

Gao, S., S. Wang, Y. Chen, Y. Zhang, and Z. Jin, 2015, Dealiased seismic data interpolation using seislet transform with low-frequency constraint: IEEE Geoscience and Remote Sensing Letters, **12**, 2150–2154.

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.

Hellman, K., and S. Boyer, 2016, A local slope approach to multidimensional trace interpolation: Proc. 75th Annu. Int. Meet., SEG, Expanded Abstracts, 4123–4127.

Jia, Y., and J. Ma, 2017, What can machine learning do for seismic data processing? An interpolation application: Geophysics, **82**, V163–V177.

Kaur, H., N. Pham, and S. Fomel, 2019, Seismic data interpolation using CycleGAN: Proc. 78th Annu. Int. Meet., SEG, Expanded Abstracts, 2202–2206.

Li, C., G. Liu, Z. Hao, S. Zu, F. Mi, and X. Chen, 2017, Multidimensional seismic data reconstruction using frequency-domain adaptive prediction-error filter: IEEE Transactions on Geoscience and Remote Sensing, **56**, 2328–2336.

Liu, G., and X. Chen, 2018, Seismic data interpolation using frequency-domain com-

plex nonstationary autoregression: Geophysical Prospecting, **66**, 478–497.

Liu, G., C. Li, Z. Guo, and Y. Rao, 2019, Irregularly sampled seismic data reconstruction using multiscale multidirectional adaptive prediction-error filter: IEEE Transactions on Geoscience and Remote Sensing, **57**, 2909–2919.

Liu, Y., and S. Fomel, 2010, OC-seislet: Seislet transform construction with differential offset continuation: Geophysics, **75**, WB235–WB245.

———, 2011, Seismic data interpolation beyond aliasing using regularized nonstationary autoregression: Geophysics, **76**, V67–V77.

Liu, Y., and B. Li, 2018, Streaming orthogonal prediction filter in t-x domain for random noise attenuation: Geophysics, **83**, F41–F48.

Mandelli, S., V. Lipari, P. Bestagini, and S. Tubaro, 2019, Interpolation and denoising of seismic data using convolutional neural networks: arXiv preprint arXiv:1901.07927.

Naghizadeh, M., and M. D. Sacchi, 2008, f-x adaptive seismic-trace interpolation: Geophysics, **74**, V9–V16.

Oliveira, D. A., R. S. Ferreira, R. Silva, and E. V. Brazil, 2018, Interpolating seismic data with conditional generative adversarial networks: IEEE Geoscience and Remote Sensing Letters, **15**, 1952–1956.

Porsani, M., 1999, Seismic trace interpolation using half-step prediction filters: Geophysics, **64**, 1461–1467.

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.

Sacchi, M. D., and T. Ulrych, 1997, Recovery of near offsets using a FX gap filling algorithm: Proc. 66th Annu. Int. Meet., SEG, Expanded Abstracts, 1096–1099.

Sherman, J., and W. J. Morrison, 1950, Adjustment of an inverse matrix corresponding to a change in one element of a given matrix: The Annals of Mathematical Statistics, **21**, 124–127.

Spitz, S., 1991, Seismic trace interpolation in the FX domain: Geophysics, **56**, 785–794.

Trickett, S., L. Burroughs, A. Milton, and R. Dyck, 2010, Rank-reduction-based trace interpolation: Proc. 80th Annu. Int. Meet., SEG, Expanded Abstracts, 3829–3833.

Wang, S. Q., X. Gao, and Z. X. Yao, 2010a, Accelerating POCS interpolation of 3D irregular seismic data with graphics processing units: Computers & Geosciences, **36**, 1292–1300.

Wang, Y., 2002, Seismic trace interpolation in the f-x-y domain: Geophysics, **67**, 1232–1239.

Wang, Y., S. Dong, and Y. Luo, 2010b, Model-based interferometric interpolation method: Geophysics, **75**, WB211–WB217.

Zhang, H., X. Yang, and J. Ma, 2020, Can learning from natural image denoising be used for seismic data interpolation: Geophysics, **85**, WA115–WA136.